STUDIES IN *FUZZINESS* AND *SOFT COMPUTING*

Studies in Fuzziness and Soft Co

Xiaodong Liu
Witold Pedrycz

Axiomat
Set Theo

Xiaodong Liu and Witold Pedrycz

Axiomatic Fuzzy Set Theory and Its Applications

Studies in Fuzziness and Soft Computing, Volume 244

Editor-in-Chief

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series can be found on our homepage: springer.com

Vol. 228. Adam Kasperski
Discrete Optimization with Interval Data, 2008
ISBN 978-3-540-78483-8

Vol. 229. Sadaaki Miyamoto,
Hidetomo Ichihashi, Katsuhiro Honda
Algorithms for Fuzzy Clustering, 2008
ISBN 978-3-540-78736-5

Vol. 230. Bhanu Prasad (Ed.)
Soft Computing Applications in Business, 2008
ISBN 978-3-540-79004-4

Vol. 231. Michal Baczynski,
Balasubramaniam Jayaram
Fuzzy Implications, 2008
ISBN 978-3-540-69080-1

Vol. 232. Eduardo Massad,
Neli Regina Siqueira Ortega,
Laécio Carvalho de Barros,
Claudio José Struchiner
Fuzzy Logic in Action: Applications in Epidemiology and Beyond, 2008
ISBN 978-3-540-69092-4

Vol. 233. Cengiz Kahraman (Ed.)
Fuzzy Engineering Economics with Applications, 2008
ISBN 978-3-540-70809-4

Vol. 234. Eyal Kolman, Michael Margaliot
Knowledge-Based Neurocomputing: A Fuzzy Logic Approach, 2009
ISBN 978-3-540-88076-9

Vol. 235. Kofi Kissi Dompere
Fuzzy Rationality, 2009
ISBN 978-3-540-88082-0

Vol. 236. Kofi Kissi Dompere
Epistemic Foundations of Fuzziness, 2009
ISBN 978-3-540-88084-4

Vol. 237. Kofi Kissi Dompere
Fuzziness and Approximate Reasoning, 2009
ISBN 978-3-540-88086-8

Vol. 238. Atanu Sengupta, Tapan Kumar Pal
Fuzzy Preference Ordering of Interval Numbers in Decision Problems, 2009
ISBN 978-3-540-89914-3

Vol. 239. Baoding Liu
Theory and Practice of Uncertain Programming, 2009
ISBN 978-3-540-89483-4

vol. 240. Asli Celikyilmaz, I. Burhan Türksen
Modeling Uncertainty with Fuzzy Logic, 2009
ISBN 978-3-540-89923-5

vol. 241. Jacek Kluska
Analytical Methods in Fuzzy Modeling and Control, 2009
ISBN 978-3-540-89926-6

Vol. 242. Yaochu Jin, Lipo Wang
Fuzzy Systems in Bioinformatics and Computational Biology, 2009
ISBN 978-3-540-89967-9

Vol. 243. Rudolf Seising (Ed.)
Views on Fuzzy Sets and Systems from Different Perspectives, 2009
ISBN 978-3-540-93801-9

Vol. 244. Xiaodong Liu and Witold Pedrycz
Axiomatic Fuzzy Set Theory and Its Applications, 2009
ISBN 978-3-642-00401-8

Xiaodong Liu and Witold Pedrycz

Axiomatic Fuzzy Set Theory and Its Applications

Springer

Author

Xiaodong Liu
Research Center of Information and Control
Dalian University of Technology
Dalian 116024
China

and

Department of Mathematics
Dalian Maritime University
Dalian 116026
China

Witold Pedrycz
University of Alberta
Dept. Electrical & Computer Engineering
9107 116 Street
Edmonton AB T6G 2V4
Canada

ISBN 978-3-642-00401-8 e-ISBN 978-3-642-00402-5

DOI 10.1007/978-3-642-00402-5

Studies in Fuzziness and Soft Computing ISSN 1434-9922

Library of Congress Control Number: 2009921143

Typeset & Cover Design: Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed in acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

It is well known that "*fuzziness*"—information granules and fuzzy sets as one of its formal manifestations— is one of important characteristics of human cognition and comprehension of reality. Fuzzy phenomena exist in nature and are encountered quite vividly within human society. The notion of a fuzzy set has been introduced by L. A., Zadeh in 1965 in order to formalize human concepts, in connection with the representation of human natural language and computing with words. Fuzzy sets and fuzzy logic are used for modeling imprecise modes of reasoning that play a pivotal role in the remarkable human abilities to make rational decisions in an environment affected by uncertainty and imprecision. A growing number of applications of fuzzy sets originated from the "empirical-semantic" approach. From this perspective, we were focused on some practical interpretations of fuzzy sets rather than being oriented towards investigations of the underlying mathematical structures of fuzzy sets themselves. For instance, in the context of control theory where fuzzy sets have played an interesting and practically relevant function, the practical facet of fuzzy sets has been stressed quite significantly.

However, fuzzy sets can be sought as an abstract concept with all formal underpinnings stemming from this more formal perspective. In the context of applications, it is worth underlying that membership functions do not convey the same meaning at the operational level when being cast in various contexts. As a consequence, when we look carefully at the literature on fuzzy sets, including Zadeh's own papers, there is no profound uniformity as to the interpretation of what a membership grade stands for. This situation has triggered some critical comments outside the fuzzy set community and has resulted in a great deal of misunderstanding within the field of fuzzy sets itself. Most negative statements expressed in the literature raised the question of interpretation and elicitation membership grades. Thus the questions of the semantics and the empirical foundations as well as the measurement of fuzzy sets remain partially unresolved. As of now, this is perhaps still a somewhat under-developed facet of fuzzy set theory.

Let us begin our discussions from the following example. What is your perception of the height of a person? An NBA basketball player describes that some person is not "*tall*" and a ten year old child describes that the same person is very "*tall*". Because the people the NBA basketball player often meets are different from the people the child meets, i.e., the "data" they observed are drawn from different probability spaces. They may have different interpretations (membership functions or membership grades) for the same linguistic concept "*tall*". Therefore the interpretations of fuzzy sets are strongly dependent on both the semantics of the concepts and the distribution of the observed data. Fuzzy sets call for efficient calibration mechanisms. Thus when we consider the interpretation of a concept, the distributions of the data the concept is applied to must be taken into account when determining the membership functions of the fuzzy concepts. In real world applications, the conventional membership functions are usually provided according to the subjective knowledge or perception of the observer. These membership functions have not accounted the particular distribution of the observed data.

When moving into the age of machine intelligence and automated decision-making, we have to deal with both the subjective imprecision of human perception-based information described in natural language and the objective uncertainty of randomness universally existing in the real world. There is a deep-seated tradition in science of dealing with uncertainty–whatever its form and nature–through the use of probability theory. What we see is that standard probability theory comes with a number of strengths and limitations. To a significant extent, standard probability theory cannot deal with information described in natural language; that is, to put it simply, standard probability theory does not have natural language processing capability. A basic problem with standard probability theory is that it does not address partiality of truth. The principal limitation is that standard probability provides no tools for operating on information that is perception-based and is described in a natural language. This incapability is rooted in the fact that perceptions are intrinsically imprecise, reflecting the bounded ability of sensory organs, and ultimately the brain, to resolve detail and store information. Zadeh has also claimed that "probability must be used in concert with fuzzy logic to enhance its effectiveness. In this perspective, probability theory and fuzzy logic are complementary rather than competitive."

It is this statement that has motivated our proposal and a comprehensive study of Axiomatic Fuzzy Set (AFS) whose aim is to explore how fuzzy set theory and probability theory can be made to work in concert, so that uncertainty of randomness and of imprecision can be treated in a unified and coherent manner. In AFS theory—the studies on how to convert the information in the observed data into fuzzy sets (membership functions), the membership functions and logic operations of fuzzy concepts are determined by both the distribution of raw data and semantics of the fuzzy concepts through the AFS structures a kind of mathematical description of data structures and AFS algebras a kind of semantic methodology of fuzzy concepts. Since the membership functions of fuzzy concepts in AFS theory always emphasizes the data set they apply to and there are such complicated forms of the descriptions and representations for the attributes of the raw data in the real world applications, hence the raw data are regularized to be AFS structures by two axioms. AFS is mainly with respect to AFS structure of the data and AFS approach mainly studies fuzzy concepts, membership functions and fuzzy logic on the AFS structure of the data, in stead of the raw data. AFS theory is a rigorous and unified mathematical theory which is evolved from two axioms of AFS structure and three natural language assumptions of AFS algebra.

An AFS structure is a triple (M, τ, X) which is a special combinatoric system, where X is the universe of discourse, M is a set of some simple (or elementary) concepts on X (e.g., linguistic labels on the features such as "*large*", "*medium*", "*small*") and $\tau : X \times X \rightarrow 2^M$ is a mathematical description of the relationship between the distributions of the original data and semantic interpretations of the simple concepts in M. The AFS algebra is a family of completely distributive lattices generated by the sets such as X and M. A large number of complex fuzzy concepts on X and their logic operations can be expressed by few simple concepts in M via the coherence membership functions of the AFS algebras and the AFS structures.

Since AFS theory was proposed in 1998 (*Journal of Mathematical Analysis and Applications*, *Fuzzy Sets and Systems*), a number of interesting developments in the theory and applications have been reported. For instance, the topological structures of AFS algebra and AFS structure were presented in 1998 (*Journal of Mathematical Analysis and Applications*), some combinatoric properties of AFS structures were introduced in 1999 (*Fuzzy Sets and Systems*). Further algebraic properties of the AFS algebra have been explored in 2004 (*Information Sciences*), the fuzzy clustering analysis based on AFS theory were proposed in 2005 (*IEEE Transactions on Systems, Man and Cybernetics Part B*), the representations and fuzzy logic operations of fuzzy concepts under framework of AFS theory were outlined in 2007 (*Information Sciences*), the relationships between AFS algebra and Formal Concept Analysis (FCA) were demonstrated in 2007 (*Information Sciences*), fuzzy decision trees under the framework of AFS theory were discussed in 2007 (*Applied Soft Computing*), AFS fuzzy clustering analysis was applied to management strategic analysis in 2008 (*European Journal of Operational Research*), concept analysis via rough set and AFS algebra was presented in 2008 (*Information Sciences*), fuzzy classifier designs based on AFS theory were proposed in 2008 (*Journal of Industrial and Management Optimization*), fuzzy rough sets under the framework of AFS theory were discussed in 2008 (*IEEE Transactions on Knowledge and Data Engineering*) ... etc. Last years saw a rapid growth of the development of the AFS theory. We may witness (maybe not always that clearly and profoundly) that AFS approach tends to permeate a number of significant endeavors. The reason is quite straightforward. In a nutshell, AFS approaches has established a bridge connecting the real world problems with many abstract mathematical theories and human natural language interpretations of fuzzy concepts via the AFS structures of the data and AFS algebras of the human natural language. Since, so far, all application algorithms in AFS approaches imitate human cognitive process with a set of objects by some given fuzzy concepts and attributes; hence AFS delivers new approaches to knowledge representations and inference that is essential to the intelligent systems. The theory offers a far more flexible and powerful framework for representing human knowledge and studying the large-scale intelligent systems in real world applications.

While the idea of AFS has been advocated and spelled out in the realm of fuzzy sets, a fundamental formal framework of AFS based on algebra, combinatorics, measure theory and probability theory has been gradually formed. Undoubtedly, AFS theory has systematically established a rigorous mathematical theory to answer the basic question of the measurement for membership functions of fuzzy concepts and set up the foundations of fuzzy sets for its future developments.

The successful applications of AFS theory show that the theory cannot only serve as the mathematical foundation of fuzzy sets, but also is applicable and practically viable to model human concepts and their logic operations. The discovered inherent relationships of AFS theory with formal concept analysis and rough sets provide a great potential to be explored even further.

In this book, the theory and application results of AFS achieved in a more than a decade are put into a systematic, a rigorous, and unified framework. The book is designed to introduce the AFS in both its rigorous mathematical theory and its flexible application methodology. The material of the monograph is structured into three main sections:

1. mathematical fundamentals which introduce some elementary mathematical notations and underlying structural knowledge about the subject matter;
2. rigorous mathematical theory for the readers who are interested in the mathematical facet of the AFS theory;
3. the applications and case studies which are of particular interest to the readers involved in the applications of the theory.

The last two parts can be studied independently to a very high extent as the algorithms (methods) coming directly from the main mathematical results presented in part 2 are clearly discussed and explained though some detailed examples. Thus there should not be any difficulties for the reader who wishes to directly proceed with part 3. We anticipate that the level of detail at which the material is presented makes this book a useful reference for many researchers working in the area of fuzzy sets and their applications.

The first part of this book, which consists of Chapter 1 (Fundamentals) and Chapter 2 (Lattices), is devoted to a detailed overview of the fundamental knowledge which is required for the rigorous exposure of the mathematical material covered in the second part. This part makes the book self-contained to a significant extent. We would like to note though that most of material in this part is not required for the readers who are predominantly interested in the applications of the AFS theory.

The second part of this book, which consists of Chapter 3 (Boolean Matrices and Binary Relation, Chapter 4 (AFS Logic, AFS Structure and Coherence Membership Functions), Chapter 5 (AFS Algebras and Their Representations of Membership Degrees), introduces the rigorous mathematical facet of the AFS theory and develops a suite of theorems and results which can be directly exploited in the studies on real world applications.

The third part of this book, which consists of Chapter 6 (AFS Fuzzy Rough Sets), Chapter 7 (AFS Topology and Its Applications), Chapter 8 (AFS Formal Concept and AFS Fuzzy Formal Concept Analysis), Chapter 9 (AFS Fuzzy Clustering Analysis), and Chapter 10 (AFS Fuzzy Classifiers), covers various applications of the theory results developed in the second part. The chapters in this part are all independent and each chapter focuses on some direct application of some results discussed in the second part of the book.

The studies on the AFS theory (primarily reported in journal publications and conference proceedings) have attracted interest of the community working within the boundaries of the technology of fuzzy sets. The underlying concept of AFS could be of interest to a far broader audience. Having this in mind, there are a number of key objectives of this book:

- ✓ To present a cohesive framework of the AFS by defining its main research objectives and specifying underlying tasks;
- ✓ To discuss individual technologies of AFS in this uniform setting and formalizing the key tasks stemming from AFS theory (that concerns the measurement of fuzzy sets by taking both fuzziness (subjective imprecision) and randomness (objective uncertainty) into account and their applications to the real world problems);
- ✓ To provide the reader with a well-thought and carefully introduced host of algorithmic methods available in AFS framework.

The intent is to produce a highly self - contained volume. The reader is provided with the underlying material on AFS theory as well as exposed to the current developments where it finds the most visible applications. Furthermore the book includes an extensive and annotated bibliography - an indispensable source of information to everybody seriously pursuing research in this rapidly developing area. Chapters come with a number of open-ended problems that might be of interest to a significant sector of the readership. The book exhibits the following features:

- Comprehensive, authoritative and up-to-date publication on AFS theory (a self-contained volume providing coverage of AFS from its mathematical foundations through methodology and algorithms to a representative set of applications).

- Coverage of detailed mathematical proofs: a complete and rigorous mathematical theory based on the underlying axioms of the AFS structure and natural language assumptions of the AFS algebra itself. The theory is oriented to link the findings to essential and well-delineated goals of practical relevance.
- Coverage of detailed algorithms, complete description of underlying experiments, and a thorough comparative analysis illustrated with the aid of complete numeric examples coming from a broad spectrum of problems stemming from information systems.
- Breadth of exposure of the material ranging from fundamental ideas and concepts to detailed, easy to follow examples; the top - down approach fully supports a systematic and in-depth comprehension of the material.
- Self - containment of the material; the book will include all necessary prerequisites so that it will appeal to a broad audience that may be diverse in terms of background and research interests.
- Exercises of different level of complexity following each chapter that help the reader reflect and build upon key conceptual and algorithmic points raised in the text. Such exercises can be of significant help to an instructor offering courses on this subject. The open problems following each chapter provide some further research topics for the readers.
- An extended and fully updated bibliography and a list of WWW resources being an extremely valuable source of information in pursuing further studies.

The audience of this book is diversified. The material could be of interest to researchers and practitioners (primarily engineers, mathematics, computer scientists, managers) interested in fuzzy set theory and applications, graduate students in electrical and computer engineering, software engineering, mathematics, computer science, operations research and management. The material will not only be advantageous to the readership in the area of fuzzy sets but also the readers in the area of mathematics, rough sets, formal concept analysis (FCA) and probabilistic methods.

The book can be used to some extent in graduate courses on intelligent systems, fuzzy sets, data mining, rough sets, formal concept analysis and data analysis. The book can be either viewed as a primary text or a reference material depending upon a way in which the subject matter becomes covered.

Edmonton, Canada
October 2008

Xiaodong Liu
Witold Pedrycz

Acknowledgements

Parts of the research reported in this book were supported by the following foundations:

- National Science Foundations in China under Grant 60174014, 60575039, 60534010, 60875032
- National Key Basic Research Development Program of China under Grant 2002CB312200
- Natural Science Foundation of Liaoning Province under Grant 20062169
- Science Foundation of China Post Doctor under Grant 2005037013
- Canada Research Chair (CRC) Program
- Natural Science and Engineering Research Council of Canada (NSERC)
- Australian Research Council Fellowship scheme.
- Science Foundations of Dalian Maritime University.
- Science Foundations of Dalian University of Technology.

The authors would also like to acknowledge contributions of all the collaborators who got involved in research projects reported in this book.

Contents

Part I: Required Preliminary Mathematical Knowledge

Part I
Required Preliminary Mathematical Knowledge

Chapter 1
Fundamentals

The main objective of this chapter is to introduce some preliminaries regarding essential mathematical notions and mathematical structures that have been commonly encountered in the theory of topological molecular lattices, fuzzy matrices, AFS (Axiomatic Fuzzy Set) structures and AFS algebras. The proofs of some theorems or propositions which are not too difficult to be proved are left to the reader as exercises.

1.1 Sets, Relations and Maps

Set theory provides important foundations of contemporary mathematics. Even if one is not particularly concerned what sets actually *are*, sets and set theory still form a powerful language for reasoning about mathematical objects. The use of the theory has spilled over into a number of related disciplines. In this section, we recall and systematize various standard set-theoretical notations and underlying constructs [6], proceeding with the development of the subject as far as the study of *maps* and *relations* is concerned.

1.1.1 Sets

We view the idea of set as a collection of objects as being a fairly obvious and quite intuitive. For many purposes we want to single out such a collection for attention, and it is convenient to be able to regard it as a single set. Usually we name a set by associating with some meaningful label so that later on we can easily refer to it. The objects which have been collected into the set are then called its *members* , or *elements*, and this relationship of membership is designated by the "*included in*" symbol $\in$. Thus, $a \in X$ is reads as 'a is a member of the set X' or just 'a is in the set X'.

An important point worth stressing here is that everything we can know about a set is provided by being told what members it is composed of. Put it another way, 'two sets are equal if and only if they have the same members'. This is referred to as

X. Liu and W. Pedrycz: Axiomatic Fuzzy Set Theory and Its Applications, STUDFUZZ 244, pp. 3–60.
springerlink.com

the *axiom of extensionality* since it tells us that how far a set 'extends' is determined by what its members are.

The other main principle of set formation is called the axiom of *comprehension*, which states that 'for any property we can form a set containing precisely those objects with the given property'. This is a powerful principle which allows us to form a great variety of sets for all sorts of purposes. Unfortunately, with too full an interpretation of the word 'property' this gives rise to the *Russell's paradox*: Let the property $p(x)$ be $x \notin x$, that is, 'x is not a member of itself' and R be a set containing precisely those objects with the property $p(x)$. This means that for every y, $y \in R$ if and only if $y \notin y$. If this is true for every y, then it must be true when R is substituted for y, so that $R \in R$ if and only if $R \notin R$. Since one of $R \in R$ and $R \notin R$ must hold, we arrive at an obvious contradiction which is the crux of the Russel paradox. This means that the axiom of comprehension must be restricted in some way and for most purposes we encounter there is no difficulty to accomplish that.

There are several quite different ways of specifying sets. In general we enumerate the elements of the set by including those in curly brackets $\{\}$. If the number of elements forming the set is not large, those can be explicitly listed in entirety. For example, $\{7, 8, 9, 10\}$ is a set with four members (i.e., a set containing precisely those objects with the property: 'the natural number between 6 and 11). Often a set could be listed in this manner, but this could be time consuming. One may use a dot notation ... to indicate what the missing elements are, provided that this mechanism is clearly spelled out or is fully understood. Thus the set $\{1, 2, 3, 4,\ldots,1000\}$ is the set of all integers from 1 to 1000 inclusive. In other cases we are not at position to list the entire set, for the reason that it is infinite, so if we are to use this form of listing, dots are essential. For instance, the set $\{2, 3, 5, 7, 11, 13, 17, 19,\ldots\}$ is the set of all prime numbers, which is an infinite set. We are assuming that it is clear which particular set we are intended in on the basis of the elements actually listed.

Since problems often arise when making attempts to list all the members of a set, it is better to use a defining property to specify it where possible. As we hinted, there is a definite ambiguity in defining P by the implied listing above, and the same set $\{2, 3, 5, 7, 11, 13, 17, 19,\ldots\}$ may be better expressed as

$$\{n|n \text{ is a positive prime number}\},$$

which we read as 'the set of all n such that n is a prime number'. Here the specification of the set is totally unambiguous. The use of the notation is justified by the axiom of comprehension: here the property is 'n is a prime number', and the notation tells us to collect together all those numbers fulfilling the property-which by the axiom of comprehension, is a set.

We first need to define formally the following two relationships, which allow us to order and equate sets: we say that the set A is *contained in* the set B or is a *subset of* B (or B *contains* A) and denote this as $A \subseteq B$ (or $B \supseteq A$) if every element x in A is also in B. Symbolically, we can write this statement as $x \in A \Rightarrow x \in B$, where the symbol '$\Rightarrow$' is read as 'implies'. The statement $A = B$ is equivalent to the two statements $A \subseteq B$ and $B \subseteq A$. Symbolically, $A = B \Leftrightarrow A \subseteq B$ and $B \subseteq A$, where symbol '⇔' reads 'if and only if'. If $A \subseteq B$ and $A \neq B$ we write $A \subset B$ and say that A is a *proper subset of* B . Alternatively, we can write $B \supset A$.

Given any two sets A and B, we have the following elementary set operations:

Union: The *union* of A and B, written $A \cup B$, is the set of elements that belong to either A or B or both. Symbolically,

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}.$$

Intersection: The *intersection* of A and B, written $A \cap B$, is the set of elements that belong to both A and B. Symbolically,

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

Complementation: The *complement* of the set A, written A', is the set of all elements that are not in A. Symbolically,

$$A' = \{x \mid x \notin A\}.$$

We begin our discussion with a brief survey of some maps and equivalence relations. Prior to that we list follows.

Empty set: We denote by $\varnothing$ the set without any members. That there is only one such set follows from the axiom of extensionality, in a vacuous way. When we know what members a set has, that set is then completely determined.

Power set: The set of all subsets of a set X is called its *power set*, written 2^X. Symbolically,

$$2^X = \{Y \mid Y \subseteq X\}.$$

If X has n members, then 2^X has 2^n members.

Difference: For two sets X and Y, the *difference set* of X and Y is the set of all members of$X$ which are not in Y, written $X - Y$. Symbolically,

$$X - Y = \{z \mid z \in X \text{ and } z \notin Y\}.$$

Disjointness: Two sets X and Y are said to be *disjoint* if their intersection is empty, i.e., $X \cap Y = \varnothing$.

Partition: Let S be a set and Γ be a collection of non-empty subsets of S. Γ is called a *partition* of S if the union of the sets in Γ is the whole of S and distinct sets in Γ are disjoint. Symbolically,

$$\bigcup_{A \in \Gamma} A = S \text{ and } A \cap B = \varnothing \text{ for } A, B \in \Gamma, A \neq B.$$

Each subset A in the partition Γ is called a *block* of Γ.

Cardinality: The number of members of the set X is denoted by $|X|$. This definition makes sense only if X is finite (that is consists of a finite number of elements).

The elementary set operations can be combined, somewhat akin to the way addition and multiplication operations can be combined. As long as we are careful in doing that, we can treat sets as if they were numbers. We can now state the following useful properties of set operations.

Theorem 1.1. *For any three sets A,B, and C, the following assertions hold:*

1. $A \cup B = B \cup A$, $A \cap B = B \cap A$; *(Commutativity)*
2. $A \cup (B \cup C)$=$(A \cup B) \cup C$, $A \cap (B \cap C)$=$(A \cap B) \cap C$; *(Associativity)*
3. $A \cap (B \cup C)$=$(A \cap B) \cup (A \cap C)$, $A \cup (B \cap C)$=$(A \cup B) \cap (A \cup C)$;*(Distributive Laws)*
4. $(A \cup B)' = A' \cap B'$, $(A \cap B)' = A' \cup B'$. *(DeMorgan's Laws)*

Proof. The proof of most of this theorem is left as a list of exercises. To illustrate the underlying technique, however, we will prove the Distributive Law: $A \cap (B \cap C)$=$(A \cap B) \cap C$.

To prove that two sets are equal, it must be demonstrated that each set contains the other. Formally, then

$$A \cap (B \cup C) = \{x \mid x \in A \text{ and } x \in (B \cap C)\};$$

$$(A \cap B) \cup (A \cap C) = \{x \mid x \in A \cap B \text{ or } x \in A \cap C\}.$$

We first show that $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$. Let $x \in (A \cap (B \cup C))$. By the definition of intersection, it must be that $x \in (B \cup C)$, that is, either $x \in B$ or $x \in C$. Since x also must be in A, we have either $x \in (A \cap B)$ or $x \in (A \cap C)$; therefore,

$$x \in ((A \cap B) \cup (A \cap C)),$$

and the containment is established.

Now assume $x \in ((A \cap B) \cup (A \cap C))$. This implies that $x \in (A \cap B)$ or $x \in (A \cap C)$. If $x \in (A \cap B)$, then x is in both A and B. Since $x \in B$, $x \in (B \cup C)$ and thus $x \in (A \cap (B \cup C))$. If, on the other hand, $x \in (A \cap C)$, the argument is similar, and we again conclude that $x \in (A \cap (B \cup C))$. Thus, we have established $((A \cap B) \cup (A \cap C)) \subseteq A \cap (B \cup C)$, showing containment in the other direction and, hence, proving the Distributive Law. □

The operations of union and intersection can be extended to finite or infinite collections of sets as well. If $A_1, A_2, A_3, \ldots, A_n$ is a collection of sets, then

$$\bigcup_{i=1}^{n} A_i = \bigcup\nolimits_{1 \le i \le n} A_i = \{x \mid x \in A_i \; for\ some\ i = 1,2,...,n\},$$

$$\bigcap_{i=1}^{n} A_i = \bigcap\nolimits_{1 \le i \le n} A_i = \{x \mid x \in A_i \; for\ all\ i = 1,2,...,n\}.$$

If I is an index set (a set of elements to be used as indices) and A_i, $i \in I$, is a collection of sets, then the operations of the union and intersection of the sets are as follows:

$$\bigcup\nolimits_{i \in I} A_i = \{x \mid x \in A_i \; for\ some\ i \in I\},$$

$$\bigcap\nolimits_{i \in I} A_i = \{x \mid x \in A_i \; for\ all\ i \in I\}.$$

The index set I can be any finite, infinite, or uncountable set. For instance, let I=$\{1, 2, \ldots, n\}$. Then

$$\bigcup_{i=1}^{n} A_i = \bigcup_{i\in\{1,2,\ldots,n\}} A_i, \quad \bigcap_{i=1}^{n} A_i = \bigcap_{i\in\{1,2,\ldots,n\}} A_i.$$

If the index set I=$\{A_1, A_2, \ldots, A_n\}$, then

$$\bigcup_{i=1}^{n} A_i = \bigcup_{i\in I} A_i, \quad \bigcap_{i=1}^{n} A_i = \bigcap_{i\in I} A_i.$$

For example, we take the index set Γ=$\{$all positive real numbers$\}$ and $A_a = (0,a) = \{x \mid 0 < x \leq a\}$, then $\bigcup_{a\in\Gamma} A_a = (0,\infty)$ is an uncountable union. For any collection of sets Γ and the set B, the distributive laws carry over to arbitrary intersections and unions:

$$B \cap (\bigcup_{A\in\Gamma} A) = \bigcup_{A\in\Gamma} (B \cap A), \quad B \cup (\bigcap_{A\in\Gamma} A) = \bigcap_{A\in\Gamma} (B \cup A)$$

The proofs of them are left as exercise.

Let A and B be two sets. The *Cartesian product* set of A and B is the set of pairs $(a,b), a\in A, b\in B$, written $A\times B$. Symbolically, $A\times B = \{(a,b) \mid a\in A, b\in B\}$.

The sets A and B need not be distinct. In the set $A\times B$, the elements (a, b) and (c, d) are regarded as equal if and only if $a = c$ and $b = d$. It is important to extend the notion of Cartesian product of two sets to the product of any finite number of sets. If $S_1, S_2, \ldots, S_r$ are any sets, then $\prod_{i=1}^{\gamma} S_i$ or $S_1 \times S_2 \times \ldots \times S_r$ is defined to be the set of r-tuples $(s_1, s_2, \ldots, s_r)$ where the ith component $s_i \in S_i$. Equality is defined by

$$(s_1, s_2, \ldots, s_r) = (e_1, e_2, \ldots, e_r)$$

if and only if $s_i = e_i$ for every i. If all the $S_i = S$ then we write S^r for $\prod_{1\leq i\leq r} S_i$.

The *disjoint union* of I and J is a kind of union of I and J in which every element in I and every element in J are always regarded as different elements,denoted as $I \sqcup J$. That is, $I \sqcup J = (I\times\{1\}) \cup (J\times\{2\})$.

1.1.2 Relations

We say that a binary relation is defined on a set S, if given any ordered pair (a, b) of elements of S, we can determine whether or not a is in the given relation to b. For example, we have the relation of order " $\leq$ " in the set of real numbers. Given two real numbers a and b, presumably we can determine whether or not $a \leq b$. Another order relation is the lexicographic ordering of words, which determines their position in a dictionary. Still another example of a relation is the first-cousin relation among people (a and b have a common grandparent). To abstract the essential element from these situations and similar ones, we are led to define in a formal way a *binary relation* R on a set S to be simply any subset of the product set $S\times S$.

Definition 1.1. Let S be a set. R is called a *binary relation* on S if $R \subseteq S \times S$. For $a,b \in S$, if $(a,b) \in R$, then we say that "a is in the relation R to b" or "a and b have the relation R"; otherwise, "a is not in the relation R to b" or "a and b have not the relation R".

Definition 1.2. Let S be a set. A binary relation R on S is called an *equivalence relation* if the following conditions hold for any a,b,c in S:

1. $(a,a) \in R$ (reflexive property)
2. $(a,b) \in R \Rightarrow (b,a) \in R$ (symmetry)
3. $(a,b) \in R$ and $(b,c) \in R \Rightarrow (a,c) \in R$ (transitivity)

An example of an equivalence relation is obtained by letting S to be the set of points in the plane and defining $(a,b) \in R$ if a and b lie on the same horizontal line for $a,b \in S$. Another example of an equivalence relation R on the same S is obtained by stipulating that $(a,b) \in R$ is a and b are equidistant from the origin O.

Theorem 1.2. *An equivalence relation R on the set S determines a partition Γ of S and a partition Γ on S determines an equivalence relation R on S.*

Proof. First, suppose the equivalence relation R is given. If $a \in S$ we let

$$R(a) = \{b \in S | (a,b) \in R\}.$$

We call $R(a)$ the *equivalence class relative to R determined by a*. Since $(a,a) \in R$, $a \in R(a)$, hence every element of S is contained in an equivalence class and so $\bigcup_{a \in S} R(a) = S$. We note next that $R(a) = R(b)$ if and only if $(a,b) \in R$. First, let $(a,b) \in R$ and let $c \in R(a)$. Then $(c,a) \in R$ and so, by condition 3, $(c,b) \in R$. Then $c \in R(a)$. Then $R(a) \subseteq R(b)$. Also, by condition 2, $(b,a) \in R$ and so $R(b) \subseteq R(a)$. Hence $R(a) = R(b)$. Conversely, suppose $R(a) = R(b)$. Since $a \in R(a) = R(b)$ we see that $(a,b) \in R$, by the definition of $R(b)$. Now suppose $R(a)$ and $R(b)$ are not disjoint and let $c \in R(a) \cap R(b)$. Then $(c,a) \in R$ and $(c,b) \in R$. Hence $R(a) = R(b) = R(c)$. We therefore see that distinct sets in the set of equivalence classes are disjoint. Hence $\{R(a) | a \in S\}$ is a partition of S.

Conversely, let Γ be any partition of the set S. Then, if $a \in S$, a is contained in one and only one $A \in \Gamma$. We define a relation R_Γ by specifying that $(a, b) \in R_\Gamma$ if and only if a and b are contained in the same $A \in \Gamma$. Clearly this relation is reflexive, symmetric, and transitive. Hence R_Γ is an equivalence relation. It is clear also that the equivalence class $R(a)$ of a relative to R_Γ is the subset A in the partition Γ containing a. Hence the partition associated with R_Γ is the given Γ. It is equally clear that if R is a given equivalence relation and the partition Λ=$R(a) | a \in S\}$, then the equivalence relation R_Λ in which elements are equivalent if and only they are contained in the same $R(a)$ is the given relation R. □

If R is an equivalence relation on S, the associated partition $\Gamma = \{R(a) | a \in S\}$ is called the *quotient set of S relative to the relation R*, written S/R . We emphasize again that S/R is no a subset of S. In the quotient set S/R, each equivalence class $R(a)$ is an element of S/R, $R(a)$ and $R(b)$ are the same element if $(a, b) \in R$.

Example 1.1. Let Z be the set of all integers. On the Cartesian product set (or briefly product set) $Z \times (Z - \{0\})$, the binary relation R is defined as follows: for (a,b), $(c,d) \in Z \times (Z - \{0\})$,

$((a,b),(c,d)) \in R \Leftrightarrow ad = bc$, where ad and bd are the multiplication of the numbers.

It is obvious that $((a,b),(a,b)) \in R$ and $((a,b),(c,d)) \in R \Rightarrow ((c,d),(a,b)) \in R$. If $((a,b),(c,d)) \in R$, $((c,d),(e,f)) \in R$, then $ad = bc, cf = de$. We have $adf = bcf = bde$. Since $d \neq 0$, hence $af = be$ and $((a,b),(e,f)) \in R$. Therefore R is an equivalence relation on $Z \times (Z - \{0\})$. If for each $(a,b) \in Z \times (Z - \{0\})$, (a,b) represents a rational number a/b, then any (c,d) in the equivalence class relative to R determined by (a,b) represents the same rational number a/b and $(Z \times (Z - \{0\}))/R$, the quotient set of $Z \times (Z - \{0\})$ relative to the relation R is the set of all rational numbers.

1.1.3 Maps

Let A, B be two sets given, a *map* of A into B is a correspondence rule φ such that $\forall a \in A$, there exits a $a' \in B$ that a' corresponds to a. a' is called the *image of a under* φ, denoted by $\varphi(a)$; a is called the *inverse image* of a'. The set A is called *domain* of φ and the set B *co-domain* (*range*) of φ. Usually, the above facts are denoted by

$$\varphi : A \longrightarrow B,\ a \longmapsto a' = \varphi(a)$$

If S is a subset of A, then we write $\varphi(S)=\{\varphi(a) \mid a \in S\}$ and call this the *image of S under* φ. In particular, we have $\varphi(A)$, which is called the *image* (or *range*) *of the map*. We will denote this also as Imφ. If $T \subseteq B$, the subset

$$\varphi^{-1}(T) = \{x | x \in A \text{ and } \varphi(x) \in T\}$$

of A is called *completely inverse image of T under* φ. For $t \in B$,

$$\varphi^{-1}(t) = \{x \mid x \in A \textit{ and } \varphi(x) = \varphi(t)\}.$$

It is clear that$\varphi^{-1}(T)=\bigcup_{t \in T} \varphi^{-1}(t)$.

If A_1 is a subset of A and φ is a map of A into B, then we get a map of A_1 to B by restricting the domain to A_1.This map is called *restriction of* φ *to* A_1 and denoted by $\varphi|A_1$. Turning things around we will say that a map φ of A to B is an *extension of the map* ψ of A_1 to B if $\psi = \varphi|A_1$.

Two maps φ: $A \to B$ and ψ: $C \to D$ are said to be equal (denoted $\varphi = \psi$) if and only if $A=C$, $B=D$ and $\varphi(a)=\psi(a)$ for any $a \in A$. A map φ: $A \to B$ is called *surjective* if Im$\varphi = B$, that is, if the range coincides with the co-domain. φ: $A \to B$ is *injective* if distinct elements of A have distinct images in B, that is, if $a_1 \neq a_2 \Rightarrow \varphi(a_1) \neq \varphi(a_2)$. If φ is both injective and surjective, it is called *bijective* (or φ is said to be one-to-one correspondence between A and B).

Example 1.2. For sets $A = \{a, b, c\}$ and $B = \{1,2,3\}$, we define the maps f and g as

$$f : a \longmapsto 1, b \longmapsto 1, c \longmapsto 2$$

$$g : a \longmapsto 1, b \longmapsto 2$$

Then f is a mapping A into B, Imf=$f(A)$=$\{1, 2\}$, and $f^{-1}(1)$=$\{a, b\}$. However g is not a mapping A into B, because there exits no image of c under g. Clearly, f is neither a surjective nor a injective.

Example 1.3. Let Z and N be integers set and natural number set respectively, we define the maps φ, ψ and γ as following

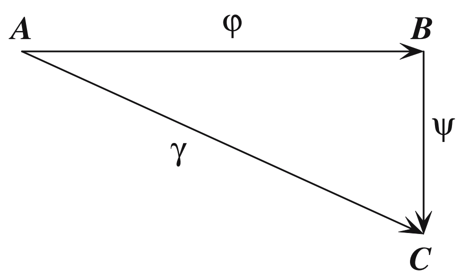
$$\begin{array}{ccc} \varphi : Z \longrightarrow N & \psi : Z \longrightarrow Z & \gamma : Z \longrightarrow Z \\ n \longmapsto |n|+1 & n \longmapsto 2n & n \longmapsto n+1 \end{array}$$

Then φ is a surjective, ψ is a injective, γ is a bijective, and $\psi \neq \gamma$.

Let φ: $A \to B$ and ψ: $B \to C$. Then we define the map γ: $A \to C$ as the map having the domain A and the co-domain C, by definition

$$(\psi\varphi)(a) = \psi(\varphi(a)) \ (\forall a \in A).$$

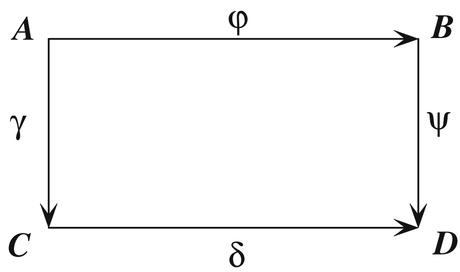
We call this map the *composite (or product, or sometimes resultant) of* φ *and* ψ (ψ following φ). It is often useful to indicate the relation γ=$\psi\varphi$ by saying that the

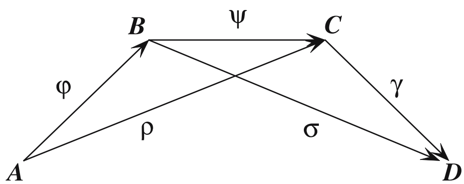


is *commutative*. Similarly, we express the fact that $\psi\varphi$=$\delta\gamma$ for φ: $A \to B$, ψ: $B \to D$, γ: $A \to C$, δ: $C \to D$ by saying that the rectangle triangle is commutative. Composition of maps satisfies the *associative law*: if φ: $A \to B$, ψ: $B \to C$, and γ: $C \to D$, then $\gamma(\psi\varphi)$=$(\gamma\psi)\varphi$. We note first that both of these maps have the same domain A and the same co-domain D. Moreover, for any $a \in A$ we have

$$(\gamma(\psi\varphi))(a) = \gamma((\psi\varphi)(a) = \gamma(\psi(\varphi(a)))$$

$$((\gamma\psi)\varphi)(a) = (\gamma\psi)(\varphi(a)) = \gamma(\psi(\varphi(a)))$$





so $\gamma(\psi\varphi)$ and $(\gamma\psi)\varphi$ are identical. This can be illustrated by shown the above diagram.

The associative law amounts to the statement that if the triangles *ABC* and *BCD* are commutative then the whole diagram is commutative.

Example 1.4. For any set A one defines the identity map 1_A (or 1 if A is clear) as

$$\begin{aligned} 1_A &: A \to A \\ a &\mapsto a \ (\forall a \in A). \end{aligned}$$

This map is a mapping A into A.

We now state the following important results:

(1) If φ: $A \to B$ one checks immediately that $1_B\varphi$=$\varphi 1_A$.
(2) φ: $A \to B$ is bijective if and only if there exists a map ψ: $B \to A$ such that $\psi\varphi$=1_A and $\varphi\psi$=1_B.

The map ψ satisfying $\psi\varphi$=1_A and $\varphi\psi$=1_B is unique since if ψ': $B \to A$ satisfies the same conditions, $\psi'\varphi$=1_A , $\varphi\psi'$=1_B, then

$$\psi' = 1_A\psi' = (\psi\varphi)\psi' = \psi(\varphi\psi') = \psi 1_B = \psi.$$

We will now denote ψ as φ^{-1} and call this the *inverse* of the (bijective) map φ. Clearly the foregoing result shows that φ^{-1} is bijective and $(\varphi^{-1})^{-1}$=φ.

As a first application of the criterion for bijectivity we give a formal proof of a fact which is fairly obvious anyhow: the product of two bijective maps is bijective. For, Let φ: $A \to B$ and ψ: $B \to C$ be bijective. Then we have the inverses φ^{-1}: $B \to A$ and ψ^{-1}: $C \to B$ and the composite map $\varphi^{-1}\psi^{-1}$: $C \to A$. Moreover,

$$(\psi\varphi)(\varphi^{-1}\psi^{-1}) =((\psi\varphi)\varphi^{-1})\psi^{-1} = (\psi(\varphi\varphi^{-1}))\psi^{-1} = \psi\psi^{-1} =1_C.$$

Also,

$$(\varphi^{-1}\psi^{-1})(\psi\varphi)=\varphi^{-1}(\psi^{-1}(\psi\varphi))=\varphi^{-1}((\psi^{-1}\psi)\varphi)=\varphi^{-1}\varphi=1_A.$$

Hence $\varphi^{-1}\psi^{-1}$ is an inverse of $\psi\varphi$, that is

$$(\psi\varphi)^{-1} = \varphi^{-1}\psi^{-1} \tag{1.1}$$

This important formula has been called the "*dressing-undressing principle*": what goes on in dressing comes off in the reverse order in undressing (e.g., socks and shoes).

The proofs of the following statements remain as exercises. Let f be a map of X into Y and A, B be any subsets of X and U, V be any subsets of Y. Then the following assertions hold.

(1)$f(A\cup B) = f(A)\cup f(B)$, $f(A\cap B) \subseteq f(A)\cap f(B)$;
(2)$f^{-1}(U\cup V) = f^{-1}(U)\cup f^{-1}(V)$, $f^{-1}(U\cap V) \subseteq f^{-1}(U)\cap f^{-1}(V)$.

It is important to extend the notion of the Cartesian product of two sets to the product of any finite or infinite number of sets using concept of mapping. Consider to the Cartesian product $A_1 \times A_2$ of two sets A_1, A_2, where I=$\{1, 2\}$ is subscript set (or indexing set). For every $(a_1, a_2) \in A_1 \times A_2$, there exists a unique map $f$from Iinto $A_1 \cup A_2$ such that $f(1)$=$a_1 \in A_1$, $f(2)$=$a_2 \in A_2$. Inversely, for every map f: $I \to A_1 \cup A_2$ which satisfies $f(1)$=$a_1 \in A_1$, $f(2)$=$a_2 \in A_2$, there exists a unique element (a_1, a_2) in $A_1 \times A_2$ such that a_1=$f(1)$, a_2=$f(2)$. Therefore, there exists an one-to-one correspondence between the maps satisfied above properties and elements in $A_1 \times A_2$. It follows that, we can extend the notion of the Cartesian product as follows:

Definition 1.3. Let $\{A_i|i \in I\}$ be family of sets and I ($\neq \varnothing$) indexing set (finite or infinite). The *Cartesian product* of A_i $(i \in I)$ is defined as

$$\{f|f : I \to \cup_{i\in I}A_i, \forall i \in I, f(i) \in A_i\},$$

and denoted by $\prod_{i\in I} A_i$.

It is clear that $f \in \prod_{i\in I} A_i$ is determined by image of $\{f(i)|i \in I\}$. If a_i=$f(i)$, we always consider the $f$as the set $\{a_i|i \in I\}$. Thus, when I=$\{1, 2, \ldots, n\}$, we have

$$\prod_{i\in I} A_i = \{(a_1,a_2,\ldots,a_n)|a_i \in A_i, i = 1,2,\ldots,n\}.$$

Let A be a set and $A \neq \varnothing$, the map $f$of $A\times A$ into A is called an *algebra operation* (*binary operation or binary composition*) of A, that is, $\forall a$, $b \in A$, there exists a unique element c in A such that $f(a, b)$=c, denoted by $a\cdot b$=c (or ab=c). For example, in 2^A, the power set of a set A, we have the algebra operation $\cap$, $\cup$ (i.e., for any C, $D \in 2^A$, $(C, D) \to C\cap D$ and $(C, D) \to C\cup D$). Also, let Q be the set of rational number, then additive, subtraction and multiplication of numbers all are algebra operations of Q, but division of numbers does not, for zero cannot be a divisor.

The concept of algebra operations can be extended to any finite number of sets. Let A be a non-vacuous set and n natural numbers, then a *n-ary operation of* A is a map f from $A \times A \times \ldots \times A$ to A. An *algebraic system* is a non-vacuous set with n-ary operation. For example, if "$\cdot$", "$*$" are binary operation of A, algebraic system A is denoted usually by $(A, \cdot, *)$. Moreover, $(Q, +, -, \times)$ and $(2^A, \cap, \cup, ')$ are algebraic systems.

We consider some important connections between maps and equivalence relations. Suppose φ: $A \to B$. Then we can define a relation R_φ in A by specifying that $aR_\varphi b$ if and only if for a, $b \in A$, $\varphi(a)$=$\varphi(b)$. It is clear that this is an equivalence relation in A. If $c \in B$, c=$\varphi(a)$ for some $a \in A$, then $\varphi^{-1}(c)$=$\varphi^{-1}(\varphi(a))$=$\{ b|$ $\varphi(b)$=$\varphi(a)\}$ and this is just the equivalence class $R_\varphi(a)$ in A determined by the element a. We shall refer to this subset of A also as the *fiber over the element* $c \in \mathrm{Im}\varphi$. The set of these fibers constitutes the partition of A determined by $R\varphi$ that is, they are the elements of the quotient set A/R_φ .

In the general, φ: $A \to B$ defines a map φ^* of A/R_φ into B: abbreviating $R_\varphi(a) = \varphi^{-1}(\varphi(a))$. We simply define φ^* by writing down

$$\varphi^*(R_\varphi(a)) = \varphi(a) \tag{1.2}$$

Since $R_\varphi(a) = R_\varphi(b)$ if and only if $\varphi(a) = \varphi(b)$, it is clear that the right-hand side is independent of the choice of the element a in $R_\varphi(a)$ and so, indeed, we do have a map. We call φ^* the map of A/R_φ induced by φ. This is injective since $\varphi^*(R_\varphi(a))$ = $\varphi^*(R_\varphi(b))$ gives $\varphi(a) = \varphi(b)$ and this implies $R_\varphi(a) = R_\varphi(b)$, by the definition of R_φ. Of course, if φ is injective to begin with, then $aR_\varphi b$ ($\varphi(a) = \varphi(b)$) implies a = b. In this caseA/R_φ can be identified with A and φ^* can be regarded as the same as φ.

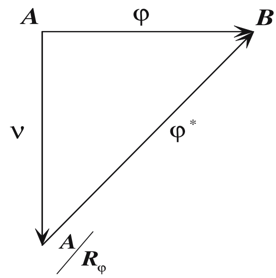
Let $\nu : A \to A/R_\varphi$ be a map for any $a \in A$, $\nu(a) = R_\varphi(a)$ (i.e., the natural map of A to A/R_φ). We now observe that $\varphi^*(\nu(a)) = \varphi^*(R_\varphi(a)) = \varphi(a)$. Hence we have the factorization of the given map as a product of the natural map ν of A to A/R_φ

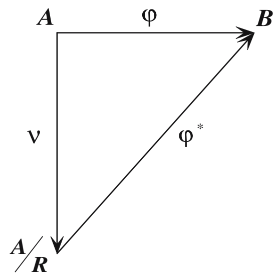
$$\varphi = \varphi^* \nu \tag{1.3}$$

and the induced map φ^* of A/R_φ to B. The map φ^* is injective and ν is surjective. The relation (1.3) is equivalent to the commutativity of the diagram A/R_φ to B. The map φ^* is injective and ν is surjective. The relation (1.3) is equivalent to the commutativity of the diagram.

Since ν is surjective it is clear that $\mathrm{Im}\varphi = \mathrm{Im}\varphi^*$. Hence φ^* is bijective if and only if φ is surjective. We remark finally that φ^* is the only map which can be defined from A/R_φ to B to make the above commutative diagram. Let ψ: $A/R_\varphi \to B$ satisfy $\psi\nu = \varphi$. Then $\psi(R_\varphi(a)) = \psi(\nu(a)) = \varphi(a)$. Hence $\psi = \varphi^*$.

There is a useful generalization of these simple considerations. Suppose we are given a map φ: $A \to B$ and an equivalence relation Ron A. We shall say that φ is *compatible with* R if aRb for a, b in A implies $\varphi(a) = \varphi(b)$. In this case, we can define a map φ^* of A/R to B by $\varphi^* : R(a) \to \varphi(a)$. Clearly this is well defined, and if ν denotes the natural surjection $a \longmapsto R(a)$, then $\varphi = \varphi^* \nu$, that is, we have the commutativity.





In this case the map φ^* need not be injective. In fact φ^* is injective if and only if the equivalence relation $R = R_\varphi$. We now call attention to the map ν of A into A/E defined by

$$\nu : a \to R(a).$$

We call this the *natural map* of A to the quotient set A/R. Clearly, ν is surjective.

1.1.4 Countable Sets

A set is finite if and only it can be put into one-to-one map with a set of the form $\{p|p \in N \text{ and } p < q\}$ for some $q \in N$, where N is the set of non-negative integers.

Definition 1.4. A set A is called *countably infinite* if and only if it can be put into one-to-one map with the set N of non-negative integers; A set is *countable* if and only if it is either finite or countably infinite, otherwise it is called *uncountable* .

Theorem 1.3. *A subset of a countable set is countable.*

Its proof remains as an exercise.

Theorem 1.4. *If $\mathscr{A}$ is a countable family of countable sets, then $\cup_{A \in \mathscr{A}} A$ is countable.*

Proof. Because $\mathscr{A}$ is countable there is a one-to-one map F whose domain is a subset of N and whose range is $\mathscr{A}$. Since $F(p) \in \mathscr{A}$ is countable for each p in

N, it is possible to find an one-to-one map G_p on a subset of $\{p\}\times N$ whose range is the countable set $F(p)$. Consequently there is a one-to-one map on a subset of $N \times N$ whose range is $\cup_{A\in\mathscr{A}}A$, and the problem reduces to showing that $N \times N$ is countable. The key to this proof is the observation that, if we think of $N \times N$ as lying in the upper left to lower right contain only a finite number of members of $N \times N$. Explicitly, for n in N, let

$$B_n = \{(p,q) \mid (p,q) \in N \times N, \text{ and } p+q = n\ \}.$$

Then B_n contains precisely $n+1$ points, and the union $\cup_{n\in Z}B_n$ is $N \times N$. A one-to-one map on N with range $N \times N$ may be constructed by choosing first the members of B_0, next those of B_1 and so on. The explicit definition of such a function is left as an exercise. □

Corollary 1.1. *The set Q of all rational numbers is a countable set.*

Its proof remains as an exercise.

Corollary 1.2. *The set R of all real numbers is uncountable.*

Proof. Before we prove R is uncountable, we first prove that the interval (0, 1) is uncountable. We suppose that the set of real numbers in (0, 1) is countable and assume a one-to-one map has been set between the set of non-negative integers $N$and (0, 1). We indicate the correspondence the following diagram:

$$1 \leftrightarrow 0.a_{11}a_{12}a_{13}a_{14}\ldots$$

$$1 \leftrightarrow 0.a_{21}a_{22}a_{23}a_{24}\ldots$$

$$1 \leftrightarrow 0.a_{31}a_{32}a_{33}a_{34}\ldots$$

$$\cdots\cdots$$

$$\cdots\cdots$$

Where each a_{ij} represents a digit, i.e., $0\leq a_{ij} \leq 9$, and where it is assumed that where we have two alternate choices for the decimal expression of real number, as for example in the case where 2/10 could be written either as 0.2000... or as 0.1999..., we always choose the one that ends in a tring of zeros. Now this one-to-one correspondence is such that to every positive integer there correspondence some real number in (0, 1) and conversely to each real number in (0, 1) there correspondence some integer. Consequently the infinite list of decimals given above is complete in the sense that every real number of (0, 1) occurs somewhere in the list. If , then, we can produce a real number in (0, 1) which is not in this list we shall have a contradiction, and this is precisely what we set out to do. We define b=0.$b_1b_2b_3\ldots$ as follows: if a_{ii} is 5 let b_i=6, if $a_{ii} \neq 5$ let b_i=5. Now it is clear that b is not equal to any one of the decimals in our list for it differs from the nth one at the nth place. Also it is clear that b is between 5/9 and 2/3, so that $b \in (0, 1)$. This contradiction then shows that there cannot exist such a one-to-one correspondence between N and (0, 1), and the set of real number in (0, 1) is uncountable. Since (0, 1) is a subset of the set of real number R, hence R is uncountable because of Theorem 1.3. □

1.1.5 Partially Ordered Sets, Directed Sets and Nets

The most general concept we consider in this section is that of a partially ordered set. We recall that a binary relation on a set S is a subset R of the product set $S \times S$. We say that a is in the relation R to b if and only if $(a,b) \in R$.

Definition 1.5. *A partially ordered set* $(S, \leq)$ (simply, S) is a set S together with a binary relation $R_{\leq}$ on S satisfying the following conditions, where for $a, b \in S$, $a \leq b$ simply denotes $(a, b) \in R_{\leq}$:

(1)$a \leq a$ for any $a \in S$; (*reflexivity*)
(2)If $a \leq b$ and $b \leq a$, then $a = b$; (*anti-symmetry*)
(3)If $a \leq b$ and $b \leq c$, then $a \leq c$. (*transitivity*)

Also, "$\leq$" is called a *partially ordered relation* on set S.

Let $(S, \leq)$ be a partially ordered set, and $a, b \in S$. If $a \leq b$ and $a \neq b$, then we write $a < b$. Also we write $a \geq b$ as an alternative for $b \leq a$ and $a > b$ for $b < a$. If $a \leq b$ or $b \leq a$, then we say that is *comparable* with a and b. In general we may have neither $a \leq b$ nor $b \leq a$ holding for the pair of elements a, $b \in S$, denoted by $a||b$, and then we say that is *not comparable* or *uncomparable* between a and b.

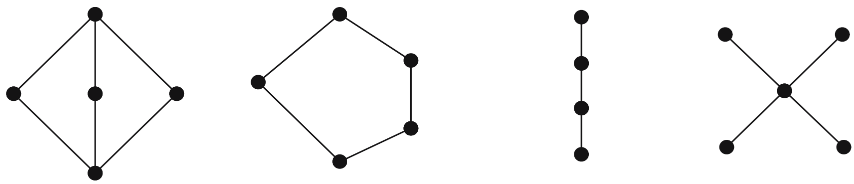
Example 1.5. Let 2^S be the power set of a set S. If $A \leq B$ for subsets A and B means $A \subseteq B$, then $(S, \subseteq)$ is a partially ordered set. Let N be a set of natural numbers. If $a \leq b$ for natural numbers a and b means $a|b$ (a is a divisor of b), then $(N, |\ )$ is a partially ordered set.

Definition 1.6. Let $(S, \leq)$ be a partially ordered set. If we do have $a \leq b$ or $b \leq a$ for every pair (a, b), in other word, every pair (a, b) is comparable, then we call $(S, \leq)$ (simply, S) a *linear ordered set* (or *a chain* , or a *totally ordered set*), and call the "$\leq$" *linear* (or *totally*) *order* on set S.

Example 1.6. Let N be a set of natural numbers. If $a \leq b$ for natural numbers a and b means that a is less than b or equals b, then $(N, \leq)$ is a linear ordered set.

Let $\leq$ be a partially ordered relation on the set S and N a subset of S, then *inverse relation* $\leq^{-1}$ (*or* $\geq$) of $\leq$ is also a partially ordered relation on S. *Induced relation* $\leq^N$ of $\leq$ on N is a partially ordered relation on N such that for a, $b \in N$, $a \leq^N b \Leftrightarrow a \leq b$. We say partially ordered set $(S, \leq^{-1})$ and $(N, \leq^N)$ to be *dual ordered subset* of partially ordered set $(S, \leq)$, respectively. If $(N, \leq^N)$ is a chain, that is, N with partially ordered relation $\leq^N$ became a chain, then we say N to be a chain in S. Clearly, if $(S, \leq)$ is a chain, then it is dual and partially ordered subset are also chain, respectively. Let C be a chain in S, we call C a *maximal chain* in S if D is a chain in S and $C \subseteq D$, then $C = D$.

In a partially ordered set $(S, \leq)$ the relation "$<$" can be expressed in terms of a relation of *covering* . We say that a_1 is a *cover* of a_2 (or a_1 is a *prime over* of a_2, or a_2 is a *prime under* of a_1) if $a_2 < a_1$ and there exists no u such that $a_2 < u < a_1$, denote by $a_2 \prec a_1$. It is clear that $a < b$ in partially ordered set if and only if there exists a sequence $b = a_1, a_2, \ldots, a_n = a$ such that each a_i is a cover of a_{i+1}. The notion of cover suggests a way of representing partially ordered set S by a diagram (Hasse diagram). We represent the elements of S by dots. If a_1 is a cover of a_2 then we place a_1 above a_2 and connect the two dots by a straight line. Then $b < a$ if and only if there is a descending broken line connecting a to b. If a and b are not comparable, that is $a||b$, then no line connects a and b. Some examples of Hasse diagrams of partially ordered sets are shown below.



The third example, represent a totally ordered set.

We now define quasi-ordered relation which is weaker than partially ordered relation.

Definition 1.7. *A quasi-ordered set* $(S, \leq)$ *is a* set S together with a binary relation $R_{\leq}$ on S satisfying the following conditions, where for $a, b \in S$, $a \leq b$ simply denotes $(a, b) \in R_{\leq}$:

(1) $a \leq a$; (*reflexivity*)
(2) If $a \leq b$ and $b \leq c$, then $a \leq c$. (*transitivity*)

"$\leq$" is called a *quasi-ordered relation* on set S.

It is clear that a partially ordered relation must be a quasi-ordered relation. But the converse statement is not always true. For example, in the set of real numbers R, we define a binary relation $\leq'$ as follows

$$a \leq' b \Leftrightarrow |a| \leq' |b|, (\forall a, b \in R)$$

Then $\leq'$ is a quasi-ordered relation without being a partially ordered relation.

Definition 1.8. Let $(S, \leq)$ be a quasi-ordered set. $(S, \leq)$ is called a *quasi-linear ordered set* if we do have $a \leq b$ or $b \leq a$ for pair of elements $a, b \in S$.

Definition 1.9. Let $(S, \leq)$ be a partially ordered set, A a non-empty subset of S, and $a \in A$. If we have $x \leq a$ for every $x \in A$, then a is called a *maximum element* of A. If there exist no $y \in A$ such that $a < y$ $(a \neq y)$, then a is called a *maximal element* of A. Dually, we can define *minimum element* and *minimal element* of A.

It is clear that a maximum (minimum) element must be a maximal (minimal) element. But in general the converse of the result is not true. In particular, the maximum element of the partially ordered set $(S, \leq)$ (if it exists) is called *identity element* of S, denoted as I or 1, the minimum element of the partially ordered set $(S, \leq)$ (if it exists) is called *zero element* of S, denoted as O or 0. In a partially ordered set there may be more than one maximal (minimal) element, but there is only one maximum (minimum) element if it has. In virtue of the following theorem, we know that every finite chain must have a maximal (minimal) element.

Theorem 1.5. *Let $(S, \leq)$ be a partially ordered set and A be a non-empty subset of S. Then the following assertions hold:*

(1) If A has maximum (minimum) element, then the maximal (minimal) element is unique.
(2) If A is finite subset of S, then there must exist a maximal (minimal) element.
(3) If A is a chain in S (e.g. linear ordered subset), then maximal (minimal) element of A (if it exists) must be maximum (minimum) element.

Proof. We only present the proof of (2) here while the others are left as exercise. Let $A = \{a_1, a_2, \ldots, a_n\}$, we define sequence

$$m_1, m_2, \ldots, m_n$$

of elements in S, such that $m_1 = a_1$ and

$$m_k = \begin{cases} a_k, & if \;\; m_{k-1} < a_k \\ m_{k-1}, & otherwise \end{cases} \tag{1.4}$$

It is clear that m_n is a maximal element of A. Similarly, we can prove that A has the minimal element. □

Let $(S, \leq)$ be a partially ordered set (resp. chain). If S is a finite set, then $(S, \leq)$ is called a *finite partially ordered set* (resp. *finite chain*). If not, $(S, \leq)$ is called a *infinite partially ordered set* (resp. *infinite chain*). Let $(S, \leq)$ be a partially ordered set together with maximal element I and minimal element O. Then prime over (or cover) of O is called an *atom* of S, and prime under of I is called a *dual atom* of S.

Definition 1.10. Let D be a non-empty set and $R_{\geq}$ be a binary relation on D. For $a, b \in D$, $a \geq b$ simple denotes $(a, b) \in R_{\geq}$. $R_{\geq}$ is said to *direct* the set D if it satisfies the following conditions:

(1) if $m, n, p \in D$ such that $m \geq n$ and $n \geq p$, then $m \geq p$;
(2) if $m \in D$, then $m \geq m$;
(3) if $m, n \in D$, then there is p in D such that $p \geq m$ and $p \geq n$.

We say that m *follows* in and n *precedes* m if and only if $m \geq n$.

The family of all finite subsets of a set is directed by $\supseteq$. A *directed set* is a pair $(D, \geq)$ such that $\geq$ directs the set D. This sometimes called a directed system. A *net* is a pair $(S, \geq)$ such that S is a map and $\geq$ directs $D$the domain of S, which is simply written as $\{S_n \mid n \in D, \geq\}$. A net $\{S_n \mid n \in D, \geq\}$ is *in* a set A if and only if $S_n \in A$ for all $n \in D$; it is *eventually in A* if and only if there is an element m of D such that, if $n \in D$ and $n \geq m$, then $S_n \in A$; it is *frequently in A* if and only if for each m in D there is n in D such that $n \geq m$ and $S_n \in A$. If $\{S_n \mid n \in D, \geq\}$ is frequently in A, then the set E of all members $n$of D such that $S_n \in A$ has the property: for each there is p in E such that $p \geq m$. Such subsets of D are called *cofinal* . Each cofinal subset E of D is also directed by $\geq$ because for elements m and n of E there is p in D such that $p \geq m$ and $p \geq n$, and there is then an element q of E which follows p.

1.1.6 Maximal Conditions and Minimal Conditions

Let $(S, \leq)$ be a partially set. We now consider the following conditions.

A. Minimal Condition: *Every non-empty subset of S must have minimal elements.*

B. Descending Chain Condition: *For every sequence of elements* $\{a_i \mid i = 1, 2, \ldots\}$, *if*

$$a_1 \geq a_2 \geq \ldots \geq a_n \geq \ldots,$$

then there exists a positive integer m such that $a_m = a_{m+n}$, $n = 1, 2, \ldots$.

C. Inductive Condition: *For any property* ε, *if*

(1) *Every minimal element* (*if it exists*) *has property* ε,
(2) *For* $\forall a, x \in S$, $x < a$, *x has property* $\varepsilon \Rightarrow$ *a has property* ε.

Then every element in S has also property ε.
The duality hold; we have

A′. Maximal Condition: *Every non-empty subset of S must have maximal elements.*

B′. Descending Chain Condition: *For every sequence of elements* $\{a_i \mid i = 1, 2, \ldots\}$, *if*

$$a_1 \leq a_2 \leq \ldots \leq a_n \leq \ldots,$$

then there exists a positive integer m such that $a_m = a_{m+n}$, $n = 1, 2, \ldots$.

C′. Dual Inductive Condition: *For any property* ε, *if*

(1) *Every maximal element* (*if it exists*) *has property* ε,
(2) *For* $\forall a, x \in S$, $x > a$, *x has property* $\varepsilon \Rightarrow$ *a has property* ε.

Then every element in S has also property ε.
The relation among those condition terms from the following

Theorem 1.6. *Conditions A, B and C (dually, Conditions A′, B′ and C′) are equivalent for any partially set* $(S, \leq)$.

Proof. A $\Rightarrow$ C. Let S satisfy the minimal condition (i.e. A), ε be a property. The premises of inductive condition (i.e. C) is satisfied. Let

$$M = \{a \mid a \in S \text{ and } a \text{ has no property } \varepsilon\},$$

then $M \subseteq S$. If $M \neq \varnothing$, there exists a minimal element $a \in M$ by A, but a is not a minimal element in S from the premises of inductive condition. However, if $x < a$ and $x \in P$, then $x \notin M$ and has property ε, consequently, a has also property ε from the premise (2) of inductive condition. This contradicts that $a \in M$ and $M = \varnothing$. Thus inductive condition (i.e. C) holds.

C⇒B. Let S satisfy inductive condition (i.e. C). Definite: $a(a \in S)$ has **property** ε if and only if for every descending chain

$$a = a_1 \geq a_2 \geq \ldots \geq a_n \geq \ldots,$$

there exists a positive integer m such that $a_m = a_{n+n}$, $n = 1, 2, \ldots$. It is clear that every minimal element (if it exists) in S has property ε. Let $a \in S$, and $\forall x \in S$, if $x < a$, then x has property ε, then a has property ε. Consequently, every element in S has property ε by inductive condition (i.e. C), that is, descending chain condition (i.e. B) holds.

B⇒A. Let descending chain condition (i.e. B) holds. Let us suppose that minimal condition (i.e. A) does not hold, then there exists a non-empty subset N of P such that N has no minimal element. It is obvious that N is an infinite set. Let $a_1 \in S$, then a_1 is not the minimal element of N, consequently, there exists a_2 such that $a_1 > a_2$. Since a_2 is not the minimal element of N, there exists a_3 such that $a_1 > a_3 \cdots\cdots$. It follows that, there is a sequence of elements of N such that

$$a_1 > a_2 > \ldots > a_n > \ldots,$$

and this contradicts that descending chain condition (i.e. B) holds.

Similarly, we can show that conditions A′, B′ and C′ are equivalent. □

Definition 1.11. A linear ordered set (or a chain) which satisfies maximal condition is called a *well ordered set* . Its linear ordered relation is called *well order*.

Example 1.7. $(N, \geq)$, the set of natural numbers N ordered by usually order relation of numbers $\geq$, becomes a well ordered set.

For a well ordered set $(S, \leq)$, it is easy to verify that every partially ordered subset of S is also a well ordered set, and there exists a unique minimal element (it is also minimum element) in S. Moreover, we have the following theorems (their proofs are left to the reader).

Theorem 1.7. *Let $(S, \leq)$ be a partially ordered set. S satisfies minimal condition if and only if every chain in S is a well ordered set.*

Here, we present the axiom of choice, and line up, without proof, some theorems, which are equivalent to the axiom of choice. They will play a fundamental role in what follows.

Theorem 1.8. *(Axiom of choice) Let $P*(S) = 2^S - \varnothing$. Then there exists a mapping φ: $P*(S) \to A$ such that $\varphi(T) \in T$ for every $T \in P*(S)$.*

Theorem 1.9. *(Zermelo) For any set S, there exists a linear order* $\geq$ *such that* $(S, \leq)$ *is a well ordered set.*

Theorem 1.10. *(Hausdorff) For any partially set* $(S, \leq)$*, every chain in S is included in some maximal chain.*

Theorem 1.11. *(Kuratowski-Zorn) Let* $(S, \leq)$ *be a partially ordered set. If every chain in S has an upper bound in S, then every element of S is included in some maximal element of S.*

1.2 Topological Spaces

Our purpose here is to study what is ordinarily called point set topology [4]. Point set topology is one of the fundamentals that we study the theory of topological molecular lattices, the topology structures on the AFS structure and the applications to the pattern recognition.

The development of general topology has followed an evolutionary development which occurs frequently in mathematics. One begins by observing similarities and recurring arguments in several situations which superficially seem to bear little resemblance to each other. We then attempt to isolate the concepts and methods which are common to the various examples, and if the analysis has been sufficiently penetrating we may find a theory containing many or all of our examples, which in itself seems worthy of study. It is in precisely this way, after much experimentation, that the notation of a topology space was developed. It is a natural product of a continuing consolidation, abstraction, and extension process. Each such abstraction, if it is to contain the examples from which it was derived in more than a formal way, must be tested to find whether we have really found the central ideas involved. In this case we want to find whether a topological space, at least under some reasonable restrictions, must necessarily be one of the particular concrete spaces from which the notation is derived. The "standard" examples with which we compare spaces are Cartesian products of unit intervals and metric spaces.

1.2.1 Neighborhood Systems and Topologies

In a certain sense, a neighborhood of a point x is a set of points which lie "*close*" to the point. For example, you have the sets of "*close friends*", "*common friends*", "*friends*" etc. to describe your relationship with your friends. According to the opinion, the notion of a neighborhood system of a point x of the set is defined as the following Definition 1.13 which is abstracted from Euclidean space.

Definition 1.12. Let X be a set and $\mathscr{T}$ be a family of subsets of X (i.e., $\mathscr{T} \subseteq 2^X$). The pair $(X, \mathscr{T})$ is call a *topological space*, if the following conditions hold.

(1) The union of any number of sets of $\mathscr{T}$ is again in $\mathscr{T}$;
(2) The intersection of any two of sets of $\mathscr{T}$ is again in $\mathscr{T}$;

(3) $X \in \mathscr{T}$;
(4) $\varnothing \in \mathscr{T}$.

It is frequently convenient to say simply that X is a topological space, rather than having to specify the topology, $\mathscr{T}$, and having to write $(X, \mathscr{T})$, since, more often than not, we are interested not so much in a particular topology, but further in properties that any topology still possesses. We shall thus feel free to omit any specific mention of the topology unless it is important to the context to emphasize a particular topology, or to distinguish between different topologies.

The members of the topology $\mathscr{T}$ are called *open* relative to $\mathscr{T}$, or $\mathscr{T}$*–open*, or if only one topology is under consideration, simply open sets. Let$X$ be a set. By Definition 1.2.1, we know that $(X, \mathscr{T})$ is a topological space if $\mathscr{T} = \{X, \varnothing\}$. This is not a very interesting topology, but it occurs frequently enough to deserve a name; it is called the *indiscrete* (or *trivial*) topology for X, and $(X, \mathscr{T})$ is then an *indiscrete topological space*). At the other extreme is the family of all subsets of X, $\mathscr{T} = 2^X$, which is the *discrete topology* for X, and $(X, \mathscr{T})$ is then a *discrete topological space*. If $\mathscr{T}$ is the discrete topology, then every subset of the space is open.

The discrete and indiscrete topologies for a set X are respectively the largest and the smallest topology for X. If $\mathscr{T}_1$ and $\mathscr{T}_2$ are topologies for X, then, following the convention for arbitrary families of sets, $\mathscr{T}_1$ is smaller than $\mathscr{T}_2$ if and only if $\mathscr{T}_1 \subseteq \mathscr{T}_2$. In other case, it is also said that $\mathscr{T}_1$ is coarser than $\mathscr{T}_2$ and $\mathscr{T}_2$ is finer than $\mathscr{T}_1$. The space $(X,\mathscr{T})$ is called a finite topological space if X is a finite set. Otherwise, $(X,\mathscr{T})$ is called an infinite topological space.

Definition 1.13. Let X be a set and $(X, \mathscr{T})$ be a topological space. A set $U \subseteq X$ is called a *neighborhood* of a point $x \in X$ if and only if there exists an open set $V \in \mathscr{T}$ such that $x \in V \subseteq U$. The family of all neighborhoods of the point x is called the *neighborhood system* of x, written $\mathscr{U}_x$.

A neighborhood of a point need not be an open set, but every open set is a neighborhood of each of its points. Each neighborhood of a point contains an open neighborhood of the point. The following theorem shows that a topology for a set X can be generated by the neighborhood systems of the points in X.

Theorem 1.12. *A set is open if and only if it contains a neighborhood of each of its points.*

Proof. The union U of all open subsets of a set A is surely an open subset of A. If A contains a neighborhood of each of its points, then each member x of A belongs to some open subset of A and hence $x \in U$. In this case $A = U$ and therefore A is open. On the other hand, if A is open it contains a neighborhood (namely, A) of each of its points. □

Example 1.8. Let R be the real line. The usual topology for R is the family of all those sets which contain an open interval about each of their points (e.g., $(a,b) = \{x \mid a < x < b\}$). Define

$$\mathscr{U}_x = \{U|x \in (a,b) \subseteq U \text{ for some } a,b \in R, a < b\}.$$

Then $\mathscr{U}_x$ is a neighborhood system at x and generates the usual topology for R according to Theorem 1.12.

Example 1.9. Let X be a set, and let $\mathscr{U}_x$=$\{U \mid x \in U, U \in 2^X\}$ for each $x \in X$, then $\mathscr{U}_x$ is a neighborhood system at x, and the topology thus generated is the discrete topology for X.

Example 1.10. Let X be a set, and let $\mathscr{U}_x = \{X\}$ for each $x \in X$, then $\mathscr{U}_x$ is a neighborhood system at x, and the topology thus generated is the trivial (indiscrete) topology for X.

Theorem 1.13. *If $\mathscr{U}$ is the neighborhood system of a point, then finite intersections of members of $\mathscr{U}$, and each set which contains a member of $\mathscr{U}$ belongs to it.*

Proof. If U and V are neighborhoods of a point x, there are open neighborhoods U_0 and V_0 contained in U and V respectively. Then $U \cap V$ contains the open neighborhood $U_0 \cap V_0$ and is hence a neighborhood of x. Thus the intersection of two (and hence of any finite number of) members of U is a member. If a set U contains a neighborhood of a point x it contains an open neighborhood of x and is consequently itself a neighborhood. □

1.2.2 Limit Points Closure of a Set and Closed Sets

Now we introduce first the notions of limit points and the derived set of a set, via the following manner.

Definition 1.14. Let X be a topological space and A be a subset of X. Then the point $x \in X$ is said to be a *limit point* (sometimes called *accumulation point* or *cluster point*) of A, provided that for every $U \in \mathscr{U}_x$, $U \cap A$ contains a point $y \neq x$.

Example 1.11. Let us study the real plane.

1. In the real plane with the usual topology, any point of the form (0, y) is a limit point of the set $D = \{(x,y) \mid x > 0\}$.
2. On the real line with the usual topology, a as well as b is a limit point of the interval (a, b).
3. Let X be a nonempty set with the discrete topology, let $A \subseteq X$ and let $x \in X$, then x is not a limit point of A.

Definition 1.15. Let X be a topological space and let $A \subseteq X$. The *derived set of A* , written A^d, is the set of all $x \in X$ such that x is a limit point of A.

Example 1.12. On the real line with the usual topology, let $A = (a,b)$, then A^d = $[a,b]$. Let $B = \{x \mid 0 < x \leq 1 \text{ or } x = 2\}$, then B^d = [0, 1]. By Definition. 1.14 and Definition. 1.15, it is easy to prove the following theorem.

Theorem 1.14. *In any topological space, the following assertions hold.*

(1) if $A \subseteq B$, then $A^d \subseteq B^d$.
(2) $(\varnothing^d)^d = \varnothing$.

The closure of a set A in a topological space X has a number of interesting properties. They are presented though the following definition.

Definition 1.16. Let X be a topological space, and let $A \subseteq X$. The *closure of* A, written A^-, is the set $A \cup A^d$.

It is clear that the intersection of the members of the family of all closed sets containing A is the closure of A.

Theorem 1.15. *Let X be a topological space, and let $A \subseteq X$, then $A^{--} = A^-$.*

Proof. By Definition 1.16 $A^{--} = A^- \cup (A^-)^d$. We show that $(A^-)^d \subseteq A^-$. Let $x \in (A^-)^d$, and suppose $x \notin A^- = A \cup A^d$, then $x \notin A$ and further, since also $x \notin A^d$, by Definition 1.2.3 there exists some $U \in \mathscr{U}_x$, such that $U \cap A = \varnothing$. Select O, open, such that $x \in O \subseteq U$, then $O \in \mathscr{U}_x$, and further since $O \cap A \subseteq U \cap A = \varnothing$, we have $O \cap A = \varnothing$. Now since $x \in (A^-)^d$, $O \cap A^-$ contains some point $y \neq x$. Thus $y \in A^-$, and since $O \cap A = \varnothing$, $y \in A^d$. Since O is open, $O \in \mathscr{U}_x$, by Definition 1.14, thus there exists $z \neq y$ such that $z \in O \cap A$. This, however, contradicts $O \cap A = \varnothing$, consequently $x \in A^-$. This completes the proof that $(A^-)^d \subseteq A^-$. Finally $A^{--} = A^- \cup (A^-)^d$, since $(A^-)^d \subseteq A^-$. □

Theorem 1.16. *Let X be a topological space and let $A \subseteq X$ and $B \subseteq X$, then the following assertions hold.*

(1) If $A \subseteq B$, then $A^- \subseteq B^-$.
(2) $(A \cap B)^- \subseteq A^- \cap B^-$.
(3) $(A \cup B)^- = A^- \cup B^-$.

Proof. (1) If $A \subseteq B$, then $A^d \subseteq B^d$ by Theorem 1.14. Consequently, $A^- = A \cup A^d \subseteq B \cup B^d = B^-$.

(2) Let $x \in (A \cap B)^-$, and let $U \in \mathscr{U}_x$. Then $U \cap (A \cup B) \neq \varnothing$. Consequently, neither $U \cap A$ nor $U \cap B$ is empty, and $x \in A^-$ and $x \in B^-$, hence $x \in A^- \cap B^-$. Therefore $(A \cap B)^- \subseteq A^- \cap B^-$.

(3) Since $A \subseteq A \cup B$ and $B \subseteq A \cup B$, we have by part (1) of this theorem that $A^- \subseteq (A \cup B)^-$ and $B^- \subseteq (A \cup B)^-$, that is $A^- \cup B^- \subseteq (A \cup B)^-$. Now let $x \in (A \cup B)^-$, and suppose $x \notin A^-$ and $x \notin B^-$, then there exist U, $V \in \mathscr{U}_x$ such that $U \cap A = \varnothing$ and $U \cap B = \varnothing$. Now $U \cap V \in \mathscr{U}_x$ and

$$U \cap V \cap (A \cup B) = (U \cap V \cap A) \cup (U \cap V \cap B) \subseteq (U \cap A) \cup (V \cap B) = \varnothing,$$

but this contradicts $x \in (A \cup B)^-$. Consequently either $x \in A^-$ or $x \in B^-$, when $x \in A^- \cup B^-$, and finally $(A \cup B)^- = A^- \cup B^-$. □

Theorem 1.17. *Let X be a topological space, and let $A \subseteq X$ be open. Let $B \subseteq X$, then $A \cap B^- \subseteq (A \cap B)^-$.*

Proof. Let

$$x \in A \cap B^- = A \cap (B \cap B^d) = (A \cap B) \cup (A \cap B^d)$$

If $x \in A \cap B$, then

$$x \in (A \cap B) \cup (A \cap B)^d = (A \cap B)^-$$

Assume $x \in A \cap B^d$. Since $A \in \mathscr{U}_x$, if $U \in \mathscr{U}_x$, $A \cap U \in \mathscr{U}_x$ by Theorem 1.14. Now since $x \in B^d$, $(A \cap U) \cap B$ contains a point $y \neq x$. Thus for each $U \in \mathscr{U}_x$, $U \cap (A \cap B) = (A \cap U) \cap B$ contains a point $y \neq x$, whence

$$x \in (A \cap B)^d \subseteq (A \cap B)^-$$

In either case, $x \in (A \cap B)^-$, whence $A \cap B^- \subseteq (A \cap B)^-$. □

It is perhaps reasonable to ask at this point if there are sets which already contain all their limit points, and if such sets have any interesting and distinctive properties. We now set about investigating some of the properties of such sets.

Definition 1.17. Let X be a topological space, let $A \subseteq X$, then A is said to be closed provided $A = A^-$.

Remark 1.1. In any space, the sets $\varnothing$ and X (the whole space) are invariably closed, and for that matter also invariably open.

It is important to remember that open and closed are not antithetical for sets in a topological space, namely that a set may be both open and closed at the same time, which it may be open but not closed, that it may be closed but not open, and that it may be neither open nor closed. Consequently we have that, if we are in a position to want to prove a set closed, it will do us no good whatever to prove it is not open. It is, of course, equally true that it does no good to prove a set is not closed if our object is to prove that it is open. However, open and closed sets are related as is shown in the following

Theorem 1.18. *A set A, in a topological space X, is closed if and only if its complement, A', is open.*

Proof. Let A be closed, and let $x \in A'$. Since $x \notin A = A^-$, there exists a neighborhood $U \in \mathscr{U}_x$ such that $U \cap A = \varnothing$. Consequently, $U \subseteq A'$, when by Theorem 1.14, $A' \in \mathscr{U}_x$, and by Theorem 1.12, A' is open.

Conversely, let A' be open, and let $x \in A^-$. Suppose that $x \in A'$, then by Definition 1.13, $A' \in \mathscr{U}_x$, and so $A' \cap A \neq \varnothing$. This is clearly a contradiction, consequently $x \in A$, and $A^- \subseteq A$. However, since for any set $A \subseteq A^-$, we have $A = A^-$, and A is closed. □

A simple application of this theorem, together with DeMorgan's rule, establishes the following

Corollary 1.3. *In any topological space, the following assertions hold.*

(1) The intersection of any number of closed sets is closed, and
(2) The union of any two (hence any finite number) of closed sets is closed.

Proof. (1) Let A be an indexing set, and for each $\alpha \in A$ let $C\alpha$ be closed. Further let $C = \cap_{\alpha\in A} C_\alpha$, then $C' = (\cap_{\alpha\in A} C_\alpha)' = \cup_{\alpha\in A} C'_\alpha$, and since by Theorem 1.18 every C'_α is open, so also is $\cup_{\alpha\in A} C'_\alpha$ by Definition 1.12. Thus C'_α is open, again by Theorem 1.18, C is closed. □

1.2.3 Interior and Boundary

Here is another operator defined on the family of all subsets of a topological space, which is very intimately related to the closure of a set.

Definition 1.18. A point x of a subset A of a topological space is called an *interior point* of A if and only if A is a neighborhood of x, and the set of all interior point of A is said to be the *interior* of A, denoted A^0.

Theorem 1.19. *let A be a subset of a topological space X. Then the interior A^0 of A is open and is the largest open subset of A. A set A is open if and only if $A = A^0$. A set of all points of A which are not points of accumulation of $X - A$ is precisely A^0. The closure of $X - A$ is $X - A^0$.*

Proof. If a point x belongs to the interior of a set A, then x is a member of some open subset U of A. Every member of U is also a member of A^0, and A^0 consequently contains a neighborhood of each of its points and is therefore open by Theorem 1.12. If V is an open subset of A and $y \in V$, then A is a neighborhood of y and so $y \in A^0$. Hence A^0 contains each open subset of A and it is therefore the largest open subset of A. If A is open, then A is surely identical with the largest open subset of A; hence A is open if and only if $A = A^0$. Assume that x is a point of A such that is not an accumulation point of $X - A$. There is a neighborhood U of x which does not intersect $X - A$ and is therefore a subset of A. Then A is a neighborhood of x and $x \in A^0$. On the other hand, A^0 is a neighborhood of each of its points and A^0 does not intersect $X - A$, so that no point of A^0 is an accumulation point of $X - A$. Finally, since A^0 consists of the points of A which are not accumulation points of $X - A$, hence the complement of A^0, $X - A^0$, is precisely the set of all points which are either point of $X - A$ or accumulation points of $X - A$; that is, $X - A^0$ is the closure of $X - A$ by Definition 1.16. □

The preceding result can be stated as $(A^0)'$=$(A')^-$, and, it follows, by taking complements, that A^0=$((A')^-)'$. Thus the interior of A is the complement of the closure of the complement of A. If A is replaced by its complement it follows that A^-=$((A')^0)'$, so that the closure of a set is the complement of the interior of the complement. If X is an indiscrete space the interior of every set except X itself is empty. If X is a discrete space, then each set is open and closed and consequently identical with its interior and with its closure. If X is the set of real numbers with the usual topology, then the interior of the set of all integers is empty; the interior of closed interval is the open interval with the same endpoints. The interior of the set of rational numbers is empty, and the closure of the interior of this set is consequently empty.

Definition 1.19. Let A be a set of a topological space X. The set of all points which are interior to neither A nor $X - A$ is said to be the *boundary* of A. Equivalently, x is a point of the boundary if and only if each neighborhood of x intersects both A and $X - A$.

It is clear that the boundary of A is identical with the boundary of $X - A$. If X is indiscrete space and A is neither X nor empty, then the boundary of A is X, while if X is discrete space the boundary of every subset is empty. The boundary of an interval of real numbers, in usual topology for the real numbers, is the set whose only members are the endpoints of the interval, regardless of whether the interval is open, closed, or half-open. The boundary of the set of rational numbers, or the set of irrational, is the set of all real numbers. It is not difficult to discover the relations between boundary, closure, and interior. The following theorem, whose proof remains as an exercise, summarizes the facts.

Theorem 1.20. *Let A be a subset of a topological space X and let $b(A)$ be the boundary of A. Then $b(A) = A^- \cap (X - A)^- = A^- - A^0$, $X - b(A) = A^0 \cup (X - A)^0$, $A^- = A \cup b(A)$ and $A^0 = A - b(A)$. A set is closed if and only if it contains its boundary. A set is open if and only if it is disjoint from its boundary.*

1.2.4 Bases Countability Axioms Separability

In defining the usual topology for the set of real numbers we began with the family $\mathscr{B}$ of open intervals, and formed this family constructed the topology $\mathscr{T}$. The same method is useful in other situations and we now examine the construction in detail.

Definition 1.20. A *base (or basis) for a topology* , $\mathscr{T}$, of a space X is $\mathscr{B}$ a subset of $\mathscr{T}$ and for each $x \in X$ and each $U \in \mathscr{U}_x$, there exists $V \in \mathscr{B}$ such that $x \in V \subseteq U$. The sets of $\mathscr{B}$ are called *basic sets* and $\mathscr{B}$ is said to be a base for the topology $\mathscr{T}$. Let $\mathscr{B}_x$ be a subset of $\mathscr{T}$ and $x \in X$. If each $U \in \mathscr{U}_x$, there exists $V \in \mathscr{B}_x$ such that $x \in V \subseteq U$, then the sets of $\mathscr{B}_x$ are called basic sets of the neighborhood system of x and $\mathscr{B}_x$ is said to be a base for the neighborhood system of x.

The family of open intervals is a base (or basis) for usual topology of the real numbers, in view of the definition of the usual topology and the fact that open intervals are open relative to this topology. The following simple characterization of bases is frequently used as an alternative definition of base.

Corollary 1.4. *A subfamily $\mathscr{B}$ of a topology $\mathscr{T}$ for X is a base for $\mathscr{T}$ if and only if each member of $\mathscr{T}$ is the union of members of $\mathscr{B}$.*

Proof. Suppose that $\mathscr{B}$ is a base for the topology $\mathscr{T}$ and that $U \in \mathscr{T}$. Let V be the union of all members of $\mathscr{B}$ which are subsets of U and suppose that $x \in U$. Then there is W in $\mathscr{B}$ such that $x \in W \subseteq U$ by Definition 1.20, and consequently $x \in V$. So that $U \subseteq V$. Since V is surely a subset of U, hence $V = U$. Conversely, suppose $\mathscr{B} \subseteq \mathscr{T}$ and each member of $\mathscr{T}$ is the union of members of $\mathscr{B}$. If $U \in T$, then U is the union of members of a subfamily of $\mathscr{B}$, and for each $x$in U there is V in $\mathscr{B}$ such that $x \in V \subseteq U$. Consequently $\mathscr{B}$ is a base for $\mathscr{T}$ by Definition 1.20. □

Example 1.13. Let X a nonempty set, and let

$$\mathscr{B}_1 = \{\{x\} \mid x \in X\}, \mathscr{B}_2 = \{X\}$$

then $\mathscr{B}_1$ is a base for the discrete topology of X, and $\mathscr{B}_2$ is a base for the trivial topology of X.

Although this is a very convenient method for the construction of topology, a little caution is necessary because not every family of sets is the base for a topology. The reason for this situation is made clear by the following example and theorem.

Example 1.14. Let $X = \{0,1,2\}$, A=$\{0, 1\}$ and B=$\{1, 2\}$. If $S = \{X, A, B, \varnothing\}$, then cannot be the base for any topology because the union of members of is always a member of . Therefore if S were the base of a topology then that topology would have to be S itself, but S is not a topology due to $A \cap B \notin S$.

We shall enable to distinguish bases from other families of subsets, as in fact is shown by the following theorem.

Theorem 1.21. *A family, $\mathscr{B}$, of subsets of a set X is a base for some topology, $\mathscr{T}$, of X if and only if both the following assertion (1) and (2) hold.*

(1) $X = \cup_{B \in \mathscr{B}} B$;
(2) For each $x \in X$ and each pair $U, V \in \mathscr{B}$, for which $x \in U$ and $x \in V$, there exists $W \in \mathscr{B}$ such that $x \in W \subseteq U \cap V$.

Proof. Let $\mathscr{B}$ be a basis for some topology, $\mathscr{T}$, for X, and let $x \in X$. Then there exists $U \in \mathscr{U}_x$, such that $x \in U$, and by Definition 1.20 there exists $B_x \in \mathscr{B}$, such that $x \in B_x \subseteq U$. Clearly $X \subseteq \cup_{x \in X} B_x \subseteq X$, hence $X = \cup_{x \in X} B_x$ and condition (1) is met.

Let $U, V \in \mathscr{B}$, $x \in U$ and $x \in V$, define $Q = U \cap V$, then since by Definition 1.20, U and V are both open, so also is Q, whence $Q \in \mathscr{U}_x$. Consequently, by Definition 1.20, there exists $W \in \mathscr{B}$ such that $x \in W \subseteq Q = U \cap V$, and condition (2) is met.

Conversely, suppose $\mathscr{B}$ satisfies both (1) and (2). Let $\mathscr{T}$ be the family of all unions of members of $\mathscr{B}$. A union of members of $\mathscr{T}$ is itself a union of members of $\mathscr{B}$ and is therefore a members of $\mathscr{T}$, and it is only necessary to show that the intersection of two members U and V of $\mathscr{T}$ is a member of $\mathscr{T}$. If $x \in U \cap V$, then we may choose U_0 and V_0 in $\mathscr{B}$ such that $x \in U_0 \subseteq U$ and $x \in V_0 \subseteq V$, and then a member W of $\mathscr{B}$ such that $x \in W \subseteq U_0 \cap V_0 \subseteq U \cap V$ by (2). Consequently $U \cap V$ is the union of members of $\mathscr{B}$, and $\mathscr{T}$ is a topology according to Definition 1.12. □

We have just seen that an arbitrary family S of sets may fail to be the base for any topology. With persistence we vary the question and enquire whether there is a unique topology which is, in some sense, generated by S. Such a topology should be a topology for the set X which is the union of the members of S, and each member of S should be open relative to the topology; that is, S should be a subfamily of the topology. This raises the question: Is there a smallest topology for X which contains S? The following simple result will enable us to exhibit this smallest topology.

Theorem 1.22. *If S is any non-empty family of sets. Then the family of all finite intersections of members of S is the base for a topology for the set $X = \bigcup_{U \in S} U$.*

Proof. Suppose that S is a family of sets. Let $\mathscr{B}$ be the family of finite intersections of members of S. Then the intersection of two members of $\mathscr{B}$ is again a member of $\mathscr{B}$ and, from Theorem 1.21, $\mathscr{B}$ is the base for a topology. □

A family S of sets is a *subbase* for a topology $\mathscr{T}$ if and only if the family of finite intersections of members of S is a base for $\mathscr{T}$ (equivalently, if and only if each member of $\mathscr{T}$ is the union of finite intersections of members of S). In the view of the preceding theorem every non-empty family S is the subbase for some topology, and this topology is, of course, uniquely determined by S. It is the smallest topology containing S (That is, it is a topology containing S and is a subfamily of every topology containing S).

There will generally be many different bases and subbases for a topology and the most appropriate choice may depend on the problem under consideration. One rather natural subbase for the usual topology for real numbers is the family of half-infinite open intervals; that is, the family of sets of the form $(-\infty, a)=\{x \mid x < a\}$ or $(a, +\infty)=\{x \mid x > a\}$. Each open interval is the intersection of two such sets, and this family is consequently a subbase. A space whose topology has a countable base has many useful properties.

Definition 1.21. A space X, which has a base, $\mathscr{B}$, which is a countable family, i.e., $\mathscr{B} = \{B_i \mid i = 1, 2, \dots\}$ is said to satisfy the *second axiom of countability* or sometimes more simply to have a *countable basis* . We also speak of *spaces being second countable*, meaning thereby that they have a countable basis, or satisfy the second axiom of countability.

Example 1.15. Let R be the real numbers with the usual topology, and let

$$\mathscr{B}_1 = \{(a, b) \mid a, b \text{ rational numbers}, a < b\}$$

then $\mathscr{B}_1$ is a countable basis for R and R is second countable.

Theorem 1.23. *If A is an uncountable subset of a space whose topology has a countable base, then some point of A is an accumulation point of A.*

Proof. Suppose that no point of A is an accumulation point and that $\mathscr{B}$ is a countable base. By Definition 1.14, we know that for each x in A there is an open set containing no point of A other than x. Since $\mathscr{B}$ is a base we may choose B_x in $\mathscr{B}$ such that $B_x \cap A = \{x\}$. There is then a one-to-one correspondence between the points of A and the members of a subfamily of $\mathscr{B}$, and A is therefore countable. It contradicts to the assumption that A is an uncountable set. Therefore some point of A must be an accumulation point of A. □

Definition 1.22. The set A is said to be *dense* in the set B if $A^- \supseteq B$. If A is dense in the whole space, X, we say that A is everywhere dense, or sometimes, if there is no chance for misunderstanding, simply dense. If there exists $A \subseteq X$, A is countable such that $A^- = X$, then space X is said to be separable.

A separable space may fail to satisfy the second axiom of countability. For example, let X be an uncountable set with the topology consisting of the empty set and the complements of all finite subsets of X. Then every non-finite set is dense because it intersects every non-empty open set. On the other hand, suppose that there is a countable base $\mathscr{B}$ and let x be a fixed point of X. The intersection of the family of all open sets to which x belongs must be $\{x\}$, because the complement of every other point is open. It follows that the intersection of those members of the base $\mathscr{B}$ to which x belongs is $\{x\}$. But the complement of this countable intersection is the union of a countable number of finite sets, which is equal to $X-\{x\}$, hence countable, and this is a contradiction. There is no difficulty in showing that a space with a countable base is separable. The relation between second countable spaces and separable spaces is given by the following theorem.

Theorem 1.24. *Let X be a topological space with a countable basis, then X is separable.*

Proof. Let $\mathscr{B}=\{B_i|\ i=1,2,\dots\}$ be a countable base for X, and define $A=\{x_i|\ x_i\in B_i,\ i=1,2,\dots\}$, i.e., $A\cap B_i=\{x_i\}$. We prove now that $A^-=X$. Let $x\in X$, if $x=x_i$ for some i, then $x\in A\subseteq A^-$, so assume that $x\neq x_i$ for each i. Let $U\in\mathscr{U}_x$, then there exists $B_i\in\mathscr{B}$, such that $x\in B_i\subseteq U$. Now $x_i\in B_i\subseteq U$, and $x_i\neq x$, thus every neighborhood U of x contains a point of A distinct from x, whence $x\in A^-$. Thus $X\subseteq A^-$, and since $A\subseteq X$, $A^-\subseteq X^-=X$, we have $A^-=X$, and A is the required countable dense subset. □

1.2.5 Subspace Separation and Connected Sets

Definition 1.23. Let $(X,\ \mathscr{T}_X)$ be a topological space, and let $Y\subseteq X$. The *relative topology* of $\mathscr{T}_X$ to Y $\mathscr{T}_Y$ is defined to be the family of all intersections of members of $\mathscr{T}_X$ with Y; that is, $U\in\mathscr{T}_Y$ if and only if there exists $V\in\mathscr{T}_X$ such that $U=V\cap Y$. The topological space $(X,\ \mathscr{T}_Y)$ is called a subspace of the space $(X,\ \mathscr{T}_X)$.

It is not difficulty to see that $\mathscr{T}_Y$ is actually a topology. Each member U of relative topology $\mathscr{T}_Y$ is said to be open in Y. It is worth noticing that, in the above definition, a subset Y of a space X is not necessarily a subspace. Only if the topology of Y agrees with relative topology of $\mathscr{T}_X$ to Y, then Y is called a subspace.

Example 1.16. Let $N=\{(x,\ y)\mid x,\ y$ real numbers, $y\geq 0\}$, i.e., N is the closed upper half of the real plane, and let

$$N^0=\{(x,\ y)\mid x,\ y \text{ real numbers}, y>0\}.$$

For $\{x,\ y\}\in N$, define $\mathscr{V}_{(x,y)}=\mathscr{U}_{(x,y)}\cap N$ if $y>0$, where $\mathscr{U}_{(x,y)}$ is the neighborhood system for $(x,\ y)$ in the usual topology for the real plane, and define

$$\mathscr{V}_{(x,y)}=\{V\mid V\supseteq(U\cap N^0)\cup\{(x,\ y)\}\}$$

for $U\in\mathscr{U}_{(x,y)}$ if y=0. Let

$$\mathscr{T}=\{\mathscr{V}_{(x,y)}|\ (x,\ y)\in N\},$$

with $\mathscr{V}_{(x,y)}$ as so defined, then $(N, \mathscr{T})$ is a topological space. N is not a subspace of the real plane with the usual topology, nor is the real line with its usual topology a subspace of N. The real line with the discrete topology is a subspace of N, where we think of the real line here as the set $R=\{(x, y) \mid y = 0\}$.

The following theorem formulates a criterion to recognize the closed sets, closure and accumulation points in a subspace.

Theorem 1.25. *Let $(X, \mathscr{T}_X)$ be a topological space and $(Y, \mathscr{T}_Y)$ a subspace of X. Then the following assertions hold.*

(1) The set A is closed in space Y if and only if it is the intersection of Y and a closed set in space X;
(2) A point y of space Y is an accumulation point of $A \subseteq Y$ if and only if it is an accumulation point of A in space X;
(3) The closure of A in space Y is the intersection of Y and the closure of A in space X.

Proof. The set A is closed in Y if and only if its relative complement $Y - A$ is of the form $V \cap Y$ for some open set V in space X, but this is true if and only if $A=(X - V) \cap Y$ for some open set V in space X. This proves (1). (2) follows directly from the definition of the relative topology and the definition of accumulation point. The closure of A in space Y is the union of A and the set of its accumulation points in space Y, and hence by (2) it is the intersection of Y and the closure of A, thus (3) holds. □

Definition 1.24. Let $(X, \mathscr{T})$ be a topological space and A, B be subsets of X. Two sets A and B are called separated in space X if and only if both $A^- \cap B$ and $A \cap B^-$ are empty. $(X, \mathscr{T})$ is called a *connected topological space* if and only if X is not the union of two nonempty separated subsets. A subset Y of X is called *connected* if and only if the topological space Y with the relative topology is connected. A component of a topological space is a maximal connected subset; that is, a connected subset which is properly contained in no other connected subset.

The separation involves the closure operation in X. However, the apparent dependence on the space X is illusory, for A and B are separated X if and only if neither A nor B contains a point or an accumulation point of the other. This condition may be restated in terms of the relative topology for $A \cup B$, in view of (2) of Theorem 1.25, as both A and B are closed in subspace $A \cup B$ (or equivalently A (or B) is both open and closed in subspace $A \cup B$) and A and B are disjoint. A set Y is connected if and only if the only subsets of Y which are both open and closed in subspace Y are Y and the empty set.

Example 1.17. The open interval (0, 1) and (1, 2) are disjoint subsets of R the real numbers and there is a point, 1, belonging to the closure of both (0, 1) and (1, 2) in the usual topology of R. However, (0, 1) is not disjoint with the closed interval $[1, 2]=(1, 2)^-$ because of 1, which is a member of [1, 2], is an accumulation point of (0, 1).

The real numbers, with the usual topology are connected (exercise 8), but the rational numbers, as a subspace of the usual topology of R, are not connected. For any irrational number $a \in R$ the sets $\{x \mid x < a\}$ and $\{x \mid x > a\}$ are separated in the subspace the rational numbers.

Theorem 1.26. *The closure of a connected set is connected.*

Proof. Suppose that Y is a connected subset of a topological space and that $Y^- = A \cup B$, where A and B are both open and closed in Y^-, A and B are disjoint. Then each of $A \cap Y$ and $B \cap Y$ is open and closed in Y, and since Y=$(A \cap Y) \cup (B \cap Y)$ and Y is connected, one of these two sets must be empty. Suppose that $B \cap Y$ is empty. Then Y is a subset of A and consequently Y^- is a subset of A because A is closed in Y^-. Hence B is empty, and it follows that Y^- is connected. □

Theorem 1.27. *Let $\mathscr{A}$ be a family of connected subsets of a topological space. If no two members of $\mathscr{A}$ are separated, then $\bigcup_{A \in \mathscr{A}} A$ is connected.*

Its proof remains as an exercise. If a space is connected, then it is its only component. If a space is discrete, then each component consists of a single point. Of course, there are many spaces which are not discrete which have components consisting of a single point, for instance, the space of rational numbers, as the subspace of the usual topology for the real numbers.

Theorem 1.28. *Each connected subset of a topological space is connected in a component, and each component is closed. If A and B are distinct components of a space, then A and B are separated.*

Its proof remains as an exercise. It is well to end our remarks on components with a word of caution. If two points, x and y, belong to the same component of a topological space, then they always lie in the same part of a separation of the space.

1.2.6 Convergence and Hausdorff Spaces

In this section, it will turn out that the topology of space can be described completely in terms of convergence. We also characterize those notions of convergence which can be described as convergence relative to some topology. Sequential convergence furnishes the pattern on which the theory is developed, and we therefore list a few definitions and theorems on sequences to indicate this pattern. These will be particular cases of the theorems proved later. A sequence is a map on the set N of the non-negative integers. The value of a sequence S at $n \in N$ is denoted, interchangeably, by S_n or $S(n)$. By Definition 1.10, we can verify that N is a directed set for the order of integers and the sequence $\{S_n \mid n \in N\}$ is a net. A sequence S is in a set A if and only if $S_n \in A$for each $n \in N$, and S is eventually in A if and only if there is $m \in N$ such that $S_n \in A$ whenever $n \geq m$. A sequence S is frequently in a set A if and only if for each non-negative integer m there is an integer n such that $n \geq m$ and $S_n \in A$. This is precisely the same thing as saying that S is not eventually in A'.

Definition 1.25. Let X be a topological space, let $x \in X$, and let $\{x_n |\ n= 1, 2, \dots\}$ be a sequence of points in X. The sequence $\{x_n\}$ is said to *converge* to x, and x is said to be a *limit* of the sequence $\{x_n\}$ if and only if for each $U \in \mathscr{U}_x$, there exists an integer m such that $n \geq m$ implies $x_n \in U$. A point s is a cluster point of a sequence S if and only if S is frequently in each neighborhood of s.

Let S be a sequence. T is said to be a subsequence of sequence S if and only if there is a sequence I of non-negative integers such that $T(i) = S(I(i))$ for each $i \in N$. It is clear that each cluster point of a sequence is a limit point of a subsequence, and conversely each limit point of a subsequence is a cluster point of a sequence.

Definition 1.26. Let $(X, \mathscr{T})$ be a topological space and $(S, \geq)$ be a net in X. The net $(S, \geq)$ is said to *converge* to s relative to $\mathscr{T}$ if and only if it is eventually in each neighborhood of s for topology $\mathscr{T}$.

It is easy to describe the accumulation points of a set, the closure of a set, and in fact the topology of a space in terms of convergence.

Theorem 1.29. *Let $(X, \mathscr{T})$ be a topological space. Then the following assertions hold.*

(1) A point s is an accumulation of a subset A of A if and only if there is a net in $A - \{s\}$ which converges to s;
(2) A point s belongs to the closure of a subset A of X if and only if there is a net in A converging to s;
(3) A subset A of X is closed if and only if no net in A converges to a point of $A - \{s\}$.

Proof. If s is an accumulation point of A, then for each neighborhood U of s there is a point $s_U \in U - \{s\}$. U$_s$, the neighborhoods of s, is directed by $\subseteq$, and if $U, V \in \mathscr{U}_s$ such that $V \subseteq U$, then $s_U \in V \subseteq U$. The net $\{\ s_U, U \in \mathscr{U}_s, \subseteq\}$, therefore converges to s. On the other hand, if a net in $A - \{s\}$ converges to s, then this net has point in every neighborhood of s and $A - \{s\}$ surly intersects each neighborhood of s. This establishes the statement (1). To prove (2), recall that the closure of a set A consists of A and together with all accumulation points of A. For each accumulation points s of A there is, by the preceding, a net in A converging to s; for each point s of Aany net whose point at every element of its domain is s converges to s. Therefore each point of the closure of A has a net in A converging to it. Conversely, if there is a net in A converging to s, then every neighborhood of s intersects A and s belongs to the closure of A. Assertion (3) is now obvious. $\square$

Example 1.18. Let E be the real plane, and define for $\varepsilon > 0$,

$$S_\varepsilon(x, y) = \{(u, v) \mid (u, v) \in E,\ |x - u| < \varepsilon\}$$

and

$$\mathscr{U}_{(x,y)} = \{U \mid U \supseteq S_\varepsilon(x, y) \text{ for some } \varepsilon > 0\}.$$

By Definition 1.13 we can verify that $\mathscr{U}_{(x,y)}$ is a neighborhood system for each $(x, y) \in E$. Then $\{\mathscr{U}_{(x,y)} | (x, y) \in E\}$ determines a topology $\mathscr{T}$ for E by Theorem 1.12. Let $\{(x_n, y_n)\}$ be a sequence in E, with the topology, $\mathscr{T}$, and let (x_0, y_0) be a limit of the sequence $\{(x_n, y_n)\}$, then (x_0, z) for any z is also a limit of the sequence $\{(x_n, y_n)\}$, and observe thus that limits of sequence need not be unique.

We have noticed that, in general, a net in a topological space may converging to several different points. Thus introduce a new kind of space, with a rather stronger structure.

Definition 1.27. Let $(X, \mathscr{T})$ be a topological space, then $\mathscr{T}$ is said to be a *Hausdorff (T_2−space, or separated space) topology* for X, provided that for each pair, x, y, with $x \neq y$, of points of X, there exist $U \in \mathscr{U}_x$, $V \in \mathscr{U}_y$ such that $U \cap V = \varnothing$. If this condition is satisfied we call X a *Hausdorff* (T_2 or separated) space.

Example 1.19. The real line with the usual topology is a Hausdorff space. The real plane with the usual topology also is a Hausdorff space. But the real plane with the topology defined in Example 1.17 is not a Hausdorff space.

Hausdorff spaces have the property that limits of sequences are unique, as in fact is shown by the following Theorem whose proof remains as a exercise.

Theorem 1.30. *A topological space is a Hausdorff space if and only if each net in the space converges to at most one point.*

It is of some interest to know when a topology can be described in terms of sequences alone, not only because it is a convenience to have a fixed domain for all nets, but also because there are properties of sequences which fail to generalize. The most important class of topological spaces for which sequential convergence is adequate are those satisfying the first countability axiom: the neighborhood system of each point has countable base. That is, for each point x of the space X there is a countable family of neighborhoods of x such that every neighborhood of x contains some member of family. In this case we may replace "net" by "sequence" in almost all of the preceding theorems.

Theorem 1.31. *Let X be a topological space satisfying the first axiom of countability. Then the following assertions hold.*

(1) A point s is an accumulation point of a set A of X if and only if there is a sequence in $A - \{s\}$ which converges *to s.*
(2) A set A is open if and only if each sequence which converges to a point of A is eventually in A.
(3) If s is a cluster point of a sequence S there is a subsequence of S converging to s.

Proof. Suppose that s is an accumulation point of a subset A of X, and that U_0, $U_1, \ldots U_2, \ldots$ is a sequence which is a base for the neighborhood system of s. Let $V_n = \cap_{0 \le i \le n} U_i$. Then the sequence $V_0, V_1, \ldots V_2, \ldots$ is also a sequence which is a base for the neighborhood system of s and, moreover, $V_{n+1} \subseteq V_n$ for each n. For each n select a point $s_n \in V_n \cap (A - \{s\})$, thus obtaining a sequence $\{s_n, n \in N\}$ which evidently converges to s. This establishes half of (1), and the converse is obvious. If A is a subset of X which is not open, then there is a sequence in $X - A$ which converges to a point of A. Such a sequence surely fails to be eventually in A, and part (2) follows. Finally, suppose that s is a cluster point of a sequence S and that $V_0, V_1, \ldots V_2, \ldots$ is a sequence which is a base for the neighborhood system of s such that $V_{n+1} \subseteq V_n$ for each n. For every non-negative integer I, choose N_i such that $N_i \geq i$ and s_{Ni} belongs to V_i. Then surely $\{s_{Ni}, n \in N\}$ is a subsequence of S which converges to s. □

1.2.7 Various Special Types of Topological Spaces

In this section, we shall investigate briefly a variety of special types of topological spaces. These spaces will play a fundamental role in the application of topology to the pattern recognition.

Definition 1.28. Let X be a set, $Y \subseteq X$, and let

$$\{D_\alpha | \ \alpha \in A, \textit{an indexing set}\}$$

be a family of subsets of X, then $\{D_\alpha\}$ is called a *cover* or *covering* for Y provided $\bigcup_{\alpha \in A} D_\alpha \supseteq Y$.

Definition 1.29. Let X be a topological space, then X is said to be *compact* provided each open cover of X contains a finite cover. (Here “open” refers to a property of the set D_α, while “finite” refers to a property of the indexing set A.)

A simple, but quite useful, consequence of Definition 1.29 comes as the following theorem.

Theorem 1.32. *Let X be a topological space, then X is compact if and only if for each family of closed sets $\{C_\alpha | \ \alpha \in A\}$ of X, $\bigcap_{\alpha \in A} C_\alpha = \varnothing$ implies that there exists $F \subseteq A$, F finite, such that $\bigcap_{\alpha \in F} C_\alpha = \varnothing$.*

Proof. Let X be compact, and let $\{C_\alpha | \ \alpha \in A\}$ be a family of closed sets with vacuous intersection. Define $O_\alpha = C'_\alpha$, then O_α is open, and

$$\bigcup_{\alpha \in A} O_\alpha = \bigcup_{\alpha \in A} C'_\alpha = [\bigcap_{\alpha \in A} C_\alpha]' = \varnothing' = X,$$

hence by Definition 1.29 there exists $F \subseteq A$, F finite, such that $\cup_{\alpha \in F} O_\alpha = X$, and

$$\varnothing = X' = [\cap_{\alpha \in F} O_\alpha]' = \cap_{\alpha \in F} O'_\alpha = \cap_{\alpha \in F} C_\alpha.$$

Similarly, the converse can be proved. □

Definition 1.30. Let X be a set and $\{D_\alpha | \ \alpha \in A\}$ be a family of subsets of X. Then $\{D_\alpha\}$ is said to have the *finite intersection property* provided that for any finite, non-empty subset F of A, $\bigcap_{\alpha \in F} D_\alpha \neq \varnothing$.

Theorem 1.33. *A topological space, X, is compact if and only if for any family $\{D_\alpha |\ \alpha \in A\}$ of closed sets with the finite intersection property.*

Theorem 1.34. *Let X be a topological space, then X is compact if and only if any family $\{D_\alpha |\ \alpha \in A\}$ of subsets of X with the finite intersection property has the further property that $\bigcap_{\alpha \in A} D_\alpha^- \neq \varnothing$.*

We now introduce a sequence of axioms, called separation axioms, into the definition of a space as follows

Definition 1.31. Let $(X, \mathscr{T})$ be a topological space, then X is said to be a T_i *space*, provided it satisfies Axiom T_i, $i = 0, 1, 2, 3, 4$, where the axioms are as follow.

Axiom T$_0$: For each x and $y \in X$, $x \neq y$, either there exists $U \in \mathscr{U}_x$ such that $y \notin U$, or there exists $V \in \mathscr{U}_y$ such that $x \notin V$.
Axiom T$_1$: For each x and $y \in X$, $x \neq y$, there exist $U \in \mathscr{U}_x$ and $V \in \mathscr{U}_y$ such that $y \notin U$ and $x \notin V$.
Axiom T$_2$: For each x and $y \in X$, $x \neq y$, there exist $U \in \mathscr{U}_x$ and $V \in \mathscr{U}_y$ such that $U \cap V = \varnothing$.
Axiom T$_3$: For each $x \in X$ and each closed set $C \subseteq X$, $x \notin C$, there exist $U \in \mathscr{U}_x$ and $O \in T$ such that $C \subseteq O$ and $O \cap U = \varnothing$.
Axiom T$_4$: For each pair of closed disjoint sets, $C, D \subseteq X$, there exists a pair O_1, $O_2 \in T$ such that $C \subseteq O_1$, $D \subseteq O_2$, and $O_1 \cap O_2 = \varnothing$.

A space which is a T_2 space is called a Hausdorff space. A space which is at one and the same time a T_1 and a T_2 space is called a *regular* space. A space which is at one and the same time a T_2 and a T_4 space is called a *normal* space.

Theorem 1.35. *A space X is a T_1 space if and only if each point is closed.*

Proof. Let $x \in X$, then for each $y \in X$, $y \neq x$, select $U \in \mathscr{U}_y$ such that $x \notin U$. Then $X - \{x\} \supseteq U$, whence $X - \{x\} \in \mathscr{U}_y$ for each $y \in X - \{x\}$, whence by Definition 1.12, $X - \{x\}$ is open, and $\{x\}$ is closed. Conversely, let $x \in X$, $y \in X$, and $x \neq y$, then since $\{x\}$ is closed, $X - \{x\}$ is open, and since $y \in X - \{x\}$, $X - \{x\} \in \mathscr{U}_y$, and by an identical argument $X - \{y\} \in \mathscr{U}_x$, whence X is a T_1 space. □

Theorem 1.36. *Each of the following properties of topological spaces is stronger than the next: Normality, Regularity, T_2 (Hausdorff), T_1, T_0, in the sense that if a space satisfies the definitions of any one of these properties, it also satisfies the definitions for all of the following ones as well.*

1.3 Metrization

In this section the elementary properties of metric and pseudo-metric spaces are developed, and necessary and sufficient conditions are given under which a space is copy of a metric space or of a subspace of the Cartesian product of intervals.

1.3.1 Continuous Functions

The words "*function*", "*map*", "*correspondence*", "*operator*" and "*transformation*" are synonymous.

Definition 1.32. A map f of a topological space $(X, \mathscr{T}_1)$ into a topological space $(Y, \mathscr{T}_2)$ is *continuous* if and only if the inverse of each open set is open. More precisely, f is continuous with respect to $\mathscr{T}_1$ and $\mathscr{T}_2$ if and only if $f^{-1}(U) \in \mathscr{T}_1$ for each $U \in \mathscr{T}_2$.

The concept of continuity depends on the topology of both the range and the domain space, but we follow the usual practice of suppressing all mention of the topologies when confusion is unlikely. The following is a list of conditions, each equivalent to continuity; it is useful because it is frequently necessary to prove functions continuous.

Theorem 1.37. *If X and Y are topological space and f is a function on X to Y, then the following statements are equivalent.*

(1) The function f is continuous.
(2) The inverse of each closed set is closed.
(3) The inverse of each member of a subbase for the topology for Y is open.
(4) For each x in X the inverse of every neighborhood of $f(x)$ is a neighborhood of x.
(5) For each x in X and each neighborhood U of $f(x)$ there is a neighborhood V of x such that $f(V) \subseteq U$.
(6) For each net S (or $\{S_n, n \in D\}$) in X which converges to a point s, the composition $f \cdot S$ (i.e., $\{f(S_n), n \in D\}$) converges to $f(s)$.
(7) For each subset A of X the image of closure is a subset of the closure of the image; that is, $f(A^-) \subseteq f(A)^-$.
(8) For each subset B of Y, $f^{-1}(B)^- \subseteq f^{-1}(B^-)$.

Proof. (6)⇒ (7) Assuming (6), let A be a subset of X and $s \in A^-$. Then there is a net Sin A which converges to s, and $f{\cdot}S$ converges to $f(s)$, which is therefore a member of $f(A)^-$. Hence $f(A^-) \subseteq f(A)^-$.

(7)⇒ (8) Assuming (7), if $A = f^{-1}(B)$, then $f(A^-) \subseteq f(A)^- \subseteq B^-$ and hence $A^- \subseteq f^{-1}(B^-)$. That is $f^{-1}(B)^- \subseteq f^{-1}(B^-)$.

(8)⇒ (2) Assuming (8), if B is a closed subset of Y, then $f^{-1}(B)^- \subseteq f^{-1}(B^-) = f^{-1}(B)$ is therefore closed. The proofs of other parts remain as exercises. □

There is also a localized form of continuity which is useful. A function f on a topological space X to a topological space Y is *continuous at point x* if and only if the inverse under f of each neighborhood of $f(x)$ is a neighborhood of x. A *homeomorphism* , or *topological transformation*, is a continuous one-to-one map of a topological space X onto a topological space Y such that f^{-1} is also continuous. If there exists a homeomorphism of one space onto another, the two spaces are said to be homeomorphic and each is a homeomorphism of the other. Consequently the collection of topological spaces can be divided into equivalence classes such that each topological space is homeomorphic to every member of its equivalence class and to these spaces only. Two topological spaces are topologically equivalent if and only if they are homeomorphic.

Two discrete spaces, X and $Y$with finite number of elements, are homeomorphic if and only if there is a one-to-one function on X onto Y, that is, if and only if X and Yhave the same number of elements. This is also true for the indiscrete topologies of X and Y. The set of all real numbers, with the usual topology, is homeomorphic to the open interval (0, 1), with the relative topology, for the function $f(x)=\frac{(2x-1)}{x(x-1)}$ which is easily proved to be a homeomorphism. However, the interval (0, 1) is not homeomorphic to (0, 1)∪(1, 2). Because if f were a homeomorphism (or, in fact, just a continuous function) on (0,1) with range then (0, 1)∪(1, 2), then f^{-1}((0, 1)) would be a proper open and closed subset of (0, 1), and (0, 1) is connected. This demonstration was achieved by noticing that one of the space is connected, the other is not, and the homeomorphism of the space is connected, the other is not, and the homeomorphism of a connected space is again connected. A property which when possessed by a topological space is also possessed by each homeomorphism is a *topological invariant*.

1.3.2 Metric and Pseudo-metric Spaces

We concentrate now to a rather special sort of topological space, one in which there is defined a distance function or a pseudo-distance function, so that we can say what the distance between points is. In a sense these spaces, so-called metric spaces, are rather special, since, as well turn out in the sequel, they will enjoy properties of the nature we have already discussed, but under less restrictive hypotheses than more general spaces. On the other hand, metric spaces are still quite general, since all the common spaces of analysis are metric spaces. We define a metric space in two stages. We start with the following facts.

Definition 1.33. Let X be a set and $\rho : X\times X \to R$ be a function (not necessarily continuous) of $X\times X$ into the non-negative real numbers R^+. ρ is called a *distance* on X and (X,ρ) is called *a metric space* if for any $x, y, z \in X$, ρ satisfies the following conditions.

(1) $\rho(x, y)$=$\rho(y, x)$;
(2) $\rho(x, y)$+$\rho(y, z) \geq \rho(x, z)$; (*triangle inequality*)
(3) $\rho(x, y)$=0 if $x=y$, and
(4) if $\rho(x, y)$=0, then $x=y$.

ρ is called a *pseudo-distance* on X and (X, ρ) is called *a pseudo-metric space* if ρ satisfies only (1), (2) and (3).

Every set is a metric set if ρ is defined as follows: $\rho(x, y)$=1 for $x\neq y$, and $\rho(x, y)$=0 for $x=y$. It is a simple verification that with this definition of ρ, any set X is a metric set. In order to generate the topology for X via the metric, we give the following definition.

Definition 1.34. Let $(X,\ \rho)$ be a metric (pseudo metric) space. If r is a positive number, then for $x \in X$, the following sets $S_r(x)$ and $S_r^-(x)$ are called the *open sphere* of ρ–radius r about x and *closed sphere* of ρ–radius r about x respectively

$$S_r(x) = \{y \mid \rho(x,y) < r\ for\ y \in X\},$$
$$S_r^-(x) = \{y \mid \rho(x,y) \leq r\ fory \in X\}.$$

X with the following $\mathscr{B}_\rho$ as basis is called a *metric space* (*pseudo metric space*). The topology so generated is called a *metric topology* (*pseudo metric topology*) generated by ρ

$$\mathscr{B}_\rho = \{S_r(x) \mid x \in X,\ r > 0\}.$$

Now only the rankest of amateurs at mathematics tries to prove definitions; however, the above definition makes some assertions which must be verified. In particular, it is asserted that B_ρ is a basis, and this is perhaps not so evident without some proof. We deal with this minor matter in the follows.

Theorem 1.38. *Let $(X,\ \rho)$ be a metric (pseudo metric) space. Then the set $\mathscr{B}_\rho$ $=\{S_r(x) \mid x \in X,\ r > 0\}$ is a basis for some topology for X.*

Proof. We shall apply Theorem 1.21. Since $x \in S_r(x)$ for each $x \in X$clearly we have $X = \cup_{B \in \mathscr{B}} B$. Now let $x \in X$ and let $S_{r_1}(x_1)$, $S_{r_2}(x_2) \in \mathscr{B}$ such that $x \in S_{r_1}(x_1) \cap S_{r_2}(x_2)$. It is clear that $\rho(x,x_1) = d_1 < r_1$ and $\rho(x,x_2) = d_2 < r_2$. Let e=min$\{r_1 - d_1,$ $r_2 - d_2\}$. We consider $S_e(x)$. For any $y \in S_e(x)$, we have $\rho(x,\ y) = e$. Now

$$\rho(y,x_1) \leq \rho(y,x) + \rho(x,x_1) < e + d_1 \leq r_1 - d_1 + d_1 = r_1,$$

So that $y \in S_{r_1}(x_1)$. Similarly, we can prove that $y \in S_{r_2}(x_2)$. Thus we have found $S_e(x) \in \mathscr{B}$ such that

$$x \in S_e(x) \subseteq S_{r_1}(x_1) \cap S_{r_2}(x_2),$$

and by Theorem 1.21, $\mathscr{B}_\rho$ is a basis. □

It can happen that a space X is already given us, and we may wish to know if it is possible to define a metric, ρ, such that the topology generated by the metric, using Definition 1.33, is in fact the same as the original topology. We define a space with this desirable property as follows.

Definition 1.35. Let $(X,\ \mathscr{T})$ be a topological space. If it is possible to define a distance (pseudo-distance) ρ such that the metric topology (pseudo-topology) generated by ρ coincides with $\mathscr{T}$, then X is said to be a *metrizable* (*pseudo metrizable*) topological space.

1.3.3 Some Properties of Metrizable and Pseudo Metrizable Space

One of the more interesting problems is to decide what sorts of topological spaces are metrizable or pseudo metrizable. Before we tackle this problem, however, let us explore some of the properties of metric spaces. Metric spaces have a fairly strong structure, and this is demonstrated by the following theorems.

Theorem 1.39. *Every metric space is a Hausdorff space.*

Proof. Let $x, y \in X$, with $x \neq y$, then $\rho(x, y) = d > 0$. Let ε=$d/2$ and let $U = S_\varepsilon(x)$, $V = S_\varepsilon(y)$, then U and V are open sets, and we need only show $U \cap V = \varnothing$. Suppose that $U \cap V \neq \varnothing$. Then there exists a $z \in U \cap V$ and since $x \in U$, $\rho(z,x) < \varepsilon$. Similarly $\rho(y,z) < \varepsilon$. Thus

$$\rho(x,y) \leq \rho(x,z) + \rho(y,z) < 2\varepsilon = d,$$

And $d = \rho(x,y) < d$. This palpable contradiction shows that $U \cap V = \varnothing$. Therefore X is Hausdorff. □

Theorem 1.40. *Every metrizable (pseudo metrizable) topological space is a T_4 space.*

Proof. We need thus only verify that if A and B are closed disjoint subset of X, then there exist open sets U, V such that $A \subseteq U$, $B \subseteq U$, and $U \cap V = \varnothing$.

Now for each $a \in A$, a is not a limit point of B, for if it were we should have $a \in B$, since B is closed. It contradicts A and B are disjoint. Thus for each $a \in A$, there exists an $\varepsilon_a > 0$ such that $S_{\varepsilon a}(a) \cap B = \varnothing$. Let $U_a = S_{\varepsilon a/2}(a)$. Similarly for each $b \in B$, there is an $\varepsilon_b > 0$ such that $S_{\varepsilon b}(b) \cap A = \varnothing$. Let $V_b = S_{\varepsilon b/2}(b)$. Finally let

$$U = \bigcup_{a\in A} U_a\ ,\ V = \bigcup_{b\in B} V_b\ ,$$

Then U and V are open.

We show now that $U \cap V = \varnothing$. Suppose that $U \cap V \neq \varnothing$, and let $x \in U \cap V$, then $x \in U_a$ for some $a \in A$ and $x \in V_b$ for some $b \in B$. Since $\rho(x,a) < \varepsilon_a/2$, $\rho(x,b) < \varepsilon_b/2$, hence

$$\rho(a,b) \leq \rho(a,x) + \rho(x,a) < (\varepsilon_a + \varepsilon_b)/2.$$

If $\varepsilon_a \leq \varepsilon_b$, then $\rho(a,b) < (\varepsilon_a + \varepsilon_b)/2 \leq \varepsilon_b$ and $a \in S_{\varepsilon b}(b)$. So that $a \in S_{\varepsilon b}(b) \cap A \neq \varnothing$, contrary to the definition of $S_{\varepsilon b}(b)$. On the other hand, if $\varepsilon_a \geq \varepsilon_b$, then $\rho(a,b) < (\varepsilon_a + \varepsilon_b)/2 \leq \varepsilon_a$ and $b \in S_{\varepsilon a}(a)$. So that $b \in S_{\varepsilon a}(a) \cap B \neq \varnothing$, contrary to the choice of $S_{\varepsilon a}$. It must be the case that $U \cap V = \varnothing$, and hence that X is T_4. □

By Theorem 1.39 we already know that if X is metrizable, then it is a Hausdorff space, hence T_2. Furthermore Theorem 1.41 shows that every metrizable topological space is a normal space. In what follows, we establish that every regular second countable space is metrizable. We require a good bit of machinery before we are ready to prove this result, and this is the purpose of the following theorem.

Theorem 1.41. *A T_1 space X is regular if and only if for each $x \in X$ and each $U \in \mathscr{U}_x$ there is a $V \in \mathscr{U}_x$ such that $V^- \subseteq U$.*

Theorem 1.42. *A T_1 space X is normal if and only if for each closed set C and each open set U such that $C \subseteq U$, there exists an open set V such that $C \subseteq V \subseteq V^- \subseteq U$.*

Definition 1.36. Let

$\mathscr{H} = \{y \mid y=\{y_n, n \in N^+\}$, y_n *is a real number for each in such that* $\sum_{n \in N} y_n^2 < \infty \}$

i.e., is the collection of all sequences of real numbers such that the series formed from the squares of the terms of the sequence is a convergent series. For $x, y \in \mathscr{H}$, the distance ρ on $\mathscr{H}$ is defined as follows

$$\rho(x,y) = \left(\sum_{n \in N+} (x_n - y_n)^2 \right)^{1/2}.$$

Then the resulting metrizable topological space is called a *Hilbert space*. Here $N^+ = N - \{0\}$.

The proofs of the following statements remain as exercises. The $\rho(x, y)$ in Definition 1.36 is a distance on $\mathscr{H}$; The subspace of $\mathscr{H}$ defined by

$$E^1 = \{x \mid x \in \mathscr{H}, x=\{x_n, n \in N^+\}, x_n=0 \text{ for } n > 1 \}$$

is homeomorphic to the real line with the usual topology; More generally $E^n \subseteq \mathscr{H}$, defined by

$$E^n = \{x \mid x \in \mathscr{H}, x=\{x_i, i \in N^+\}, x_i=0 \text{ for } i > n \}$$

is homeomorphic to R^n the n-dimension Euclidean space with the usual topology. The subspace $\mathscr{H}'$ of $\mathscr{H}$ defined by

$$\mathscr{H}' = \{x \mid x \in \mathscr{H}, x=\{x_n, n \in N^+\}, 0 \leq x_n \leq 1/n \text{ for each } n \in N^+ \}$$

is called the *Hilbert cub* (or *Hilbert parallelotrope*). Let I_n=[0, 1] for each $n \in N^+$, i.e., the unit interval with the relative topology inherited from the real numbers, and let $I^{N+} = \prod_{i \in I} I_i$, then $\mathscr{H}'$ is homeomorphic to I^{N+}. The key of the proof of the metrization theorem is to show that every second countable regular space is homeomorphic to a subset of Hilbert space (in fact of the Hilbert cube), that is, the space is metrizable. Now in order to define the appropriate mapping of our space X into $\mathscr{H}'$, we need to specify the terms, y_n, of the sequence $y \in \mathscr{H}'$ which is to be the image point of some preselected point $x \in X$. We thus need some mechanism for associating a sequence of real numbers $\{y_n \mid 0 \leq y_n \leq 1/n\}$ with each point of our space.

We wish to exploit the regularity of X, and specifically to make use of Theorem 1.41, which tells us that for each open set, consequently for each basic neighborhood B_i, of a point $x \in X$, there is another open set, which we may choose as a second basic neighborhood, such that $x \in B_j \subseteq B_j^- \subseteq B_i$. Then B_j^- and $X - B_i$, will be disjoint closed sets. If we consider all pairs of basic neighborhoods, (B_j, B_i) of X such that $B_j^- \subseteq B_i$, which set of pairs is countable (by the second countability of X), and if we can associate with each such pair a real-valued function λ_i such that $0 \le \lambda_i(x) \le 1$, we would at least be partially on our way. In order to accomplish this we prove first the following:

Lemma 1.1. *(Urysohn) Let X be a normal space, A, B closed disjoint subsets of X, then exists a map $f\colon X \to I$, I=[0, 1] such that $f(A)$=0, $f(B)$=1.*

Proof. Stage 1. Let $X - B = G_1$, an open set, since B is closed, then $A \subseteq G_1$ since $A \cap B = \varnothing$. By Theorem 1.42, there exists an open set $G_{1/2}$ such that $A \subseteq G_{1/2} \subseteq G^-_{1/2} \subseteq G_1$.

Stage 2. Again by Theorem 1.42, there exist open sets $G_{1/4}$ and $G_{3/4}$ such that

$$A \subseteq G_{1/4} \subseteq G^-_{1/4} \subseteq G_{1/2} \subseteq G^-_{1/2} \subseteq G_{3/4} \subseteq G^-_{3/4} \subseteq G_1.$$

Stage 3. Once again by Theorem 1.42, there exist open sets $G_{1/8}$, $G_{3/8}$, $G_{5/8}$ and $G_{7/8}$ such that

$$A \subseteq G_{1/8} \subseteq G^-_{1/8} \subseteq G_{1/4} \subseteq G^-_{1/4} \subseteq G_{3/8} \subseteq G^-_{3/8} \subseteq G_{1/2} \subseteq G^-_{1/2} \subseteq G_{5/8} \subseteq G^-_{5/8} \subseteq G_{3/4} \subseteq G^-_{3/4} \subseteq G_{7/8} \subseteq G^-_{7/8} \subseteq G_1$$

and so forth up to.

Stage N. By Theorem 1.42 for each odd integer $2i-1$, $1 \le 2i-1 \le 2^N-1$, there exists an open set $G_{(2i-1)/2^N}$ such that $A \subseteq G_{1/2^N}$ and $G^-_{(2i-2)/2^N} \subseteq G_{(2i-1)/2^N} \subseteq G^-_{(2i-1)/2^N} \subseteq G_{2i/2^N}$.

By induction we construct for each dyadic fraction number, t, between 0 and 1, i.e., for each fraction whose denominator is 2^n, $n \ge 0$, an open G_t such that if t and t' are two dyadic fractions that $t < t'$ if and only if $G^-_t \subseteq G_{t'}$.

Now for $x \in X$, define

$$f(x) = \begin{cases} \inf\limits_{x \in G_t} t & x \notin B \\ 1 & x \in B \end{cases}$$

Observe that $A \subseteq G_t$ for all t, thus $f(x) = 0$ if $x \in A$, and also note that $0 \le f(x) \le 1$.

We are left with the task of showing f continuous. Let us examine the structure of $f^{-1}([0,y))$ for $0 < y \le 1$. New $f(x) \in [0,y)$ provided $0 \le f(x) < y$, and since the dyadic fractions are dense in [0, 1] there exists a dyadic fraction t_0 such that

$$f(x) = \inf_{x \in G_t} t < t_0 < y$$

Consequently $x \in G_{t_0}$. On the other hand, if $t_0 < y$ and $x \in G_{t_0}$, $f(x) \in [0,y)$. We thus see that $f^{-1}([0,\, y))=\cup_{t<y} G_t$. Since G_t is open for each t, $f^{-1}([0,\, y))$ is open for each y. By a similar argument it is clear that $f^{-1}((y,\, 1])=\cup_{t>y}(X - G_t)$, where $0 \le y < 1$. Since $G_t \subseteq G^-_t$, $X - G^-_t \subseteq X - G_t$ for each t. Thus

$$\cup_{t>y}(X - G_t) \supseteq \cup_{t>y}(X - G^-_t).$$

How ever, if $x \in \cup_{t>y}(X - G_t)$, then there exists a $t > y$ such that $x \in X - G_t$, and again by the density of the dyadic fractions in [0,1] we may select t', a dyadic fraction, such that $t > t' > y$, then $G_{t'}^{-} \subseteq G_t$, and $X - G_{t'}^{-} \supseteq X - G_t$, so that $x \in X - G_{t'}^{-}$ for some $t' > y$, whence $x \in \cup_{t>y}(X - G_t^{-})$. Thus $\cup_{t>y}(X - G_t) \subseteq \cup_{t>y}(X - G_t^{-})$ and finally

$$\cup_{t>y}(X - G_t) = \cup_{t>y}(X - G_t^{-}).$$

Consequently, since each G_t^{-} is closed, hence each $X - G_t^{-}$ is therefore open, and we see that $f^{-1}((y, 1])$=$\bigcup_{t>y}(X - G_t^{-})$ is open.

Now let U be some open set in [0,1], such that $f(x) \in U$, then there exists a basic set V in [0,1] such that $f(x) \in V \subseteq U$, and the basic set V in the relative topology of [0,1] has one of the following forms: $[0,y), 0 < y < 1; (y,1], 0 < y < 1; (y_1,y_2), 0 < y_1 < y_2 < 1; [0,1]$. If $V = [0,y)$, then $f^{-1}(V)$ is open by what we have proved above. If $V = (y,1]$, then $f^{-1}(V)$ is open by what we have proved above, also. If $V = (y_1,y_2)$, then $V = [0,y_2) \cap (y_1,1]$ and $f^{-1}(V) = f^{-1}([0,y_2)) \cap f^{-1}((y_1,1])$ is open as the intersection of open sets, whence f is continuous. □

Theorem 1.43. *Every second countable regular space is homeomorphic to a subset of the Hilbert cube.*

Proof. Let $(X, \mathscr{T})$ be a second countable regular topological space and $\mathscr{B} = \{B_i \mid i = 1,2,\ldots\}$ be the countable basis. It is clear that $(X, \mathscr{T})$ is normal, thus by Theorem 1.42, there exists pairs (B_i, B_j) of elements of $\mathscr{B}$ such that $B_i^{-} \subseteq B_j$. Since $\mathscr{B}$ is countable, the collection of all such pairs is again countable. Let us call it $\mathscr{P} = \{P_n \mid n = 1,2,\ldots\}$ where $P_n = (B_i^n, B_j^n)$ and $B_i^{n-} \subseteq B_j^n$. Now since $B_i^{n-} \cap (X - B_j^n) = \varnothing$ and both B_i^{n-} and $(X - B_j^n)$ are closed, we may define, by Lemma 1.1, a map $f_n : X \to I$=[0,1] such that $f_n(B_i^{n-})$=0, $f_n(X - B_j^n)$=1. Finally define f: $X \to \mathscr{K}'$, the Hilbert cube, by for $x \in X$,

$$f(x) = \{f_n(x)/n \mid n = 1,2,\ldots\}.$$

Since for each $x \in X$, $0 \leq f_n(x) \leq 1$, hence $\sum_{n\in N+} (f_n(x)/n)^2 < \infty$ and $f(x) \in \mathscr{K}'$.

First we show f is one-to-one map. Let $x \neq y$, then since X is Hausdorff by Theorem 1.36, there exist open sets, which we may choose as basic sets, A_1, A_2 such that $x \in A_1$, $y \in A_2$, and $A_1 \cap A_2 = \varnothing$. Further, since X is normal we can find $A_0 \in \mathscr{B}$, such that $x \in A_0 \subseteq A_0^{-} \subseteq A_1$, then $x \in A_0^{-}$ and $y \in X - A_2$, and the pair $(A_0,A_1) \in \mathscr{P}$, i.e., for some n, $(A_0,A_1) = (B_i^n,B_j^n)$. Thus $f_n(x) = f_n(B_i^{n-}) = f_n(A_0^{-}) = 0$, while $f_n(y) = f_n(X - B_j^n) = f_n(X - A_1)$=1. This implies that $f(x) \neq f(y)$, since $f(x)$ differs from $f(y)$ at the nth place.

Now we prove that f is continuous. Let $x \in X$, and let $\varepsilon > 0$. We wish to construct $U \in \mathscr{U}_x$ in space X such that for any $y \in U$, $\rho(f(x), f(y)) < \varepsilon$ in metric subspace $\mathscr{K}'$. First, since for any point $y \in X$, $0 \leq f_n(y) \leq 1$, we have that $| f_n(x) - f_n(y)|^2 \leq 1$. Now the finite series $\sum_{n\in N+} n^{-2}$ converges, thus for N sufficiently large,

$$\sum_{n=N}^{\infty} n^{-2} < \frac{\varepsilon^2}{2}$$

thus
$$\sum_{n=N}^{\infty} |f_n(x) - f_n(y)|^2 n^{-2} \leq \sum_{n=N}^{\infty} n^{-2} < \frac{\varepsilon^2}{2}$$
Now let $k < N$, then the function $f_k : X \to I$ is continuous and there exists a $U_k \in \mathscr{U}_x$ such that $y \in U_k$ implies
$$|f_k(x) - f_k(y)| < \frac{k\varepsilon}{(2(N-1))^{1/2}}$$
Or
$$\frac{|f_k(x) - f_k(y)|^2}{k^2} < \frac{\varepsilon^2}{2(N-1)}.$$
Now let $U = \cap_{1 \leq k \leq N-1} U_k$, then if $y \in U$,
$$\sum_{n=1}^{\infty} \frac{|f_n(x) - f_n(y)|^2}{n^2} = \sum_{n=1}^{N-1} \frac{|f_n(x) - f_n(y)|^2}{n^2} + \sum_{n=N}^{\infty} \frac{|f_n(x) - f_n(y)|^2}{n^2}$$
$$< (N-1)\frac{\varepsilon^2}{2(N-1)} + \frac{\varepsilon^2}{2} = \varepsilon^2$$

And finally $\rho(f(x), f(y)) < \varepsilon$. Therefore f is continuous.

Finally we must show that f is an open mapping. Let Ube open in X, and let $x \in U$, then there exist $B_i, B_j \in \mathscr{B}$ such that
$$x \in B_i \subseteq B_i^- \subseteq B_j \subseteq U$$
by the normality of X and the fact that $\mathscr{B}$ is a basis. Thus the pair $(B_i, B_j) \in \mathscr{P}$, say $(B_i, B_j) = (B_i^n, B_j^n)$. Then $f_n(x) = f_n(B_i^{n-})$=0, and since $X - U \subseteq X - B_j^n$, hence $f_n(X - U) = f_n(X - B_j^n)$=1. Thus for any $y \in X - U$,
$$\rho(f(x), f(y)) = (\sum_{n=1}^{\infty} \frac{|f_n(x) - f_n(y)|^2}{n^2})^{1/2} \geq (\frac{|f_n(x) - f_n(y)|^2}{n^2})^{1/2} = \frac{1}{N}$$

So that if $V = S_{1/n}(f(x)) \subseteq \mathscr{K}'$, $z \in V$, then we have $\rho(f(x), z) < 1/n$ and $f^{-1}(z) \in U$, because if $f^{-1}(z) \in X - U$, then $\rho(f(x), f(f^{-1}(z))) \geq 1/n$ and it is a contradiction. Thus $f^{-1}(V) \subseteq U$, and $x \in V \subseteq f(U)$, whence $f(U)$ is open. Consequently we have proved that f is a one-to-one continuous open mapping, and f is a homeomorphism. □

1.4 Measures

In this section, we just briefly introduce some concepts and results of measure theory [7] which are necessary for the representation of fuzzy concepts in the AFS theory.

1.4.1 Algebras and Sigma-Algebras

Definition 1.37. Let X be an arbitrary set. A collection $\mathscr{A}$ of subsets of X is an *algebra* on X if the following conditions are satisfied

(1) $X \in \mathscr{A}$;
(2) For each set A that belongs to $\mathscr{A}$ the set A' belongs to $\mathscr{A}$;
(3) For each finite sequence $A_1, A_2, \ldots, A_n \in \mathscr{A}$, $\bigcup_{1 \le i \le n} A_i$, $\bigcap_{1 \le i \le n} A_i \in \mathscr{A}$.

Remark 1.2. In conditions (2), (3) we have required that $\mathscr{A}$ be closed under complementation, under the formation of finite unions, and under the formation of finite intersections. It is easy to check that closure under complementation and closure under the formation of finite unions together imply closure under the formation of finite intersections (use the fact that $\bigcap_{1 \le i \le n} A_i = (\bigcup_{1 \le i \le n} A_i')'$. Thus we could have defined an algebra using only $\bigcup_{1 \le i \le n} A_i \in \mathscr{A}$ in conditions (3).

Definition 1.38. Let X be an arbitrary set. A collection $\mathscr{A}$ of subsets of X is a σ-*algebra* on X if the following conditions are satisfied

(1) $X \in \mathscr{A}$.
(2) For each set A that belongs to $\mathscr{A}$ the set A' belongs to $\mathscr{A}$;
(3) For each infinite sequence $\{A_i,\, n \in N^+\}$ of sets that belong to $\mathscr{A}$, $\bigcup_{n \in N+} A_i$, $\bigcap_{n \in N+} A_i \in \mathscr{A}$.

A subset of X is called $\mathscr{A}$*-measurable* if it belongs to $\mathscr{A}$.

Remark 1.3. A σ-algebra on X is a family of subsets of X that contains X and is closed under complementation, under the formation of countable unions, and under the formation of countable intersections. As in the case of algebras, we could have used only $\bigcup_{n \in N+} A_i \in \mathscr{A}$ or $\bigcap_{n \in N+} A_i \in \mathscr{A}$ in conditions (3).

Next we consider ways of constructing σ-algebras.

Theorem 1.44. *Let X be a set. Then the intersection of an arbitrary non-empty family of σ-algebras on X is a σ-algebra on X.*

Proof. Let $\mathscr{A}$ be a non-empty family of σ-algebras on X, and let $\mathscr{A}$ be the intersection of the σ-algebras that belong to $\mathscr{L}$. It is enough to check that $\mathscr{A}$ contains X, is closed under complementation, and is closed under the formation of countable unions. The set X belongs to $\mathscr{A}$, since it belongs to each σ-algebra that belongs to $\mathscr{L}$. Now suppose that $A \in \mathscr{A}$. Each σ-algebra that belongs to $\mathscr{L}$ contains A and so contains A'; thus A' belongs to the intersection $\mathscr{A}$ of these σ-algebras. Finally, suppose that $\{A_i,\, n \in N^+\}$ is a sequence of sets that belong to $\mathscr{A}$, and hence to each σ-algebra in $\mathscr{L}$. Then $\cup_{n \in N+} A_i$ belongs to each σ-algebra in $\mathscr{L}$, and so to $\mathscr{A}$. □

Remark 1.4. The union of a family of σ-algebras can fail to be a σ-algebra.

Definition 1.39. Let $\mathscr{A}$ is a σ-algebra on X that includes $\mathscr{F} \subseteq 2^X$, and that every σ-algebra on X that includes $\mathscr{F}$ also includes $\mathscr{A}$, then $\mathscr{A}$ is called the *σ-algebra generated* by $\mathscr{F}$, and denoted by $\sigma(\mathscr{F})$. We also say that $\mathscr{A}$ is the smallest σ-algebra on X that includes $\mathscr{F}$.

It is evident that the smallest σ-algebra on X that includes $\mathscr{F}$ is unique.

Corollary 1.5. *Let X be a set, and let $\mathscr{F}$ be a family of subsets of X. Then there is a smallest σ-algebra on X that includes $\mathscr{F}$.*

We now use the preceding corollary to define an important family of σ-algebras. The *Borel σ-algebra* on R^d is the σ-algebra on R^d generated by the collection of open subsets of R^d, and is denoted by $\mathscr{B}(R^d)$. The *Borel subsets of R^d* are those that belong to $\mathscr{B}(R^d)$. In case $d = 1$, one generally writes $\mathscr{B}(R)$ in place of $\mathscr{B}(R^1)$.

Theorem 1.45. *The σ-algebra $\mathscr{B}(R)$ of Borel subsets of T is generated by each of the following collections of sets:*

(1) The collection of all closed subsets of R.
(2) The collection of all subintervals of R of the form $(-\infty, b]$.
(3) The collection of all subintervals of R of the form $(a, b]$.

Proof. Let $\mathscr{B}_1$, $\mathscr{B}_2$ and $\mathscr{B}_3$ be the σ-algebras generated by the collections of sets in parts (1), (2), and (3) of the theorem. We shall show that $\mathscr{B}(R) \supset \mathscr{B}_1 \supset \mathscr{B}_2 \supset \mathscr{B}_3$, and then that $\mathscr{B}_3 \supset \mathscr{B}(R)$; this will establish the theorem. Since $\mathscr{B}(R)$ includes the family of open subsets of R and is closed under complementation, it includes the family of closed subsets of R; thus it includes the σ-algebra generated by the closed subsets of R, namely $\mathscr{B}_1$. The sets of the form $(-\infty, b]$ are closed and so belong to $\mathscr{B}_1$; consequently $\mathscr{B}_2 \subset \mathscr{B}_1$. Since

$$(a,b] = (-\infty, b] \cap (-\infty, a]'$$

each set of the form $(a,b]$ belongs to $\mathscr{B}_2$; thus $\mathscr{B}_3 \subset \mathscr{B}_2$. Finally, note that each open subinterval of R is the union of a sequence of sets of the form $(a, b]$, and that each open subset of R is the union of a sequence of open intervals. Thus each open subset of R belongs to $\mathscr{B}_3$, and so $\mathscr{B}(R) \subset \mathscr{B}_3$. □

We should note the following properties of the σ-algebra $\mathscr{B}(R)$, it is largely these properties that explain the importance of $\mathscr{B}(R)$.

1. It contains virtually every subset of R that is of interest in analysis.
2. It is small enough that it can be dealt with in a fairly constructive manner.

Theorem 1.46. *The σ-algebra $\mathscr{B}(R^d)$ of Borel subsets of R^d is generated by each of the following collections of sets.*

(1) The collection of all closed subsets of R^d;

(2) The collection of all closed half-spaces in R^d that have the form

$$\{(x_1, x_2, \ldots, x_d) \mid x_i \leq b\}$$

for some index i and some b in R;

(3) The collection of all rectangles in R^d that have the form

$$\{(x_1, x_2, \ldots, x_d) \mid a_i < x_i \leq b_i \; for \; i = 1, 2, \ldots, d\}.$$

Proof. This theorem can be proved with essentially the argument that was used for Theorem 1.45, and so most of the proof is omitted. To see that the σ-algebra generated by the rectangles of part (3) is included in the σ-algebra generated by the half-spaces of part (2), note that each strip that has the form

$$\{(x_1, x_2, \ldots, x_d) \mid a_i < x_i \leq b_i\}$$

for some i is the deference of two of the half-spaces in part (2), and that each of the rectangles in part (3) is the intersection of d such strips. □

A sequence $\{A_i\}$ of sets is called *increasing* if $A_i \subseteq A_{i+1}$ holds for each i, and *decreasing* if $A_i \supset A_{i+1}$ holds for each i.

Theorem 1.47. *Let X be a set, and let $\mathscr{A}$ be an algebra on X. Then $\mathscr{A}$ is a σ-algebra if either the following (1) or (2) holds.*

(1) $\mathscr{A}$ is closed under the formation of unions of increasing sequences of sets;

(2) $\mathscr{A}$ is closed under the formation of intersections of decreasing sequences of sets.

Proof. First suppose that condition (1) holds. Since $\mathscr{A}$ is an algebra, we can check that it is a σ-algebra by verifying that it is closed under the formation of countable unions. Suppose that $\{A_i\}$ is a sequence of sets that belong to $\mathscr{A}$. For each n let $B_n = \bigcup_{1 \leq i \leq n} A_i$. The sequence $\{B_n\}$ is increasing, and, since $\mathscr{A}$ is an algebra, each B_n belongs to $\mathscr{A}$; thus assumption (1) implies that $\bigcup_{n \in Z+} B_n$ belongs to $\mathscr{A}$. However, $\bigcup_{n \in Z+} A_n$ is equal to $\bigcup_{n \in Z+} B_n$, and so belongs to $\mathscr{A}$. Thus $\mathscr{A}$ is closed under the formation of countable unions, and so is a σ-algebra.

Now suppose that condition (2) holds. It is enough to check that condition (1) holds. If $\{A_i\}$ is an increasing sequence of sets that belong to $\mathscr{A}$, then $\{A_i'\}$ is a decreasing sequence of sets that belong to $\mathscr{A}$, and so condition (2) implies that $\bigcap_{i \in Z+} A_i'$ belongs to $\mathscr{A}$. Since $\bigcup_{i \in Z+} A_i = (\bigcap_{i \in Z+} A_i')'$, it follows that $\bigcup_{i \in Z+} A_i$ belongs to $\mathscr{A}$. Thus condition (1) follows from condition (2), and the proof is complete. □

1.4.2 Measures

A *set function* is a function whose domain is a class of sets. An extended real valued set function μ defined on a family E of sets is *additive* if, whenever A, $B \in E$, $A \cup B \in E$, and $A \cap B = \varnothing$, then $\mu(A \cup B) = \mu(A) + \mu(B)$.

Definition 1.40. Let X be a set, and let $\mathscr{A}$ be a σ-algebra on X. A function μ whose domain is the σ-algebra $\mathscr{A}$ and whose values belong to the extended half-line [0, $+\infty$) is said to be *countably additive* if it satisfies

$$\mu(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k)$$

for each infinite sequence $\{A_i\}$ of disjoint sets that belong to $\mathscr{A}$. (Since $\mu(A_i)$ is non-negative for each i, the sum $\sum_{k=1}^{\infty} \mu(A_i)$ always exists, either as a real number or as $+\infty$) A *measure* (or a *countably additive measure*) on $\mathscr{A}$ is a function $\mu : \mathscr{A} \to [0, +\infty)$ that satisfies $\mu(\varnothing)$=0 and is countably additive.

Definition 1.41. Let $\mathscr{A}$ be an algebra (not necessarily a σ-algebra) on the set X. A function μ whose domain is $\mathscr{A}$ and whose values belong to $[0, +\infty)$ is *finitely additive* if it satisfies

$$\mu(\bigcup_{1 \le i \le n} A_i) = \sum_{1 \le i \le n} \mu(A_i)$$

for each finite sequence $A_1, A_2, \ldots, A_n$ of disjoint sets that belong to $\mathscr{A}$. A finitely additive measure on the algebra $\mathscr{A}$ is a function $\mu : \mathscr{A} \to [0, +\infty)$ that satisfies $\mu(\varnothing)$=0 and is finitely additive.

It is easy to check that every countably additive measure is finitely additive: simply extend the finite sequence $A_1, A_2, \ldots, A_n$ to an infinite sequence $\{A_i\}$ by letting $A_i = \varnothing$ if $i > n$, and then use the fact that $\mu(\varnothing)$=0. There are, however, finitely additive measures that are not countably additive. Finite additivity might at first seem to be a more natural property than countable additivity. However countably additive measures on the one hand seem to be sufficient for almost all applications, and on the other hand support a much more powerful theory of integration than do finitely additive measures. Thus we shall follow the usual practice, and devote almost all of our attention to countably additive measures.

Definition 1.42. Let X be a set, $\mathscr{A}$ be a σ-algebra on X, and μ is a measure on $\mathscr{A}$, then the triple $(X, \mathscr{A}, \mu)$ is called a *measure space* . Likewise, if X is a set and if $\mathscr{A}$ is a σ-algebra on X, then the pair $(X, \mathscr{A})$ is called a *measurable space*. If $(X, \mathscr{A}, \mu)$ is a measure space, then one says that μ is a *measure on* $(X, \mathscr{A})$, or, if the σ-algebra $\mathscr{A}$ is clear from context, a *measure on* X.

Theorem 1.48. *Let $(X, \mathscr{A}, \mu)$ be a measure space, and let A and B be subsets of X that belong to $\mathscr{A}$ and satisfy $A \subset B$. Then $\mu(A) \le \mu(B)$. If in addition A satisfies $\mu(A) < +\infty$, then $\mu(B - A) = \mu(B) - \mu(A)$.*

Proof. The sets A and $B - A$ are disjoint and satisfy $B = A \cup (B - A)$, thus the additivity of μ implies that $\mu(B) = \mu(A) + \mu(B - A)$. Since $\mu(B - A) \ge 0$, it follows that $\mu(B) \ge \mu(A)$. In case $\mu(A) < +\infty$, the relation $\mu(B) - \mu(A) = \mu(B - A)$ also follows. □

Let μ be a measure on a measurable space $(X, \mathscr{A})$. Then μ is a *finite measure* if $\mu(X) < +\infty$, and is a *σ-finite measure* if X is the union of a sequence $A_1, A_2, \ldots$ of sets that belong to $\mathscr{A}$ and satisfy $\mu(A_i) < +\infty$ for each i. More generally, a set in $\mathscr{A}$ is *σ-finite under* μ if it is the union of a sequence of sets that belong to $\mathscr{A}$ and have finite measure under μ. The measure space $(X, \mathscr{A}, \mu)$ is also called *finite* or *σ-finite* if μ is *finite or σ-finite*. The following theorems give some elementary but useful properties of measures.

Theorem 1.49. *Let $(X, \mathscr{A}, \mu)$ be a measure space. If $\{A_k\}$ is an arbitrary sequence of sets that belong to $\mathscr{A}$, then*

$$\mu(\bigcup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} \mu(A_k)$$

Proof. Define a sequence $\{B_k\}$ of subsets of X by letting $B_1 = A_1$ and letting $B_k = A_k - \cup_{1 \leq i \leq k-1} A_i$ if $k > 1$. Then each B_k belongs to A and is a subset of the corresponding A_k, and so satisfies $\mu(B_k) \leq \mu(A_k)$. Since, in addition, the sets B_k are disjoint and satisfy $\bigcup_{k=1}^{\infty} B_k = \bigcup_{k=1}^{\infty} A_k$, it follows that

$$\mu(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(B_k) \leq \sum_{k=1}^{\infty} \mu(A_k) \qquad \square$$

Theorem 1.50. *Let $(X, \mathscr{A}, \mu)$ be a measure space. The the following assertions hold.*

(1) If $\{A_k \mid k \in Z^+\}$ is an increasing sequence of sets that belong to $\mathscr{A}$, then

$$\mu(\bigcup_{k=1}^{\infty} A_k) = \lim_{k \to \infty} \mu(A_k).$$

(2) If $\{A_k \mid k \in Z^+\}$ is a decreasing sequence of sets that belong to $\mathscr{A}$, and if $\mu(A_n) < +\infty$ holds for some n, then

$$\mu(\bigcap_{k=1}^{\infty} A_k) = \lim_{k \to \infty} \mu(A_k).$$

Proof. (1) If we write $E_0 = \varnothing$, then

$$\begin{aligned}
\mu(\bigcup_{k=1}^{\infty} A_k) &= \mu(\bigcup_{k=1}^{\infty} (A_k - A_{k-1})) = \sum_{k=1}^{\infty} \mu(A_k - A_{k-1}) \\
&= \lim_{n \to \infty} \sum_{k=1}^{n} \mu(A_k - A_{k-1}) = \lim_{n \to \infty} \mu(\bigcup_{k=1}^{n} (A_k - A_{k-1})) \\
&= \lim_{n \to \infty} \mu(A_n)
\end{aligned}$$

(2) If $\mu(A_m) < \infty$, then $\mu(A_n) \leq \mu(A_m) < \infty$ for $n \geq m$, and $\{A_m - A_n \mid n \geq m\}$ is an increasing sequence, that is

$$\begin{aligned}\mu(A_m) - \mu(\bigcap_{n=1}^{\infty} A_n) &= \mu(A_m - \bigcap_{n=1}^{\infty} A_n) = \mu(\bigcup_{n=1}^{\infty}(A_m - A_n)) \\ &= \lim_{n\to\infty} \mu(A_m - A_n) = \lim_{n\to\infty} (\mu(A_m) - \mu(A_n)) \\ &= \mu(A_m) - \lim_{n\to\infty} \mu(A_n)\end{aligned}$$

Since $\mu(A_m) < \infty$, the proof is complete. □

We shall say that an extended real valued set function μ defined on a family ε is *continuous from below* at a set ε if for every increasing sequence $\{E_k\}$ of sets in ε for which $\bigcup_{k=1}^{\infty} E_k = E$, $\lim_{k\to\infty} \mu(E_k) = \mu(E)$. Similarly μ is *continuous from above* at E if for every decreasing sequence $\{E_k\}$ of sets in ε for which $|\mu(E_m)| < \infty$ for at least one value of m and for which $\bigcap_{n=1}^{\infty} E_n = E$, we have $\lim_{k\to\infty} \mu(E_k) = \mu(E)$. Theorem 1.50 assert that if μ is a measure, then μ is continuous from above and from below and had the following partial converse, which is sometimes useful for checking that a finitely additive measure is in fact countably additive.

Theorem 1.51. *Let $(X, \mathscr{A})$ be a measurable space, and let μ be a finitely additive measure on $(X, \mathscr{A})$. If either continuous from below at every E in $\mathscr{A}$, or continuous from above at $\varnothing$, then μ is a measure on $\mathscr{A}$.*

Proof. Let $\{E_n\}$ be a disjoint sequence of sets in $\mathscr{A}$, whose union, E, is also in $\mathscr{A}$ and write $F_n = \bigcup_{1\leq i\leq n} E_i$, $G_n = E - F_n$. If μ is continuous from below, then since $\{F_n\}$ is an increasing sequence of sets in $\mathscr{A}$ with $\bigcup_{k=1}^{\infty} F_k = E$, hence

$$\mu(E) = \lim_{n\to\infty} \mu(F_n) = \lim_{n\to\infty} \textstyle\sum_{1\leq i\leq n} \mu(E_i) = \sum_{i=1}^{\infty} \mu(E_i)$$

If μ is continuous from above at $\varnothing$, then, since $\{G_n\}$ is a decreasing sequence of sets in $\mathscr{A}$ with $\bigcap_{n=1}^{\infty} G_n = \varnothing$, and μ is finitely additive, we have

$$\begin{aligned}\mu(E) &= \mu(F_n \cup G_n) = \left(\sum_{1\leq i\leq n} \mu(E_i)\right) + \mu(G_n) \\ &= \lim_{n\to\infty} \sum_{1\leq i\leq n} \mu(E_i) + \lim_{n\to\infty} \mu(G_n) = \sum_{1\leq i\leq\infty} \mu(E_i).\end{aligned}$$ □

1.5 Probability

The subject of probability theory is the foundation upon which all of statistics [1, 3] is built, providing a means for modeling populations, experiments, or almost anything else that could be considered related to a random phenomenon. Through these models, statisticians are able to draw inferences about populations, inferences based on examination of only a part of the whole.

1.5.1 Probability Space and Probability Measure

The set, S, of all possible outcomes of a particular experiment is called the *sample space* for the experiment. An *event* is any collection of possible outcomes of an experiment, that is, any subset of S (including S itself).

Consider the experiment of rolling an ordinary six-sided die and observing the number x=1, 2, ..., 5, or 6 showing on the top face of the die. "*the number x is even*", "*it is less than* 4", "*it is equal to* 6" each such statement corresponds to a possible outcome of the experiment. From this point of view there are as many events associated with this particular experiment as there are combinations of the first six positive integers taken any number at time. If, for the sake of aesthetic completeness and later convenience, we consider also the impossible event, "the number x is not equal to any of the first six positive integers," then there are altogether 2^6 admissible events associated with the experiment of the rolling die. We write the set $\{2, 4, 6\}$ for the event "*x is even*", $\{1, 2, 3\}$ for "*x is less than* 4" and so on. An event is a set, and its opposite event is the complementary set; mutually exclusive events are disjoint sets, and an event consisting of the simultaneous occurrence of two other events is a set obtained by intersecting two other sets.

For situations arising in modern theory and practice, and even for the more complicated gambling games, it is necessary to make an additional assumption. This assumption is that the system of events is closed under the formation of countably infinites unions, or, in the technical language we have already used σ-algebra. When we ask "what is the probability of a certain event ? ", we expect the answer to be a number, a number associated with the event. In other words, probability is a numerically valued function P of the event E, which is in $\mathscr{A}$ the σ-algebra on a sample space X. On intuitive and practical grounds we demand that the number $P(E)$ should give information about the occurrence habits of the event E.

If, to begin with, $P(E)$ is to represent the proportion of times that E is expected to occur, then $P(E)$ must be a non negative real number, in fact a number in the unit interval [0,1]. If E and F are mutually exclusive events–say E=$\{1\}$ and $F = \{2,4,6\}$ in the example of the die–then the proportion of times that the union $E \cup F$ occurs is clearly the sum of the proportion associated with E and F separately. It follows therefore that the function P cannot be completely arbitrary; it is necessary to subject it to the condition of additivity, that is to require that if $E \cap F = \varnothing$, then $P(E \cup F)$ should be equal to $P(E) + P(F)$. Since the certain event X occurs every time, we should also require that $P(X)$=1. To sum up: numerical probability is a measure P on an σ-algebra $\mathscr{S}$ of subsets of a set X, such that $P(X) = 1$. We are now in a position to define a probability function.

Definition 1.43. Given a sample space X and an σ-algebra $\mathscr{S}$ on X, *a probability function* is a function P with domain $\mathscr{S}$ that satisfies the following conditions

(1) $P(A) \geq 0$ for all $A \in \mathscr{S}$.
(2) $P(X)$=1.
(3) For each infinite sequence $\{A_i\}$ of disjoint sets that belong to $\mathscr{S}$

$$P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k).$$

The measure space $(X,\mathscr{S},P)$ is called a probability space and the measure P is called a probability measure.

The three properties given in Definition 1.43 are usually referred to as the Axioms of Probability (or the Kolmogorov Axioms). Any function P that satisfies the Axioms of Probability is called a probability function. The axiomatic definition makes no attempt to tell what particular function P to choose; it merely requires P to satisfy the axioms. For any sample space many different probability functions can be defined. Which one reflects what is likely to be observed in a particular experiment is still to be discussed. The following gives a common method of defining a legitimate probability function.

Theorem 1.52. *Let $X=\{x_1,x_2,\dots,x_n\}$ be a finite set. Let $\mathscr{S}$ be any σ-algebra on X. Let $p_1,p_2,\dots,p_n$ be nonnegative numbers that sum to 1. For any $A\in\mathscr{S}$, define $P(A)$ by $P(A)=\sum_{i:x_i\in A}p_i$. (The sum over an empty set is defined to be 0.) Then P is a probability function on $\mathscr{S}$. This remains true if $X=\{x_1,x_2,\dots\}$ is a countable set.*

Proof. We will give the proof for finite X. For any $A\in\mathscr{S}, P(A)\geq 0$ because every $p_i\geq 0$. Thus, Axiom 1 is true. Now $P(X)=\sum_{1\leq i\leq n}p_i=1$. Thus, Axiom 2 is true. Let $A_1,A_2,\dots,A_k$ denote pair wise disjoint events. ($\mathscr{S}$ contains only a finite number of sets, so we need consider only finite disjoint unions.) Then

$$P(\bigcup_{i=1}^{k}A_i)=\sum_{j:x_j\in\bigcup_{i=1}^{k}A_i}p_j=\sum_{i=1}^{k}\sum_{j:x_j\in A_i}p_j=\sum_{i=1}^{k}P(A_i).$$

The first and third equalities are true by the definition of $P(A)$. The disjointedness of the A_is ensures that the second equality is true, because the same p_js appear exactly once on each side of the equality. Thus, Axiom 3 is true and Kolmogorov's Axioms are satisfied. □

1.5.2 Some Useful Properties of Probability

From the axioms of Probability we can build up many properties of the probability function, properties that are quite helpful in the calculation of more complicated probabilities.

Theorem 1.53. *Let $(X,\mathscr{S},P)$ be a probability space and A, B be any sets in $\mathscr{S}$. Then the following assertions hold.*

(1) $P(A')=1-P(A)$;
(2) $P(A)\leq 1$;
(3) $P(\varnothing)=0$;
(4) $P(B-A)=P(B)-P(A\cap B)$;
(5) $P(A\cup B)=P(A)+P(B)-P(A\cap B)$;
(6) *If* $A\subseteq B$, *then* $P(A)\leq P(B)$.

Proof. (1) The sets A and A' form a partition of the sample space X, that is, $X = A \cup A'$. Therefore, $P(A \cup A') = P(X)$=1 by the second Axiom. Also, A and A' are disjoint, so by the third axiom, $P(A \cup A') = P(A) + P(A')$. Therefore we have (1). Since $P(A') \geq 0$, (2) is immediately implied by (1). To prove (3), we use a similar argument on $X = X \cup \varnothing$. Since X, $\varnothing \in$S and they are disjoint, we have

$$1 = P(X) = P(X \cup \varnothing) = P(X) + P(\varnothing)$$

and thus (3) holds.

The proofs of (4)-(6) remain as exercises. □

Theorem 1.54. *Let $(X, \mathscr{S}, P)$ be a probability space and A be any set in $\mathscr{S}$. Let $\{C_i\}$ be an infinite sequence of the sets in $\mathscr{S}$ which is a partition of the sample space X and $\{A_i\}$ be any sequence of sets in $\mathscr{S}$. Then the following assertions hold.*

(1) $P(A) = \sum_{k=1}^{\infty} P(A \cap C_k)$;
(2) $P(\bigcup_{k=1}^{\infty} A_k) \leq \bigcup_{k=1}^{\infty} P(A_k)$.

Definition 1.44. Let $(X, \mathscr{S}, P)$ be a probability space and the events A, B be any sets in $\mathscr{S}$. If $P(B) > 0$, then the *conditional probability* of A given B, written $P(A|B)$, is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Note that what happens in the conditional probability calculation is that B becomes the sample space: $P(B|B)$=1. The intuition is that our original sample space, X, has been updated to B. All further occurrences are then calibrated with respect to their relation to B. In particular, note what happens to conditional probabilities of disjoint sets. Suppose A and B are disjoint, so $P(A \cap B)$=0. It then follows that $P(A|B) = P(B|A)$=0.

Theorem 1.55. *(Bayes'Rule) Let $(X, \mathscr{S}, P)$ be a probability space and A be any sets in $\mathscr{S}$. Let $\{C_i\}$ be an infinite sequence of the sets in S which is a partition of the sample space X. Then for each $i = 1, 2, \ldots$,*

$$P(C_i|A) = \frac{P(A|C_i)P(C_i)}{\sum_{i=1}^{\infty} P(A|C_i)P(C_i)}$$

Proof. By Theorem 1.54, we know that $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$. For each i=1, 2, ..., we have $P(A \cap C_i) = P(A|C_i)P(C_i)$ from Definition 1.43. Therefore

$$P(C_i|A) = \frac{P(C_i \cap A)}{P(A)} = \frac{P(A|C_i)P(C_i)}{\sum_{i=1}^{\infty} P(A|C_i)P(C_i)}.$$ □

Definition 1.45. Let $(X, \mathscr{S}, P)$ be a probability space. Then a finite sequence of events in $\mathscr{S}$, $\{A_i|i$=1,2,..., $n\}$ are mutually *independent* if for any subsequence $\{B_j \mid j = 1, 2, \ldots, q\} \subseteq \{A_i|i = 1, 2, \ldots, n\}$,

$$P(\bigcap_{j=1}^{q} B_j) = \prod_{j=1}^{q} P(B_j).$$

In many experiments it is easier to deal with a summary variable than with the original probability structure. For example, in an opinion poll, we might decide to ask 50 people whether agree or disagree with a certain issue. If we record a "1" for agree and "0" for disagree, the sample space for this experiment has 2^{50} elements, each an ordered string of 1s and 0s of length 50. We should be able to reduce this to a reasonable size! It may be that the only quantity of interest is the number of people who agree (equivalently, disagree) out of 50 and, if we define a variable X=number of 1s recorded out of 50, we have captured the essence of the problem. Note that the sample space for X is the set of integers $\{0,1, 2, \ldots, 50\}$ and is much easier to deal with than the original sample space. In general, we have the following definition.

Definition 1.46. Let $(X, \mathscr{S}, P)$ be a probability space. *A random variable* ξ is a map from the sample space X into the set of real numbers R. The *distribution function* of a random variable, ξ, denoted by $F_\xi(x)$, is defined as follows: for any $x \in R$, $F_\xi(x) = P(\xi \le x)$. For a discrete random variable, ξ, its *probability mass function*, denoted by $f_\xi(x)$, is given by $f_\xi(x) = P(\xi = x)$ for all x. For a continuous random variable, ξ, its *probability density function*, denoted by $f_\xi(x)$, is the function satisfies

$$F_\xi(x) = \int_{-\infty}^{x} f_\xi(t)dt \ for\ all\ x \in R.$$

A fundamental concept in the analysis of univariate data is the probability density function. Let ξ be a random variable that has probability density function $f(x)$. A motivation for the construction of a nonparametric estimate of the density function can be found using the definition of the density function.

$$f(x) = \frac{d}{dx}F(x) = \lim_{h\to 0} \frac{F(x+h) - F(x)}{h} \tag{1.5}$$

where $F(x)$ is the distribution function of the random variable ξ. Let $\{x_1, x_2, \ldots, x_n\}$ represent a random sample of size n from the density f. A natural finite-sample analog of (1.5) is to divide the line into a set of k equalized bins with small bin width h and to replace $F(x)$ with the empirical distribution function $F^-(x) = |\{x_i \mid x_i \le x\}|/n$. This leads to the histogram estimate of the density within a given bin: for $x \in (b_j, b_{j+1}]$,

$$f^-(x) = \frac{|\{x_i | b_j \le x_i \le b_{j+1}\}|/n}{h} = \frac{n_j}{nh}, \tag{1.6}$$

where $(b_j, b_{j+1}]$ defines the boundaries of the jth bin, n_j is the number of the observed samples in the jth bin $h = b_{j+1} - b_i$. What is needed is some way to evaluate $f^-(x)$ as an estimator of $f(x)$. One way to evaluate$f^-(x)$ is via mean integrated squared error (MISE) shown as follows (refer to [3]).

$$\int_{-\infty}^{+\infty} (f^-(u) - f(u))^2 du = \frac{1}{nh} + \frac{h^2 \int_{-\infty}^{+\infty} f'^2(u)du}{12} + O(n^{-1}) + O(h^3),$$

providing that $f'(x)$ is absolutely continuous and square integrable. In order for the estimator to be consistent, the bins must get narrower, with the number of observed samples per bin getting larger, as $n \to \infty$; that is, $h \to 0$ with $nh \to \infty$.

All the above definitions and results can be generalized to the multivariate case, where there are n random variables. Corresponding to Definition 1.46, the values of the joint probability distribution of n random variables $\xi_1, \xi_2, \ldots, \xi_n$ are given by

$$F(x_1, x_2, \ldots, x_n) = P(\xi_1 \le x_1, \xi_2 \le x_2, \ldots, \xi_n \le x_n)$$

for $-\infty < x_1, x_2, \ldots, x_n < \infty$. In the continuous case, probabilities are again obtained by integrating the joint probability density, and the joint distribution function is given by

$$F(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(t_1, t_2, \ldots, t_n) dt_1 dt_2 \ldots dt_n$$

for $-\infty < x_1, x_2, \ldots, x_n < \infty$. Also, partial differentiation yields

$$f(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \ldots \partial x_n} F(x_1, x_2, \ldots, x_n),$$

wherever these partial derivatives exist.

The *normal distribution* is in many ways the cornerstone of modern statistical theory. It was investigated first in the eighteenth century when scientists observed an astonishing degree of regularity in error of measurement. They found that the patterns (distributions) that they observed could be closely approximated by continuous curves, which they referred to as “normal curves of errors and attributed to the law of chance.

Definition 1.47. Let $(X, \mathscr{S}, P)$ be a probability space. n random variables ξ_1, $\xi_2, \ldots, \xi_n$ have a *joint normal distribution* if and only if for $-\infty < x_1, x_2, \ldots, x_n < \infty$,

$$F(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)} dt_1 dt_2 \ldots dt_n, \tag{1.7}$$

where $x = (x_1, x_2, \ldots, x_n)^T$, Σ is a $n \times n$ symmetry positive definite matrix and μ is a n-dimension vector in R^n.

Let $X_1, X_2, \ldots, X_l$ be the observed samples of size l from a probability space $(X, \mathscr{S}, P)$ with a joint normal distribution and $X \subseteq R^n$. Let matrix $X = (X_1, X_2, \ldots, X_l)$. If $l \ge n$, then by the Maximum Likelihood Estimate method in [1], the parameters Σ and μ in (1.7) can be estimated as follows:

$$\mu = \frac{1}{l} \sum_{i=1}^{l} X_i, \quad \Sigma = \frac{1}{l} X H X^T, \tag{1.8}$$

where $H = I - l^{-1}J$, I is the identical matrix and J is a $n \times n$ matrix whose entries are all 1.

1.6 Combinatoric Systems

In this section, we will give the definitions and some simple properties of combinatoric systems in [2]. By using them, the structure of data will be described in the AFS theory.

Definition 1.48. Let V and E be sets and disjoint, $f : E \to 2^V$. Then the triple (V, f, E) is called a *system*. The elements of E are called the *blocks* of the system (V, f, E), the elements of V are called the *vertices* of the system (V, f, E). If $x \in f(e)$, we say that the block e contains the vertex x, or that x and e are incident with each other.

Let $\Lambda = (V, f, E)$ and $\Omega = (W, g, F)$ be systems. The systems Λ and Ω are called *isomorphic systems* if there exist bijections, $p : E \to F, q : V \to W$ such that $q(f(e)) = g(p(e))$ for all $e \in E$. The pair (p, q) is called a *system-isomorphism.*

Definition 1.49. Given two systems $\Lambda_i = (V_i, f_i, E_i)$ for i=1, 2 where $V_1 \cap V_2 = \varnothing = E_1 \cap E_2$, the system $\Lambda = (V_1 \cup V_2, f, E_1 \cup E_2)$ where $f(e) = f_i(e_i)$ for $e \in E_i$ is called the *direct sum system* of Λ_1 and Λ_2 and denoted by $\Lambda_1 \oplus \Lambda_2$.

Since the direct sum of system is commutative and associative, Definition 1.49 may be extended to any finite number of systems Λ_i=(V_i, f_i, E_i) for i=1, 2, ..., k, as long as $V_i \cap V_j = \varnothing = E_i \cap E_j$, for any $i \neq j$. the system Λ=($V_1 \cup ... \cup V_k$, f, $E_1 \cup ... \cup E_k$), where $f(e) = f_i(e_i)$ for $e \in E_i$, $i = 1, 2, ..., k$, is called the direct sum of Λ_1, ... Λ_k and denoted by $\oplus_{1 \leq i \leq k} \Lambda_i$. Each Λ_i is called a direct summand of Λ. The system ($\varnothing$, f, $\varnothing$) is called the *trivial system*. Clearly Λ itself and the trivial system are always direct summands of Λ.

Definition 1.50. A system Λ is called a *connected system* if Λ itself and the trivial system are its only direct summands. A connected nontrivial summand of Λ is called a *component* of Λ.

Definition 1.51. Let $\Lambda = (V, f, E)$ be a system. $x_i \in V$ for $i = 0, 2, \ldots, n$ and $e_i \in E$, $i = 1, 2, \ldots, n-1$. A sequence $x = x_0, e_1, x_2, e_3, \ldots, e_{n-1}, x_n = y$ is called a $x - y$ *path* if $\{x_i, x_{i+2}\} \subseteq f(e_{i+1})$ for $i = 0, 2, \ldots, n-1$. $s \in V$, $t \in E$, if a sequence $s = s_0, t_1, s_2, t_3, \ldots, s_{n-1}, t_n = t$ in which $s_{n-1} \in f(t_n)$ and $\{s_{i-1}, s_{i+1}\} \subseteq f(t_i)$ for $i = 1, 2, \ldots, n-2$, or a sequence $t = t_0, s_1, t_2, s_3, \ldots, t_{n-1}, s_n = s$ in which $s_1 \in f(t_0)$ and $\{s_i, s_{i+2}\} \subseteq f(t_{i+1})$ for $i = 1, 2, \ldots, n-2$, then the sequence is called a $s - t$ path. We always define there exists a $x - x$ path for any $x \in V \cup E$.

Proposition 1.1. *Let $\Lambda = (V, f, E)$ be a system. The binary relation R on $V \cup E$ is an equivalence relation if $(s, t) \in R \Leftrightarrow$ there exists a $s - t$ path in Λ for $s, t \in V \cup E$.*

Proof. By the definition, we know that there exists a $x - x$ path for any $x \in V \cup E$. Also by the definition, we know that if there exist $x - y$ path then there also exists a $y - x$ path for $x, y \in V \cup E$. Suppose there exist the path

$$x-y, x=x_0, y_1, x_2, y_3, ..., x_{n-1}, y_n=y$$

and the path

$$y-z, y=y_0, z_1, y_2, z_3, ..., y_{n-1}, z_n=z$$

in Λ. If $y \in V$, then

$$x=x_0, y_1, x_2, y_3, ..., x_{n-1}, y_n=y=y=y_0, z_1, y_2, z_3, ..., y_{n-1}, z_n=z$$

is a $x-z$ path in Λ. If $y \in E$, by the definition, we know that $\{x_{n-1}, z_1\} \in f(y)$, then

$$x=x_0, y_1, x_2, y_3, ..., x_{n-1}, y_n=y=y_0, z_1, y_2, z_3, ..., y_{n-1}, z_n=z$$

is also a $x-z$ path in Λ. □

Theorem 1.56. *The component partition of system $\Lambda=(V,f,E)$ is the partition of the equivalence relation of Proposition 1.1 restricted to V.*

Proof. Assume $\Lambda_{V_1}, \Lambda_{V_2}, \ldots, \Lambda_{V_K}$ are the components of $\Lambda=(V,f,E)$ and let $F_i=\{e \in E|\ f(e) \subseteq V_i\}$. Let $s \in V_i$ and $t \in V_j$ for some $i \neq j$. Suppose $s=s_0, s_1, s_2, s_3, ..., s_{n-1}, s_n=t$ is a $s-t$ path, and let s_k be the last term in the path in $V_i \cup F_i$. If s_k is a vertex, then $s_k \in f(s_{k+1})$ where $s_k \in V_i$ and $s_{k+1} \notin F_i$. Since $\Lambda_{V_1}, \Lambda_{V_2}, \ldots, \Lambda_{V_K}$ are the components of Λ, $s_{k+1} \in F_q$ for some $q \neq i$and $f(s_{k+1}) \subseteq F_q$, i.e., $f(s_{k+1}) \cap V_i = \varnothing$. This is clearly impossible. If s_k is a block, then $s_{k+1} \in f(s_k)$, but $f(s_k) \subseteq V_i$ while$s_k \notin F_i$ which is impossible. We conclude that there exists no $s-t$path.

Now suppose $s, t \in V_i$ for some i. Let

$$S=\{r \in V_i \cup F_i \mid \text{ there is a } s-r \text{ path}\}.$$

Observe that if $r \in S \cap F_i$, then $f(r) \subseteq S$ and hence $f(r) \subseteq S \cap V_i$. On the other hand if $r \in F_i-(S \cap F_i)$, then $f(r) \cap S = \varnothing$, i.e., $f(r) \subseteq V_i-(S \cap V_i)$. We conclude that

$$\Omega_1=(S \cap V_i, f|_{S \cap F_i}, S \cap F_i),\ \ \Omega_2=(V_i-(S \cap V_i), f|_{F_i-S \cap F_i}, F_i-(S \cap F_i))$$

are both well-defined subsystems of Λ. Furthermore $\Lambda_{V_i}=\Omega_1 \oplus \Omega_2$. However Λ_{V_i}, being a component, is connected. Hence Ω_1 or Ω_2 is trivial. Since $s \in S$, Ω_1 is not trivial, hence Ω_2 is trivial. Thus $V_i-(S \cap V_i)=\varnothing$, $t \in V_i \subseteq S$ and s is equivalent to t. □

By Theorem 1.56, we immediately have the following corollary.

Corollary 1.6. *Let $\Lambda=(V,f,E)$ be a system. The following three conditions are equivalent:*

(1) Λ is connected.
(2) $f(e) \neq \varnothing$ for all $e \in E$, and for every $s, t\ \in V$ there is a $s-t$ path.
(3) For every $s, t \in V \cup E$ there is a $s-t$ path.

Exercises

Exercise 1.1. For any three sets A, B, and C, show the following assertions hold:

(a) $A \cup B = B \cup A, A \cap B = B \cap A$; (*Commutativity*)
(b) $A \cup (B \cup C)=(A \cup B) \cup C, A \cap (B \cap C)=(A \cap B) \cap C$; (*Associativity*)
(c) $(A \cup B)' = A' \cap B'$, $(A \cap B)' = A' \cup B'$. (*DeMorgan's Laws*)

Exercise 1.2. For any collection of sets Γ and the set B, show the distributive laws carry over to arbitrary intersections and unions:

$$B \cap (\bigcup_{A\in\Gamma} A) = \bigcup_{A\in\Gamma} (B \cap A), B \cup (\bigcap_{A\in\Gamma} A) = \bigcap_{A\in\Gamma} (B \cup A).$$

Exercise 1.3. Let f be a map of X into Y and A, B be any subsets of X. Let U, V be any subsets of Y. Show the following assertions hold

(a)$f(A \cup B) = f(A) \cup f(B)$, $f(A \cap B) \subseteq f(A) \cap f(B)$;
(b) $f^{-1}(U \cup V) = f^{-1}(U) \cup f^{-1}(V)$, $f^{-1}(U \cap V) \subseteq f^{-1}(U) \cap f^{-1}(V)$.

Exercise 1.4. Prove that a subset of a countable set is countable.

Exercise 1.5. Prove that the set Q of all rational numbers is a countable set.

Exercise 1.6. Let $(S, \leq)$ be a partially ordered set and A a non-empty subset of S. Show the following assertions hold:

(a) If A has maximum (minimum) element, then the maximum (minimum) element is unique.

(b) If A is a chain in S (e.g. linear ordered subset), then maximal (minimal) element of A (if it exists) must be maximum (minimum) element.

Exercise 1.7. Let A be a subset of a topological space X and let $b(A)$ be the boundary of A. Show $B(A) = A^- \cap (X - A)^- = A^- - A^0$, $X - b(A) = A^0 \cup (X - A)^0$, $A^- = A \cup b(A)$ and $A^0 = A - b(A)$. And prove that a set is closed if and only if it contains its boundary. Furthermore prove that a set is open if and only if it is disjoint from its boundary.

Exercise 1.8. Prove that the real numbers, with the usual topology are connected.

Exercise 1.9. Let a be a family of connected subsets of a topological space. If no two members of a are separated, prove that $\cup_{A\in\mathsf{a}} A$ is connected.

Exercise 1.10. Prove that each connected subset of a topological space is connected in a component and each component is closed. If A and B are distinct components of a space, prove that A and B are separated.

Exercise 1.11. Prove that a topological space is a Hausdorff space if and only if each net in the space converges to at most one point.

Exercise 1.12. Prove that a topological space X is compact if and only if for any family $\{D_\alpha \mid \alpha \in A\}$ of closed sets with the finite intersection property.

Exercise 1.13. Let X be a topological space, then X is compact if and only if any family $\{D_\alpha \mid \alpha \in A\}$ of subsets of X with the finite intersection property has the further property that $\cap_{\alpha \in A} D_\alpha^- \neq \varnothing$.

Exercise 1.14. Prove that each of the following properties of topological spaces is stronger than the next: Normality, Regularity, T_2 (Hausdorff), T_1, T_0, in the sense that if a space satisfies the definitions of any one of these properties, it also satisfies the definitions for all of the following ones as well.

Exercise 1.15. If X and Y are topological space and f is a function on X to Y, show the following statements are equivalent.

(a) The function f is continuous.
(b) The inverse of each closed set is closed.
(c) The inverse of each member of a subbase for the topology for Y is open.
(d) For each x in X the inverse of every neighborhood of $f(x)$ is a neighborhood of x.
(e) For each x in X and each neighborhood U of $f(x)$ there is a neighborhood V of x such that $f(V) \subseteq U$.
(f) For each net S (or $\{S_n,\, n \in D\}$) in X which converges to a point s, the composition $f \cdot S$ (i.e., $\{f(S_n),\, n \in D\}$) converges to $f(s)$.
(g) For each subset A of X the image of closure is a subset of the closure of the image; that is, $f(A^-) \subseteq f(A)^-$.
(h) For each subset B of Y, $f^{-1}(B)^- \subseteq f^{-1}(B^-)$.

Exercise 1.16. Prove that a T_1 space X is regular if and only if for each $x \in X$ and each $U \in \mathscr{U}_x$ there is a $V \in \mathscr{U}_x$ such that $V^- \subseteq U$.

Exercise 1.17. Prove that a T_1 space X is normal if and only if for each closed set C and each open set U such that $C \subseteq U$, there exists an open set V such that $C \subseteq V \subseteq V^- \subseteq U$.

Exercise 1.18. Prove that the $\rho(x,y)$ in Definition 1.36 is a distance on $\mathscr{H}$; The subspace of $\mathscr{H}$ defined by

$$E^1 = \{x \mid x \in \mathscr{H},\, x = \{x_n, n \in N^+\},\, x_n = 0 \text{ for } n > 1\ \}$$

is homeomorphic to the real line with the usual topology; More generally $E^n \subseteq \mathscr{H}$, defined by

$$E^n = \{x \mid x \in \mathscr{H},\, x = \{x_i, i \in N^+\},\, x_i = 0 \text{ for } i > n\ \}$$

is homeomorphic to R^n the n-dimension Euclidean space with the usual topology.

Exercise 1.19. Let $(X, \mathscr{S}, P)$ be a probability space and A, B be any sets in $\mathscr{S}$. Show the following assertions

(a) $P(B-A) = P(B) - P(A \cap B)$;
(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
(c) If $A \subseteq B$, then $P(A) \leq P(B)$.

Exercise 1.20. Let $(X, \mathscr{S}, P)$ be a probability space and A be any set in $\mathscr{S}$. Let $\{C_i\}$ be an infinite sequence of the sets in $\mathscr{S}$ which is a partition of the sample space X and $\{A_i\}$ be any sequence of sets in $\mathscr{S}$. Show the following assertions.

(a) $P(A) = \sum_{k=1}^{\infty} P(A \cap C_k)$;
(b) $P(\bigcup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} P(A_k)$.

References

1. George, C., Roger, L.B.: Statistical Inference, Wadsworth Group (2002)
2. Graver, J.E., Watkins, M.E.: Combinatorics with Emphasis on the Theory of Graphs. Springer, New York (1977)
3. Simonoff, J.S.: Smoothing Methods in Statistics. Springer, New York (1996)
4. John, L.K.: General Topology. Springer, New York (1955)
5. John, D.B.: Elements of Point Set Topology. Prentice-Hall, Inc., Englewood Cliffs (1964)
6. Jacobson, N.: Basic Algebra I. W H Freeman and Company, New York (1985)
7. Halmos, P.R.: Measure Theory. Springer, New York (1974)

Chapter 2
Lattices

This chapter offers a concise introduction to lattices, Boolean algebras, topological molecular lattices and shows main relations between them. For details, the readers may refer to [1, 3, 2]. Our purpose is to familiarize the readers with the concepts and fundamental results, which will be exploited in further discussion. Some results listed without proofs is left to the reader.

2.1 Lattices

An element u of a partially ordered set S is an *upper bound* of a subset A of S if $u \geq a$ for every $a \in A$. The element u is a least upper bound or supremum of A (denoted by $\sup A$) if u is an upper bound of A and $u \leq v$ for every upper bound v of A. It is clear from anti-symmetry of Definition 1.5 that if a $\sup A$ exists, then it is unique. In a similar fashion one defines lower bounds and greatest lower bounds or infimum of a set A(denoted by $\inf A$). Also if $\inf A$ exists, then it is unique. The set of all of the upper bounds (resp. lower bounds) of A is denoted by $M_a A$ (resp. $M_i A$). We now introduce the following.

Definition 2.1. *A lattice* is a partially ordered set in which any two elements have a least upper bound (supremum) and a greatest lower bound (infimum).

Let a partially ordered set L be a lattice, we denote the least upper bound of a and b by $a \vee b$ ("a cup b" or "a union b") and the greatest lower bound by $a \wedge b$ ("a cap b" or "a meet b"). And the lattice is briefly denoted as $(L, \vee, \wedge)$.

Proposition 2.1. *Let $\leq$ be the ordered relation of the partially ordered set $(L, \leq)$ and $(L, \vee_1, \wedge_1)$ be a lattice. If $\leq^{-1}$ is the inverse relation of $\leq$, then the partially ordered set $(L, \leq^{-1})$ is a lattice $(L, \vee_2, \wedge_2)$, where for any $a, b \in L$,*

$$a \vee_2 b = a \wedge_1 b, \quad a \wedge_2 b = a \vee_1 b.$$

The latices $(L, \vee_1, \wedge_1)$ and $(L, \vee_2, \wedge_2)$ are called dual lattices. *$(L, \vee_2, \wedge_2)$, the dual lattice of $(L, \vee_1, \wedge_1)$, also briefly denoted as L^{-1}.*

X. Liu and W. Pedrycz: Axiomatic Fuzzy Set Theory and Its Applications, STUDFUZZ 244, pp. 61–94.
springerlink.com

Proposition 2.2. *Let L be a lattice. For any* $a,b,c \in L$*,* $(a \vee b) \vee c \geq a,b,c$*; if* $v \geq a,b,c$*, then* $v \geq (a \vee b), c$ *and so* $v \geq (a \vee b) \vee c$*. Hence* $(a \vee b) \vee c$ *is a supremum of* a,b,c*.*

Remark 2.1. One shows that any finite set of elements of a lattice has a supremum. Similarly, any finite subset has an infimum. We denote the supremum and infimum of $a_1, a_2, \ldots, a_n$ by

$$a_1 \vee a_2 \vee \ldots \vee a_n$$

and

$$a_1 \wedge a_2 \wedge \ldots \wedge a_n$$

respectively.

Example 2.1. Let S be a set, then $(2^S, \subseteq)$ is a lattice. Where $\vee$ and $\wedge$ are defined by $\cup$ and $\cap$, respectively. $(2^S, \subseteq)$ is called power set lattice on set S.

Proposition 2.3. *Any totally ordered set is a lattice.*

Proof. Let $(L, \leq)$ be a totally ordered set. For $\forall a,b \in L$, we have either $a \geq b$ or $b \geq a$. In the first case, $a \vee b = a$ and $a \wedge b = b$. If $b \geq a$ then $a \vee b = b$ and $a \wedge b = a$. Thus $(L, \leq)$ is a lattice by Definition 2.1.

Definition 2.2. Let L be a lattice with finite elements, then L is called a *finite lattice*. Otherwise, L is called an *infinite lattice*.

Theorem 2.1. *Let X be a non-empty partially ordered set. Then X is a lattice if and only if every non-empty subset of X has a least upper bound and a greatest lower bound.*

Proof. It is straightforward by Definition 2.1. □

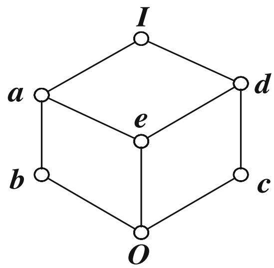
Definition 2.3. A subset X of a lattice L is called a *sublattice* if it is closed under the operation $\vee$ and $\wedge$, that is, $a \vee b \in X$ and $a \wedge b \in X$ for $\forall a,b \in X$.

The following facts which remain as exercises are clear from Definition 2.3.

Proposition 2.4. *Let L be a lattice. Then the following assertions hold.*

(1) Empty set $\varnothing$ *is a sublattice of lattice L;*
(2) Unit set $\{a\}$ $(a \in L)$ *is a sublattice of lattice L;*
(3) The intersection of any sublattice of lattice L is a sublattice of L.

Remark 2.2. It is evident that a sublattice is a lattice relative to the induced operations. On the other hand, a subset of a lattice may be a lattice relative to the partial ordered relation $\leq$ defined in L without being a sublattice.



For example, a fact that L is a lattice can be visualized in the form of the above diagram. Let $X = L - \{a, b, c\}$ and $Y = L - \{e\}$, then X is a sublattice of L, and Y is a lattice without being a sublattice of L.

If a is a fixed element of a lattice L, then the subset of elements x such that $x \geq a(x \leq a)$ is clearly a sublattice. If $a \leq b$, the subset of elements $s \in L$ such that $a \leq x \leq b$ is a sublattice. We call such a sublattice an *interval sublattice* and we denote it as $I[a, b]$.

Theorem 2.2. *Let $(L, \leq)$ be a lattice, $x, y \in L$. Then the following conditions are equivalent:*

(1) $x \leq y$;
(2) $x \wedge y = x$;
(3) $x \vee y = y$.

Proof. If $x \leq y$, then x is a lower bound of $\{x, y\}$. Let z be any lower bound of $\{x, y\}$, then $z \leq x$, that is, x is a greatest lower bound of $\{x, y\}$, and so $x \wedge y = x$. Conversely, if $x \wedge y = x$, we must have $x \leq y$. Thus, the condition (1) is equivalent to the condition (2).

In virtue of the principle of duality, the condition (1) is equivalent to the condition (3). □

In particular, if a lattice $(L, \leq)$ has identity element I and zero element O, for $\forall x \in L$, we have the following relationships

$$x \wedge O = O, x \vee O = x, x \wedge I = x, x \vee I = I$$

The following results can be directly proved by resorting to the definitions and are left to the reader.

Theorem 2.3. *Let $(L, \leq)$ be a lattice, $\forall x, y, z \in L$ we have the follows:*

L1. $x \wedge x = x$, $x \vee x = x$. *(Idempotency)*
L2. $x \wedge y = y \wedge x$, $x \vee y = y \vee x$. *(Commutativity)*
L3. $x \wedge (y \wedge z) = (x \wedge y) \wedge z$, $x \vee (y \vee z) = (x \vee y) \vee z$. *(Associativity)*
L4. $x \wedge (x \vee y) = x = x \vee (x \wedge y)$. *(Absorption)*

Corollary 2.1. *In lattice $(L, \leq)$, operations of union and meet are order-preserving. That is, $\forall x, y, z \in L$ we have*

$$x \leq y \Longleftrightarrow x \wedge z \leq y \wedge z, x \vee z \leq y \vee z.$$

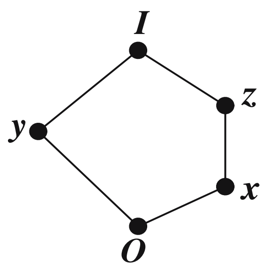
Corollary 2.2. *In lattice $(L, \leq)$, distributive inequality holds. That is, $\forall x, y, z \in L$ we have*

(1) $x \wedge (y \vee z) \geq (x \wedge y) \vee (x \wedge z)$;
(2) $x \vee (y \wedge z) \leq (x \vee y) \wedge (x \vee z)$.

Corollary 2.3. *In a lattice $(L, \leq)$, modular inequality holds. That is, $\forall x, y, z \in L$ we have*

$$x \leq z \;\Rightarrow\; x \vee (y \wedge z) \leq (x \vee y) \wedge z$$

Generally, the equality does not hold in the relationships presented in Corollary 2.2 and Corollary 2.3. The lattice L, which is presented in the form of the following diagram as shown below, is an example for this. In fact, the algebraic characteristics



of lattice are completely described by L1∼L4, as we demonstrate in the form of the following theorem.

Theorem 2.4. *Let L be any set in which there are defined two binary operation $\vee$ and $\wedge$ satisfying the condition* L1∼L4 *of Theorem* 2.3*. Then the following assertions hold.*

(1) $\forall x, y \in L,\ x \wedge y = x \Leftrightarrow x \vee y = y$;
(2) The L is a lattice relative to the following definition of $\leq$

$$x \leq y \Longleftrightarrow x \wedge y = x$$

and that $x \vee y$ and $x \wedge y$ are the supremum and infimum of x and b in this lattice.

Proof. (1) If $x \wedge y = x$, we have $x \vee y = (x \wedge y) \vee y = y$ by L2 and L4. Conversely, if $x \vee y = y$, we have $x \wedge y = x \wedge (x \vee y) = x$.

(2) Since $x \wedge x = x$ we have $x \geq x$ so reflexivity holds. If $x \geq y$and $y \geq x$, then we have $x \vee y = x$ and $y \vee x = y$. Since $x \vee y = y \vee x$ this gives $x = y$, which proves anti-symmetry. Next assume that $x \geq y$ and $y \geq z$. Then $x \vee y = x$ and $y \vee z = y$. Hence

$$x \vee z = (x \vee y) \vee z = x \vee (y \vee z) = x \vee y = x$$

which means that $x \geq z$. Hence transitivity is valid.

Since $(x \vee y) \wedge x = x$, by L4, $x \vee y \geq x$. Similarly, $x \vee y \geq y$. Now let z be an element such that $z \geq x$ and $z \geq y$. Then $x \vee z = z$ and $y \vee z = z$. Hence

$$(x \vee y) \vee z = x \vee (y \vee z) = x \vee z = z$$

so $z \geq x \vee y$. Thus $x \vee y$ is a supremum of a and b in L. By duality, $x \wedge y$ is an infimum of a and b. This completes the verification that a set L with binary operations satisfying L1∼L4 is a lattice and $x \vee y$ and $x \wedge y$ are the supremum and infimum in this lattice. □

Let us recall that a lattice (in virtue of Theorem 2.3 and Theorem 2.4) is an algebraic system with two algebra operation (denoted as $\vee$ and $\wedge$), which satisfies the the properties of idempotency, commutativity, associativity and absorption (axioms L1∼L4). These axioms may be regarded as an equivalent definition of lattices. Similarly, we may define a concept of semi-lattice.

Definition 2.4. An algebraic system with an algebra operation $\wedge$ or $\vee$ which satisfies axioms L1∼L3 is called a *semi-lattice*.

Let $(S, \leq)$ be a partially ordered set. We can prove that, if there exists infimum $x \wedge y$ (resp. supremum $x \vee y$) for any $x, y \in S$, then the algebraic system $(S, \wedge)$ (resp. $(S, \vee)$) is a semi-lattice, which is called *meet semi-lattice* (resp. *union semi-lattice*). Conversely, let the algebraic system $(S, \circ)$ be a semi-lattice, We define binary relation $\leq$ as follows

$$x \leq y \Longleftrightarrow x \circ y = x \quad (\text{or } y \circ x = y)$$

Then $(S, \leq)$ is a partially ordered set, and there exists a infimum $x \wedge y = x \circ y$ (or supremum $x \vee y = x \circ y$).

Theorem 2.5. *Let $(S, \leq)$ is a partially ordered set. Then we have*

(1) $(S, \leq)$ is a lattice $\Leftrightarrow$ $(S, \leq)$ is both meet semi-lattice and union semi-lattice;
(2) If $(S, \leq)$ is finite meet (union) semi-lattice with identity element I (or zero element O), then $(S, \leq)$ is a lattice.

Proof. It is clear that (1) holds by Theorem 2.34. We now demonstrate that (2) holds. Let $(S, \leq)$ be a finite meet semi-lattice with identity element I. For $\forall x, y \in S$, $M_a\{x, y\} \neq \varnothing$ by $I \in M_a\{x, y\}$, where $M_a\ \{x, y\}$ is the set of upper bound of $\{x, y\}$. Since $(S, \leq)$ satisfies maximal condition and minimal condition, there exists minimal element z_0 of $M_a\{x, y\}$. It follows that, $\forall z' \in M_a\{x, y\}$, $z_0 \wedge z'$ exists and $z_0 \wedge z' \in M_a\{x, y\}$ since S is a meet semi-lattice. Also, $z_0 \leq z'$ by minimally of z_0, that is, z_0 is minimum element of $M_a\{x, y\}$. Thus $z_0 = x \vee y$. Consequently, $(S, \leq)$ is a union semi-lattice. It follows that $(S, \leq)$ is a lattice by (1).

According to the principle of duality, if $(S, \leq)$ is a finite union semi-lattice with zero element O, then $(S, \leq)$ is a lattice. □

We now introduce the definition of a complete lattice.

Definition 2.5. A partially ordered set L is called a *complete lattice* if every subset $A = \{a_i \mid i \in I\}$ of L has a supremum and infimum.

By the definition of complete lattice, we can directly obtain the following.

Theorem 2.6. *Let L be a lattice. We have the following assertions.*

(1) L is a complete lattice if and only if L^{-1} dual of L) is complete;
(2) If L is a finite lattice, then L is complete.

We denote supremum and infimum of $\{a_i \mid i \in I\}$ by $\bigvee_{i \in I} a_i$ and $\bigwedge_{i \in I} a_i$, respectively. If the set $\{a_i \mid i \in I\}$ coincides with the underlying set of the lattice L then $O \equiv \wedge_{i \in I} a_i$ is the least element of L and $I \equiv \vee_{i \in I} a_i$ is the greatest element of $L : O \leq a$ and $I \geq a$ for every $a \in L$. The following comes as a useful criterion for recognizing whether a given partially ordered set is complete lattice.

Theorem 2.7. *A partially ordered set with a greatest element I such that every non-vacuous subset has a greatest lower bound is a complete lattice. Dually, a partially ordered set with a least element O such that every non-vacuous subset has a least upper bound is a complete lattice.*

Proof. Assuming the first set of hypotheses we have to show that any A=$\{a_i \mid i \in I\}$ has a supremum. Since $I \geq a_\alpha$ the set B of upper bounds of A is non-vacuous. Let b = inf$B$. Then it is clear that b = sup$A$. The second statement follows in virtue of symmetry. □

The definition of a lattice provided by means of the axioms L1∼L4 makes it natural to define a *homomorphism* of a lattice L into a lattice L' to be a map φ: $a \to \varphi(a)$ such that $\varphi(a \vee b) = \varphi(a) \vee \varphi(b)$ and $\varphi(a \wedge b) = \varphi(a) \wedge \varphi(b)$. In this case if $a \geq b$ then we have $a \vee b = a$; hence $\varphi(a) \vee \varphi(b) = \varphi(a)$ and $\varphi(a) \geq \varphi(b)$. A map between partially ordered sets having this property is called *order preserving*. Thus we have shown that a lattice homomorphism is order preserving. However, the converse need not hold. A bijective homomorphism of lattices is called an *isomorphism*. These can be characterized by order preserving properties, as we see in the following

Theorem 2.8. *A bijective map $\varphi(a)$ of a lattice L onto a lattice L' is a lattice isomorphism if and only if it and its inverse are order preserving.*

Proof. We have seen that if $a \to \varphi(a)$ is a lattice isomorphism, then this map is order preserving. It is also apparent that the inverse map is an isomorphism of L' into L so it is order preserving. Conversely, suppose $a \to \varphi(a)$ is bijective and as well as its inverse is order preserving. This means that $a \geq b$ in L if and only if $\varphi(a) \geq \varphi(b)$ in L'. Let $d = a \vee b$. Then $d \geq a$, b, so $\varphi(d) \geq \varphi(a)$, $\varphi(b)$. Let $\varphi(e) \geq \varphi(a)$, $\varphi(b)$ and let e be the inverse image of $\varphi(e)$. Then $e \geq a$, b. Hence $e \geq d$ and $\varphi(e) \geq \varphi(d)$. Thus we have shown that $\varphi(d) = \varphi(a) \vee \varphi(b)$. In a similar fashion we can show that $\varphi(a \wedge b) = \varphi(a) \wedge \varphi(b)$. □

Finally, we give the concept of ideal of lattice.

Definition 2.6. Let $(L, \leq)$ be a lattice (resp. union semi-lattice). If J a subset of L which satisfies the following conditions, then J is called an *ideal* (resp. *union ideal*) of L.

(1) $a \vee b \in J (\forall a, b \in J)$;
(2) $\forall a \in J, x \in L, x \leq a$ implies $x \in J$.

Dually, we can define a *dual ideal* (resp. *meet ideal*) of lattice (resp. meet semi-lattice). If J is an ideal (or dual ideal) of a lattice L, and $J \neq \varnothing$, $J \neq L$, we call J a *proper ideal* (or *proper dual ideal*) of L. Clearly, $\varnothing$ and L are ideal (or dual ideal) of L, we call them *usual ideal* (or *usual dual ideal*) of L. It is easy to verify that every ideal (or dual ideal) of lattice L is a sublattice of L.

Theorem 2.9. *Let L be a lattice, and J a subset of L. Then the following assertions hold.*

(1) J is an ideal (dual ideal) if and only if $a \vee b \in J$ *(reap.* $a \wedge b \in J$*)* $\Longleftrightarrow$ $a \in J$ *and* $b \in J$, $\forall a, b \in L$;
(2) If J is a sublattice of L, then J is an ideal (resp. dual ideal) of L if and only if $\forall a \in J$, $b \in L$, *we have* $a \wedge b \in J$ *(resp.* $a \vee b \in J$*);*
(3) Arbitrary intersections of ideal (or dual ideal) of L is also ideal (or dual ideal) of L; finite intersections of proper ideal (or dual ideal) of L is also proper ideal (or proper dual ideal) of L.

2.2 Distributive Lattices

Let L be a lattice. We now formulate the following two distributive laws:

D1 $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z), \forall x, y, z \in L$

and its dual

D2 $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z), \forall x, y, z \in L$

Theorem 2.10. *For any lattice L, condition* D1 *is equivalent to condition* D2.

Proof. Let L be a lattice and D1 hold in L. For $\forall x, y, z \in L$, we have

$$\begin{aligned}
(x \vee y) \wedge (x \vee z) &= ((x \vee y) \wedge x) \vee ((x \vee y) \wedge z) && (by\ \ condition\ D1) \\
&= x \vee ((x \vee y) \wedge z) && (by\ \ L2\ \ and\ L4\ of\ Theorem\ 2.3) \\
&= x \vee ((x \wedge z) \vee (y \wedge z)) && (by\ condition\ D1) \\
&= (x \vee (x \wedge z)) \vee (y \wedge z) && \\
&= x \vee (y \wedge z) && (by\ L4\ of\ Theorem\ 2.3)
\end{aligned}$$

which is D2. Dually D2 implies D1. □

Definition 2.7. A lattice L in which these distributive laws hold is called *distributive lattice*.

From Theorem 2.10, in fact, a lattice L is distributive lattice as long as it satisfies one of D1 and D2. There are some important examples of distributive lattices. Firstly, the lattice 2^S of subsets of a set S is distributive. Secondly, we have the following

Lemma 2.1. *Any linear ordered set is a distributive lattice.*

Proof. We wish to establish D1 for any three elements x, y, z. We distinguish two cases (1) $x \geq y$, $x \geq z$, (2) $x \leq y$ or $x \leq z$. In (1) we have $x \wedge (y \vee z) = y \vee z$ and $(x \wedge y) \vee (x \wedge z) = y \vee z$. In (2) we have $x \wedge (y \vee z) = x$ and$(x \wedge y) \vee (x \wedge z) = x$. Hence in both cases D1 holds. □

Example 2.2. Let N be a set of natural numbers. If $a \leq b$ for natural numbers a and b means $a|b$ (a is a divisor of b), then $(N, |\ )$ is a distributive lattice.

From Example 2.2, $(N, |\ )$ is a partially ordered set. In this example, $x \vee y = (x, y)$ the g.c.d. (*greatest common divisor*) of x and y and $x \wedge y = [x, y]$ the l.c.m. (*least common multiple*) of x and y. Also, if we write $x = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}$, $y = p_1^{b_1} p_2^{b_2} \dots p_k^{b_k}$ where the p_i are distinct primes and the a_i and b_i are non-negative integers, the $(x, y) = \prod_{1 \leq i \leq k} p_i^{\min(a_i,\, b_i)}$, $[x, y] = \prod_{1 \leq i \leq k} p_i^{\max(a_i,\, b_i)}$. Hence if $z = p_1^{c_1} p_2^{c_2} \dots p_k^{c_k}$, c_i non-negative integral, then

$$[x, (y, z)] = \prod_{1 \leq i \leq k} p_i^{\max(a_i,\, \min(b_i,\, c_i))}$$

and

$$([x, y], [x, z]) = \prod_{1 \leq i \leq k} p_i^{\min(\max(a_i,\, b_i),\, \max(a_i,\, c_i))}$$

Now the set of non-negative integers with the natural order is totally ordered and $\max(a_i, b_i) = a_i \vee b_i$, $\min(a_i, b_i) = a_i \wedge b_i$ in this lattice. Hence, the distributive law D2 in this lattice leads to the relation

$$\max(a_i, \min(a_i, c_i)) = \min(\max(a_i, b_i), \max(a_i, c_i))$$

Then we have

$$[x, (y, z)] = ([x, y], [x, z])$$

which is D1 for the lattice of positive integers ordered by divisibility.

The following results whose proofs is left as exercises are evident from the definition of distributive lattice.

Theorem 2.11. *Let L and $L_i (i \in I)$ be lattice. Then the following assertions hold.*

(1) L is a distributive lattice if and only if L^{-1} (dual of L) is distributive lattice;
(2) $\prod_{i \in I} L_i$ is distributive lattice if and only if L_i ($\forall i \in I$) is distributive lattice.
(3) If L is a distributive lattice, then sublattices of L are also distributive lattice.

Definition 2.8. Let L be a lattice. If for $\forall x, y, z \in L$, we have

$$x \leq z \Longrightarrow x \vee (y \wedge z) = (x \vee y) \wedge z, \qquad \text{(modular law)}$$

then L is called a *modular lattice* .

It is clear that a distributive lattice is a modular lattice, that is, if L is a lattice, and D1 and D2 hold for $\forall x, y, z \in L$, then $x \leq z \Rightarrow x \vee (y \wedge z) = (x \vee y) \wedge z$.

Theorem 2.12. *Let L be a lattice, $x,y,z \in L$. Then the following conditions are equivalent.*

(1) L is a modular lattice;
(2) $x \wedge (y \vee (x \wedge z)) = (x \wedge y) \vee (x \wedge z)$, $(\forall x,y,z \in L)$;
(3) $x \leq y, z \wedge x = z \wedge y$ and $z \vee x = z \vee y \Rightarrow x = y$, $(\forall x,y,z \in L)$.

Proof. (1) $\Rightarrow$ (2). Let L be a modular lattice. Then $x \leq z \Rightarrow x \vee (y \wedge z) = (x \vee y) \wedge z$ for $\forall x,y,z \in L$. Since $x \wedge z \leq x$, we have

$$(x \wedge y) \vee (x \wedge z) = (x \wedge z) \vee (y \wedge x) = ((x \wedge z) \vee y) \wedge x = x \wedge (y \vee (x \wedge z)).$$

(2) $\Rightarrow$ (3). Suppose (2) holds. Form L2, L4 of Theorem 2.3 and Definition 2.8, if $x \leq y, z \wedge x = z \wedge y$ and $z \vee x = z \vee y$, we have

$$x = x \vee (z \wedge x) = x \vee (z \wedge y) = (x \vee z) \wedge y = (y \vee z) \wedge y = y.$$

Thus (3) holds.

(3) $\Rightarrow$ (1). Let $a = x \vee (y \wedge z)$, $b = (x \vee y) \wedge z$. If $x \leq z$, then

$$b \wedge y \;\geq\; a \wedge y \geq (y \wedge z) \wedge y = y \wedge z = (x \vee y) \wedge z \wedge y = b \wedge y,$$

that is, $b \wedge y = a \wedge y$. Dually, we have $b \vee y = a \vee y$. Thus $a = b$ by (3), that is that (1) holds. □

Theorem 2.13. *Let L be a lattice. Then the following conditions are equivalent.*

(1) L is a distributive lattice;
(2) $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$, $(\forall x, y, z \in L)$;
(3) $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$, $(\forall x, y, z \in L)$;
(4) $(x \wedge y) \vee (y \wedge z) \vee (z \wedge x) = (x \vee y) \wedge (y \vee z) \wedge (z \vee x)$, $(\forall x, y, z \in L)$;
(5) $x \wedge (y \vee z) \leq (x \wedge y) \vee (x \wedge z)$, $(\forall x, y, z \in L)$;
(6) $x \wedge (y \vee z) \leq (x \wedge y) \vee z$, $(\forall x, y, z \in L)$;
(7) $z \wedge x = z \wedge y$ and $z \vee x = z \vee y \Rightarrow x = y$ $(\forall x, y, z \in L)$.

Proof. From Theorem 2.10, Definition 2.7 and Corollary 2.2, we have (5) $\Leftrightarrow$ (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3).

(5) $\Rightarrow$ (6). Since $x \wedge z \leq z$, by Corollary 2.1, it is obvious that $x \wedge (y \vee z) \leq (x \wedge y) \vee (x \wedge z) \leq (x \wedge y) \vee z$ for $\forall x, y, z \in L$. It follows that (5) $\Rightarrow$ (6). Conversely, suppose (6) holds. Then

$$x \wedge (y \vee z) = x \wedge (x \wedge (y \vee z)) \leq \; x \wedge ((x \wedge y) \vee z)$$

(6) $\Rightarrow$ (7). For $\forall x,y,z \in L$, if $z \wedge x = z \wedge y$ and $z \vee x = z \vee y$, we have

$$x = x \wedge (z \vee x) = x \wedge (z \vee y) \;\leq (x \wedge z) \vee y = (z \wedge y) \vee y = y$$

$$y = y \wedge (z \vee y) = y \wedge (z \vee x) \;\leq (y \wedge z) \vee x = (z \wedge x) \vee x = x$$

by absorption (L4 of Theorem 2.3) and (6). Thus $x = y$.

(7) ⇒ (5). Let $a = x \wedge (y \vee z)$, $b = (x \wedge y) \vee (x \wedge z)$. From L2, L4 of Theorem 2.3, Corollary 2.2 and Corollary 2.3, we have

$$a \wedge y = x \wedge (y \vee z) \wedge y = x \wedge ((y \vee z) \wedge y) = x \wedge y$$

$$x \wedge y \geq [x \wedge (y \vee z)] \wedge y \geq b \wedge y = [(x \wedge y) \vee (x \wedge z)] \wedge y \geq (x \wedge y) \vee (x \wedge z \wedge y) \geq (x \wedge y)$$

That is, $a \wedge y = b \wedge y$. Dually, we have $b \vee y = a \vee y$. Thus $a = b$ by (7), that is that (5) holds.

(3) ⇒ (4). By (3), for $\forall x, y, z \in L$, we have

$$\begin{aligned}(x \wedge y) \vee (y \wedge z) \vee (z \wedge x) &= [(x \wedge y) \vee (y \wedge z) \vee z] \wedge [(x \wedge y) \vee (y \wedge z) \vee x] \\ &= [(x \wedge y) \vee ((y \vee z) \wedge z)] \wedge [(y \wedge z) \vee ((x \vee y) \wedge x)] \\ &= \ldots = (x \vee y) \wedge (y \vee z) \wedge (z \vee x)\end{aligned}$$

That is, (4) holds.

(4) ⇒ (3). Suppose (4) holds. If $z \leq x$, for $\forall x, y, z \in L$, we have

$$(x \wedge y) \vee (y \wedge z) \vee (z \wedge x) = (x \wedge y) \vee (y \wedge z) \vee z = (x \wedge y) \vee z$$

$$(x \vee y) \wedge (y \vee z) \wedge (z \vee x) = (x \vee y) \wedge (y \vee z) \wedge x = x \wedge (y \vee z)$$

Thus,

$$z \leq x \;\Rightarrow\; x \wedge (y \vee z) = (x \wedge y) \vee z.$$

That is L is a modular lattice. Consequently, in virtue of the modular law (Definition 2.8),

$$x \wedge [(x \wedge y) \vee (y \wedge z) \vee (z \wedge x)] = (x \wedge y \wedge z) \vee (x \wedge y) \vee (z \wedge x) = (x \wedge y) \vee (z \wedge x)$$

$$x \wedge [(x \vee y) \wedge (y \vee z) \wedge (z \vee x)] = x \wedge (y \vee z) \wedge (z \vee x) = x \wedge (y \vee z).$$

That is, $x \wedge (y \vee z) = (x \wedge y) \vee (z \wedge x)$ by

$$(x \wedge y) \vee (y \wedge z) \vee (z \wedge x) = (x \vee y) \wedge (y \vee z) \wedge (z \vee x). \qquad \square$$

We now discuss irreducible decomposition of elements of the distributive lattice.

Definition 2.9. Let L be a lattice and $a \in L$. a is called a $\vee$-*irreducible element* (simply, *irreducible element*) if $\forall\, x, y \in L$, we have $a = x \vee y \;\Rightarrow a = x$ or $a = y$. Dually, we can define $\wedge$–*irreducible element*.

If element a of lattice L can be represented by union of some $\vee$–irreducible elements x_i $(i = 1, 2, \ldots, r)$, that is,

$$a = x_1 \vee x_2 \vee \ldots \vee x_r \tag{2.1}$$

then (2.1) is called a $\vee$–*irreducible decomposition* of a (simply, *irreducible decomposition of* a). If any x_i cannot be omitted in 2.1, that is, $\vee_{j \neq i} x_j < a$, then we say such a irreducible decomposition to be *incompressible*.

Theorem 2.14. *Let L be a lattice which satisfies minimal condition. Then every element of L has a $\vee$–irreducible decomposition.*

Proof. $\forall a \in L$. If a is $\vee$–irreducible element, then the result holds evidently. Otherwise, let $a = x \vee y$ and $x, y < a (x, y \in L)$. If there exist $\vee$–irreducible decompositions of x and y, so does a. It follows that, if there does not exist $\vee$–irreducible decomposition of a, so does at the least one of x and y, and $x, y < a$. We can assume without any loss of generality that, there does not exist $\vee$–irreducible decompositions of x. Analogously to the above proof, we have that there exists x_1 such that x_1 has not $\vee$–irreducible decomposition and $x_1 < x$. .... In this way we obtain a series of elements $x_i (i = 1, 2, \dots)$ which have not $\vee$–irreducible decompositions such that $a > x > x_1 > \dots > x_n > \dots$, and this contradicts that L satisfies minimal condition. Thus, there exist $\vee$–irreducible decompositions of a. □

Definition 2.10. Let L be a lattice and $a \in L$. a is called a *strong $\vee$–irreducible element* if $\forall\, x, y \in L$, we have

$$a \leq x \vee y \;\Rightarrow\; a \leq x \text{ or } a \leq y.$$

It is clear that strong $\vee$–irreducible element must be $\vee$–irreducible element. But converse of the result is not true in general. For distributive lattice, we have the result as follows

Theorem 2.15. *Let L be a distributive lattice and $a \in L$. Then a is strong $\vee$–irreducible element if and only if a is $\vee$–irreducible element.*

Proof. If a is a $\vee$–irreducible element, $a \leq x \vee y$, then $a = a \wedge (x \vee y) = (a \wedge x) \vee (a \wedge y)$. Consequently, $a = a \wedge x$ or $a = a \wedge y$, that is, $a \leq x$ or $a \leq y$. Thus, a is strong $\vee$–irreducible element. Conversely, if a is strong $\vee$–irreducible element, it is clear to be $\vee$–irreducible element. □

From Theorem 2.14 and Theorem 2.15, in a distributive lattice, every element can be represented by union of some strong $\vee$–irreducible elements. Conversely, we have the following

Theorem 2.16. *Let L be a lattice. If every element in L can be represented by union of some strong $\vee$–irreducible elements, then L is a distributive.*

Proof. Since there are strong $\vee$–irreducible elements $p_i \in L\,(i = 1, 2, \dots, r)$ such that $a \wedge (b \vee c) = p_1 \vee p_2 \vee \dots \vee p_r$ for $\forall a, b, c \in L$, we have $p_i \leq a$ and $p_i \leq b \vee c$ $(1 \leq i \leq r)$. Also, p_i is strong $\vee$–irreducible element so $p_i \leq a \wedge b$ or $p_i \leq a \wedge c$. Consequently, $p_i \leq (a \wedge b) \vee (a \wedge c)$, and $a \wedge (b \vee c) = p_1 \vee p_2 \vee \dots \vee p_r \leq (a \wedge b) \vee (a \wedge c)$. It follows from Theorem 2.10 that L is a distributive lattice. □

The following result (whose proof remains as an exercise) shows that incompressible $\vee$–irreducible decomposition is unique in distributive lattice.

Theorem 2.17. *Let L be a distributive lattice and $a \in L$. If there are two incompressible $\vee$–irreducible decomposition of a*

$$a = x_1 \vee x_2 \vee \ldots \vee x_r = y_1 \vee y_2 \vee \ldots \vee y_s,$$

then $r = s$, *and we have* $x_i = y_i$ $(i = 1, 2, \ldots, r)$ *by properly adjusting the subscripts.*

We now discuss the infinite distributive laws in complete lattices.

Definition 2.11. Let L be a complete lattice. L is called a $\wedge$-*infinite distributive lattice* if the following D3 is satisfied.

D3: $a \wedge (\vee_{x \in N} x) = \vee_{x \in N} (a \wedge x)$ for any element a and non-empty subset N of L.

Dually, L is called a $\vee$-*infinite distributive lattice* if the following D4 is satisfied.

D4: $a \vee (\wedge_{x \in N} x) = \wedge_{x \in N} (a \vee x)$ for any element a and non-empty subset N of L.

If L satisfies both D3 and D4, we call L an *infinite distributive lattice*.

Remark 2.3. D3 and D4 are called $\vee$–infinite distributive law and $\wedge$–infinite distributive law, respectively. Clearly that $\wedge$–infinite distributive lattices and $\vee$–infinite distributive lattices must be distributive lattices.

It is easy to prove the following theorem by the commutativity.

Theorem 2.18. *In a* $\wedge$*–infinite distributive lattice L, we have*

$$(\bigvee_{x \in M} x) \wedge (\bigvee_{y \in N} y) = \bigvee_{x \in M, y \in N} (x \wedge y), \quad (\forall M, N \subseteq L).$$

In a $\vee$*–infinite distributive lattice L, we have*

$$(\bigwedge_{x \in M} x) \vee (\bigwedge_{y \in N} y) = \bigwedge_{x \in M, y \in N} (x \vee y), \quad (\forall M, N \subseteq L).$$

Remark 2.4. (1) Condition D3 and D4 are not equivalent. This is different from the relation between D1 and D2. (2) Generally, a completely distributive lattice (that is complete lattice satisfied distributive laws) is always not necessary infinite distributive lattice.

2.3 Boolean Algebra

Historically, Boolean algebras were the first lattices to be studied. They were introduced by Boole to formalize the calculus of propositions. The most important instances of Boolean algebras are the lattices of subsets of any set.

Definition 2.12. Let L be a lattice which has identity element I and zero element O, $x, y \in L$. If $x \wedge y = O$, $x \vee y = I$, then y is called a *complement* of x. If every element of L has complements, then L is called a *complemented lattice*.

The following result holds.

Theorem 2.19. *Let L and* L_i $(i \in I)$ *be lattices. Then the following assertions hold.*

(1) L is a complemented lattice if and only if L^{-1} *is a complemented lattice.*
(2) $\prod_{i \in I} L_i$ *is a complemented lattice if and only if every* L_i $(\forall i \in I)$ *is a complemented lattice.*

In a lattice L, the complement of an element of L (if it exists) is not always necessary unique, such as presented in the following diagrams:

But the complement of an element of a distributive lattice, if it exists, is unique.

Definition 2.13. *A Boolean algebra* (or *Boolean lattice*) is a lattice with an identity element I and zero element O which is distributive and complemented.

A collection of subsets of S which is closed under union and intersection, contains S and $\varnothing$, and the complement of any set in the collection is a Boolean algebra. The following theorem outlines the most important elementary properties of complements in a Boolean algebra.

Theorem 2.20. *The complement of any element x of a Boolean algebra B is uniquely determined (such a complement of x denote by x').*

Proof. Let $x \in B$ and let x' and x_1 satisfy $x \vee x' = I$, $x \wedge x_1 = O$. Then

$$x_1 = x_1 \wedge I = x_1 \wedge (x \vee x') = (x_1 \wedge x) \vee (x_1 \wedge x') = x_1 \wedge x'$$

Hence, if in addition, $x \vee x_1 = I$ and $x \wedge x' = O$, then $x' = x' \wedge x_1$, and so $x' = x_1$. This proves the uniqueness of the complement. It is clear that x is the complement of x'. The proof of theorem has been completed. □

Theorem 2.21. *Let B be a Boolean algebra and $x, y \in B$. Then we have the following assertions.*

(1) $x \wedge x' = O$, $x \vee x' = I$.
(2) $x'' = x$.
(3) $(x \vee y)' = x' \wedge y'$, $(x \wedge y)' = x' \vee y'$.
(4) $O' = I$, $I' = O$.
(5) $x \wedge y = O \Leftrightarrow x \leq y'$.
(6) $x \leq y \Leftrightarrow y' \leq x'$.

Proof. Here, we just prove (3) and (5). From commutativity, associativity and absorption (L2∼L4 of Theorem 2.3) we have

$$\begin{aligned}(x \wedge y) \wedge (x' \vee y') &= ((x \wedge y) \wedge x') \vee ((x \wedge y) \wedge y') = ((x \wedge x') \wedge y) \vee (x \wedge (y \wedge y')) \\ &= (O \wedge y) \vee (x \wedge O) = O \vee O = O.\end{aligned}$$

Similarly, we have $(x \wedge y) \vee (x' \vee y') = I$. Thus, we have $(x \wedge y)' = x' \vee y'$. In this way, we can prove that $(x \vee y)' = x' \wedge y'$. □

Let L be a lattice with an identity element I and zero element O, and B be a sublattice of L. If every element x of B has a complement x', and $x' \in B$, then we call B a *Boolean subalgebra* of L. It is evident that Boolean subalgebra must be a Boolean algebra. But converse is not true in general, that is, if sublattice B of L is Boolean algebra, B is always not necessary Boolean subalgebra of L. For instance, let L be a Boolean algebra, then interval sublattice $I[a,b](O \neq a \leq b \neq I)$ is a Boolean algebra, but is not a Boolean subalgebra of L.

The distributive lattice with an identity element I and zero element O includes the greatest Boolean subalgebra.

Theorem 2.22. *Let L be a distributive lattice with an identity element I and zero element O, and* $A = \{x \mid x \in L, x \text{ has a complement } x'\}$*. Then A is sublattice of L, and A is a greatest Boolean subalgebra of L, further.*

Proof. For any $x, y \in A$, it is clear that $x', y' \in A$. Similar to Theorem 2.21, we can prove that $x \wedge y$ and $x \vee y$ have complements, and $(x \wedge y)' = x' \vee y'$, $(x \vee y)' = x' \wedge y'$. It follows that $x \wedge y \in A$, $x \vee y \in A$. Thus, A is a sublattice of L. Clearly, A is a greatest Boolean subalgebra of L. □

Theorem 2.23. *Every complete Boolean algebra must be an infinite distributive lattice.*

Proof. Let L be a complete Boolean algebra, $\varnothing \neq M \subseteq L$, $a \in L$. We denote $u = \bigvee_{x \in M}(a \wedge x)$, a' is a complement of a, then

$$(a \wedge x) \vee a' \leq u \vee a' (\forall x \in M),$$

and by the property of distributivity we have

$$(a \wedge x) \vee a' = (a \vee a') \wedge (x \vee a') = I \wedge (x \vee a') = x \vee a'$$

Consequently,

$$x \leq x \vee a' \leq u \vee a' (\forall x \in M),$$

that is, $\bigvee_{x \in M} x \leq u \vee a'$. It follows that,

$$a \wedge (\bigvee_{x \in M} x) \leq a \wedge (u \vee a') = a \wedge u \leq u$$

According to the following principle of minimal and maximal, we have $u \leq a \wedge (\bigvee_{x \in M} x)$. Thus,

$$a \wedge (\bigvee_{x \in M} x) = u = \bigvee_{x \in M}(a \wedge x)$$

Similarly, we can prove that

$$a \vee (\bigvee_{x \in M} x) = u = \bigvee_{x \in M}(a \vee x)$$

This completes the proof of theorem. □

Remark 2.5. (1) **Principle of minimal and maximal** Let $(S, \leq)$ be a partially ordered set, $\{a_{ij} \mid a_{ij} \in S, j \in T_i\}$ $(i \in I)$ be a family of subset of S, $T = \prod_{i \in I} T_i$. Then

$$\bigwedge_{i \in I} (\bigvee_{j \in T_i} a_{ij}) \geq \bigvee_{f \in T} (\bigwedge_{i \in I} a_{ij})$$

(If sup and inf exist in above formula).

(2) Complete Boolean algebra is always not necessary a completely distributive lattice (see Definition 2.17). However, A.Tarski had proved the following

Theorem 2.24. *Let L be a complete Boolean algebra. Then the following conditions are equivalent.*

(1) L is a completely distributive lattice.
(2) L is isomorphism to power set lattice on certain set.

Now we will introduce an algebraic system, Boolean ring, which is closely related with Boolean algebra.

Definition 2.14. *A ring* is an algebraic system consisting of a non-empty set R together with two binary operations "+", "·" in R and two distinguished elements 0, $1 \in R$ such that

(1) $(R, +, 0)$ is an abelian group.
(2) $(R, \cdot, 1)$ is a semigroup.
(3) The distributive laws hold, that is, for all $x, y, z \in R$, we have

$$x \cdot (y+z) = x \cdot y + x \cdot z, (y+z) \cdot x = y \cdot x + z \cdot x.$$

Generally, in a ring R, the product $x \cdot y$ of x and y is denoted simply by xy .

Remark 2.6. In Definition 2.14, condition (1) is equivalent to following conditions (A1∼A5):

A1 $x + y \in R$;
A2 $x + y = y + x$;
A3 $(x+y)+z = x+ (y+z)$;
A4 For every $x \in R$ there exists *a zero element* $0 \in R$ such that

$$x + 0 = x = 0 + x;$$

A5 For every $x \in R$ there exists an *inverse element*-$x \in R$ such that

$$x + (-x) = 0 = (-x) + x.$$

Condition (2) is equivalent to following conditions (M1∼M3):

M1 $xy \in R$;
M2 $(xy)z = x(yz)$;
M3 There exists a *unit element* (or *identity element*) $1 \in R$ such that

$$x1 = x = 1x \quad \text{for any } x \in R.$$

Definition 2.15. Let R be a ring. If $xy = yx$ for all $x, y \in R$, then R is called a *commutative ring*. If $x^2 = xx = x$, then x is called an *idempotent element*.

Definition 2.16. A ring is called *Boolean ring* if all of its elements are idempotent.

Theorem 2.25. *Let R be a Boolean ring. Then the following assertions hold.*

(1) $x + x = 0$ *(that is,* $x = -x$*), for* $\forall x \in R$*;*
(2) $xy = yx$*, for* $\forall x, y \in R$*.*

Proof. (1) For any $x, y \in R$, we have $x^2 = x$ and $y^2 = y$. Consequently, we have

$$(x+y)^2 = x^2 + xy + yx + y^2 = x + xy + yx + y = x + y$$

that is, $xy + yx = 0$. Let $x = y$, then $xy = yx = x^2 = x$. It follows that $x + x = x^2 + x^2 = 0$.
(2) For any $x, y \in R$, we have $xy + xy = 0 = xy + yx$, that is, $xy = yx$. □

The next two theorems reveal the close relationship between Boolean algebras and Boolean rings.

Theorem 2.26. *Let L be a Boolean algebra together with identity element I and zero element O. We define the binary operations "+" and "·" as follows:*

$$x + y = (x \wedge y') \vee (x' \wedge y) \text{ and } xy = x \wedge y,\ \forall x, y, z \in R$$

Then $(L, +, \cdot)$ *is Boolean ring, where I is unit element of ring* $(L, +, \cdot)$ *and O is zero element.*

Proof. For any $x, y, z \in R$, we have

(1) $x + y = y + x$.
(2)
$$\begin{aligned}(x+y)+z &= [((x \wedge y') \vee (x' \wedge y)) \wedge z'] \vee [((x \wedge y') \vee (x' \wedge y))' \wedge z] \\ &= (x \wedge y' \wedge z') \vee (x' \wedge y \wedge z') \vee (x' \wedge y' \wedge z) \vee (x \wedge y \wedge z)\end{aligned}$$

$$\begin{aligned}x+(y+z) &= [x \wedge ((y \wedge z') \vee (y' \wedge z))'] \vee [x' \wedge ((y \wedge z') \vee (y' \wedge z))] \\ &= (x \wedge y' \wedge z') \vee (x' \wedge y \wedge z') \vee (x' \wedge y' \wedge z) \vee (x \wedge y \wedge z)\end{aligned}$$

Thus $(x+y)+z = x+ (y+z)$. Similarly $(xy)z = x(yz)$.
(3) $x(y+z) = x \wedge ((y \wedge z') \vee (y' \wedge z)) = (x \wedge y \wedge z') \vee (x \wedge y' \wedge z)$
$xy + xz = [(x \wedge y) \wedge (x \wedge z)'] \vee [(x \wedge y)' \wedge (x \wedge z)] = (x \wedge y \wedge z') \vee (x \wedge y' \wedge z)$
Thus $x(y+z) = xy + xz$. Similarly, $(y+z)x = yx + zx$.
(4) For all $x \in R$, $x + O = O + x$, that is, O is the zero element.
(5) For all $x \in R$, $x + x = O$, that is, x is the inverse element of x.

It follows that L is a ring. Also, for all $x \in R$, $x^2 = xx = x$, we have that L is a Boolean ring. □

Theorem 2.27. *Let* $(R, +, \cdot)$ *be a Boolean ring, I and O unit element and zero element respectively. We definite binary operations "*$\wedge$*" and "*$\vee$*" as follows:*

$$x \wedge y = xy \text{ and } x \vee y = x + y + xy, \forall x, y, z \in R$$

Then $(R, \vee, \wedge)$ is a Boolean lattice (i.e. Boolean algebra), 0 and 1 are zero element and identity element of lattice $(R, \vee, \wedge)$ respectively, and $x' = x + 1$ for $\forall\, x \in R$.

Proof. From Theorem 2.25, $(R, +, \cdot)$ is a commutative ring. Consequently, for any $x, y, z \in R$, we have

(1) $x \wedge y = y \wedge x$, $x \vee y = y \vee x$.
(2) $x \wedge x = x$, $x \vee x = x$.
(3) $(x \wedge y) \wedge z = (x \wedge y) \wedge z)$,

$$\begin{aligned}(x \vee y) \vee z &= (x + y + xy) + z + (x + y + xy)z \\ &= a + (y + z + yz) + x(y + z + yz) \\ &= x \vee (y \vee z).\end{aligned}$$

(4) $x \wedge (x \vee y) = x(x + y + xy) = x^2 + xy + x^2 y = x$. Similarly, $x \vee (x \wedge y) = x$. Thus, $(R, \vee, \wedge)$ is a lattice by Theorem 2.4.

(5) $x \wedge (y \vee z) = x(y + z + yz) = xy + xz + xyz = xy + xz + xyxz = (x \wedge y) \vee (x \wedge z)$. Thus, $(R, \vee, \wedge)$ is a distributive lattice by Definition 2.7.

(6) Since $x \wedge 0 = 0$ and $x \wedge 1 = x$ for any $x \in R$, 0 and 1 are zero element and identity element of lattice $(R, \vee, \wedge)$ respectively, and we have

$$x \wedge (x + 1) = x(x + 1) = x^2 + x = 0, \forall x \in R$$

$$x \vee (x + 1) = x + (x + 1) + x(x + 1) = 1, \forall x \in R$$

by Theorem 2.25. Thus, $x + 1$ is complement of x, that is, $x' = x + 1$. From (1)∼(6), we have that $(R, \vee, \wedge)$ is a Boolean lattice (or Boolean algebra). □

2.4 Completely Distributive Lattices

In this section, we will introduce concepts and properties of completely distributive lattices and minimal families, and give a theorem concerning the structure of CD lattices. Finally, introduced is the generalized order-homomorphism on completely distributive lattices.

Definition 2.17. Let L be a complete lattice. If for any family $\{a_{ij} \mid j \in J_i\}(i \in I$, $a_{ij} \in L$, I and J_i are subscript sets), we have

CD1: $\bigwedge_{i \in I}(\bigvee_{j \in J_i} a_{ij}) = \bigvee_{f \in \prod_{i \in I} J_i}(\bigwedge_{i \in I} a_{if(i)})$,
CD2: $\bigvee_{i \in I}(\bigwedge_{j \in J_i} a_{ij}) = \bigwedge_{f \in \prod_{i \in I} J_i}(\bigvee_{i \in I} a_{if(i)})$,

then L is called a *completely distributive lattice* (briefly, CD lattice). Where CD1 and CD2 is called completely distributive laws.

It has been proved that a CD lattice must be an infinite distributive lattice, and the power set lattice on a set is a CD lattice. G.N.Raney had proved the following results (refer to [1]) which is left to the reader as an exercise.

Theorem 2.28. *In complete lattices,* CD1 *is equivalent to* CD2.

Corollary 2.4. *A complete lattice L is a CD lattice if and only if one of* CD1 *and* CD2 *holds.*

Lemma 2.2. ***(Principle of Minimal and maximal)*** *Let $(S, \leq)$ be a partially ordered set. For any family $\{x_{ij} \mid j \in J_i\}$ $(i \in I, x_{ij} \in L$, I and J_i are indexing sets),*

$$\bigwedge_{i\in I}(\bigvee_{j\in J_i} x_{ij}) \geq \bigvee_{f\in\prod_{i\in I}J_i}(\bigwedge_{i\in I} x_{if(i)}).$$

Proof. For any $i_0 \in I$ and $f \in \prod_{i\in I} J_i$, it is clear that

$$\bigvee_{j\in J_i} x_{ij} \geq x_{i_0 f(i_0)} \geq \bigwedge_{i\in I} x_{i,f(i)}.$$

Consequently,

$$\bigwedge_{i\in I}(\bigvee_{j\in J_i} x_{ij}) \geq \bigwedge_{i\in I} x_{if(i)} \quad (\forall f \in \prod_{i\in I} J_i).$$ □

Theorem 2.29. *Every complete chain must be a CD lattice.*

Proof. Let L be a complete chain, $\{x_{ij} \mid j \in J_i\}$ $(i \in I$, $x_{ij} \in L$, I and J_i are indexing sets) be a subset family of L, and

$$a = \bigwedge_{i\in I}(\bigvee_{j\in J_i} x_{ij}) \text{ and } b = \bigvee_{f\in\prod_{i\in I}J_i}(\bigwedge_{i\in I} x_{if(i)}).$$

Then $b \leq a$ by the principle of minimal and maximal (Lemma 2.2). Now we show that $b \geq a$.

(1) If $y < a (y \in L)$, then for every $i \in I$, there is a $j_i \in J_i$ such that $x_{ij_i} > y$.

(2) If y is a prime under of a, then $a \leq x_{ij_i}$ for every $i \in I$. Consequently, there exists $f \in \prod_{i\in I} J_i$, $f(i) = j_i$ such that $a \leq \wedge_{i\in I} x_{if(i)} \leq b$.

(3) In the case, y is not a prime under of a for every $y < a$. $a = \vee\{y \in L \mid y < a\}$. By (1), there exists a $f \in \prod_{i\in I} J_i$ such that $y \leq \wedge_{i\in I} x_{if(i)}$ for every $y < a$. It follows that,

$$a = \vee\{y \in L \mid y < a\} \leq \bigvee_{f\in\prod_{i\in I}J_i}(\bigwedge_{i\in I} x_{if(i)}) = b.$$

Thus, $a = b$, that is that L is a CD lattice. □

Now we introduce the notion of minimal families on complete lattices.

Definition 2.18. Let L be a complete lattice, $a \in L$, $B \subset L$. B is called a *minimal family* of a if $B \neq \varnothing$ and

(1) $\sup B = a$;
(2) $\forall A \subset L$, $\sup A \geq$ a implies that $\forall x \in B$, there exists $y \in A$ such that $y \geq x$.

Example 2.3. Let $L = [0,1]$ be the lattice with the order of the numbers. Then $\forall a \in (0,1], [0,a)$ is a minimal family of a and $\{0\}$ is the minimal family of 0.

Example 2.4. Let $L = 2^X$ be the lattice with the order of set inclusion $\subseteq$. Where X is a non-empty set. Then $\forall E \in L, E \subset X$, $\{\{e\} | \ e \in E\}$ is a minimal family of E, and $\{\varnothing\}$ is the minimal family of $\varnothing$.

The following result whose proof remains as exercise is straightforward by Definition 2.13.

Theorem 2.30. *Let L be a complete lattice, $a \in L$. Then the unions of minimal families of a are minimal families of a as well. Especially, if a has a minimal family, then a has a* greatest minimal family*, i.e., the union of all maximal families of a, denoted by $\beta(a)$.*

The following property of the greatest minimal family $\beta(a)$ remains as an exercise: $\beta(a)$ is a lower set, i.e., if $x \in \beta(a)$, then for any $y \leq x$, $y \in \beta(a)$. The next theorem shows that the CD lattice can be constructed by minimal families.

Theorem 2.31. *Let L be a complete lattice. Then L is a CD lattice if and only if $\forall \alpha \in L$, a has a minimal family, and hence, $\beta(a)$ exists.*

Proof. Let L be a CD lattice and $a \in L$. Let $B = \{B \subset L | \sup B \geq a\} = \{B_i \mid i \in I\}$ and $\forall i \in I$, $B_i = \{a_{ij} \mid j \in J_i\}$. Let

$$B = \{ \bigwedge_{i \in I} a_{if(i)} | \ f \in \prod_{i \in I} J_i \}.$$

Then it is easy to prove that B is a minimal family of a.

Conversely, suppose that $\forall a \in L$, a has a minimal family. In what follows, we prove that condition CD1 is valid. In fact, let

$$a = \bigwedge_{i \in I} \bigvee_{j \in J_i} a_{ij}$$

Then it is clear that

$$a \geq \bigvee_{f \in \prod_{i \in I} J_i} (\bigwedge_{i \in I} a_{if(i)}).$$

To prove the inverse of the above inequality, let $\beta(a)$ be the greatest minimal family of a and $x \in \beta(a)$. Since for any $i \in I$,

$$\bigvee_{j \in J_i} a_{ij} \geq a,$$

hence from the definition of $\beta(a)$ that $\forall i \in I$ there exists a $j = f(i) \in J_i$ such that $a_{ij} = a_{if(i)} \geq x$ and $\bigwedge_{i \in I} a_{if(i)} \geq x$. Hence

$$\bigvee_{f \in \prod_{i \in I} J_i} (\bigwedge_{i \in I} a_{if(i)}) \geq \bigvee_{x \in \beta(a)} x = a.$$

This completes the proof. □

Let L be a CD lattice. Definite map, $\beta : L \to 2^L$ such that $a \mapsto \beta(a)$, $\forall a \in L$, where $\beta(a)$ is the greatest minimal family of element a of L. Then β is a well defined map, we say β to be the *minimal map* with respect to L.

Theorem 2.32. *Let L be a CD lattice, $a, b \in L$ and $a \leq b$. Then $\beta(a) \subseteq \beta(b)$.*

Proof. Suppose that $y \in \beta(a)$. We will prove that $y \in \beta(b)$. For this purpose, we prove that $\beta^*(b) = \beta(b) \cup \{y\}$ is a minimal family of b. In fact, since $b \geq a \geq y$, we have

$$\sup \beta^*(b) = \sup \beta(b) = b.$$

Next, suppose that $B \subset L$ and $\sup B \geq b$ and x is any fixed element of $\beta^*(b)$. If $x \in \beta(b)$, then there exists $z \in B$ such that $z \geq x$; if $x = y$, then $x \in \beta(a)$ and $\sup B \geq a$, hence there exists $z \in B$ such that $z \geq x$. This shows that $\beta^*(b)$ is a minimal family of b and hence $\beta^*(b) \subset \beta(b)$. Thus $y \in \beta(b)$. □

Theorem 2.33. *Let L be a CD lattice and $\forall i \in I$, $a_i \in L$. Then $\beta(\bigvee_{i\in I} a_i) = \bigcup_{i\in I} \beta(a_i)$.*

Proof. Let $a = \bigvee_{i\in I} a_i$, we only need to prove $\beta(a) = \bigcup_{i\in I} \beta(a_i)$. In fact, it follows from Theorem 2.23 that $\beta(a) \supset \bigcup_{i\in I} \beta(a_i)$. On the other hand, suppose that $b \in \beta(a)$. Since

$$\sup(\bigcup_{i\in I} \beta(a_i)) = \sup_{i\in I}(\sup \beta(a_i)) = \sup_{i\in I}\ a_i = a,$$

By Definition 2.13, we know that there exists $c \in \bigcup_{i\in I} \beta(a_i)$ such that $c \geq b$, say $c \in \beta(a_{i_0})$. Then we have

$$b \in \beta(a_{i_0}) \subset \bigcup_{i\in I} \beta(a_i),$$

because $\beta(a_{i_0})$ is a lower set. Hence $\beta(a) = \bigcup_{i\in I} \beta(a_i)$. □

The next theorem whose proof remains as an exercise can be proved with the aid of Theorem 2.32, 2.33 and the use of maximal mapping.

Theorem 2.34. *Let L be a CD lattice and β: $L \to 2^L$ be the minimal map with respect to L. Then the following assertions hold.*

(1) $\beta(O) = \{O\}$;
(2) $\forall a \in L$, $\beta(a) \subset \beta(I)$;
(3) *β is a union-preserving map, that is,* $\beta(\bigvee_{i\in I} a_i) = \bigcup_{i\in I} \beta(a_i)$.

In what follows, we introduce the dual concept and properties of minimal families.

Definition 2.19. Let L be a complete lattice, $a \in L$, $A \subset L$. A is called a *maximal family* of a if $A \neq \varnothing$ and

(1) $\inf A = a$;
(2) $\forall B \subset L$, $\inf B \leq a$ implies that $\forall x \in A$, there exists $y \in B$ such that $y \leq x$.

The following results (whose proofs remain as exercises) are similar to Theorem 2.21 and Theorem 2.25.

Theorem 2.35. *Let L be a complete lattice, $a \in L$. Then the unions of maximal families of a are maximal families of a as well. Especially, if a has a maximal family, then a has a* greatest maximal family, *i.e., the union of all maximal families of a, denoted by $\alpha(a)$.*

Theorem 2.36. *Let L be a CD lattice and $\alpha : L \to 2^L$ be the* maximal map *with respect to L, i.e., $\alpha : L \to 2^L$ such that $a \mapsto \alpha(a)$, $\forall a \in L$, where $\alpha(a)$ greatest minimal family of element a. Then the following assertions hold.*

(1) $\alpha(I) = \{I\}$;
(2) $\forall a \in L$, $\alpha(a) \subset \alpha(O)$;
(3) α *is a* $\wedge - \cup$ *map, that is,* $\alpha(\bigwedge_{i\in I} a_i) = \bigcup_{i\in I} \alpha(a_i)$.

In the following, the set $\{x \in L \mid x \leq a,\ a \in L\}$ will be denoted by the symbol $\downarrow a$.

Lemma 2.3. *Let L be a CD lattice, a, $b \in L$ and $b \in \alpha(a)$. Then there exists $c \in L$ such that $c \in \alpha(a)$ and $b \in \alpha(c)$.*

Corollary 2.5. *Let L be a CD lattice, $a, b \in L$ and $b \in \alpha(a)$. Then there exists a sequence $c_1, c_2, \ldots$ in L such that*

$$c_1 \in \alpha(a), c_{k+1} \in \alpha(c_k), k = 1, 2, \ldots \tag{2.2}$$

and

$$b \in \alpha(c_n), n = 1, 2, \ldots \tag{2.3}$$

Lemma 2.4. *Let L be a CD lattice, $a, b \in L$ and $b \in \alpha(a)$. Then there exists an ideal I in L such that*

(1) $a \in I \subset \downarrow b$;
(2) $\forall x \in L \backslash I$, *there exists a minimal element m of $L \backslash I$ such that $x \geq m$.*

Proof. Let $c_1, c_2, \ldots$ be the sequence given in Corollary 2.5 and $I = \bigcup_{n=1}^{\infty} \downarrow c_n$. On account of conditions 2.2 and 2.3 we know that I is an ideal in L and $\alpha \in I \subset \downarrow b$. Suppose $x \in L \backslash I$. Then $\{x\}$ is a chain contained in $L \backslash I$, and by Theorem 1.11, there exists a maximal chain $\varphi \subset L \backslash I$ which contains $\{x\}$. Let m be the infimum of φ in the complete lattice L, viz., $m = \inf_L \varphi$. We only need to prove $m \in L \backslash I$, or equivalently, $m \notin I$.

In fact, suppose that $m \in I$; then there exists $k \in N$ such that $m \in \downarrow c_k$, i.e., $m \leq c_k$. Let

$$B = \{y \vee c_k \mid y \in \varphi\}.$$

Then

$$\inf B = \bigwedge_{y\in\varphi}(y \vee c_k) = (\bigwedge_{y\in\varphi}) \vee c_k = m \vee c_k = c_k.$$

It follows from 2.2 that there exists $y \vee c_k \in B$ such that $y \vee c_k \leq c_{k+1}$. Hence $y \leq c_{k+1}$ and so $y \in \downarrow c_{k+1} \subset I$, contradicting the fact that $y \in \varphi \subset L \backslash I$. □

Definition 2.20. Let L be a lattice, then the non-null $\vee$–irreducible elements of L is called a *molecule* , and M denote set of all molecule of L. In the sequel, if L is a CD lattice, we prefer to call L a *molecular lattice* and write it in the form $L(M)$.

Only the molecular lattices contain full molecules, as is shown by the following:

Theorem 2.37. *Let L be a molecular lattice. Then each element of L is a union of $\vee$–irreducible elements. Consequently, for any $x \in L$, we have*

$$x = \vee\{a \mid a \leq x, \text{ a is a molecule of } L\}.$$

Proof. For $l \in L$, let

$$\pi(l) = \{x \in L \mid x \leq l \text{ and } x \text{ is } \vee\text{–irreducible}\}.$$

Then $\sup \pi(l) \leq l$, hence we only need to prove that $\sup \pi(l) \geq l$.

In fact, suppose that $a = sup\pi(l) \geq l$ is not true. Then there exists $b \in \alpha(a)$ such that $b \not\geq l$. Let I be the ideal defined in Lemma 2.4. Then $a \in I \subset \downarrow b$. Since $l \notin \downarrow b$, we have $l \in L \backslash I$. By Lemma 2.4, there is a minimal element m in $L \backslash I$ satisfying $m \leq l$. m is a $\vee$–irreducible element. Indeed, if $m \leq x \vee y$ and $m \not\leq x$, $m \not\leq y$, then

$$m = (m \wedge x) \vee (m \wedge y),$$

and $m \wedge x \neq m$, $m \wedge y \neq m$. Since m is a minimal element in $L \backslash I$, it follows that $m \wedge x \notin L \backslash I$, $m \wedge y \notin L \backslash I$, i.e., $m \wedge x \in I$, $m \wedge y \in I$ and so

$$m = (m \wedge x) \vee (m \wedge y) \in I$$

because I is an ideal, contradicting the fact that $m \in L \backslash I$. This shows that m is $\vee$–irreducible, hence $m \in \pi(l)$ and so $m \leq \sup \pi(l) = a$. But this implies that $m \in I$ because I is a lower set. This contradicts the fact that $m \in L \backslash I$. Hence $a = \sup \pi(l) \geq l$ and the proof has been completed. □

Let L be a CD lattice and $a \in L$. By Theorem 2.37 if $\forall x \in \beta(a)$, $[x]$ denotes the set of all $\vee$–irreducible elements which are smaller than or equal to x, then $x = \sup[x]$. Let

$$\beta^*(a) = \cup\{[x] \mid x \in \beta(a)\}.$$

Then is a minimal family of a.

Definition 2.21. Let L be a complete lattice, $a \in L$, $B \subset L$. B is said to be a *standard minimal family* of a if B is a minimal family of a and members of B are $\vee$–irreducible elements.

The following theorem whose proof remains as an exercise can be proved by virtue of Theorem 2.31 and Theorem 2.37.

Theorem 2.38. *Let L be a complete lattice. Then L is a CD lattice if and only if $\forall a \in L$, a has a standard minimal family.*

Example 2.5. Let $L = I^X$. Consider the element $x_\lambda \in L$ defined by

$$x_\lambda(t) = \begin{cases} \lambda, \ t = x \\ 0, \ t \neq x \end{cases} \tag{2.4}$$

It is clear that M=$\{x_\lambda \mid x \in X, \ \lambda \in (0, \ 1]\}$ is the set of all non-zero $\vee$–irreducible elements, and $\forall f \in I^X$,

$$\beta^*(f) = \{x_\lambda \mid f(x) \neq 0, 0 < \lambda < f(x)\}$$

is a standard minimal family of f.

Finally, we introduce the notion of generalized order-homomorphism on molecular lattices (i.e. CD lattices).

Definition 2.22. Let L_1 and L_2 be molecular lattices and $f : L_1 \to L_2$ a map. f will be called a *generalized order-homomorphism*, or briefly, *a* GOH, if the following conditions are satisfied.

(1) $f(O) = O$;
(2) f is union-preserving;
(3) f^{-1} is union-preserving, where $\forall b \in L_2$, $f^{-1}(b) = \vee\{a \in L_1 \mid f(a) \leq b\}$.

Example 2.6. (1) Let L_1=2^X, $L_2 = 2^Y$ and f: $X \to Y$ be a usual map. Then f induces a map f: $2^X \to 2^Y$ by letting $f(A)$=$\{f(x) \mid x \in A\}$ for any $A \in 2^X$. It is clear that f is a GOH.

(2) Let L_1, L_2 be molecular lattices and define

$$f(a) = O, \forall a \in L_1.$$

Then $f^{-1}(b)$=1, $\forall b \in L_2$ and f: $L_1 \to L_2$ is a GOH.

From now on we only consider GOHs which map non-zero elements into non-zero elements.

Theorem 2.39. *Let L_1 and L_2 be molecular lattices and $f : L_1 \to L_2$ a GOH. Then the following assertions hold.*

(1) f and f^{-1} are order-preserving;
(2) $f^{-1}f(a) \geq a$, $\forall a \in L_1$;
(3) $ff^{-1}(b) \leq b$, $\forall b \in L_2$;
(4) $f(a) \leq b$ if and only if $a \leq f^{-1}(b)$;
(5) $f(a) = \wedge\{b \in L_2 \mid f^{-1}(b) \geq a\}$, $\forall a \in L_1$;
(6) f^{-1}: $L_2 \to L_1$ is intersection-preserving.

Proof. It is easy to prove (1)—(5) (which are left as exercises); here we only give the proof of (6). By (4) we know that the following statements are equivalent;

(i) $a \leq f^{-1}(b_i)$, $i \in I$;
(ii) $f(a) \leq \bigwedge_{i\in I} b_i$;
(iii) $\forall i \in I$, $f(a) \leq b_i$;
(iv)$a \leq \bigwedge_{i\in I}\ f^{-1}(b_i)$.

It follows that, $a \leq f^{-1}(\bigwedge_{i\in I} b_i)$ if and only if $a \leq \bigwedge_{i\in I} f^{-1}(b_i)$. Since a is arbitrary, it follows that $f^{-1}(\bigwedge_{i\in I} b_i) = \bigwedge_{i\in I} f^{-1}(b_i)$, viz., f^{-1} is intersection-preserving. □

Theorem 2.40. *Let L_1 and L_2 be molecular lattices and $f : L_1 \to L_2 a$ GOH. Then the following assertions hold.*

(1) If a is $\vee$–irreducible in L_1, then $f\ (a)$ is $\vee$–irreducible in L_2;
(2) If B is a minimal family of a in L_1, then $f(B)$ is a minimal family of $f(a)$ in L_2;
(3) If B^* is a standard minimal family of a in L_1, then $f(B^*)$ is a standard minimal family of $f(a)$ in L_2.

Proof. (1) Let $a \in L_1$ be $\vee$–irreducible and $f(a) \leq b \vee c$. Then

$$a \leq f^{-1}(b \vee c) = f^{-1}(b) \vee f^{-1}(c).$$

Since a is $\vee$–irreducible we have $a \leq f^{-1}(b)$ or $a \leq f^{-1}(c)$, and hence

$$f(a) \leq ff^{-1}(b) \leq b \text{ or } f(a) \leq ff^{-1}(c) \leq c.$$

This proves that $f(a)$ is $\vee$–irreducible in L_2.

(2) Let B be a minimal family of a in L_2, then $\sup f(B) = f(\sup B) = f(a)$ in L_2. Suppose that $C \subset L_2$, $\sup C \geq f(a)$ and $y \in f(B)$. Then there exists $x \in B$ such that $f(x) = y$. Since

$$\sup f^{-1}(C) = f^{-1}(\sup C) \geq f^{-1}f(a) \geq a,$$

by the meaning of B we know that there exists $z \in f^{-1}(C)$ such that $x \leq z$. Now $y = f(x) \leq f(z) \in C$. This proves that $f(B)$ is a minimal family of $f(a)$ in L_2.

(3) follows directly from (1) and (2). □

Theorem 2.41. *Let $L_1(M_1)$,$L_2(M_2)$ be molecular lattices and $g : M_1 \to M_2$ a map. Then g can be extended to a GOH f: $L_1 \to L_2$ if and only if $\forall m \in M_1$, g maps the standard minimal family of m into the standard minimal family of $g(m)$.*

Proof. The necessity is a consequence of Theorem 2.40 and we only need to show the sufficiency.

Suppose that g maps standard minimal families into standard minimal families. It is easy to prove that g is order-preserving. $\forall a \in L_1$, let $\beta(a)$ be the greatest minimal family of a in L_1, and let

$$\beta^*(a) = \{m \in M_1 \mid m \leq b \ \ for \ \ some \ \ b \in \beta(a)\}.$$

Then $\beta^*(a)$ is completely determined by a and is clearly a standard minimal family of a. By the hypothesis we know that $\forall m \in M_1$, $g(\beta^*(m))$ is a standard minimal family of $g(m)$, i.e.,

$$g(\beta^*(m)) \subset M_2, \sup g(\beta^*(m) = g(m)$$

and if $C \subset L_2$, $\sup C \geq g(m)$, then $\forall y \in g(\beta^*(m))$, there exists $z \in C$ such that $y \leq z$.

Now $\forall a \in L_1$ define

$$f(a) = \begin{cases} \vee g(\beta^*(a)), & if \ \ a \neq 0 \\ 0, & if \ \ a = 0 \end{cases} \tag{2.5}$$

Then f: $L_1 \to L_2$ is a GOH.

In fact, suppose that $a = \vee_{i\in I} a_i$ where a, $a_i \in L_1$. It is clear that $\forall i$, $f(a) \geq f(a_i)$. Hence $f(a) \geq \vee_{i\in I} f(a_i)$. On the other hand, $\forall m \in \beta^*(a)$, by virtue of the fact that

$$\sup_{i\in I} \vee \beta^*(a_i) = \sup_{i\in I} \ a_i = a.$$

We know that there exists some i_0 and $m' \in \beta^*(a_{i_0})$ such that $m \leq m'$. Hence

$$g(m) \leq g(m') \leq \vee g(\beta^*(a_{i_0})) = f(a_{i_0})$$

and therefore

$$f(a) = \vee g(\beta^*(a)) \leq \vee f(a_i).$$

This proves that f is union-preserving.

By the condition that $\forall m \in M_1$, $g(\beta^*(m))$ is a standard minimal family of $g(m)$ we know that

$$\forall m \in M_1, f(m) = \vee g(\beta^*(m)) = g(m).$$

This shows that $f : L_1 \to L_2$ is an extension of g: $M_1 \to M_2$.

Let us turn to the inverse of f. On account of Theorem 2.1 it suffices to prove that f^{-1} is union-preserving. First we prove that

$$f^{-1}(b) = \vee\{a \in L_1 |\ f(a) \leq b\} = \vee\{m \in M |\ g(m) \leq b\}.$$

In fact, since

$$\{\beta^*(a) \mid \vee g(\beta^*(a)) \leq b\} \subset \{m \in M |\ g(m) \leq b\},$$

we have

$$\vee\beta^*(a) \leq \vee\{m \in M_1 \mid g(m) \leq b\}$$

whenever $f(a) \leq b$. Hence

$$f^{-1}(b) = \vee\{\vee\beta^*(a) \mid a \in L_1 \text{ and } f(a) \leq b\} \leq \vee\{m \in M_1 \mid g(m) \leq b\}$$

and hence

$$f^{-1}(b) = \vee\{m \in M_1 \mid g(m) \leq b\}.$$

Now suppose that $b = \vee_{i\in I} b_i$ where b, $b_i \in L_2$, and $m \in M_1$ satisfies $g(m) \leq b$. Consider the standard minimal family $\beta^*(m)$ of m. By the condition given in this theorem $g(\beta^*(m))$ is a standard minimal family of $g(m)$, and so $\forall m' \in \beta^*(m)$, since $\sup_{i\in I} b_i = b \geq g(m)$ it follows that there exists an i such that $g(m') \leq b_i$. Hence

$$m' \leq g^{-1}(b_i) \leq \vee_{i\in I} g^{-1}(b_i)$$

and therefore

$$m = \sup \beta^*(m) \leq \vee_{i\in I} g^{-1}(b_i)$$

whenever $g(m) \leq b$. This shows that

$$f^{-1}(b) = \vee\{m \in M_1 \mid g(m) \leq b\} \leq \vee_{i\in I} f^{-1}(b_i),$$

hence $f^{-1}(b) = \vee_{i\in I} f^{-1}(b_i)$ because the opposite inequality $f^{-1}(b) \geq \vee_{i\in I} f^{-1}(b_i)$ is obviously true. This completes the proof. □

Let L_1, L_2 and L_3 be molecular lattices and $f : L_1 \to L_2$, $g : L_2 \to L_3$ GOHs. Then $gf(O) = O$, gf is union-preserving and

$$\begin{aligned}(gf)^{-1}(c) &= \vee\{a \in L_1 \mid (gf)(a) \leq c\} = \vee\{a \in L_1 \mid g(f(a)) \leq c\} \\ &= \vee\{a \in L_1 \mid f(a) \leq g^{-1}(c)\} = f^{-1}(g^{-1}(c)).\end{aligned}$$

Hence $(gf)^{-1}$ is union-preserving and therefore $gf : L_1 \to L_3$ is a GOH. Moreover, the identity map $I : L \to L$ is a GOH. Hence we have:

Theorem 2.42. *Let $f : L_1 \to L_2$ be a mapping. Then the following assertions are equivalent.*

(1) f is an isomorphism;
(2) f is a bijective GOH;
(3) f^{-1} is a bijective GOH.

Proof. Suppose that the GOH $f : L_1 \to L_2$ is a bijection. Then it is easy to prove that f is intersection-preserving and hence f is an isomorphism.

Conversely, if f: $L_1 \to L_2$ is an isomorphism, then (f is bijection and) f is a GOH. In fact, it is easy to prove that under this condition $f(a)$=b if and only if $a = f^{-1}(b)$ where $f^{-1}(b)$ has the meaning given in the beginning of this section. In other words, for any isomorphism f: $L_1 \to L_2$, the two inverses, one of which is given in the beginning of this section and the other given in set theory, coincidentally. Now f^{-1} is also an isomorphism, so that $(f^{-1})^{-1} = f$ is union-preserving. Hence f^{-1} is a GOH because $f^{-1}(O) = O$ is true. □

2.5 Topological Molecular Lattices

In this section, we introduce the theory of topological molecular lattices [3, 4]. This theory is to treat the theories of point set topology, fuzzy topology and L—fuzzy topology in a unified way.

In what follows, we often use capital letters A, B, P, ... to denote elements of a molecular lattice $L(M)$ and use small letters a, b, m, ... to denote elements of the set of molecules M and call them "*points*".

Definition 2.23. Let $L(M)$ be a molecular lattice, $\eta \subset L$. η is said to be a *closed topology*, or briefly, *co-topology*, if $O, I \in \eta$ and η is closed under finite unions (i.e., $\vee$) and arbitrary intersections (i.e., $\wedge$), elements of η will be called closed elements, $(L(M), \eta)$ will be called a *topological molecular lattice*, or briefly, TML.

Example 2.7. (1) Let $(X, \mathscr{U})$ be a topological space. Then the lattice 2^X with the order of set inclusion is a molecular lattice and $(2^X, \eta)$ is a TML, where η=$\{X \backslash U \mid U \in \mathscr{U}\}$. Closed elements of $(2^X, \eta)$ are closed sets of $(X, \mathscr{U})$.

(2) Let I=[0,1] and X be a set. Let (I^X, δ) be a L–fuzzy topological space. Then the lattice I^X, in which $\xi, \zeta \in I^X$, $\xi \leq \zeta \Leftrightarrow$ for any $x \in X$, $\xi(x) \leq \zeta(x)$, is a molecular lattice and (I^X, η) is a TML, where η=$\{A' \mid A \in \delta\}$ and $A'(x) = 1 - A(x)$ for all $x \in X$. Closed elements of (I^X, η) are closed fuzzy sets of (I^X, δ).

(3) Let L be a lattice and X be a set. Let (L^X, δ) be a L–fuzzy topological space. Then it is a TML of which the closed elements are closed L–fuzzy sets.

Definition 2.24. Let $(L(M), \eta)$ be a TML, $a \in M$, $P \in \eta$ and $a \nleq P$. Then P is called a *remote-neighborhood* of a, and the set of all remote-neighborhoods (briefly, *R-neighborhoods*) of a will be denoted by $\eta(a)$.

Since a is $\vee$–irreducible, it is easy to verify that $P \in \eta(a)$ and $Q \in \eta(a)$ imply that $P \vee Q \in \eta(a)$. Moreover, it is clear that $P \in \eta(a)$, $Q \in \eta$ and $Q \leq P$ imply that $Q \in \eta(a)$. Hence $\eta(a)$ is an ideal in the complete lattice η.

Definition 2.25. Let $(L(M), \eta)$ be a TML, $A \in L$. Then the intersection of all closed elements containing A will be called the *closure* of A and denoted A^-.

Theorem 2.43. *Let $(L(M), \eta)$ be a TML. Then the following assertions hold.*

(2) $\forall A \in L$, $A \leq A^-$;
(2) O^-=O;
(3) $(A^-)^-$=A^-;
(4) $(A \vee B)^-$=$A^- \vee B^-$.

Definition 2.26. Let $(L(M), \eta)$ be a TML, $A \in L$, $a \in M$. Then a is called an *adherence point* of A if $\forall P \in \eta(a)$ we have $A \nleq P$. If a is an adherence point of A and $a \nleq A$, or $a \leq A$ and for each point $b \in M$ satisfying $a \leq b \leq A$ we have $A \nleq b \vee P$, then a is called *an accumulation point* of A. The union of all accumulation points of A will be called the *derived element* of A and denoted by A^d.

The element O has no adherence point because $\forall a \in M$, $\forall P \in \eta(a) \Rightarrow O \leq P$.

Theorem 2.44. *Let $(L(M), \eta)$ be a TML, $A \in L$, $a \in M$. Then the following assertions hold.*

(1) a is an adherence point of A if and only if $a \leq A^-$;
(2) A^- equals the union of all adherence points of A;

(3) $A^-{=}A \vee A^d$*;*
(4) $(A^d)^- \leq A^-$.

Proof. (1) By Definition 2.26, a is an adherence point of A if and only if $P \in \eta(a) \Rightarrow A \nleq P$, or equivalently, $A \leq P \Rightarrow P \notin \eta(a)$ for every closed P. This implies that $a \leq A^-$.

(2) We have only to consider the case $A \neq O$. By Theorem 2.37 we have $A^-{=}\vee\{a \in M |\ a \leq A^-\}$, and by (1), this means that A^- is the union of all its adherence points.

(3) We need only prove that $A^- \leq A \vee A^d$. In fact, if for some point $a \leq A^-$ and $a \nleq A$, then by (1) and Definition 2.26 we know that $a \leq A^d$.

(4) If $a \leq (A^d)^-$, then by (1) and Definition 2.26 we know that $\forall p \in \eta(a), A^d \nleq P$. Hence there exists an accumulation point b of A such that $b \nleq P$, which means $P \in \eta(b)$. But b is an adherence point of A, hence $A \nleq P$. This proves that a is an adherence point of A. □

Corollary 2.6. *An element A of* $(L(M), \eta)$ *is closed if and only if for each point* $a \nleq A$*, there exists* $P \in \eta(a)$ *such that* $A \leq P$*.*

Definition 2.27. Let $L(M)$ be a molecular lattice, $A \in L$, $A \neq O$, $m \in M$. Then m is called a *component* of A if (i) $m \leq A$, and (ii) $u \in M$, $u \geq m$ and $u \leq A$ imply that $u = m$. Components of I will be called *maximal points*. Here O and I are the minimum and maximum of $L(M)$.

Example 2.8. (1) If $L = 2^X$, $A \in L$, $A \neq 0$, then every point of A is a component of A. (2) If $L = I^X$, $A \in L$, $A \neq O$, then a point x_λ is a component of A if and only if $A(x) = \lambda$. In (1) *and* (2), if m_1 and m_2 are different components of A, then $m_1 \wedge m_2 = O$.

Theorem 2.45. *Let L be a molecular lattice,* $A \in L$*,* $A \neq O$*,* $a \in M$ *and* $a < A$*. Then A has at least a component m such that* $a \leq m$*.*

Proof. Let φ be a chain in L. We say that φ is in A if $\forall x \in \varphi$, $x \leq A$, in symbols, $\varphi \leq A$. Consider the family of chains

$$\mathscr{C} = \{\varphi \mid a \in \varphi \subset M,\ \varphi \leq A\}.$$

Since $\{a\} \in \mathscr{C}$ we have $\mathscr{C} \neq \varnothing$. Assume that

$$\varphi_1 \leq \varphi_2 \text{ if and only if } \varphi_1 \subset \varphi_2,$$

then $\mathscr{C}$ becomes a poset. It is clear that every totally ordered subset of $\mathscr{C}$ has an upper bound, and hence there exists a maximal element $\varphi_0 \in \mathscr{C}$. Let $m = \sup \varphi$. Then $a \leq m \leq A$ and one readily checks that m is $\vee$–irreducible, i.e., $m \in M$, and m is a component of A such that $a \leq m$. □

Remark 2.7. The component mentioned in Theorem 2.45 may not be unique. For example, let $L{=}\{O, I, a, m, u\}$, define $a \leq m$, $a \leq u$ and suppose m and u are incomparable. Then $L(M)$ is a molecular lattice where $M = \{a, m, u\}$. Now $A{=}I$ has two different components m and u such that both $a \leq m$ and $a \leq u$ are true.

Theorem 2.46. *Let $L(M)$ be a molecular lattice, $A \in L$. Then for each point $a \leq A$, A has a unique component $m(a,A)$ such that $a \leq m(a,A)$ if and only if different components of A are disjoint, i.e., their intersections are equal to O.*

The proof is left to the reader.

Theorem 2.47. *Let $(L(M), \eta)$ be a TML, $\forall A \in L$, different components of A are disjoint. Then the derived element of every element is closed if and only if the derived element of every point is closed.*

Proof. We only consider the sufficiency. The proof of necessity remains as a exercise. Suppose that $a \in M$and $a \leq (A^d)^-$. We have to prove that $a \leq A^d$. If $a \nleq A$, then by virtue of the fact that $a \leq A^-$we have $a \leq A^d$. Hence we may assume that $a \leq A$. Let m=$m(a, A)$ be the unique component of A such that $a \leq m$.

(i) $a \leq m^{d\cdot}$ Let $\beta^*(a)$ be the standard minimal family of a. Then we only need to prove that $\forall x \in \beta^*(a)$, $a \leq A^d$. In fact, $\forall x \in \beta^*(a)$, since $m^d = \vee\{y \mid y$ is an accumulation point of $m\} \geq a$, the point m has an accumulation point d_x such that $d_x \geq x$. By the meaning of d_x one readily verifies that $d_x \nleq m$, hence $d_x \nleq A$ (because otherwise there will be at least two components of A containing the same point x). On the other hand, $d_x \leq m^- \leq A^-$, and hence d_x is an accumulation point of A which proves that $x \leq d_x \leq A^d$.

(ii) $a \nleq m^d$. Since m^d is closed we have $m^d \in \eta(a)$. Suppose that $P \in \eta(a)$ and let P_1=$m^d \vee P$. Then $P_1 \in \eta(a)$. Note that $a \leq (A^d)^-$, i.e., a is an adherence point of A^d. We have $A^d \nleq P_{1,}$ so that A has an accumulation point c such that $c \nleq P_1$ and so $P_1 \in \eta(a)$. If $c \nleq m^-$, then $m^- \vee P_1 \in \eta(c)$ and hence $a \nleq m^- \vee P_1$; if $c \leq m^-$, then it follows by the facts $c \nleq P_1$ and $m^d \leq P_1$ that $c \nleq m^d$, hence $c \leq m \leq A$. Moreover, by the meaning of c and the fact that $c \leq A$ we know that $a \nleq m \vee P_1$. This proves that a is an accumulation point of A, i.e., $a \leq A^d$. □

Definition 2.28. Let $(L(M), \eta)$ be a TML and $\{A_i\}$ be a subset of L. Then $\{A_i\}$ is said to be *locally finite* if every point a has a R-neighborhood P_a such that $A_i \nleq P_a$ holds for at most a finite number of i.

This definition is evidently a generalization of the one available in general topology.

Theorem 2.48. *Let $\{A_i \mid i \in I\}$ be a locally finite family. Then the following assertions hold.*

(1) $\{A_i^- \mid i \in I\}$ be a locally finite family;
(2) If $B_j \leq A_j^-$ for each $j \in J \subset I$, then $\{B_j \mid j \in J\}$ be a locally finite family;
(3) $(\bigvee_{i \in I} A_i)^- = \bigvee_{i \in I} A_i^-$.

Proof. (1) Since P_a is closed we have

$$A_i \nleq P_a \text{ if and only if } A_i^- \nleq P_a.$$

(2) is trivial.

(3) $(\bigvee_{i \in I} A_i)^- \geq \bigvee_{i \in I} A_i^-$ is obviously true. Now suppose that $a \in M$ and $a \nleq \bigvee_{i \in I} A_i^-$. By virtue of (1) we know that $\{A_i^-\}$ is locally finite, and hence there exists

$P_a \in \eta(a)$ such that $A_i^- \leq P$ holds for all but a finite number of i's, say $i_1, i_2, \ldots, i_n$. Since $a \not\leq A_{ik}^-$, $k = 1, 2, \ldots, n$, there exist $p_k \in \eta(a)$ such that

$$A_{ik} \leq p_k, k = 1, 2, \ldots, n.$$

Put $P = (\bigvee_{k=1}^{n} P_k) \vee P_a$. Then $P \in \eta(a)$ and $\vee A_i \leq P$. Hence $a \not\leq (\bigvee_{i \in I} A_i)^-$. This proves (3). □

Definition 2.29. Let $L(M)$ be a molecular lattice and D a directed set and $S : D \to M$ a map. Then S is called a *molecular net* in L and denoted $S = \{s(n) \mid n \in D\}$. S is said to be in $A \in L$, if $\forall n \in D$, $s(n) \leq A$.

Definition 2.30. Let $(L(M), \eta)$ be a TML, $S=\{s(n) \mid n \in D\}$ a molecular net and a point a. a is said to be a *limit point* of S (or S converges to a; in symbols, $S \to a$), if $\forall P \in \eta(a)$, $s(n) \not\leq P$ is eventually true. a is said to be a *cluster point* of S(or S accumulates to a; in symbols, $S \infty a$), if $\forall P \in \eta(a)$, $s(n) \not\leq P$ is frequently true. The union of all limit points and all cluster points of S will be denoted by lim$S$ and ad$S$, respectively.

A limit point of S is a cluster point of S but not vice versa.

Corollary 2.7. *(1) Suppose that*

$$S=\{s(n), n \in D\} \to a(S \infty a),$$

$T=\{T(n), n \in D\}$ is a molecular net with the same domain as S and $\forall n \in D$, $T(n) \geq S(n)$ holds. Then $T \to a(T \infty a)$.

(2) Suppose that $S \to a(S \infty a)$ and $b \leq a$. Then $S \to b(S \infty a)$.

A subset ξ of η is called a *base* of η if every element of η is an intersection of elements of ξ. A subset ζ of η is called a *subbase* of η if the set consisting of all finite unions of elements of ζ forms a base of η. The proof of the following results are left as exercises.

Theorem 2.49. *Let $(L(M), \eta)$ be a TML. ξ and ζ a base and a subbase for η respectively, S a molecular net and $a \in M$. Then the following assertions hold.*

(1) $S \to a$ if and only if $\forall P \in \eta(a) \cap \zeta$, S is eventually not in P;
(2) $S \infty a$ if and only if $\forall P \in \eta(a) \cap \xi$, S is frequently not in P;

Theorem 2.50. *Let $(L(M), \eta)$ be a TML, S a molecular net, $a \in M$, and $\beta^*(a)$ a standard minimal family of a. Then $S \to a(S \infty a)$ if and only if $\forall x \in \beta^*(a)$, $S \to a(S \infty a)$.*

Theorem 2.51. *Let $(L(M), \eta)$ be a TML, S a molecular net, $a \in M$. Then the following assertions hold.*

(1) a is a limit point of S if and only if $a \leq$ lim$S$;
(2) a is a cluster point of S if and only if $a \leq$ad$S$.

Proof. (1) We only prove the sufficiency and the necessity is left to the reader. Suppose that $a \leq \lim S$ and $\beta^*(a)$ is a standard minimal family of a. Since

$$\lim S = \sup\{y|\ y \text{ is a limit point of } S\} \geq a,$$

$\forall x \in \beta^*(a)$, there exists a limit point y of S such that $x \leq y$. By Corollary 2.7(2), $S \to x$, and by Theorem 2.50 we have $S \to a$.

(2) The proof is similar to that of (1) and is omitted. □

Theorem 2.52. *Let $(L(M), \eta)$ be a TML, $A \in L$, $a \in M$. Then the following assertions hold.*

(1) If there exists in A a molecular net which accumulates to a, then $a \leq A^-$;
(2) If $a \leq A^-$, then there exists in A a molecules net which converges to a.

Proof. (1) Suppose S=$\{s(n),\ n \in D\}\infty a$ and $\forall n \in D$, $s(n) \leq A$. Then $\forall P \in \eta(a)$, $A \not\leq P$ because of the fact that $s(n) \not\leq P$ is frequently true, i.e., for any $n \in D$, there always exists $n_0 \in D$ such that $s(n_0) \not\leq P$. Hence $a \leq A^-$.

(2) Suppose that $a \leq A^-$. Then $\forall P \in \eta(a)$ there exists a point $s(P)$ such that $s(P) \leq A$ and $s(P) \not\leq P$. Define S=$\{s(P) \mid P \in \eta(a)\}$. Then S is a molecular net in A because of the fact that $\eta(a)$ is a directed set in which the order is defined by inclusion. Clearly, $S \to a$. □

Definition 2.31. Let $S = \{s(n),\ n \in D\}$ and $T = \{T(m), m \in E\}$ be two molecular nets. T is called a *subnet* of S if there exists a mapping $N : E \to D$ such that

(1) T=$S \circ N$.
(2) $\forall n \in D$, there exists $m \in E$ such that $N(k) \geq n$ whenever $m \leq k \in E$.

Theorem 2.53. *Let $(L(M), \eta)$ be a TML, S a molecular net, $a \in M$. Then $S\infty a$ if and only if S has a subnet T which converges to a.*

Proof. The sufficiency follows from the definition of subnet, and we only prove the necessity. Suppose that $S\infty a$. Then $\forall P \in \eta(a)$ and $\forall n \in D$, there exists $f(P, n) \in D$ such that

$$f(P, n) \geq n \text{ and } S(f(P,n)) \not\leq P.$$

Let

$$E = \{(f(P,n), P) \mid P \in \eta(a),\ n \in D\},$$

and define

$$(f(P_1, n_1), P_1) \leq (f(P_2, n_2), P_2) \text{ if and only if } n_1 \leq n_2 \text{ and } P_1 \leq P_2.$$

Then E is a directed set. Let

$$T(f(P,n), P) = S(f(P,n), P),$$

then T is a subnet of S and $T \to a$. □

Exercises

Exercise 2.1. Let $\leq$ be a partially ordered relation on the set S and N a subset of S. Show the following

(1) Inverse relation $\leq^{-1}$ (or $\geq$) of $\leq$, i.e., $a \leq^{-1} b \Leftrightarrow \ b \leq a$, is also a partially ordered relation on S.

(2) Induced relation $\leq^{N}$ of $\leq$ on N, i.e., for a, $b \in N$, $a \leq^{N} b \Leftrightarrow \ a \leq b$, is a partially ordered relation on N.

Exercise 2.2. Let $(S, \leq)$ be a partially ordered set and A a non-empty subset of S. Show that the following assertions hold.

(1) If A has maximum (minimum) element, then the maximum (minimum) element is unique.

(2) If A is a chain in S (e.g. linear ordered subset), then maximal (minimal) element of A (if it exists) must be maximum (minimum) element.

Exercise 2.3. Let $(S, \leq)$ be a partially ordered set. Prove that S satisfies minimal condition if and only if every chain in S is a well ordered set.

Exercise 2.4. (Axiom of choice) Let $P^*(S) = 2^S - \varnothing$. Prove that there exists a mapping φ: $P^*(S) \ \rightarrow A$ such that $\varphi(T) \in T$ for every $T \in P^*(S)$.

Exercise 2.5. (Zermelo) Prove that for any set S, there exists a linear order $\geq$ such that $(S, \geq)$ is a well ordered set.

Exercise 2.6. (Hausdorff) Prove that for any partially set $(S, \leq)$, every chain in S is included in some maximal chain.

Exercise 2.7. (Kuratowski-Zorn) Let $(S, \leq)$ be a partially ordered set. Prove that if every chain in S has an upper bound in S, then every element of S is included in some maximal element of S.

Exercise 2.8. Let $\leq$ be the ordered relation of the partially ordered set $(L, \leq)$ and $(L, \vee_1, \wedge_1)$ be a lattice. Prove that if $\leq^{-1}$ is the inverse relation of $\leq$, then the partially ordered set $(L, \leq^{-1})$ is a lattice $(L, \vee_2, \wedge_2)$, where for any $a, b \in L$,

$$a \vee_2 b = a \wedge_1 b, \ \ a \wedge_2 b = a \vee_1 b.$$

The latices $(L, \vee_1, \wedge_1)$ and $(L, \vee_2, \wedge_2)$.

Exercise 2.9. Let L be a lattice. Show the following

(a) Empty set $\varnothing$ is a sublattice of lattice L.
(b) Unit set $\{a\}$ $(a \in L)$ is a sublattice of lattice L.
(c) The intersection of any sublattice of lattice L is a sublattice of L.

Exercise 2.10. Let $(L, \leq)$ be a lattice. For any $x, y, z \in L$, show the following

L1 $x \wedge x = x$, $x \vee x = x$. (Idempotency)
L2 $x \wedge y = y \wedge x$, $x \vee y = y \vee x$. (Commutativity)

L3 $x \wedge (y \wedge z)$=$(x \wedge y) \wedge z$, $x \vee (y \vee z)$=$(x \vee y) \vee z$. (Associativity)
L4 $x \wedge (x \vee y)$=x=$x \vee (x \wedge y)$. (Absorption)

Exercise 2.11. In lattice $(L, \leq)$, prove that operations of union and meet are order-preserving. That is, $\forall x, y, z \in L$ the following assertions hold.

$$x \leq y \Leftrightarrow \ x \wedge z \leq y \wedge z, x \vee z \leq y \vee z.$$

Exercise 2.12. Let L be a lattice. show the following

(1) L is a complete lattice if and only if $L^{\cdot -1}$ (dual of L) is complete.
(2) If L is a finite lattice, then L is complete.

Exercise 2.13. Let $L$and L_i $(i \in I)$ be lattice. Show the following

(a) L is a distributive lattice if and only if $L^{\cdot -1}$ (dual of L) is distributive lattice.
(b) $\prod_{i \in I} L_i$ is distributive lattice if and only if L_i $(\forall i \in I)$ is distributive lattice.
(c) If L is a distributive lattice, then sublattices of L are also distributive lattice.

Exercise 2.14. In complete lattices, prove that CD1 is equivalent to CD2.

Exercise 2.15. Let Lbe a lattice and $a \in L$. Prove that $\beta(a)$ is a lower set, i.e., if $x \in \beta(a)$, then for any $y \leq x$, $y \in \beta(a)$.

Exercise 2.16. Let L be a CD lattice and $\beta : L \to 2^L$ be the minimal map with respect to L. Then the following assertions hold.

(1) $\beta(O)$=$\{O\}$.
(2) $\forall a \in L$, $\beta(a) \subset \beta(I)$.
(3) β is a union-preserving map, that is, $\beta(\bigvee_{i \in I} a_i)$=$\bigcup_{i \in I} \beta(a_i)$.

Exercise 2.17. Let L be a complete lattice, $a \in L$. Prove that the unions of maximal families of a are maximal families of a as well. Especially, if a has a maximal family, then a has a greatest maximal family, i.e., the union of all maximal families of a, denoted by $\alpha(a)$.

Exercise 2.18. Let L be a CD lattice and $\alpha : L \to 2^L$ be the maximal map with respect to$L$, i.e., $\alpha : L \to 2^L$ such that $a \mapsto \alpha(a)$, $\forall a \in L$, where $\alpha(a)$ greatest minimal family of element a. Show the following

(a) $\alpha(I)$=$\{I\}$.
(b) $\forall a \in L$, $\alpha(a) \subset \alpha(O)$.
(c) α is a $\wedge - \cup$ map, that is, $\alpha(\bigwedge_{i \in I} a_i)$=$\bigcup_{i \in I} \alpha(a_i)$.

Exercise 2.19. Let L be a CD lattice, a, $b \in L$and$b \in \alpha(a)$. Then there exists $c \in L$ such that $c \in \alpha(a)$ and $b \in \alpha(c)$.

Exercise 2.20. Let L be a CD lattice, a, $b \in L$ and $b \in \alpha(a)$. Prove that there exists a sequence $c_1, c_2, \ldots$ in L such that

$$c_1 \in \alpha(a),\ c_{k+1} \in \alpha(c_k),\ k\text{=}1, 2, \ldots$$

$$b \in \alpha(c_n),\ n\text{=}1, 2, \ldots$$

Exercise 2.21. Let L_1 and L_2 be molecular lattices and $f : L_1 \to L_2$ a GOH. Prove that the following assertions hold.

(a) f and f^{-1} are order-preserving.
(b) $f^{-1}f(a) \geq a, \forall a \in L_1$.
(c) $ff^{-1}(b) \leq b, \forall b \in L_2$.
(d) $f(a) \leq b$ if and only if $a \leq f^{-1}(b)$.
(e) $f(a)$=$\wedge\{b \in L_2 | f^{-1}(b) \geq a\}, \forall a \in L_1$.

Exercise 2.22. Let $(L(M), \eta)$ be a TML. $\forall A \in L$, different components of A are disjoint. Prove that the derived element of every element is closed if and only if the derived element of every point is closed.

Exercise 2.23. Suppose that

$$S=\{s(n), n \in D\} \rightarrow a(S\infty a),$$

T=$\{T(n), n \in D\}$ is a molecular net with the same domain as S and $\forall n \in D$, $T(n) \geq S(n)$ holds. Show $T \rightarrow a(T\infty a)$. Suppose that $S \rightarrow a(S\infty a)$ and $b \leq a$. Show $S \rightarrow b(S\infty a)$.

Exercise 2.24. Let $(L(M), \eta)$ be a TML. ξ and ζ a base and a subbase for η respectively, S a molecular net and $a \in M$. shown the following

(a) $S \rightarrow a$ if and only if $\forall P \in \eta(a) \cap \zeta$, S is eventually not in P.
(b) $S\infty a$ if and only if $\forall P \in \eta(a) \cap \xi$, S is frequently not in P.

Exercise 2.25. Let $(L(M), \eta)$ be a TML, S a molecular net, $a \in M$, and $\beta^*(a)$ a standard minimal family of a. Prove that $S \rightarrow a(S\infty a)$ if and only if $\forall x \in \beta^*(a)$, $S \rightarrow a(S\infty a)$.

Exercise 2.26. Let $(L(M), \eta)$ be a TML, S a molecular net, $a \in M$. show the following

(a) a is a limit point of S if and only if $a \leq \lim S$.
(b) a is a cluster point of S if and only if $a \leq \mathrm{ad} S$.

Exercise 2.27. Let B be a Boolean algebra and $x, y \in B$. Show the following

(a) $x \wedge x'$=O, $x \vee x'$=I.
(b) x''=x.
(c) O'=I, I'=O.
(d) $x \leq y \Leftrightarrow y' \leq x'$.

References

1. Birkhoff, G.: Lattice Theory, 3rd edn. American Mathematics Society Providence, RI (1967)
2. Jacobson, N.: Basic Algebra I. W H Freeman and Company, New York (1985)
3. Wang, G.J.: Theory of Topological Molecular Lattices. Fuzzy Sets and Systems 47, 351–376 (1992)
4. Pu, B.M., Liu, Y.M.: Fuzzy Topology I, Neighborhood structure of Fuzzy Point and Moore-Smith Convergence. Journal of Mathematical Analysis and Applications 76, 571–599 (1980)

Part II
Methodology and Mathematical Framework of AFS Theory

Chapter 3
Boolean Matrices and Binary Relations

Considering the three types of information-driven tasks where graded membership plays a role: classification and data analysis, decision-making problems, and approximate reasoning, Dubois gave the corresponding semantics of the membership grades, expressed in terms of similarity, preference, and uncertainty [1]. For a fuzzy concept ξ in the universe of discourse X, by comparison of the graded membership (Dubois interpretation of membership degree), an "empirical relational membership structure " $\langle X, R_\xi \rangle$ is induced [5, 6], where $R_\xi \subseteq X \times X$ is a binary relation on X, $(x,y) \in R_\xi$ if and only if an observer, an expert, judges that "x belongs to ξ at some extent and the degree of x belonging to ξ is at least as large as that of y. The fundamental measurement of the gradual-set membership function can be formulated as the construction of homomorphisms from an "empirical relational membership structure", $\langle X, R_\xi \rangle$, to a "numerical relational membership structure", $\langle \{\mu_\xi(x) \mid x \in X\}, \leq \rangle$.

In this chapter, we present some new results that help us to analyze and study the structures of concepts via mathematical tools such as binary relations, Boolean matrices and lattices. Some properties of the binary relations are represented and explored by Boolean matrix theory and lattice theory. This is possible as there exists a one to one correspondence between the Boolean matrices and the binary relations and a lattice can be established on the set of the Boolean matrices.

Based on the main theorems in this chapter, we will prove, in next chapter, that any fuzzy concept ξ in a finite set X can be represented by some very simple concepts on X. Thus AFS theory offers a great deal of modeling capabilities which help model both the mathematical structures and the semantics of human concepts.

In what follows, we will define the Boolean matrices and their operations. Here the Boolean algebra $\{0,1\}$ is an algebra system with operations, +, $\cdot$ such that 0+0=0$\cdot$0=0$\cdot$1=1$\cdot$0=0, 0+1=1+1=1$\cdot$1=1.

Definition 3.1. A $m \times n$ matrix on the Boolean algebra $\{0, 1\}$ is called a *Boolean matrix*. The set of all $m \times n$ Boolean matrices is denoted by $M(B)_{m\times n}$. Let $A = (a_{ij}) \in M(B)_{m\times n}$. Then a_{ij} is called a (i, j) *element*. If all $a_{ij} = 0$, then we call A

X. Liu and W. Pedrycz: Axiomatic Fuzzy Set Theory and Its Applications, STUDFUZZ 244, pp. 97–110.
springerlink.com

a *zero matrix*. If all $a_{ij} = 1$, then we call A a *universal Boolean matrix*. If $i = j \Rightarrow a_{ij} = 1$ and $i \neq j \Rightarrow a_{ij} = 0$, then A is called a *unit matrix* (or *identity matrix*).

Definition 3.2. Let $A = (a_{ij}), B = (b_{ij}) \in M(B)_{m\times n}$. Then $C = A + B = (c_{ij}) \in M(B)_{m\times n}$, is defined as $c_{ij} = a_{ij} + b_{ij}$ and called the *sum of A and B*; Let $A = (a_{ij}) \in M(B)_{m\times q}, B = (b_{ij}) \in M(B)_{q\times n}$. Then $C = AB = (c_{ij}) \in M(B)_{m\times n}$, is defined as follows

$$c_{ij} = \sum_{k=1}^{q} a_{ik} b_{kj},$$

and called the *product of A and B*.

Boolean matrices and their sum and product have almost all the properties of the matrices expressed for real numbers. The readers can refer to [3] for further details.

Example 3.1. Let

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Then

$$A + B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}, AB = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

by Definition 3.2.

Definition 3.3. The *binary operation* "$*$" in $M(B)_{m\times n}$ defined as follows

$$C = M * N = (c_{ij}) \in M(B)_{m\times n} \text{ such that } c_{ij} = m_{ij} n_{ij},$$

is called $*$ *product of Boolean matrices*, where $N = (n_{ij}), M = (m_{ij}) \in M(B)_{m\times n}$.

It is clear that for any $H, N, M \in M(B)_{m\times n}$, $H * (N * M) = (H * N) * M$; $M * N = N * M$, $H * (N + M) = H * N + H * M$.

Their proofs remain as exercise.

Definition 3.4. The in $M(B)_{m\times n}$ is defined as follows

$$A = (a_{ij}) \leq B = (b_{ij}) \Leftrightarrow a_{ij} = 1 \text{ implies } b_{ij} = 1$$

where $A, B \in M(B)_{m\times n}$.

It can prove that $\leq$ is a partially ordered relation and $(M(B)_{m\times n}, \leq)$ is a lattice in which $A \vee B = A + B$, $A \wedge B = A * B$ for any $A, B \in M(B)_{m\times n}$. Furthermore $(M(B)_{m\times n}, \leq)$ is a distributive lattice. It is convenient to study Boolean matrices via the techniques of lattice theory.

Definition 3.5. Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set and R be a binary relation on set X. A Boolean matrix $M_R = (r_{ij}) \in M(B)_{n\times n}$ is called *correspondent matrix of R* if

$$r_{ij} = 1 \Leftrightarrow (x_i, x_j) \in R, x_i, x_j \in X.$$

It is clear that there exists a one to one correspondence between the Boolean matrices and the binary relation via Definition 3.5. If R is a quasi-ordered relation on X, then correspondent matrix M_R is called a *quasi-ordered Boolean matrix*. The following lemma gives a sufficient and necessary condition which characterizes quasi-ordered Boolean matrix. It is convenient for us to verify whether a binary relation is a quasi-order relation.

Theorem 3.1. *Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set, R be a binary relation on X, and M_R be correspondent matrix of R. Then R is quasi-ordered relation (or, M_R is a quasi-ordered Boolean matrix) if and only if*

$$M_R^2 = M_R \text{ (i.e., } M_R \text{ is idempotent) and } M_R + I = M_R$$

Proof. Let M_R=(m_{ij}) and M_R^2=(n_{ij}). The proof of the necessity condition. Since M_R is a quasi-ordered Boolean matrix, it is clear that $M_R + I$=M_R by Definition 3.3 and

$$M_R^2 = I^2 + 2M_R + M_R^2 = I + M_R + M_R^2 \geq M_R.$$

Also, as n_{ij}=1 implies 1=$n_{ij} = \sum_{1\leq k\leq n} m_{ik}m_{kj}$. Thus $\exists k \in \{1,\ 2,\ \ldots,\ n\}$ such that $m_{ik}m_{kj}$=1. Consequently, m_{ik}=1 and m_{kj}=1. Since M_R is a quasi-ordered matrix, we have that m_{ik}=1 and m_{kj}=1 $\Leftrightarrow$ $(x_i,\ x_k) \in R$ and $(x_k,\ x_j)\ \in R$. It follows that $(x_i,\ x_j)\ \in R$ by transitivity, this implies that m_{ij}=1 and $M_R \geq M_R^2$. Thus, M_R^2=M_R by "$\geq$" is a partially ordered relation.

The proof of the sufficiency. Since M_R is a correspondent matrix of R and $M_R + I$=M_R, M_R is a Boolean matrix whose elements on the main diagonal are 1. Consequently, R satisfies reflexivity by Definition 3.5. Also, if $(a_i,\ a_k) \in R$ and $(a_k,\ a_j) \in R$ for $\forall i,\ j,\ k \in \{1,\ 2,\ \ldots,\ n\}$, i.e., m_{ik}=m_{kj}=1, then we have m_{ij}=n_{ij}=$\sum_{1\leq k\leq n} m_{ik}m_{kj}$=1 by M_R^2=M_R. It follows that R satisfies transitivity. Thus, R is a quasi-ordered relation. □

By Theorem 3.1, we have an equivalent definition of quasi-ordered Boolean matrix as follows: $M \in M(B)_{n\times n}$ is called a *quasi-ordered Boolean matrix* if

$$M^2 = M \text{ (i.e., } M \text{ is idempotent) and } M + I = M.$$

Theorem 3.2. *Let M, N be two idempotent Boolean matrices of order n. If $I + N = N$, $I + M$=M, then $(M * N)^2$=$M * N$.*

Proof. It is clear that M and N are quasi-ordered Boolean matrices. Let M=(m_{ij}) and M=(n_{ij}), then $M * N$=$(m_{ij}n_{ij})$. Consequently,

$$(M * N)^2 = (b_{ij}) = (\textstyle\sum_{1\leq k\leq n} m_{ik}n_{ik}m_{kj}n_{kj}) = (\sum_{1\leq k\leq n} m_{ik}m_{kj}n_{ik}n_{kj}).$$

If b_{ij}=1, then $\exists k \in \{1,\ 2,\ \ldots,\ n\}$ such that $m_{ik}m_{kj}n_{ik}n_{kj}$=1, that is, m_{ik}=1, m_{kj}=1, n_{ik}=1, and n_{kj}=1. It follows that, m_{ij}=1 and n_{ij}=1 by transitivity, that is, $m_{ij}n_{ij}$=1. If b_{ij}=0, it is clear that m_{ik}=0, m_{kj}=0, n_{ik}=0, or n_{kj}=0 for all $k \in \{1,\ 2,\ \ldots,\ n\}$. Consequently, m_{ij}=0 or n_{ij}=0, that is, $m_{ij}n_{ij}$=0. These facts imply that $(M * N)^2 = M * N$. □

Definition 3.6. Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set and R be a quasi-ordered relation on X. Then R is called a *quasi-linear ordered relation* if for any $x, y \in X$, the following conditions are satisfied.

(1) $(x,x) \in R$;
(2) either $(x,y) \in R$ or $(y,x) \in R$.

The correspondent matrix M_R of R is called a *quasi-linear ordered Boolean matrix*. Also, if R is a linear ordered relation on X such that

$$M_R = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \Delta$$

then R is called *canonical linear ordered relation*, Δ is called *canonical linear ordered matrix.*

Theorem 3.3. *Let $M \in M(B)_{n\times n}$. If $M^2 = M$ and $\Delta \leq M$, then M is the matrix as follows*

$$\begin{bmatrix} J_{11} & J_{12} & \cdots & J_{1k} \\ 0 & J_{22} & \cdots & J_{3k} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & J_{kk} \end{bmatrix},$$

Where J_{ij} are the matrices whose elements are all 1, i.e., the universal Boolean matrix.

Proof. Since $I \leq \Delta \leq M$, we have that $M + I = M$. Hence M is a quasi-ordered matrix by Theorem 3.1. Thus M is a quasi-linear ordered matrix by $\Delta \leq M$. From Definition 3.6 we have that, the quasi-ordered relation R which corresponds to M is a quasi-linear ordered relation. We now can assume without any loss of generality that $M = (m_{ij})$, $m_{ij} = 1$ for $i \geq j$. Let us consider the principle sub-block $M(j, j+1, \ldots, i)$ of M. Since $r \leq s$ implies $m_{rs} = 1$ by $\Delta \leq M$, in addition $m_{ij} = 1$, we have $m_{jk} = 1$ and $m_{li} = 1$, for any $k, l \in \{j, j+1, \ldots, i\}$. It follows that, $m_{lk} = m_{li} m_{ij} m_{jk} = 1$ from transitivity. This completes the proof of the theorem. □

By Theorem 3.1, we can verify that the following Boolean matrix is a quasi-linear ordered matrix. We call the matrix which is the following form

$$\begin{bmatrix} J_{11} & J_{12} & \cdots & J_{1k} \\ 0 & J_{22} & \cdots & J_{3k} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & J_{kk} \end{bmatrix} \tag{3.1}$$

canonical quasi-linear ordered form. In what follows, we will prove the main theorem which show that each quasi-ordered relation can be represented by some quasi-linear ordered relations.

Definition 3.7. Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set and R be a quasi-ordered relation on X. Let S_X be the *set of all minimal elements of* X. Then for each $x \in S_X$, we define,

$$\mathbf{m}_x = \{C \mid C \subseteq X, C \text{ is a maximum quasi-order chain which contains } x\},$$

$\mathbf{m}_x$ is called the set of *maximum quasi-order chains at x.*

Theorem 3.4. *Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set and R be a quasi-ordered relation on X. Then there exist quasi-linear orders $R_1, R_2, \ldots, R_r$ such that*

$$M_R = M_{R_1} * M_{R_2} * \ldots * M_{R_r} = * \textstyle\prod_{i=1}^{r} M_{R_i}.$$

Proof. It is obvious that for any $x \in X$, $(x, u) \in R \Leftrightarrow x \in \cup\{B \mid B \in \mathbf{m}_u\}$. Since u is the minimal element and the definition of $\mathbf{m}_u$, hence for any $x \in X - \cup\{B \mid B \in \mathbf{m}_u\}$, for any $y \in \cup\{B \mid B \in \mathbf{m}_u\}$, we have $(x, u) \notin R$, $(x, y) \notin R$, $(u, x) \notin R$.

Let $u \in X$ and $A \in \mathbf{m}_u$. For any $a \in \cup\{B \mid B \in \mathbf{m}_u\} - A$, we define $A_a = \{x \mid x \in A, (a, x) \in R\}$. Since $(a, u) \in R$, hence $A_a \neq \varnothing$. Because A is a quasi-linear chain, A_a is a quasi-linear chain. Let LA_a be the maximum element of A_a (if there exist more than one maximum elements, let any one of them be LA_a). For any $a \in A$, we define $LA_a = a$. For any $A \in \mathbf{m}_u$, we construct a binary relation R_A on X as follows:

For any $x, y \in X = (\cup\{B \mid B \in \mathbf{m}_u\}) \cup (X - \cup\{B \mid B \in \mathbf{m}_u\})$, we define

$$(x, y) \in R_A \Leftrightarrow (LA_x, LA_y) \in R, \text{ if } y, x \in \cup\{B \mid B \in \mathbf{m}_u\};$$
$$(x, y) \in R_A \text{ and } (y, x) \in R_A, \text{ if } x, y \in X - \cup\{B \mid B \in \mathbf{m}_u\};$$
$$(x, y) \in R_A, \text{ if } x \in \cup\{B \mid B \in \mathbf{m}_u\} \text{ and } y \in X - \cup\{B \mid B \in \mathbf{m}_u\}.$$

Since sets $\cup\{B \mid B \in \mathbf{m}_u\}$ and $(X - \cup\{B \mid B \in \mathbf{m}_u\})$ are the partition of set X, hence the above binary relation R_A is defined well. Furthermore, we prove that R_A is a quasi-linear order on X. For any $x, y, z \in X$, suppose $(x, y), (y, z) \in R_A$. If $z \in X - \cup\{B \mid B \in \mathbf{m}_u\}$, then $(x, z) \in R_A$; if $z \in \cup\{B \mid B \in \mathbf{m}_u\}$, then by the definition of R_A, we have $x, y \in \cup\{B \mid B \in \mathbf{m}_u\}$ and $(LA_x, LA_y) \in R$ and $(LA_y, LA_z) \in R$. Because R is a quasi-ordered relation on X, $(LA_x, LA_z) \in R \Rightarrow (x, z) \in R_A$. For any $x, y \in X$, one can check either $(x, y) \in R_A$ or $(y, x) \in R_A$. Therefore R_A is a quasi-linear order.

Next, we prove that for $x, y \in X$, $(x, y) \in R \Rightarrow (x, y) \in R_A$. For $x, y \in X$, suppose $(x, y) \in R$. If $x, y \in \cup\{B \mid B \in \mathbf{m}_u\}$ or $x, y \in X - \cup\{B \mid B \in \mathbf{m}_u\}$, then by the definitions of R_A, we know that $(x, y) \in R_A$. In the case that $x \in \cup\{B \mid B \in \mathbf{m}_u\}$ and $y \in X - \cup\{B \mid B \in \mathbf{m}_u\}$, we also have $(x, y) \in R_A$ by the above definition of R_A. Assume that $y \in \cup\{B \mid B \in \mathbf{m}_u\}$ and $x \in X - \cup\{B \mid B \in \mathbf{m}_u\}$. Since $y \in \cup\{B \mid B \in \mathbf{m}_u\}$, hence $(y, u) \in R$. By $(x, y) \in R$, we have $(x, u) \in R \Rightarrow x \in \cup\{B \mid B \in \mathbf{m}_u\}$. It contradicts the assumption $x \in X - \cup\{B \mid B \in \mathbf{m}_u\}$. Therefore for $x, y \in X$, $(x, y) \in R \Rightarrow (x, y) \in R_A$. This implies that for any $A \in \mathbf{m}_u$, any $u \in S_X$, the following assertions hold.

$$M_R \leq M_{R_A}, \;\; M_R \leq * \prod_{A \in m_u, u \in S_X} M_{R_A}. \tag{3.2}$$

Let $* \prod_{A \in m_u, u \in S_X} M_{R_A} = N = (n_{ij})$, $\;M_R = (m_{ij})$. From the definition of S_X, we have

$$X = \cup\{B \mid B \in \mathbf{m}_u, u \in S_X\}.$$

Suppose m_{hk}=0. This means $(x_h, x_k) \notin R$. Since

$$x_h, x_k \in X = \cup\{B \mid B \in \mathbf{m}_u, u \in S_X\},$$

hence for x_h, x_k, there are the following situations:

1) there exists $A \in \mathbf{m}_u$ such that $x_h, x_k \in A$;
2) there exist $A, B \in \mathbf{m}_u$ such that $x_h \in A$, $x_k \in B$;
3) there exist $A \in \mathbf{m}_u$, $B \in \mathbf{m}_v$ for some $u, v \in S_X$, $u \neq v$ such that $x_h \in A$, $x_k \in B$.

We will prove that $n_{hk} = 0$ in any situations. This implies that $N \leq M_R$.

1) From the definition of R_A, we know that $(x_h, x_k) \in R_A \Leftrightarrow (x_h, x_k) \in R$. By (3.2), we have n_{hk}=0.

2) Assume that $(x_h, x_k) \in R_B$. By the definition of LB_{x_h}, we have (x_h, LB_{x_h}), $(LB_{x_h}, x_k) \in R$. This implies that $(x_h, x_k) \in R$ and contradicts to $m_{hk} = 0$. By (3.2), we have $n_{hk} = 0$.

3) If $x_h \in \cup\{B \mid B \in \mathbf{m}_v\}$, then it is the same as situation 2); We can suppose $x_h \in X - \cup\{B \mid B \in \mathbf{m}_v\}$, by the definition of R_B, we know $(x_h, x_k) \notin R_B$, and $n_{hk} = 0$. Finally we have $M_R \geq N \geq M_R \Rightarrow M = N$. □

In what follows, applying the lattice theory, we will study the representation of the quasi-linear ordered relation by some more simple binary relations whose correspondence Boolean matrices are idempotent prime matrices.

Definition 3.8. Let $B, X, Y \in M(B)_{n\times n}$, and $B^2 = B$. B is called an *idempotent prime matrix* if

$$B = X * Y = X \wedge Y, X^2 = X \text{ and } Y^2 = Y \Rightarrow B = X \text{ or } B = Y.$$

Let $A \in M(B)_{n\times n}$. We define

$$\begin{aligned} p(A) &= \{M_R \in M(B)_{n\times n} \mid R \textit{ is a linear ordered relation such that } M_R \leq A\}; \\ i(A) &= \{M \in M(B)_{n\times n} \mid M^2 = M \textit{ and } M \leq A\}; \\ li(A) &= \{M \in i(A) \mid M \textit{ is a maximal element of } i(A)\}. \end{aligned}$$

The proof of the following results is left to the reader.

Theorem 3.5. *Let $A, B \in M(B)_{n\times n}$, and P is a permutation matrix of order n. We have the following assertions.*

(1) If $A \leq B$ and $M \in M(B)_{n\times n}$, then $MA \leq MB$ and $AM \leq BM$;
(2) $P(A * B)P^T = (PAP^T) * (PBP^T)$;
(3) $p(PAP^T) = \{PM_RP^T \mid M_R \in p(A)\} \triangleq Pp(A)P^T$;
(4) $p(A * B) = p(A) \cap p(B)$;
(5) $i(A + B) \supseteq i(A) \cup i(B)$;
(6) $i(PAP^T) = \{PMP^T \mid M \in i(A)\}$.

Theorem 3.6. *Let $A \in M(B)_{n\times n}$, and $A^2 = A$. If $p(A) \neq \varnothing$, then there exists a permutation matrix P such that PAP^T is the canonical quasi-linear order form.*

Proof. Since $p(A) \neq \varnothing$, we can assume that $M_R \in p(A)$, M_R corresponds to a linear order R of X=$\{x_1, x_2, \ldots, x_n\}$. It is obvious that any two linear ordered relations of a finite set are isomorphisms, therefore there exists a permutation matrix P such that $PM_RP^T = \Delta \leq PAP^T$. It follows (see Theorem 3.3) that PAP^T is the standard quasi-linear ordered form. □

Lemma 3.1. *The Boolean matrices shown as follows*

$$P\begin{bmatrix} J_{11} & J_{12} \\ 0 & J_{22} \end{bmatrix}P^T \tag{3.3}$$

and

$$P\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix}P^T \tag{3.4}$$

are the idempotent prime elements of $M(B)_{n\times n}$. Here P is any permutation Boolean matrix and all J_{ij} are the same as those in Theorem 3.3.

Proof. Suppose that A is the form given by (3.3), that is,

$$A = P\begin{bmatrix} J_{11} & J_{12} \\ 0 & J_{22} \end{bmatrix}P^T$$

Let A=$X * Y$ and $X, Y \in M(B)_{n\times n}$, X^2=X, Y^2=Y. We have

$$N = \begin{bmatrix} J_{11} & J_{12} \\ 0 & J_{22} \end{bmatrix} = (P^TXP) * (P^TYP)$$

It follows that, $N \leq P^TXP$, $N \leq P^TYP$, and so P^TXP and P^TYP are the canonical quasi-linear ordered forms by Theorem 3.3. Therefore either P^TXP or P^TYP must be equal to N. Let $P^TXP = N$, we have $X = PNP^T = A$. Suppose that B is the form shown in (3.4), that is,

$$B = P\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix}P^T$$

Let B=$X * Y$ and $X, Y \in M(B)_{n\times n}$, X^2=X, Y^2=Y. We have

$$N = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix} = (P^TXP) * (P^TYP)$$

It follows that $N \leq P^T XP$, $N \leq P^T YP$, and so $P^T XP$ and $P^T YP$ are the forms shown as

$$W = \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & \cdots & 1 & 1 \\ b_1 & b_2 & \cdots & b_{n-1} & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} J & a \\ b & c \end{bmatrix} = V^2 = \begin{bmatrix} J+ab & Ja+ca \\ bJ+cb & ba+c \end{bmatrix}$$

where a=$[1,\ldots,1]^T$, b=$[b_1, b_2, \ldots, b_{n-1}]$. If $b \neq [0,\ldots,0]$, then $bJ+cb = [1,1,\ldots,1]$. If $b = [0,\ldots,0]$, then $bJ+cb = [0,\ldots,0]$. This implies that $P^T XP$ and $P^T YP$ take on one of the following forms

$$\begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & \cdots & 1 & 1 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & \cdots & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & \cdots & 1 & 1 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

Therefore either $P^T XP$ or $P^T YP$ must be equal to N. Let $P^T XP = N$, we have $X = PNP^T = B$, and this completes the proof. □

Definition 3.9. Let $A = (a_{ij}) \in M(B)_{n\times n}$ be a canonical quasi-linear ordered form shown as (3.1). $subtr(A)$ the *sub-trace of A* is defined as follows.

$$\text{subtr}(A)=(a_{21}\ a_{32} \ldots a_{n(n-1)})$$

which is a $n-1$ dimension Boolean vector.

It is obvious that subtr($A * B$)=subtr(A)$*$subtr(B), subtr($A+B$)=subtr(A)+subtr(B). Let $A, B \in M(B)_{n\times n}$ and A, B be idempotent matrices such that $\Delta \leq B, \Delta \leq A$. If subtr($A$)=subtr($B$). Then we have A=B by Theorem 3.3. Therefore, the following map

$$\text{subtr: } \{B \in M(B)_{n\times n} |\ B \text{ is an idempotent matrix such that } \Delta \leq B\} \to M(B)_{n-1},$$

is an isomorphism from the algebra system $(M(B)_{n\times n}, +, *)$ to the algebra system $(M(B)_{n-1}, +, *)$, where $M(B)_{n-1}$ is the set of $n-1$ dimension vectors.

Theorem 3.7. *Let A^2=$A \in M(B)_{n\times n}$. If $p(A) \neq \varnothing$, then there exists a unique group of idempotent prime elements A_1, A_2, ..., A_k such that $A = A_1 * A_2 * \ldots * A_k$ which is irreducible.*

Proof. Let $M \in p(A)$, M be a linear ordered matrix. There exists a permutation matrix P such that PMP^T=$\Delta \leq PAP^T$. From Theorem 3.3, we know that PAP^T is of the following form

$$\begin{bmatrix} J_{11} & J_{12} & \cdots & J_{1k} \\ 0 & J_{21} & \cdots & J_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & J_{kk} \end{bmatrix}$$

Where J_{ii} is square matrix of order r_i (i = 1, 2, ..., k),

$$subtr(PAP^T) = (\overbrace{1\cdots 1}^{r_1-1}0\overbrace{1\cdots 1}^{r_2-1}\cdots 0\overbrace{1\cdots 1}^{r_k-1}),$$

there exists a unique vector sequence $V_1, V_2, \ldots, V_k$ such that every $\text{subtr}^{-1}(V_i)$ (i=1, 2, ..., k) is an idempotent prime matrix and

$$\text{subtr}(PAP^T)=V_1 * V_2 * \ldots * V_k \ (V_i=(1\ 1 \ldots 1\ 0\ 1 \ldots 1),\ i = 1,2,\ldots,k).$$

Therefore

$$PAP^T=\text{subtr}^{-1}(V_1)*\ \text{subtr}^{-1}(V_2) * \ldots * \ \text{subtr}^{-1}(V_k)$$

$$A=(P^T\text{subtr}^{-1}(V_1)P) * (P^T\text{subtr}^{-1}(V_2)P)*\ldots*(P^T\ \text{subtr}^{-1}(V_k)P)$$

Each $P^T\text{subtr}^{-1}(V_i)P$ is an idempotent prime element (i=1, 2, ..., k). Assume A=$B_1 * B_2 * \ldots * B_r$ which is irreducible. Then

$$\text{subtr}(PAP^T)=(P^T\text{subtr}^{-1}(B_1)P) * (P^T\text{subtr}^{-1}(B_2)P)*\ldots*(P^T\text{subtr}^{-1}(B_r)P)$$

Since PB_iP^T is standard quasi-linear ordered form and idempotent prime element (i=1, 2, ..., k), hence for every V_i there exists a unique PB_wP^T such that subtr (PB_wP^T)=V_i (i=1, 2, ..., k). This implies that $P^T\text{subtr}^{-1}(V_i)P$=$B_w$. This completes the proof of the theorem. □

Theorem 3.7 implies that each quasi-linear ordered relation can be uniquely represented by some simpler quasi-linear ordered relations whose correspondence matrices are forms shown in the form (3.3) and (3.4). The following theorem shows that each binary relation R on X which satisfies $(x, x) \in R$ for any $x \in X$ can be represented by some quasi-ordered relations.

Theorem 3.8. *Let $M \in M(B)_{n\times n}$ and $I+M$=M. Then there exist $A_1, A_2, \ldots, A_r$ which are idempotent Boolean matrices such that $M = A_1 + A_2 + \ldots + A_r$.*

Proof. Let the Schein rank of M is s [3]. Then there exist $n \times s$ Boolean matrix U=(a_1, a_2, ..., a_s) and $s \times n$ Boolean matrix V=(b_1, b_2, ..., b_s)^T such that $M = UV = a_1b_1 + a_2b_2 + \ldots + a_sb_s$. Thus we have $I+M$=$I+a_1b_1+I+a_2b_2+\ldots+I+a_sb_s$=$M$. Let A_i=$I+a_ib_i$, i=1, 2, ..., s.

Then

$$A_iA_i=I+a_ib_i+a_ib_ia_ib_i.$$

Suppose that a_i=(a_{i1}, ..., a_{in}), b_i=(b_{i1} ,..., b_{in}). One has

$$A_iA_i=I+a_igb_i+a_ib_i=A_i,$$

where a_ib_i=$a_{i1}b_{i1}+\ldots+a_{in}b_{in}$=$g$. □

Theorem 3.9, which can be directly proved by Theorem 3.4, Theorem 3.7 and Theorem 3.8, states that each binary relation R on X which satisfies $(x, x) \in R$ for any $x \in X$ can be represented by some quasi-linear ordered relations whose correspondence Boolean matrices shown as (3.3) and (3.4). This result implies that any concept can be represented by the concepts whose binary relations are as simple as the binary relations shown as (3.3) and (3.4).

Theorem 3.9. *Let $M \in M(B)_{n\times n}$ and $I+M$=M. Then there exist quasi-linear orders R_{ij}, i=1,2,...,r, j=1,2,...,q_i such that*

$$M = \sum_{i=1}^{r}(*\prod_{k=1}^{q_i} M_{R_{ik}})$$

Example 3.2. Let X=$\{x_1, x_2, x_3, x_4\}$ be the set of four persons. Concept ξ is "beautiful". We also know that x_1 is the most beautiful of x_1, x_2, x_3. Both x_2 and x_3 are more beautiful than x_4. x_2 and x_3 are incomparable. x_4 is more beautiful than x_1. Thus we have the following correspondence Boolean matrix of the binary relation R of the concept ξ.

$$M_R = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

By Theorem 3.8, we have

$$A_1 = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, A_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

$$M_R = A_1 + A_2 + A_3, A_i^2 = A_i, I + A_i = A_i, i\text{=}1, 2, 3.$$

According to Theorem 3.1, we know that the binary relations R_i on X corresponding to A_i, $i = 1,2,3$ are quasi-ordered relations on X. Applying Theorem 3.4, we have the following quasi-linear order decompositions of R_1, R_2, R_3. First we get the set of all minimal elements of X under the quasi-ordered relation R_1, S_X=$\{x_2, x_3, x_4\}$ and the sets of $\mathbf{m}_u$, the maximum quasi-order chains at $u \in S_X$ as follows. $\mathbf{m}_{x_2}$=$\{ \{x_1, x_2\} \}$, $\mathbf{m}_{x_3}$=$\{ \{x_1, x_3\} \}$, $\mathbf{m}_{x_4}$=$\{ \{x_4\} \}$. For each $A \in \mathbf{m}_u$, the quasi-linear ordered relation R_A can be constructed by the method in Theorem 3.4 as follows

$R_{\{x_1,x_2\}}$=$\{(x_1, x_1), (x_2, x_2), (x_3, x_3), (x_4, x_4), (x_3, x_4), (x_4, x_3), (x_1, x_3), (x_1, x_4), (x_2, x_3), (x_2, x_4), (x_1, x_2)\}$,

$R_{\{x_1,x_3\}}$=$\{(x_1, x_1), (x_2, x_2), (x_3, x_3), (x_4, x_4), (x_2, x_4), (x_4, x_2), (x_1, x_2), (x_1, x_4), (x_3, x_2), (x_3, x_4), (x_1,x_3)\}$,

$R_{\{x_4\}}$=$\{(x_1, x_1), (x_2, x_2), (x_3, x_3), (x_4, x_4), (x_1, x_2), (x_1, x_3), (x_2, x_1), (x_2, x_3), (x_3, x_1), (x_3, x_2), (x_4,x_1), (x_4, x_2), (x_4, x_3)\}$.

The Boolean matrices corresponding to these quasi-linear orders are

$$M_{R\{x_1,x_2\}} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, M_{R\{x_1,x_3\}} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}, M_{R\{x_4\}} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$$M_{R\{x_1,x_2\}} * M_{R\{x_1,x_3\}} * M_{R\{x_4\}} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = A_1$$

For the quasi-order R_2, we have S_X={ x_1, x_4}, $\mathbf{m}_{x_1}$={ {x_1 } }, $\mathbf{m}_{x_4}$={ {x_2, x_4}, {x_3, x_4} }. For each $A \in \mathbf{m}_u$, the quasi-linear ordered relation R_A can be constructed by the method in Theorem 3.4 as follows

$R_{\{x_2,x_4\}}$={(x_1, x_1), (x_2, x_2), (x_3, x_3), (x_4, x_4), (x_2, x_3), (x_4, x_3), (x_3, x_4), (x_2, x_1), (x_3, x_1), (x_4, x_1), (x_2, x_4)},
$R_{\{x_3,x_4\}}$={(x_1, x_1), (x_2, x_2), (x_3, x_3), (x_4, x_4), (x_3, x_2), (x_4, x_2), (x_2, x_4), (x_2, x_1), (x_3, x_1), (x_4, x_1), (x_3,x_4)},
$R_{\{x_1\}}$={(x_1, x_1), (x_2, x_2), (x_3, x_3), (x_4, x_4), (x_1, x_2), (x_1, x_3), (x_1, x_4), (x_2, x_3), (x_2, x_4), (x_3, x_2), (x_3,x_4), (x_4, x_2), (x_4, x_3)}.

Furthermore we have

$$M_{R\{x_2,x_4\}} * M_{R\{x_3,x_4\}} * M_{R\{x_1\}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} = A_2$$

For the quasi-order R_3, we have S_X={x_2, x_3, x_1}, $\mathbf{m}_{x_1}$={ {x_4, x_1 } }, $\mathbf{m}_{x_2}$={{x_2}}, $\mathbf{m}_{x_3}$={ {x_3}}. For each $A \in \mathbf{m}_u$, the quasi-linear ordered relation R_A can be constructed by the method in Theorem 3.4 as follows

$R_{\{x_2\}}$={(x_1, x_1), (x_2, x_2), (x_3, x_3), (x_4, x_4), (x_3, x_4), (x_3, x_1), (x_1, x_3), (x_1, x_4), (x_4, x_1), (x_4, x_3), (x_2, x_1), (x_2, x_3), (x_2, x_4)},
$R_{\{x_3\}}$={(x_1, x_1), (x_2, x_2), (x_3, x_3), (x_4, x_4), (x_2, x_4), (x_2, x_1), (x_1, x_2), (x_1, x_4), (x_3, x_2), (x_4, x_1), (x_4,x_2), (x_3, x_4), (x_3,x_1)},
$R_{\{x_1,x_4\}}$={(x_1, x_1), (x_2, x_2), (x_3, x_3), (x_4, x_4), (x_3, x_2), (x_2, x_3), (x_1, x_2), (x_1, x_3), (x_4, x_2), (x_4, x_3), (x_4,x_1) }.

Finally we have

$$M_{R\{x_2\}} * M_{R\{x_3\}} * M_{R\{x_4,x_1\}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} = A_3$$

We should point out that the above representation is not unique. The following is a different representation of the binary relation R of the "beautiful". Let

$$A_1 = \begin{bmatrix} 1&1&1&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1 \end{bmatrix}, A_2 = \begin{bmatrix} 1&0&0&0\\ 0&1&0&1\\ 0&0&1&0\\ 0&0&0&1 \end{bmatrix}, A_3 = \begin{bmatrix} 1&0&0&0\\ 0&1&0&0\\ 0&0&1&1\\ 0&0&0&1 \end{bmatrix}, A_4 = \begin{bmatrix} 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 1&0&0&1 \end{bmatrix}$$

We can verify the followings:

$$M_R = A_1 + A_2 + A_3 + A_4,\ A_i^2 = A_i,\ I + A_i = A_i,\ i\text{=}1,\ 2,\ 3,\ 4.$$

The binary relations on X, R_i, corresponding to A_i, $i = 1,2,3,4$ are quasi-orders on X. Similarly, we have the following quasi-linear order decompositions of A_1, A_2, A_3, A_4 via the method presented in Theorem 3.4.

$$A_1 = \begin{bmatrix} 1&1&1&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1 \end{bmatrix} = \begin{bmatrix} 1&1&1&1\\ 0&1&1&1\\ 0&0&1&1\\ 0&0&1&1 \end{bmatrix} * \begin{bmatrix} 1&1&1&1\\ 0&1&0&1\\ 0&1&1&1\\ 0&1&0&1 \end{bmatrix} * \begin{bmatrix} 1&1&1&0\\ 1&1&1&0\\ 1&1&1&0\\ 1&1&1&1 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 1&0&0&0\\ 0&1&0&1\\ 0&0&1&0\\ 1&0&0&1 \end{bmatrix} = \begin{bmatrix} 1&1&1&1\\ 0&1&1&1\\ 0&1&1&1\\ 0&1&1&1 \end{bmatrix} * \begin{bmatrix} 1&1&0&1\\ 1&1&0&1\\ 1&1&1&1\\ 1&1&0&1 \end{bmatrix} * \begin{bmatrix} 1&0&1&0\\ 1&1&1&1\\ 1&0&1&0\\ 1&0&1&1 \end{bmatrix}$$

$$A_3 = \begin{bmatrix} 1&0&0&0\\ 0&1&0&0\\ 0&0&1&1\\ 0&0&0&1 \end{bmatrix} = \begin{bmatrix} 1&1&1&1\\ 0&1&1&1\\ 0&1&1&1\\ 0&1&1&1 \end{bmatrix} * \begin{bmatrix} 1&0&1&1\\ 1&1&1&1\\ 1&0&1&1\\ 1&0&1&1 \end{bmatrix} * \begin{bmatrix} 1&1&0&0\\ 1&1&0&0\\ 1&1&1&1\\ 1&1&0&1 \end{bmatrix}$$

$$A_4 = \begin{bmatrix} 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 1&0&0&1 \end{bmatrix} = \begin{bmatrix} 1&0&1&1\\ 1&1&1&1\\ 1&0&1&1\\ 1&0&1&1 \end{bmatrix} * \begin{bmatrix} 1&1&0&1\\ 1&1&0&1\\ 1&1&1&1\\ 1&1&0&1 \end{bmatrix} * \begin{bmatrix} 1&1&1&0\\ 0&1&1&0\\ 0&1&1&0\\ 1&1&1&1 \end{bmatrix}$$

Let $M\text{=}M + I \in M(B)_{n\times n}$, we define

$$ri(M)\text{=}\inf\{l \mid M\text{=}A_1 + A_2 + \ldots + A_l,\ \text{each } A_i \text{ is an idempotent matrix } (i = 1,2,\ldots,l)\}.$$

There is an open problem: Is the shortest sum of the idempotent matrices of any Boolean matrix unique? The following theorem addresses the related problems.

Theorem 3.10. *Let $M, N \in M(B)_{n\times n}$. Then the following assertions hold*

(1) $ri(M) \leq |li(M)|$.
(2) $M = N$ if and only if $li(M) = li(N)$.
(3) $li(M) = li(N)$ if and only if $i(M) = i(N)$.

Proof. (1) By Theorem 3.8, we have that there exist idempotent elements A_1, $A_2, \ldots, A_r$ ($r = ri(M)$) such that

$$M{=}A_1 + A_2 + \ldots + A_r \Rightarrow A_i \leq M \Rightarrow A_i \in i(M).$$

Hence, there exist $B_j \in li(M)$ for every A_i such that $A_i \leq B_i$, i=1, 2, …, r. Suppose $li(M) = \{B_1, B_2, \ldots, B_l\}$, there exists a mapping $f : \{1,2,\ldots,r\} \to \{1,2,\ldots,l\}$ such that $A_i \leq B_{f(i)}$. Thus, we have

$$B_{f(1)} + B_{f(2)} + \ldots + B_{f(r)} \leq M = A_1 + A_2 + \ldots + A_r \leq B_{f(1)} + B_{f(2)} + \ldots + B_{f(r)}.$$

Therefore $M = B_{f(1)} + B_{f(2)} + \ldots + B_{f(r)}$. It follows that, f is injective by the definition of $ri(M)$.

(2) Let $li(M) = li(N)$. By (1), there exist $B_i \in li(M) = li(N)$ ($i = 1,2,\ldots,r$) such that

$$M = B_1 + B_2 + \ldots + B_r \leq N.$$

Similarly we have $N \leq M$.

(3) $i(M) = i(N) \Rightarrow li(M) = li(N) \Rightarrow M = N \Rightarrow i(M) = i(N)$. □

Exercises

Exercise 3.1. For any H, N, $M \in M(B)_{m\times n}$, show that the following assertions hold
$H * (N * M){=}(H * N) * M$;
$M * N = N * M$,
$H * (N + M) = H * N + H * M$.

Exercise 3.2. Show that($M(B)_{m\times n}, \vee, \wedge$) is a distributive lattice if $A \vee B{=}A + B$, $A \wedge B{=}A * B$ for any A, $B \in M(B)_{n\times n}$.

Exercise 3.3. Let A, $B \in M(B)_{n\times n}$, and P is a permutation matrix of order n. Prove the following

(a) If $A \leq B$ and $M \in M(B)_{n\times n}$, then $MA \leq MB$ and $AM \leq BM$.
(b) $P(A * B)P^T = (PAP^T) * (PBP^T)$.
(c) $p(PAP^T) = \{PM_RP^T \mid M_R \in p(A)\} \triangleq Pp(A)P^T$.
(d) $p(A * B){=}p(A) \cap p(B)$.
(e) $i(A + B) \supseteq i(A) \cup i(B)$.
(f) $i(PAP^T) = \{PMP^T \mid M \in i(A)\}$.

Exercise 3.4. Let$M \in M(B)_{n\times n}$ and $I+M$=M. Shown that there exist quasi-linear ordersR_{ij}, $i=1,2,\ldots,r, j=1,2,\ldots,q_i$, whose correspondence Boolean matrices shown as (3.3) and (3.4), such that

$$M=\sum_{i=1}^{r}(*\prod_{k=1}^{q_i}M_{R_{ik}}).$$

Open Problems

Problem 3.1. Is the shortest sum of the idempotent matrices of any Boolean matrix unique?

References

1. Dubois, D., Prade, H.: The Three Semantics of Fuzzy Sets. Fuzzy Sets and Systems 90, 141–150 (1997)
2. Jacobson, N.: Basic Algebra I. W H Freeman and Company, New York (1985)
3. Kim, K.H.: Boolean Matrix Theory and Applications. Marcel Dekker, New York (1982)
4. Liu, X.D.: The Structure of Fuzzy Matrices. International Journal of Fuzzy Mathematics 2, 311–325 (1994)
5. Liu, X.D., Chai, T.Y., Wang, W.: Approaches to the Representations and Logic Operations for Fuzzy Concepts in the Framework of Axiomatic Fuzzy Set Theory I. Information Sciences 177, 1007–1026 (2007)
6. Turksen, I.B.: Measurement of Membership Functions and Their Acquisition. Fuzzy Sets and Systems 40, 5–38 (1991)

Chapter 4
AFS Logic, AFS Structure and Coherence Membership Functions

In this chapter, we start with an introduction to EI algebra and AFS structure. Then the coherence membership functions of fuzzy concepts for AFS fuzzy logic for the AFS structure are proposed and a new framework of determining coherence membership functions is developed by taking both fuzziness (subjective imprecision) and randomness (objective uncertainty) into account. Singpurwalla's measure of the fuzzy events in a probability space has been applied to explore the proposed framework. Finally, the consistency, stability, efficiency and practicability of the proposed methodology are illustrated and studied via various numeric experiments. The investigations in this chapter open a door to explore the deep statistic properties of fuzzy sets. In this sense, they may offer further insights as to the to a role of natural languages in probability theory.

The aim of this chapter is to develop a practical and effective framework supporting the development of membership functions of fuzzy concepts based on semantics and statistics completed with regard to fuzzy data. We show that the investigations concur with the main results of the Singpurwalla's theory [44].

4.1 AFS Fuzzy Logic

The notion of a fuzzy set has been introduced in [51] in order to formalize the measurement of human concepts on numerical scales, in connection with the representation of human natural language and computing with words. Fuzzy sets and fuzzy logic are used for modeling imprecise modes of reasoning that play an essential role in the remarkable human ability to make rational decisions in an environment of uncertainty and imprecision ([55]).

As moving further into the age of machine intelligence and automated decision-making, we have to deal with both the subjective imprecision of human perception-based information described in natural language and the objective uncertainty of randomness universally existing in the real world. Zadeh has claimed that "probability must be used in concert with fuzzy logic to enhance its effectiveness. In this perspective, probability theory and fuzzy logic are complementary rather than competitive" [57]. In this chapter, we explore how the imprecision of natural language

X. Liu and W. Pedrycz: Axiomatic Fuzzy Set Theory and Its Applications, STUDFUZZ 244, pp. 111–166.
springerlink.com

and the randomness of observed data can be made to work in concert, so that uncertainty of randomness and of imprecision can be treated in a unified and coherent manner. Additionally, this treatment opens the door to enlarge the role of natural languages in probability theory.

In the real world applications, "conventional" membership functions are usually formed based on by the user's intuition. But these membership functions cannot be directly used in the fuzzy observation model because they do not offer us assurance to meet the requirement as the fuzzy event.

Recently, some new theories have been developed to interpret the membership functions. The authors in [5, 7, 6] proposed a coherent conditional probability which is looked on as a general non-additive "uncertainty" measure $m(.) = P(E|.)$ of the conditioning events. This gives rise to a clear, precise and rigorous mathematical frame, which allows to define fuzzy subsets and to introduce in a very natural way the counterparts of the basic continuous t-norms and the corresponding dual t-conorms, bound to the former by coherence. Some new approaches to De Finetti's coherence criterion [17, 41] provide more powerful results as to further exploration of these problems. So far, the applicable semantic aspects of fuzzy concepts and their fuzzy logic operations have not been fully discussed in the framework of conditional probability theory.

Singpurwalla and Booker developed a line of argument that demonstrates that probability theory has a sufficiently rich structure for incorporating fuzzy sets within its framework [44]. Thus probabilities of fuzzy events can be logically induced. The philosophical underpinnings that make this happen are a subjectivistic interpretation of probability, an introduction of Laplace's famous genie, and the mathematics of encoding expert testimony. Singpurwalla and Booker provide a real advance in our understanding of fuzzy sets, by identifying a sensible connection between membership functions and likelihood, and thereby probability. However, Singpurwalla and Booker focus on the interpretation of the probability measure of a fuzzy event with a predetermined membership function and have not discussed the problem of how to determine the membership functions of fuzzy concepts based on the theory they developed.

AFS theory [25, 23, 27, 30, 35] focuses on the study of determining the membership functions and logic operations of fuzzy concepts through AFS structures—a sort of mathematical descriptions of data structure and AFS algebras—a kind of completely distributivity lattices generated by a set and some simple (elementary) concepts defined on it. The AFS theory is based on the following essential observation: concepts are extracted from what an individual observed and the interpretation of a concept is strongly dependent on the observed "data" i.e., the world an individual observed or the background knowledge. Different people may give greatly different interpretations for the same concept due to their different observations. For instance, an NBA basketball player may describe that a person is not *"tall"* while a ten year old child may describe that the same person is very *"tall"*. Because the "data" the NBA basketball player observed, i.e., the people the NBA basketball player often meets are quite different from those the child often meets. In AFS theory—the studies on how to convert the information in the observed data into fuzzy sets (membership functions), the determination of membership functions of fuzzy concepts always emphasizes the data set they apply to. Considering that there are such complicated forms of the descriptions and representations for the attributes of the raw data in the real world applications, the raw data are "regularized" to be AFS structures by two axioms. AFS is mainly studied with respect to AFS structure of the data and AFS theory mainly studies fuzzy concepts, membership functions and fuzzy logic on the AFS structure of the data, instead of the raw data that are available through experiments.

An AFS structure is a triple (M, τ, X) which is a special family of combinatorical systems [11], where X is the universe of discourse, M is a set of some simple (or elementary) concepts on X (e.g., linguistic labels on the features such as "*large*","*medium*","*small*") and $\tau : X \times X \to 2^M$, which satisfies two axioms, is a mathematical description of the relationship of the distributions of the raw data and the semantics of the simple concepts in M. An AFS algebra is a family of completely distributive lattices generated by the sets such as X and M. Using the AFS algebras and the AFS structures, a great number of complex fuzzy concepts on X and their logic operations can be expressed by a few simple concepts in M. Liu, Pedrycz and Zhang gave the complement operation of the fuzzy concepts in EI algebra EM–a sort of the AFS algebra, thus a fuzzy logic system called AFS fuzzy logic has been developed [27].

4.1.1 EI Algebra

In [23], defined was a family of completely distributivity lattices, referred to as AFS algebras, and denoted as EI, EII, ..., EI^n, $E^{\#}I$, $E^{\#}II$, ..., $E^{\#}I^n$ algebras. The authors applied them to study the semantics of expressions and the representations of fuzzy concepts. The following example serves as an introductory illustration of the EI algebra.

Example 4.1. Let $X = \{x_1, x_2, ..., x_{10}\}$ be a set of 10 people and their features (attributes) which are described by real numbers (age, height, weight, salary, estate), Boolean values (gender) and the ordered relations (hair black, hair white, hair yellow), see Table 4.1; here the number i in the "*hair color*" columns which corresponds to some $x \in X$ implies that the hair color of x has ordered ith following our perception of the color by our intuitive perception. For example, the numbers in the column *"hair black"* implies some order (denoted here by $>$)

$$x_7 > x_{10} > x_4 = x_8 > x_2 = x_9 > x_5 > x_6 = x_3 = x_1$$

When moving from right to left, the relationship states how strongly the hair color resembles black color. In this order, $x_i > x_j$ (e.g., $x_7 > x_{10}$) states that the hair of x_i is closer to the black color than the color of hair the individual x_j. The relationship $x_i = x_j$ (e.g., $x_4 = x_8$) means that the hair of x_i looks as black as the one of x_j. A concept on X may associate to one or more features. For instance, the fuzzy concept "*tall*" associates a single feature "*height*" and the fuzzy concept "*old white hair males*" associates three features "*age*", "*hair color_black*" and "*gender_male*".

Many concepts may associate with a single feature. For instance, the fuzzy concepts "*old*", "*young*" and "*about 40 years old*" all associate to feature "*age*". Let $M = \{m_1, m_2, ..., m_{12}\}$ be the set of fuzzy or Boolean concepts on X and each $m \in M$ associate to a single feature. The following terms are used here m_1 : *"old people"*, m_2 : *"tall people"*, m_3 : *"heavy people"*, m_4 : *"high salary"*, m_5 : *"more estate"*, m_6 :*"male"*, m_7 : *"female"*, m_8 : *"black hair people"*, m_9 : *"white hair people"*, m_{10} : *"yellow hair people"*, m_{11} : "*young people*", m_{12} : "*the people about 40 years old*". The elements of M are viewed as "*elementary*" (or "*simple*") concepts.

Table 4.1 Descriptions of features

		appearance		wealth		gender		hair color		
	age	*height*	*weigh*	*salary*	*estate*	*male*	*female*	*black*	*white*	*yellow*
x_1	20	1.9	90	1	0	1	0	6	1	4
x_2	13	1.2	32	0	0	0	1	4	3	1
x_3	50	1.7	67	140	34	0	1	6	1	4
x_4	80	1.8	73	20	80	1	0	3	4	2
x_5	34	1.4	54	15	2	1	0	5	2	2
x_6	37	1.6	80	80	28	0	1	6	1	4
x_7	45	1.7	78	268	90	1	0	1	6	4
x_8	70	1.65	70	30	45	1	0	3	4	2
x_9	60	1.82	83	25	98	0	1	4	3	1
x_{10}	3	1.1	21	0	0	0	1	2	5	3

For each set of concepts, $A \subseteq M$, $\prod_{m\in A} m$ represents a conjunction of the concepts in A. For instance, $A = \{m_1, m_6\} \subseteq M$, $\prod_{m\in A} m = m_1m_6$ representing a new fuzzy concept "*old males*" which associates to the features of age and gender. $\sum_{i\in I}(\prod_{m\in A_i} m)$, which is a formal sum of the concepts $\prod_{m\in A_i} m, A_i \subseteq M, i \in I$, is the disjunction of the conjunctions represented by $\prod_{m\in A} m$'s (i.e., the disjunctive normal form of a formula representing a concept). For example, we may have $\gamma = m_1m_6 + m_1m_3 + m_2$ which translates as "*old males*" or "*heavy old people*" or "*tall people*". (The "+" denotes here a disjunction of concepts). While M may be a set of fuzzy or Boolean (two-valued) concepts, every $\sum_{i\in I}(\prod_{m\in A_i} m), A_i \subseteq M, i \in I$, has a well-defined meaning such as the one we have discussed above. By a straightforward comparison of

$$m_3m_8 + m_1m_4 + m_1m_6m_7 + m_1m_4m_8$$

and

$$m_3m_8 + m_1m_4 + m_1m_6m_7,$$

we conclude that the expressions are equivalent. Considering the terms on left side of the expressions, for any x, the degree of x belonging to the fuzzy concept represented by $m_1m_4m_8$ is always less than or equal to the degree of x belonging to the fuzzy concept represented by m_1m_4. Therefore, the term $m_1m_4m_8$ is redundant when forming the left side of the fuzzy concept and the expressions are equivalent in semantics. In practice, when we form complex concepts using some simple concepts like what we have discussed above, we always accept the following axioms of natural language:

1) The repeat of a concept can be reduced in the product $\prod$ (e.g., $m_1m_4m_8$ is equivalent to $m_1m_4m_1m_8$);
2) The sum $\sum$ is commutative (e.g., $m_5m_6+m_5m_8$ is equivalent to $m_5m_8+m_5m_6$) and the product is also commutative (e.g., m_5m_8 is equivalent to m_8m_5);
3) The product distributes over the sum (e.g.,$m_5m_6+m_5m_8$ is equivalent to $m_5(m_6+m_8)$).

Let us take into consideration two expressions of the form $\alpha : m_1m_4+m_2m_5m_6$, and $\nu : m_5m_6+m_5m_8$. Under the above axioms, the semantic contents of the fuzzy concepts "α *or* ν" and "α *and* ν" can be expressed as follows
"α *or* ν": $m_1m_4+m_2m_5m_6+m_5m_6+m_5m_8$ equivalent to

$$m_1m_4+m_5m_6+m_5m_8,$$

"α *and* ν": $m_1m_4m_5m_6+m_2m_5m_6+m_1m_4m_5m_8+m_2m_5m_6m_8$ equivalent to

$$m_1m_4m_5m_6+m_2m_5m_6+m_1m_4m_5m_8.$$

The semantics of the logic expressions such as "*equivalent to*", "*or*" and "*and*" as expressed by $\sum_{i\in I}(\prod_{m\in A_i}m), A_i \subseteq M, i\in I$ can be formulated in terms of the EI algebra in the following manner.

Let M be a non-empty set. The set EM^* is defined by

$$EM^* = \left\{ \sum_{i\in I}(\prod_{m\in A_i} m) \mid A_i \subseteq M, i\in I, I \text{ is any no empty indexing set} \right\}. \quad (4.1)$$

Definition 4.1. ([23]) Let M be a non-empty set. A binary relation R on EM^* is defined as follows: for $\sum_{i\in I}(\prod_{m\in A_i}m)$, $\sum_{j\in J}(\prod_{m\in B_j}m)\in EM^*$,

$\left[\sum_{i\in I}(\prod_{m\in A_i}m)\right] R \left[\sum_{j\in J}(\prod_{m\in B_j}m)\right] \iff$ (i) $\forall A_i$ $(i\in I)$, $\exists B_h$ $(h\in J)$ such that $A_i\supseteq B_h$; (ii) $\forall B_j$ $(j\in J)$, $\exists$ A_k $(k\in I)$, such that $B_j\supseteq A_k$.

It is clear that R is an equivalence relation. The *quotient set,* EM^*/R is denoted as EM. Notice that any element of EM is an equivalence class. Let $[\sum_{i\in I}(\prod_{m\in A_i}m)]_R$ be the set of all elements which are equivalent to $\sum_{i\in I}(\prod_{m\in A_i}m)\in EM^*$, and $[\sum_{i\in I}(\prod_{m\in A_i}m)]_R\in EM$. In general, for any $\xi,\zeta\in EM^*$, ξ,ζ are equivalent under R means $\xi\in[\zeta]_R$, $\zeta\in[\xi]_R$, and $[\zeta]_R=[\xi]_R$. If $\sum_{i\in I}(\prod_{m\in A_i}m)$ is not specified in EM^*, then the equivalent class $[\sum_{i\in I}(\prod_{m\in A_i}m)]_R$ is always denoted as $\sum_{i\in I}(\prod_{m\in A_i}m)$ for the sake of simplicity and the notation $\sum_{i\in I}(\prod_{m\in A_i}m)=\sum_{j\in J}(\prod_{m\in B_j}m)$ means that $[\sum_{i\in I}(\prod_{m\in A_i}m)]_R=[\sum_{j\in J}(\prod_{m\in B_j}m)]_R$. Thus the semantics they represent are equivalent. In Example 4.1, for $\xi = m_3m_8+m_1m_4+m_1m_6m_7+m_1m_4m_8$, $\zeta =$

$m_3m_8+m_1m_4+m_1m_6m_7 \in EM$, by Definition 4.1 we have $\xi=\zeta$. In what follows, each $\sum_{i\in I}(\prod_{m\in A_i} m) \in EM$ is called a fuzzy concept.

Proposition 4.1. *Let M be a non-empty set. If $\sum_{i\in I}(\prod_{m\in A_i} m) \in EM$, $A_t \subseteq A_s$, $t,s\in I$, $t\neq s$, then*

$$\sum_{i\in I}(\prod_{m\in A_i} m)=\sum_{i\in I-\{s\}}(\prod_{m\in A_i} m).$$

Its proof is left as an exercise.

Theorem 4.1. *([23]) Let M be a non-empty sets. Then $(EM,\vee,\wedge)$ forms a completely distributive lattice under the binary compositions $\vee$ and $\wedge$ defined as follows: for any $\sum_{i\in I}(\prod_{m\in A_i} m)$, $\sum_{j\in J}(\prod_{m\in B_j} m) \in EM$,*

$$\sum_{i\in I}(\prod_{m\in A_i} m)\vee\sum_{j\in J}(\prod_{m\in B_j} m)=\sum_{k\in I\sqcup J}(\prod_{m\in C_k} m)\triangleq\sum_{i\in I}(\prod_{m\in A_i} m)+\sum_{j\in J}(\prod_{m\in B_j} m), \quad (4.2)$$

$$\sum_{i\in I}(\prod_{m\in A_i} m)\wedge\sum_{j\in J}(\prod_{m\in B_j} m)=\sum_{i\in I, j\in J}(\prod_{m\in A_i\cup B_j} m), \quad (4.3)$$

where for any $k\in I\sqcup J$ (the disjoint union of I and J, i.e., every element in I and every element in J are always regarded as different elements in $I\sqcup J$), $C_k=A_k$ if $k\in I$, and $C_k=B_k$ if $k\in J$.

Remark 4.1. $(EM,\vee,\wedge)$ is a completely distributive lattice meaning that the following important properties hold.

- For $\sum_{i\in I}(\prod_{m\in A_i} m), \sum_{j\in J}(\prod_{m\in B_j} m) \in EM$,

$$\sum_{i\in I}(\prod_{m\in A_i} m)\geq\sum_{j\in J}(\prod_{m\in B_j} m)\Leftrightarrow \forall B_j,(j\in J),\exists A_k,(k\in I) \text{ such that } B_j\supseteq A_k. \quad (4.4)$$

- Let I be any indexing set and $\sum_{j\in I_i}(\prod_{m\in A_{ij}} m)\in EM, i\in I$. Then

$$\bigvee_{i\in I}(\sum_{j\in I_i}(\prod_{m\in A_{ij}} m))=\sum_{i\in I}\sum_{j\in I_i}(\prod_{m\in A_{ij}} m), \quad (4.5)$$

$$\bigwedge_{i\in I}(\sum_{j\in I_i}(\prod_{m\in A_{ij}} m))=\sum_{f\in\prod_{i\in I} I_i}(\prod_{m\in\cup_{i\in I}A_{if(i)}} m). \quad (4.6)$$

- Let I, J_i be non-empty indexing sets, $i\in I$. For any $\lambda_{ij}\in EM, i\in I, j\in J_i$, the following CD1 and CD2 hold.

$$\bigwedge_{i\in I}(\bigvee_{j\in J_i}\lambda_{ij})=\bigvee_{f\in\prod_{i\in I}J_i}(\bigwedge_{i\in I}\lambda_{if(i)}), \quad (4.7)$$

$$\bigvee_{i\in I}(\bigwedge_{j\in J_i}\lambda_{ij})=\bigwedge_{f\in\prod_{i\in I}J_i}(\bigvee_{i\in I}\lambda_{if(i)}). \quad (4.8)$$

Proof. First, we prove that $\vee$, $\wedge$ are binary compositions. Let $\sum_{i\in I_1}(\prod_{m\in A_{1i}} m) = \sum_{i\in I_2}(\prod_{m\in A_{2i}} m)$, $\sum_{j\in J_1}(\prod_{m\in B_{1j}} m) = \sum_{j\in J_2}(\prod_{m\in B_{2j}} m) \in EM$. (4.2) can be directly verified based on Definition 4.1. From (4.3),we have

$$\sum_{i\in I_1}(\prod_{m\in A_{1i}} m) \wedge \sum_{j\in J_1}(\prod_{m\in B_{1j}} m) = \sum_{i\in I_1, j\in J_1}(\prod_{m\in A_{1i}\cup B_{1j}} m),$$
$$\sum_{i\in I_2}(\prod_{m\in A_{2i}} m) \wedge \sum_{j\in J_2}(\prod_{m\in B_{2j}} m) = \sum_{i\in I_2, j\in J_2}(\prod_{m\in A_{2i}\cup B_{2j}} m).$$

Since $\sum_{i\in I_1}(\prod_{m\in A_{1i}} m) = \sum_{i\in I_2}(\prod_{m\in A_{2i}} m)$, $\sum_{j\in J_1}(\prod_{m\in B_{1j}} m) = \sum_{j\in J_2}(\prod_{m\in B_{2j}} m)$, hence by Definition 4.1, for any $A_{1i}\cup B_{1j}, i\in I_1, j\in J_1$, there exist A_{2k}, $B_{2l}, k\in I_2, l\in J_2$ such that $A_{1i}\supseteq A_{2k}, B_{1j}\supseteq B_{2l}$. Thus $A_{1i}\cup B_{1j}\supseteq A_{2k}\cup B_{2l}$. Similarly, for any $A_{2i}\cup B_{2j}, i\in I_2, j\in J_2$, there exist A_{1q}, $B_{1e}, q\in I_1, e\in J_1$, such that $A_{2i}\cup B_{2j}\supseteq A_{1q}\cup B_{1e}$. This implies that

$$\sum_{i\in I_1, j\in J_1}(\prod_{m\in A_{1i}\cup B_{1j}} m) = \sum_{i\in I_2, j\in J_2}(\prod_{m\in A_{2i}\cup B_{2j}} m)$$

and $\wedge$ is a binary composition. Theorem 2.4 states that two binary compositions satisfying the condition L1-L4 of Theorem 2.3 are lattice operations. For any $\sum_{i\in I}(\prod_{m\in A_i} m), \sum_{j\in J}(\prod_{m\in B_j} m), \sum_{u\in U}(\prod_{m\in C_u} m) \in EM$, we can directly verify that $\vee$, $\wedge$ satisfy L1-L3 of Theorem 2.3 by the definitions (which remains as an exercise).

In the following, we prove that $\vee$, $\wedge$ satisfy L4 of Theorem 2.3. By Proposition 4.1, we have

$(\sum_{i\in I}(\prod_{m\in A_i} m) \vee \sum_{j\in J}(\prod_{m\in B_j} m)) \wedge \sum_{i\in I}(\prod_{m\in A_i} m)$
$= \sum_{i,j\in I}(\prod_{m\in A_i\cup A_j} m) + \sum_{i\in I, j\in J}(\prod_{m\in A_i\cup B_j} m)$
$= \sum_{i\in I}(\prod_{m\in A_i} m) + \sum_{i\in I, j\in J}(\prod_{m\in A_i\cup B_j} m) = \sum_{i\in I}(\prod_{m\in A_i} m)$.
$(\sum_{i\in I}(\prod_{m\in A_i} m) \wedge \sum_{j\in J}(\prod_{m\in B_j} m)) \vee \sum_{i\in I}(\prod_{m\in A_i} m)$
$= \sum_{i\in I, j\in J}(\prod_{m\in A_i\cup B_j} m) + \sum_{i\in I}(\prod_{m\in A_i} m) = \sum_{i\in I}(\prod_{m\in A_i} m)$.
Therefore $\vee$, $\wedge$ satisfy L4 of Theorem 2.3 and $(EM, \vee, \wedge)$ is a lattice.

$$\sum_{i\in I}(\prod_{m\in A_i} m) \geq \sum_{j\in J}(\prod_{m\in B_j} m) \Leftrightarrow \sum_{i\in I}(\prod_{m\in A_i} m) \vee \sum_{j\in J}(\prod_{m\in B_j} m) = \sum_{i\in I}(\prod_{m\in A_i} m), \tag{4.9}$$

if and only if $\forall B_j, (j\in J)$, $\exists A_k, (k\in I)$ such that $B_j\supseteq A_k$.

In the following, we prove that $(EM, \vee, \wedge)$ is a complete lattice. Let $\sum_{j\in I_i}(\prod_{m\in A_{ij}} m) \in EM, i\in I$. We prove that $\bigvee_{i\in I}(\sum_{j\in I_i}(\prod_{m\in A_{ij}} m))$, $\bigwedge_{i\in I}(\sum_{j\in I_i}(\prod_{m\in A_{ij}} m)) \in EM$. It is obvious that the following relationships are satisfied

$$\sum_{j\in I_i}(\prod_{m\in A_{ij}} m) \leq \sum_{i\in I}\sum_{j\in I_i}(\prod_{m\in A_{ij}} m), \ \ \forall i\in I,$$
$$\sum_{j\in I_i}(\prod_{m\in A_{ij}} m) \geq \sum_{f\in \prod_{i\in I} I_i}(\prod_{m\in \cup_{i\in I} A_{if(i)}} m), \ \ \forall i\in I.$$

For $\sum_{u\in U}(\prod_{m\in B_u} m) \in EM$, if $\sum_{j\in I_i}(\prod_{m\in A_{ij}} m) \leq \sum_{u\in U}(\prod_{m\in B_u} m), \forall i \in I$, then $\forall A_{i_0 j_0}$, $i_0 \in I, j_0 \in J_{i_0}$, there exists $u_0 \in U$ such that $A_{i_0 j_0} \supseteq B_{u_0}$. Therefore by (4.9), we have $\sum_{i\in I}\sum_{j\in I_i}(\prod_{m\in A_{ij}} m) \leq \sum_{u\in U}(\prod_{m\in B_u} m)$. This implies that $\sum_{i\in I}\sum_{j\in I_i}(\prod_{m\in A_{ij}} m) \in EM$ is the least upper bound or supremum of the set $\{\sum_{j\in I_i}(\prod_{m\in A_{ij}} m) \in EM \mid i \in I\}$, i.e.,

$$\bigvee_{i\in I}(\sum_{j\in I_i}(\prod_{m\in A_{ij}} m)) = \sum_{i\in I}\sum_{j\in I_i}(\prod_{m\in A_{ij}} m). \tag{4.10}$$

For $\sum_{u\in U}(\prod_{m\in B_u} m) \in EM$, if $\sum_{j\in I_i}(\prod_{m\in A_{ij}} m) \geq \sum_{u\in U}(\prod_{m\in B_u} m), \forall i \in I$, then $\forall B_{u_0}$, $u_0 \in U$ and $\forall i_0 \in I$, there exists $j_{i_0} \in I_{i_0}$ such that $B_{u_0} \supseteq A_{i_0 j_{i_0}}$. This implies that for any $u_0 \in U$ there exists $f_{u_0} \in \prod_{i\in I} I_i$ such that $f_{u_0}(i_0) = j_{i_0}, \forall i_0 \in I$ and $B_{u_0} \supseteq \cup_{i\in I} A_{i f_{u_0}(i)}$. Therefore by (4.9), we have

$$\sum_{f\in\prod_{i\in I} I_i}(\prod_{m\in\cup_{i\in I} A_{if(i)}} m) \geq \sum_{u\in U}(\prod_{m\in B_u} m).$$

Thus $\sum_{f\in\prod_{i\in I} I_i}(\prod_{m\in\cup_{i\in I} A_{if(i)}} m) \in EM$ is the greatest lower bound or infimum of the set $\{\sum_{j\in I_i}(\prod_{m\in A_{ij}} m) \in EM \mid i \in I\}$, i.e.,

$$\bigwedge_{i\in I}(\sum_{j\in I_i}(\prod_{m\in A_{ij}} m)) = \sum_{f\in\prod_{i\in I} I_i}(\prod_{m\in\cup_{i\in I} A_{if(i)}} m). \tag{4.11}$$

By Definition 2.5, we know that $(EM, \vee, \wedge)$ is a complete lattice.

For any $\gamma, \zeta, \eta, \in EM$, the following properties (D1 and D2) can be directly verified by Definition 4.1; the proof remains as an exercise.

D1: $\gamma \wedge (\zeta \vee \eta) = (\gamma \wedge \zeta) \vee (\gamma \wedge \eta)$;
D2: $\gamma \vee (\zeta \wedge \eta) = (\gamma \vee \zeta) \wedge (\gamma \vee \eta)$.

Therefore by Definition 2.5, we know that $(EM, \vee, \wedge)$ is also a distributive lattice.

In the following, we prove that $(EM, \vee, \wedge)$ is a completely distributive lattice, i.e., satisfying CD1 and CD2 in Definition 2.17. By Corollary 2.4 we know that a complete lattice L is a completely distributive lattice if and only if one of CD1 and CD2 holds. Hence we just prove the lattice $(EM, \vee, \wedge)$ satisfies that CD1.

Let $\lambda_{ij} = \sum_{u\in U_{ij}}(\prod_{m\in A_u^{ij}} m) \in EM$, $i \in I$, $j \in J_i$, U_{ij} is a non-empty indexing set. For any $f \in \prod_{i\in I} J_i$,we know that $\forall k \in I$, since $f(k) \in J_k$,. Hence for any $f \in \prod_{i\in I} J_i$ and any $k \in I$ we have

$$\bigwedge_{i\in I}\lambda_{if(i)} \leq \lambda_{kf(k)} \leq \bigvee_{j\in J_k}\lambda_{kj}.$$

Since $\forall k \in I, \forall f \in \prod_{i\in I} J_i, \bigwedge_{i\in I}\lambda_{if(i)} \leq \bigvee_{j\in J_k}\lambda_{kj}$, hence for any $f \in \prod_{i\in I} J_i$,

$$\bigwedge_{i\in I}\lambda_{if(i)} \leq \bigwedge_{k\in I}(\bigvee_{j\in J_k}\lambda_{kj}).$$

Furthermore

$$\bigvee_{f\in\prod_{i\in I}J_i}(\bigwedge_{i\in I}\lambda_{if(i)}) \le \bigwedge_{i\in I}(\bigvee_{j\in J_i}\lambda_{ij}). \tag{4.12}$$

By (4.10) and (4.11), we have

$$\begin{aligned}\bigvee_{f\in\prod_{i\in I}J_i}(\bigwedge_{i\in I}\lambda_{if(i)}) &= \bigvee_{f\in\prod_{i\in I}J_i}(\bigwedge_{i\in I}(\sum_{u\in U_{if(i)}}(\prod_{m\in A_u^{if(i)}}m))\\ &= \sum_{f\in\prod_{i\in I}J_i}(\sum_{h\in\prod_{i\in I}U_{if(i)}}(\prod_{m\in\cup_{i\in I}A_{h(i)}^{if(i)}}m)).\end{aligned} \tag{4.13}$$

$$\begin{aligned}\bigwedge_{i\in I}(\bigvee_{j\in J_i}\lambda_{ij}) &= \bigwedge_{i\in I}(\sum_{j\in J_i}\sum_{u\in U_{ij}}(\prod_{m\in A_u^{ij}}m))\\ &= \bigwedge_{i\in I}(\sum_{u\in\bigsqcup_{k\in J_i}U_{ik}}(\prod_{m\in E_u^i}m))\\ &= \sum_{g\in\prod_{i\in I}(\bigsqcup_{k\in J_i}U_{ik})}(\prod_{m\in\cup_{i\in I}E_{g(i)}^i}m),\end{aligned} \tag{4.14}$$

where for any $u\in\bigsqcup_{k\in J_i}U_{ik}$, $E_u^i=A_u^{ij}$ if $u\in U_{ij}$. For any $g_0\in\prod_{i\in I}(\bigsqcup_{k\in J_i}U_{ik})$, since $g_0(i)\in\bigsqcup_{k\in J_i}U_{ik}, i\in I$, hence for any $i\in I$, there exists $k_i\in J_i$ such that $g_0(i)\in U_{ik_i}$. This implies that if we define $f_0(i)=k_i\in J_i$, $i\in I$, then $f_0\in\prod_{i\in I}J_i$, $g_0(i)\in U_{if_0(i)}$, $g_0\in\prod_{i\in I}U_{if_0(i)}$ and $E_{g_0(i)}^i=A_{g_0(i)}^{if_0(i)}$, for any $i\in I$. Thus, considering the right sides of (4.13) and (4.14), for any $g_0\in\prod_{i\in I}(\bigsqcup_{k\in J_i}U_{ik})$, there exist $f_0\in\prod_{i\in I}J_i$ such that $g_0\in\prod_{i\in I}U_{if_0(i)}$ and

$$\bigcap_{i\in I}E_{g_0(i)}^i=\bigcap_{i\in I}A_{g_0(i)}^{if_0(i)}.$$

By (4.9), we have

$$\bigvee_{f\in\prod_{i\in I}J_i}(\bigwedge_{i\in I}\lambda_{if(i)}) \ge \bigwedge_{i\in I}(\bigvee_{j\in J_i}\lambda_{ij}). \tag{4.15}$$

Therefore the following CD1 hold from (4.12)

$$\bigwedge_{i\in I}(\bigvee_{j\in J_i}\lambda_{ij}) = \bigvee_{f\in\prod_{i\in I}J_i}(\bigwedge_{i\in I}\lambda_{if(i)}). \tag{4.16}$$

By Corollary 2.4 we know that

$$\bigvee_{i\in I}(\bigwedge_{j\in J_i}\lambda_{ij}) = \bigwedge_{f\in\prod_{i\in I}J_i}(\bigvee_{i\in I}\lambda_{if(i)}). \tag{4.17}$$

and $(EM,\vee,\wedge)$ is a completely distributive lattice. □

In Example 4.1, for $\psi = m_1m_4 + m_2m_5m_6, \nu = m_5m_6 + m_5m_8 \in EM$, in virtue (4.2) and (4.3) we know that the semantic content of the fuzzy concepts " ψ *or* ν" and "ψ *and* ν" are expressed as "$\psi \vee \nu$" and "$\psi \wedge \nu$", respectively. The algebra operations carried out on them come in the form:

$$\begin{aligned}
\psi \vee \nu &= m_1m_4 + m_2m_5m_6 + m_5m_6 + m_5m_8 \\
&= m_5m_6 + m_5m_8 + m_1m_4, \\
\psi \wedge \nu &= m_1m_4m_5m_6 + m_2m_5m_6 + m_1m_4m_5m_8 + m_2m_5m_6m_8 \\
&= m_1m_4m_5m_6 + m_2m_5m_6 + m_1m_4m_5m_8.
\end{aligned}$$

$(EM, \vee, \wedge)$ is called the *EI (expanding one set M) algebra* over M—one type of AFS algebra. For $\psi = \sum_{i \in I}(\prod_{m \in A_i} m)$, $\vartheta = \sum_{j \in J}(\prod_{m \in B_j} m) \in EM$, $\psi \le \vartheta \Longleftrightarrow \psi \vee \vartheta = \vartheta \Leftrightarrow \forall A_i \ (i \in I), \exists B_h \ (h \in J)$ such that $A_i \supseteq B_h$. In light of the interpretation of concepts $\prod_{m \in A_i} m$ and $\prod_{m \in B_h} m$, we know that for any $x \in X$, the membership degree of x belonging to $\prod_{m \in A_i} m$ is always less than or equal to that of $\prod_{m \in B_h} m$ considering $A_i \supseteq B_h$, i.e., the stricter the constraint of a concept, the lower degree of x belongs to the concept. Therefore the membership degree of x belonging to concept ψ is always less than or equal to that of ϑ for all $x \in X$ due to $\psi \le \vartheta$. For instance, in Example 4.1, $\psi = m_1m_4 + m_2m_5m_6$ and $\vartheta = m_5m_6 + m_5m_8 + m_1m_4$. By (4.4), we have $\psi \vee \vartheta = \vartheta$, i.e., $\psi \le \vartheta$. In the sense of the underlying semantics we have

ψ states " *old high salary people*" or "*tall male with more estate*",
ϑ reads "*old high salary people*" or "*male with more estate*" or "*black hair people with more estate*",

since the membership degree of x belonging to the concept "*tall male with more estate*" is always less than or equal to that of "*male with more estate*" for all $x \in X$, hence the membership degree of any x belonging to ψ is less than or equal to that of ϑ.

Theorem 4.2. *([27]) Let M be a set and $g : M \to M$ be a map satisfying $g(g(m)) = m$ for all $m \in M$. If the operator $^g : EM \to EM$ is defined as follows*

$$\left(\sum_{i \in I}(\prod_{m \in A_i} m)\right)^g = \bigwedge_{i \in I}\left(\bigvee_{m \in A_i} g(m)\right) = \bigwedge_{i \in I}\left(\sum_{m \in A_i} g(m)\right). \tag{4.18}$$

Then for any $\alpha, \beta \in EM$, "g" has the following properties:

1) $(\alpha^g)^g = \alpha$;
2) $(\alpha \vee \beta)^g = \alpha^g \wedge \beta^g$, $(\alpha \wedge \beta)^g = \alpha^g \vee \beta^g$;
3) $\alpha \le \beta \Rightarrow \alpha^g \ge \beta^g$.

i.e. the operator "^{g}" is an order reversing involution *(or* conversely ordered involutory mapping*) in the EI algebra* $(EM, \vee, \wedge)$.

Proof. First we prove that the operator "^{g}" is a map from EM to EM. Let $\zeta = \sum_{i\in I}(\prod_{m\in A_i} m), \eta = \sum_{j\in J}(\prod_{m\in B_j} m) \in EM$ and $\zeta = \eta$. By $\zeta = \eta$, we know that for any $A_i, i \in I$ there exists $B_k, k \in J$ such that $A_i \supseteq B_k$. This implies that for any $i \in I$,

$$\bigvee_{m\in A_i} g(m) = \sum_{m\in A_i} g(m) \geq \sum_{m\in B_k} g(m) = \bigvee_{m\in B_k} g(m) \geq \bigwedge_{j\in J}(\bigvee_{m\in B_j} g(m)) = \eta^g.$$

Furthermore we have $\zeta^g = \bigwedge_{i\in I}\bigvee_{m\in A_i} g(m) \geq \eta^g$. Similarly we can prove $\zeta^g \leq \eta^g$. Thus $\zeta^g = \eta^g$ and the operator "^{g}" is a map.

1) For any $\sum_{i\in I}(\prod_{m\in A_i} m) \in EM$, because $(EM, \vee, \wedge)$ is a completely distributive lattice, i.e., it satisfies (4.7) and (4.8), we have

$$\begin{aligned}
((\sum_{i\in I}(\prod_{m\in A_i} m))^g)^g &= (\bigwedge_{i\in I}(\bigvee_{m\in A_i} g(m)))^g \\
&= (\bigvee_{f\in\prod_{i\in I} A_i} (\bigwedge_{i\in I} g(f(i))))^g \\
&= \bigwedge_{f\in\prod_{i\in I} A_i} (\bigvee_{i\in I} g(g(f(i))) \\
&= \bigwedge_{f\in\prod_{i\in I} A_i} (\bigvee_{i\in I} f(i)) \\
&= \bigvee_{i\in I}(\bigwedge_{m\in A_i} m) = \sum_{i\in I}(\prod_{m\in A_i} m).
\end{aligned}$$

2) For any $\sum_{i\in I}(\prod_{m\in A_i} m), \sum_{j\in J}(\prod_{m\in B_j} m) \in EM$, we have

$$\begin{aligned}
(\sum_{i\in I}(\prod_{m\in A_i} m)\bigvee\sum_{j\in J}(\prod_{m\in B_j} m))^g &= \left(\bigwedge_{i\in I}(\bigvee_{m\in A_i} g(m))\right)\bigwedge\left(\bigwedge_{j\in J}(\bigvee_{m\in B_j} g(m))\right) \\
&= (\sum_{i\in I}(\prod_{m\in A_i} m))^g\bigwedge(\sum_{j\in J}(\prod_{m\in B_j} m))^g.
\end{aligned}$$

In the a completely distributive lattice $(EM, \vee, \wedge)$, one can verify that $\bigwedge_{i\in I}(\alpha \vee \beta_i) = \alpha \vee (\bigwedge_{i\in I} \beta_i)$ for any $\alpha, \beta_i \in EM, i \in I$. Its proof is left to the reader. Thus we have

$$\begin{aligned}
(\sum_{i\in I}(\prod_{m\in A_i} m)\bigwedge\sum_{j\in J}(\prod_{m\in B_j} m))^g &= (\sum_{i\in I, j\in J}(\prod_{m\in A_i\cup B_j} m))^g \\
&= \bigwedge_{i\in I, j\in J}(\bigvee_{m\in A_i\cup B_j} g(m)) \\
&= \bigwedge_{i\in I, j\in J}\left((\bigvee_{m\in A_i} g(m))\bigvee(\bigvee_{m\in B_j} g(m))\right) \\
&= \bigwedge_{i\in I}\left(\bigwedge_{j\in J}\left((\bigvee_{m\in A_i} g(m))\bigvee(\bigvee_{m\in B_j} g(m))\right)\right)
\end{aligned}$$

$$= \bigwedge_{i\in I}\left((\bigvee_{m\in A_i} g(m))\bigvee\left(\bigwedge_{j\in J}(\bigvee_{m\in B_j} g(m))\right)\right)$$
$$= \left(\bigwedge_{i\in I}(\bigvee_{m\in A_i} g(m))\right)\bigvee\left(\bigwedge_{j\in J}(\bigvee_{m\in B_j} g(m))\right)$$
$$= (\sum_{i\in I}(\prod_{m\in A_i} m))^g \bigvee (\sum_{j\in J}(\prod_{m\in B_j} m))^g.$$

3) For $\alpha,\beta \in EM$, $\alpha \le \beta \Leftrightarrow \alpha\wedge\beta = \alpha$. From 2) we have $\alpha^g = (\alpha\wedge\beta)^g = \alpha^g \vee \beta^g$. Thus $\alpha^g \ge \beta^g$. Therefore the operator "^{g}" is an order reversing involution in the EI algebra $(EM,\vee,\wedge)$. □

If m' stands for the negation of the concept $m\in M$ and $m''=m$, then for any $\zeta \in EM$, "$'$" defined as (4.18) is an order reversing involution in the EI algebra $(EM,\vee,\wedge)$. Thus for any $\zeta = \sum_{i\in I}(\prod_{m\in A_i} m) \in EM$, ζ' defined as the follow (4.19) stands for the logical negation of ζ.

$$\zeta' = \bigwedge_{i\in I}(\bigvee_{m\in A_i} m') = \bigwedge_{i\in I}(\sum_{m\in A_i} m'). \tag{4.19}$$

$(EM,\vee,\wedge,')$ is called an *AFS fuzzy logic system*. In Example 4.1, for $\gamma = m_1m_6 + m_1m_3 + m_2 \in EM$, by (4.19), we have

$$\begin{aligned}\gamma' &= (m_1m_6+m_1m_3+m_2)' \\ &= (m_1'+m_6')\wedge(m_1'+m_3')\wedge m_2' \\ &= (m_1'+m_6'm_3')\wedge m_2' \\ &= m_1'm_2'+m_2'm_3'm_6'\end{aligned}$$

γ', which is the logical negation of $\gamma = m_1m_6+m_1m_3+m_2$, reads as "*not old and not tall people*" or "*not tall and not heavy females*".

The AFS fuzzy logic system $(EM,\vee,\wedge,')$ being regarded as a completely distributive lattice not only provides a sound mathematical tool to study the fuzzy concepts in *EM* and to construct their membership functions and logic operations, but also ensures us that they are the fuzzy sets of some well-understood underlying semantics.

For M being a set of few fuzzy or Boolean concepts, a large number of fuzzy concepts can be expressed by the elements of *EM* and the fuzzy logic operations can be implemented by the operations $\vee$, $\wedge$ and $'$ available in the *EI* algebra system $(EM,\vee,\wedge,')$, even though we have not specified the membership functions of the fuzzy concepts in *EM*. In other words, the expressions and the fuzzy logic operations of the fuzzy concepts in *EM* just focus on the few simple concepts in M and the semantics of the fuzzy concepts in *EM*. As long as we can determine the fuzzy logic operations of these few concepts in M, the fuzzy logic operations of all concepts in *EM* can also be determined. Thus, not only will the accuracy of the representations and the fuzzy logic operations of fuzzy concepts be improved in comparison with the fuzzy logic directly equipped with some t-norms and a negation operator, but also the complexity of determining membership functions and their logic operations for the complex fuzzy concepts in EM will be alleviated. Let us stress that the complexity of human concepts is a direct result of the combinations of a few relatively simple concepts. It is obvious that the simpler the concepts in M, the more accurately and conveniently the membership functions and the fuzzy logic operations of the fuzzy concepts in EM will be determined. A collection of a few concepts in M plays a similar role to the one of a "*basis*" used in linear vector spaces. In what follows, we elaborate on a suite of "*simple concepts*" which can be regarded as a "*basis*".

4.1.2 Simple Concepts and Complex Concepts

In this section, we first recall the interpretations of graded membership [8]. Then we discuss the complexity of the concepts through the analysis of the binary relation structures of concepts.

Identifying the three types of information-driven tasks where graded membership plays a role: classification and data analysis, decision-making problems, and approximate reasoning, Dubois gave three views at the semantics of the membership grades, respectively, in terms of similarity, preference, and uncertainty [8]. In more detail, considering the degree of membership $\mu_F(u)$ of an element u in a fuzzy set F, defined on a referential U, three interpretations of this degree are sought:

Degree of similarity: $\mu_F(u)$ is the degree of proximity of u to prototype elements of F. Historically, this is the oldest semantics of membership grades since Bellman advocated the interest of the fuzzy set concept in pattern classification from the very inception of the theory [1].

Degree of preference: F represents a set of more or less preferred objects (or values of a decision variable x) and $\mu_F(u)$ represents an intensity of preference in favor of object u, or the feasibility of selecting u as a value of x. Fuzzy sets then represent criteria or flexible constraints. This view is the one later put forward by Bellman and Zadeh in [2].

Degree of uncertainty: This interpretation was proposed by Zadeh in [53] when he introduced the possibility theory and developed his theory of approximate reasoning [54]. $\mu_F(u)$ is then the degree of possibility that a parameter x has value u, given that all that is known about it is that "x is F".

For a fuzzy set A on the universe of discourse X, by comparison of the graded membership (e.g., Dubois's interpretations of membership degree), Turksen induced an "empirical relational membership structure"[46], $< X, R_A >$, where $R_A \subseteq X \times X$ is a binary relation on X, $(x,y) \in R_A \Leftrightarrow$ an observer, an expert, judges that " x is at least as much a member of the fuzzy set A as y" or "x's degree of membership in A is at least as large as y's degree of membership in A. The fundamental measurement of the gradual-set membership function can be formulated as the construction of homomorphisms from an "empirical relational membership structure", $< X, R_A >$, to a "numerical relational membership structure", $\langle \{\mu_A(x) | x \in X\}, \geq \rangle$.

Furthermore we formulated these ideas as the following definitions of the binary relations to represent concepts.

Definition 4.2. ([30]) Let ζ be any concept on the universe of discourse X. R_ζ is called the *binary relation (i.e., $R_\zeta \subset X \times X$) of the concept* ζ if R_ζ satisfies: $x, y \in X$, $(x,y) \in R_\zeta \Leftrightarrow x$ belongs to concept ζ at some extent (or x is a member of ζ) and the degree of x belonging to ζ is larger than or equal to that of y, or x belongs to concept ζ at some degree and y does not at all.

For instance, according to the value of each $x \in X$ on the feature of age shown in Table 4.1, we have the binary relations R_ζ, $R_{\zeta'}$, R_γ of the fuzzy concepts ζ: "*old*", ζ': "*not old*", γ: "*the person whose age is about 40 years*" as follows:

$$R_\zeta = \{ (x,y) \mid (x,y) \in X \times X,\ age_x \geq age_y \},$$
$$R_{\zeta'} = \{ (x,y) \mid (x,y) \in X \times X,\ age_x \leq age_y \},$$
$$R_\gamma = \{ (x,y) \mid (x,y) \in X \times X,\ |age_x - 40| \leq |age_y - 40| \},$$

where age_x is the age of x. Note that $(x,x) \in R_\eta$ implies that x belongs to η at some degree and that $(x,x) \notin R_\eta$ implies that x does not belong to η at all. For the fuzzy concept m_5: "*more estate*" in Example 4.1, by feature "*estate*" and Definition 4.2, we have $(x_5, x_5) \in R_{m_5}$ although the estate of x_5 is just 2, and $(x_2, x_2) \notin R_{m_5}$ because the estate of x_2 is 0. For a Boolean concept ξ, $(x,x) \in R_\xi$ implies that x belongs to concept ξ. For instance, the concept m_6: "*male*" in Example 4.1, considering the feature "*male*" and Definition 4.2, we have $(x_1,y),(x_4,y),(x_5,y),(x_7,y),(x_8,y) \in R_{m_6}$ and $(x_2,y),(x_3,y),(x_6,y),(x_9,y),(x_{10},y) \notin R_{m_6}$ for any $y \in X$. In real world applications, the comparison of the degrees of a pair x and y belonging to a concept can be obtained through the use of the values of the feature or by relying on human intuition, even though the membership function of the fuzzy concept has not been specified by the degrees in [0, 1] or a lattice in advance. For instance, we can obtain the binary relation R_{m_8} for fuzzy concept m_8: "*black hair people*" in Example 4.1, just by comparing each pair of people' hair and expressing our intuitive judgment. Based on Table 4.1 and following this intuitive assessment, we can construct the binary relation R_m of each concept $m \in M$ being used in Example 4.1.

Compared with "empirical relational membership structure", $< X, R_A >$ defined by Turksen in [46], Definition 4.2 stresses that if $(x,y) \in R_A$ then x must be a member of A to some extent or x's degree of membership in A is not equal to zero. By the definition of $< X, R_A >$ in [46], we know that if both x and y are not members of A then x's degree of membership in A is at least as large as y's degree of membership in A and $(x,y) \in R_A$. It is unnatural and un-convenient for us to derive membership functions from the binary relations.

Definition 4.3. ([28, 30]) Let X be a set and R be a binary relation on X. R is called a *sub-preference relation* on X if for x, y, $z \in X$, $x \neq y$, R satisfies the following conditions:

1. If $(x,y) \in R$, then $(x,x) \in R$;
2. If $(x,x) \in R$ and $(y,y) \notin R$, then $(x,y) \in R$;

3. If (x,y), $(y,z) \in R$, then $(x,z) \in R$;
4. If $(x,x) \in R$ and $(y,y) \in R$, then either $(x,y) \in R$ or $(y,x) \in R$.

A concept ζ is called a *simple concept* on X if R_ζ is a sub-preference relation on X. Otherwise ζ is called a *complex concept* on X. R_ζ is the binary relation of ζ defined by Definition 4.2.

In [46] Turksen defined a "*weak order*" relation R on X as follows: for all $x,y,z \in X$, R is called a weak order relation on X, if the following axioms are satisfied:

1. Connectedness: Either $(x,y) \in R$ or $(y,x) \in R$;
2. Transitivity: If (x,y), $(y,z) \in R$, then $(x,z) \in R$.

Indeed weak order is equivalent to the *preference relation* defined by Kim in [16]. It can be proved that for any sub-preference relation R on X there exists a preference relation $\overline{R}$ such that $R \subseteq \overline{R}$. So that the binary relations defined in Definition 4.3 are called sub-preference relations. For a sub-preference relation R on X, if $\forall x \in X$, $(x,x) \in R$ then R is a weak order. Thus the sub-preference relation is a generalization of the weak order.

The essential difference between a simple concept and a complex concept on a set X is in that all elements belonging to a simple concept at some degree are comparable (i.e., they form a linear order or total order) and there exists a pair of different elements belonging to a complex concept at some degree such that their degrees of belonging to the complex concept are incomparable. For example, let X be a set of people. Also assume that X contains two disjoint subsets Y and Z of male and female respectively. It is easy to see that, if we consider incomparable the elements of Y with those of Z, then the property of being "*beautiful*" is a simple concept if restricted to Y or Z, while it is a complex concept if applied to the whole set X. In fact, if $x,y \in X$, $x \in Y$ and $y \in Z$, then the degree of x,y belonging to "*beautiful*" may be incomparable although both x and y may belong to "*beautiful*" at some degree, i.e., $(x,x),(y,y) \in R_{beautiful}$, $(y,x) \notin R_{beautiful}$, $(x,y) \notin R_{beautiful}$. This implies that the fourth condition of Definition 4.3 is not satisfied and "*beautiful*" is a complex concept on X. By Table 4.1 and Definition 4.3, one can verify that each concept $m \in M$ in Example 4.1 is a simple concept. Many concepts associated with more than a single feature are complex concepts. In Example 4.1, let fuzzy concept $\beta = m_1 m_2 \in EM$ meaning "*tall old people*". One can verify that $(x_8,x_8),(x_4,x_4) \in R_\beta$, but neither (x_8,x_4) nor (x_4,x_8) in R_β. This implies that the fourth condition of Definition 4.3 is not satisfied by R_β and therefore β is also a complex concept. The fuzzy concept $\gamma = m_1 + m_2 \in EM$ reads as "*old people*" or "*tall people*". By the data shown in Table 4.1, i.e., x_8: age=70, height=1.6; x_1: age=20, height=1.9; x_4: age=80, height=1.8, we have $(x_8,x_1) \in R_\gamma$ because x_8 is older than x_1 and $(x_1,x_4) \in R_\gamma$ because x_1 is taller than x_4. But x_8 is neither older nor taller than x_4, i.e., $(x_8,x_4) \notin R_\gamma$. Thus the binary relation R_γ does not satisfy the third condition of Definition 4.3 and therefore concept γ is a complex concept.

It is clear that with any simple concept ζ, X can be divided into three classes.

$$\begin{aligned} T_\zeta &= \{x \in X | (x,y) \in R_\zeta, \forall y \in X\}, \\ O_\zeta &= \{x \in X | (x,x) \notin R_\zeta\}, \\ F_\zeta &= X - T_\zeta - O_\zeta. \end{aligned} \tag{4.20}$$

Let $M_\zeta = (r_{ij})_{n\times n}$ be the correspondent Boolean matrix of R_ζ defined by Definition 3.5. The above (4.20) implies that there exist a permutation Boolean matrix P such that

$$M_\zeta = P \begin{bmatrix} J_{11} & J_{12} & J_{13} \\ O_{21} & N_{22} & J_{23} \\ O_{31} & O_{32} & O_{33} \end{bmatrix} P^T, \tag{4.21}$$

where $J_{11}, J_{12}, J_{13}, J_{23}$ are Boolean matrices whose elements are all 1; O_{21}, O_{31}, O_{32}, O_{33} are Boolean matrices whose elements are all 0 and N_{22} is a square sub-block Boolean matrix such that $N_{22}+I = N_{22}, N_{22}^2 = N_{22}$ (refer to Theorem 3.1). Moreover, according to the semantics of the fuzzy concepts, ζ' the negation of the simple concept ζ can be constructed by its correspondent Boolean matrix $M_{\zeta'}$ shown as follows:

$$M_{\zeta'} = P \begin{bmatrix} O_{11} & O_{12} & O_{13} \\ J_{21} & N_{22}^T & O_{23} \\ J_{31} & J_{32} & J_{33} \end{bmatrix} P^T, \tag{4.22}$$

where $O_{11}, O_{12}, O_{13}, O_{23}$ are the Boolean matrices whose elements are all 0; J_{21}, J_{31}, J_{32}, J_{33} are Boolean matrices whose elements are all 1. The dimensions of O_{11}, O_{12}, O_{13}, O_{23} are the same as J_{11}, J_{12}, J_{13}, J_{23}, respectively and the dimensions of J_{21}, J_{31}, J_{32}, J_{33} are the same as O_{21}, O_{31}, O_{32}, O_{33}, respectively.

For each element in T_ζ, its degree belongingness to concept ζ is 1 if concept ζ is represented by an ordinary fuzzy set and its degree belonging to concept ζ is the maximum element of lattice L if concept ζ is represented by a L-fuzzy set; for each element in O_ζ, its degree belongingness (membership) to concept ζ is 0 if concept ζ is represented by an ordinary fuzzy set and its degree belonging to concept ζ is the minimum element of lattice L if concept ζ is represented by a L-fuzzy set; the elements in F_ζ belong to concept ζ at different degrees in open interval (0,1) if concept ζ is represented by an ordinary fuzzy set and the degrees of elements in F_ζ belonging to concept ζ form a linearly ordered chain in lattice L if concept ζ is represented by a L-fuzzy set. Consequently, concept ζ is a Boolean concept if $F_\zeta = \emptyset$.

Definition 4.4. Let $X = \{x_1, x_2, ..., x_n\}$ and R_ζ be the binary relation of concept ζ on X. Let $M_\zeta = (r_{ij})_{n\times n}$ be the correspondent Boolean matrix of R_ζ defined by Definition 3.5. Then the concept ζ is called an *atomic fuzzy concept* if there exists a permutation Boolean matrix P such that

$$PM_\zeta P^T = \begin{bmatrix} J_{11} & J_{12} \\ O_{21} & J_{22} \\ O_{31} & O_{32} \end{bmatrix}, \tag{4.23}$$

where $J_{11}, J_{12}, J_{22}, O_{21}, O_{22}, O_{32}$ are sub-block Boolean matrices of appropriate dimensionless and each element in J_{11}, J_{12}, J_{22} is 1 and each element in O_{21}, O_{22}, O_{32} is equal to 0.

By Definition 4.4, we can easily verify that each atomic fuzzy concept is a simple concept. Similarly X is also divided into three classes $T_\zeta, O_\zeta, F_\zeta$ (refer to (4.20)) by an atomic fuzzy concept. For each element in T_ζ, its degree belonging to concept ζ is 1 if concept ζ is represented by an ordinary fuzzy set; for each element in F_ζ, its degree belonging to concept ζ is 0; Compared to the ordinary simple concept, for the atomic fuzzy concept ζ, the degrees of all elements in F_ζ belonging to ζ are equal to one value in (0,1). By Theorem 3.3, we know that for a simple concept γ, if $M_\gamma > M_\zeta$ and ζ is an atomic fuzzy concept, then $F_\gamma = \varnothing$ and the concept γ is a Boolean concept. This implies that the atomic fuzzy concepts are the simplest of the simple concepts except Boolean concepts.

Let X be a set and ζ be any concept on X. Let R_ζ be the binary relation of the simple concept ζ defined by Definition 4.2 and $M_\zeta = (r_{ij})_{n\times n}$ be the correspondent Boolean matrix of R_ζ defined by Definition 3.5. By the definitions, we can verify that $r_{ii} = 0 \Leftrightarrow r_{ij} = 0$ for all $j = 1, 2, ..., n$. Thus there exists a permutation Boolean matrix P such that

$$M_\zeta = P \begin{bmatrix} N & J \\ O_1 & O_2 \end{bmatrix} P^T, \tag{4.24}$$

where N is a Boolean matrix such that $N + I = N$, J is a universal Boolean matrix, i.e.,whose elements are all 1, O_1 and O_2 are zero matrices. One can also verify that the concept ζ on a set X is a simple concept if and only if N is the correspondent Boolean matrix of a quasi-linear order, i.e., $N^2 = N, N + I = N$. By Definition 4.3, we can verify that any quasi-linear order on a set is a sub-preference relation on the set.The proofs of the conclusions remain as exercises.

Theorem 4.3. *([28]) For any fuzzy concept η on a finite set X, there exists M a set of simple concepts on X and a fuzzy concept $\xi = \sum_{i\in I} \prod_{m\in A_i} m \in EM$ such that $R_\eta = R_\xi$ provided that for $x, y \in X$, $(x,y) \in R_\xi \Leftrightarrow \exists k \in I$ such that $(x,y) \in R_{A_k}$ (i.e., $\forall m \in A_k, (x,y) \in R_m$).*

Proof. Let $M_\eta = (r_{ij})_{n\times n}$ be the corresponding Boolean matrix of R_η expressed by Definition 3.5 for concept η. From the above discussion, we know that there exists a Boolean permutation matrix P such that

$$M_\eta = P \begin{bmatrix} N & J \\ O_1 & O_2 \end{bmatrix} P^T,$$

where N is a Boolean matrix such that $N + I = N$.

According to Theorem 3.9, we have that for the Boolean matrix N, there exist quasi-linear orders $R_{ij}, i = 1, 2, ..., r, j = 1, 2, ..., q_i$ such that

$$N = \sum_{i=1}^{r} \left(* \prod_{k=1}^{q_i} M_{R_{ik}} \right).$$

Furthermore, by Theorem 3.5 we have

$$M_\eta = P\begin{bmatrix} N & J \\ O_1 & O_2 \end{bmatrix} P^T = P\begin{bmatrix} \sum_{i=1}^{r}\left(*\prod_{k=1}^{q_i} M_{R_{ik}}\right) & J \\ O_1 & O_2 \end{bmatrix} P^T$$
$$= \sum_{i=1}^{r}\left(*\prod_{k=1}^{q_i} P\begin{bmatrix} M_{R_{ik}} & J \\ O_1 & O_2 \end{bmatrix} P^T\right). \tag{4.25}$$

By (4.24), we also know that the concept corresponding to $P\begin{bmatrix} M_{R_{ik}} & J \\ O_1 & O_2 \end{bmatrix} P^T$ for each i,k is a simple concept. Let m_{ik} be the simple concept corresponding to $P\begin{bmatrix} M_{R_{ik}} & J \\ O_1 & O_2 \end{bmatrix} P^T$ and $M = \{m_{ik} | i = 1,2,...,r, j = 1,2,...,q_i\}$. If $\xi = \sum_{i=1}^{r}(\prod_{k=1}^{q_i} m_{ik})$ in EM, then by (4.25) we have $R_\eta = R_\xi$. □

This implies that any fuzzy concept on X can be expressed by some simple concepts on X and the AFS fuzzy logic system $(EM, \vee, \wedge, ')$ forms a comprehensive platform capturing the mathematical structures and the semantics of concepts used by humans.

4.1.3 AFS Structure of Data

Many data sets involve a mixture of quantitative and qualitative feature variables, some examples have been shown in Table 4.1. Beside quantitative features, qualitative features, which could be further classified as nominal and ordinal features, are also commonly encountered. Taking advantage of AFS structure of data, the qualitative features evaluated by information-based criteria such as human perception-based information, gain ratio, symmetric uncertainty, order, binary relation, etc can be applied to determine the membership functions of fuzzy concepts.

Definition 4.5. ([23, 25]) Let X, M be sets and 2^M be the power set of M. Let $\tau : X \times X \to 2^M$. (M, τ, X) is called an *AFS structure* if τ satisfies the following axioms:

AX1: $\forall (x_1, x_2) \in X \times X$, $\tau(x_1, x_2) \subseteq \tau(x_1, x_1)$;
AX2: $\forall (x_1, x_2)$, $(x_2, x_3) \in X \times X$, $\tau(x_1, x_2) \cap \tau(x_2, x_3) \subseteq \tau(x_1, x_3)$.

X is called *universe of discourse*, M is called a *concept set* and τ is called a *structure*.

For an AFS structure (M, τ, X), if we define $f_\tau(x,y) = \tau(x,y) \cap \tau(y,y), \forall (x,y) \in X \times X$, then $(M, f_\tau, X \times X)$ is a combinatoric system [11].

Let X be a set of objects and M be a set of simple concepts on X. If $\tau : X \times X \to 2^M$ is defined as follows: for any $(x,y) \in X \times X$

$$\tau(x,y) = \{m \mid m \in M, (x,y) \in R_m\} \in 2^M, \tag{4.26}$$

where R_m is the binary relation of simple concept $m \in M$ (refer to Definition 4.2 and Definition 4.3). Then (M, τ, X) is an AFS structure. The proof goes as follows. For any $(x_1, x_2) \in X \times X$, if $m \in \tau(x_1, x_2)$, then by (4.26) we know $(x_1, x_2) \in$

R_m. Because each $m \in M$ is a simple concept, we have $(x_1,x_1) \in R_m$, i.e., $m \in \tau(x_1,x_1)$. This implies that $\tau(x_1,x_2) \subseteq \tau(x_1,x_1)$ and AX1 of Definition 4.5 holds. For $(x_1,x_2),(x_2,x_3) \in X \times X$, if $m \in \tau(x_1,x_2) \cap \tau(x_2,x_3)$, then $(x_1,x_2),(x_2,x_3) \in R_m$. Since m is a simple concept, so by Definition 4.3, we have $(x_1,x_3) \in R_m$, i.e., $m \in \tau(x_1,x_3)$. This implies $\tau(x_1,x_2) \cap \tau(x_2,x_3) \subseteq \tau(x_1,x_3)$ and AX2 of Definition 4.5 holds. Therefore (M,τ,X) is an AFS structure. In light of the above discussion, an AFS structure based on a data set can be established by (4.26), as long as each concept in M is a simple concept on X.

Let us continue with Example 4.1, in which $X = \{x_1,x_2,...,x_{10}\}$ is the set of 10 people and their features are shown in Table 4.1. $M = \{m_1,m_2,\ ...,\ m_{12}\}$ is the set of simple concepts shown in Example 4.1. By Table 4.1 and Definition 4.3, one can verify that each concept $m \in M$ is a simple concept. Thus for any $x,y \in X$, $\tau(x,y)$ is well-defined by (4.26). For instance, we have

$$\tau(x_4,x_4) = \{m_1,m_2,m_3,m_4,m_5,m_6,m_8,m_9,m_{10},m_{11},m_{12}\}$$
$$\tau(x_4,x_7) = \{m_1,m_2,m_6,m_9,m_{10}\}$$

by comparing the feature values of x_4,x_7 shown in Table 4.1 as follows:

	age	*height*	*weigh*	*salary*	*estate*	*m.*	*f.*	*black*	*white*	*yellow*
x_4	80	1.8	73	20	80	1	0	3	4	2
x_7	45	1.7	78	268	90	1	0	1	6	4

Similarly, we can obtain $\tau(x,y)$ for other $x,y \in X$. Finally, we arrive at the AFS structure (M,τ,X) of the data shown in Table 4.1.

Let M be a set of simple concepts on X and g be a map $g : X \times X \rightarrow 2^M$. In general g may not be guaranteed to satisfy AX1, AX2 of Definition 4.5. Making use of the following theorem g can be converted into τ such that (M,τ,X) becomes an AFS structure.

Definition 4.6. Let M, $X = \{x_1,x_2,...,x_n\}$ be finite sets and $g : X \times X \rightarrow 2^M$. $M_g = (m_{ij})_{n\times n}$ is called a *Boolean matrix of the map g* if $m_{ij} = g(x_i,x_j) \in 2^M$. For M_g,M_h, the Boolean matrices of the maps g,h:$X \times X \rightarrow 2^M$, $M_g + M_h = (g(x_i,x_j) \cup h(x_i,x_j))_{n\times n}$, $M_gM_h = (q_{ij})_{n\times n}, q_{ij} = \bigcup_{1\leq k\leq n} (g(x_i,x_k) \cap h(x_k,x_j)), i,j = 1,2,...,n$.

Theorem 4.4. *Let M, $X = \{x_1,x_2,...,x_n\}$ be finite sets and $g : X \times X \rightarrow 2^M$. Then g is a structure of an AFS structure, that is, g satisfies AX1, AX2 of Definition 4.5 if and only if*

$$M_g^2 = M_g \text{ and } \bigcup_{1\leq j\leq n} m_{ij} \subseteq m_{ii}, i = 1,2,...,n.$$

Its proof remains as an exercise.

Based on the criteria presented in this theorem, one can establish an AFS structure (M,τ,X) if g does not satisfy AX1 and AX2 in Definition 4.5. One may note that (M,τ,X) is the mathematical abstraction of the complex relationships existing among objects in X with the attributes in M. This implies that the information contained in databases and human intuition are aggregated to (M,τ,X) from which we can obtain the fuzzy sets and AFS fuzzy logic operations for the fuzzy concepts expressed by the elements in EM.

4.1.4 Coherence Membership Functions of the AFS Fuzzy Logic and the AFS Structure of Data

In this section, we discuss how to determine the membership functions for the fuzzy concepts in EM according to the AFS structure (M,τ,X) of the data and the semantics of the concepts. At the same time, the membership functions are consistent with both the AFS logic system $(M,\vee,\wedge,')$ in the sense of the underlying semantics and the distribution of the data.

Definition 4.7. Let (M,τ,X) be an AFS structure of a data set X. For $x\in X$, $A\subseteq M$, the set $A^{\tau}(x)\subseteq X$ is defined as follows.

$$A^{\tau}(x)=\{y\mid y\in X,\tau(x,y)\supseteq A\}. \tag{4.27}$$

For $\zeta\in EM$, let $\mu_{\zeta}:X\to[0,1]$ be the membership function of the concept ζ. $\{\mu_{\zeta}(x)\mid\zeta\in EM\}$ is called a set of *coherence membership functions* of the AFS fuzzy logic system $(EM,\vee,\wedge,')$ and the AFS structure (M,τ,X), if the following conditions are satisfied.

1. for $\alpha,\beta\in EM$, if $\alpha\leq\beta$ in lattice $(EM,\vee,\wedge,')$, then $\mu_{\alpha}(x)\leq\mu_{\beta}(x)$ for any $x\in X$;
2. for $x\in X,\eta=\sum_{i\in I}(\prod_{m\in A_i} m)\in EM$, if $A_i^{\tau}(x)=\varnothing$ for all $i\in I$ then $\mu_{\eta}(x)=0$;
3. for $x,y\in X,A\subseteq M$, $\eta=\prod_{m\in A} m\in EM$, if $A^{\tau}(x)\subseteq A^{\tau}(y)$, then $\mu_{\eta}(x)\leq\mu_{\eta}(y)$; if $A^{\tau}(x)=X$ then $\mu_{\eta}(x)=1$.

The following proposition stresses that the coherence membership functions are consistent with the AFS logic system $(M,\vee,\wedge,')$ in terms of the underlying semantics.

Proposition 4.2. *Let M be a set of simple concepts on X and (M,τ,X) be an AFS structure defined as (4.26). Let $\{\mu_{\zeta}(x)\mid\zeta\in EM\}$ be a set of coherence membership functions of $(EM,\vee,\wedge,')$ and (M,τ,X). Then for any $\alpha,\beta\in EM$, any $x\in X$,*

$$\mu_{\alpha\vee\beta}(x)\geq max\{\mu_{\alpha}(x),\mu_{\beta}(x)\},\;\;\mu_{\alpha\wedge\beta}(x)\leq min\{\mu_{\alpha}(x),\mu_{\beta}(x)\} \tag{4.28}$$

Proof. In lattice $(EM,\vee,\wedge,')$, for any $\alpha,\beta\in EM$, we have $\alpha\vee\beta\geq\alpha$, $\alpha\vee\beta\geq\beta$ and $\alpha\wedge\beta\leq\alpha$, $\alpha\wedge\beta\leq\beta$. Using condition 1 of Definition 4.7, for any $x\in X$, one has

$$\mu_{\alpha\vee\beta}(x)\geq\mu_{\alpha}(x),\quad\mu_{\alpha\vee\beta}(x)\geq\mu_{\beta}(x),$$
$$\mu_{\alpha\wedge\beta}(x)\leq\mu_{\alpha}(x),\quad\mu_{\alpha\wedge\beta}(x)\leq\mu_{\beta}(x).$$

This implies that (4.28) holds. □

Proposition 4.3. *Let M be a set of Boolean concepts on X and (M,τ,X) be an AFS structure defined as (4.26). Let $\{\mu_\zeta(x) \mid \zeta \in EM\}$ be a set of coherence membership functions of $(EM,\vee,\wedge,')$ and (M,τ,X). Then for any $\xi,\eta \in EM$, any $x \in X$ the following assertions hold:*

1. *either $\mu_\xi(x)=1$ or $\mu_\xi(x)=0$;*
2.

$$\mu_{\xi\wedge\eta}(x) = min\{\mu_\xi(x),\mu_\eta(x)\},$$
$$\mu_{\xi\vee\eta}(x) = max\{\mu_\xi(x),\mu_\eta(x)\},$$
$$\mu_{\xi'}(x) = 1-\mu_\xi(x),$$

i.e., $(EM,\vee,\wedge,')$ is degenerated to a Boolean logic system $(2^X,\cap,\cup,')$.

Proof. 1. From Definition 4.3, we know that for each $m \in M$ since it is a Boolean concept, hence either $\{m\}^\tau(x)=X$ or $\{m\}^\tau(x)=\varnothing$. For any $A \subseteq M$ and any $x \in X$, by (4.27), we have $A^\tau(x)=\bigcap_{m\in A}\{m\}^\tau(x)$. Thus either $A^\tau(x)=X$ or $A^\tau(x)=\varnothing$. By conditions 2,3 of Definition 4.7, either $\mu_{\prod_{m\in A} m}(x)=1$ or $\mu_{\prod_{m\in A} m}(x)=0$. For any $\xi=\sum_{i\in I}\prod_{m\in A_i} m \in EM$,$x \in X$, if there exists $k \in I$ such that $A_k^\tau(x) \neq \varnothing$ then $\mu_{\prod_{m\in A_k} m}(x)=1$. Thus by condition 1 of Definition 4.7 and $\prod_{m\in A_k} m \leq \xi$, we have $1=\mu_{\prod_{m\in A_k} m}(x) \leq \mu_\xi(x) \leq 1$, i.e., $\mu_\xi(x)=1$. If for any $i \in I$, $A_k^\tau(x)=\varnothing$ then by condition 2 of Definition 4.7 $\mu_\xi(x)=0$.

2. Let $\xi=\sum_{i\in I}(\prod_{m\in A_i} m), \eta=\sum_{j\in J}(\prod_{m\in B_j} m) \in EM$ and $x \in X$. In the case that one of $\mu_\xi(x)$ and $\mu_\eta(x)$ is 0, by Proposition 4.2, we have

$$0 \leq \mu_{\xi\wedge\eta}(x) \leq min\{\mu_\xi(x),\mu_\eta(x)\}=0.$$

If $\mu_\xi(x)=1$ and $\mu_\eta(x)=1$, then by the above proof, we know that there exist $k \in I, h \in J$ such that $A_k^\tau(x)=X, B_h^\tau(x)=X$. By (4.3), we have

$$\xi\wedge\eta = \sum_{i\in I, j\in J}(\prod_{m\in A_i\cup B_j} m).$$

Thus for $k \in I, h \in J$, $(A_k\cup B_h)^\tau(x)=A_k^\tau(x)\cap B_h^\tau(x)=X$. By conditions 1, 3 of Definition 4.7, one has $1=\mu_{\prod_{m\in A_k\cup B_h} m}(x) \leq \mu_{\xi\wedge\eta}(x) \leq 1$, i.e., $\mu_{\xi\wedge\eta}(x)=1$. Thus we prove that $\mu_{\xi\wedge\eta}(x)=\min\{\mu_\xi(x), \ \mu_\eta(x)\}$. Similarly, we also can prove that $\mu_{\xi\vee\eta}(x)=max\{\mu_\xi(x),\mu_\eta(x)\}$ and $\mu_{\xi'}(x)=1-\mu_\xi(x)$. □

Theorem 4.5 provides a constructive method to define coherence membership functions in which both the distribution of the data and the semantics of the fuzzy concepts are taken into account.

Theorem 4.5. *Let M be a set of simple concepts on X and (M,τ,X) be an AFS structure defined as (4.26). Let S be a σ-algebra over X such that for any $m \in M$ and any $x \in X$, $\{m\}^\tau(x) \in S$. For each simple concept $\gamma \in M$, let $\mathscr{M}_\gamma$ be a measure over S with $0 \leq \mathscr{M}_\gamma(U) \leq 1$ for all $U \in S$ and $\mathscr{M}_\gamma(X)=1$. $\{\mu_\zeta(x) \mid \zeta \in EM\}$ is a set of coherence membership functions of $(EM,\vee,\wedge,')$ and (M,τ,X), if for each* *concept* $\zeta = \sum_{i\in I}(\prod_{m\in A_i} m) \in EM$, $\mu_\zeta : X \to [0,1]$ *is defined as follows: for any* $x \in X$

$$\mu_\zeta(x) = \sup_{i\in I}\left(\prod_{\gamma\in A_i} \mathscr{M}_\gamma(A_i^\tau(x))\right), \tag{4.29}$$

or

$$\mu_\zeta(x) = \sup_{i\in I}\left(\inf_{\gamma\in A_i} \mathscr{M}_\gamma(A_i^\tau(x))\right). \tag{4.30}$$

Proof. Let $\alpha = \sum_{i\in I}(\prod_{m\in A_i} m)$, $\beta = \sum_{j\in J}(\prod_{m\in B_j} m) \in EM$ and $\alpha \le \beta$ in lattice $(EM,\vee,\wedge,')$. By Theorem 4.1, we know that for any A_i $(i \in I)$, there exist B_h $(h \in J)$ such that $A_i \supseteq B_h$. By (4.27), we have $A_i^\tau(x) \subseteq B_h^\tau(x)$ for any $x \in X$. Thus for any $i \in I$,

$$\begin{aligned}\prod_{\gamma\in A_i} \mathscr{M}_\gamma(A_i^\tau(x)) &\le \prod_{\gamma\in A_i} \mathscr{M}_\gamma(B_h^\tau(x))\\ &\le \prod_{\gamma\in B_h} \mathscr{M}_\gamma(B_h^\tau(x))\\ &\le \mu_\beta(x).\end{aligned}$$

Furthermore, one has

$$\mu_\alpha(x) = \sup_{i\in I}\prod_{\gamma\in A_i} \mathscr{M}_\gamma(A_i^\tau(x)) \le \mu_\beta(x).$$

Thus condition 1 of Definition 4.7 holds.

Since $\mathscr{M}_\gamma(\varnothing) = 0$ for any $\gamma \in M$ hence condition 2 of Definition 4.7 holds.

For $x,y \in X, A \subseteq M$, $\eta = \prod_{m\in A} m \in EM$, if $A^\tau(x) \subseteq A^\tau(y)$, then for any $\gamma \in A$, $\mathscr{M}_\gamma(A^\tau(x)) \le \mathscr{M}_\gamma(A^\tau(y))$. This implies that $\mu_\eta(x) \le \mu_\eta(y)$. Furthermore, since $\mathscr{M}_\gamma(X) = 1$, hence condition 3 of Definition 4.7 holds. Therefore $\{\mu_\zeta(x) \mid \zeta \in EM\}$ is the set of coherence membership functions of $(EM,\vee,\wedge,')$ and (M,τ,X). □

In theory, as long as for each $\gamma \in M$ and any $U \in S$, $0 \le \mathscr{M}_\gamma(U) \le 1$ and $\mathscr{M}_\gamma(U) = 1$, the functions defined by (4.29) or (4.30) are coherence membership functions. In real world applications, the measure $\mathscr{M}_\gamma$ can be constructed according to the semantic meaning of the simple concept γ and may have various interpretations depending on the specificity of the problem at hand. In general, $\mathscr{M}_\gamma(A^\tau(x))$ measures the degree of set $A^\tau(x)$ supporting the claim: "*x belongs to γ*".

The *coherence* applied to membership functions of fuzzy concepts in *EM* to denote the membership functions which respect the semantic interpretations expressed by the fuzzy concepts, the logic relationships among the fuzzy concepts in AFS logic systems $(EM,\vee,\wedge,')$ and the distribution of the data. In what follows we explain the conditions of Definition 4.7 and show that the *coherence* approach to the fuzzy sets in AFS framework is crucial.

Condition 1: For $\alpha,\beta \in EM$, let $\alpha = \beta$, i.e., $\alpha \le \beta$ and $\alpha \ge \beta$. From condition 1, we have that $\mu_\alpha(x) = \mu_\beta(x)$ for any $x \in X$. Thus condition 1 ensures that the membership functions of the concepts in *EM* with equivalent meanings are identical.

For instance, in Example 4.1, one can verify that the semantic interpretations of $\xi = m_3m_8 + m_1m_4 + m_1m_6m_7 + m_1m_4m_8$ and $\zeta = m_3m_8 + m_1m_4 + m_1m_6m_7$ are equivalent. Thus the coherence membership functions of them are identical. By condition 1 and Proposition 4.2, we know that for coherence membership functions of any fuzzy concepts $\alpha, \beta \in EM$, any $x \in X$,

$$\mu_{\alpha\vee\beta}(x) \geq max\{\mu_\alpha(x), \mu_\beta(x)\}, \ \ \mu_{\alpha\wedge\beta}(x) \leq min\{\mu_\alpha(x), \mu_\beta(x)\}.$$

Thus the coherence membership functions of the fuzzy concepts in EM are consistent with their AFS fuzzy logic operations "$\vee$" (OR), "$\wedge$" (AND)," $'$" (NO) in $(EM, \vee, \wedge, ')$. For instance, in Example 4.1,

$\psi = m_1m_4 + m_2m_5m_6$ states " *old high salary people*" or "*tall male with more estate*",
$\vartheta = m_1m_4 + m_5m_6 + m_5m_8$ reads "*old high salary people*" or "*male with more estate*" or "*black hair people with more estate*",

By (4.2), we have $\psi \vee \vartheta = \vartheta$, i.e., $\psi \leq \vartheta$. Since the constraint of the concept "*tall male with more estate*" is stricter than that of "*male with more estate*", hence in terms of semantics, the membership degree of x belonging to "*tall male with more estate*" must lower than or equal to that of "*male with more estate*" for all $x \in X$.

Condition 2: It ensures that the fuzzy logic system $(EM, \vee, \wedge, ')$ is consistent with the Boolean logic. For $A \subseteq M$, we know that $A^\tau(x)$ is the set of all $y \in X$ such that the degrees of y belonging to the concept $\prod_{m\in A} m$ is less than or equal that of x by Definition 4.5 and formula (4.27). Since the AFS structure (M, τ, X) is determined by the distribution of the data, hence $A^\tau(x)$ is determined by both the distribution of data and the semantics of the simple concepts in A. $A_i^\tau(x) = \varnothing$ implies that there exists a simple concept in A_i such that x does not belong to. Thus the membership function $\mu_\eta(x)$ of concept $\eta = \sum_{i\in I}(\prod_{m\in A_i} m)$ has to $\mu_\eta(x) = 0$ if for any $i \in I$, $A_i^\tau(x) = \varnothing$. This also ensures that for any fuzzy concept $\eta \in EM$, $\mu_\eta(x) \geq 0$ for any $x \in X$.

Condition 3: It ensures that the coherence membership functions and their fuzzy logic operations observe both the distributions of the original data and the semantic interpretations of the fuzzy concepts. For example, what is your image (perception) of a person? If an NBA basketball player describes that the person is not *"tall"* and a ten year old child describes that the same person is very *"tall"*. Because the people the NBA basketball player often meets are different from the people the child meets, i.e., there are different data they observed. They may have different interpretations (membership functions) for the same concept *"tall"* due to the data sets drawn from different distribution probability space. Therefore the interpretations of concepts are strongly dependent on both the semantics of the concepts and the distribution of the observed data. The distributions of the observed data must be considered in the determining of the membership functions of the fuzzy concepts. Given (4.27), by Definition 4.5, we know that $A^\tau(x) \subseteq X$ is the set of all elements in X whose degrees of belonging to concept $\prod_{m\in A} m$ are less than or equal to that of x. $A^\tau(x) = \bigcap_{m\in A}\{m\}^\tau(x)$ is determined by both the semantics of the simple concepts in A and the AFS structure (M,τ,X) of the dataset which is dominated by a certain probability distribution, i.e., different distribution of the observed data may have different $A^{\tau}(x)$ for the same A and x. It is clear that condition 3 ensures that the larger the set $A^{\tau}(x)$, the larger the degree x belongingness (membership) to $\prod_{m\in A} m$ will be. For instance, in Example 4.1, Table 4.1 can be regarded as a set of samples randomly drawn from a certain population. Let $A=\{m_1,m_2\}\subseteq M$. Concept m_1m_2 states "*old and tall people*". In term of the feature values, *age* and *height* shown in Table 4.1

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
age	20	13	50	80	34	37	45	70	60	3
height	1.9	1.2	1.7	1.8	1.4	1.6	1.7	1.65	1.82	1.1

we have $A^{\tau}(x_5)=\{x_2,x_5,x_{10}\}\subseteq A^{\tau}(x_4)=\{x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_{10}\}$. Therefore the degree of x_5 belonging to concept m_1m_2 must be less than or equal to that of x_4 according to the given data.

For $A,B\subseteq M$, it is clear that $(A\cup B)^{\tau}(x)=A^{\tau}(x)\cap B^{\tau}(x)$. For a measure $\mathscr{M}$ on X, we know that $\mathscr{M}(A^{\tau}(x))$ and $\mathscr{M}(B^{\tau}(x))$ are not sufficient to determine $\mathscr{M}((A\cup B)^{\tau}(x))$ which is dependent on the distributions of the samples in the sets $A^{\tau}(x)$ and $B^{\tau}(x)$. As both the distributions of the original data and the semantic meanings of fuzzy concepts are taken into consideration, $\mu_{\eta}(x),\mu_{\zeta}(x)$, the coherence membership functions of fuzzy concepts $\eta=\prod\limits_{m\in A} m,\zeta=\prod\limits_{m\in B} m\in EM$ based on some measures on X(refer to Theorem 4.5), are not sufficient to determine $\mu_{\eta\wedge\zeta}(x)$, which is the membership degree of x belonging to the conjunction of η and ζ. This stands in a sharp contrast with the existing fuzzy logic systems equipped with some t-norm, in which $\mu_{\eta\wedge\zeta}(x)=T(\mu_{\eta}(x),\mu_{\zeta}(x))$ is fully determined by the membership degrees $\mu_{\eta}(x)$ and $\mu_{\zeta}(x)$ which is independent from the distribution of the original data. Hence, the constructed coherence membership functions and the logic operations in Theorem 4.5 include more information of the distributions of the original data and the semantic interpretations, i.e., it becomes more objective and less subjective.

4.2 Coherence Membership Functions via AFS Logic and Probability

In this section, we discuss the construction of coherence membership functions and provide interpretations of the measure $\mathscr{M}_{\gamma}$ (in Theorem 4.5) for the simple concept γ in the setting of probability theory. Thus the coherence membership functions based on the semantics of fuzzy concepts and the statistic characteristics of the observed data can be established for applications. Furthermore, the imprecision of natural language and the randomness of observed data can be put into work together, so that uncertainty of randomness and of imprecision can be treated in a unified and coherent manner.

We consider the following setting for the representation of subjective imprecision and the objective randomness. There is a "probability measure space", $(\Omega,\mathscr{F},\mathscr{P})$

of possible instances X or the observed samples based on which the coherence membership functions of fuzzy concepts in EM may be defined, where M is a set of simple concepts on X selected for the specificity of the problem at hand. We assume that different instances in X may be encountered at different frequencies. A convenient way to model this is to assume the probability distribution $\mathscr{P}$ defines the probability of encountering each instance in X (e.g.,$\mathscr{P}$ might assign a higher probability to encountering 19-year-old people than 109-year-old people). Notice $\mathscr{P}$ says nothing about the degree of x belonging to a concept $\zeta \in EM$; $\mathscr{P}(x)$ only determines the probability that x will be encountered. Let Ω be the universe of discourse and $F = \{f_1, f_2, \cdots, f_n\}$ be the set of all features on the objects in Ω. For any $x = (v_1, v_2, \cdots, v_n) \in \Omega$, $1 \leq i \leq n$, $v_j = f_j(x)$ is the value of x on the feature f_j. In general, the observed data X is drawn from Ω with a sampling density function $p(x) = \frac{d\mathscr{P}(x)}{dx}$. Let $X = \{x_1, x_2, ..., x_h\} \subseteq \Omega$ and N_x be the number of times that $x \in X$ is observed as a sample. If Ω is a discrete (i.e., countable) set, then $\frac{N_x}{|X|} \rightarrow \mathscr{P}(x)$ as the set X approaching to Ω.

4.2.1 Coherence Membership Functions Based on the Probability Measures

Let (M, τ, X) be an AFS structure of the observed data X and M be a set of simple concepts on X. Suppose that for any $m \in M$ and any $x \in X$, $\{m\}^\tau(x) \in \mathscr{F}$. For each simple concept $\gamma \in M$, assume that the measure $\mathscr{M}_\gamma$ for the simple concept γ under probability space be defined as follows: for any $U \in \mathscr{F} \cap 2^X$,

$$\mathscr{M}_\gamma(U) = \mathscr{P}(U) \approx \frac{\sum_{x \in U} N_x}{|X|}. \tag{4.31}$$

Considering the coherence membership functions defined by (4.30) in Theorem 4.5, we have

$$\mu_\zeta(x) = \sup_{i \in I} \left(\inf_{\gamma \in A_i} \mathscr{M}_\gamma(A_i^\tau(x)) \right) = \sup_{i \in I} \left(\mathscr{P}(A_i^\tau(x)) \right). \tag{4.32}$$

In virtue of (4.27), we have $A_i^\tau(x) = \bigcap_{m \in A_i} \{m\}^\tau(x)$. If for $m \in A_i$, $\{m\}^\tau(x)$ as events are pairwise independent, then $\mathscr{P}(A_i^\tau(x)) = \prod_{m \in A_i} \mathscr{P}(\{m\}^\tau(x))$. For any simple concept $m \in A_i$, by (4.31), we have the membership function of m formed as follows: for any $x \in X$, $\mu_m(x) = \mathscr{P}(\{m\}^\tau(x))$. Thus by (4.32), for $\eta = \prod_{m \in A_i} m$, i.e., the conjunction of the concepts $m \in A_i$, its membership function defined by (4.30) is degenerated to the fuzzy logic equipped by the product t-norm and max t-conorm as follows: for any $x \in X$,

$$\mu_\eta(x) = \mathscr{P}(A_i^\tau(x)) = \prod_{m \in A_i} \mathscr{P}(\{m\}^\tau(x)) = \prod_{m \in A_i} \mu_m(x). \tag{4.33}$$

This implies that the membership functions in the "conventional" fuzzy logic systems equipped by the product t-norm and max t-conorm are coherence membership functions when for all $m \in M$, $\{m\}^\tau(x)$ as events are pairwise independent in the probability space for every $x \in X$. This condition is too strict to be applied to the real world applications.

The measure $\mathscr{M}_m$ for the simple concept m defined by (4.31) just evaluates the occurring frequency of $\{m\}^\tau(x) \in \mathscr{F} \cap 2^X$ as an event, i.e., the random uncertainty. In fact, the degree of the set $\{m\}^\tau(x)$ supporting the claim: "*x belongs to m*" is determined by both the occurring frequency of the even $\{m\}^\tau(x)$ and the relationship of the elements in $\{m\}^\tau(x)$ with the semantics expressed by m. For instance, in Example 4.1, let $A = \{m_1, m_2\}$. The fuzzy concept $\eta = m_1 m_2$ states "*old and tall people*". The feature values, *age* and *height* are shown in Table 4.1 as follows:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
age	20	13	50	80	34	37	45	70	60	3
height	1.9	1.2	1.7	1.8	1.4	1.6	1.7	1.65	1.82	1.1

Given the AFS structure (M, τ, X) of Table 4.1 defined as (4.26), we have $A^\tau(x_7) = \{x_2, x_5, x_6, x_7, x_{10}\}$ and $A^\tau(x_8) = \{x_2, x_5, x_6, x_8, x_{10}\}$. From the measures for simple concepts m_1, m_2 defined by (4.31) under the assumption that each sample $x \in X$ has the same probability $\mathscr{P}(x)$ (i.e., the uniform distribution $\mathscr{P}$), one has

$$\mathscr{M}_{m_i}(A^\tau(x_7)) = \mathscr{P}(A^\tau(x_7)) = \mathscr{P}(A^\tau(x_8)) = \mathscr{M}_{m_i}(A^\tau(x_8)), \quad i = 1, 2$$

and the membership degrees of x_7 and x_8 belonging to $m_1 m_2$*: "old and tall people"* defined by (4.32) are equal. However, from our intuitive point of view, the degree of a person 70 years and 1.65 belonging to $m_1 m_2$*: "old and tall people"* should be larger than that of a person 40 years and 1.7, although 1.7 is a little higher than 1.65. Since the difference between ages of $x_8 \in A^\tau(x_8)$ and $x_7 \in A^\tau(x_7)$ is much greater than that of heights of them, hence the measures of simple concepts m_1, m_2 should satisfy

$$\mathscr{M}_{m_1}(A^\tau(x_8)) - \mathscr{M}_{m_1}(A^\tau(x_7)) > \mathscr{M}_{m_2}(A^\tau(x_8)) - \mathscr{M}_{m_2}(A^\tau(x_7)) > 0.$$

This implies that the measures for the simple concepts defined by the probability shown as (4.31) have not sufficiently considered the distributions of the feature values of the data, although $A^\tau(x) = \bigcap_{m \in A} \{m\}^\tau(x)$ is determined by both the semantic meanings of the simple concepts in A and the AFS structure (M, τ, X) of the dataset which is dominated by a certain probability distribution.

4.2.2 Coherence Membership Functions Based on Weight Functions and Probability Distributions

In this section, we propose the measures of simple concepts which are constructed according to both the semantics of the simple concepts and the probability distribution of the feature values of the data, i.e., the measure $\mathscr{M}_\gamma$ is induced by a function $\rho_\gamma : X \to [0, \infty)$ as will be clarified in the forthcoming definition.

Definition 4.8. ([35]) Let ν be a simple concept on X, $\rho_\nu : X \to R^+ = [0,\infty)$. ρ_ν is called a *weight function of the simple concept* ν if ρ_ν satisfies the following conditions:

1. $\rho_\nu(x) = 0 \Leftrightarrow (x,x) \notin R_\nu, x \in X$,
2. $\rho_\nu(x) \geq \rho_\nu(y) \Leftrightarrow (x,y) \in R_\nu, x,y \in X$,

where R_ν is the binary relation of the concept ν (refer to Definition 4.2).

Let ρ_γ be the weight function of simple concept $\gamma \in M$. In continuous case: Let X be a set, $X \subseteq R^n$. For each $\gamma \in M$, ρ_γ is integrable on X under Lebesgue measure with $0 < \int_X \rho d\mu < \infty$. S ($S \subseteq 2^X$) is the set of Borel sets in X. For all $U \in S$, we define a measure $\mathscr{M}_\gamma$ over S as follows:

$$\mathscr{M}_\gamma(U) = \frac{\int_U \rho_\gamma d\mathscr{P}}{\int_X \rho_\gamma d\mathscr{P}}. \tag{4.34}$$

In discrete case, the definition is formulated as follows. Let X be a finite set and $S \subseteq 2^X$ be a σ-algebra over X. For any $U \in S$, a measure $\mathscr{M}_\gamma$ over σ-algebra S is expressed in the form

$$\mathscr{M}_\gamma(U) = \frac{\sum_{x\in U} \rho_\gamma(x)\mathscr{P}(x)}{\sum_{x\in X} \rho_\gamma(x)\mathscr{P}(x)}. \tag{4.35}$$

It is clear that the measure $\mathscr{M}_\gamma$ defined by (4.34) and (4.35) satisfies Theorem 4.5. The weight function $\rho_\gamma(x)$ of a simple concept may have various interpretations depending on the specificity of the problem at hand. In general, $\rho_\gamma(x)$ weights degree of x supporting the claim that "*s is γ*" if the sample x is observed and the degree of x belonging to γ is less than or equal to that of s. For example, if γ: "*old people*" given the Table 4.1 as the observed data, the weight of the person x_8 who is 70 years to support the claim that "*s is an old person*" is larger than that of the person x_7 who is 40 years if the age of the person s is larger than both x_7 and x_8.

Following the line of the Singpurwalla's theory [44], the weight function $\rho_\gamma(x)$ is interpreted as $\mathscr{P}_{\mathscr{D}}(x \in \gamma)$ which is $\mathscr{D}$'s personal probability that x is classified in γ. Here we mainly apply the weight functions $\rho_\gamma(x)$, $\gamma \in M$ to reduce the influence of less essential samples and increase the influence of more essential ones when determining the membership functions with the use of (4.29) or (4.30). In other words, $\rho_\gamma(x)$, $\gamma \in M$ weight the referring value of every observed sample in X for the determining of the membership functions of fuzzy concepts in EM.

For example, if $\Omega \subseteq R^n$ and $m_{i1}, m_{i2}, m_{i3}, m_{i4} \in M, i = 1,2,...,n$, are the fuzzy concepts, *"small"* , *"medium"*, *"not medium"*, *"large"* associating to the feature f_i, respectively, then the weight functions of them can be defined according to the observed data $X = \{x_1, x_2, ..., x_h\} \subseteq \Omega$ and the semantic-oriented interpretations of the simple concepts in M as follows:

$$\rho_{m_{j1}}(x_i) = \frac{h_{j1} - f_j(x_i)}{h_{j1} - h_{j2}}, \tag{4.36}$$

$$\rho_{m_{j2}}(x_i) = \frac{h_{j4} - |f_j(x_i) - h_{j3}|}{h_{j4} - h_{j5}}, \tag{4.37}$$

$$\rho_{m_{j3}}(x_i) = \frac{|f_j(x_i) - h_{j3}| - h_{j5}}{h_{j4} - h_{j5}}, \tag{4.38}$$

$$\rho_{m_{j4}}(x_i) = \frac{f_j(x_i) - h_{j2}}{h_{j1} - h_{j2}}, \tag{4.39}$$

with the semantics of the terms of "*small*", "*medium*, "*not medium*", and "*large*", respectively, where $j = 1,2,...,n$,

$$\begin{aligned}
h_{j1} &= \max\{f_j(x_1), f_j(x_2), \cdots, f_j(x_h)\},\\
h_{j2} &= \min\{f_j(x_1), f_j(x_2), \cdots, f_j(x_h)\},\\
h_{j3} &= \frac{f_j(x_1) + f_j(x_2) + \cdots + f_j(x_h)}{h},\\
h_{j4} &= \max\{|f_j(x_k) - h_{j3}| \mid k = 1,2,\cdots,h\},\\
h_{j5} &= \min\{|f_j(x_k) - h_{j3}| \mid k = 1,2,\cdots,h\}.
\end{aligned}$$

By Definition 4.8 and the interpretations of $m_{i1}, m_{i2}, m_{i3}, m_{i4}$, one can verify that $\rho_{m_{ik}}$, $j = 1,2,...,n$ are the weight functions of simple concept $m_{ik} \in M$. In general, the weight function ρ_ν of a simple concept ν associating to a feature f_i is subjectively defined by users according to the data distribution on the feature f_i and the semantical interpretation of the simple concept ν. It is obvious that for a given simple concept γ we can define many different functions $\rho_\gamma : X \to [0, +\infty)$ such that satisfies the weight function conditions shown in Definition 4.8. The diversity of the weight functions of a simple concept results from the subjective imprecision of human perception of the observed data. However, his diversity is bounded or constrained by the sub-preference relation of the individual simple concept defined by Definition 4.3 according to the semantics of the natural language. This is rooted in the fact that perceptions are intrinsically imprecise, reflecting the bounded ability of sensory organs. In order to provide a tool for representing and managing an infinitely complex reality, the weight functions for simple concepts are mental constructs with the subjective imprecision (i.e., subjectively constructing the functions satisfying Definition 4.8 for the concerned simple concepts). But the constructs of the weight function for an individual simple concept γ have to observe the objectivity in nature, which is the sub-preference relation R_γ (refer to Definition 4.3) objectively determined by the observed data X and the semantics of γ. In other words, the subjective imprecisions of the weight function of γ are constrained by the objectivity of R_γ. The multi-options of the weight functions just reflect the subjective imprecisions of the perceptions of the observed data. In what follows, we construct the coherence membership functions using the weight functions of simple concepts according to the probability distribution of the data.

Theorem 4.6. *Let $(\Omega, \mathscr{F}, \mathscr{P})$ be a probability measure space and M be the set of some simple concepts on Ω. Let ρ_γ be the weight function of the simple concept $\gamma \in M$ (refer to Definition 4.8). $X \subseteq \Omega$, X is a finite set of the observed samples drawn from the probability space $(\Omega, \mathscr{F}, \mathscr{P})$. Let (M, τ, Ω) and $(M, \tau|_X, X)$ be the*

AFS structures defined as (4.26). If for any $m \in M$ and any $x \in \Omega$, $\{m\}^{\tau}(x) \in \mathscr{F}$, then the following assertions hold:

1. $\{\mu_{\zeta}(x) \mid \zeta \in EM\}$ is a set of coherence membership functions of $(EM, \vee, \wedge, ')$ and (M, τ, Ω), $(M, \tau|_X, X)$, provided the membership functions of each concept $\zeta = \sum_{i \in I}(\prod_{m \in A_i} m) \in EM$ defined as follows: for any $x \in X$

$$\mu_{\xi}(x) = \sup_{i \in I} \prod_{\gamma \in A_i} \frac{\sum_{u \in A_i^{\tau}(x)} \rho_{\gamma}(u) N_u}{\sum_{u \in X} \rho_{\gamma}(u) N_u}, \quad \forall x \in X, \tag{4.40}$$

$$\mu_{\xi}(x) = \sup_{i \in I} \prod_{\gamma \in A_i} \frac{\int_{A_i^{\tau}(x)} \rho_{\gamma}(t) d\mathscr{P}(t)}{\int_{\Omega} \rho_{\gamma}(t) d\mathscr{P}(t)}, \quad \forall x \in \Omega, \tag{4.41}$$

or

$$\mu_{\xi}(x) = \sup_{i \in I} \inf_{\gamma \in A_i} \frac{\sum_{u \in A_i^{\tau}(x)} \rho_{\gamma}(u) N_u}{\sum_{u \in X} \rho_{\gamma}(u) N_u}, \quad \forall x \in X, \tag{4.42}$$

$$\mu_{\xi}(x) = \sup_{i \in I} \inf_{\gamma \in A_i} \frac{\int_{A_i^{\tau}(x)} \rho_{\gamma}(t) d\mathscr{P}(t)}{\int_{\Omega} \rho_{\gamma}(t) d\mathscr{P}(t)}, \quad \forall x \in \Omega, \tag{4.43}$$

where N_u is the number of times that $u \in X$ is observed as a sample.

2. The membership function defined by (4.40) or (4.42) converges to the membership function defined by (4.41) or (4.43) respectively for all $x \in \Omega$ as $|X|$ approaches to infinity, provided that for every $\gamma \in M$ $\rho_{\gamma}(x)$ is continuous on Ω and X is a set of samples randomly drawn from the probability space $(\Omega, \mathscr{F}, \mathscr{P})$.

Proof. 1. It can be directly proved by Theorem 4.5.
2. Let $p(x)$ be the density function of the probability space $(\Omega, \mathscr{F}, \mathscr{P})$. Since X is a set of samples randomly drawn from $(\Omega, \mathscr{F}, \mathscr{P})$. Hence by formulas (1.5) and (1.6) for any $x \in X$, we have

$$p(x) = \lim_{|X| \to \infty, S(\Delta_x) \to 0} \frac{|X_{\Delta_x}|}{|X| S(\Delta_x)}. \tag{4.44}$$

Here $x \in \Delta_x \subseteq \Omega$, $S(\Delta_x)$ is the size of the small space Δ_x, X_{Δ_x} is the set of the drawn samples in X falling into Δ_x in which a sample is regarded as n different samples if it is observed n times.

For any $i \in I, \gamma \in A_i$ in (4.40) and (4.42), $x \in X$, assume that Ω is divided into q small subspaces $\Delta_j \in \mathscr{F}, j = 1, ..., q$ such that for any j either $\Delta_j \subseteq A_i^{\tau}(x)$ or $\Delta_j \cap A_i^{\tau}(x) = \varnothing$. Let $\mathscr{I}_{A_i^{\tau}(x)} = \{\Delta_j \mid \Delta_j \subseteq A_i^{\tau}(x), j = 1, 2, ..., q\}$. Let Δ_{max} be the maximum size of $S(\Delta_j), j = 1, 2, ..., q$ and Δ_u be the small space Δ_j such that $u \in \Delta_j$. In virtue of (4.44), we have

$$\begin{aligned}
&\lim_{|X|\to\infty} \frac{\sum_{u\in A_i^\tau(x)} \rho_\gamma(u) N_u}{\sum_{u\in X} \rho_\gamma(u) N_u} \\
&= \lim_{|X|\to\infty,\Delta_{max}\to 0} \frac{\sum_{u\in X_\Delta,\Delta\in\mathscr{I}_{A_i^\tau(x)}} \rho_\gamma(u)}{\sum_{u\in X_{\Delta_j},1\le j\le q} \rho_\gamma(u)} \\
&= \lim_{|X|\to\infty,\Delta_{max}\to 0} \frac{\sum_{\Delta_u\in\mathscr{I}_{A_i^\tau(x)}} \rho_\gamma(u)|X_{\Delta_u}|}{\sum_{\Delta_u\in\{\Delta_j|,1\le j\le q\}} \rho_\gamma(u)|X_{\Delta_u}|} \\
&= \lim_{|X|\to\infty,\Delta_{max}\to 0} \frac{\sum_{\Delta_u\in\mathscr{I}_{A_i^\tau(x)}} \rho_\gamma(u)\frac{|X_{\Delta_u}|}{|X|S(\Delta_u)}S(\Delta_u)}{\sum_{\Delta_u\in\{\Delta_j|,1\le j\le q\}} \rho_\gamma(u)\frac{|X_{\Delta_u}|}{|X|S(\Delta_u)}S(\Delta_u)} \\
&= \frac{\int_{A_i^\tau(x)} \rho_\gamma(t)d\mathscr{P}(t)}{\int_\Omega \rho_\gamma(t)d\mathscr{P}(t)}. \quad (in\ virtue\ of\ (4.44))
\end{aligned}$$

Therefore the membership function defined by (4.40) or (4.42) converges to that defined by (4.41) or (4.43), respectively for all $x \in \Omega$ as $|X|$ approaches infinity. □

Theorem 4.6 defines the membership functions and their fuzzy logic operations on the observed data and the whole space by taking both the fuzziness (subjective imprecision: the uncertainty of $\rho_\gamma(x)$ due to the individual different interpretations of the simple concepts) and the randomness (objective uncertainty: the uncertainty of $A_i^\tau(x)$, N_x due to randomly observe of the samples) into account. Since the lattice $(EM, \vee, \wedge, ')$ is closed under the AFS fuzzy logic operations $\vee$, $\wedge$, $'$, hence Theorem 4.6 also gives the membership functions of all fuzzy logic operations of fuzzy concepts in EM. According to the observed data or the probability distribution of the space, (1) of Theorem 4.6 provides a very applicable and simple method to construct coherence membership functions by the weight functions of the simple concepts which can be flexibly and expediently defined to represent the individual perceptions. The following practical aspects of the applications of AFS and probability framework to the real world can be ensured by (2) of Theorem 4.6 for a large sample set.

- The membership functions and the fuzzy logic operations determined by the observed data drawn from a probability space (i.e. defined by (4.40) or (4.42)) will be consistent with ones determined by the probability distribution (i.e., defined by (4.41) or (4.43)).
- The results via the AFS fuzzy logic based on the membership functions and their logic operations determined by different data sets drawn from the same probability space (i.e. defined by (4.40) or (4.42)) will be stable and consistent.
- The laws discovered based on the membership functions and their logic operations determined by the observed data drawn from a probability space (i.e. defined by (4.40) or (4.42)) can be applied to the whole space via the membership functions of the concepts determined by the probability distribution (i.e., defined by (4.41) or (4.43)).

Thus uncertainty of randomness and of imprecision can be treated in a unified and coherent manner under the AFS and probability framework and it offers a new avenue to explore the statistical properties of fuzzy set theory and to a major enlargement of the role of natural languages in probability theory. In the last section of this chapter, we will test these via the experimental studies completed for a well-known the Iris data set.

Concerning the applicability, it is clear that the membership functions and their fuzzy logic operations can be easily obtained according to the finite observed data set X by exploiting (4.40) or (4.42). However those realized according to the probability distribution on the whole space Ω by (4.41) or (4.43) involving the computations of high dimension integral may be too complicated. The essential advantage of the AFS and probability framework is in that what is easily discovered by the simple computations of the membership functions determined by the finite data X observed from a system can be applied to predict and describe the behavior of the system on the whole apace Ω by the computations of the integral for the continuous functions $\rho_\gamma(x)$. In order to make the framework more applicable, in Section 4.4, the coherence membership functions based on Gaussion weight functions and multi-normal distributions has been exhaustively discussed. Concerning the theory, in the following section, we fit AFS and probability framework into Singpurwalla's theory [44] to make its theoretical foundation more stable.

Finally, we show some links between the AFS approach and Lawry's Label Semantics which also defines fuzziness in terms of a probability measure. Label semantics [20] is a framework for linguistic reasoning based on a random set model that uses degrees of appropriateness of a label to describe a given example. In such systems, fuzzy labels provide a high-level mechanism of discretization and interpretation of modelling uncertainty. In label semantics, labels are assumed to be chosen from a finite predefined set of labels and the set of appropriate labels for a value is defined as a random set-valued function from a *population of individuals* into the set of subsets of labels which are the labels the population of individuals consider appropriate to describe the value. Furthermore, appropriateness degrees of a value belonging to a label is defined according to the mass assignment on labels.

In AFS theory, currently, we just study how to determine the membership function of a concept based on the data drawn from one probability space which can be regards as "*one of a population of individuals' description of the value*". A probability space can be regarded as an individual knowledge. The membership functions of a concept based on the data drawn from some different probability spaces which can be regarded as "*a population of individuals' description of the value* has remained as an open problem. Different probability spaces can be regarded as different individuals's knowledge. Thus the AFS theory can be expanded under the framework of label semantics and the label semantics may be explored in virtue of the AFS theory.

4.3 Coherence Membership Functions and the Probability Measures of Fuzzy Events

In this section, we discuss the coherence membership functions defined by Theorem 4.6 under the probability measures of fuzzy events proposed and developed by Zadeh, Singpurwalla and Booker [44, 52] and fit AFS and probability framework into the Singpurwalla's theory.

4.3.1 *The Probability Measures of Fuzzy Events*

Zadehs article titled "Probability Measures of Fuzzy Events" [52] suggested how to expand the scope of applicability of probability theory to include fuzzy sets. His construction proceeds along the following line. Let $(\Omega, \mathscr{F}, \mathscr{P})$ be a "probability measure space" . Recall that x, an outcome of ε, is a member of A, and assume for now that A is a countable Boolean set, where $A \in \mathscr{F}$, and let $I_A(x)$ be the characteristic function of A, i.e., $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ otherwise. Then it is easy to see that

$$\mathscr{P}(A) = \sum_{x \in \Omega} I_A(x) \mathscr{P}(x) \tag{4.45}$$

where $\mathscr{P}(x)$ is the probability of x. An analog of the foregoing result when A is not countable is a relationship of the form

$$\mathscr{P}(A) = \int_{\Omega} I_A(x) d\mathscr{P}(x)$$

Motivated by this (well-known) result, Zadeh has declared that the probability measure of a fuzzy subset A of Ω, which he calls a fuzzy event, is

$$\Pi(A) = \int_{\Omega} \mu_A(x) d(\mathscr{P}(x)) = E(\mu_A(x)) \tag{4.46}$$

where $\mu_A(x)$ is the membership function of A and E denotes expectation. The point to be emphasized here is that the expectation is taken with respect to the initial probability measure $\mathscr{P}$ that has been defined on the (Boolean) sets of Ω. Having defined $\Pi(A)$ as before, Zadeh proceeded to show that

$$A \subseteq B \Rightarrow \Pi(A) \leq \Pi(B) \tag{4.47}$$

$$\Pi(A \cup B) = \Pi(A) + \Pi(B) - \Pi(A \cap B) \tag{4.48}$$

$$\Pi(A + B) = \Pi(A) + \Pi(B) - \Pi(A \bullet B) \tag{4.49}$$

where $A \bullet B$ is the product (not the intersection) of A and B. Finally, A and B are declared to be independent if

$$\Pi(A \bullet B) = \Pi(A) \cdot \Pi(B),$$

and the conditional probability of A, were B to occur, denoted by $\Pi(A|B)$, is defined as

$$\Pi(A|B) = \frac{\Pi(A \bullet B)}{\Pi(B)}$$

Thus when A and B are independent,

$$\Pi(A|B) = \Pi(A)$$

Whereas the definition (4.45) has the virtue that when A is a Boolean set, $\Pi(A) = \mathscr{P}(A)$, so that the measure Π can be seen as a generalization of the measure $\mathscr{P}$, the question still remains as to whether Π is a probability measure. Properties (4.47), (4.48), and (4.49) seem to suggest that Π could indeed be viewed as a probability measure. But [44] shown that such a conclusion would be premature by the following arguments:

a) With property (4.48), the evaluation of $\Pi(A)$ and $\Pi(B)$ is enough to evaluate $\Pi(A \cap B)$, whereas with probability, the evaluation of $\mathscr{P}(A)$ and $\mathscr{P}(B)$ is not sufficient to evaluate $\mathscr{P}(A \cap B)$, unless A and B are independent. Note that because

$$\Pi(A \cap B) = E(\mu_{A \cap B}(x)) = E(\min\{\mu_A(x), \mu_B(x)\}),$$

it can be easily seen that for $\mathscr{A} = \{x \mid \mu_A(x) \leq \mu_B(x)\}$

$$\Pi(A \cup B) = \int_{x \in \mathscr{A}} \mu_B d\mathscr{P}(x) + \int_{x \in \mathscr{A}^c} \mu_A d\mathscr{P}(x)$$

b) Property (4.49) has no "analog" in probability theory, because the notions of $(A + B)$ and $(A \bullet B)$ are not part of classical set theory. More importantly, conditional probability has only been defined in terms of $(A \bullet B)$.

We agree with Singpurwalla and Booker's view that while [52] attempted at making fuzzy set theory and probability theory work together, there are some interesting points to be pursued further. In what follows, we present Singpurwalla's line of argument that is able to achieve Zadeh's goal of forming constructs such as "probability measures of fuzzy events" provided that the membership functions are predetermined. Let

$$\mathscr{P}_{\mathscr{D}}(A) = \mathscr{P}_{\mathscr{D}}(X \in A).$$

Here the generic X denotes the uncertain outcome of an experiment ε and the subscript $\mathscr{D}$ denotes the fact that what is being assessed is $\mathscr{D}$'s personal probability, that is $\mathscr{D}$'s willingness to bet. To incorporate the role of membership functions in the assessment of a probability measure of a fuzzy set A, Singpurwalla and Booker introduced a new component into the analysis – namely an expert, say $\mathscr{Z}$ (in honor of Zadeh), whose expertise lies in specifying a membership function $\mu_A(x)$ for all $x \in \Omega$, and a fuzzy set A. Singpurwalla and Booker assume that $\mathscr{D}$ has no access to any membership function of A or a membership function $\mu_A(x)$ is given by $\mathscr{Z}$. With the fuzzy set A entering the picture, $\mathscr{D}$ is confronted with both the imprecision and the uncertainty, i.e., about the membership of x in A and the other about the outcome

$X = x$. If A is a Boolean set (as is normally the case in standard probability theory), then $\mathscr{D}$ would be confronted with only the uncertainty, namely the uncertainty that $X = x$. As a subjectivist, $\mathscr{D}$ views the imprecision as simply another uncertainty, and to $\mathscr{D}$ all uncertainties can be quantified only by probability. Thus $\mathscr{D}$ specifies two probabilities:

a) $\mathscr{P}_{\mathscr{D}}(x)$, which is $\mathscr{D}$'s prior probability that an outcome of ε will be x, and
b) $\mathscr{P}_{\mathscr{D}}(x \in A)$, which is $\mathscr{D}$'s prior probability that an outcome x belongs to A.

Whereas the specification of $\mathscr{P}_{\mathscr{D}}(x)$ is an operation in standard probability theory, the assessment of $\mathscr{P}_{\mathscr{D}}(x \in A)$ raises an issue. Specifically, because $\mathscr{P}_{\mathscr{D}}(x \in A)$ is $\mathscr{D}$'s personal probability that x is classified in A.

Once $\mathscr{P}_{\mathscr{D}}(x)$ and $\mathscr{P}_{\mathscr{D}}(x \in A)$ have been specified by $\mathscr{D}$ for all $x \in \Omega$, $\mathscr{D}$ will use the law of total probability to write

$$\begin{aligned} \mathscr{P}_{\mathscr{D}}(A) &= \mathscr{P}_{\mathscr{D}}(X \in A) \\ &= \sum_{x \in \Omega} \mathscr{P}_{\mathscr{D}}(X \in A | X = x) \mathscr{P}_{\mathscr{D}}(x) \\ &= \sum_{x \in \Omega} \mathscr{P}_{\mathscr{D}}(x \in A) \mathscr{P}_{\mathscr{D}}(x) \end{aligned} \tag{4.50}$$

An analog of the foregoing result when A is not countable is a relationship of the form

$$\mathscr{P}_{\mathscr{D}}(A) = \int_{x \in \Omega} \mathscr{P}_{\mathscr{D}}(x \in A) d\mathscr{P}_{\mathscr{D}}(x) \tag{4.51}$$

which is the expected value of $\mathscr{D}$'s classification probability with respect to $\mathscr{D}$'s prior probability of X. Thus Singpurwalla and Booker gave a probability measure for a fuzzy set A that can be justified on the basis of personal (i.e., subjective) probabilities and the notion that probability is a reflection of one's partial knowledge about an event of interest in Equation (4.50) which is based on $\mathscr{D}$'s inputs alone.

No matter how to interpret $\mathscr{P}_{\mathscr{D}}(x \in \xi)$ (In AFS theory, we interpret it as $\rho(x)$ — the weight function of the simple concept ξ, refer to Definition 4.8), an assessment of this quantity is essential for developing a normative approach for assessing probability measures of fuzzy sets. In introducing $\mathscr{P}_{\mathscr{D}}(x \in A)$, Singpurwalla and Booker have in fact reaffirmed Lindley's claim that probability is able to handle any situation that fuzzy logic can [19]. But the weight functions of simple concepts expressed by Definition 4.8, which are determined by the semantics of the fuzzy concepts and the human comparisons of the associating feature values, are mathematical description of the imprecise perceptions and are different from random uncertainty in probability. So that probability itself is not able to handle the AFS fuzzy logic in AFS and probability framework (see Theorem 4.6).

The authors in [44] discussed the sensible connection between membership functions and probability, and this connection is an important contribution to a better understanding of the probability measure of the fuzzy events whose membership functions are predetermined. However, Singpurwalla and Booker have not touched the problem of how to determine membership functions for fuzzy sets based on the theory they developed. In what follows, we discuss the coherence membership functions defined by Theorem 4.6 under Singpurwalla & Booker's probability measure of fuzzy events.

4.3.2 Coherence Membership Functions Based on Probability Measure of the Fuzzy Events

In this section, we apply the probability measure of the fuzzy set [44] to induce the measures of simple concepts for the coherence membership functions in Theorem 4.6.

Given the considerations presented in Section 4.3.1, we know that $\mathscr{P}_{\mathscr{D}}(x \in \gamma)$ (refer to (4.50)) is $\mathscr{D}$'s personal probability that x is classified in γ. Thus $\mathscr{P}_{\mathscr{D}}(x \in \gamma)$ is merely a reflection of $\mathscr{D}$'s imprecision (or partial knowledge) of the boundaries of a concept. For a simple concept γ such as *"small"* , *"medium", "not medium", "large"*, by Definition 4.8, we know that $\rho_\gamma(x) = \mathscr{P}_{\mathscr{D}}(x \in \gamma), \forall x \in X$, is a weight function of γ. Because if $(x,y) \in R_\gamma$, i.e., the degree of x belonging to γ is greater than or equal to that of y, then $\mathscr{P}_{\mathscr{D}}(x \in \gamma) \geq \mathscr{P}_{\mathscr{D}}(y \in \gamma)$ and if $(x,x) \notin R_\gamma$, i.e., x does not belong to γ at all, then $\mathscr{P}_{\mathscr{D}}(x \in \gamma) = 0$. In other world, we also can regard $\mathscr{P}_{\mathscr{D}}(x \in \gamma)$ in (4.50) as an interpretation of weight function $\rho_\gamma(x)$ defined by Definition 4.8. For each simple concept $\gamma \in M$, let

$$\mathscr{P}_{\mathscr{D}}(x \in \gamma) = \rho_\gamma(x),\ \forall x \in X.$$

Let N_x be the number of times x is observed as a sample and Ω be a discrete (i.e., countable) set. Then $\frac{N_x}{|X|} \to \mathscr{P}(x)$ as the set X approaching to Ω. Thus the probability measure of fuzzy simple concept γ defined by (4.50) is expressed as

$$\begin{aligned}
\mathscr{P}_{\mathscr{D}}(\gamma) &= \sum_{x \in X} \mathscr{P}_{\mathscr{D}}(x \in \gamma)\mathscr{P}_{\mathscr{D}}(x) \\
&= \sum_{x \in X} \rho_\gamma(x)\mathscr{P}_{\mathscr{D}}(x) \\
&\approx \frac{1}{|X|}\sum_{x \in X} \rho_\gamma(x)N_x \ \ (because\ \ of\ \mathscr{P}_{\mathscr{D}}(x) \approx \frac{N_x}{|X|})
\end{aligned} \tag{4.52}$$

In (4.52), the probability measure of fuzzy event γ takes each $x \in X$ in account. Thus it is natural to define the probability measure of γ take $x \in W \subseteq X$ in account and call it the *probability measure of fuzzy simple concept γ on W* as follows:

$$\begin{aligned}
\mathscr{P}_{\mathscr{D}}(\gamma : W) &= \sum_{x \in W} \mathscr{P}_{\mathscr{D}}(x \in \gamma)\mathscr{P}_{\mathscr{D}}(x) \\
&= \sum_{x \in W} \rho_\gamma(x)\mathscr{P}_{\mathscr{D}}(x) \\
&\approx \frac{1}{|X|}\sum_{x \in W} \rho_\gamma(x)N_x
\end{aligned} \tag{4.53}$$

When W is not countable, the probability measure of fuzzy simple concept γ on $W \subseteq \Omega$ is defined as follows:

$$\mathscr{P}_{\mathscr{D}}(\gamma : W) = \int_W \rho_\gamma(t) d\mathscr{P}_{\mathscr{D}}(t) \tag{4.54}$$

In Theorem 4.6, for $A^\tau(x) \in S$ and $\gamma \in M$, $\mathscr{M}_\gamma(A^\tau(x))$ measures the degree of set $A^\tau(x)$ supporting the claim: "*x belongs to* γ". We can construct $\mathscr{M}_\gamma$ according to the probability measure of fuzzy simple concept defined by [44] (refer to (4.53), (4.54)) as follows: for $U \in S$

$$\mathscr{M}_\gamma(U) = \frac{\mathscr{P}_{\mathscr{D}}(\gamma : U)}{\mathscr{P}_{\mathscr{D}}(\gamma : X)} \approx \frac{\sum_{x \in U} \rho_\gamma(x) N_x}{\sum_{x \in X} \rho_\gamma(x) N_x}, \quad if\ U\ and\ X\ is\ countable. \tag{4.55}$$

$$\mathscr{M}_\gamma(U) = \frac{\mathscr{P}_{\mathscr{D}}(\gamma : U)}{\mathscr{P}_{\mathscr{D}}(\gamma : X)} = \frac{\int_U \rho_\gamma(t) d\mathscr{P}_{\mathscr{D}}(t)}{\int_\Omega \rho_\gamma(t) d\mathscr{P}_{\mathscr{D}}(t)}, \quad if\ U\ and\ \Omega\ is\ not\ countable. \tag{4.56}$$

Applying the measures in (4.55) and (4.56) to Theorem 4.5, we get the coherence membership functions defined by Theorem 4.6. Thus the AFS and probability framework fits into Singpurwalla's probability measure of fuzzy events.

4.4 Coherence Membership Functions of Multi-normal Distributions and Gaussion Weights

In this section, in order to make AFS and probability framework more applicable, we study the coherence membership functions based on Gaussion weight functions of the simple concepts and the multi-normal probability space. Let $\Omega = R^n$ and $\mathscr{F}$ be all set of Borel sets in R^n and the probability distribution $\mathscr{P}$ in the probability space $(\Omega, \mathscr{F}, \mathscr{P})$ is a m-normal distribution with the density function shown as follows:

$$p(x) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{\sqrt{(2\pi)^n |\Sigma_i|}} e^{-\frac{1}{2}(x-\mu_i)' \Sigma_i^{-1} (x-\mu_i)}, \tag{4.57}$$

where $x \in R^n$, Σ_i is a $n \times n$ symmetry positive definite matrix and μ_i is a n-dimension vector in R^n, $i = 1, 2, ..., m$. Let M be the set of simple concepts on Ω. $\gamma \in M$ is a simple concept with the semantic meaning: "*near to* c_γ" and γ' is a simple concept with the semantic meaning: "*not near to* c_γ", where $c_\gamma \in R^n$. For each simple concept $\gamma \in M$, $(x,y) \in R_\gamma \Leftrightarrow \|x - c_\gamma\| \leq \|y - c_\gamma\|$; $(x,y) \in R_{\gamma'} \Leftrightarrow \|x - c_\gamma\| \geq \|y - c_\gamma\|$ where R_γ,$R_{\gamma'}$ defined by Definition 4.2 is the binary relation of the concepts γ, γ'. By Definition 4.3, one can verify that for any $\gamma \in M$, γ, γ' are simple concepts. (M, τ, Ω) is an AFS structure in which τ is defined as (4.26). For a simple concept $\gamma \in M$, the weight functions are defined as follows: for any $x \in \Omega$,

$$\rho_\gamma(x) = e^{-(x-c_\gamma)' \Delta_\gamma (x-c_\gamma)}, \tag{4.58}$$

$$\rho_{\gamma'}(x) = 1 - \rho_\gamma(x), \tag{4.59}$$

where $\Delta_\gamma, \gamma \in M$ are semi-positive definite symmetry matrices.

Proposition 4.4. *Let $\Omega = R^n$ and $(\Omega, \mathscr{F}, \mathscr{P})$ be a m-normal distribution with the density function shown as (4.57). Let M be a set of simple concepts on Ω whose weight functions are defined as (4.58), (4.59). Then for any fuzzy concept $\xi = \sum_{i\in I}(\prod_{m\in A_i} m) \in EM$, the coherence membership function of ξ on Ω defined by (4.41) or (4.43) is formulated as follows: $\forall x \in \Omega$,*

$$\mu_\xi(x) = \begin{cases} \sup\limits_{i\in I} \prod\limits_{\zeta\in A_i} \Upsilon_\zeta, & for\ (4.41), \\ \sup\limits_{i\in I} \inf\limits_{\zeta\in A_i} \Upsilon_\zeta, & for\ (4.43), \end{cases} \tag{4.60}$$

Where

$$\Upsilon_\zeta = \begin{cases} \dfrac{\sum_{k=1}^{m} \frac{e^{\Gamma_{k\gamma}}}{\sqrt{\pi^n|\Sigma_k|}} \int\limits_{A_i^\tau(\{x\})} e^{-(t-\Xi_{k\gamma})'\Theta_{k\gamma}(t-\Xi_{k\gamma})}dt}{\sum_{k=1}^{m} \frac{e^{\Gamma_{k\gamma}}|\Theta_{k\gamma}|^{-\frac{1}{2}}}{\sqrt{|\Sigma_k|}}}, & \zeta = \gamma, \\ \dfrac{\mathscr{P}(A_i^\tau(\{x\})) - \frac{1}{m}\sum_{k=1}^{m} \frac{e^{\Gamma_{k\gamma}}}{\sqrt{(2\pi)^n|\Sigma_k|}} \int\limits_{A_i^\tau(\{x\})} e^{-(t-\Xi_{k\gamma})'\Theta_{k\gamma}(t-\Xi_{k\gamma})}dt}{1 - \frac{1}{m}\sum_{k=1}^{m} \frac{e^{\Gamma_{k\gamma}}|\Theta_{k\gamma}|^{-\frac{1}{2}}}{\sqrt{(2)^n|\Sigma_k|}}}, & \zeta = \gamma', \end{cases} \tag{4.61}$$

for $1 \le k \le m$, $\gamma \in M$,

$$\Xi_{k\gamma} = (\Delta_\gamma + \frac{1}{2}\Sigma_k^{-1})^{-1}(\Delta_\gamma c_\gamma + \frac{1}{2}\Sigma_k^{-1}\mu_k), \tag{4.62}$$

$$\Theta_{k\gamma} = \Delta_\gamma + \frac{1}{2}\Sigma_k^{-1}, \tag{4.63}$$

$$\begin{aligned} \Gamma_{k\gamma} = {} & c_\gamma'\Delta_\gamma c_\gamma + \frac{1}{2}\mu_k'\Sigma_k^{-1}\mu_k \\ & -(\Delta_\gamma c_\gamma + \frac{1}{2}\Sigma_k^{-1}\mu_k)'(\Delta_\gamma + \frac{1}{2}\Sigma_k^{-1})^{-1}(\Delta_\gamma c_\gamma + \frac{1}{2}\Sigma_k^{-1}\mu_k). \end{aligned} \tag{4.64}$$

Proof. Considering the membership functions defined by (4.41) and (4.43), we know that the key is to study the following integral for simple concept $\gamma \in M$ and $\Phi \subseteq \Omega$,

$$\begin{aligned} & \int\limits_\Phi \rho_\gamma(t)d\mathscr{P}(t) \\ & = \frac{1}{m}\sum_{k=1}^{m} \frac{1}{\sqrt{(2\pi)^n|\Sigma_k|}} \int\limits_\Phi e^{-(t-c_\gamma)'\Delta_\gamma(t-c_\gamma)-\frac{1}{2}(t-\mu_k)'\Sigma_k^{-1}(t-\mu_k)}dt \end{aligned} \tag{4.65}$$

Notice that for any t, any $\gamma \in M$, $(t-c_\gamma)'\Delta_\gamma(t-c_\gamma) \ge 0$ and $\frac{1}{2}(t-\mu_k)'\Sigma_k^{-1}(t-\mu_k) \ge 0$. Let us study

$$\begin{aligned}
&(t-c_\gamma)'\Delta_\gamma(t-c_\gamma)+\frac{1}{2}(t-\mu_k)'\Sigma_k^{-1}(t-\mu_k)\\
&=t'\Delta_\gamma t-c_\gamma'\Delta_k t-t'\Delta_\gamma c_\gamma+c_\gamma'\Delta_\gamma c_\gamma+\frac{1}{2}(t'\Sigma_k^{-1}t-\mu_k'\Sigma_k^{-1}t-t'\Sigma_k^{-1}\mu_k+\mu_k'\Sigma_k^{-1}\mu_k)\\
&=\left(t-(\Delta_\gamma+\frac{1}{2}\Sigma_k^{-1})^{-1}(\Delta_\gamma c_\gamma+\frac{1}{2}\Sigma_k^{-1}\mu_k)\right)'\\
&\left(\Delta_\gamma+\frac{1}{2}\Sigma_k^{-1}\right)\left(t-(\Delta_\gamma+\frac{1}{2}\Sigma_k^{-1})^{-1}(\Delta_\gamma c_\gamma+\frac{1}{2}\Sigma_k^{-1}\mu_k)\right)\\
&+c_\gamma'\Delta_\gamma c_\gamma+\frac{1}{2}\mu_k'\Sigma_k^{-1}\mu_k-(\Delta_\gamma c_\gamma+\frac{1}{2}\Sigma_k^{-1}\mu_k)'(\Delta_\gamma+\frac{1}{2}\Sigma_k^{-1})^{-1}(\Delta_\gamma c_\gamma+\frac{1}{2}\Sigma_k^{-1}\mu_k)
\end{aligned}$$

Thus by (4.62), (4.63),(4.64) and the above equation, formula (4.65) can be simplified as follows:

$$\int\limits_{\Phi}\rho_\gamma(t)d\mathscr{P}(t)=\frac{1}{m}\sum_{k=1}^{m}\frac{e^{\Gamma_{k\gamma}}}{\sqrt{(2\pi)^n|\Sigma_k|}}\int\limits_{\Phi}e^{-(t-\Xi_{k\gamma})'\Theta_{k\gamma}(t-\Xi_{k\gamma})}dt \tag{4.66}$$

Since Δ_γ is semi-positive definite and Σ_k is positive definite, hence $\Theta_{k\gamma}=\Delta_\gamma+\frac{1}{2}\Sigma_k^{-1}$ is positive definite. By the properties of density function of the normal distribution [3], for any $\mu\in R^n$ and any positive definite matrix $\Sigma\in R^{n\times n}$,

$$\int_\Omega e^{-(t-\mu)'\Sigma^{-1}(t-\mu)}dt=\pi^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}. \tag{4.67}$$

Therefore we have

$$\int_\Omega e^{-(t-\Xi_{k\gamma})'\Theta_{k\gamma}(t-\Xi_{k\gamma})}dt=\pi^{\frac{n}{2}}|\Theta_{k\gamma}|^{-\frac{1}{2}}. \tag{4.68}$$

Then by (4.66) and (4.68), we have

$$\int\limits_{\Omega}\rho_\gamma(t)d\mathscr{P}(t)=\frac{1}{m}\sum_{k=1}^{m}\frac{e^{\Gamma_{k\gamma}}|\Theta_{k\gamma}|^{-\frac{1}{2}}}{\sqrt{(2)^n|\Sigma_k|}} \tag{4.69}$$

Thus by (4.66), (4.69) and (4.41),(4.43), we have the upper one of (4.61). When $\zeta=\gamma'$, refer to (4.41),(4.43) we have

$$\begin{aligned}
\frac{\int\limits_{A_i^\tau(\{x\})}\rho_\zeta(t)d\mathscr{P}(t)}{\int\limits_{\Omega}\rho_\zeta(t)d\mathscr{P}(t)}&=\frac{\int\limits_{A_i^\tau(\{x\})}(1-\rho_\gamma(t))d\mathscr{P}(t)}{\int\limits_{\Omega}(1-\rho_\gamma(t))d\mathscr{P}(t)}\\
&=\frac{\int\limits_{A_i^\tau(\{x\})}d\mathscr{P}(t)-\int\limits_{A_i^\tau(\{x\})}\rho_\gamma(t)d\mathscr{P}(t)}{1-\int\limits_{\Omega}\rho_\gamma(t)d\mathscr{P}(t)}
\end{aligned} \tag{4.70}$$

Then by (4.66) and (4.69), we have the lower one of (4.61). Therefore, for any fuzzy concept $\xi = \sum_{i\in I}(\prod_{m\in A_i} m) \in EM$, the membership function of ξ on Ω defined by (4.41) and (4.43) is formulated as (4.60). □

In some real world applications, a simple concept often associates with a single feature. For instance, in Example 4.1, each simple concept in M associates with a single feature. Thus the weight functions defined by (4.58) and (4.59) can be simplified as follows if every $\gamma \in M$ associates with a single feature. For any $x = (x_1, x_2, ..., x_n)' \in \Omega$,

$$\rho_\gamma(x) = e^{-d_\gamma(x_i - c_\gamma)^2},\ c_\gamma, d_\gamma \in R, d_\gamma > 0, 1 \le i \le n; \tag{4.71}$$

$$\rho_{\gamma'}(x) = 1 - \rho_\gamma(x), \tag{4.72}$$

if the simple concept γ associates with feature f_i, i.e., $\Delta_\gamma = diag(0, ..., d_\gamma, ...0)$ in the weight functions defined by (4.58). Thus each simple concept $\gamma \in M$ just associates with a single feature f_i like the simple concepts described by (4.36)–(4.39). In order to express the results clearly, we introduce some symbols. For $A \subseteq M$, let

$$\begin{aligned} F(A) &= \{i \mid \gamma \in A,\ simple\ concept\ \gamma\ associates\ with\ the\ feature\ f_i\}, \\ F'(A) &= \{1, 2, ..., n\} - F(A). \end{aligned} \tag{4.73}$$

$F(A)$ is the set of the features the simple concepts in A associate with and $F(A) \cup F'(A) = \{1, 2, ..., n\}$. Let $H = \{i_1, i_2, ..., i_s\}, Q = \{j_1, j_2, ..., j_l\} \subseteq \{1, 2, .., n\}$, where $i_1 < i_2 < ... < i_s, j_1 < j_2 < ... < j_l$. For matrix $P = (p_{ij}) \in R^{n\times n}$, let

$$P_H^Q = (w_{uv}), w_{uv} = p_{i_u j_v}, 1 \le u \le s, 1 \le v \le l. \tag{4.74}$$

So P_H^Q denotes the sub-block of matrix P which is constituted by the rows $i_1, i_2, ..., i_s$ and the columns $j_1, j_2, ..., j_l$. For $x = (x_1, x_2, ..., x_n)' \in R^n$, $x^H = (x_{i_1}, x_{i_2}, ..., x_{i_s})'$. For $\Phi \subseteq \Omega \subseteq R^n$,

$$\Phi^H = \{x^H | x \in \Phi\} \tag{4.75}$$

Proposition 4.5. *Let $\Omega = R^n$ and $(\Omega, \mathscr{F}, \mathscr{P})$ be a m-normal distribution with the density function shown as (4.57). Let M be a set of simple concepts on Ω whose weight functions are defined as (4.71),(4.72). Then for any fuzzy concept $\xi = \sum_{i\in I}(\prod_{m\in A_i} m) \in EM$, the coherence membership functions of ξ on Ω defined by (4.41) or (4.43) are formulated as follows: $\forall x \in \Omega$,*

$$\mu_\xi(x) = \begin{cases} \sup\limits_{i\in I} \prod\limits_{\zeta\in A_i} \Upsilon_\zeta, & for\ (4.41), \\ \sup\limits_{i\in I} \inf\limits_{\zeta\in A_i} \Upsilon_\zeta, & for\ (4.43), \end{cases} \tag{4.76}$$

Where

$$\Upsilon_\zeta=\begin{cases}\dfrac{\sum_{k=1}^m \frac{e^{\Gamma_{k\gamma}}\left|\Theta^{F'(A_i)}_{k\gamma F'(A_i)}\right|^{-\frac{1}{2}}}{\sqrt{\pi^{|F(A)|}|\Sigma_k|}}\mathscr{I}^{A_i}_{k\gamma}(x)}{\sum_{k=1}^m \frac{e^{\Gamma_{k\gamma}}|\Theta_{k\gamma}|^{-\frac{1}{2}}}{\sqrt{|\Sigma_k|}}}, & \zeta=\gamma,\\[2ex] \dfrac{\mathscr{P}(A_i^\tau(\{x\}))-\frac{1}{m}\sum_{k=1}^m \frac{e^{\Gamma_{k\gamma}}\left|\Theta^{F'(A_i)}_{k\gamma F'(A_i)}\right|^{-\frac{1}{2}}}{\sqrt{2^n\pi^{|F(A_i)|}|\Sigma_k|}}\mathscr{I}^{A_i}_{k\gamma}(x)}{1-\frac{1}{m}\sum_{k=1}^m \frac{e^{\Gamma_{k\gamma}}|\Theta_{k\gamma}|^{-\frac{1}{2}}}{\sqrt{(2)^n|\Sigma_k|}}}, & \zeta=\gamma',\end{cases} \tag{4.77}$$

$$\mathscr{I}^{A_i}_{k\gamma}(x)=\int_{A_i^\tau(\{x\})^{F(A_i)}} e^{-(t^{F(A_i)}-\Xi^{F(A_i)}_{k\gamma})'H^{A_i}_{k\gamma}(t^{F(A_i)}-\Xi^{F(A_i)}_{k\gamma})}dt^{F(A_i)}$$

$\Xi_{k\gamma}$,$\Theta_{k\gamma}$, $\Gamma_{k\gamma}$,$1\le k\le m$, $\gamma\in M$, are shown in (4.62),(4.63) and (4.64); $i\in I$, $A_i^\tau(\{x\})^{F(A_i)}$ is defined by (4.27) and (4.75); $F(A_i),F'(A_i)$ are defined by (4.73);

$$H^{A_i}_{k\gamma}=\left(\Theta^{F(A_i)}_{k\gamma F(A_i)}-\Theta^{F'(A_i)}_{k\gamma F(A_i)}\left(\Theta^{F'(A_i)}_{k\gamma F'(A_i)}\right)^{-1}\Theta^{F(A_i)}_{k\gamma F'(A_i)}\right)$$

and $\Theta^{F(A_i)}_{k\gamma F(A_i)}$,$\Theta^{F'(A_i)}_{k\gamma F(A_i)}$, $\Theta^{F'(A_i)}_{k\gamma F'(A_i)}$, $\Theta^{F(A_i)}_{k\gamma F'(A_i)}$, $t^{F(A_i)}$ and $\Xi^{F(A_i)}_{k\gamma}$ are defined by (4.74).

Proof. In what follows, for $\gamma\in A\subseteq M$, we study $\int_{A^\tau(\{x\})}e^{-(t-\Xi_{k\gamma})'\Theta_{k\gamma}(t-\Xi_{k\gamma})}dt$ in (4.61) and let $\omega_1=(t-\Xi_{k\gamma})^{F(A)}$, $\omega_2=(t-\Xi_{k\gamma})^{F'(A)}$. Then

$$\begin{aligned}&\left(t-\Xi_{k\gamma}\right)'\Theta_{k\gamma}\left(t-\Xi_{k\gamma}\right)\\ &=\omega_1'\Theta^{F(A)}_{k\gamma F(A)}\omega_1+\omega_1'\Theta^{F'(A)}_{k\gamma F(A)}\omega_2+\omega_2'\Theta^{F(A)}_{k\gamma F'(A)}\omega_1+\omega_2'\Theta^{F'(A)}_{k\gamma F'(A)}\omega_2\\ &=\omega_1'\Theta^{F(A)}_{k\gamma F(A)}\omega_1\\ &\quad+\left(\omega_2+(\Theta^{F'(A)}_{k\gamma F'(A)})^{-1}\Theta^{F(A)}_{k\gamma F'(A)}\omega_1\right)'\Theta^{F'(A)}_{k\gamma F'(A)}\left(\omega_2+(\Theta^{F'(A)}_{k\gamma F'(A)})^{-1}\Theta^{F(A)}_{k\gamma F'(A)}\omega_1\right)\\ &\quad-\omega_1'\Theta^{F'(A)}_{k\gamma F(A)}(\Theta^{F'(A)}_{k\gamma F'(A)})^{-1}\Theta^{F(A)}_{k\gamma F'(A)}\omega_1\\ &=\omega_1'\left(\Theta^{F(A)}_{k\gamma F(A)}-\Theta^{F'(A)}_{k\gamma F(A)}(\Theta^{F'(A)}_{k\gamma F'(A)})^{-1}\Theta^{F(A)}_{k\gamma F'(A)}\right)\omega_1\\ &\quad+\left(\omega_2+(\Theta^{F'(A)}_{k\gamma F'(A)})^{-1}\Theta^{F(A)}_{k\gamma F'(A)}\omega_1\right)'\Theta^{F'(A)}_{k\gamma F'(A)}\left(\omega_2+(\Theta^{F'(A)}_{k\gamma F'(A)})^{-1}\Theta^{F(A)}_{k\gamma F'(A)}\omega_1\right).\end{aligned} \tag{4.78}$$

Thus by (4.67), we have

$$\begin{aligned}&\int_{\Omega^{F'(A)}}e^{\left(\omega_2+(\Theta^{F'(A)}_{k\gamma F'(A)})^{-1}\Theta^{F(A)}_{k\gamma F'(A)}\omega_1\right)'\Theta^{F'(A)}_{k\gamma F'(A)}\left(\omega_2+(\Theta^{F'(A)}_{k\gamma F'(A)})^{-1}\Theta^{F(A)}_{k\gamma F'(A)}\omega_1\right)}dt^{F'(A)}\\ &=\frac{\pi^{\frac{n-|F(A)|}{2}}}{\left|\Theta^{F'(A)}_{k\gamma F'(A)}\right|^{\frac{1}{2}}}\end{aligned}$$

Furthermore $\int_{A^\tau(\{x\})} e^{-(t-\Xi_{k\gamma})'\Theta_{k\gamma}(t-\Xi_{k\gamma})}dt$ in (4.61) for $\gamma \in A \subseteq M$ can be formulated as follows:

$$\int_{A^\tau(\{x\})} e^{-(t-\Xi_{k\gamma})'\Theta_{k\gamma}(t-\Xi_{k\gamma})}dt$$
$$= \frac{\pi^{\frac{n-|F(A)|}{2}}}{\left|\Theta^{F'(A)}_{k\gamma F'(A)}\right|^{\frac{1}{2}}} \int_{A^\tau(\{x\})^{F(A)}} e^{-(t^{F(A)}-\Xi^{F(A)}_{k\gamma})'H^A_{k\gamma}(t^{F(A)}-\Xi^{F(A)}_{k\gamma})}dt^{F(A)} \tag{4.79}$$

Where

$$H^A_{k\gamma} = \left(\Theta^{F(A)}_{k\gamma F(A)} - \Theta^{F'(A)}_{k\gamma F(A)} \left(\Theta^{F'(A)}_{k\gamma F'(A)}\right)^{-1} \Theta^{F(A)}_{k\gamma F'(A)}\right)$$

Thus by (4.60) and (4.79), we have the upper one of (4.77). When $\zeta = \gamma'$, by the lower one of (4.61) we have the lower one of (4.77). Therefore for any fuzzy concept $\xi = \sum_{i\in I}(\prod_{m\in A_i} m) \in EM$, the membership function of ξ on Ω defined by (4.41), (4.43) is formulated as (4.76). □

4.5 Experimental Studies

In this section, first in Section 4.5.1 the density function $p(x)$ of the probability space $(\Omega, \mathscr{F}, \mathscr{P})$ from which Iris data are drawn is estimated by the probability introduced in Section 1.5.2. Then various experiments on the Iris plant data are considered in Section 4.5.2, 4.5.3, 4.5.4 by applying the coherence membership functions defined by Theorem 4.6 and the techniques developed in Section 4.4 for the Gaussion weight functions of the simple concepts in the multi-normal probability space. In Section 4.5.2, the experiments on Iris data test the consistency of the membership functions determined by the observed data drawn from a probability space with the ones determined by the probability distribution. In Section 4.5.3, we study whether the laws discovered on the observed samples could be applied to the predictions on the whole space via the membership functions of the concepts determined by the probability distribution. The experiments in Section 4.5.4 show that the inferential results of the AFS fuzzy logic based on different data sets drawn from the same probability space are very stable and quite consistent.

4.5.1 Probability Distribution of Iris Plant Data

The well-known Iris data is provided by Fisher in 1936 [40]. these data can be represented by a 150×4 matrix $W = (w_{ij})_{150\times 4}$. The patterns are evenly distributed in three classes: C_1 iris-setosa, C_2 iris-versicolor, and C_3 iris-virginica. A vector of sample i, $(w_{i1}, w_{i2}, w_{i3}, w_{i4})$ has four features: f_1 the sepal length and f_2 the sepal width, and f_3 the petal length and f_4 the petal width (all given in centimeters). So that $X = \{x_1\ , x_2, ..., x_{150}\}$ is the set of the 150 observed samples randomly drawn from a probability space $(\Omega, \mathscr{F}, \mathscr{P})$. Let μ_{ij} and σ_{ij} be the mean and the standard variance of the values of the samples in the class C_i on the feature f_j,

Table 4.2 The mean and the standard variance of each class for iris data

class	mean μ_{ij}	standard variance σ_{ij}
C_1	(5.0060,3.4180,1.4640,0.2440)	(0.3525,0.3810,0.1735,0.1072)
C_2	(5.9360,2.7700,4.2600,1.3260)	(0.5162,0.3138,0.4699,0.1978)
C_3	(6.5880,2.9740,5.5520,2.0260)	(0.6359,0.3225,0.5519,0.2747)

$i = 1,2,3, j = 1,2,3,4$. The values of the means and standard deviations are listed in Table 4.2.

We assume that $\mathscr{P}$ has a multi-normal distribution. By formula (1.8), the density function $p(x)$ [3] of the probability space $(\Omega, \mathscr{F}, \mathscr{P})$ can be estimated via the 150 observed samples as follows:

$$p(x) = \frac{1}{3}\sum_{i=1}^{3}\frac{1}{\sqrt{(2\pi)^p|\Sigma_i|}}e^{-\frac{1}{2}(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i)} \tag{4.80}$$

where $p = 4$,

$$\mu_i = \frac{1}{50}\sum_{j=50(i-1)+1}^{50i}(w_{j1}, w_{j2}, w_{j3}, w_{j4})', \quad i = 1,2,3,$$

$$\Sigma_i = \frac{1}{50}W_i'HW_i, \quad i = 1,2,3,$$

W_1, W_2 and W_3 are the sub-block matrices of W selecting from 1th to 50th rows, from 51th to 100th rows, from 101th to 150th rows, respectively, i.e., W_i is the data of 50 samples in class C_i. $H = I - \frac{1}{50}J$, J is a 4×4 matrix whose entries are all 1. We show them in detail as follows:

$\mu_1 = (5.0060, 3.4180, 1.4640, 0.2440)^T$, $\mu_2 = (5.9360, 2.7700, 4.2600, 1.3260)^T$, $\mu_3 = (6.5880, 2.9740, 5.55202.0260)^T$;

$$\Sigma_1 = \begin{bmatrix} 0.1218 & 0.0983 & 0.0158 & 0.0103 \\ 0.0983 & 0.1423 & 0.0114 & 0.0112 \\ 0.0158 & 0.0114 & 0.0295 & 0.0056 \\ 0.0103 & 0.0112 & 0.0056 & 0.0113 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 0.2611 & 0.0835 & 0.1792 & 0.0547 \\ 0.0835 & 0.0965 & 0.0810 & 0.0404 \\ 0.1792 & 0.0810 & 0.2164 & 0.0716 \\ 0.0547 & 0.0404 & 0.0716 & 0.0383 \end{bmatrix},$$

$$\Sigma_3 = \begin{bmatrix} 0.3963 & 0.0919 & 0.2972 & 0.0481 \\ 0.0919 & 0.1019 & 0.0700 & 0.0467 \\ 0.2972 & 0.0700 & 0.2985 & 0.0478 \\ 0.0481 & 0.0467 & 0.0478 & 0.0739 \end{bmatrix}.$$

In the following sections, we study the membership functions and their logical operations defined on the set of the observed samples X and the whole space Ω with the density function $p(x)$ (refer to (4.80)) by (4.40)-(4.43) in Theorem 4.6 or Proposition 4.5.

4.5.2 Consistency of the Membership Functions on the Observed Data and the Whole Space

In this section, by applying Proposition 4.5, we compare the membership functions and their logic operations defined by (4.40) in Theorem 4.6 on the 150 observed samples in X with the ones defined by (4.41) in Theorem 4.6 according to the density function $p(x)$ shown as (4.80). Let $M = \{m_{ij} \mid 1 \le i \le 4,\ 1 \le j \le 6\}$ be the set of simple concepts for the Iris data X associating to the features. The semantics of the simple concepts $m \in M$ are expressed as follows: for c_{ij}, the mean of the values of the samples in class C_j on the feature f_i

$m_{1,1}$: *"the sepal length is about* $c_{1,1}$*"*, $m_{1,2}$ is the negation of $m_{1,1}$;
$m_{1,3}$: *"the sepal length is about* $c_{1,2}$*"*, $m_{1,4}$ is the negation of $m_{1,3}$;
$m_{1,5}$: *"the sepal length is about* $c_{1,3}$*"*, $m_{1,6}$ is the negation of $m_{1,5}$;

$m_{2,1}$: *"the sepal width is about* $c_{2,1}$*"*, $m_{2,2}$ is the negation of $m_{2,1}$;
$m_{2,3}$: *"the sepal width is about* $c_{2,2}$*"*, $m_{2,4}$ is the negation of $m_{2,3}$;
$m_{2,5}$: *"the sepal width is about* $c_{2,3}$*"*, $m_{2,6}$ is the negation oft $m_{2,5}$;

$m_{3,1}$: *"the petal length is about* $c_{3,1}$*"*, $m_{3,2}$ is the negation of $m_{3,1}$;
$m_{3,3}$: *"the petal length is about* $c_{3,2}$*"*, $m_{3,4}$ is the negation of $m_{3,3}$;
$m_{3,5}$: *"the petal length is about* $c_{3,3}$*"*, $m_{3,6}$ is the negation of $m_{3,5}$;

$m_{4,1}$: *"the petal width is about* $c_{4,1}$*"*, $m_{4,2}$ is the negation of $m_{4,1}$;
$m_{4,3}$: *"the petal width is about* $c_{4,2}$*"*, $m_{4,4}$ is the negation of $m_{4,3}$;
$m_{4,5}$: *"the petal width is about* $c_{4,3}$*"*, $m_{4,6}$ is the negation of $m_{4,5}$.

By Definition 4.3, one can verify that each $m \in M$ is a simple concept. For any $x, y \in X$, if τ is defined by (4.26) as

$$\tau(x,y) = \{m | m \in M, (x,y) \in R_m\},$$

then (M, τ, X) is an AFS structure. Let the σ-algebra on X be $S = 2^X$. For each simple concept $m_{ij} \in M$, refer to (4.71) and (4.72), the weight functions are defined as follows: for any $x \in \Omega$,

$$\rho_{m_{ij}}(x) = e^{-(2\sigma_{ik})^{-2}(f_i(x)-c_{ik})^2},\ i = 1,2,3,4,\ j = 2k-1,\ k = 1,2,3, \tag{4.81}$$

$$\rho_{m_{ij}}(x) = 1 - \rho_{m_{i(j-1)}}(x),\ i = 1,2,3,4, j = 2k,\ k = 1,2,3, \tag{4.82}$$

where σ_{ik} is the standard variance of the values of the samples in class C_k on the feature f_i and $f_i(x)$ is the value of x on the feature f_i, then by the semantics of each $m \in M$ and Definition 4.8 we can verify that $\rho_{m_{ij}}(x)$ is a weight function of the simple concept m_{ij}.

Respectively, applying the weight functions defined by (4.81) and (4.82) to formulas (4.40) in Theorem 4.6 in which $N_x = 1$ (i.e., assume that each sample in X is observed one time) and to formulas (4.76) for (4.41) in Proposition 4.5 in which the density function $p(x)$ is defined by (4.80), we can obtain the membership functions of any fuzzy concept $\xi \in EM$ on both the observed data X and the total space Ω. Refer to Figure 4.1, Figure 4.2, and Figure 4.3, in which the membership functions determined by the observed data X (defined by (4.40)) are denoted as "*observed-memb-fun*" and the ones determined by the density function $p(x)$ (refer to (4.76)) are denoted as "*total-memb-fun*", show the membership functions of fuzzy concepts $m_{1,1}$, $m_{2,1}$ and $m_{1,1}m_{2,1}$.

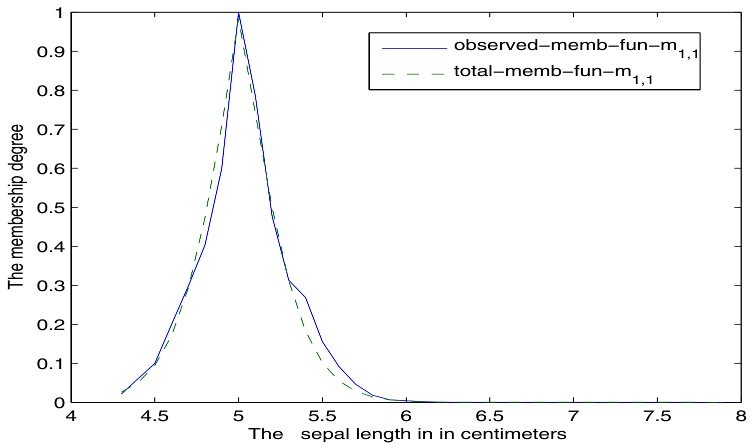


Fig. 4.1 The membership function of the fuzzy concept $m_{1,1}$ with the semantic meaning "*sepal length about 5.0061*" defined by (4.40) according to the 150 observed samples and the one defined by (4.41) according to the density function $p(x)$ in (4.80)

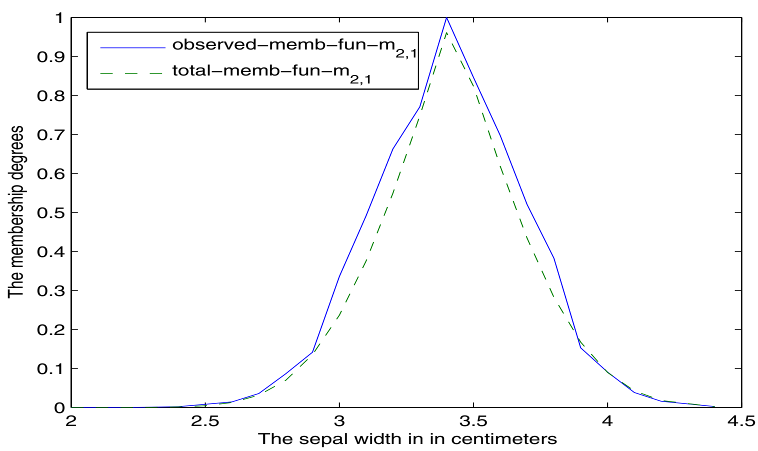


Fig. 4.2 The membership function of the fuzzy concept $m_{2,1}$ with the semantic meaning "*sepal width about 3.4180*" defined by (4.40) according to the 150 observed samples and the one defined by (4.41) according to the density function $p(x)$ in (4.80)

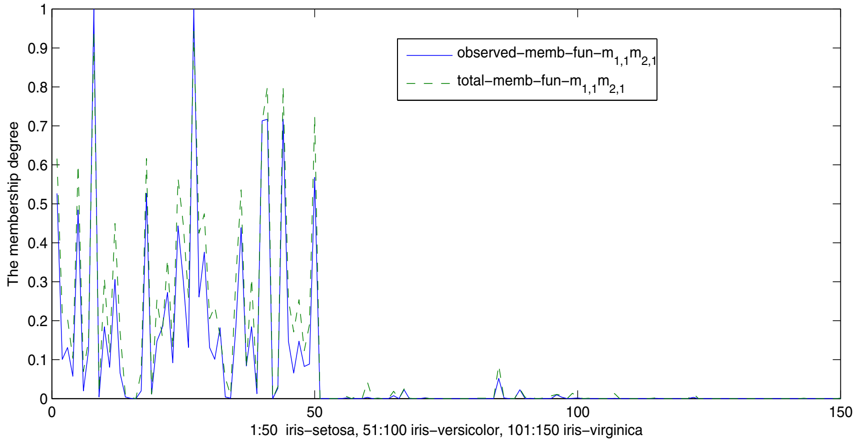


Fig. 4.3 The membership function of the fuzzy concept $m_{1,1}m_{2,1}$ with the underlying semantics "*sepal length about 5.0061 and sepal width about 3.4180*" defined by (4.40) according to 150 observed samples and the one defined by (4.41) according to the density function $p(x)$ in (4.80)

Given these the figures, one can observe that the membership functions defined by (4.40) in Theorem 4.6 according to the observed samples are consistent with the ones defined by (4.41) according to the density functions, although they are calculated by completely different methodologies (i.e., discrete one and continuous one). 2 of Theorem 4.6 ensures that the membership function defined by (4.40) will infinitely approximate to the one defined by (4.41) as the number of samples approaching to infinite. Since $(EM,\vee,\wedge,')$ is a logic system, i.e., the fuzzy concepts in EM are closed under the fuzzy logic operations $\vee$, $\wedge$, $'$. Thus the membership functions and the fuzzy logic operations of the fuzzy concepts in EM are fully determined by (4.40) or (4.41). This implies that the AFS fuzzy logic operations determined by (4.40) in Theorem 4.6 according to the observed samples are also consistent with those determined by (4.41) according to the density functions, as what is shown in Figure 4.3. Therefore, in real world applications, we can apply the knowledge and rules discovered from the observed data X to predict and analyze the system behavior on the total space Ω in virtue of Theorem 4.6. The following Section 4.5.3 shows how this approach works.

4.5.3 Universality of the Laws Discovered on the Observed Data

In the real world applications, we always predict the system behavior by the dependencies discovered based on the observed data. Thus the universality of the laws discovered on the observed data is very crucial. In AFS and probability framework, first we discover the knowledge and rules by the membership functions and their logic operations defined by (4.40) or (4.42) in Theorem 4.6 according to the observed data and describe the discovered laws by some fuzzy concepts in EM. Then the laws discovered on the observed data are generalized to the whole space via the membership functions of the fuzzy concepts defined by (4.41)or (4.43) according to the density function $p(x)$ of the space. Therefore the universality of the membership functions defined by (4.40) or (4.42) in Theorem 4.6 according to the observed data is very crucial. In this section, these issues are investigated via the fuzzy clustering applied to the Iris plant data. Fuzzy clustering problems, which have been studied in [30] by applying AFS fuzzy logic to imitate a way humans cluster data, will be exhaustively discussed in Chapter 9. For a data set $X \subseteq \Omega$, the algorithm presented in [30] (refer to AFS Fuzzy Clustering Algorithm Based on the $1/k-A$ Nearest Neighbors in Section 9.3.4) not only can cluster the samples in X into clusters $C_1, C_2, ..., C_l$, but also give a description of the cluster C_i using a fuzzy concept $\zeta_{C_i} \in EM$. The cluster labels of the samples are determined by the membership functions of $\zeta_{C_i}, i = 1, 2, ..., l$ as follows: for $x \in X$

$$q = \arg \max_{1 \leq k \leq l} \{\mu_{\zeta_{C_k}}(x)\} \Rightarrow x \in C_q. \tag{4.83}$$

In the sequel, we will test the universality of cluster rules $\zeta_{C_i}, i = 1, 2, .., l$ discovered on the observed data X by checking the clustering results determined by the membership functions of the fuzzy concepts $\zeta_{C_i}, i = 1, 2, .., l$ defined by (4.43) according to the density function $p(x)$ of the space. Let $M = \{m_{ij} \mid 1 \leq i \leq 4,\ 1 \leq j \leq 4\}$ be the simple concepts on X. The semantic interpretations of the simple concept in M are shown as follows:

$m_{1,1}$: *"short sepal length"*, $m_{1,2}$: *"mid sepal length"*, $m_{1,3}$: *"not mid sepal length"*, $m_{1,4}$: *"long sepal length"*;
$m_{2,1}$: *"narrow sepal width"*, $m_{2,2}$: *"mid sepal width"*, $m_{2,3}$: *"not mid sepal width"*, $m_{2,4}$: *"wide sepal width"*;
$m_{3,1}$: *"short petal length"*, $m_{3,2}$: *"mid petal length"*, $m_{3,3}$: *"not mid petal length"*, $m_{3,4}$: *"long petal length"*;
$m_{4,1}$: *"narrow petal width"*, $m_{4,2}$: *"mid petal width"*, $m_{4,3}$: *"not mid petal width"*, $m_{4,4}$: *"wide petal width"*.

For each $m \in M$, $x, y \in X$, $(x, y) \in R_m \Leftrightarrow x \geq_m y$. Here $x \geq_m y$ implies that the degree of x belonging to m is larger than or equal to that of y. The degrees of x, y belonging to m are always comparable by the feature values of x, y and the semantic meanings of m. By Definition 4.3, one can verify that each $m \in M$ is a simple concept. For any $x, y \in X$, if $\tau(x, y) = \{m | m \in M, (x, y) \in R_m\}$, then (M, τ, X) is an AFS structure. Let the σ-algebra on X be $S = 2^X$. For the simplicity of the integral computations, let the weight function of each simple concept $m \in M$ be simply defined as $\rho_m(x) = 1, \forall x \in \Omega$. The 150 samples in Iris plant data X are clustered into three clusters by applying the method [30] via the membership functions defined by (4.42) on the observed data X and the fuzzy concepts $\zeta_{C_1} = m_{3,1}m_{4,1} + m_{4,3}m_{4,1}$ states *"short petal length and narrow petal width"* or *"not mid petal width and narrow petal width"*; $\zeta_{C_2} = m_{3,2}m_{4,2}$ reads *"mid petal length and mid petal width"*; $\zeta_{C_3} = m_{1,4}m_{3,4} + m_{4,4}$ states *"long sepal length and long petal length" or "wide petal width"* are obtained to describe the three clusters.

6 samples x_{53} , x_{57}, x_{71}, x_{78}, x_{84}, x_{86} are incorrectly clustered by the membership functions of $\zeta_{C_1}, \zeta_{C_2}, \zeta_{C_3}$ defined by (4.42) according to the observed data which are shown as Figure 4.4. By (4.83), one knows that the different clustering results may be obtained by the different interpretations (membership functions) of the fuzzy concepts $\zeta_{C_i}, i = 1,2,3$ describing the clustering rules. In order to predict the cluster labels of all samples in whole space by the cluster rules discovered on the observed samples, the membership functions of the fuzzy concepts $\zeta_{C_i}, i = 1,2,3$ have to be redetermined by (4.43) according to the density function $p(x)$ of the space in (4.57). 5 samples: $x_{57}, x_{71}, x_{78}, x_{84}, x_{120}$ are incorrectly clustered by the membership

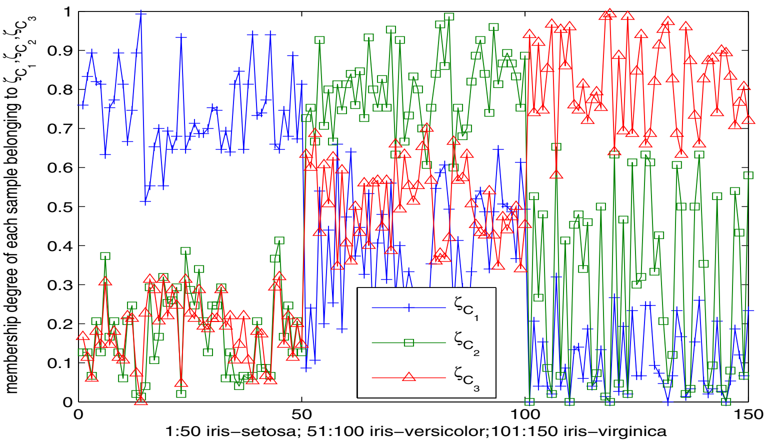


Fig. 4.4 The membership functions of $\zeta_{C_1}, \zeta_{C_2}, \zeta_{C_3}$, the fuzzy concepts describing the three clusters, defined by (4.42) according to the observed data X

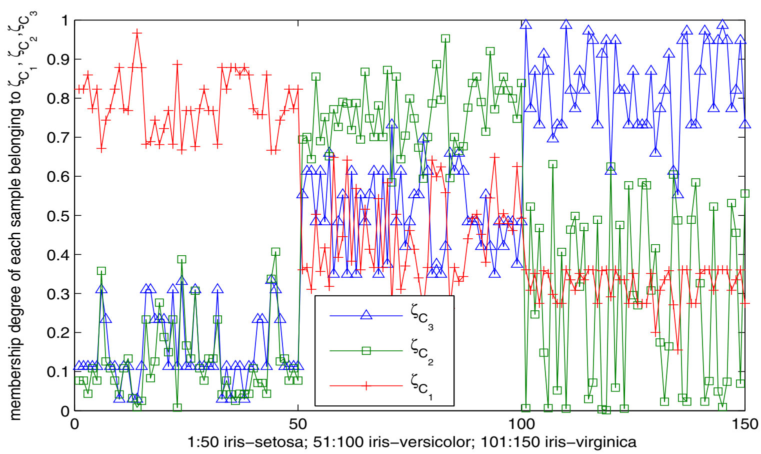


Fig. 4.5 The membership functions of $\zeta_{C_1}, \zeta_{C_2}, \zeta_{C_3}$, the fuzzy concepts describing the three clusters, obtained by (4.43) according to the density function $p(x)$ in (4.57)

functions redefined by (4.43) according to the density function $p(x)$ of the space which are shown as Figure 4.5. The comparison of Figure 4.4 with Figure 4.5 shows that the laws discovered on the observed data can be generalized to the whole space very well. Thus the rules and knowledge discovered by the membership functions defined by (4.40) or (4.42) on the observed data X can be applied to the whole space Ω through (4.41) or (4.43) and used to predict and analyze the system behavior.

4.5.4 Stability Analysis of AFS Fuzzy Logic on the Different Observed Data Sets Drawn from the Same Probability Space

In this section, we study the stability of the membership functions determined by (4.42) in Theorem 4.6 according to different observed data sets. Let X_1, X_2 be two observed data sets. For each fuzzy concept in EM, its membership functions on AFS structure (M, τ, X_1) and (M, τ, X_2) may be different due to the different data sets X_1, X_2, although every simple concept $m \in M$ has the same semantics on X_1 and X_2. For example, for a person, an NBA basketball player may describe that the person is not *"tall"* and a ten year old child may describe that the same person is very *"tall"*. Because the people the NBA basketball player often meets are different from the people the child meets, i.e., the different data they observed, they have different membership functions to describe the fuzzy concepts. This difference of the membership functions is mainly led by the observed data sets drawn from the different probability distributions. Although the interpretations (membership functions) of concept tall of two NBA basketball players may be different, it is impossible that a NBA basketball player describes the person not *"tall"* and another NBA basketball player describes the same person very *"tall"*. This implies that the membership functions of the fuzzy concept determined by the different observed data sets drawn from the same probability space should be stable. In fact, the people two NBA basketball players meet are drawn from the same probability distribution and this difference of membership functions results from the fuzziness (subjective imprecision, i.e., the weight function of simple concept γ, $\rho_\gamma(x)$) and the randomness (objective uncertainty: randomly observed the data sets X). By Theorem 4.6, we know that the uncertainty of randomness decreases to 0 as the observed data approaching to infinity and the uncertainty of fuzziness decreases to 0 as the difference of the weight functions of simple concepts approaching to 0.

In the following experiments on the Iris plant data, we analysis the stability of membership functions and AFS fuzzy logic on the different observed data sets drawn from the same probability space $(\Omega, \mathscr{F}, \mathscr{P})$. Let $M = \{m_1, m_2, ..., m_8\}$ be the set of simple concepts on the features f_3, f_4 the petal length and the petal width. The semantics of the simple concepts $m \in M$ are expressed as

m_1: *"the petal length is long"*, m_2 is the negation of m_1;
m_3: *"the petal length is middle"*, m_4 is the negation of m_3;
m_5: *"the petal width is wide"*, m_6 is the negation of m_5;
m_7: *"the petal width is middle"*, , m_8 is the negation of m_7.

By Definition 4.3, one can verify that each $m \in M$ is a simple concept. For any $x, y \in X$, if τ is defined by (4.26) as follows

$$\tau(x,y) = \{m | m \in M, (x,y) \in R_m\},$$

then (M, τ, X) is an AFS structure. Let the σ-algebra on X be $S = 2^X$. For simplicity, let the weight function of every simple concept $m \in M$, $\rho_m(x) = 1$ for any $x \in \Omega$. This implies that every sample is equally important to simple concept $m \in M$. For each fuzzy concept in EM, its membership function is defined by (4.42) in Theorem 4.6 in which $N_x = 1$ (i.e., assume that each sample in X is observed one time).

We run 4 experiments. In each experiment, first the 150 samples of the Iris plant data are randomly parted into two equal number of sample sets, i.e., $X = X_1 \cup X_2$, $X_1 \cap X_2 = \varnothing$, $|X_1| = |X_2|$. Second, the AFS structures (M, τ, X_1) and (M, τ, X_2) are separately established according the the data sets X_1, X_2. Finally, using the membership functions on (M, τ, X_1) and (M, τ, X_2), we separately apply fuzzy clustering algorithm based on AFS fuzzy logic in [30] to find the fuzzy concepts $\zeta_{C_i}, i = 1,2,3$ to describe the clusters C_1: iris-setosa, C_2: iris-versicolor, and C_3: iris-virginica in X_1 and X_2. In [30] (refer to AFS Fuzzy Clustering Algorithm Based on the $1/k - A$ Nearest Neighbors in Section 9.3.4), the author proposed an overall AFS clustering procedure

1. For each sample $x \in X$ the data set, using membership functions defined by (4.42) and the AFS fuzzy logic system $(EM, \vee, \wedge, ')$ on X, find a fuzzy concept $\xi_x \in EM$ to describe x.
2. Evaluate the degree of similarity between two samples $x, y \in X$ based on their fuzzy descriptions ξ_x, ξ_y.
3. Cluster the data X according to the degrees of the similarity of each pair of sample in X.
4. For each cluster C, find a fuzzy concept $\xi_C \in EM$ to describe the character of the samples in this cluster.
5. Each sample is clustered according to the membership functions of the fuzzy concepts describing the clusters by (4.83).

Table 4.3 shows the fuzzy concepts describing the three clusters in data X_1, X_2 in the 4 experiments using the above algorithm to the data sets. From 8 different data sets in the 4 experiments, three different fuzzy concepts $\xi_{C1}^1 = m_6 + m_2$, $\xi_{C1}^2 = m_6 + m_2 m_4$, $\xi_{C1}^3 = m_6 m_8 + m_2 m_4 + m_2 m_6$ are extracted to describe C_1: iris-setosa, one fuzzy concept $\xi_{C2} = m_3 + m_7$ is extracted to describe C_2: iris-versicolor, two fuzzy concepts $\xi_{C3}^1 = m_1 + m_5$, $\xi_{C3}^2 = m_1 + m_5 m_8$ are extracted to describe C_3: iris-virginica. By Theorem 4.1, we know that $\xi_{C1}^1 \geq \xi_{C1}^2 \geq \xi_{C1}^3$, $\xi_{C3}^1 \geq \xi_{C3}^2$.

This implies that the interpretations of the fuzzy concept discovered by the 8 different observed data sets to describe each cluster are very similar. By each triple fuzzy descriptions for the three clusters, we can obtain clustering of the 150 samples by (4.83). Thus, by the 3 fuzzy descriptions for C_1, one for C_2 and two for C_3, we have 6 triple fuzzy descriptions: $\{\xi_{C1}^1, \xi_{C2}, \xi_{C3}^1\}$, $\{\xi_{C1}^1, \xi_{C2}, \xi_{C3}^2\}$, $\{\xi_{C1}^2, \xi_{C2}, \xi_{C3}^1\}$, $\{\xi_{C1}^2, \xi_{C2}, \xi_{C3}^2\}$, $\{\xi_{C1}^3, \xi_{C2}, \xi_{C3}^1\}$, $\{\xi_{C1}^3, \xi_{C2}, \xi_{C3}^2\}$ for the clustering of the 150 Iris plant samples and their accurate rates are 96.67%, 96%, 96.67%, 96%, 96.67%,

Table 4.3 Descriptions of fuzzy concepts of the clusters in 4 experiments

	Fuzzy descriptions determined by the data sets X_1, X_2		
no of experiments	C_1: iris-setosa	C_2: iris-versicolor	C_3: iris-virginica
1th data X_1	$m_6 + m_2$	$m_3 + m_7$	$m_1 + m_5$
1th data X_2	$m_6 + m_2$	$m_3 + m_7$	$m_1 + m_5$
2th data X_1	$m_6 + m_2m_4$	$m_3 + m_7$	$m_1 + m_5$
2th data X_2	$m_6 + m_2m_4$	$m_3 + m_7$	$m_1 + m_5$
3th data X_1	$m_6 + m_2$	$m_3 + m_7$	$m_1 + m_5$
3th data X_2	$m_6m_8 + m_2m_4 + m_2m_6$	$m_3 + m_7$	$m_1 + m_5$
4th data X_1	$m_6 + m_2$	$m_3 + m_7$	$m_1 + m_5m_8$
4th data X_2	$m_6 + m_2m_4$	$m_3 + m_7$	$m_1 + m_5m_8$

96%, respectively. Figures 4.6, 4.7, 4.8 show the membership functions of the different fuzzy descriptions of clusters $C_i, i = 1,2,3$. Based on the figures, we can observe that although the fuzzy concepts discovered on the different observed data sets to describe each cluster may exhibit a very little difference, i.e., with slightly different interpretations, the membership functions are very similar. These experimental results imply that inferential results of the AFS fuzzy logic discovered from the different data sets drawn from the same probability space are very stable and quite consistent.

In this chapter, we propose an algorithm of determining membership functions and their fuzzy logic operations of fuzzy concepts according to the semantics and the statistics of the underlying data. Specially, it opens the door to explore the statistic properties of fuzzy set theory and to a major enlargement of the role of natural languages in probability theory. We prove that the membership functions defined by (4.40) or (4.42) according to the observed data converges to the one defined

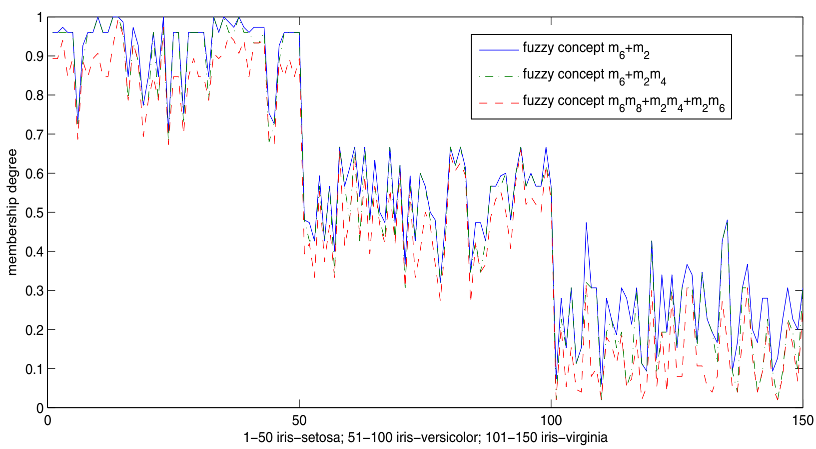


Fig. 4.6 The membership functions of the three different fuzzy descriptions of cluster C_1 discovered on the 8 different observed data sets in the 4 experiments

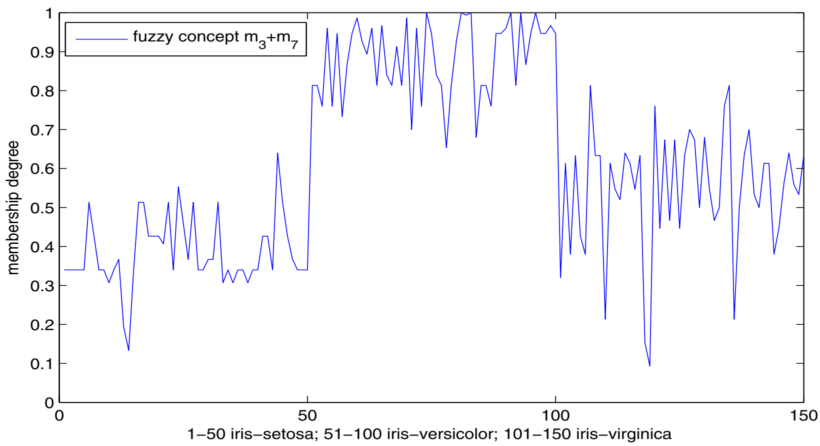


Fig. 4.7 The membership function of the fuzzy description of cluster C_2 discovered on the 8 different observed data sets in the 4 experiments

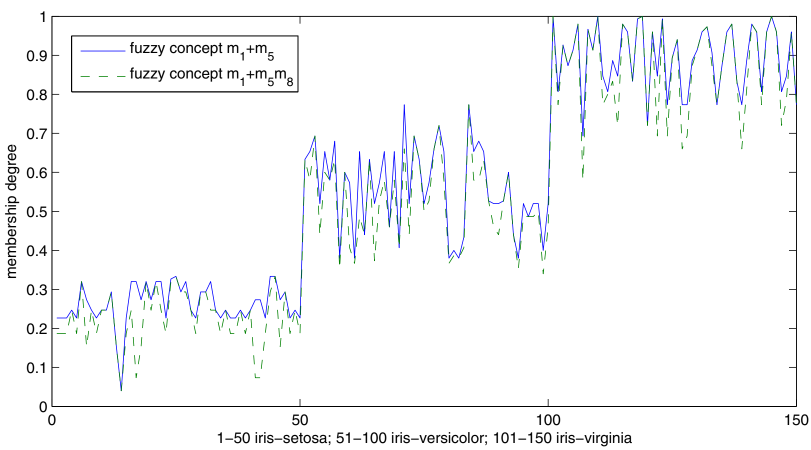


Fig. 4.8 The membership functions of the two different fuzzy descriptions of cluster C_3 discovered on the 8 different observed data sets in the 4 experiments

by (4.41) or (4.43) according to the probability distribution for all $x \in \Omega$ as $|X|$ approaching infinity. Theorem 4.6 not only provides a clear representations of imprecision and uncertainty which takes both fuzziness (subjective imprecision) and randomness (objective and uncertainty) into account and treats the uncertainty of randomness and of imprecision in a unified and coherent manner, but also gives a practical methodology of knowledge discovery and representation for data analysis. Along this approach direction, more systematic studies may be carried out in view of an organic integration of the mentioned aspects within a general framework for statistical analysis based on a wider notion of information uncertainty including fuzziness and its statistical treatment.

Exercises

Exercise 4.1. Let M be a non-empty set. If $\sum_{i\in I}(\prod_{m\in A_i} m) \in EM, A_t \subseteq A_s$, $t,s \in I$, $t \neq s$, show

$$\sum_{i\in I}(\prod_{m\in A_i} m) = \sum_{i\in I-\{s\}}(\prod_{m\in A_i} m).$$

Exercise 4.2. Prove that the binary compositions $\vee, \wedge$ in Theorem 4.1 satisfy the following properties: for any $\gamma, \zeta, \eta \in EM$

L1: $\gamma \wedge \gamma = \gamma$, $\gamma \wedge \gamma = \gamma$. (Idempotency)
L2: $\zeta \vee \eta = \eta \vee \zeta$, $\zeta \wedge \eta = \eta \wedge \zeta$. (Commutativity)
L3: $\gamma \vee (\zeta \vee \eta) = (\gamma \vee \zeta) \vee \eta$, $\gamma \wedge (\zeta \wedge \eta) = (\gamma \wedge \zeta) \wedge \eta$. (Associativity)

Exercise 4.3. For any $\gamma, \zeta, \eta, \in EM$, show the following D1 and D2 hold.

D1: $\gamma \wedge (\zeta \vee \eta) = (\gamma \wedge \zeta) \vee (\gamma \wedge \eta)$;
D2: $\gamma \vee (\zeta \wedge \eta) = (\gamma \vee \zeta) \wedge (\gamma \vee \eta)$.

Exercise 4.4. For the completely distributive lattice $(EM, \vee, \wedge)$, prove that

$$\bigwedge_{i\in I}(\alpha \vee \beta_i) = \alpha \vee (\bigwedge_{i\in I} \beta_i)$$

for any $\alpha, \beta_i \in EM, i \in I$.

Exercise 4.5. Let X be a set and ζ be any concept on X. Let R_ζ be the binary relation of the simple concept ζ defined by Definition 4.2 and $M_\zeta = (r_{ij})_{n\times n}$ be the correspondent Boolean matrix of R_ζ defined by Definition 3.5. Show the following assertions hold

1) $r_{ii} = 0 \Leftrightarrow r_{ij} = 0$ for all $j = 1,2,...,n$.
2) There exists a permutation Boolean matrix P such that

$$M_\zeta = P \begin{bmatrix} N & J \\ O_1 & O_2 \end{bmatrix} P^T,$$

where N is a Boolean matrix such that $N + I = N$, J is a universal Boolean matrix, i.e.,whose elements are all 1, O_1 and O_2 are zero matrices.
3) The concept ζ on a set X is a simple concept if and only if N is the correspondent Boolean matrix of a quasi-linear order, i.e., $N^2 = N, N + I = N$.

Exercise 4.6. Let M, $X = \{x_1, x_2, ..., x_n\}$ be finite sets and $g : X \times X \to 2^M$. Prove that g satisfies AX1, AX2 of Definition 4.5 if and only if

$$M_g^2 = M_g \textit{ and } \bigcup_{1\le j\le n} m_{ij} \subseteq m_{ii}, i = 1,2,...,n.$$

Open Problems

Problem 4.1. Find the simple and effective computing methodology for the membership functions of the fuzzy concepts defined by formulas (4.43) and (4.41) on the whole space, which also could be applied to proceed mathematical analysis.

Problem 4.2. Find the numeric computations of the high dimension integral for the membership functions of fuzzy concepts via the probability distributions when many simple concepts involve the fuzzy concepts.

Problem 4.3. Estimation of the error boundary between the membership functions of fuzzy concepts obtained by the observed data and that determined by the probability distributions.

Problem 4.4. Estimation of the error boundary between the membership functions of fuzzy concepts obtained for different weight functions of the simple concepts and analysis of the influence of subjective imprecision on the interpretations of the fuzzy concepts.

Problem 4.5. Also apart from product which has been shown in (4.33) is there any other t-norm which can (in a limited way - for conjunctions of basic expressions) be captured in the AFS model?

Problem 4.6. Are there links between the AFS theory and Lawrys Label Semantics which also defines fuzziness in terms of a probability measure?

References

1. Bellman, R., Kalaba, L., Zadeh, L.A.: Abstraction and pattern classification. J. Math. Anal. Appl. 13, 1–7 (1966)
2. Bellman, R., Zadeh, L.A.: Decision making in a fuzzy environment. Management Sci. 17, B141–B164 (1970)
3. Casella, G., Berger, R.: Statistical Inference, 2nd edn. Wadsworth Group, New York (2004)
4. Coppi, R., Gilb, M.A., Kiersc, H.A.L.: The fuzzy approach to statistical analysis. Computational Statistics and Data Analysis 51, 1–14 (2006)
5. Coletti, G., Scozzafava, R.: Conditional probability and fuzzy information. Computational Statistics and Data Analysis 51, 115–132 (2006)
6. Coletti, G., Scozzafava, R.: Conditional probability, fuzzy sets and possibility: a unifying view. Fuzzy Sets and Systems 144, 227–249 (2004)
7. Coletti, G., Scozzafava, R.: Conditioning in a coherent setting: theory and applications. Fuzzy Sets and Systems 155, 26–49 (2005)
8. Dubois, D., Prade, H.: The three semantics of fuzzy sets. Fuzzy Sets and Systems 90, 141–150 (1997)
9. Ding, R., Liu, X.D., Chen, Y.: The fuzzy clustering algorithm based on AFS topology. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS, vol. 4223, pp. 89–98. Springer, Heidelberg (2006)

10. Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatly, J., Kruger, L.: The Empire of Chance. Cambridge University Press, Cambridge (1990)
11. Graver, J.E., Watkins, M.E.: Combinatorics with Emphasis on the Theory of Graphs. Springer, New York (1977)
12. Grzegorzewski, P.: The coefficient of concordance for vague data. Computational Statistics and Data Analysis 51, 314–322 (2006)
13. Hall, M.A.: Correlation-based Feature Selection for Machine Learning. A Dissertation submit to Department of Computer Science. University of Waikato, Hamilton, NewZealand (1999)
14. Halmos, P.R.: Measure theory. Springer, New York (1974)
15. Hisdal, E. (Guest ed.) : Special Issue-Interpretation of Grades of Membership. Fuzzy Sets and Systems 25(3), 271–379 (1988)
16. Kim, K.H.: Boolean Matrix Theory and Applications. Marcel Dekker, New York (1982)
17. Kuhr, J., Mundici, D.: De Finetti theorem and Borel states in [0, 1]-valued algebraic logic. International Journal of Approximate Reasoning 46, 605–616 (2007)
18. Loginov, V.I.: Probability Treatment of Zadeh Membership Functions and Their Use in Pattern Recognition. Engineering Cybernetics, 68–69 (1966)
19. Lindley, D.V.: The Probability Approach to the Treatment of Uncertainty in Artifical Intelligence and Expert Systems. Statistical Science 2, 17–24 (1987)
20. Lawry, J.: A Framework for Linguistic Modelling. Artificial Intelligence 155, 1–39 (2006)
21. Liu, X.D.: The Structure of Fuzzy Matrices. Journal of Fuzzy Mathematics 2, 311–325 (1994)
22. Liu, X.D.: A New Mathematical Axiomatic System of Fuzzy Sets and Systems. Journal of Fuzzy Mathematics 3, 559–560 (1995)
23. Liu, X.D.: The Fuzzy Theory Based on AFS Algebras and AFS Structure. Journal of Mathematical Analysis and Applications 217, 459–478 (1998)
24. Liu, X.D.: The Topology on AFS Algebra and AFS Structure. Journal of Mathematical Analysis and Applications 217, 479–489 (1998)
25. Liu, X.D.: The Fuzzy Sets and Systems Based on AFS Structure, EI Algebra and EII algebra. Fuzzy Sets and Systems 95, 179–188 (1998)
26. Liu, X.D.: A new fuzzy model of pattern recognition and hitch diagnoses of complex systems. Fuzzy Sets and Systems 104, 289–297 (1999)
27. Liu, X.D., Pedrycz, W., Zhang, Q.L.: Axiomatics Fuzzy sets logic. In: The Proceedings of IEEE International Conference on Fuzzy Systems, vol. 1, pp. 55–60 (2003)
28. Liu, X.D., Zhu, K.J., Huang, H.Z.: The Representations of Fuzzy Concepts Based on the Fuzzy Matrix Theory and the AFS Theory. In: IEEE International Symposium on Intelligent Control, Houston, Texas, USA, October 5-8, 2003, pp. 1006–1011 (2003)
29. Liu, X.D., Zhang, Q.L.: The Fuzzy Cognitive Maps Based on AFS Fuzzy Logic. Dynamics of Continuous, Discrete and Impulsive Systems. Series A: Mathematical Analysis 11(5-6), 787–796 (2004)
30. Liu, X.D., Wang, W., Chai, T.Y.: The Fuzzy Clustering Analysis Based on AFS Theory. IEEE Transactions on Systems, Man and Cybernetics Part B 35(5), 1013–1027 (2005)
31. Liu, X.D., Liu, W.Q.: Credit Rating Analysis with AFS Fuzzy Logic. In: Wang, L., Chen, K., S. Ong, Y. (eds.) ICNC 2005. LNCS, vol. 3612, pp. 1198–1204. Springer, Heidelberg (2005)
32. Liu, X.D.: AFS Theory and Its Applications. IEEE Systems, Man and Cybernetics, Society eNewsletter 13 (December 2005), http://www.ieeesmc.org/Newsletter/Dec2005/R8xdliu.php

33. Liu, X.D., Chai, T.Y., Wang, W.: AFS Fuzzy Logic Systems and Its Applications to Model and Control. International Journal of Information and Systems Sciences 2(3), 285–305 (2006)
34. Liu, X.D.: The Development of AFS Theory Under Probability Theory. International Journal of Information and Systems Science 3(2), 326–348 (2007)
35. Liu, X.D., Chai, T.Y., Wang, W.: Approaches to the Representations and Logic Operations for Fuzzy Concepts in the Framework of Axiomatic Fuzzy Set Theory I. Information Sciences 177, 1007–1026 (2007)
36. Liu, X.D., Chai, T.Y., Wang, W.: Approaches to the Representations and Logic Operations for Fuzzy Concepts in the Framework of Axiomatic Fuzzy Set Theory II. Information Sciences 177, 1027–1045 (2007)
37. Liu, X.D., Pedrycz, W.: The Development of Fuzzy Decision Trees in the Framework of Axiomatic Fuzzy Set Logic. Applied Soft Computing 7, 325–342 (2007)
38. Liu, X.D., Zhang, L.S., Zhou, J., Zhou, K.J., Zhang, Q.L.: The Structures of EI Algebras Generated by Information Attributes. Int. J. Intelligent Systems Technologies and Applications 3(3/4), 341–355 (2007)
39. Liu, X.D.: The Membership Functions of Fuzzy Concepts via AFS Approach and Probability Theory. Fuzzy Sets and Systems (submited)
40. Merz, C.J., Murphy, P.M.: UCI Repository for Machine Learning Data-Bases. Dept. of Information and Computer Science, University of California, Irvine, CA (1996), http://www.ics.uci.edu/~mlearn/MLRepository.html
41. Mundici, D.: Bookmaking over Infinite-valued Events. International Journal of Approximate Reasoning 43, 223–240 (2006)
42. Ren, Y., Song, M.L., Liu, X.D.: New Approaches to the Fuzzy Clustering via AFS Theory. International Journal of Information and Systems Sciences 3(2), 307–325 (2007)
43. Ren, Y., Wang, X.C., Liu, X.D.: Fuzzy Clustering Approaches Based On AFS Fuzzy Logic I. In: Sixth World Congress on Intelligent Control and Automation, Dalian, China, June 21-23, 2006, vol. 5, pp. 4244–4248 (2006)
44. Singpurwalla, N.D., Booker, J.M.: Membership Functions and Probability Measures of Fuzzy Sets. Journal of the American Statistical Association 99(467), 867–877 (2004)
45. Simonoff, J.S.: Smoothing Methods in Statistics. Springer, New York (1996)
46. Turksen, I.B.: Measurement of Membership Functions and Their Acquisition. Fuzzy Sets and Systems 40, 5–38 (1991)
47. Wang, G.J.: Theory of Topological Molecular Lattices. Fuzzy Sets and Systems 47, 351–376 (1992)
48. Wang, X.C., Rui, D., Liu, X.D.: New approach to the fuzzy Classification via AFS theory. International Journal of Information and Systems Sciences 3(4), 581–593 (2007)
49. Wang, X.C., Ren, Y., Liu, X.D.: Fuzzy Clustering Approaches Based On AFS Fuzzy Logic II. In: Sixth World Congress on Intelligent Control and Automation, Dalian, China, June 21-23, 2006, vol. 5, pp. 4199–4203 (2006)
50. Zhang, L.S., Liu, X.D.: Concept lattice and AFS algebra. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS, vol. 4223, pp. 290–299. Springer, Heidelberg (2006)
51. Zadeh, L.A.: Fuzzy Sets. Information and Control 8, 338–353 (1965)
52. Zadeh, L.A.: Probability Measures of Fuzzy Events. Journal of Mathematical Analysis and Applications 23, 421–427 (1968)
53. Zadeh, L.A.: Fuzzy Sets as a Basis for a Theory of Possibility. Fuzzy Sets and Systems 1, 3–28 (1978)

54. Zadeh, L.A.: A Theory of Approximate Reasoning. In: Hayes, J.E., Michie, D., Mikulich, L.I. (eds.) Machine Intelligence, vol. 9, pp. 149–194. Elsevier, New York (1979)
55. Zadeh, L.A.: Fuzzy Logic. IEEE Trans. Comput. 35, 83–93 (1988)
56. Zadeh, L.A.: Toward a Theory of Fuzzy Information Granulation and its Centrality in Human Reasoning and Fuzzy Logic. Fuzzy Sets and Systems 90, 111–127 (1997)
57. Zadeh, L.A.: Discussion: Probability Theory and Fuzzy Logic Are Complementary Rather Than Competitive. Technometrics 37, 271–276 (1995)
58. Zhang, Y.J., Liang, D.Q., Tong, S.C.: On AFS Algebra Part I. Information Sciences 167, 263–286 (2004)
59. Zhang, Y.J., Liang, D.Q., Tong, S.C.: On AFS Algebra Part II. Information Sciences 167, 287–303 (2004)

Chapter 5
AFS Algebras and Their Representations of Membership Degrees

In this chapter, first we construct some lattices—AFS algebras using sets X and M over an AFS structure (M,τ,X) for the representation of the membership degrees of each sample $x \in X$ belonging to the fuzzy concepts in EM. Then the mathematical properties and structures of AFS algebras are exhaustively discussed. Finally, the relations, advantages and drawbacks of various kinds of AFS representations for fuzzy concepts in EM are analyzed. Some results listed without proofs are left for the reader as exercises.

5.1 AFS Algebra

In [7], the author has defined a family of completely distributive lattices AFS algebra and applied AFS algebra to study the lattice value representations for fuzzy concepts. AFS algebra includes EI^n algebras and $E^{\#}I^n$ algebras, $n = 1,2,....$

5.1.1 EI^n Algebras

In this section, we introduce an EI^n algebra. The EI algebra which is applied to study the semantics and logic of the fuzzy concepts presented in Chapter 4 is a particular type of the EI^n algebra in case $n = 1$.

Definition 5.1. ([7]) Let $X_1,...,X_n,M$ be $n+1$ non-empty sets. Then the set $EX_1...X_nM^+$ is defined as follows

$$EX_1...X_nM^+ = \left\{ \sum_{i\in I}(u_{1i}...u_{ni}A_i) \mid A_i \in 2^M, u_{ri} \in 2^{X_r}, r = 1,2,...,n, i \in I, \right.$$
$$\left. I \text{ is a non-empty indexing set} \right\}.$$

In the case $n = 0$,

$$EM^+ = \left\{ \sum_{i\in I} A_i \mid A_i \in 2^M, i \in I, I \text{ is a non-empty indexing set} \right\}.$$

X. Liu and W. Pedrycz: Axiomatic Fuzzy Set Theory and Its Applications, STUDFUZZ 244, pp. 167–224.
springerlink.com

Each $\sum_{i\in I}(u_{1i}...u_{ni}A_i)$ is an element of $EX_1...X_nM^+$ and $\sum_{i\in I}$ is a symbol expressing that element $\sum_{i\in I}(u_{1i}...u_{ni}A_i)$ is composed of items $(u_{1i}...u_{ni})A_i$, $u_{ri}\subseteq X_r, A_i\subseteq M, r=1,2,...,n, i\in I$ separated by "+" . $\sum_{i\in I}(u_{1p(i)}...u_{np(i)}A_{p(i)})$ and $\sum_{i\in I}(u_{1i}...u_{ni}A_i)$ are the same elements of $EX_1...X_nM^+$ if p is a bijection from I to I.

Definition 5.2. ([7]) Let $X_1,...,X_n,M$ be $n+1$ non-empty sets. A binary relation R on $EX_1...X_nM^+$ is defined as follows: $\forall\sum_{i\in I}(u_{1i}...u_{ni}A_i)$, $\sum_{j\in J}(v_{1j}...v_{nj}B_j)\in EX_1...X_nM^+$,

$\left[\sum_{i\in I}(u_{1i}...u_{ni}A_i)\right]R\left[\sum_{j\in J}(v_{1j}...v_{nj}B_j)\right]\Longleftrightarrow$

(i) $\forall(u_{1i}...u_{ni})A_i$ $(i\in I)$, $\exists(v_{1h}...v_{nh})B_h$ $(h\in J)$ such that $A_i\supseteq B_h$, $u_{ri}\subseteq v_{rh}$, $1\le r\le n$;

(ii) $\forall(v_{1j}...v_{nj})B_j$ $(j\in J)$, $\exists$ $(u_{1k}...u_{nk})A_k$ $(k\in I)$, such that $B_j\supseteq A_k$, $v_{rj}\subseteq u_{rk}$, $1\le r\le n$.

It is obvious that R is an equivalence relation. We denote $EX_1...X_nM^+/R$ (i.e., the quotient set) as $EX_1...X_nM$. The notation $\sum_{i\in I}(u_{1i}...u_{ni}A_i)=\sum_{j\in J}(v_{1j}...v_{nj}B_j)$ implies that the equivalence class containing $\sum_{i\in I}(u_{1i}...u_{ni}A_i)$ is the same as the equivalence class containing $\sum_{j\in J}(v_{1j}...v_{nj}B_j)$.

Proposition 5.1. *([7]) Let $X_1,...,X_n,M$ be n+1 non-empty sets. If $A_t\subseteq A_s$, $u_{rt}\supseteq u_{rs}$, $r=1,2,...,n$, $t,s\in I$, $t\neq s$, $\sum_{i\in I}(u_{1i}...u_{ni}A_i)\in EX_1...X_nM$, then*

$$\sum_{i\in I}(u_{1i}...u_{ni}A_i)=\sum_{i\in I-\{s\}}(u_{1i}...u_{ni}A_i).$$

Theorem 5.1. *([7]) Let $X_1,...X_n,M$ be $n+1$ non-empty sets. Then $(EX_1...X_nM,\vee,\wedge)$ forms a completely distributive lattice under the binary compositions $\vee$ and $\wedge$ defined as follows: $\forall\sum_{i\in I}(u_{1i}...u_{ni}A_i)$, $\sum_{j\in J}(v_{1j}...v_{nj}B_j)\in EX_1...X_nM$,*

$$\begin{aligned}\sum_{i\in I}(u_{1i}...u_{ni}A_i)\vee\sum_{j\in J}(v_{1j}...v_{nj}B_j) &= \sum_{k\in I\sqcup J}(w_{1k}...w_{nk}C_k)\\ &\triangleq\sum_{i\in I}(u_{1i}...u_{ni}A_i)+\sum_{j\in J}(v_{1j}...v_{nj}B_j),\end{aligned}\quad(5.1)$$

$$\sum_{i\in I}(u_{1i}...u_{ni}A_i)\wedge\sum_{j\in J}(v_{1j}...v_{nj}B_j)=\sum_{i\in I,j\in J}\left[(u_{1i}\cap v_{1j}...u_{ni}\cap v_{nj})(A_i\cup B_j)\right],\quad(5.2)$$

where $\forall k\in I\sqcup J$, $C_k=A_k,w_{rk}=u_{rk}$ when $k\in I$ and $C_k=B_k,w_{rk}=v_{rk}$ when $k\in J$, $r=1,2,...,n$.

Proof. We just prove the theorem in the case $n=0$ and the other cases, which remain as exercises, are similar to the proof of Theorem 4.1. Let h be a map from EM^+ to EM^* defined as follows: for any $\xi=\sum_{i\in I}A_i\in EM^+$, $h(\xi)=\sum_{i\in I}(\prod_{m\in A_i}m)$. It is clear that h is a one-to-one correspondence between EM^+ and EM^*. One can verify that the following assertions hold: for $\xi,\eta\in EM$,

1. $\xi=\eta\Leftrightarrow h(\xi)=h(\eta)$;
2. $h(\xi\vee\eta)=h(\xi)\vee h(\eta), h(\xi\wedge\eta)=h(\xi)\wedge h(\eta)$.

This implies that h is an isomorphism from $(EM^+/R,\vee,\wedge)$ to $(EM^*/R,\vee,\wedge)$ which is a completely distributive lattice. Therefore $(EM^+/R,\vee,\wedge)$ is a completely distributive lattice. □

Considering that $(EM^+/R,\ \vee,\ \wedge)$ and $(EM^*/R,\ \vee,\ \wedge)$ are isomorphism, both EM^+/R and EM^*/R are denoted as EM and called EI algebra. $(EX_1...X_nM,\vee,\wedge)$ is called the EI^{n+1} *(expanding* $n+1$ *sets* $X_1,...X_n,M$*) algebra* over $X_1,...,X_n$ and M. $X_1...X_n\varnothing$ and $\varnothing...\varnothing M$ are the maximum element and minimum element of $EX_1...X_nM$, respectively. For $\alpha=\sum_{i\in I}(u_{1i}...u_{ni}A_i)$, $\beta=\sum_{j\in J}(v_{1j}...v_{nj}B_j)\in EX_1...X_nM$, $\alpha\leq\beta \iff \forall(u_{1i}...u_{ni})A_i\ \ (i\in I)$, $\exists(v_{1h}...v_{nh})B_h\ \ (h\in J)$ such that $A_i\supseteq B_h$, $u_{ri}\subseteq v_{rh}$, $1\leq r\leq n$.

For a set, we know that the subsets of the set often contain or represent some pieces of useful information. In real world applications, instead of a single set, often many sets are involved and the information and knowledge represented by the subsets of different sets may exhibit various types of relationships. In order to study such diverse relations associated with different sets, we introduce the notation of $EX_1...X_nM^+$. Every element of $EX_1...X_nM^+$ is a "formal sum" of the terms constituted by the subsets of X_1, X_2, ..., X_n, M. For $\gamma=\sum_{i\in I}(u_{1i}...u_{ni}A_i)\in EX_1...X_nM^+$, γ can be regarded as a result of "synthesis" of the information represented by all terms $u_{1i}...u_{ni}A_i$'s. In practice, M is a set of elementary concepts, and X_1, X_2, ..., X_n are the sets associated to the concepts formed in M. For example, let X be a set of persons and M be a set of concepts such as "*male*", "*female*", "*old*", "*tall*", "*high salary*", "*black hair persons* ", "*white hair persons* ",...,etc. For $\sum_{i\in I}(u_iA_i)\in EXM^+$, every term u_iA_i, $i\in I$, expresses that the persons in set $u_i\subset X$ satisfy some "condition" described by the concepts in $A_i\subset M$. The AFS theory supports the studies on how to convert the information represented by the elements of $EX_1...X_nM^+$ in the training examples and databases into the membership functions and their fuzzy logic operations.

Proposition 5.2. *Let* $X_1,...X_n,M$ *be n+1 non-empty sets,* $EX_1...X_nM$ *and* $EX_1...X_hM$ *be* EI^{n+1}, EI^{h+1} *algebra,* $1<h<n$. $\forall\sum_{i\in I}(u_{1i}...u_{ni}A_i)\in EX_1...X_nM$, *if*

$$p\left[\sum_{i\in I}(u_{1i}...u_{ni}A_i)\right]=\sum_{i\in I}(u_{1i}...u_{hi}A_i),$$

then p *is a homomorphism from lattice* $(EX_1...X_nM,\vee,\wedge)$ *to lattice* $(EX_1...X_hM,\vee,\wedge)$.

Proof. It can be directly proved by making use of Definition 5.2 and Theorem 5.1. □

5.1.2 $E^{\#}I^n$ Algebras

In this section, we introduce $E^{\#}I^n$ algebra which has different algebraic structures from the EI^n algebra and can be applied to represent the degrees of a sample belonging to the fuzzy concepts in EM.

Definition 5.3. Let $X_1, X_2, ..., X_n$ be n non-empty sets. A binary relation $R^{\#}$ on the set

$$EX_1X_2...X_n^+ = \left\{ \sum_{i\in I} a_{1i}...a_{ni} \mid a_{ri} \in 2^{X_r}, r=1,...,n,\ I\ is\ a\ non-empty\ indexing\ set \right\}$$

is defined as follows: for $\sum_{i\in I} a_{1i}a_{2i}...a_{ni}$, $\sum_{j\in J} b_{1j}b_{2j}...b_{nj} \in EX_1X_2...X_n^+$,
$(\sum_{i\in I} a_{1i}a_{2i}...a_{ni}) R^{\#} (\sum_{j\in J} b_{1j}b_{2j}...b_{nj}) \Leftrightarrow$
(i) $\forall a_{1i}a_{2i}...a_{ni}$ $(i \in I)$, $\exists b_{1h}b_{2h}...b_{nh}$ $(h \in J)$ such that $a_{ri} \subseteq b_{rh}, r = 1,2,...n$;
(ii) $\forall b_{1j}b_{2j}...b_{nj}$ $(j \in J)$, $\exists a_{1k}a_{2k}...a_{nk}$ $(k \in I)$ such that $b_{rj} \subseteq a_{rk}$, $r = 1,2,...n$.

It is evident that $R^{\#}$ is an equivalence relation on $EX_1X_2...X_n^+$. We denote $EX_1X_2...X_n^+/R^{\#}$ (i.e. the quotient set) as $E^{\#}X_1X_2...X_n$. $\sum_{i\in I} a_{1i}a_{2i}...a_{ni} = \sum_{j\in J} b_{1j}b_{2j}...b_{nj}$ implies that $\sum_{i\in I} a_{1i}a_{2i}...a_{ni}$ and $\sum_{j\in J} b_{1j}b_{2j}...b_{nj}$ are equivalent under the equivalence relation $R^{\#}$ and the membership degrees represented by them are equal.

Proposition 5.3. *Let $X_1, X_2, ..., X_n$ be n non-empty sets. $\sum_{i\in I} a_{1i}a_{2i}...a_{ni} \in E^{\#}X_1X_2...X_n$, if $a_{ru} \subseteq a_{rv}$, $r = 1,2,...,n$, $u, v \in I$, $u \neq v$, then*

$$\sum_{i\in I} a_{1i}a_{2i}...a_{ni} = \sum_{i\in I-\{u\}} a_{1i}a_{2i}...a_{ni}.$$

Proof. It can be verified by using Definition 5.3. □

Theorem 5.2. *Let $X_1, X_2, ..., X_n$ be n non-empty sets. $(E^{\#}X_1X_2...X_n, \vee, \wedge)$ forms a completely distributive lattice under the binary compositions $\vee$ and $\wedge$ defined as follows:*

$$\sum_{i\in I} a_{1i}a_{2i}...a_{ni} \vee \sum_{j\in J} b_{1j}b_{2j}...b_{nj} = \sum_{k\in I\sqcup J} c_{1k}c_{2k}...c_{nk} \triangleq \sum_{i\in I} a_{1i}a_{2i}...a_{ni} + \sum_{j\in J} b_{1j}b_{2j}...b_{nj} \tag{5.3}$$

$$\sum_{i\in I} a_{1i}a_{2i}...a_{ni} \wedge \sum_{j\in J} b_{1j}b_{2j}...b_{nj} = \sum_{i\in I, j\in J} (a_{1i}\cap b_{1j})(a_{2i}\cap b_{2j})...(a_{ni}\cap b_{nj}) \tag{5.4}$$

where $\forall k \in I \sqcup J$, $c_{rk} = a_{rk}$, $r = 1,...,n$, if $k \in I$ and $c_{rk} = b_{rk}$, $r = 1,...,n$, if $k \in J$.

Proof. We just prove it for $E^{\#}X$. The proofs for $E^{\#}X_1X_2...X_n$ are similar and remain as exercises. First, we prove that $\vee$, $\wedge$ are binary compositions. Let $\sum_{i\in I_1} a_{1i} = \sum_{i\in I_2} a_{2i}$, $\sum_{j\in J_1} b_{1j} = \sum_{j\in J_2} b_{2j} \in E^{\#}X$. (5.3) can be directly verified Definition 5.3. By (5.4),we have

$$\sum_{i\in I_1} a_{1i} \wedge \sum_{j\in J_1} b_{1j} = \sum_{i\in I_1, j\in J_1} a_{1i} \cap b_{1j},$$
$$\sum_{i\in I_2} a_{2i} \wedge \sum_{j\in J_2} b_{2j} = \sum_{i\in I_2, j\in J_2} a_{2i} \cap b_{2j}.$$

Since $\sum_{i\in I_1} a_{1i} = \sum_{i\in I_2} a_{2i}, \sum_{j\in J_1} b_{1j} = \sum_{j\in J_2} b_{2j}$, hence for any $a_{1i} \cap b_{1j}, i \in I_1, j \in J_1$, there exist a_{2k}, $b_{2l}, k \in I_2, l \in J_2$, such that $a_{1i} \subseteq a_{2k}, b_{1j} \subseteq b_{2l}$. Therefore $a_{1i} \cap b_{1j} \subseteq a_{2k} \cap b_{2l}$. Similarly, for any $a_{2i} \cap b_{2j}, i \in I_2, j \in J_2$, there exist a_{1q}, $b_{1e}, q \in I_1, e \in J_1$, such that $a_{2i} \cap b_{2j} \subseteq a_{1q} \cap b_{1e}$. This implies that $\sum_{i\in I_1, j\in J_1} a_{1i} \cap b_{1j} = \sum_{i\in I_2, j\in J_2} a_{2i} \cap b_{2j}$ and $\wedge$ is a binary composition.

Theorem 2.4 states that two binary compositions satisfying the condition L1-L4 of Theorem 2.3 are lattice operations. For any $\sum_{i\in I} a_i$, $\sum_{j\in J} b_j$, $\sum_{k\in K} c_k \in E^{\#}X$, we can directly verify that $\vee$, $\wedge$ satisfy L1-L3 of Theorem 2.3 by the definitions.

In what follows, we prove that $\vee$, $\wedge$ satisfy L4 of Theorem 2.3. By Proposition 5.3, we have

$$\begin{aligned}
(\sum_{i\in I} a_i \vee \sum_{j\in J} b_j) \wedge \sum_{i\in I} a_i &= \sum_{i,j\in I} a_i \cap a_j + \sum_{i\in I, j\in J} a_i \cap b_j \\
&= \sum_{i\in I} a_i + \sum_{i\in I, j\in J} a_i \cap b_j = \sum_{i\in I} a_i. \\
(\sum_{i\in I} a_i \wedge \sum_{j\in J} b_j) \vee \sum_{i\in I} a_i &= \sum_{i\in I, j\in J} a_i \cap b_j + \sum_{i\in I} a_i = \sum_{i\in I} a_i.
\end{aligned}$$

Therefore $\vee$, $\wedge$ satisfy L1-L4 of Theorem 2.3 and $(E^{\#}X, \vee, \wedge)$ is a lattice. $\sum_{i\in I} a_i \geq \sum_{j\in J} b_j \Leftrightarrow \sum_{i\in I} a_i \vee \sum_{j\in J} b_j = \sum_{i\in I} a_i$. This implies that $\forall b_j, (j \in J), \exists a_k, (k \in I)$ such that $b_j \subseteq a_k$.

Next, we prove that $(E^{\#}X, \vee, \wedge)$ is a complete lattice. Let $\sum_{j\in I_i} a_{ij} \in E^{\#}X, i \in I$. We prove that $\bigvee_{i\in I}(\sum_{j\in I_i} a_{ij})$, $\bigwedge_{i\in I}(\sum_{j\in I_i} a_{ij}) \in E^{\#}X$. It is obvious that

$$\begin{aligned}
\sum_{j\in I_i} a_{ij} &\leq \sum_{i\in I} \sum_{j\in I_i} a_{ij}, \forall i \in I, \\
\sum_{j\in I_i} a_{ij} &\geq \sum_{f\in\prod_{i\in I} I_i} \bigcap_{i\in I} a_{if(i)}, \forall i \in I.
\end{aligned}$$

For $\sum_{u\in U} b_u \in E^{\#}X$, if $\sum_{j\in I_i} a_{ij} \leq \sum_{u\in U} b_u, \forall i \in I$, then $\forall a_{i_0 j_0}, i_0 \in I, j_0 \in J_{i_0}$, there exists $u_0 \in U$ such that $a_{i_0 j_0} \subseteq b_{u_0}$. Therefore $\sum_{i\in I} \sum_{j\in I_i} a_{ij} \leq \sum_{u\in U} b_u$. This implies that

$$\bigvee_{i\in I}(\sum_{j\in I_i} a_{ij}) = \sum_{i\in I} \sum_{j\in I_i} a_{ij} \in E^{\#}X. \tag{5.5}$$

For $\sum_{u\in U} b_u \in E^{\#}X$, if $\sum_{j\in I_i} a_{ij} \geq \sum_{u\in U} b_u, \forall i \in I$, then $\forall b_{u_0}, u_0 \in U$ and $\forall i_0 \in I$, there exists $j_{i_0} \in I_{i_0}$ such that $b_{u_0} \subseteq a_{i_0 j_{i_0}}$. This implies that there exists $f_{u_0} \in \prod_{i\in I} I_i$, where $f_{u_0}(i_0) = j_{i_0}, \forall i_0 \in I$, such that $b_{u_0} \subseteq \bigcap_{i\in I} a_{i f_{u_0}(i)}$. We have

$$\sum_{f\in\prod_{i\in I} I_i} \bigcap_{i\in I} a_{if(i)} \geq \sum_{u\in U} b_u$$

and

$$\bigwedge_{i\in I}(\sum_{j\in I_i} a_{ij}) = \sum_{f\in\prod_{i\in I} I_i} \bigcap_{i\in I} a_{if(i)} \in E^{\#}X. \tag{5.6}$$

Therefore $(E^{\#}X, \vee, \wedge)$ is a complete lattice.

Now, we prove that $(E^{\#}X, \vee, \wedge)$ is a completely distributive lattice. Let $\lambda_{ij} = \sum_{u\in U_{ij}} a_u^{ij} \in E^{\#}X$, $i \in I$, $j \in J_i$, U_{ij} is a non-empty indexing set. It is obvious that for any $f \in \prod_{i\in I} J_i, \forall k \in I$, since $f(k) \in J_k$, hence

$$\bigwedge_{i\in I} \lambda_{if(i)} \leq \lambda_{kf(k)} \leq \bigvee_{j\in J_k} \lambda_{kj}.$$

Since $\forall k \in I, \forall f \in \prod_{i\in I} J_i, \bigwedge_{i\in I} \lambda_{if(i)} \leq \bigvee_{j\in J_k} \lambda_{kj}$, hence for any $f \in \prod_{i\in I} J_i$,

$$\bigwedge_{i\in I} \lambda_{if(i)} \leq \bigwedge_{k\in I}(\bigvee_{j\in J_k} \lambda_{kj}).$$

Therefore

$$\bigvee_{f\in\prod_{i\in I} J_i} (\bigwedge_{i\in I} \lambda_{if(i)}) \leq \bigwedge_{i\in I}(\bigvee_{j\in J_i} \lambda_{ij}). \tag{5.7}$$

By (5.6) and (5.5), we have

$$\begin{aligned} \bigvee_{f\in\prod_{i\in I} J_i} (\bigwedge_{i\in I} \lambda_{if(i)}) &= \bigvee_{f\in\prod_{i\in I} J_i} (\bigwedge_{i\in I}(\sum_{u\in U_{if(i)}} a_u^{if(i)})) \\ &= \sum_{f\in\prod_{i\in I} J_i} (\sum_{h\in\prod_{i\in I} U_{if(i)}} \bigcap_{i\in I} a_{h(i)}^{if(i)}). \end{aligned}$$

$$\begin{aligned} \bigwedge_{i\in I}(\bigvee_{j\in J_i} \lambda_{ij}) &= \bigwedge_{i\in I}(\sum_{j\in J_i} \sum_{u\in U_{ij}} a_u^{ij}) \\ &= \bigwedge_{i\in I}(\sum_{u\in\bigsqcup_{k\in J_i} U_{ik}} e_u^i) \\ &= \sum_{g\in\prod_{i\in I}(\bigsqcup_{k\in J_i} U_{ik})} \bigcap_{i\in I} e_{g(i)}^i. \end{aligned}$$

where for any $u \in \bigsqcup_{k\in J_i} U_{ik}$, $e_u^i = a_u^{ij}$, when $u \in U_{ij}$. For any $g_0 \in \prod_{i\in I}(\bigsqcup_{k\in J_i} U_{ik})$, since $g_0(i) \in \bigsqcup_{k\in J_i} U_{ik}, i \in I$, hence for any $i \in I$, there exists $k_i \in J_i$ such that $g_0(i) \in U_{ik_i}$. This implies that if we define $f_0(i) = k_i \in J_i$, $i \in I$, then $f_0 \in \prod_{i\in I} J_i$, $g_0(i) \in U_{if_0(i)}$, $g_0 \in \prod_{i\in I} U_{if_0(i)}$ and $e_{g_0(i)}^i = a_{g_0(i)}^{if_0(i)}$, for any $i \in I$. Therefore for any $g_0 \in \prod_{i\in I}(\bigsqcup_{k\in J_i} U_{ik})$, there exist $f_0 \in \prod_{i\in I} J_i$ such that $g_0 \in \prod_{i\in I} U_{if_0(i)}$ and

$$\bigcap_{i\in I} e_{g_0(i)}^i = \bigcap_{i\in I} a_{g_0(i)}^{if_0(i)}.$$

This implies that

$$\bigvee_{f\in\prod_{i\in I} J_i} (\bigwedge_{i\in I} \lambda_{if(i)}) \geq \bigwedge_{i\in I}(\bigvee_{j\in J_i} \lambda_{ij}).$$

By (5.7), we have

$$\bigwedge_{i\in I}(\bigvee_{j\in J_i}\lambda_{ij}) = \bigvee_{f\in\prod_{i\in I}J_i}(\bigwedge_{i\in I}\lambda_{if(i)}).$$

Therefore $(E^{\#}X,\vee,\wedge)$ is a completely distributive lattice according to Theorem 2.28 □

$(E^{\#}X_1X_2...X_n,\vee,\wedge)$ is called a $E^{\#}I^n$ *algebra* over $X_1,X_2,...,X_n$. Although $E^{\#}I^n$ algebra and EI^n algebra are similar in many ways, they are not dual lattices. $E^{\#}I$ algebra is a lattice which can be applied to represent the membership degrees of the fuzzy sets in EM. We also can prove that $E^{\#}I$ algebra and EI algebra are not isomorphism. The proofs of the following properties remain as exercises: for any $\sum_{i\in I}a_i\in E^{\#}X$,

1. $\varnothing\vee\sum_{i\in I}a_i=\sum_{i\in I}a_i$, $\varnothing\wedge\sum_{i\in I}a_i=\varnothing$;
2. $X\vee\sum_{i\in I}a_i=X, X\wedge\sum_{i\in I}a_i=\sum_{i\in I}a_i$.

In $E^{\#}I$ algebra $E^{\#}X$, $\varnothing$ is the minimum element and X is the maximum element.

Proposition 5.4. *Let $X_1,...,X_n,M$ be n+1 non-empty sets. $EX_1...X_nM$ and $E^{\#}X_1...X_n$ be EI^{n+1} and $E^{\#}I^n$ algebra, respectively. $\forall\sum_{i\in I}(u_{1i}...u_{ni}A_i)\in EX_1...X_nM$, if*

$$p\left[\sum_{i\in I}(u_{1i}...u_{ni}A_i)\right]=\sum_{i\in I}u_{1i}...u_{ni},$$

then p is a homomorphism from $(EX_1...X_nM,\vee,\wedge)$ to $(E^{\#}X_1...X_n,\vee,\wedge)$.

Proof. First, we prove that p is a map. α, $\beta\in EX_1...X_nM$. Suppose $\alpha=\beta$. By the equivalence relations R and $R^{\#}$ defined in Definition 5.3 and Definition 5.2, one can verify that $p(\alpha)=p(\beta)$ in $E^{\#}X_1...X_n$.

Next, we prove that p is a homomorphism. For any $\sum_{i\in I}(u_{1i}...u_{ni}A_i)$, $\sum_{j\in J}(v_{1j}...v_{nj}B_j)\in EX_1...X_nM$,

$$\begin{aligned}
p\left[\sum_{i\in I}(u_{1i}...u_{ni}A_i)\vee\sum_{j\in J}(v_{1j}...v_{nj}B_j)\right] &= \sum_{i\in I}u_{1i}...u_{ni}+\sum_{j\in J}v_{1j}...v_{nj}\\
&= p\left[\sum_{i\in I}(u_{1i}...u_{ni}A_i)\right]\vee p\left[\sum_{j\in J}(v_{1j}...v_{nj}B_j)\right].\\
p\left[\sum_{i\in I}(u_{1i}...u_{ni}A_i)\wedge\sum_{j\in J}(v_{1j}...v_{nj}B_j)\right] &= p\left[\sum_{i\in I,j\in J}(u_{1i}\cap v_{1j}...u_{ni}\cap v_{nj}A_i\cup B_j)\right]\\
&= \sum_{i\in I,j\in J}u_{1i}\cap v_{1j}...u_{ni}\cap v_{nj}\\
&= p\left[\sum_{i\in I}(u_{1i}...u_{ni}A_i)\right]\wedge p\left[\sum_{j\in J}(v_{1j}...v_{nj}B_j)\right].
\end{aligned}$$

□

5.2 AFS Algebra Representation of Membership Degrees

Generic fuzzy sets, L-fuzzy sets or Boolean subsets of some universe of discourse X are various representation of fuzzy concepts or Boolean concepts. We regard fuzzy sets or L-fuzzy sets as different representing forms of fuzzy concepts and fuzzy sets of X mean all kinds of representing forms for fuzzy concepts. The fuzzy sets and Boolean subsets on X can be described in the following way.

For a fuzzy set ζ on universe of discourse X, any $x \in X$, either x belongs to ζ at some degree or does not belong to ζ at all, while for a Boolean subset A of X, any $x \in X$, either x belongs to A or does not belong to A at all.

In what follows, we introduce three types of L-fuzzy sets proposed in [7] whose membership degrees are in lattices EI^2, EI^3 and $E^{\#}I$ algebras, respectively.

Theorem 5.3. *([7]) Let (M,τ,X) be an AFS structure. $U \subseteq X$, $A \subseteq M$. We introduce the notation*

$$A^{\tau}(U) = \{y \mid y \in X, \tau(x,y) \supseteq A \text{ for any } x \in U\}. \tag{5.8}$$

For any given $x \in X$, if we define a mapping $\phi_x : EM \to EXM$ as follows: for any $\sum_{i\in I}(\prod_{m\in A_i} m) \in EM$,

$$\phi_x\left(\sum_{i\in I}(\prod_{m\in A_i} m)\right) = \sum_{i\in I} A_i^{\tau}(x)A_i \in EXM, \tag{5.9}$$

then ϕ_x is a homomorphism from lattice $(EM,\vee,\wedge)$ to lattice $(EXM,\vee,\wedge)$, where $A_i^{\tau}(\{x\})$ is simply denoted as $A_i^{\tau}(x)$ defined as (4.27).

For $A \subseteq M$, $x \in X, A^{\tau}(x) = \{y|y \in X, \tau(x,y) \supseteq A\}$, which is the subset of X, for any $y \in A^{\tau}(x)$, the degree of x belonging to the fuzzy concept $\prod_{m\in A} m \in EM$ is larger than or equal to that of y since $\tau(x,y) \supseteq A$. By Theorem 5.3, we know that for any given fuzzy concept $\sum_{i\in I}(\prod_{m\in A_i} m) \in EM$, we form a map $\sum_{i\in I}(\prod_{m\in A_i} m) : X \to EXM$ defined as follows: for any $x \in X$,

$$(\sum_{i\in I}(\prod_{m\in A_i} m))(x) = \sum_{i\in I} A_i^{\tau}(x)A_i \in EXM. \tag{5.10}$$

Since $(EXM,\vee,\wedge)$ is a lattice, hence the map $\sum_{i\in I}(\prod_{m\in A_i} m)$ is a L-fuzzy set (with membership degrees in lattice EXM). In this way, $\sum_{i\in I}(\prod_{m\in A_i} m)$ is L-fuzzy set on X and the membership degree of x ($x \in X$) belonging to fuzzy set $\sum_{i\in I}(\prod_{m\in A_i} m)$ is $\sum_{i\in I} A_i^{\tau}(x)A_i \in EXM$. If $\sum_{i\in I} A_i^{\tau}(x)A_i \geq \sum_{i\in I} A_i^{\tau}(y)A_i$ in lattice EXM, then the degree of x belonging to fuzzy set $\sum_{i\in I}(\prod_{m\in A_i} m)$ is larger than or equal to that of y. For fuzzy sets $\alpha = \sum_{i\in I}(\prod_{m\in A_i} m), \beta = \sum_{j\in J}(\prod_{m\in B_j} m) \in EM$, fuzzy set $\alpha \vee \beta$ and $\alpha \wedge \beta$ are logic "*or*" and "*and*" of L-fuzzy sets α and β respectively. $'$ is the negation of the fuzzy concepts in EM. Thus $(EM,\vee,\wedge,')$ is a fuzzy logic system.

In order to utilize more information to represent the membership degrees of fuzzy sets, we introduce EI^3 algebra representation for fuzzy concepts in EM below.

Let (M,τ_1,X) be an AFS structure, where X is a universe of discourse and M is a set of simple concepts. In some case, for a given $x \in X$, and m_1, $m_2 \in M$, we can compare the degree of x belonging to m_1 with that of x belonging to m_2. In order to utilize this kind of information, we can employ another AFS structure (X, τ_2, M), where $\forall (m_1, m_2) \in M \times M$, $\tau_2(m_1, m_2) = \{x \mid x \in X, m_1 \geq_x m_2\}$, where $m_1 \geq_x m_2$ means that x belongs to concept m_1 at some degree and the degree of x belonging to m_1 is larger than or equals to the degree of x belonging to m_2. It is obvious that if the representations of membership degrees of fuzzy sets can utilize both (M, τ_1, X) and (X, τ_2, M), then the representations of fuzzy concepts will be more accurate than that given just by (M, τ_1, X). The following definition expresses the conditions under which two AFS structures capture different aspect abstractions for the same original data.

Definition 5.4. ([7]) Let X, M be sets, (M, τ_1, X) and (X, τ_2, M) be AFS structures. $m \in M$, $x \in X$. If τ_1, τ_2 satisfy:

1. $x \in \tau_2(m,m) \Leftrightarrow m \in \tau_1(x,x)$;
2. $x \in \tau_2(m,m) \Rightarrow \{y \mid y \in X, \tau_1(y,x) \supseteq \{m\}\} \subseteq \tau_2(m,m)$.

Then we call (M, τ_1, X) *compatible with* (X, τ_2, M) and $((M, \tau_1, X), (X, \tau_2, M))$ is called the *cognitive space*.

Theorem 5.4. *([7]) Let X and M be sets. (M, τ_1, X) is compatible with (X, τ_2, M). $\forall x \in X$, for any $\sum_{i \in I}(\prod_{m \in A_i} m) \in EM$, if we define*

$$\phi_x\left(\sum_{i \in I}\prod_{m \in A_i} m\right) = \sum_{i \in I} A_i^{\tau_1}(x)\{x\}^{\tau_2}(A_i)A_i \in EXMM, \tag{5.11}$$

then ϕ_x is a homomorphism from lattice $(EM, \vee, \wedge)$ to lattice $(EXMM, \vee, \wedge)$, where $A_i^{\tau_1}(x)$ and $\{x\}^{\tau_2}(A_i)$ are defined by (5.8).

By Theorem 5.4, we know that for any given fuzzy concept $\sum_{i \in I}\prod_{m \in A_i} m \in EM$, we get a map $\sum_{i \in I}\prod_{m \in A_i} m : X \to EXMM$ defined as follows: for any $x \in X$,

$$(\sum_{i \in I}\prod_{m \in A_i} m)(x) = \sum_{i \in I} A_i^{\tau_1}(x)\{x\}^{\tau_2}(A_i)A_i \in EXMM. \tag{5.12}$$

Since $(EXMM, \vee, \wedge)$ is a lattice, hence by Theorem 5.4, EI^3 algebra representation for fuzzy concepts in EM is obtained and $(EM, \vee, \wedge, ')$ is a fuzzy logic system for the EI^3 algebra representing fuzzy concepts. We know EI^3 algebra represented fuzzy sets in EM can utilize more information than EII algebra representations. This implies that EI^3 algebra representation is more accurate than EII algebra representation and there may be more elements in X which cannot be compared under EI^3 algebra representation than EII algebra representation. By Proposition 5.2, one can verify that if $\sum_{i \in I} A_i^{\tau_1}(x)\{x\}^{\tau_2}(A_i)A_i \leq \sum_{i \in I} A_i^{\tau_1}(y)\{y\}^{\tau_2}(A_i)A_i$ in lattice $(EXMM, \vee, \wedge)$, then $\sum_{i \in I} A_i^{\tau_1}(x)A_i \leq \sum_{i \in I} A_i^{\tau_1}(y)A_i$ in lattice $(EXM, \vee, \wedge)$. Thus for the EI^3 algebra membership degrees of x, y defined by (5.12), we have $(\sum_{i \in I}\prod_{m \in A_i} m)(x) \leq (\sum_{i \in I}\prod_{m \in A_i} m)(y)$. This implies that the more detail information is considered for the membership degrees, the more elements in universe of discourse may not be compared.

Although EII, EI^3 algebra representations of fuzzy concepts containing great information, in some case, too many pair elements in X may not be compared. The following $E^{\#}I$ algebra representations of the membership degrees for each fuzzy concept in EM may filter (eliminate) some trivial information to make more pairs of elements comparable.

By Proposition 5.4, we know for $\alpha, \beta \in EX_1...X_nM$, if $\alpha \geq \beta$, i.e., $\alpha \vee \beta = \alpha$ in lattice $(EX_1X_2...X_nM, \vee, \wedge)$, then $p(\alpha \vee \beta) = p(\alpha) \vee p(\beta) = p(\alpha)$ i.e. $p(\alpha) \geq p(\beta)$ in $(E^{\#}X_1X_2...X_n, \vee, \wedge)$. Given an AFS structure (M, τ, X) and for each $x \in X$, we have two homomorphisms $\phi_x : EM \to EXM$ and $p : EXM \to E^{\#}X$ by Theorem 5.3 and Proposition 5.4. Since both ϕ_x and p are homomorphisms, hence the composed map of ϕ_x and p, $p \circ \phi_x : EM \to E^{\#}X$ is a homomorphism from the lattice $(EM, \vee, \wedge)$ to the lattice $(E^{\#}X, \vee, \wedge)$. For each fuzzy concept $\sum_{i\in I} \prod_{m\in A_i} m \in EM$, we get another kind of L-fuzzy set representation, i.e., the $E^{\#}I$ algebra represented membership degrees as follows: $\forall x \in X$

$$(\sum_{i\in I} \prod_{m\in A_i} m)(x) = p \circ \phi_x(\sum_{i\in I} \prod_{m\in A_i} m) = p(\sum_{i\in I} A_i^{\tau}(x)A_i) = \sum_{i\in I} A_i^{\tau}(x) \in E^{\#}X. \quad (5.13)$$

By Proposition 5.4, for $x, y \in X$, if $\sum_{i\in I} A_i^{\tau}(x)A_i \leq \sum_{i\in I} A_i^{\tau}(y)A_i$ in lattice EXM, then $\sum_{i\in I} A_i^{\tau}(x) \leq \sum_{i\in I} A_i^{\tau}(y)$ in lattice $E^{\#}X$. Therefore, compared with the EII algebra representing membership degrees, $E^{\#}I$ algebra representation is finer. Although $E^{\#}I$ algebra represented membership degrees are finer than EII algebra representations, we should notice that $E^{\#}I$ algebra representations lost some original information and are not so strict as EII algebra representations. In what follows, we apply EI^3 algebra to develop another $E^{\#}I$ algebra represented L-fuzzy sets in EM by the following theorem.

Theorem 5.5. *Let M, X be sets. $M \cap X = \varnothing$. $EXMM$ is the EI^3 algebra on X, M, M and $E^{\#}(X \cup M)$ is $E^{\#}I$ algebra on $X \cup M$. For any $\sum_{i\in I} a_i e_i A_i \in EXMM$, if we define $p(\sum_{i\in I} a_i e_i A_i) = \sum_{i\in I} a_i \cup e_i$, then p is a homomorphism from the lattice $(EXMM, \vee, \wedge)$ to the lattice $(E^{\#}(X \cup M), \vee, \wedge)$.*

Proof. First, we prove that p is a map from $EXMM$ to $E^{\#}(X \cup M)$. Let $\alpha = \sum_{i\in I} a_i e_i A_i$, $\beta = \sum_{j\in J} b_j q_j B_j \in EXMM$. Suppose $\alpha = \beta$. By the equivalence relations R and $R^{\#}$ defined in Definition 5.2 and Definition 5.3, we have that $\forall i \in I$, $\exists k \in J$ such that $a_i \subseteq b_k$, $e_i \subseteq q_k$, $B_k \subseteq A_i$ and $\forall j \in J, \exists l \in I$ such that $b_j \subseteq a_l$, $q_j \subseteq e_l$, $A_l \subseteq B_j$. This implies that $\forall i \in I$, $\exists k \in J$ such that $a_i \cup e_i \subseteq b_k \cup q_k$ and $\forall j \in J$, $\exists l \in I$ such that $b_j \cup q_j \subseteq a_l \cup e_l$. Therefore

$$p\left(\sum_{i\in I} a_i e_i A_i\right) = \sum_{i\in I} a_i \cup e_i = \sum_{j\in J} b_j \cup q_j = p\left(\sum_{j\in J} b_j q_j B_j\right)$$

in $E^{\#}(X \cup M)$ and p is a map from $EXMM$ to $E^{\#}(X \cup M)$.

Next, we prove that p is a homomorphism from the lattice $(EXMM, \vee, \wedge)$ to the lattice $(E^{\#}(X \cup M), \vee, \wedge)$. For any $\sum_{i\in I} a_i e_i A_i$, $\sum_{j\in J} b_j q_j B_j \in EXMM$,

$$p\left(\sum_{i\in I} a_i e_i A_i \vee \sum_{j\in J} b_j q_j B_j\right) = \sum_{i\in I} a_i \cup e_i + \sum_{j\in J} b_j \cup q_j$$
$$= p\left(\sum_{i\in I} a_i e_i A_i) \vee p(\sum_{j\in J} b_j q_j B_j\right).$$
$$p\left(\sum_{i\in I} a_i e_i A_i \wedge \sum_{j\in J} b_j q_j B_j\right) = p\left(\sum_{i\in I, j\in J} a_i \cap b_j e_i \cap q_j A_i \cup B_j\right)$$
$$= \sum_{i\in I, j\in J} (a_i \cap b_j) \cup (e_i \cap q_j).$$

Since $M \cap X = \varnothing$, a_i, $b_j \in 2^X$, e_i, $q_j \in 2^M$, hence for any $i \in I, j \in J$,

$$(a_i \cup e_i) \cap (b_j \cup q_j) = ((a_i \cup e_i) \cap b_j) \cup ((a_i \cup e_i) \cap q_j)$$
$$= ((a_i \cap b_j) \cup (e_i \cap b_j)) \cup ((a_i \cap q_j) \cup (e_i \cap q_j))$$
$$= (a_i \cap b_j) \cup (e_i \cap q_j).$$

Hence

$$p\left(\sum_{i\in I} a_i e_i A_i \wedge \sum_{j\in J} b_j q_j B_j\right) = \sum_{i\in I, j\in J} (a_i \cap b_j) \cup (e_i \cap q_j)$$
$$= \sum_{i\in I, j\in J} (a_i \cup e_i) \cap (b_j \cup q_j)$$
$$= \sum_{i\in I} (a_i \cup e_i) \wedge \sum_{j\in J} (b_j \cup q_j)$$
$$= p\left(\sum_{i\in I} a_i e_i A_i\right) \wedge p\left(\sum_{j\in J} b_j q_j B_j\right). \qquad \square$$

Similarly, we can verify that for $\alpha, \beta \in EXMM$, if $\alpha \le \beta$, in the lattice $(EXMM, \vee, \wedge)$ then $p(\alpha) \le p(\beta)$ in the lattice $(E^{\#}(X \cup M), \vee, \wedge)$. Given two compatible AFS structure (M, τ_1, X), (X, τ_2, M) and for each $x \in X$, we have two homomorphisms $\phi_x : EM \to EXMM$, $p : EXMM \to E^{\#}(X \cup M)$ by Theorem 5.4 and Theorem 5.5. The composed map $p \circ \phi_x$ of ϕ_x and p is a homomorphism form the lattice $(EM, \vee, \wedge)$ to the lattice $(E^{\#}(X \cup M), \vee, \wedge)$. For each fuzzy concept $\sum_{i\in I} \prod_{m\in A_i} m \in EM$, the $E^{\#}I$ algebra representing membership degree of x is defined as follows:

$$(\sum_{i\in I} \prod_{m\in A_i} m)(x) = p \circ \phi_x(\sum_{i\in I} \prod_{m\in A_i} m) = \sum_{i\in I} \left[A_i^{\tau_1}(x) \cup \{x\}^{\tau_2}(A_i)\right] \in E^{\#}(X \cup M). \quad (5.14)$$

One can verify that for fuzzy concepts $\alpha, \beta \in EM$, if $\alpha \ge \beta$ in lattice EM, then $\forall x \in X$, $\alpha(x) \ge \beta(x)$ in $E^{\#}I$ algebra.

In what follows, we study the norm of AFS algebra by which we can obtain coherence membership degrees in [0,1] interval from various AFS algebra representing fuzzy sets.

5.3 Norm of AFS Algebra and Membership Functions of Fuzzy Concepts

In this section, we study the norm of the $EI^n, (n > 1), E^\# I$ algebra in order to convert the AFS algebra -represented membership degrees to the [0,1] interval.

Definition 5.5. Let L be a set and $(L, \vee, \wedge)$ be a lattice. The map $||.|| : L \to [0,1]$ is called a *fuzzy norm of the lattice L* if $||.||$ satisfies the following conditions: for any $x, y \in L$,

1. if $x \leq y$, then $||x|| \leq ||y||$;
2. $||x \wedge y|| \leq \min\{||x||, ||y||\}$, $||x \vee y|| \geq \max\{||x||, ||y||\}$.

In what follows, we propose a special family of measure by which $EI^n, (n > 1), E^\# I$ algebra becomes lattices with norms. Such that we can convert AFS algebra represented membership degrees to [0,1] interval and at great extent to preserve the information contained in the EII representations.

Definition 5.6. (Continuous case) Let X be a set, $X \subseteq R^n$. $\rho : X \to R^+ = [0, \infty)$. ρ is integrable on X under Lebesgue measure and $0 < \int_X \rho d\mu < \infty$. S ($S \subseteq 2^X$) is the set of Borel sets in X. For all $A \in S$, we define a measure $\mathscr{M}$ over S,

$$\mathscr{M}(A) = \frac{\int_A \rho d\mu}{\int_X \rho d\mu}. \tag{5.15}$$

(Discrete case) Let X be a set, S is a σ-algebra over X. $\rho : X \to R^+ = [0, \infty)$. $0 < \sum_{x \in X} \rho(x) < \infty$. For any $A \in S$, the measure $\mathscr{M}$ over σ-algebra S on X is defined as follows,

$$\mathscr{M}(A) = \frac{\sum_{x \in A} \rho(x)}{\sum_{x \in X} \rho(x)}. \tag{5.16}$$

By Definition 1.41, we can verify that $\mathscr{M}$ defined in Definition 5.6 is a measure over X for each function ρ. Indeed for each simple concept ζ on X, according to the distributions of original data and the interpretation of ζ, ζ corresponds to a function $\rho_\zeta : X \to R^+ = [0, \infty)$, by which we can obtain the norm of AFS algebra using measure m_ζ provided by Definition 5.6.

Proposition 5.5. *Let $X_1, ...X_n, M$ be $n+1$ non-empty sets, $EX_1...X_nM$ be EI^{n+1} algebra over $X_1, ..., X_n, M$ and M be a finite set of simple concepts, S_r be a σ-algebra over X_r, $r = 1, 2, ..., n$. Let $\sigma(EX_1...X_nM)$, a subset of $EX_1...X_nM$, be defined as follows:*

$$\sigma(EX_1...X_nM) = \left\{ \sum_{i \in I} (u_{1i}...u_{ni}A_i) \in EX_1...X_nM \mid u_{ri} \in S_r,\ r = 1,\ ...,\ n,\ \forall i \in I \right\}. \tag{5.17}$$

Then $(\sigma(EX_1...X_nM), \vee, \wedge)$ is a sublattice of $(EX_1...X_nM, \vee, \wedge)$, i.e., $\xi \vee \eta$, $\xi \wedge \eta \in \sigma(EX_1...X_nM)$ for all $\xi, \eta \in \sigma(EX_1...X_nM)$.

Proposition 5.6. *Let $X_1,...X_n,M$ be $n+1$ non-empty sets, $EX_1...X_nM$ be EI^{n+1} algebra over $X_1,...,X_n,M$ and M be a finite set of simple concepts, S_r be a σ-algebra over X_r, $r=1,2,...,n$. For each simple concept $\zeta \in M$, let $\mathscr{M}_\zeta$ be the measure defined by Definition 5.6 for ρ_ζ. Then the map $||.|| : \sigma(EX_1...X_nM) \to [0,1]$ defined as follows is a fuzzy norm of the lattice $(\sigma(EX_1...X_nM),\vee,\wedge)$: for any $\sum_{i\in I}(u_{1i}...u_{ni}A_i) \in \sigma(EX_1...X_nM)$,*

$$||\sum_{i\in I}(u_{1i}...u_{ni}A_i)|| = \sup_{i\in I}\left(\inf_{m\in A_i}\prod_{1\le r\le n}\mathscr{M}_m(u_{ri})\right) \in [0,1]. \tag{5.18}$$

or

$$||\sum_{i\in I}(u_{1i}...u_{ni}A_i)|| = \sup_{i\in I}\left(\prod_{m\in A_i,1\le r\le n}\mathscr{M}_m(u_{ri})\right) \in [0,1]. \tag{5.19}$$

For $u_{1k}...u_{nk}A_k$, $k\in I$,if $A_k=\varnothing$, define

$$\inf_{m\in A_k}\prod_{1\le r\le n}\mathscr{M}_m(u_{rk}) = \prod_{m\in A_i,1\le r\le n}\mathscr{M}_m(u_{ri}) = \prod_{1\le r\le n}\left[\max_{m\in M}\{\mathscr{M}_m(u_{rk})\}\right]. \tag{5.20}$$

Proof. We prove that what (5.18) defines is a fuzzy norm. In virtue of (5.18) and (5.20), for any $A\subseteq M$, $A\neq\varnothing, \forall u_k \subseteq X_k$, $k=1,...,n$, we have

$$||(u_1...u_n)A|| = \inf_{m\in A}\prod_{1\le r\le n}\mathscr{M}_m(u_r) \le \prod_{1\le r\le n}\left[\max_{m\in M}\{\mathscr{M}_m(u_r)\}\right] = ||(u_1...u_n)\varnothing||.$$

Let $\xi=\sum_{i\in I}(u_{1i}...u_{ni}A_i), \eta=\sum_{j\in J}(v_{1j}...v_{nj}B_j)\in\sigma(EX_1...X_nM)$. $||\xi\vee\eta|| = \max\{||\xi||,||\eta||\}$ can be directly proved by the (5.18). Now we prove $||\xi\vee\eta|| \le \min\{||\xi||,||\eta||\}$. By Theorem 5.1, we have

$$\begin{aligned}||\sum_{i\in I}(u_{1i}...u_{ni}A_i)\wedge\sum_{j\in J}(v_{1j}...v_{nj}B_j)|| &= ||\sum_{i\in I,j\in J}\left[(u_{1i}\cap v_{1j}...u_{ni}\cap v_{nj})(A_i\cup B_j)\right]||,\\ &= \sup_{i\in I,j\in J}\{||(u_{1i}\cap v_{1j}...u_{ni}\cap v_{nj})(A_i\cup B_j)||\}.\end{aligned}$$

For any $i\in I, j\in J$, we have

$$\begin{aligned}||(u_{1i}\cap v_{1j}...u_{ni}\cap v_{nj})(A_i\cup B_j)|| &= \inf_{m\in A_i\cup B_j}\prod_{1\le r\le n}\mathscr{M}_m(u_{ri}\cap v_{rj})\\ &\le \inf_{m\in A_i\cup B_j}\prod_{1\le r\le n}\mathscr{M}_m(u_{ri}) \le \inf_{m\in A_i}\prod_{1\le r\le n}\mathscr{M}_m(u_{ri})\\ &\le ||\sum_{i\in I}(u_{1i}...u_{ni}A_i)||. \end{aligned}\tag{5.21}$$

Similarly, we can show that for any $i\in I, j\in J$,

$$||(u_{1i}\cap v_{1j}...u_{ni}\cap v_{nj})(A_i\cup B_j)|| \le ||\sum_{j\in J}(v_{1j}...v_{nj}B_j)||. \tag{5.22}$$

By (5.21) and (5.22), we have

$$||\sum_{i\in I}(u_{1i}...u_{ni}A_i)\wedge\sum_{j\in J}(v_{1j}...v_{nj}B_j)|| = \sup_{i\in I,j\in J}\{||(u_{1i}\cap v_{1j}...u_{ni}\cap v_{nj})(A_i\cup B_j)||\}$$
$$\leq \min\{||\sum_{i\in I}(u_{1i}...u_{ni}A_i)||, ||\sum_{j\in J}(v_{1j}...v_{nj}B_j)||\}.$$

Suppose that $\xi \leq \eta$ in lattice $\sigma(EX_1...X_nM)$. By Theorem 5.1 we know that $\forall(u_{1i}...u_{ni})A_i$, $(i\in I)$, $\exists(v_{1k}...v_{nk})B_k$, $(k\in J)$ such that $u_{ri}\subseteq v_{rk}, A_i\supseteq B_k$. By (5.18), one knows that $||(u_{1i}...u_{ni})A_i|| \leq ||(v_{1k}...v_{nk})B_k||$. This implies that $||\xi|| \leq ||\eta||$. Therefore $||.||$ defined by (5.18) satisfies Definition 5.5 and it is a fuzzy norm. □

Proposition 5.7. *Let X be a set and $E^{\#}X$ be the $E^{\#}I$ algebra over X. S is the σ-algebra on X and $\mathscr{M}$ is a measure on S with $\mathscr{M}(X)=1$. Let $\sigma(E^{\#}X)$, a subset of $E^{\#}X$, be defined as follows:*

$$\sigma(E^{\#}X) = \{\sum_{i\in I}a_i \in E^{\#}X \mid a_i\in S,\ \forall i\in I\}. \tag{5.23}$$

Then the following assertions hold:

1. *$(\sigma(E^{\#}X),\vee,\wedge)$ is a sublattice of the lattice $(E^{\#}X,\vee,\wedge)$;*
2. *the map $||.|| : \sigma(E^{\#}X)\rightarrow[0,1]$ defined as follows is a fuzzy norm of the lattice $(\sigma(E^{\#}X),\vee,\wedge)$: for any $\gamma=\sum_{i\in I}a_i\in\sigma(E^{\#}X)$,*

$$||\gamma|| = \sup_{i\in I}\{\mathscr{M}(a_i)\}. \tag{5.24}$$

Proof. Let $\xi=\sum_{i\in I}a_i, \eta=\sum_{j\in J}b_j\in\sigma(E^{\#}X)$. $||\xi\vee\eta|| = max\{||\xi||,||\eta||\}$ can be directly proved by the (5.24). Now we prove $||\xi\vee\eta|| \leq min\{||\xi||,||\eta||\}$. By Theorem 5.2, we have

$$||\sum_{i\in I}a_i\wedge\sum_{j\in J}b_j)|| = ||\sum_{i\in I,j\in J}(a_i\cap b_j)|| = \sup_{i\in I,j\in J}\{||a_i\cap b_j||\}$$
$$\leq \sup_{i\in I,j\in J}\{\min\{||a_i||,||b_j||\}\} \leq \min\left\{||\sum_{i\in I}a_i||, ||\sum_{j\in J}b_j||\right\}.$$

Suppose $\xi\leq\eta$ in lattice $\sigma(E^{\#}X)$. By Theorem 5.2 we know that $\forall a_i$, $(i\in I)$, $\exists b_k$, $(k\in J)$ such that $a_i\subseteq b_k$. By (5.24), one knows that $||A_i||\leq||B_k||$. This implies that $||\sum_{i\in I}a_i||\leq||\sum_{j\in J}b_j||$. Therefore $||.||$ defined by (5.24) satisfies Definition 5.5 and it is a fuzzy norm. □

In general, the weight function ρ_ξ defined by Definition 4.8 can be applied to construct the measure defined by Definition 5.6 for a simple concept ξ. For a fuzzy concept $\zeta=\sum_{i\in I}\prod_{m\in A_i}m\in EM$,the membership function $\mu_\zeta(x)$ can be obtained by $||.||$ a fuzzy norm of AFS algebra as follows:

$$\mu_\zeta(x) = ||(\sum_{i\in I}\prod_{m\in A_i}m)(x)||, \forall x\in X, \tag{5.25}$$

where $(\sum_{i\in I}\prod_{m\in A_i} m)(x)$ is the AFS algebra represented membership degree defined by one of (5.10), (5.12),(5.13) and (5.14).

Theorem 5.6. *Let X be a set and M be a set of simple concepts on X. Let (M,τ,X) be an AFS structure and $||.||$ be a fuzzy norm of an AFS algebra. For any fuzzy concept $\xi \in EM$, let $\xi(x)$ is the AFS algebra representating membership degree by any one of (5.10), (5.12),(5.13) and (5.14). Then $\{\mu_\xi(x) \mid \xi \in EM\}$ is the set of coherence membership functions of the AFS fuzzy logic system $(EM,\vee,\wedge,')$ and the AFS structure (M,τ,X).*

The following example shows how to construct the coherence membership functions of the fuzzy concepts in EM by using the norm of the AFS algebra.

Example 5.1. Let $X = \{x_1,x_2,...,x_{10}\}$ be a set of 10 people and their features (attributes) which are described by real numbers (age, height, weight, salary, estate), Boolean values (gender) and the ordered relations (hair black, hair white, hair yellow), see Table 5.1; there the number i in the "*hair color*" columns which corresponds to some $x \in X$ implies that the hair color of x has ordered ith following our perception of the color by our intuitive perception. Let $M = \{m_1,m_2, ..., m_{10}\}$ be the set of fuzzy or Boolean concepts on X and each $m \in M$ associate to a single feature. Where m_1 : *"old people"*, m_2 : *"tall people"*, m_3 : *"heavy people"*, m_4 : *"high salary"*, m_5 : *"more estate"*, m_6 :*"male"*, m_7 : *"female"*, m_8 : *"black hair people"*, m_9 : *"white hair people"*, m_{10} : *"yellow hair people"*.

For each numerical attribute $m \in M$, $\rho_m(x)$ is equal to the value of x on the attribute, for each Boolean attribute m and each attribute m described by a sub-preference relation $\rho_m(x) = 1 \Leftrightarrow x$ possesses attribute m at some extent. For example, $\rho_{m9}(x_7) = 0$ implies that x_7 has not white hair. For each $m \in M$, let $\rho_{m'}(x) = \max_{x\in X}(\rho_m(x)) - \rho_m(x)$ for $x \in X$, where m' is the negation of simple concept m. By Definition 4.8, we can verify that each ρ_m is the weight function of concept m. Table 5.2 shows each weight function of the simple concept in M. Table 5.3 shows the membership functions obtained by the norm of the lattice

Table 5.1 Description of features

	appearance			wealth		gender		hair color		
	age	*height*	*weigh*	*salary*	*estate*	*male*	*female*	*black*	*white*	*yellow*
x_1	20	1.9	90	1	0	1	0	6	1	4
x_2	13	1.2	32	0	0	0	1	4	3	1
x_3	50	1.7	67	140	34	0	1	6	1	4
x_4	80	1.8	73	20	80	1	0	3	4	2
x_5	34	1.4	54	15	2	1	0	5	2	2
x_6	37	1.6	80	80	28	0	1	6	1	4
x_7	45	1.7	78	268	90	1	0	1	6	4
x_8	70	1.65	70	30	45	1	0	3	4	2
x_9	60	1.82	83	25	98	0	1	4	3	1
x_{10}	3	1.1	21	0	0	0	1	2	5	3

Table 5.2 Weight functions of simple concepts in M

ρ_{m_i}	ρ_{m_1}	ρ_{m_2}	ρ_{m_3}	ρ_{m_4}	ρ_{m_5}	ρ_{m_6}	ρ_{m_7}	ρ_{m_8}	ρ_{m_9}	$\rho_{m_{10}}$
x_1	20	1.9	90	1	0	1	0	1	1	1
x_2	13	1.2	32	0	0	0	1	1	1	1
x_3	50	1.7	67	140	34	0	1	1	1	1
x_4	80	1.8	73	20	80	1	0	1	1	1
x_5	34	1.4	54	15	2	1	0	1	1	1
x_6	37	1.6	80	80	28	0	1	1	1	1
x_7	45	1.7	78	268	90	1	0	1	0	1
x_8	70	1.65	70	30	45	1	0	1	1	1
x_9	60	1.82	83	25	98	0	1	1	1	1
x_{10}	3	1.1	21	0	0	0	1	1	1	1

Table 5.3 Membership functions obtained by the norm of the lattice EXM defined by (5.18)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
$\mu_{\alpha_1}()$	.0874	.0388	.4903	1.0	.1699	.2597	.3689	.8058	.6359	.0073
$\mu_{\alpha_2}()$	0	0	.1698	.5013	.0053	.0796	.7401	.2891	1.0	0
$\mu_{\alpha_3}()$	1	0	0	1	1	0	1	1	0	0
$\mu_{\alpha_4}()$	0	0	0	.5013	.0053	0	.3183	.2891	0	0
$\mu_{\alpha_5}()$	.0017	0	.5371	.5013	.0276	.2953	.3689	.2891	1	0
$\mu_{\alpha'_5}()$	.6750	1	.0609	.0431	.5124	.1504	.2448	.1012	0	1

Table 5.4 Membership functions obtained by the norm of the lattice $E^{\#}X$ defined by (5.24)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
$\mu_{\alpha_1}()$	.3000	.2000	.7000	1.0	.4000	.5000	.6000	.9000	.8000	.1000
$\mu_{\alpha_2}()$	0	0	.6000	.8000	.4000	.5000	.9000	.7000	1.0	0
$\mu_{\alpha_3}()$	1	0	0	1	1	0	1	1	0	0
$\mu_{\alpha_4}()$	0	0	0	.8000	.4000	0	.6000	.7000	0	0
$\mu_{\alpha_5}()$	.3000	0	.9000	.8000	.4000	.8000	.6000	.7000	1	0
$\mu_{\alpha'_5}()$	.8000	1	.2000	.3000	.7000	.3000	.5000	.2000	0	1

$(EXM,\vee,\wedge)$ defined by (5.18) for the weight functions in Table 5.2. Table 5.4 shows the membership functions obtained by the norm of the lattice $(E^{\#}X,\vee,\wedge)$ defined by (5.24) for $\forall a\subseteq X$, $\mathscr{M}(a)=|a|$. Where $\alpha_1=m_1$, $\alpha_2=m_5$, $\alpha_3=m_6$, $\alpha_4=m_1m_5m_6$, $\alpha_5=m_1m_4m_6+m_1m_5m_6+m_4m_7+m_5m_7$, $\alpha'_5\in EM$. By (4.19), we have $\alpha'_5=m'_6m'_7+m'_4m'_5+m'_1m'_7$.

Compared to t-norm and t-conorm, if $\forall\alpha,\beta\in EX_1...X_nM$, define $\mathscr{T}(\alpha,\beta)=||\alpha\wedge\beta||$, where $||.||$ is the norm of $EX_1...X_nM$ or $E^{\#}X$, then $\mathscr{T}$ has the similar property to the one of any t-norm and satisfies the following: $\alpha,\beta,\gamma,\delta\in EX_1...X_nM$,

$$||O\wedge O||=0,\ ||\alpha\wedge 1||=||1\wedge\alpha||=||\alpha||,\ ||\alpha\wedge\beta||=||\beta\wedge\alpha||,$$
$$||\alpha\wedge\beta||\le||\gamma\wedge\delta||,\ if\ \alpha\le\gamma,\beta\le\delta\ in\ lattice\ EX_1...X_nM\ or\ in\ lattice\ E^{\#}X,$$

where 1 is the maximum element $(X_1...X_n)\varnothing$ in $EX_1...X_nM$ or X in $E^{\#}X$ and O is the minimum element $(\varnothing...\varnothing)M$ in $EX_1...X_nM$ or $\varnothing$ in $E^{\#}X$. Similarly, if $\forall\alpha,\beta \in EX_1...X_nM$, define $\mathscr{S}(\alpha,\beta) = ||\alpha \vee \beta||$, then $\mathscr{S}$ has the similar properties to t-conorm.

For each fuzzy concept $\gamma \in EM$, the membership degree of each x belonging to γ can be represented by the lattice EXM, $EXMM$, $E^{\#}X$ or $E^{\#}(X \cup M)$, and by the norms of these lattice representations, four kinds of $[0,1]$ representations can be obtained. If the original information allows us to establish one AFS structure (M,τ,X) instead of two compatible AFS structures, then we have four kinds of representations of the membership degrees which are all in agreement and in the following order from left to right more and more finer (i.e. more and more elements in X are comparable) while more and more original information becomes lost.

1. EXM representations→ $E^{\#}X$ representations;
2. The norm of EXM representations→the norm of $E^{\#}X$ representations;

If the original information allows us to establish two compatible AFS structures (M,τ_1,X), (X,τ_2,M), then we have eight kinds of representations of the membership degrees which are all consistent and in the following order from left to right more and more finer (i.e. more and more elements in X are comparable) while more and more original information are lost.

1. $EXMM$ representations→ $E^{\#}(X \cup M)$ representations
2. $EXMM$ representations→ EXM representations→ $E^{\#}X$ representations.
3. The norm of $EXMM$ representations→The norm of $E^{\#}(X \cup M)$ representations
4. The norm of $EXMM$ representations→The norm of EXM representations→The norm of $E^{\#}X$ representations.

For a real world problem, the type of representation to be employed depends on how much original information is provided, how much the original information we have to preserve and how many elements in the universe of discourse we want to compare their membership degrees by including in decision making. Thus for the fuzzy concepts in EM, this approach provides 8 kinds of representations of the membership degrees, i.e., EXM, $EXMM$, $E^{\#}X$, $E^{\#}(X \cup M)$ algebra and four kinds of [0,1] interval representations and their fuzzy logic operations, which are automatically determined by the distribution of original data and the semantic interpretations of the simple concepts in M. All these representation are harmonic and consistent. That is, for $x,y \in X$ and any fuzzy concept $\zeta \in EM$, if the membership degrees of x and y belonging to ζ are comparable in two different AFS representing forms, then the orders of x and y belonging to ζ in the two different AFS representing forms must be the same.

5.4 Further Algebraic Properties of AFS Algebra

In this section, we first elaborate on the order of the EI algebra. Second, the mathematical properties and algebraic structures of EI^n algebra have been exhaustively explored and the expressions of both $\wedge$-irreducible elements in the molecular lattice

$(EM, \vee, \wedge)$ and $(EX_1...X_nM, \vee, \wedge)$ are developed. Next the standard minimal family of elements in $(EM, \vee, \wedge)$ and $(EX_1...X_nM, \vee, \wedge)$ is formulated. It is proved that neither $(EM, \vee, \wedge)$ nor $(EX_1...X_nM, \vee, \wedge)$ is a fuzzy lattice.

5.4.1 Order of *EI* Algebra

In this section, we study the order of *EI* algebra generated by some finite elements [21]. The results shown that AFS algebra leads to structures. It involves difficult combinatoric problems such as the Sperner class [6].

Definition 5.7. Let M be a finite set. The number of the elements in EM is called the *order of the EI algebra* of EM, denoted as $O(EM)$.

Let M be the set of simple concepts on a set X. $O(EM)$ is the number of complex concepts generated by the simple concepts in M, which are not equivalent in *EI* algebra EM,, i.e., they have different semantic interpretations.

Definition 5.8. Let M be any set, $\sum_{i\in I} A_i \in EM$. $\sum_{i\in I} A_i$ is called *irreducible*, if for any $u \in I$, $\sum_{i\in I} A_i \neq \sum_{i\in I, i\neq u} A_i$, otherwise $\sum_{i\in I} A_i$ is called *reducible*.

For any irreducible fuzzy concept $\alpha = \sum_{i\in I} A_i \in EM$, if β is a fuzzy concept obtained by omitting an item A_i from α, then α, β are different fuzzy concepts.

Lemma 5.1. *([1, 3, 7])Let M be a set. For $\sum_{i\in I} A_i, \sum_{j\in J} B_j \in EM$, if $\sum_{i\in I} A_i = \sum_{j\in J} B_j$, $\sum_{i\in I} A_i$ and $\sum_{j\in J} B_j$ are both irreducible, then $\{B_j | j \in J\} = \{A_i | i \in I\}$.*

Its proof is left as an exercise. According to Lemma 5.1, we arrive at the following definition.

Definition 5.9. Let M be a set and $\sum_{i\in I} A_i \in EM$. $|\sum_{i\in I} A_i|$, the *set of irreducible items* of $\sum_{i\in I} A_i$ is defined as follows.

$$|\sum_{i\in I} A_i| \triangleq \{A_i \mid i \in I,\ for\ any\ j \in I, i \neq j,\ A_i \not\supseteq A_j\}.$$

$||\sum_{i\in I} A_i|| = |\{A_i \mid i \in I,\ for\ any\ j \in I, i \neq j,\ A_i \not\supseteq A_j\}|$ is called the *length* of $\sum_{i\in I} A_i$. Thus $||.||$ is a map from EX to the set of natural numbers.

For the fuzzy concept $\gamma = \sum_{j\in J} B_j \in EM$, $||\gamma||$ is the number of items in γ which cannot be omitted. Let M be a finite set and $|M| = n$. The set $C_k(EM)$ is defined as follows: for $k = 1, 2, ..., n$.

$$C_k(EM) = \{\alpha \mid \alpha \in EM^+, \alpha\ is\ irreducible,\ ||\alpha|| = k\}, \tag{5.26}$$

i.e., $C_k(EM)$ is the set of k length irreducible elements in EM. The number of elements in $C_k(EM)$, $|C_k(EM)|$ is denoted by $l_k(n)$. It is straightforward to notice

$$O(EM) = \sum_{1\leq i\leq\infty} |C_i(EM)|. \tag{5.27}$$

Theorem 5.7. *Let* $M = \{m_1, m_2, \cdots, m_n\}$*, then* $|C_1(M)| = l_1(n) = 2^n$.

Theorem 5.8. *Let* $M = \{m_1, m_2, \cdots, m_n\}$*, then*

$$|C_2(EM)| = C^2_{2^n-1} + 2^{n+1} - 3^n - 1,$$

where $C^k_m = \frac{m!}{k!(m-k)!}$.

Proof. Let $B = \{\alpha | \alpha = A_1 + A_2, A_1 \neq A_2, A_1, A_2 \in 2^M - \{\varnothing\}, \alpha \text{ is reducible}\}$. Let

$$B_i = \{\alpha | \alpha = A_1 + A_2, A_1 \subset A_2, |A_1| = i, A_1, A_2 \in 2^M - \{\varnothing\}\},$$

for $1 \leq i \leq n-1$. For $|A_1| = 1$, we have

$$|B_1| = C^1_n C^1_{n-1} + C^1_n C^2_{n-1} + \cdots + C^1_n C^{n-1}_{n-1} = C^1_n (2^{n-1} - 1).$$

For $|A_1| = 2$, we have

$$|B_2| = C^2_n C^1_{n-2} + C^2_n C^2_{n-2} + \cdots + C^2_n C^{n-2}_{n-2} = C^2_n (2^{n-2} - 1)$$

Similarly, for $3 \leq |A_1| \leq n-2$, it is easy to verify that

$$|B_k| = C^k_n C^1_{n-k} + C^k_n C^2_{n-k} + \cdots + C^k_n C^{n-k}_{n-k} = C^k_n (2^{n-k} - 1).$$

If $|A_1| = n-1$, then $|A_2| = n$ and $|B_{n-1}| = C^{n-1}_n C^1_1$. It is obvious that $B = \bigcup_{1 \leq i \leq n-1} B_i$, $B_i \cap B_j = \varnothing$, $i \neq j$, $i, j = 1, 2, \cdots, \quad n-1$. Thus

$$\begin{aligned} |B| &= |\bigcup_{1 \leq i \leq n-1} B_i| = \sum_{i=1}^{n-1} |B_i| = \sum_{i=1}^{n-1} (C^i_n 2^{n-i} - C^i_n) \\ &= \sum_{i=1}^{n-1} C^i_n 2^{n-i} - \sum_{i=1}^{n-1} C^i_n = (2+1)^n - 2^n - 1 - (2^n - 2) \\ &= -2^{n+1} + 3^n + 1. \end{aligned}$$

Let $L = \{A_1 + A_2 | A_1, A_2 \in 2^M\}$. For any $\alpha = A_1 + A_2 \in L$, we consider the following cases

case 1: α is reducible, i.e., one of A_1 and A_2 contains another, this means that $\alpha \in B$.
case 2: α is irreducible, i.e., each one of A_1 and A_2 cannot contain another.

This implies that $\alpha \in C_2(EM)$. Therefore L can be expressed as $C_2(EM) \cup B$, $C_2(EM) \cap B = \varnothing$. Now, we have $|C_2(EM)| = |L| - |B|$. As $|L| = C^2_{2^n-1}$ therefore, we get that

$$|C_2(EM)| = C^2_{2^n-1} + 2^{n+1} - 3^n - 1. \qquad \square$$

For $k \geq 3$, we do not know how many elements in $C_k(EM)$ we have so far and this remains as an open problem. However we have the following theorem to estimate $|C_k(EM)|$.

Theorem 5.9. *Let M be set and $|M| = n$. If $C_k(EM)$ is denoted as $l_k(n)$, then*

$$l_k(n+1) \geq \sum_{m=0}^{n-2} (2^k-1)^m \Bigg(C^k_{2^{n-m}-1} - \sum_{1\leq i,j\leq n-m-2}^{i+j\leq n-m} C^{k-3}_{2^{n-m}-4} C^i_{n-m-j} C^j_{n-m} (2^{n-m-(i+1)}-1) + l_{k-1}(n-m) \Bigg),$$

where $n \geq 2, k \geq 3$.

Proof. Let $M = \{m_1, m_2, ..., m_n\}$ and $M_1 = \{m_1, m_2, ..., m_{n+1}\}$. For any k, the set J is defined as follows:

$$J = \left\{\alpha \mid \alpha = A_1 + A_2 + \cdots + A_k,\ \ \alpha \in EM^+, A_i \neq A_j,\ i \neq j, i, j = 1, 2, \cdots, k\right\}.$$

It is clear that $C_k(M) \subseteq J$. In what follows, J is divided into three disjoint parts $C_k(EM), C$ and D. Put

$$C = \left\{ \alpha = \sum_{i=1}^{k} A_i \in J \ \mid\ A_i \neq \varnothing, \exists i_1, i_2, i_3 \in \{1, 2, \cdots, k\},\ \ A_{i_1} \subset A_{i_2} \subset A_{i_3} \right\},$$

We now pick $\alpha \in J$ such that α is reducible and $\alpha \notin C$, such elements form a set D. It is easy to show that

$$J = C_k(EM) \cup C \cup D, \tag{5.28}$$

and $C_k(EM),\ C,\ D$ are pairwise disjointed.

Next we give a partition of the set $C_k(EM_1)$. It is obvious that $C_k(EM) \subseteq C_k(EM_1)$. For all $\alpha \in C_k(EM_1) - C_k(EM), \alpha = A_1 + A_2 + \cdots + A_k$,so there exists $A_i (1 \leq i \leq k)$ such that $m_{n+1} \in A_i$. If there is exactly one $A_i (1 \leq i \leq k)$ such that $\{m_{n+1}\} = A_i$, then by removing $\{m_{n+1}\}$ from $\alpha = A_1 + A_2 + \cdots + A_k$, we get

$$\alpha_1 = A_1 + A_2 + \cdots + A_{i-1} + \cdots A_{i+1} + \cdots + A_k$$

It is evident that α_1 is irreducible and $\alpha_1 \in C_{k-1}(EM)$. Let

$$J_1 = \{\alpha \mid \alpha = \beta + \{m_{n+1}\},\ \ \beta \in C_{k-1}(EM)\}.$$

Then we have that $|J_1| = |C_{k-1}(EM)|$. If there are more than one $A_i (1 \leq i \leq k)$ such that $m_{n+1} \in A_i$,by removing m_{n+1} from $A_i (1 \leq i \leq k)$,we obtain an element (denoted by $\alpha - \{m_{n+1}\}$) from EM. Let

$$J_2 = \{\alpha \mid \alpha = A_1 + A_2 + \cdots + A_k,\ \alpha \in C_k(EM_1),\ \alpha - \{m_{n+1}\} \in C_k(EM)\}.$$

This implies that for any α from J_2, $\alpha - \{m_{n+1}\}$ is irreducible, and

$$|J_2| = (C^1_k + C^2_k + \cdots + C^k_k)\,|C_k(EM)| = (2^k - 1)\,|C_k(EM)|.$$

If $\alpha - \{m_{n+1}\}$ is reducible, we put

$$J_3 = \{\alpha \mid \alpha = A_1 + A_2 + \cdots + A_k, \alpha \in C_k(M_1), \alpha - \{m_{n+1}\} \text{ is reducible}\}.$$

Thus, we get a disjoint union

$$C_k(EM_1) = C_k(EM) \cup J_1 \cup J_2 \cup J_3. \tag{5.29}$$

In the sequel, we need to show that $|J_3| \geq |D|$.

For all $\alpha \in D$, as α is reducible, therefore, there exist $i, j \in \{1, 2, \cdots, k\}$ such that $i \neq j$ and $A_i \subset A_j$. Adding m_{n+1} to all such A_i (it also denoted by A_i), thus, $A_i \subset A_j$ cannot stand in EM_1^+, then we get an element α_1 from EM_1^+. We claim that $\alpha_1 \in J_3$. It suffices to show that $\alpha_1 \in C_k(EM_1)$ and $\alpha_1 - \{m_{n+1}\}$ is reducible.

It is clear that $\alpha_1 - \{m_{n+1}\}$ is reducible, if α_1 is reducible, then there exist $i, j \in \{1, 2, \cdots, k\}$ such that $i \neq j$ and $A_i \subset A_j$. If $m_{n+1} \in A_i$, but $m_{n+1} \notin A_j$ then from the formation of α_1, we know that this is impossible, so we have that $m_{n+1} \in A_i$, and $m_{n+1} \in A_j$, thus there exists A_l such that $A_j \subset A_l$. Omitting m_{n+1} from A_i, A_j, A_l, we get that $A_i \subset A_j \subset A_l$, this is in conflict with $\alpha \notin C$, thus $\alpha_1 \in C_k(EM_1)$. Therefore, a monomorphic mapping $\alpha \rightarrow \alpha_1$ is established. This implies that $|J_3| \geq |D|$. Considering (5.28) and (5.29), we have

$$|J_3| \geq |D| = |J| - |C_k(M)| - |C|,$$
$$|C_k(M_1)| = |C_k(M)| + |J_1| + |J_2| + |J_3|.$$

Also

$$|C_k(M_1)| \geq |J| + |J_1| + |J_2| - |C|. \tag{5.30}$$

For the convenience,we denote $|C|$ by $N_3(n)$. Then it follows that

$$l_k(n+1) - (2^k - 1)l_k(n) \geq C^k_{2^n - 1} + l_{k-1}(n) - N_3(n). \tag{5.31}$$

Furthermore

$$l_k(n) - (2^k - 1)l_k(n-1) \geq C^k_{2^{n-1} - 1} + l_{k-1}(n-1) - N_3(n-1) \tag{5.32}$$
$$\cdots\cdots\cdots$$
$$l_k(3) - (2^k - 1)l_k(2) \geq C^k_{2^2 - 1} + l_{k-1}(2) \tag{5.33}$$

By multiplying both sides of the $(i+1)$ th inequality by $(2^k - 1)^i$, $i = 1, 2, \cdots, n-2$ (from (5.31) to (5.33)), we obtain $n-2$ inequalities, After summing them up, we get

$$l_k(n+1) \geq \sum_{i=1}^{n-2} (2^k - 1)^i (C^k_{2^{n-i} - 1} - N_3(n) + l_{k-1}(n-i)) \tag{5.34}$$

Now we formulate a upper boundary of $N_3(n)$. For all $\alpha \in C, \alpha$ can be presented as $\alpha = A_1 + A_2 + A_3 + \cdots + A_k$ where $A_1 \subset A_2 \subset A_3$. From this, we construct a set B as follows

$$B = \left\{ \alpha = A_1 + A_2 + A_3 \in EM^+ \mid A_i \neq \varnothing,\ A_1 \subset A_2 \subset A_3 \right\}.$$

If $|A_1| = 1$, $|A_2| = 2$, $|A_3| = 3, 4, \cdots, n$, then

$$\begin{aligned} |B| &= C_n^1 C_{n-1}^1 (C_{n-2}^1 + C_{n-2}^2 + C_{n-2}^3 + \cdots + C_{n-2}^{n-2}) \\ &= C_n^1 C_{n-1}^1 (2^{n-2} - 1). \end{aligned}$$

If $|A_1| = 1$, $|A_2| = 3$, $|A_3| = 4, 5, \cdots, n$, then

$$\begin{aligned} |B| &= C_n^1 C_{n-1}^2 (C_{n-3}^1 + C_{n-3}^2 + C_{n-3}^3 + \cdots + C_{n-3}^{n-3}) \\ &= C_n^1 C_{n-1}^2 (2^{n-3} - 1) \end{aligned}$$

$\cdots\cdots$

If $|A_1| = 1$, $|A_2| = n-1$, $|A_3| = n$, then $|B| = C_n^1 C_{n-1}^{n-2} C_1^1$.

$\cdots\cdots$

If $|A_1| = 2$, $|A_2| = 3$, $|A_3| = 4, 5, \cdots, n$, then

$$\begin{aligned} |B| &= C_n^2 C_{n-2}^1 (C_{n-3}^1 + C_{n-3}^2 + C_{n-3}^3 + \cdots + C_{n-3}^{n-3}) \\ &= C_n^2 C_{n-2}^1 (2^{n-3} - 1). \end{aligned}$$

If $|A_1| = 2$, $|A_2| = 4$, $|A_3| = 5, 6, \cdots, n$, then

$$\begin{aligned} |B| &= C_n^2 C_{n-2}^2 (C_{n-4}^1 + C_{n-4}^2 + C_{n-4}^3 + \cdots + C_{n-4}^{n-4}) \\ &= C_n^2 C_{n-2}^2 (2^{n-4} - 1). \end{aligned}$$

$\cdots\cdots$

If $|A_1| = 2$, $|A_2| = n-1$, $|A_3| = n$, then $|B| = C_n^2 C_{n-2}^{n-3} C_1^1$.

$\cdots\cdots$

Thus A_1, A_2, A_3 are completely defined in this way. Now $\alpha \in C$, we select the other $k-3$ items from the set $2^M - \{\varnothing, A_1, A_2, A_3\}$. As $\left|2^M - \{\varnothing, A_1, A_2, A_3\}\right| = 2^n - 4$ and $|A_1| \le n-2$, then we get

$$N_3(n) \le \sum_{1 \le i,j \le n-2}^{i+j \le n} C_{2^n-4}^{k-3} C_n^j C_{n-j}^i (2^{n-(i+j)} - 1),$$

where $n \ge 2, k \ge 3$. As $|J| = C_{2^n-1}^k$, We rewrite (5.34) as follows

$$\begin{aligned} l_k(n+1) \ge \sum_{m=0}^{n-2} (2^k - 1)^m \Bigg(& C_{2^{n-m}-1}^k \\ & - \sum_{1 \le i,j \le n-m-2}^{i+j \le n-m} C_{2^{n-m}-4}^{k-3} C_{n-m-j}^i C_{n-m}^j (2^{n-m-(i+j)} - 1) + l_{k-1}(n-m) \Bigg), \end{aligned}$$

where $n \ge 2$, $k \ge 3$. $\square$

Thus by (5.27) and (5.9), we can obtain the estimation of $O(EM)$.

Definition 5.10. ([6]) Let Γ be the set of some subsets of the finite set $\{1,2,\cdots,n\}$ (called n-set for short, denoted by $[1,n]$). If it satisfied $\forall A,B \in \Gamma, A \nsubseteq B$ and $B \nsubseteq A$, then Γ is called a *Sperner class* on n-set $[1,n]$.

Theorem 5.10. *If Γ is a Sperner class of the n-set $[1,n]$, then $|\Gamma| \leq C_n^{[n/2]}$, where $[n/2]$ is the least integer larger than $n/2$.*

Proof. Let Γ be a Sperner class of the n-set $[1,n]$. First, we prove that

$$\sum_{F\in\Gamma} \frac{1}{C_n^{|F|}} \leq 1$$

by the induction with respect to n. Assume that $[1,n] \notin \Gamma$. For $i \in [1,n]$, let $\Gamma(i) = \{F \in \Gamma \mid i \notin F\}$. It is obvious that $\Gamma(i)$ is a Sperner class of the $(n-1)$-set $[1,n]-\{i\}$. By the assumption of the induction, we have

$$\sum_{F\in\Gamma(i)} \frac{1}{C_{n-1}^{|F|}} \leq 1. \tag{5.35}$$

Because (5.35) holds for every $i \in [1,n]$, we arrive at the the sum of (5.35) for all $i \in [1,n]$ shown as follows.

$$\sum_{i=1}^{n} \sum_{F\in\Gamma(i)} \frac{1}{C_{n-1}^{|F|}} \leq n. \tag{5.36}$$

Checking the number of times $1/C_{n-1}^{|\gamma|}$ occur at the left side of (5.36), we know that $1/C_{n-1}^{|\gamma|}$ appears one time as long as $i \notin F$, thus in total it appears $(n-|F|)$ times. Therefore we have

$$\sum_{F\in\Gamma} \frac{n-|F|}{C_{n-1}^{|F|}} \leq n. \tag{5.37}$$

Each side of (5.37)is now divided by n. By noting that $(n-|F|)/(nC_{n-1}^{|F|}) = 1/C_{n-1}^{|F|}$, we have

$$\sum_{F\in\Gamma} 1/C_n^{|F|} \leq 1. \tag{5.38}$$

It is clear that $C_n^{[n/2]}$ is the biggest one of all combinatorial numbers C_n^r. From (5.38), it follows that

$$|\Gamma|\frac{1}{C_n^{[n/2]}} = \sum_{F\in\Gamma} \frac{1}{C_n^{[n/2]}} \leq \sum_{F\in\Gamma} \frac{1}{C_{n-1}^{|F|}} \leq 1. \tag{5.39}$$

Therefore $|\Gamma| \leq C_n^{[n/2]}$. □

Theorem 5.11. *Let $M = \{m_1, m_2, \cdots, m_n\}$. Then the longest fuzzy concept in EM is $C_n^{[n/2]}$, where $[n/2]$ is the least number which is larger than $n/2$.*

Proof. It follows straightforward by Definition 5.9 and Theorem 5.10. □

For M a set of simple concepts, the fuzzy concept $\gamma = \sum_{i=1}^{C_n^{[n/2]}} A_i \in EM$ is the longest of all fuzzy concepts in EM, where $|A_i| = [n/2]$ for all $1 \le i \le C_n^{[n/2]}$ and $||\gamma|| = C_n^{[n/2]}$. From the above discussions, we know that an enormously great number of fuzzy sets can be represented by a few simple concepts using the EI algebra. Interestingly, EI algebra has a very rich structure and AFS fuzzy logic is an important mathematical tool to study human-centric facets of intelligent systems.

5.4.2 Algebraic Structures of EI Algebra

In this section, we further discuss the properties of lattice $(EM, \vee, \wedge)$ and mainly focus on the following issues:

1. The structure of the set of irreducible elements in EM;
2. The standard minimal family of an element in EM;
3. We will prove that $(EM, \vee, \wedge)$ is a new type of molecular lattice, which is neither a Boolean algebra nor a fuzzy lattice. However, the sublattice $(SEM, \vee, \wedge)$ is a fuzzy lattice, where

$$SEM = \left\{ \sum_{i \in I} A_i \mid A_i \in 2^M - \{\varnothing\}, i \in I, I \text{ is any indexing set} \right\} \subseteq EM.$$

Definition 5.11. Let L be a molecular lattice. L is called a *fuzzy lattice* if there exists a map $\sigma : L \to L$ such that for any $x, y \in L$ the following conditions are satisfied:

1. $x \le y \Rightarrow y^\sigma \le x^\sigma$;
2. $x = (x^\sigma)^\sigma$.

Theorem 5.12. *Let $(L, \vee, \wedge, \sigma)$ be a fuzzy lattice, then the strong De Morgan Law holds, that is, for any $a_t \in L, t \in T$,*

$$\left(\bigvee_{t \in T} a_t \right)^\sigma = \bigwedge_{t \in T} a_t^\sigma, \quad \left(\bigwedge_{t \in T} a_t \right)^\sigma = \bigvee_{t \in T} a_t^\sigma.$$

Proposition 5.8. *Let M be a non-empty set and $\sum_{i \in I} A_i, \sum_{j \in J} B_j \in EM$. Then the following assertions hold:*

1. $\sum_{i \in I} A_i \wedge \sum_{j \in J} B_j = M \Rightarrow A_i \supseteq \bigcup_{j \in J} B'_j$, $\forall i \in I$ *and* $B_j \supseteq \bigcup_{I \in I} A'_i$, $\forall j \in J$;
2. $\sum_{i \in I} A_i \vee \sum_{j \in J} B_j = \varnothing \Rightarrow \sum_{i \in I} A_i = \varnothing$ *or* $\sum_{j \in J} B_j = \varnothing$.

Proof. Let $\sum_{i \in I} A_i \wedge \sum_{j \in J} B_j = M$, we have $\sum_{i \in I, j \in J} A_i \cup B_j = M$, consequently, for any $i \in I, j \in J$, $A_i \cup B_j \supseteq M$, i.e., $A_i \cup B_j = M$. It follows that for any $i \in I$, $B_j \supseteq A'_i$,

$\forall j \in J$, that is $B_j \supseteq \bigcup_{i\in I} A_i' = (\bigcap_{i\in I} A_i)'$, $\forall j \in J$; and for any $j \in J$, $A_i \supseteq B_j'$, $\forall i \in I$, that is $A_i \supseteq \bigcup_{j\in J} B_j' = (\bigcap_{j\in J} B_j)'$, $\forall i \in I$.

Let $\sum_{i\in I} A_i \vee \sum_{j\in J} B_j = \varnothing$, consequently, we have either $\exists i_0 \in I$ such that $\varnothing \supseteq A_{i_0}$, i.e., $A_{i_0} = \varnothing$ or $\exists j_0 \in J$ such that $\varnothing \supseteq B_{j_0}$, i.e., $B_{j_0} = \varnothing$. Thus $\sum_{i\in I} A_i = \varnothing$ or $\sum_{j\in J} B_j = \varnothing$. □

In what follows, we will show that the lattice $(EM, \vee, \wedge)$ is not a Boolean algebra if $|M| > 1$. Assume that it is a Boolean algebra. By Definition 2.13, for any $\sum_{i\in I} A_i \in EM$, there exists $\sum_{j\in J} B_j \in EM$ such that $\sum_{i\in I} A_i \vee \sum_{j\in J} B_j = \varnothing$ and $\sum_{i\in I} A_i \wedge \sum_{j\in J} B_j = M$. We have either $\sum_{i\in I} A_i = \varnothing$ or $\sum_{j\in J} B_j = \varnothing$ by Proposition 5.8. We can assume without loss of generality, that $\sum_{i\in I} A_i = \varnothing$. Then $\exists u \in I$ such that $A_u = \varnothing$. From Proposition 5.8 again, we have $\varnothing = A_u \supseteq \bigcup_{j\in J} B_j'$. It follows that $\forall j \in J$, $B_j = M$, i.e., $\sum_{j\in J} B_j = M$. This implies that $(EM, \vee, \wedge)$ is not a Boolean algebra if $|M| > 1$. However, it is easy to prove that $(EM, \vee, \wedge)$ is a Boolean algebra if $|M| = 1$.

Theorem 5.13. *Let M be a non-empty set and $(EM, \vee, \wedge)$ be the EI algebra over M. Then $\{A \mid A \in 2^M\}$ is the set of all strong $\vee$—irreducible elements in EM (refer to Definition 2.10)*

Theorem 5.14. *Let M be a non-empty set and $(EM, \vee, \wedge)$ be the EI algebra over M. Then $\{\sum_{m\in A}\{m\} \mid A \subseteq M\}$ is the set of all $\wedge$—irreducible elements in EM.*

Proof. First, for any $\forall A \in 2^M$, we prove that $\sum_{m\in A}\{m\}$ is a $\wedge$—irreducible element in *EM*. For $\sum_{i\in I} A_i, \sum_{j\in J} B_j \in EM$, if $\sum_{m\in A}\{m\} = \sum_{i\in I} A_i \wedge \sum_{j\in J} B_j$, we have

$$\sum_{m\in A}\{m\} \le \sum_{i\in I} A_i \ \ and \ \ \sum_{m\in A}\{m\} \le \sum_{j\in J} B_j.$$

Furthermore, we have

$$\sum_{m\in A}\{m\} \vee \sum_{i\in I} A_i = \sum_{i\in I} A_i \ \ and \ \ \sum_{m\in A}\{m\} \vee \sum_{j\in J} B_j = \sum_{j\in J} B_j.$$

Assume that $\sum_{m\in A}\{m\} \ne \sum_{i\in I} A_i$ *and* $\sum_{m\in A}\{m\} \ne \sum_{j\in J} B_j$. This implies that there exist $I_0 \subseteq I$ and $J_0 \subseteq J$ such that $A_i \cap A = B_i \cap A = \varnothing, \forall i \in I_0, j \in J_0$ and

$$\sum_{m\in A}\{m\} \vee \sum_{i\in I_0} A_i = \sum_{i\in I} A_i \ \ and \ \ \sum_{m\in A}\{m\} \vee \sum_{j\in J_0} B_j = \sum_{j\in J} B_j.$$

Consequently, we have

$$\begin{aligned}
\sum_{m\in A}\{m\} &= \left(\sum_{i\in I} A_i\right) \wedge \left(\sum_{j\in J} B_j\right) \\
&= \left(\sum_{m\in A}\{m\} + \sum_{i\in I_0} A_i\right) \wedge \left(\sum_{m\in A}\{m\} + \sum_{j\in J_0} B_j\right) \\
&= \left[\left(\sum_{m\in A}\{m\}\right) \wedge \left(\sum_{m\in A}\{m\}\right)\right] + \left[\left(\sum_{i\in I_0} A_i\right) \wedge \left(\sum_{m\in A}\{m\}\right)\right]
\end{aligned}$$

$$+\left[\left(\sum_{m\in A}\{m\}\right)\wedge\left(\sum_{j\in J_0}B_j\right)\right]+\left[\left(\sum_{i\in I_0}A_i\right)\wedge\left(\sum_{j\in J_0}B_j\right)\right]$$
$$=\sum_{m\in A}\{m\}+\sum_{m\in A,i\in I_0}(\{m\}\cup A_i)+\sum_{m\in A,j\in J_0}(\{m\}\cup B_j)+\sum_{i\in I_0,j\in J_0}(A_i\cup B_j)$$
$$=\sum_{m\in A}\{m\}+\sum_{i\in I_0,j\in J_0}(A_i\cup B_j).$$

Therefore, it follows that $\forall i\in I_0,\ j\in J_0,\ \exists m\in A$ such that $A_i\cup B_j\supseteq\{m\}$, that is, $\exists m\in A$ such that $m\in A_i\cup B_j$, this contradicts that $A_i\cap A=B_j\cap B=\varnothing$. Thus,

$$\sum_{m\in A}\{m\}=\sum_{i\in I}A_i \ or\ \sum_{m\in A}\{m\}=\sum_{j\in J}B_j.$$

Secondly, we prove that each $\wedge$—irreducible element must be in

$$\{\sum_{m\in A}\{m\}\mid A\in 2^M\}.$$

Suppose that $\sum_{i\in I}A_i$ is a $\wedge$—irreducible element in EM, and $\exists u\in I,|A_u|>1$. Let $B=A_u-\{m\}$ and $m\in A_u$. Since

$$\left[\sum_{i\in I-\{u\}}A_i+B\right]\wedge\left[\sum_{i\in I-\{u\}}A_i+\{m\}\right]$$
$$=\left(\sum_{i\in I-\{u\}}A_i\right)\wedge\left(\sum_{i\in I-\{u\}}A_i\right)+B\wedge\left(\sum_{i\in I-\{u\}}A_i\right)+\left(\sum_{i\in I-\{u\}}A_i\right)\wedge\{m\}+B\wedge\{m\}$$
$$=\sum_{i\in I-\{u\}}A_i+\sum_{i\in I-\{u\}}(A_i\cup B)+\sum_{i\in I-\{u\}}(A_i\cup\{x\})+A_u=\sum_{i\in I}A_i.$$

By the assumption, we have either

$$\textstyle\sum_{i\in I}A_i=\sum_{i\in I-\{u\}}A_i+B \text{ or } \sum_{i\in I}A_i=\sum_{i\in I-\{u\}}A_i+\{m\}.$$

However, it contradicts the facts that

$$\sum_{i\in I}A_i\neq\sum_{i\in I-\{u\}}A_i+B\ \ and\ \ \sum_{i\in I}A_i\neq\sum_{i\in I-\{u\}}A_i+\{m\}.$$

Thus the set of $\wedge$—irreducible elements in EM is $\{\sum_{m\in A}\{m\}\mid A\subseteq M\}$ and this completes the proof. □

By Theorem 5.13 and Theorem 5.14, it is straightforward to note that $\varnothing$ is both $\vee$—irreducible element and $\wedge$—irreducible element in EM. However, M is a $\vee$—irreducible element but not a $\wedge$—irreducible element in EM. For example, let $A,A'\in EM$ and $A\subset M$, we have that $M=A\cup A'=A\wedge A'$ but $A\neq M$ and $A'\neq M$, i.e., M is not a $\wedge$—irreducible element in EM.

Theorem 5.15. *Let M be a non-empty set, $\sum_{i\in I}A_i\in EM$, $\mathscr{P}_i=\{B\in 2^M|B\supseteq A_i\}$. Then $\mathscr{B}=\bigcup_{i\in I}\mathscr{P}_i$ is a standard minimal family of $\sum_{i\in I}A_i$ in molecular lattice $(EM,\vee,\wedge)$.*

Proof. Clearly, each element in $\mathscr{B}$ is a $\vee$—irreducible element in EM. By Theorem 5.1 and Proposition 5.1, we have $\sup \mathscr{B} = \sum_{i \in I} A_i$. Also, let

$$\mathscr{C} = \{\sum_{j \in J_k} A_{kj} \mid k \in K,\ K \text{ and } J_k \text{ are indexing sets}\} \subseteq EM,$$

then $\sum_{k \in K} \sum_{j \in J_k} A_{kj}$ is supremum of $\mathscr{C}$. Assume that $\sum_{i \in I} A_i \leq \sum_{k \in K} \sum_{j \in J_k} A_{kj}$. For any $B \in \mathscr{B}$ there exists $i \in I$ such that $B \supseteq A_i$. For this A_i there exist $k_0 \in K$, $j_0 \in J_{k_0}$ such that $A_i \supseteq A_{j_0 k_0}$, this leads to the fact that $B \supseteq A_i \supseteq A_{j_0 k_0}$. Thus $B \leq \sum_{j \in j_{k_0}} A_{k_0 j}$.

To sum up, $\mathscr{B}$ is a standard minimal family of $\sum_{i \in I} A_i$ by Definition 2.21. With this observation, the proof is completed. □

Let $\sum_{i \in I} A_i \in EM$, the set $\mathscr{M}(\sum_{i \in I} A_i)$ is defined as follows:

$$\mathscr{M}(\sum_{i \in I} A_i) = \{A \in 2^M, A \leq \sum_{i \in I} A_i\}. \tag{5.40}$$

Proposition 5.9. *Let M be a non-empty set and $\sum_{i \in I} A_i, \sum_{j \in J} B_j \in EM$. Then the following assertions hold:*

1. $\sum_{i \in I} A_i = \sup \mathscr{M}(\sum_{i \in I} A_i)$;
2. $\sum_{i \in I} A_i \leq \sum_{J \in J} B_j \Leftrightarrow \mathscr{M}(\sum_{i \in I} A_i) \subseteq \mathscr{M}(\sum_{j \in J} B_j)$;
3. $\sum_{i \in I} A_i \wedge \sum_{j \in J} B_j = M \Leftrightarrow \mathscr{M}(\sum_{i \in I} A_i) \cap \mathscr{M}(\sum_{j \in J} B_j) = \{M\}$;
4. $\mathscr{M}(\sum_{i \in I} A_i)$ *is the greatest of those standard minimal families of* $\sum_{i \in I} A_i$.

Proof.
1. Clearly, $\sum_{i \in I} A_i$ is a upper bound of $\mathscr{M}(\sum_{i \in I} A_i)$. Assume that $\sum_{k \in K} C_k$ is another upper bound of $\mathscr{M}(\sum_{i \in I} A_i)$. We have $A \leq \sum_{k \in K} C_k$ for $\forall A \in \mathscr{M}(\sum_{i \in I} A_i)$. Since $A_i \in \mathscr{M}(\sum_{i \in I} A_i)$ for $\forall i \in I$, $A_i \leq \sum_{k \in K} C_k$. It follows that, $\sum_{i \in I} A_i \leq \sum_{k \in K} C_k$. This implies that $\sum_{i \in I} A_i = \sup \mathscr{M}(\sum_{i \in I} A_i)$.

2. Let $\sum_{i \in I} A_i \leq \sum_{J \in J} B_j$. Since for $\forall A \in \mathscr{M}(\sum_{i \in I} A_i)$, we have that $A \leq \sum_{i \in I} A_i$. Thus $\exists i_0 \in I$ such that $A \supseteq A_{i_0}$. $\sum_{i \in I} A_i \leq \sum_{J \in J} B_j$ implies that for $\forall i \in I$, $\exists j \in J$ such that $A_i \supseteq B_j$, it follows that $\exists j_0$ such that $A \supseteq A_{i_0}$, that is, $A \leq \sum_{J \in J} B_j$. So we have that $A \in \mathscr{M}(\sum_{j \in J} B_j)$ and $\mathscr{M}(\sum_{i \in I} A_i) \subseteq \mathscr{M}(\sum_{j \in J} B_j)$.

Conversely, if $\mathscr{M}(\sum_{i \in I} A_i) \subseteq \mathscr{M}(\sum_{j \in J} B_j)$, then for $\forall i \in I$, we have that $A_i \in \mathscr{M}(\sum_{j \in J} B_j)$. Consequently, $\exists j_0 \in J$ such that $A_i \supseteq B_{j_0}$, that is, $\sum_{i \in I} A_i \leq \sum_{J \in J} B_j$.

3. Suppose that $\sum_{i \in I} A_i \wedge \sum_{j \in J} B_j = M$. By Proposition 5.8, we have that $A_i \supseteq \bigcup_{j \in J} B'_j$ for all $i \in I$. Let $\mathscr{M}(\sum_{i \in I} A_i) \cap \mathscr{M}(\sum_{j \in J} B_j) = P$, then for $\forall A \in P$ we have

$$A \in \mathscr{M}(\sum_{i \in I} A_i) \Rightarrow A \supseteq \bigcup_{j \in J} B'_j \Rightarrow A \supseteq B'_j, \forall j \in J.$$

Also $A \in \mathscr{M}(\sum_{j \in J} B_j)$ implies that $A \supseteq B_j$ for $\forall j \in J$, it follows that,

$$A \supseteq B_j \cup B'_j = M \Rightarrow A = M.$$

That is, $\mathscr{M}(\sum_{i \in I} A_i) \cap \mathscr{M}(\sum_{j \in J} B_j) = \{M\}$.

Conversely, if $\sum_{i \in I} A_i \wedge \sum_{j \in J} B_j = \sum_{k \in K} C_k \neq M$, we have that $C_k \subset M$ for $\forall k \in K$ and $\exists k_0$, $i_0 \in I$, $j_0 \in J$ such that $C_{k_0} \supseteq (A_{i_0} \cup B_{j_0})$. Consequently, $C_{k_0} \supseteq A_{i_0}$ and $C_{k_0} \supseteq$

A_{i_0}. That is, $C_{k_0} \le \sum_{i\in I} A_i$ and $C_{k_0} \le \sum_{j\in J} B_j$. It follows that $C_{k_0} \in \mathscr{M}(\sum_{i\in I} A_i) \cap \mathscr{M}(\sum_{j\in J} B_j)$ and this contradicts that $\mathscr{M}(\sum_{i\in I} A_i) \cap \mathscr{M}(\sum_{j\in J} B_j) = \{M\}$.

4. From 1., it can be concluded that $\sum_{i\in I} A_i = \sup \mathscr{M}(\sum_{i\in I} A_i)$. Now, let

$$\mathscr{C} = \left\{ \sum_{j\in J_k} A_{kj} \,|k \in K, K, J_k \textit{ are indexing sets} \right\} \subseteq EM.$$

Then $\sum_{k\in K} \sum_{j\in J_k} A_{kj}$ is the supremum of $\mathscr{C}$. If $\sum_{i\in I} A_i \le \sum_{k\in K} \sum_{j\in J_k} A_{kj}$, for $\forall A \in \mathscr{M}(\sum_{i\in I} A_i)$, we have

$$A \le \sum_{i\in I} A_i \le \sum_{k\in K} \sum_{j\in J_k} A_{kj}.$$

Consequently, $\exists k_0 \in K$, $j_0 \in J_{k_0}$ such that $A \supseteq A_{k_0 j_0}$, it follows that, $A \le \sum_{j\in J_{k_0}} A_{kj}$. Also, each element in $\mathscr{M}(\sum_{i\in I} A_i)$ is a $\vee$—irreducible element in EM, so we have that $\mathscr{M}(\sum_{i\in I} A_i)$ is a standard minimal family of $\sum_{i\in I} A_i$.

Now, by Theorem 5.13, we see that $\mathscr{M}(\sum_{i\in I} A_i)$ is the greatest of those standard minimal family of $\sum_{i\in I} A_i$. This completes the proof. □

The above discussion shows that $(EM, \vee, \wedge)$ is a molecular lattice but not a Boolean algebra when $|M| > 1$. In point of fact, $(EM, \vee, \wedge)$ is a new type of molecular lattice which differs from Boolean algebra and fuzzy lattice based on the following theorem.

Theorem 5.16. *Let M be a set and $|M| > 1$, then $(EM, \vee, \wedge)$ is not a fuzzy lattice.*

Proof. Suppose that $(EM, \vee, \wedge)$ is a fuzzy lattice. Then there exists a conversely ordered involutory map $\sigma : EM \to EM$ such that the strong De Morgan law holds in EM by. It follows that $M^\sigma = \varnothing$, $\varnothing^\sigma = M$ since $\varnothing$ and M are identity element and zero element of EM, respectively.

Let $A \in 2^M$, $A \ne \varnothing$, $A \ne M$ and $A' = M - A$. We have $A \wedge A' = M$. Consequently, by De Morgan law, we have that

$$A^\sigma + (A')^\sigma = (A \wedge A')^\sigma = M^\sigma = \varnothing.$$

Next, without any loss of generality, we assume that $A^\sigma = \sum_{i\in I} A_i$ and $(A')^\sigma = \sum_{j\in J} B_j$, it follows that

$$\varnothing = A^\sigma + (A')^\sigma = \sum_{i\in I} A_i + \sum_{j\in J} B_j = \sum_{k\in I \bigsqcup J} C_k.$$

Where, $\exists u \in I \bigsqcup J$ such that $\varnothing \supseteq C_u$, i.e., $C_u = \varnothing$. According to Proposition 5.8, we have

$$A^\sigma = \sum_{i\in I} A_i = \varnothing, \ \ \textit{when } u \in I,$$

or

$$(A')^\sigma = \sum_{j\in J} B_j = \varnothing, \ \ \textit{when } u \in J.$$

If $A^{\sigma}=\varnothing$, we have that $A=M$ and this contradicts that $A\neq M$; if $(A')^{\sigma}=\varnothing$, we have that $A'=M$, i.e.,$A=\varnothing$, which contradicts that $A\neq\varnothing$. Thus, $(EM,\vee,\wedge)$ is not a fuzzy lattice and this completes the proof. □

Proposition 5.10. *Let M be a non-empty set and $(EM,\vee,\wedge)$ be the EI algebra over M. Let*

$$SEM=\left\{\sum_{i\in I}A_i \mid A_i\in 2^M-\{\varnothing\}, i\in I, I \text{ is anyindexing set}\right\}\subseteq EM.$$

Then $(SEM,\vee,\wedge)$ be a sublattice of $(EM,\vee,\wedge)$ with minimum element M and maximum element $\sum_{m\in M}\{m\}$. Furthermore $(SEM,\vee,\wedge)$ is a molecular lattice.

Theorem 5.17. *Let M be a non-empty set and $(EM,\vee,\wedge)$ be the EI algebra over M. $(SEM,\vee,\wedge)$ is not a Boolean algebra if $|M|>2$. Nevertheless, if $|M|=2$, i.e., $M=\{m_1,m_2\}$, then $(SEM,\vee,\wedge)$ is a Boolean algebra.*

Proposition 5.11. *Let M be a non-empty set and $(EM,\vee,\wedge)$ be the EI algebra over M. In lattice $(SEM,\vee,\wedge)$, $M-\{m\}$ $(\forall m\in M)$ is an atom of SEM and $\sum_{m\in M-\{e\}}\{m\}$ $(\forall e\in M)$ is a dual atom of SEM.*

Proof. First, we prove that $M-\{m\}$ is an atom of SEM for any $m\in M$. Let $\sum_{i\in I}A_i\in SEM$ and $M-\{m\}\geq\sum_{i\in I}A_i\geq M$. Then $A_i\supseteq M-\{m\}$ for any $i\in I$, that is, for any $i\in I$, either $A_i=M-\{m\}$ or $A_i=M$. If there exists $u\in I$ such that $A_u=M-\{m\}$, we have $A_i\supseteq A_u$ for any $i\in I$. Consequently, $\sum_{i\in I}A_i=M-\{m\}$ by Definition 5.2. If $\forall i\in I$, $A_i=M$, we have $\sum_{i\in I}A_i=M$. Thus, for any $m\in M$, any $\sum_{i\in I}A_i\in SEM$, $M-\{m\}\geq\sum_{i\in I}A_i\geq M\Rightarrow\sum_{i\in I}A_i=M-\{m\}$ or $\sum_{i\in I}A_i=M$. This implies that $M-\{m\}$ $(\forall m\in M)$ is an atom of SEM.

Secondly, we prove that $\sum_{m\in M-\{e\}}\{m\}$ $(\forall e\in M)$ is a dual atom of SEM. For any $e\in M$, let $\sum_{i\in I}A_i\in SEM$ and $\sum_{m\in M}\{m\}\geq\sum_{i\in I}A_i\geq\sum_{m\in M-\{e\}}\{m\}$. Then for any $m\in M-\{e\},\exists u\in I$ such that $\{m\}\supseteq A_u$, that is $A_u=\{m\}$, and for any $i\in I,\exists m\in M$ such that $A_i\supseteq\{m\}$. If $A_i\neq\{e\}$ for any $i\in I$, we have $\sum_{i\in I}A_i=\sum_{m\in M-\{e\}}\{m\}$. If $\exists u\in I$ such that $A_u=\{e\}$, we have $\sum_{i\in I}A_i=\sum_{m\in M}\{m\}$. Thus for any $e\in M$, if $\sum_{m\in M}\{m\}\geq\sum_{i\in I}A_i\geq\sum_{m\in M-\{e\}}\{m\}$, then either $\sum_{i\in I}A_i=\sum_{m\in M}\{m\}$ or $\sum_{i\in I}A_i=\sum_{m\in M-\{e\}}\{m\}$. Thus $\sum_{m\in M-\{e\}}\{m\}$ $(\forall e\in M)$ is a dual atom of SEM. With this observation the proof is complete. □

Proposition 5.12. *Let M be a non-empty set and $(EM,\vee,\wedge)$ be the EI algebra over M. If $|M|>2$, then in lattice $(SEM,\vee,\wedge)$, the following assertions hold: for $\sum_{i\in I}A_i,\sum_{j\in J}B_j\in SEM$,*

1. *if $\sum_{i\in I}A_i\wedge\sum_{j\in J}B_j=M$, then $A_i\supseteq\bigcup_{j\in J}B'_j$, $\forall i\in I$ and $B_j\supseteq\bigcup_{I\in I}A'_i$, $\forall j\in J$;*
2. *if $\sum_{i\in I}A_i\vee\sum_{j\in J}B_j=\sum_{m\in M}\{m\}$, $\sum_{i\in I}A_i\wedge\sum_{j\in J}B_j=M$, then either $\sum_{i\in I}A_i=\sum_{m\in M}\{m\},\sum_{j\in J}B_j=M$ or $\sum_{j\in J}B_j=\sum_{m\in M}\{m\},\sum_{i\in I}A_i=M$.*

Proof. 1. The proof of 1. is the same as that shown in Proposition 5.8.

2. For $\sum_{i\in I}A_i,\sum_{j\in J}B_j\in SEM$, let $\sum_{i\in I}A_i\vee\sum_{j\in J}B_j=\sum_{m\in M}\{m\}$ and $\sum_{i\in I}A_i\wedge\sum_{j\in J}B_j=M$. For any $m\in M$ either $\exists u\in I$ such that $A_u=\{m\}$ or $\exists v\in J$ such that

$B_v = \{m\}$. Now, without any loss in generality, we may assume that $A_u = \{m\}$, then by 1. we have

$$\{m\} \supseteq \bigcup_{j \in J} B_j' = (\bigcap_{i \in J} B_j)' \Rightarrow \bigcap_{i \in J} B_j \supseteq M - \{m\} \Rightarrow B_j \supseteq M - \{m\}, \forall j \in J.$$

By $|M| > 2$, we know that $|B_j| \geq |M - \{m\}| > 1$. Whence for any $m \in M$ there exists $k \in I$ such that $A_k = \{m\}$, that is $\sum_{i \in I} A_i = \sum_{m \in M} \{m\}$. From 1., we also have

$$B_j \supseteq \bigcup_{i \in I} A_i' = (\bigcap_{i \in I} A_i)' = \varnothing' = M,\ \forall j \in J.$$

Thus $\sum_{j \in J} B_j = M$.

Similarly, if $\exists v \in J$ such that $B_v = \{m\}$, then we can prove that $\sum_{j \in J} B_j = \sum_{m \in M} \{m\}$ and $\sum_{i \in I} A_i = M$. This completes the proof. □

Theorem 5.18. *Let M be a non-empty set and $(EM, \vee, \wedge)$ be the EI algebra over M. Then $(SEM, \vee, \wedge)$ is a fuzzy lattice.*

Proof. For any $\sum_{i \in I} A_i, \sum_{j \in J} B_j \in SEM$, let

$$A_i = \{m_{iu} \mid u \in I_i\},\ i \in I,\ \ B_j = \{m_{jv} \mid v \in J_j\}, j \in J.$$

Since $(SEM, \vee, \wedge)$ is a molecular lattice, hence $\sigma : SEM \to SEM$ defined as follows is a map.

$$(\sum_{i \in I} A_i)^\sigma = \bigwedge_{i \in I} \left(\sum_{u \in I_i} \{m_{iu}\} \right) = \sum_{f \in \prod_{i \in I} I_i} \left(\bigcup_{i \in I} \{m_{if(i)}\} \right) \tag{5.41}$$

$$(\sum_{j \in J} B_j)^\sigma = \bigwedge_{j \in J} \left(\sum_{v \in J_j} \{m_{jv}\} \right) = \sum_{f \in \prod_{j \in J} J_j} \left(\bigcup_{j \in J} \{m_{jf(j)}\} \right) \tag{5.42}$$

Now,it is sufficient to show that "σ " satisfies Definition 5.11. If $\sum_{j \in J} B_j \geq \sum_{i \in I} A_i$, then $\forall i \in I, \exists k_i \in J$ such that

$$A_i = \{m_{iu} \mid u \in I_i\} \supseteq B_{k_i} = \{m_{k_i v} \mid v \in J_{k_i}\}.$$

Thus we have a map $\Psi : I \to J$, for any $i \in I$, $\Psi(i) = k_i$ and a map $\Phi_{k_i} : J_{k_i} \to I_i$ such that for any $v \in J_{k_i}$, $m_{i\Phi(v)} = m_{k_i v}$. Furthermore, let $\mathscr{J}_1 = \Psi(I)$ and $\mathscr{J}_2 = J - \mathscr{J}_1$. We have $J = \mathscr{J}_1 \cup \mathscr{J}_2$, $\mathscr{J}_1 \cap \mathscr{J}_2 = \varnothing$ and $\mathscr{J}_1 \neq \varnothing$. It is clear that $\Psi^{-1}(\mathscr{J}_1) = I$. Thus $\prod_{j \in J} J_j = \left(\prod_{j \in \mathscr{J}_1} J_j \right) \times \left(\prod_{j \in \mathscr{J}_2} J_j \right)$ and in (5.42), for any $g \in \prod_{j \in J} J_j$ we have

$$\bigcup_{j \in J} \{m_{jg(j)}\} = \left(\bigcup_{j \in \mathscr{J}_1} \{m_{jg_1(j)}\} \right) \cup \left(\bigcup_{j \in \mathscr{J}_2} \{m_{jg_2(j)}\} \right), \tag{5.43}$$

where $g_1 \in \prod_{j\in\mathscr{J}_1} J_j$ and $g_2 \in \prod_{j\in\mathscr{J}_2} J_j$ such that for any $j \in \mathscr{J}_1$, $g(j) = g_1(j)$ and for any $j \in \mathscr{J}_2$, $g(j) = g_2(j)$. For any $j \in \mathscr{J}_1$, we have

$$m_{\Psi^{-1}(j)\Phi_j(g_1(j))} = m_{jg_1(j)}. \tag{5.44}$$

Since $\bigcup_{j\in\mathscr{J}_1} \Psi^{-1}(j) = \Psi^{-1}(\mathscr{J}_1) = I$, hence for any $i \in I$ there exists a $j \in \mathscr{J}_1$ such that $i = \Psi^{-1}(j)$. Let $f \in \prod_{i\in I} I_i$ be defined as follows: for any $i \in I$, $f(i) = \Phi_j(g_1(j)) \in I_i$. Then by (5.44) and (5.43), we have

$$\bigcup_{i\in I}\{m_{if(i)}\} = \bigcup_{j\in\mathscr{J}_1}\{m_{jg_1(j)}\} \subseteq \left(\bigcup_{j\in\mathscr{J}_1}\{m_{jg_1(j)}\}\right) \cup \left(\bigcup_{j\in\mathscr{J}_2}\{m_{jg_2(j)}\}\right) = \bigcup_{j\in J}\{m_{jg(j)}\}.$$

This implies that $\left(\sum_{j\in J} B_j\right)^\sigma \le (\sum_{i\in I} A_i)^\sigma$. "$\sigma$ " satisfies condition 1 of Definition 5.11.

Next, by Theorem 5.1, we have

$$\begin{aligned}\left(\left(\sum_{i\in I} A_i\right)^\sigma\right)^\sigma &= \left(\sum_{f\in\prod_{i\in I} I_i}\left(\bigcup_{i\in I}\{m_{if(i)}\}\right)\right)^\sigma \\ &= \bigwedge_{f\in\prod_{i\in I} I_i}\left(\sum_{i\in I}\{m_{if(i)}\}\right) \\ &= \sum_{i\in I}\left(\bigcup_{u\in I_i}\{m_{iu}\}\right) = \sum_{i\in I} A_i,\end{aligned}$$

that is,"σ " satisfies condition 2 of Definition 5.11. With this our proof is complete. □

5.4.3 Algebraic Structures of EI^n Algebra

It is worth noting that EI^n algebra is more general algebraic class, which includes EI algebra. Of course, the algebraic structure of EI^n algebra is more complicated than that of the EI algebra. In this section, similarly to the study of the EI algebra, we will complete an exhaustive study of properties of the lattice $(EX_1...X_nM, \vee, \wedge)$.

The following proposition is a straightforward consequence of the already introduced definitions.

Proposition 5.13. *Let X_1, ..., X_n, M be n+1 non-empty sets and $(EX_1...X_nM, \vee, \wedge)$ be the EI^{n+1} algebra over X_1, ..., X_n, M. For any $\alpha = \sum_{i\in I}(u_{1i}...u_{ni}A_i), \beta = \sum_{j\in J}(v_{1i}...v_{ni}B_j) \in EX_1...X_nM$, the following assertions hold:*

1. $\alpha \wedge \beta = \beta \wedge \alpha$, $\alpha \vee \beta = \beta \vee \alpha$; *(Commutativity)*
2. $(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma)$, $(\alpha \vee \beta) \vee \gamma = \alpha \vee (\beta \vee \gamma)$; *(Associativity)*
3. $(\alpha \wedge \beta) \vee \alpha = \alpha$, $(\alpha \vee \beta) \wedge \alpha = \alpha$; *(Absorbance)*

4. $\alpha \wedge (\beta \vee \gamma) = (\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$, $\alpha \vee (\beta \wedge \gamma) = (\alpha \vee \beta) \wedge (\alpha \vee \gamma)$; *(Distributivity)*
5. $\alpha \wedge \alpha = \alpha$, $\alpha \vee \alpha = \alpha$. *(Idempotence)*

Proposition 5.14. *Let X_1, X_2,..., X_n, M be $n+1$-non-empty sets and $(EX_1X_2...X_nM$, $\vee$, $\wedge)$ be the EI^{n+1} algebra over X_1, X_2,..., X_n, M. Let*

$$SEX_1X_2...X_nM = \Big\{ \sum_{i \in I} (u_{1i}...u_{ni}A_i) \mid A_i \in 2^M - \{\varnothing\}, \\ i \in I, u_{ri} \in 2^{X_r}, r = 1,2,...,n, i \in I, I \text{ is anyindexing set} \Big\} \subseteq EM.$$

Then $(SEX_1X_2...X_nM, \vee, \wedge)$ is a sublattice of $(EX_1X_2...X_nM, \vee, \wedge)$ with minimum element $\varnothing...\varnothing M$ and maximum element $\sum_{m \in M} X_1...X_n\{m\}$. Moreover $(SEX_1X_2...X_nM$, $\vee$, $\wedge)$ is a molecular lattice.

Proposition 5.15. *Let X_1, X_2,..., X_n,M be $n+1$-non-empty sets and $(EX_1X_2...X_nM$, $\vee$, $\wedge)$ be the EI^{n+1} algebra over X_1, X_2,..., X_n, M. In lattice $(SEX_1X_2...X_nM, \vee, \wedge)$,*

$$\mathscr{A} = \{\varnothing...\varnothing(M - \{m\}) \mid m \in M\} \cup \{\varnothing...x_k...\varnothing M \mid x_k \in X_k, k = 1,2,...,n\}$$

is the set of all atoms of the lattice $SEX_1X_2...X_nM$, and

$$\mathscr{A}^d = \Big\{ \sum_{m \in M - \{e\}} (X_1...X_n\{m\}) + [X_1...(X_k - \{x_k\})...X_n\{e\}] \mid x_k \in X_k, \\ e \in M, \ 1 \le k \le n \Big\}$$

is the set of all dual atoms of the lattice $SEX_1X_2...X_nM$.

Proof. First, we prove that $\mathscr{A}$ is an atom of $SEX_1X_2...X_nM$. Clearly, for any $m \in M$, any $x_k \in X_k, k = 1,2,...,n$,we have

$$\varnothing...\varnothing(M - \{m\}) \ge \varnothing...\varnothing M, \ \ and \ \ \varnothing...\varnothing(M - \{m\}) \ne \varnothing...\varnothing M;$$
$$\varnothing...\{x_k\}...M \ge \varnothing...\varnothing M, \ \ and \ \ \varnothing...\{x_k\}...M \ne \varnothing...\varnothing M.$$

If there exists $\sum_{i \in I}(u_{1i}...u_{ni}A_i) \in SEX_1X_2...X_nM$ such that

$$\varnothing...\varnothing(M - \{m\}) \ge \sum_{i \in I}(u_{1i}...u_{ni}A_i),$$

then for any $i \in I$, we have $A_i \supseteq M - \{m\}$, $u_{ri} \subseteq \varnothing$ $(r = 1,2,...,n)$. It follows that, $A_i \supseteq M$ or $A_i = M - \{m\}$, $u_{ri} = \varnothing$ $(r = 1,2,...,n)$. That is, $\sum_{i \in I}(u_{1i}...u_{ni}A_i) = \varnothing...\varnothing M$ or $\sum_{i \in I}(u_{1i}...u_{ni}A_i) = \varnothing...\varnothing(M - \{m\})$.

If there exists $\sum_{i \in I}(u_{1i}...u_{ni}A_i) \in SEX_1X_2...X_nM$ such that $\varnothing...\{x_k\}...\varnothing M \ge \sum_{i \in I}(u_{1i}...u_{ni}A_i)$, then for any $i \in I$, we have $A_i \supseteq M$, $u_{ki} \subseteq \{x_k\}$, $u_{ri} \subseteq \varnothing$ $(r = 1,2,...,n, r \ne k)$. It follows that

$$A_i = M, \ u_{ri} = \varnothing \ (r = 1,2,...,n, r \ne k), u_{ki} = \varnothing \ or \ \{u_{ki}\} = \{x_k\}.$$

That is, $\sum_{i\in I}(u_{1i}...u_{ni}A_i) = \varnothing...\varnothing M$ or $\sum_{i\in I}(u_{1i}...u_{ni}A_i) = \varnothing...\{x_k\}...\varnothing M$. Therefore

$$\mathscr{A} = \{\varnothing...\varnothing(M-\{m\}) \mid m\in M\} \cup \{\varnothing...x_k...\varnothing M \mid x_k\in X_k, k=1,2,...,n\}$$

is the set of all atoms of the lattice $SEX_1X_2...X_nM$.

Secondly, let $e\in M$, $\sum_{i\in I}(u_{1i}...u_{ni}A_i) \in SEX_1X_2...X_nM$. If

$$\begin{aligned}\sum_{m\in M}(X_1...X_n\{m\}) &\geq \sum_{i\in I}(u_{1i}...u_{ni}A_i)\\ &\geq \sum_{m\in M-\{e\}}(X_1...X_n\{m\}) + [X_1...(X_k-\{x_k\})...X_n\{e\}],\end{aligned}$$

then for $m\in M-\{e\}$, we have $\sum_{i\in I}(u_{1i}...u_{ni}A_i) \geq \sum_{m\in M-\{e\}}(X_1...X_n\{m\}) + [X_1...(X_k-\{x_k\})...X_n\{e\}] \Rightarrow \forall m\in M = (M-\{e\})\cup\{e\}, \exists i_m\in I$ such that $A_m \subseteq \{m\}, u_{ri_m} \supseteq X_r$ and $A_e \subseteq \{e\}, u_{ri_e} \supseteq (X_k-\{x_k\})$, $r=1,2,...,n$. That is, $A_m = \{m\}, u_{ri_m} = X_r$ for $m\in M, m\neq e$ and $A_e=\{e\}$, either $u_{ri_e} = (X_k-\{x_k\})$ or $u_{ri_e} = X_k$, $r=1,2,...,n$. This implies that

$$\sum_{i\in I}(u_{1i}...u_{ni}A_i) = \sum_{m\in M}(X_1...X_n\{m\}) + \sum_{i\in I}(u_{1i}...u_{ni}A_i)$$

or

$$\sum_{i\in I}(u_{1i}...u_{ni}A_i) = \sum_{i\in I}(u_{1i}...u_{ni}A_i) + \sum_{m\in M-\{e\}}(X_1...X_n\{m\}) + X_1...(X_k-\{x_k\})...X_n\{e\}.$$

Furthermore by Proposition 5.1, we conclude that

$$\sum_{i\in I}(u_{1i}...u_{ni}A_i) = \sum_{m\in M}(X_1...X_n\{m\})$$

or

$$\sum_{i\in I}(u_{1i}...u_{ni}A_i) = \sum_{m\in M-\{e\}}(X_1...X_n\{m\}) + X_1...(X_k-\{x_k\})...X_n\{e\}.$$

Thus

$$\mathscr{A}^d = \left\{\sum_{m\in M-\{e\}}(X_1...X_n\{m\}) + [X_1...(X_k-\{x_k\})...X_n\{e\}] \mid x_k\in X_k, e\in M,\ 1\leq k\leq n\right\}$$

is the set of all dual atoms of the lattice $SEX_1...X_nM$, and this completes the proof. □

Proposition 5.16. *Let $X_1, \ldots, X_n$, M be n+1 non-empty sets and $(EX_1...X_nM, \vee, \wedge)$ be the EI^{n+1} algebra over X_1, ..., X_n, M. For $\alpha = \sum_{i\in I}(u_{1i}...u_{ni}A_i), \beta = \sum_{j\in J}(v_{1i}...v_{ni}B_j) \in EX_1...X_nM$, the following assertions hold:*

1. *if* $\alpha \wedge \beta = (\varnothing...\varnothing M)$, *then* $A_i \supseteq \bigcup_{j\in J} B'_j$ $(\forall i \in I)$, $B_j \supseteq \bigcup_{i\in I} A'_i$ $(\forall j \in J)$, *and* $u_{ri} \cap v_{rj} = \varnothing, (\forall i \in I, \forall j \in J, r = 1, 2, ..., n)$;
2. *if* $\alpha \vee \beta = (X_1...X_n\varnothing)$, *then* $\alpha = (X_1...X_n\varnothing)$ *or* $\beta = (X_1...X_n\varnothing)$.

Proof. Let $\alpha = \sum_{i\in I}(u_{1i}...u_{ni}A_i)$, $\beta = \sum_{j\in J}(v_{1i}...v_{ni}B_j)$. If $\alpha \wedge \beta = (\varnothing...\varnothing M)$, then

$$\sum_{i\in I, j\in J} [(u_{1i} \cap v_{1j})...(u_{ni} \cap v_{nj})(A_i \cup B_j)] = (\varnothing...\varnothing M).$$

Consequently, $\forall i \in I,\ j \in J$, we have

$$A_i \cup B_j \supseteq M, \quad u_{ri} \cap v_{rj} \subseteq \varnothing\ (r = 1, 2, \ldots, n).$$

This is,

$$A_i \cup B_j = M, \quad u_{ri} \cap v_{rj} = \varnothing\ (r = 1, 2, \ldots, n).$$

It follows that, $A_i \supseteq B'_j$ $(\forall i \in I)$, $B_j \supseteq A'_i$ $(\forall j \in J)$ and $u_{ri} \cap v_{rj} = \varnothing$ $(\forall i \in I, \forall j \in J, r = 1, 2, \ldots, n)$. This is, $A_i \supseteq \bigcup_{j\in J} B'_j$ $(\forall i \in I)$, $B_j \supseteq \bigcup_{i\in I} A'_i$ $(\forall j \in J)$ and $u_{ri} \cap v_{rj} = \varnothing$ $(\forall i \in I, \forall j \in J, r = 1, 2, \ldots, n)$.

Let $\alpha \vee \beta = (X_1...X_n\varnothing)$. Consequently, we have that either $\exists i_0 \in I$ such that $\varnothing \supseteq A_{i_0}$, $u_{ri_0} \supseteq X_r$ $(r = 1, 2, ..., n)$, this is $\varnothing = A_{i_0}$, $u_{ri_0} = X_r$ $(r = 1, 2, ..., n)$; or $\exists j_0 \in J$ such that $\varnothing \supseteq B_{j_0}$, $v_{rj_0} \supseteq X_r$ $(r = 1, 2, ..., n)$, this is $\varnothing = B_{j_0}$, $v_{rj_0} = X_r$ $(r = 1, 2, ..., n)$. Thus, $\alpha = (X_1...X_n\varnothing)$ or $\beta = (X_1...X_n\varnothing)$. This completes the proof. □

Let $\alpha = \sum_{i\in I}(u_{1i}...u_{ni}A_i), \beta = \sum_{j\in J}(v_{1i}...v_{ni}B_j)$. From Proposition 5.16, when $\alpha \vee \beta = (X_1 \ldots X_n\varnothing)$ and $\alpha \wedge \beta = (\varnothing \ldots \varnothing M)$, we have that either $\alpha = (X_1 \ldots X_n\varnothing)$ or $\beta = (X_1 \ldots X_n\varnothing)$. Now, without any loss in generality we assume that $\alpha = (X_1 \ldots X_n\varnothing)$, then $\exists k \in I$ such that

$$A_k = \varnothing, u_{rk} = X_r\ (r = 1, 2, \ldots, n).$$

It follows that $\beta = (\varnothing \ldots \varnothing M)$, from

$$\varnothing = A_k \supseteq \bigcup_{j\in J} B'_j \Rightarrow B_j = M,\ \forall j \in J,$$

and

$$u_{rk} \cap v_{rj} = \varnothing\ (r = 1, 2, \ldots, n) \Rightarrow v_{rj} = \varnothing\ (r = 1, 2, \ldots, n), \forall j \in J.$$

So we have that $(EX_1 \ldots X_nM, \vee, \wedge)$ is not a Boolean algebra when there exist at least two non-empty sets among M, X_r $(r = 1, 2, \ldots, n)$.

Theorem 5.19. *Let* $X_1, \ldots, X_n$, M *be n+1 non-empty sets and* $(EX_1...X_nM, \vee, \wedge)$ *be the* EI^{n+1} *algebra over* X_1, ..., X_n, M. *Then*

$$\{u_1...u_nA \mid A \in 2^M, u_r \in 2^{X_r},\ r = 1, 2, ..., n\}$$

is the set of all $\vee$*—irreducible elements in* $EX_1...X_nM$.

Theorem 5.20. *Let $X_1, \ldots, X_n$, M be n+1 non-empty sets and $(EX_1...X_nM, \vee, \wedge)$ be the EI^{n+1} algebra over X_1, ..., X_n, M. Then*

$$\{\sum_{m\in A}(u_1(m)...u_n(m)\{m\} \mid A \in 2^M, u_r(m) \in 2^{X_r}, r = 1,2,...,n\}$$

is the set of all $\wedge$—irreducible elements in $EX_1...X_nM$.

Proof. Let $\mathscr{P} = \{\sum_{m\in A}(u_1(m)...u_n(m)\{m\} \mid A \in 2^M, u_r(m) \in 2^{X_r}, r=1,2,...,n\}$, $\gamma = \sum_{m\in A}(u_1(m)...u_n(m)\{m\}) \in \mathscr{P}$, $\alpha = \sum_{i\in I}(u_{1i}...u_{ni}A_i)$ and $\beta = \sum_{j\in J}(v_{1i}...v_{ni}B_j)$.

First, we prove that γ is a $\wedge$—irreducible element in $EX_1...X_nM$. As a matter of fact, if there exist $\alpha, \beta \in EX_1...X_nM$ such that $\gamma = \alpha\wedge\beta$, from Proposition 5.13, we have

$$\gamma + \alpha = (\alpha\wedge\beta) + \alpha = \alpha,\ \gamma + \beta = (\alpha\wedge\beta) + \beta = \beta.$$

In the sequel, we have $\gamma \leq \alpha$, $\gamma \leq \beta$. We assume that $\gamma \neq \alpha$, $\gamma \neq \beta$, we have

$$\alpha = \gamma + \alpha_0,\ \beta = \gamma + \beta_0,$$

where $\alpha_0 = \sum_{i\in I_0}(u_{1i}...u_{ni}A_i)$ and either $\forall i \in I_0$, $A_i \cap A = \varnothing$ or $\exists r \in \{1,2,\ldots,n\}$, $\exists m \in A$ such that $u_{ri} \not\subseteq u_r(m)$, $\beta_0 = \sum_{j\in J_0}(v_{1i}...v_{ni}B_j)$ and either $\forall j \in J_0$, $B_j \cap A = \varnothing$ or $\exists r \in \{1,2,\ldots,n\}$, $\exists m \in A$ such that $v_{ri} \not\subseteq u_r(m)$.

Furthermore, from Proposition 5.13 and Theorem 5.1, we have

$$\begin{aligned}\gamma &= \alpha\wedge\beta = (\gamma+\alpha_0)\wedge(\gamma+\beta_0)\\ &= (\gamma\wedge\gamma) + (\alpha_0\wedge\gamma) + (\gamma\wedge\beta_0) + (\alpha_0\wedge\beta_0)\\ &= \sum_{m\in A}(u_1(m)...u_n(m)\{m\}) + \sum_{m\in A, i\in I_0}([u_1(m)\cap u_{1i}]...[u_n(m)\cap u_{ni}](\{m\}\cup A_i))\\ &+ \sum_{m\in A, j\in J_0}([u_1(m)\cap v_{1i}]...[u_n(m)\cap v_{ni}](\{x\}\cup B_j))\\ &+ \sum_{i\in I_0, j\in J_0}([u_{1i}\cap v_{1j}]...[u_{ni}\cap v_{nj}](A_i\cup B_j)).\end{aligned}$$

and $u_r(m)\cap u_{ri} \subseteq u_r(m)$, $\{m\}\cup A_i \supseteq \{m\}$; $u_r(m)\cap v_{rj} \subseteq u_r(m)$, $\{m\}\cup B_j \supseteq \{m\}$ for $\forall i \in I_0$, $j \in J_0$, $r = 1,2,...,n$, it follows that

$$\begin{aligned}\sum_{m\in A}(u_1(m)...u_n(m)\{m\}) = &\sum_{m\in A}(u_1(m)...u_n(m)\{m\})\\ &+ \sum_{i\in I_0, j\in J_0}([u_{1i}\cap v_{1j}]...[u_{ni}\cap v_{nj}](A_i\cup B_j))\end{aligned}$$

Consequently, $\gamma = \gamma + \delta$ (where $\delta = \sum_{i\in I_0, j\in J_0}([u_{1i}\cap v_{1j}]...[u_{ni}\cap v_{nj}](A_i\cup B_j)))$, i.e., $\gamma \geq \delta$. Therefore, we have that $\forall i \in I_0$, $j \in J_0$, $\exists m \in A$ such that $A_i\cup B_j \supseteq \{m\}$ and $u_r(m) \supseteq u_{ri}\cap v_{rj}$ $(r = 1,2,\ldots,n\ )$, i.e., $m \in A_i$ or $m \in B_j$; $u_r(m) \supseteq u_{ri}$ and $u_r(m) \supseteq v_{rj}$. This contradicts that $\gamma \neq \alpha$ and $\gamma \neq \beta$. Thus, we have $\gamma = \alpha\wedge\beta \Rightarrow \gamma = \alpha$ or $\gamma = \beta$.

Secondly, we prove that each $\wedge$—irreducible element must be in $\mathscr{P}$. Suppose that $\sum_{i\in J}(u_{1i}...u_{ni}A_i)$ is a $\wedge$—irreducible element in $EX_1...X_nM$, and $\exists k \in I$, $|A_k| > 1$. Let $B = A_k - \{x\}$ and $x \in A_k$. Since

$$\begin{aligned}
&\left[\left(\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)\right)+(u_{1k}...u_{nk}B)\right]\wedge\left[\left(\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)\right)+(u_{1k}...u_{nk}\{x\})\right]\\
&=\left[\left(\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)\right)\wedge\left(\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)\right)\right]\\
&\quad+\left[(u_{1k}...u_{nk}B)\wedge\left(\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)\right)\right]\\
&\quad+\left[\left(\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)\right)\wedge(u_{1k}...u_{nk}\{x\})\right]\\
&\quad+(u_{1k}...u_{nk}B)\wedge(u_{1k}...u_{nk}\{x\})\\
&=\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)+\sum_{i\in I-\{k\}}[(u_{1i}\cap u_{1k})...(u_{ni}\cap u_{nk})(A_i\cup B)]\\
&\quad+\sum_{i\in I-\{k\}}[(u_{1i}\cap u_{1k})...(u_{ni}\cap u_{nk})(A_i\cup\{x\})]+(u_{1k}...u_{nk}A_k)
\end{aligned}$$

and for $\forall i \in I-\{k\}$,

$$A_i\cup B\supseteq A_i,\ A_i\cup\{x\}\supseteq A_i,\ u_{ri}\cap u_{rk}\subseteq u_{ri}(r=1,2,\ldots,n),$$

it follows that

$$\begin{aligned}
&\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)\\
&=\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)+\sum_{i\in I-\{k\}}[(u_{1i}\cap u_{1k})...(u_{ni}\cap u_{nk})(A_i\cup B)]\\
&+\sum_{i\in I-\{k\}}[(u_{1i}\cap u_{1k})...(u_{ni}\cap u_{nk})(A_i\cup\{x\})],
\end{aligned}$$

that is,

$$\begin{aligned}
&\left[\left(\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)\right)+(u_{1k}...u_{nk}B)\right]\wedge\left[\left(\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)\right)+(u_{1k}...u_{nk}\{x\})\right]\\
&=\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)+(u_{1k}...u_{nk}A_k)=\sum_{i\in I}(u_{1i}...u_{ni}A_i).
\end{aligned}$$

Consequently, from $\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)$ is a $\wedge$—irreducible element in $EX_1\ldots X_nM$, we should have that either

$$\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)=\left(\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)\right)+(u_{1k}...u_{nk}B)$$

or

$$\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)=\left(\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)\right)+(u_{1k}...u_{nk}\{x\}).$$

But, from $B\subset A_k$, $\{x\}\subset A_k$, in fact

$$\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)\neq\left(\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)\right)+(u_{1k}...u_{nk}B)$$

and

$$\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)\neq\left(\sum_{i\in I-\{k\}}(u_{1i}...u_{ni}A_i)\right)+(u_{1k}...u_{nk}\{x\}),$$

This is a palpable contradiction.Therefore, it is concluded that each $\wedge$—irreducible element must be in $\mathscr{P}$. With this the proof is complete. □

Theorem 5.21. *Let* $X_1, \ldots, X_n$, *M be n+1 non-empty sets and* $(EX_1...X_nM,\vee,\wedge)$ *be the* EI^{n+1} *algebra over* X_1, ..., X_n, *M. For* $\sum_{i\in I}(u_{1i}...u_{ni}A_i)\in EX_1...X_nM$, *let*

$$\mathscr{P}_i=\{(v_1...v_nB)\mid B\supseteq A_i, v_r\subseteq u_{ri}, r=1,2,\ldots,n\}\subseteq EX_1...X_nM,$$

$S_i\subseteq\mathscr{P}_i(i\in I)$. *Then* $\mathscr{B}=\bigcup_{i\in I}\mathscr{P}_i$ *is a standard minimal family of* $\sum_{i\in I}(u_{1i}...u_{ni}A_i)$ *in molecular lattice* $(EX_1...X_nM,\vee,\wedge)$.

From the above considerations, we know that $(EX_1\ldots X_nM,\vee,\wedge)$ is a molecular lattice but not a Boolean algebra. With analogy lattice $(EM,\vee,\wedge)$, we have the following theorem.

Theorem 5.22. *Let* $X_1, \ldots, X_n$, *M be n+1 non-empty sets and* $|M|>1$. *Then* $(EX_1\ldots X_nM,\vee,\wedge)$ *is not a fuzzy lattice.*

Proof. First, from Theorem 5.1, $\forall\sum_{i\in I}(u_{1i}...u_{ni}A_i),\sum_{j\in J}(v_{1i}...v_{ni}B_j)\in EX_1\ldots X_nM$, we have

$$\begin{aligned}&\sum_{i\in I}(u_{1i}...u_{ni}A_i)\geq\sum_{j\in J}(v_{1i}...v_{ni}B_j)\\ \Leftrightarrow&\left[\sum_{i\in I}(u_{1i}...u_{ni}A_i)\right]\vee\left[\sum_{j\in J}(v_{1i}...v_{ni}B_j)\right]=\sum_{i\in I}(u_{1i}...u_{ni}A_i)\\ \Leftrightarrow&\left[\sum_{i\in I}(u_{1i}...u_{ni}A_i)\right]\wedge\left[\sum_{j\in J}(v_{1i}...v_{ni}B_j)\right]=\sum_{j\in J}(v_{1i}...v_{ni}B_j).\end{aligned}$$

Suppose that $(EX_1\ldots X_nM,\vee,\wedge)$ is a fuzzy lattice. Then, by Definition 5.11 and Theorem 5.12, there exists a conversely ordered involutory mapping $\sigma:EX_1\ldots X_nM\to EX_1\ldots X_nM$ such that strong De Morgan law holds in $EX_1\ldots X_nM$. It follows that

$(\varnothing...\varnothing M)^\sigma = (X_1 \dots X_n\varnothing)$ and $(X_1 \dots X_n\varnothing)^\sigma = (\varnothing...\varnothing M)$, because $(X_1 \dots X_n\varnothing)$ and $(\varnothing...\varnothing M)$ are identity element and zero element of $EX_1 \dots X_nM$, respectively.

Let $A \in 2^M$, $A \neq \varnothing$, $A \neq M$, $A' = M - A$, $u_r \in 2^{X_r}$, $u_r \neq \varnothing$, $u_r \neq X_r, u'_r = M - A$, $u_r \in 2^{Xr}$, $u_r \neq \varnothing$, $u_r \neq X_r$, $u'_r = X_r - u_r$, $r = 1, 2, \dots, n$. Then, by Definition 5.11, we have that

$$\begin{aligned}(u_1 \dots u_nA)^\sigma + (u'_1...u'_nA')^\sigma &= [(u_1 \dots u_nA) \wedge (u'_1...u'_nA')]^\sigma = (\varnothing...\varnothing M)^\sigma \\ &= (X_1...X_n\varnothing).\end{aligned}$$

Now, without any loss in generality, we assume that

$$(u_1...u_nA)^\sigma = \sum_{i\in I}(u_{1i}...u_{ni}A_i),\ (u'_1...u'_nA')^\sigma = \sum_{j\in J}(v_{1j}...v_{nj}B_j).$$

It follows that

$$\begin{aligned}(X_1...X_n\varnothing) &= (u_1 \dots u_nA)^\sigma + (u'_1...u'_nA')^\sigma \\ &= \sum_{i\in I}(u_{1i}...u_{ni}A_i) + \sum_{j\in J}(v_{1j}...v_{nj}B_j) \\ &\triangleq \sum_{k\in I\sqcup J}(w_{1k}...w_{nk}C_k).\end{aligned}$$

Therefore, $\exists l \in I \sqcup J$ such that $\varnothing \supseteq C_l$ and $X_r \subseteq w_{ri}$ $(r = 1, 2, ..., n)$, that is, $C_l = \varnothing$ and $X_r = w_{ri}$ $(r = 1, 2, ..., n)$. From Proposition 5.16, we have that either

$$(u_1...u_nA)^\sigma = \sum_{i\in I}(u_{1i}...u_{ni}A_i) = (X_1...X_n\varnothing),\ \ when\ i \in I,$$

or

$$(u_1...u_nA)^\sigma = \sum_{j\in J}(v_{1i}...v_{ni}B_j) = (X_1...X_n\varnothing),\ \ when\ j \in J.$$

If $(u_1 \dots u_nA)^\sigma \neq (X_1 \dots X_n\varnothing)$, we have that $(u'_1...u'_nA') = (\varnothing \dots \varnothing M)$ and this contradicts that $A \neq \varnothing$, $A \neq M$, $u_r \neq \varnothing$, $u_r \neq X_r$ $(r = 1, 2, \dots, n)$. If $(u_1...u_nA)^\sigma \neq (X_1...X_n\varnothing)$, we have that $(u_1 \dots u_nA)^\sigma = (X_1 \dots X_n\varnothing)$ and this also contradicts that $A \neq \varnothing, A \neq M$, $u_r \neq \varnothing$, $u_r \neq X_r$ $(r = 1, 2, \dots, n)$. Thus, $(EX_1 \dots X_nM, \vee, \wedge)$ is not a fuzzy lattice and this completes the proof. □

So far, whether $(SEX_1 \dots X_nM, \vee, \wedge)$ is a fuzzy lattice has not been proved. However, the following theorem shows that the sublattice $(SEX_1^- \dots X_n^-M, \vee, \wedge)$ is a fuzzy lattice, where

$$SEX_1^- \dots X_n^-M = \{\sum_{i\in I}(u_1...u_nA_i)\ \ \mid A_i \in 2^M - \{\varnothing\},\ u_r \in 2^{X_r},$$
$$r = 1, 2, ..., n, I\ is\ any\ indexing\ set\}.$$

It is clear that $(SEX_1^- \dots X_n^-M, \vee, \wedge)$ is a sublattice of $(SEX_1 \dots X_nM, \vee, \wedge)$.

Theorem 5.23. *Let $X_1, \ldots, X_n$, M be n+1 non-empty sets and $|M| > 1$. Then $(SEX_1^- \ldots X_n^- M, \vee, \wedge)$ is a fuzzy lattice.*

Proof. For any $\sum_{i\in I}(u_1...u_nA_i), \sum_{j\in J}(v_1...v_nB_j) \in SEX_1^- \ldots X_n^- M$, let

$$A_i = \{m_{ip} \mid p \in I_i\},\ i \in I,\ \ B_j = \{m_{jq} \mid q \in J_j\}, j \in J.$$

Since $(SEX_1^- \ldots X_n^- M, \vee, \wedge)$ is a molecular lattice, hence $\sigma : SEX_1^- \ldots X_n^- M \to SEX_1^- \ldots X_n^- M$ defined as follows is a map.

$$[\sum_{i\in I}(u_1...u_nA_i)]^\sigma = \bigwedge_{i\in I}\left(\sum_{m\in A_i}(u_1'...u_n'\{m\})\right) \tag{5.45}$$

$$= \sum_{f\in\prod_{i\in I} I_i}\left(u_1'...u_n'\bigcup_{i\in I}\{m_{if(i)}\}\right) \tag{5.46}$$

$$[\sum_{j\in J}(v_1...v_nB_j)]^\sigma = \sum_{f\in\prod_{j\in J} J_j}\left(v_1'...v_n'\bigcup_{j\in J}\{m_{jf(j)}\}\right) \tag{5.47}$$

Now, it is sufficient to show that "σ " satisfies Definition 5.11. If $\sum_{i\in I}(u_1...u_nA_i) \le \sum_{j\in J}(v_1...v_nB_j)$, then $\forall i \in I, \exists k_i \in J$ such that

$$A_i = \{m_{ip} \mid p \in I_i\} \supseteq B_{k_i} = \{m_{k_iq} \mid q \in J_{k_i}\},\ \ u_r \subseteq v_r,\ r = 1,2,...,n.$$

Thus we have a map $\Psi : I \to J$, for any $i \in I$, $\Psi(i) = k_i$ and a map $\Phi_{k_i} : J_{k_i} \to I_i$ such that for any $v \in J_{k_i}$, $m_{i\Phi(v)} = m_{k_iv}$. Furthermore, let $\mathscr{J}_1 = \Psi(I)$ and $\mathscr{J}_2 = J - \mathscr{J}_1$. We have $J = \mathscr{J}_1 \cup \mathscr{J}_2$, $\mathscr{J}_1 \cap \mathscr{J}_2 = \varnothing$ and $\mathscr{J}_1 \ne \varnothing$. It is clear that $\Psi^{-1}(\mathscr{J}_1) = I$. Thus $\prod_{j\in J} J_j = \left(\prod_{j\in\mathscr{J}_1} J_j\right) \times \left(\prod_{j\in\mathscr{J}_2} J_j\right)$ and in (5.47), for any $g \in \prod_{j\in J} J_j$ we have

$$\bigcup_{j\in J}\{m_{jg(j)}\} = \left(\bigcup_{j\in\mathscr{J}_1}\{m_{jg_1(j)}\}\right) \cup \left(\bigcup_{j\in\mathscr{J}_2}\{m_{jg_2(j)}\}\right), \tag{5.48}$$

where $g_1 \in \prod_{j\in\mathscr{J}_1} J_j$ and $g_2 \in \prod_{j\in\mathscr{J}_2} J_j$ such that for any $j \in \mathscr{J}_1$, $g(j) = g_1(j)$ and for any $j \in \mathscr{J}_2$, $g(j) = g_2(j)$. For any $j \in \mathscr{J}_1$, we have

$$m_{\Psi^{-1}(j)\Phi_j(g_1(j))} = m_{jg_1(j)}.$$

Since $\bigcup_{j\in\mathscr{J}_1}\Psi^{-1}(j) = \Psi^{-1}(\mathscr{J}_1) = I$, hence for any $i \in I$ there exists a $j \in \mathscr{J}_1$ such that $i = \Psi^{-1}(j)$. Let $f \in \prod_{i\in I} I_i$ be defined as follows: for any $i \in I$, $f(i) = \Phi_j(g_1(j)) \in I_i$. Then in virtue of (5.49) and (5.48) we have

$$\bigcup_{i\in I}\{m_{if(i)}\}=\bigcup_{j\in\mathscr{J}_1}\{m_{jg_1(j)}\}\subseteq\left(\bigcup_{j\in\mathscr{J}_1}\{m_{jg_1(j)}\}\right)\cup\left(\bigcup_{j\in\mathscr{J}_2}\{m_{jg_2(j)}\}\right)=\bigcup_{j\in J}\{m_{jg(j)}\}.$$

Furthermore, by $u_r\subseteq v_r,\ r=1,2,...,n$, we have $u'_r\supseteq v'_r,\ r=1,2,...,n$. This implies that $\left[\sum_{j\in J}(v_1...v_nB_j)\right]^{\sigma}\leq[(\sum_{i\in I}(u_1...u_nA_i)]^{\sigma}$. "$\sigma$" satisfies condition 1 of Definition 5.11.

Next, by Theorem 5.1, we have

$$\begin{aligned}\left[\left(\sum_{i\in I}(u_1...u_nA_i)\right)^{\sigma}\right]^{\sigma}&=\left(\sum_{f\in\prod_{i\in I}I_i}\left(u'_1...u'_n\bigcup_{i\in I}\{m_{if(i)}\}\right)\right)^{\sigma}\\&=\bigwedge_{f\in\prod_{i\in I}I_i}\left(\sum_{i\in I}(u_1...u_n\{m_{if(i)}\})\right)\\&=\sum_{i\in I}\left((u_1...u_n\bigcup_{u\in I_i}\{m_{iu}\})\right)=\sum_{i\in I}A_i,\end{aligned}$$

that is,"σ" satisfies condition 2 of Definition 5.11. With this the proof is completed. □

5.4.4 Algebraic Structures of $E^{\#}I^n$ Algebra

In this section, the further exploration of the algebraic properties of $E^{\#}I^n$ algebra are exhaustively discussed. First,the expressions of special elements such as $\wedge-$irreducible elements, $\vee-$irreducible elements, atoms and dual atoms, are given in $E^{\#}I$ and $E^{\#}I^n$ algebra. Then, it is proved that $E^{\#}I^n$ algebra is a new structure which is different from EI^n algebra.

In what follows, if $(S,\vee,\wedge)$ is a AFS algebra, by the symmetric property of operation $\vee$, $\wedge$, when the two operations are exchanged, we denote the one as $(^*S,\vee,\wedge)$ or briefly as $^*S$. Now the definition of the dual lattice $(^*EM,\vee,\wedge)$ of $(EM,\vee,\wedge)$ is as follows.

Theorem 5.24. *Let M be a non-empty set.If we define binary operations $\vee,\wedge$ as follows. For $\sum_{i\in I}A_i,\sum_{j\in J}B_j\in EM$,*

$$\sum_{i\in I}A_i\vee\sum_{j\in J}B_j=\sum_{i\in I,j\in J}(A_i\cup B_j),\tag{5.49}$$

$$\sum_{i\in I}A_i\wedge\sum_{j\in J}B_j=\sum_{k\in I\sqcup J}C_k=\sum_{i\in I}A_i+\sum_{j\in J}B_j.\tag{5.50}$$

Then $(^*EM,\vee,\wedge)$ is a molecular lattice, in which for $\sum_{i\in I}A_i,\sum_{j\in J}B_j\in EM$, $\sum_{i\in I}A_i\leq\sum_{j\in J}B_j$ if and only if for any B_j $(j\in J)$ there exists A_k $(i\in I)$ such that $B_j\supseteq A_k$.

Its proof, which remains as an exercise, is similar to that of Theorem 4.1. $(^*EM,\vee,\wedge)$ is called the $^*EI$ algebra over set M.

Similarly, by the symmetric property of operation "$\vee$, $\wedge$", the dual lattice $({}^*E^\#X, \vee, \wedge)$ can be defined if only we alter the operations "$\vee$, $\wedge$". Thus for $\sum_{i\in I} a_i$, $\sum_{j\in J} b_j \in E^\#X$, $\sum_{i\in I} a_i \le \sum_{j\in J} b_j \Longleftrightarrow \forall b_j (j \in J)$ there exists $a_k (k \in I)$ such that $a_k \supseteq b_j$. $({}^*E^\#X, \vee, \wedge)$ is called ${}^*E^\#I$ algebra over set X. In the similar manner, we have ${}^*EI^n$ algebra and ${}^*E^\#I^n$ algebra.

In what follows, we establish isomorphism between the lattices EX and $E^\#X$. Thus the properties about the lattice $(E^\#X, \vee, \wedge)$ can be directly derived from those about $(EX, \vee, \wedge)$. By Proposition 5.10, we know that $(SEM, \vee, \wedge)$ is a sublattice of the lattice $(EM, \vee, \wedge)$. It is not difficult to check that the lattice $(SE^\#X, \vee, \wedge)$ is a sublattice of $(E^\#X, \vee, \wedge)$, where

$$SEM = \left\{ \sum_{i\in I} A_i \mid A_i \in 2^M - \{\varnothing\} \right\},$$

$$SE^\#X = \left\{ \sum_{i\in I} a_i \mid a_i \in 2^X - \{X\} \right\}.$$

Theorem 5.25. *Let X be a non-empty set. Let $(EX, \vee, \wedge)$ and $(E^\#X, \vee, \wedge)$ be the EI algebra and $E^\#I$ algebra over X. Then the lattice $(EX, \vee, \wedge)$ and the lattice $(E^\#X, \vee, \wedge)$ are isomorphism.*

Proof. For any $\sum_{i\in I} a_i \in E^\#X$, the map $\varphi : E^\#X \to EX$ is defined as follows.

$$\varphi(\sum_{i\in I} a_i) = \sum_{i\in I} a_i'.$$

First, we prove that φ is a map from $E^\#X$ to EX. For $\sum_{i\in I} a_i$, $\sum_{j\in J} b_j \in E^\#X$, suppose that $\sum_{i\in I} a_i = \sum_{j\in J} b_j$. Then we have

$$\sum_{i\in I} a_i \le \sum_{j\in J} b_j \ \ and \ \ \sum_{i\in I} a_i \ge \sum_{j\in J} b_j.$$

As

$$\textstyle\sum_{i\in I} a_i \le \sum_{j\in J} b_j \Longleftrightarrow \forall i \in I, \exists k \in J \text{ such that } a_i \subseteq b_k, \text{ that is } b_k' \subseteq a_i'.$$

This implies that the following assertions hold in the lattice $(EX, \vee, \wedge)$

$$\varphi\left(\textstyle\sum_{i\in I} a_i\right) = \sum_{i\in I} a_i' \le \sum_{j\in J} b_j' = \varphi\left(\sum_{j\in J} b_j\right).$$

Similarly,we get

$$\textstyle\sum_{i\in I} a_i \ge \sum_{j\in J} b_j \Longleftrightarrow \varphi\left(\sum_{i\in I} a_i\right) \ge \varphi\left(\sum_{j\in J} b_j\right).$$

Next,

$$\textstyle\sum_{i\in I} a_i = \sum_{j\in J} b_j \Longleftrightarrow \varphi\left(\sum_{i\in I} a_i\right) = \varphi\left(\sum_{j\in J} b_j\right).$$

Therefore,φ is a map from $E^\#X$ to EX.

Let $\sum_{i\in I} a_i,\ \sum_{j\in J} b_j \in E^{\#}X$. It is clear that

$$\varphi\left(\sum_{i\in I} a_i \vee \sum_{j\in J} b_j\right) = \varphi\left(\sum_{i\in I} a_i\right) \vee \varphi\left(\sum_{j\in J} b_j\right).$$

From Theorem 5.2, we get

$$\varphi\left(\sum_{i\in I} a_i \wedge \sum_{j\in J} b_j\right) = \varphi(\sum_{i\in I, j\in J} (a_i \cap b_j)) = \sum_{i\in I, j\in J} (a_i \cap b_j)'.$$

By the De Morgan law and Theorem 4.1, it follows that

$$\varphi\left(\sum_{i\in I} a_i \wedge \sum_{j\in J} b_j\right) = \sum_{i\in I, j\in J} (a_i \cap b_j)' = \sum_{i\in I, j\in J} a_i' \cup b_j' = \varphi\left(\sum_{i\in I} a_i\right) \wedge \varphi\left(\sum_{j\in J} b_j\right).$$

Therefore,φ is an isomorphism from $(E^{\#}X, \geq)$ to $(EX, \geq)$. The proof is complete. □

The following theorem whose proof is left as an exercise can be proved in a similar way as discussed earlier.

Theorem 5.26. *Let X be a non-empty set. Let $(^*EX, \vee, \wedge)$ and $(^*E^{\#}X, \vee, \wedge)$ be the $^*EI$ algebra and $^*E^{\#}I$ algebra over X. Then the following assertions hold.*

(1) The sublattices $(SE^{\#}X, \vee, \wedge)$ and $(SEX, \vee, \wedge)$ are isomorphism;
(2) The sublattices $(^*SE^{\#}X, \vee, \wedge)$ and $(^*SEX, \vee, \wedge)$ are isomorphism;
(3) The lattices $(^*EX, \vee, \wedge)$ and $(^*E^{\#}X, \vee, \wedge)$ are isomorphism.

Thus by Theorem 5.26, the corresponding properties of $E^{\#}I$ algebra and $^*EI$ algebra can be directly obtained in virtue of the ideas we have for EI algebra. However, in what follows, we will show that $E^{\#}I^n$ is a new algebra families which are quite different from EI^n algebra. Given this, it is necessary to discuss the properties of the $E^{\#}I^n$ algebra.

Theorem 5.27. *Let $X_1, ..., X_n$ be n non-empty sets and $(E^{\#}X_1...X_n, \vee, \wedge)$ be the $E^{\#}I^n$ algebra over the sets $X_1, ..., X_n$. Then the following assertions hold.*

(1) The set of all $\vee$-irreducible elements in $E^{\#}X_1...X_n$ is

$$\mathscr{I}^{\vee} = \left\{u_1...u_n \mid u_k \in 2^{X_k}, k = 1, 2, \cdots n\right\}.$$

(2) The set of all $\wedge$-irreducible elements in $E^{\#}X_1...X_n$ is

$$\mathscr{I}^{\wedge} = \left\{\sum_{x\in D_k} [X_1 \cdots (X_k - \{x\}) \cdots X_n] \ \mid\ D_k \subseteq X_k, k = 1, 2, ..., n\right\}.$$

Here $\sum_{x\in D_k} [X_1 \cdots (X_k - \{x\}) \cdots X_n] \triangleq X_1 X_2 \cdots X_n$ if for some $k \in \{1, 2, ..., n\}$, $D_k = \varnothing$.

(3) The set of all atoms in $E^{\#}X_1...X_n$ is

$$\mathscr{A}=\{\varnothing\cdots\varnothing\{x_k\}\varnothing\cdots\varnothing \mid x_k\in X_k, k=1,2,\cdots,n\}\cup\{\varnothing\cdots\varnothing\}.$$

(4) The set of all dual atom in $E^{\#}X_1...X_n$ is

$$\mathscr{A}^d=\left\{\sum_{1\le k\le n}\sum_{x\in X_k}[X_1\cdots(X_k-\{x\})\cdots X_n]\right\}.$$

Proof. (1) It directly results from Definition 2.9.

(2) Let $\sum_{i\in I}(u_{1i}...u_{ni})\in E^{\#}X_1...X_n$ be a $\wedge-$irreducible element. Furthermore, let $\sum_{i\in I}(u_{1i}...u_{ni})$ be irreducible, i. e., for any $k\in I$, $\sum_{i\in I}(u_{1i}...u_{ni})\ne\sum_{i\in I-\{k\}}(u_{1i}...u_{ni})$. First, we prove that

$$|X_k-u_{ki}|\le 1,\forall k\in\{1,2,\cdots,n\},\forall i\in I.$$

Suppose there exists a $i_0\in I$ such that $|X_k-u_{ki_0}|\ge 2$, for some $k\in\{1,2,\cdots,n\}$. Let $x_{i_0},y_{i_0}\in X_k-u_{ki_0}$, that is $x_{i_0},y_{i_0}\notin u_{ki_0}$. Let $\alpha,\beta\in E^{\#}X_1...X_n$ be given as

$$\alpha=\sum_{i\in I-\{i_0\}}(u_{1i}...u_{ni})+[u_{1i_0}...u_{k-1,i_0}(\{x_{i_0}\}\cup u_{ki_0})u_{k+1,i_0}\cdots u_{ni_0}],$$
$$\beta=\sum_{i\in I-\{i_0\}}(u_{1i}...u_{ni})+[u_{1i_0}...u_{k-1,i_0}(\{y_{i_0}\}\cup u_{ki_0})u_{k+1,i_0}\cdots u_{ni_0}].$$

In lattice $(E^{\#}X_1...X_n,\vee,\wedge)$, it is clear that $\sum_{i\in I}(u_{1i}...u_{ni})\le\alpha$, $\sum_{i\in I}(u_{1i}...u_{ni})\le\beta$. Since $\sum_{i\in I}(u_{1i}...u_{ni})$ is irreducible, hence $\sum_{i\in I}(u_{1i}...u_{ni})\ne\alpha$, $\sum_{i\in I}(u_{1i}...u_{ni})\ne\beta$. According to Theorem 5.2, one has

$$\begin{aligned}\alpha\wedge\beta=&\sum_{i\in I-\{i_0\}}(u_{1i}...u_{ni})+[\sum_{i\in I-\{i_0\}}(u_{1i}...u_{ni})]\wedge[u_{1i_0}...(\{x_{i_0}\}\cup u_{ki_0})\cdots u_{ni_0}]\\&+[\sum_{i\in I-\{i_0\}}(u_{1i}...u_{ni})]\wedge[u_{1i_0}...(\{y_{i_0}\}\cup u_{ki_0})\cdots u_{ni_0}]\\&+[u_{1i_0}...(\{x_{i_0}\}\cup u_{ki_0})\cdots u_{ni_0}]\wedge[u_{1i_0}...(\{y_{i_0}\}\cup u_{ki_0})\cdots u_{ni_0}]\\&=\sum_{i\in I}(u_{1i}...u_{ni}).\end{aligned}$$

This contradicts the assumption that $\sum_{i\in I}(u_{1i}...u_{ni})$ is an $\wedge$–irreducible element. Therefore, we have

$$|X_k-u_{ki}|\le 1,\forall k\in\{1,2,\cdots,n\},\forall i\in I.$$

This implies that $|X_k-u_{ki_0}|=0$ or 1. If for some $i_0\in I$, there exist at least two $k_1,k_2\in\{1,2,\cdots,n\}$ such that $|X_{k_1}-u_{k_1i_0}|=|X_{k_2}-u_{k_2i_0}|=1$. Without loosing generality, let $k_1=1,k_2=2$. Then, let $u_{1i_0}=X_1-\{a_{1i_0}\}$ and $u_{2i_0}=X_2-\{a_{2i_0}\},a_{1i_0}\in X_1,a_{2i_0}\in X_2$. Consequently, it follows that

$$\sum_{i\in I}(u_{1i}...u_{ni}) = \sum_{i\in I-\{i_0\}}(u_{1i}...u_{ni}) + (X_1-\{a_{1i_0}\})(X_2-\{a_{2i_0}\})u_{3i_0}...u_{ni_0}.$$

Let γ, ν be in $E^{\#}X_1...X_n$,

$$\gamma = \sum_{i\in I-\{i_0\}}(u_{1i}...u_{ni}) + (X_1-\{a_{1i_0}\})X_2u_{3i_0}...u_{ni_0},$$
$$\nu = \sum_{i\in I-\{i_0\}}(u_{1i}...u_{ni}) + X_1(X_2-\{a_{2i_0}\})u_{3i_0}...u_{ni_0}.$$

We have

$$\lambda \wedge \nu = \sum_{i\in I}(u_{1i}...u_{ni}).$$

It is clear that $\sum_{i\in I}(u_{1i}...u_{ni}) \neq \gamma$ and $\sum_{i\in I}(u_{1i}...u_{ni}) \neq \nu$ because $\sum_{i\in I}(u_{1i}...u_{ni})$ is irreducible. This conflicts with the assumption that $\sum_{i\in I}(u_{1i}...u_{ni})$ is a $\wedge$– irreducible element in lattice $(E^{\#}X_1...X_n,\vee,\wedge)$. Thus for any $i\in I$ there exists a unique $k_i\in\{1,2,...,n\}$ such that $|X_{k_i}-u_{k_i i}|=1$ and $|X_j-u_{ji}|=0$ for any $j\neq k_i, j\in\{1,2,...,n\}$. In other words, $u_{k_i i}=X_{k_i}-\{x_i\}$ for some $x_i\in X_{k_i}$ and for all $j\neq k_i, j\in\{1,2,...,n\}$, $u_{ji}=X_j$. This implies that

$$\sum_{i\in I}(u_{1i}...u_{ni}) = \sum_{k\in K}\sum_{x\in D_k}[X_1\cdots(X_k-\{x\})\cdots X_n],$$

where $K\subseteq\{1,2,...,n\}$, $D_k=\{x_i \mid i\in I_k\}\subseteq X_k, k=1,2,...,n$, $I_u=\{i \mid k_i=u\}$, $u=1,2,...,n$. Assume that $|K|\geq 2$. Without loosing generality, let $1,2\in K$.

$$\alpha = \sum_{k\in K}\sum_{x\in D_k}[X_1\cdots(X_k-\{x\})\cdots X_n] + [\{x_i \mid i\in I_1\}\varnothing...\varnothing] \ \ \in E^{\#}X_1\cdots X_n,$$
$$\beta = \sum_{k\in K}\sum_{x\in D_k}[X_1\cdots(X_k-\{x\})\cdots X_n] + [\varnothing\{x_i \mid i\in I_2\}\varnothing...\varnothing] \ \ \in E^{\#}X_1\cdots X_n.$$

It is clear that $\sum_{i\in I}(u_{1i}...u_{ni})\neq\alpha$ and $\sum_{i\in I}(u_{1i}...u_{ni})\neq\beta$. However, by Theorem 5.2, we have $\sum_{i\in I}(u_{1i}...u_{ni})=\alpha\wedge\beta$. This contradicts the assumption that $\sum_{i\in I}(u_{1i}...u_{ni})$ is a $\wedge$– irreducible element in lattice $(E^{\#}X_1...X_n,\vee,\wedge)$. Therefore $|K|=1$ and $\sum_{i\in I}(u_{1i}...u_{ni})\in\mathscr{I}^{\wedge}$.

Conversely, let $\gamma=\sum_{i\in I}(u_{1i}...u_{ni})\in\mathscr{I}^{\wedge}$. We prove that $\sum_{i\in I}(u_{1i}...u_{ni})$ is a $\wedge$–irreducible element in the lattice $(E^{\#}X_1...X_n,\vee,\wedge)$. If $\gamma\neq X_1X_2...X_n$, then for each $i\in I$ there exist $k_i\in\{1,2,...,n\}$ such that $u_{k_i i}=X_{k_i}-\{x_i\}$ for some $x_i\in X_{k_i}$ and for all $j\neq k_i, j\in\{1,2,...,n\}$, $u_{ji}=X_j$. For $\alpha=\sum_{l\in L}(w_{1l}\cdots w_{nl})$, $\beta=\sum_{j\in J}(v_{1j}\cdots v_{nj})\in E^{\#}X_1\cdots X_n$, assume that γ,α,β are all irreducible and $\gamma=\alpha\wedge\beta$. So $\gamma=\alpha\wedge\beta\leq\alpha$ and $\gamma=\alpha\wedge\beta\leq\beta$. Since $X_1X_2...X_n$ is the maximum element of the lattice $(E^{\#}X_1...X_n,\vee,\wedge)$, hence if $\gamma=X_1X_2...X_n$ then $\gamma=\alpha=\beta$.

In what follows, we prove that γ is a $\wedge$–irreducible element provided that $\gamma\neq X_1X_2...X_n$. Because $\gamma\leq\alpha$ and $\gamma\leq\beta$, for any $i\in I$ there exist $l_i\in L$ and $j_i\in J$ such that

$$u_{ki} \subseteq w_{kl_i},\ \ u_{ki} \subseteq v_{kj_i},\ \ k = 1,2,...,n.$$

Considering, for all $j \neq k_i, j \in \{1,2,...,n\}$, $u_{ji} = X_j$ and $u_{k_i i} = X_{k_i} - \{x_i\}$ for some $x_i \in X_{k_i}$, we have

$$\begin{aligned} & u_{ki} = w_{kl_i} = v_{kj_i} = X_k,\ \ k \neq k_i, k = 1,2,...,n; \\ & u_{k_i i} = X_{k_i} - \{x_i\} \subseteq w_{k_i l_i},\ \ u_{k_i i} = X_{k_i} - \{x_i\} \subseteq v_{k_i j_i}. \end{aligned} \tag{5.51}$$

If there exists $t \in I$ such that $w_{k_t l_t} = X_{k_t}$ or $v_{k_t j_t} = X_{k_t}$, then $\alpha = X_1X_2...X_n$ or $\beta = X_1X_2...X_n$. This implies that $\gamma = \alpha \wedge \beta = \alpha$ or $\gamma = \alpha \wedge \beta = \beta$. Thus γ is a $\wedge$–irreducible element. If for any $i \in I$, $w_{k_i l_i} \neq X_{k_i}$ and $v_{k_i j_i} \neq X_{k_i}$, then $u_{k_i i} = X_{k_i} - \{x_i\} = w_{k_i l_i},\ u_{k_i i} = X_{k_i} - \{x_i\} = v_{k_i j_i}$ for all $i \in I$. If $L - \{l_i \mid i \in I\} = \varnothing$ or $J - \{j_i \mid i \in I\} = \varnothing$ then $\gamma = \alpha$ or $\gamma = \beta$ and γ is a $\wedge$–irreducible element. Assume that $L - \{l_i \mid i \in I\} \neq \varnothing$ and $J - \{j_i \mid i \in I\} \neq \varnothing$. This implies that

$$\begin{aligned} \alpha &= \sum_{i \in I}(u_{1i}...u_{ni}) + \sum_{l \in L - \{l_i \mid i \in I\}}(w_{1l} \cdots w_{nl}), \\ \beta &= \sum_{i \in I}(u_{1i}...u_{ni}) + \sum_{j \in J - \{j_i \mid i \in I\}}(v_{1j} \cdots v_{nj}). \end{aligned}$$

Let $I_u = \{i \mid k_i = u\},\ u = 1,2,...,n$. Because both α and β are irreducible and (5.51), for any $l \in L - \{l_i \mid i \in I\}$ there exists $u \in \{1,2,...,n\}$ such that $\{x_i \mid i \in I_u\} \subseteq w_{ul}$ and for any $j \in J - \{j_i \mid i \in I\}$ there exists $q \in \{1,2,...,n\}$ such that $\{x_i \mid i \in I_q\} \subseteq v_{qj}$. Furthermore x_q. By the definition of the set $\mathscr{I}^{\wedge}$, we have $q = u$ and $\{x_i \mid i \in I_u\} \subseteq w_{ul} \cap v_{qj}$. Then for any $i \in I$ either $u_{qi} \supseteq w_{ul} \cap v_{qj}$ nor $u_{qi} \subseteq w_{ul} \cap v_{qj}$. This implies that $\gamma \neq \alpha \wedge \beta$ and contradicts to the assumption $\gamma = \alpha \wedge \beta$. Therefore $L - \{l_i \mid i \in I\} = \varnothing$ or $J - \{j_i \mid i \in I\} = \varnothing$. That follows $\gamma = \alpha$ or $\gamma = \beta$. Finally we prove that any $\gamma = \sum_{i \in I}(u_{1i}...u_{ni}) \in \mathscr{I}^{\wedge}$ is a $\wedge$–irreducible element in the lattice $(E^{\#}X_1...X_n, \vee, \wedge)$.

(3) Let $\gamma = \sum_{i \in I}(u_{1i} \cdots u_{ni}) \in E^{\#}X_1 \cdots X_n$ be an atom in $E^{\#}X_1 \cdots X_n$ and γ be irreducible. It is straightforward to note

$$\sum_{i \in I}(u_{1i} \cdots u_{ni}) \neq \varnothing\varnothing \cdots \varnothing.$$

If I contains more than one element, choose $i_0 \in I$. Let $J = I - \{i_0\}$.Thus we have

$$\varnothing\varnothing \cdots \varnothing < \sum_{i \in J}(u_{1i} \cdots u_{ni}) < \sum_{i \in I}(u_{1i} \cdots u_{ni}). \tag{5.52}$$

It contradicts that γ is $\wedge$–irreducible element. So we derive that $|I| = 1$, it follows that $\sum_{i \in I}(u_{1i}...u_{ni})$ has the form $u_1...u_n$. If there exists $k_0 \in \{1,2,\cdots,n\}$ such that $u_{k_0} \neq \varnothing_{k_0}$ and $|u_{k_0}| \geq 2$. Let $x_{k_0} \in u_{k_0}$, we have

$$\varnothing \cdots \varnothing < u_1...u_{k_0-1}(u_{k_0} - \{x_{k_0}\})u_{k_0+1}...u_n < u_1...u_n.$$

It is also contradicts that γ is $\wedge$–irreducible element. Thus we arrive that for $k = 1,2,\cdots,n$, either $|u_k| = 0$ or 1. Furthermore, if there exist two u_{k_1}, u_{k_2} such that $|u_{k_1}| = 1, |u_{k_2}| = 1$, then it follows that

$$\varnothing\cdots\varnothing < \varnothing\cdots\varnothing\{x_{k_2}\}\varnothing\cdots\varnothing < \varnothing\cdots\varnothing\{x_{k_1}\}\varnothing\cdots\varnothing\{x_{k_2}\}\varnothing\cdots\varnothing < u_1...u_n.$$

This implies that that there exists unique $k_0 \in \{1,2,...,n\}$ such that $|u_{k_0}| = 1$ and for all $k \neq k_0,\ k \in \{1,2,...,n\}$, $u_k = \varnothing$. Therefore,we have that $\gamma = u_1...u_n \in \mathscr{A}$.

Conversely, we prove that any $\gamma \in \mathscr{A}$ is an atom element, that is

$$\gamma = \varnothing\cdots\varnothing\{x_k\}\varnothing\cdots\varnothing,\ x_k \in X_k,\ k = 1,2\cdots n.$$

Let $\sum_{j\in J}(v_{1j}...v_{nj}) \in E^{\#}X_1...X_n, k = 1,2,\cdots,n$, such that

$$\sum_{j\in J}(v_{1j}...v_{nj}) \leq \varnothing\cdots\varnothing\{x_k\}\varnothing\cdots\varnothing \tag{5.53}$$

From the Theorem 5.2, for any $j \in J$, we have that

$$v_{1j} \subseteq \varnothing,...,v_{k-1,j} \subseteq \varnothing,\ v_{kj} \subseteq \{x_k\},\ v_{k+1,j} \subseteq \varnothing,...,v_{nj} \subseteq \varnothing.$$

It follows that

$$\sum_{j\in J}(v_{1j}...v_{nj}) = \varnothing\cdots\varnothing\{x_k\}\varnothing\cdots\varnothing \ \ or\ \ \sum_{j\in J}(v_{1j}...v_{nj}) = \varnothing\cdots\varnothing.$$

In either case,we know that, $\varnothing\cdots\varnothing\{x_i\}\varnothing\cdots\varnothing$ is an atom in $E^{\#}X_1...X_n$. Therefore any $\gamma \in \mathscr{A}$ is an atom element.

(4) Let $\gamma = \sum_{i\in I}(u_{1i}...u_{ni}) \in E^{\#}X_1...X_n$ be a dual atom in $E^{\#}X_1...X_n$ and γ is irreducible. It is clear that

$$\sum_{i\in I}(u_{1i}...u_{ni}) \neq X_1...X_n. \tag{5.54}$$

If for some $k \in \{1,2,\cdots,n\}, \exists i_0 \in I$ such that $|X_k - u_{ki_0}| \geq 2$, then choose $x_k \notin u_{ki_0}, x_k \in X_k$, and $k = 1,2,\cdots,n$. Let $J = I - \{i_0\}$. Then we have

$$\sum_{i\in I}(u_{1i}...u_{ni}) < \sum_{i\in J}(u_{1i}...u_{ni}) + \left[u_{1i_0}...u_{k-1i_0}(u_{ki_0}\cup\{x_k\})u_{k+1i_0}\cdots u_{ni_0}\right] < X_1...X_n.$$

This contradicts the assumption that $\sum_{i\in I}(u_{1i}...u_{ni})$ is a dual atom. It implies that

$$|X_k - u_{ki}| = 0 \ \ or\ \ |X_k - u_{ki}| = 1,\ \forall i \in I,\ \forall k \in \{1,2\cdots n\}.$$

If there exists some $i_0 \in I$ such that for all $k = 1,2,\cdots,n$, $|X_k - u_{ki_0}| = 0$, then we have $u_{1i_0}\cdots u_{ni_0} = X_1\cdots X_n$. It forces $\sum_{i\in I}(u_{1i}\cdots u_{ni}) = X_1\cdots X_n$. This is in conflict with the assumption. Therefore, $\forall i \in I$, there exists certain $k_i \in \{1,2,\cdots,n\}$ such that $|X_{k_i} - u_{k_i i}| = 1$. It means that there exists $x_{k_i i} \in X$ such that $u_{k_i i} = X_{k_i} - \{x_{k_i i}\}$. If $\forall i \in I$, there exists unique $k_i \in \{1,2,\cdots,n\}$ such that $|X_{k_i} - u_{k_i i}| = 1$, then from the discussion above, we know that

$$\sum_{i\in I}(u_{1i}\cdots u_{ni}) = \sum_{1\le k\le n}\sum_{x\in D_k}[X_1\cdots(X_k-\{x\})\cdots X_n],\ \ D_k\subseteq X_k,\ \ k=1,2,...,n.$$

If for some $k_0 \in \{1,2,...,n\}$, $X_{k_0} - D_{k_0} \neq \varnothing$, then for $x_{k_0} \in X_{k_0} - D_{k_0}$ we have

$$\begin{aligned}\sum_{i\in I}(u_{1i}\cdots u_{ni}) &< \sum_{1\le k\le n}\sum_{x\in D_k}[X_1\cdots(X_k-\{x\})\cdots X_n] + \left[X_1\cdots X_{k_0-1}\{x_{k_0}\}X_{k_0+1}\cdots X_n\right]\\ &< X_1\cdots X_n.\end{aligned}$$

This contradicts the assumption that $\sum_{i\in I}(u_{1i}...u_{ni})$ is a dual atom. Therefore $X_{k_0} - D_{k_0} \neq \varnothing$ and the dual atom element $\gamma \in \mathscr{A}$.

Conversely, we prove that any $\gamma \in \mathscr{A}$ is a dual atom element. Let $\sum_{i\in I}(u_{1i}...u_{ni}) \in E^{\#}X_1...X_n$ such that

$$\sum_{1\le k\le n}\sum_{x\in X_k}[X_1\cdots(X_k-\{x\})\cdots X_n] < \sum_{i\in I}(u_{1i}...u_{ni}) < X_1\cdots X_n. \tag{5.55}$$

Suppose $\sum_{i\in I}(u_{1i}...u_{ni}) \neq X_1...X_n$, then we have that every item $u_{1i}...u_{ni} \neq X_1...X_n$ for all $i \in I$. Thus for any $i \in I$ there exists $k_i \in \{1,2,\cdots,n\}$ such that $X_{k_i} \neq u_{k_i i} \subseteq X_{k_i}$, it follows that, there exists element $x_{k_i i} \in X_{k_i}$ such that $x_{k_i i} \notin u_{k_i i}$. This implies that $u_{k_i i} \subseteq X_{k_i} - \{x_{k_i i}\}$, which follows that

$$u_{1i} \subseteq X_1,\ ...,\ u_{k_i i} \subseteq X_{k_i} - \{x_{k_i i}\},\ ...,\ u_{ni} \subseteq X_n,\ \ \forall i \in I.$$

This means that

$$\sum_{i\in I}(u_{1i}\cdots u_{ni}) \le \sum_{1\le i\le n}\sum_{j\in J_i}[X_1\cdots(X_i-\{x_{ij}\})\cdots X_n]$$

which contradicts (5.55). Therefore, we have

$$\sum_{i\in I}(u_{1i}\cdots u_{ni}) = \sum_{1\le k\le n}\sum_{x\in X_k}[X_1\cdots(X_k-\{x\})\cdots X_n].$$

This indicates that $\sum_{1\le k\le n}\sum_{x\in X_k}[X_1\cdots(X_k-\{x\})\cdots X_n]$ is a dual atom in $E^{\#}X_1\cdots X_n$. □

Example 5.2. Let $X_1 = \{a,b\}, X_2 = \{c,d\}$,then $X_1\{c\}+X_1\{d\}+\{a\}X_2+\{b\}X_2$ is the only dual atom element in $E^{\#}X_1X_2$,while $X_1\{c\}$, $X_1\{d\}$, $\{a\}X_2$, $\{b\}X_2$, $X_1\{c\}+X_1\{d\}$, $\{a\}X_2+\{b\}X_2$ are all $\wedge-$irreducible elements in $E^{\#}X_1X_2$.

In what follows, we will show that $E^{\#}I^n (n \ge 2)$ algebra $E^{\#}X_1...X_n$ is indeed a new structure which is quite different from EI^{n+1} algebra $EY_1...Y_mM$, where $X_1,...,X_n,Y_1,...,Y_m,M$ are sets.

Lemma 5.2. *Let $(L_1,\ge),(L_2,\ge)$ be two lattices and f be an isomorphism from L_1 to L_2. Then for $x \in L_1$, x is a $\vee-$irreducible element in L_1 if and only if $f(x)$ is a $\vee-$irreducible element in L_2; and x is a $\wedge-$irreducible element in L_1 if and only if $f(x)$ is a $\wedge-$irreducible element in L_2.*

Theorem 5.28. *Let $X_1,...,X_n,Y_1,...,Y_m,M$ be $n+m+1$ non-empty sets. Then there is no isomorphism from $(EY_1...Y_mM,\vee,\wedge)$ to $(E^{\#}X_1...X_n,\vee,\wedge)$ for any $n \geq 1$ and $m \geq 1$.*

Its proof remains as an exercise. The proceeding theorem states that under no circumstances the two structures are the same, so $(E^{\#}X_1...X_n,\vee,\wedge)$ is the different algebra structure from $(EX_1...X_nM,\vee,\wedge)$.

Theorem 5.29. *Let $X_1,...,X_n$ be n non-empty sets and $(E^{\#}X_1...X_n,\vee,\wedge)$ be the $E^{\#}I^n$ algebra over $X_1,...,X_n$. Then the following assertions hold.*

(1) If there exists $i_0 \in \{1,2,\cdots,n\}$ such that $|X_{i_0}|>1$, then $(E^{\#}X_1...X_n,\vee,\wedge)$ is not a fuzzy lattice.
(2) If for all $i \in \{1,2,\cdots,n\}$ such that $|X_i|=1$, then $E^{\#}X_1...X_n$ is a boolean algebra.

Let $X_1,X_2,...,X_n$ be n non-empty sets and $(E^{\#}X_1...X_n,\vee,\wedge)$ be the $E^{\#}I^n$ algebra over $X_1,...,X_n$. The subset $SE^{\#}X_1...X_n \subseteq E^{\#}X_1...X_n$ is defined as follows.

$$SE^{\#}X_1...X_n = \left\{ \sum_{i\in I}(A_{1i}...A_{ni}) | A_{ki} \in 2^{X_k} - \{X_k\}, k=1,2,\cdots,n \right\}. \tag{5.56}$$

It is clear that $(SE^{\#}X_1...X_n,\vee,\wedge)$ is a sublattice of the lattice $(E^{\#}X_1...X_n,\vee,\wedge)$. Along the direction we have discussed for $(SEX_1...X_n,\vee,\wedge)$, the similar algebraic properties of the lattice $(E^{\#}X_1...X_n,\vee,\wedge)$ could be explored, which are left as open problems.

5.5 Combinatoric Properties of the AFS Structure

In this section, we study some combinatoric properties of AFS structures which can be applied to the analysis of complex systems. As we have already pointed out that the AFS structure is a special combinatoric system. The combinatoric techniques outlined in Section 1.6 are considered to study the AFS structure.

Definition 5.12. Let X and M be sets, (M,τ,X) be an AFS structure. Let $V=M$, $E=X\times X$. If the map $f_\tau : E \to V$ is defined as follows: for any $x_1,x_2 \in X$, $f_\tau((x_1,x_2)) = \tau(x_1,x_2) \cap \tau(x_2,x_2)$. Then $(M,f_\tau,X\times X)$ is a system described by Definition 1.48. It is called the *combinatoric system induced by the AFS structure* (M,τ,X), and denoted as $(M,f_\tau,X\times X)$.

Considering $f_\tau((x_1,x_2)) = \tau(x_1,x_2) \cap \tau(x_2,x_2)$ in Definition 5.12, $\tau(x_1,x_2) \cap \tau(x_2,x_2)$ is the block (a set of simple concepts in M) associating to (x_1,x_2) which describes the relationship from x_1 to x_2. So the simple concepts in $\tau(x_1,x_2)$ which x_2 does not belong have to be excluded from $\tau(x_1,x_2)$. Thus x_2 belongs to any simple concept in the block and the membership degree of x_1 belonging to any simple concept in the block is larger than or equal to that of x_2. Definition 5.12 builds a link between the AFS theory and the combinatorics. In virtue of this important association, many developed combinatoric techniques can be applied to analysis AFS structure of data. As an example, in what follows, we present how to decompose a complex AFS structure to some simple AFS structures via combinatoric theory.

Definition 5.13. Let X and M be sets, (M,τ,X) be an AFS structure. (M,τ,X) is called a *connected AFS structure* if $(M,f_\tau,X\times X)$, the system induced by AFS structure (M,τ,X), is a connected system. $V\subseteq M, U\subseteq X$, if $(V,f_\tau,U\times U)$ is a connected component of $(M,f_\tau,X\times X)$, then sub-AFS structure $(V,\tau|_{U\times U},U)$ is called a *connected component* of (M,τ,X).

Definition 5.14. Let X and M be sets and (M,τ,X) be an AFS structure. If $M=\bigcup_{i\in I}V_i$, $X=\bigcup_{i\in I}U_i$, $V_i\neq\varnothing$, $V_i\cap V_j=\varnothing$, $U_i\cap U_j=\varnothing$, $i\neq j, i,j\in I$, $(V_i,\tau|_{U_i\times U_i},U_i)$ is the connected components of (M,τ,X), then the *direct sum of connected components of* $(V_i,\tau|_{U_i\times U_i},U_i)$, $i\in I$ is defined as follows.

$$(M,\tau,X)=\otimes_{i\in I}(V_i,\tau|_{U_i\times U_i},U_i).$$

By Theorem 1.56, we know that this direct sum for any given AFS structure is unique. For any $i\in I$, since $(V_i,f_\tau,U_i\times U_i)$ is a connected component of $(M,f_\tau,X\times X)$, hence for any $x,y\in U_i$,

$$V_i\supseteq f_\tau(x,y)=\tau(x,y)\cap\tau(y,y)\neq\varnothing.$$

This implies that for any $x,y\in U_i$ in the AFS structure $(V_i,\tau|_{U_i\times U_i},U_i)$ there exist some simple concepts in V_i such that x is associated to y. However, for $i,j\in I$, $i\neq j$, for any $x\in U_i$ in the AFS structure $(V_i,\tau|_{U_i\times U_i},U_i)$ and $y\in U_j$ in the AFS structure $(V_j,\tau|_{U_j\times U_j},U_j)$,

$$f_\tau(x,y)=\tau(x,y)\cap\tau(y,y)=\varnothing.$$

This implies that there does not exists any simple concept in M such that an object in U_i in $(V_i,\tau|_{U_i\times U_i},U_i)$ is associated to an object in U_j in $(V_j,\tau|_{U_j\times U_j},U_j)$. Thus the AFS structure of a complex system will be decomposed into some independent sub AFS structures which are the most simple AFS structures and cannot be decomposed further.

Theorem 5.30. *Let X and M be sets, (M,τ,X) be an AFS structure. Then there exist $V_i\subseteq M, U_i\subseteq X$, $i\in I$ such that*

$$(M,\tau,X)=\otimes_{i\in I}(V_i,\tau|_{U_i\times U_i},U_i).$$

Proof. In virtue of Theorem 1.56, for the system $(M,f_\tau,X\times X)$, we know that there exist $V_i\subseteq M$, $E_i\subseteq X\times X$, $i\in I$, such that

$$V_i\cap V_j=\varnothing,\ E_i\cap E_j=\varnothing,\ i\neq j,i,j\in I, M=\bigcup_{i\in I}V_i,\ X\times X=\bigcup_{i\in I}E_i,$$

and

$$(M,f_\tau,X\times X)=\oplus_{i\in I}(V_i,f|_{E_i},E_i),$$

where $(V_i, f|_{E_i}, E_i)$ is the maximum connected sub-system of system $(M, f_\tau, X \times X)$. Let

$$i_1(E_i) = \{x \mid \exists y \in X, (x,y) \in E_i\} \subseteq X,$$
$$i_2(E_i) = \{y \mid \exists x \in X, (x,y) \in E_i\} \subseteq X.$$

If $V_i \neq \varnothing$, then $\forall x \in i_1(E_i), \exists y \in i_2(E_i)$, such that

$$f_\tau((x,y)) = \tau(x,y) \cap \tau(y,y) \subseteq V_i.$$

Since $V_i \neq \varnothing$ and $(V_i, f|_{E_i}, E_i)$ is the maximum connected sub-system of system $(M, f_\tau, X \times X)$, hence

$$f_\tau((x,y)) = \tau(x,y) \cap \tau(y,y) \neq \varnothing.$$

If $(x,x) \notin E_i$, then there exists j such that $(x,x) \in E_j$ and

$$f_\tau((x,x)) = \tau(x,x) \cap \tau(x,x) = \tau(x,x) \subseteq V_j.$$

By AX1 of Definition 4.5, we know that the following conclusions hold $f_\tau((x,y)) \subseteq \tau(x,y) \subseteq \tau(x,x)$. Therefore we have

$$V_i \cap V_j \supseteq \tau(x,x) \cap f_\tau((x,y)) = \tau(x,x) \cap \tau(x,y) \cap \tau(y,y) = f_\tau((x,y)) \neq \varnothing.$$

This fact contradicts $V_i \cap V_j = \varnothing$. Therefore $(x,x) \in E_i$ and $x \in i_2(E_i)$. It also implies that $i_1(E_i) \subseteq i_2(E_i)$ if $V_i \neq \varnothing$. If $V_i \neq \varnothing$, then $\forall y \in i_2(E_i), \exists x \in i_1(E_i)$, such that

$$f_\tau((x,y)) = \tau(x,y) \cap \tau(y,y) \subseteq V_i.$$

Since $(V_i, f|_{E_i}, E_i)$ is the maximum connected sub-system of system $(M, f_\tau, X \times X)$ and $V_i \neq \varnothing$, hence

$$f_\tau((x,y)) = \tau(x,y) \cap \tau(y,y) \neq \varnothing.$$

If $(y,y) \notin E_i$, then there exists j such that $(y,y) \in E_j$, and

$$f_\tau((y,y)) = \tau(y,y) \cap \tau(y,y) = \tau(y,y) \subseteq V_j.$$

We have

$$V_i \cap V_j \supseteq \tau(y,y) \cap f_\tau((x,y)) = \tau(y,y) \cap \tau(x,y)\tau(y,y) = f_\tau((x,y)) \neq \varnothing.$$

It contradicts $V_i \cap V_j = \varnothing$. Therefore $(y,y) \in E_i$ and $y \in i_1(E_i)$. This implies that $i_2(E_i) \subseteq i_1(E_i)$ if $V_i \neq \varnothing$. Thus we prove $i_1(E_i) = i_2(E_i)$ if $V_i \neq \varnothing, i \in I$. Let $i_1(E_i) = U_i$, i.e., $E_i = i_1(E_i) \times i_2(E_i) = U_i \times U_i$. On the other hand, if $V_i = \varnothing$. Because $(V_i, f_\tau|_{E_i}, E_i)$ is the maximum connected sub-system of system $(M, f_\tau, X \times X)$, E_i is a singleton set $E_i = \{(x,y)\}$. This implies $\tau(x,y) \cap \tau(y,y) = \varnothing$. For the AFS structure $(V_i, \tau|_{U_i \times U_i}, U_i)$, let $(x,y) \in U_i \times U_i$. Since $\tau(x,x) \subseteq V_i$, hence $\tau(x,y) \subseteq \tau(x,x) \subseteq V_i$

from AX1 of Definition 4.5. This shows that for each $i \in I$, $(V_i, \tau|_{U_i \times U_i}, U_i)$ is a sub-AFS structure of (M, τ, X) and $(V_i, f|_{E_i}, E_i)$, which is a connected component of the AFS structure (M, τ, X), is the combinatoric system of the sub AFS structure $(V_i, \tau|_{U_i \times U_i}, U_i)$. Therefore we have

$$(M, \tau, X) = \otimes_{i \in I} (V_i, \tau|_{U_i \times U_i}, U_i)$$

and the proof is complete. □

Definition 5.15. Let X and Y be sets. Let the AFS structure (Y, τ_1, X) be compatible with AFS structure $(X, \tau_2, Y))$ (refer to Definition 5.4). If

$$\begin{aligned}(Y, \tau_1, X) &= \oplus_{i \in I} (Y_i^{(1)}, \tau_1|_{X_i^{(1)} \times X_i^{(1)}}, X_i^{(1)}),\\ (X, \tau_2, Y) &= \oplus_{i \in J} (X_i^{(2)}, \tau_2|_{Y_i^{(2)} \times Y_i^{(2)}}, Y_i^{(2)}),\end{aligned}$$

and $I = J$, $X_i^{(2)} = X_i^{(1)} = X_i, Y_i^{(2)} = Y_i^{(1)} = Y_i$, for any $i \in I = J$, then the direct sum of the cognitive space $((Y, \tau_1, X), (X, 2, Y))$ is defined as follows.

$$((Y, \tau_1, X), (X, \tau_2, Y)) = \oplus_{i \in I} ((Y_i, \tau_1|_{X_i \times X_i}, X_i), (X_i, \tau_2|_{Y_i \times Y_i}, Y_i)).$$

Theorem 5.31. *Let X and Y be sets. Let the AFS structure (Y, τ_1, X) be compatible with AFS structure $(X, \tau_2, Y))$ (refer to Definition 5.4). Then there exist $X_i \subseteq X$, $Y_i \subseteq Y$, $i \in I$ such that*

$$((Y, \tau_1, X), (X, \tau_2, Y)) = \oplus_{i \in I} ((Y_i, \tau_1|_{X_i \times X_i}, X_i), (X_i, \tau_2|_{Y_i \times Y_i}, Y_i)).$$

And $\forall x \in X_j$, $\forall \zeta = \sum_{u \in U} (\prod_{m \in A_u} m) \in E(Y - Y_i)$, $i, j \in I, i \neq j$, the following assertions hold

(1) $\sum_{u \in U} (\prod_{m \in A_u} m)(x) = \sum_{u \in U} A_u^{\tau_1}(x) A_u = \sum_{u \in U} \varnothing A_u \in EXY.$

(2) $\sum_{u \in U} (\prod_{m \in A_u} m)(x) = \sum_{u \in U} A_u^{\tau_1}(x) \{x\}^{\tau}(A_u) A_u = \sum_{u \in U} \varnothing\varnothing A_u \in EXYY.$

(3) $\mu_\zeta(x) = 0$ *if* $\mu_\zeta(.)$ *is a coherence membership function.*

Proof. Because of Theorem 5.30, for (Y, τ_1, X) and (X, τ_2, Y), we have

$$\begin{aligned}(Y, \tau_1, X) &= \oplus_{i \in I} (Y_i^{(1)}, \tau_1|_{X_i^{(1)} \times X_i^{(1)}}, X_i^{(1)}),\\ (X, \tau_2, Y) &= \oplus_{i \in J} (X_i^{(2)}, \tau_2|_{Y_i^{(2)} \times Y_i^{(2)}}, Y_i^{(2)}),\end{aligned}$$

Suppose $X_i^{(1)} \cap X_j^{(2)} \neq \varnothing$ and $x \in X_i^{(1)} \cap X_j^{(2)}$. Because each $(Y_i^{(1)}, \tau_1|_{X_i^{(1)} \times X_i^{(1)}}, X_i^{(1)})$, $i \in I$, is connected, $\forall y \in X_i^{(1)}$, $\tau_1(x,x) \subseteq Y_i^{(1)}$, $\tau_1(y,y) \subseteq Y_i^{(1)}$. Then there exist $m_1 \in \tau_1(x,x) \neq \varnothing$, $m_2 \in \tau_1(y,y) \neq \varnothing$ and a $m_1 - m_2$ path in the system

$$(Y_i^{(1)}, f_{\tau_1}|_{X_i^{(1)} \times X_i^{(1)}}, X_i^{(1)} \times X_i^{(1)}),$$

written as $v_0 = m_1, (x_1, y_1), v_1, (x_2, y_2), v_2, (x_3, y_3), ..., (x_g, y_g), v_g = m_2$ and $x_1 = x, y_g = y$. By the AX1 of Definition 4.5, we know $\tau_1(x_i, y_i) \subseteq \tau_1(x_i, x_i) \subseteq Y_i^{(1)} \Rightarrow \tau_1(x_i, y_i) \cap \tau_1(y_i, y_i) \subseteq \tau_1(x_i, x_i)$. Therefore $v_0 = m_1$, (x_1, x_1), v_1, (x_2, x_2), v_2, (x_3, x_3),..., (x_g, x_g), $v_g = m_2$ is also a $m_1 - m_2$ path and $v_{i-1}, v_i \in \tau_1(x_i, x_i)$, $i = 1, 2, ..., g$. Because (Y, τ_1, X) is compatible with (X, τ_2, Y), $x_i \in \tau_2(v_{i-1}, v_{i-1})$, $x_i \in \tau_2(v_i, v_i), i = 1, 2, ..., g$. So that $x, (v_0, v_0), x_2, (v_2, v_2), ..., x_{g-1}, (v_g, v_g), y$, is a $x - y$ path in the system $(X, f_{\tau_2}, Y \times Y)$. Since $(X_j^{(2)}, \tau_2|_{Y_j^{(2)} \times Y_j^{(2)}}, Y_j^{(2)})$ is connected, hence $y \in X_j^{(2)}$. This means $X_i^{(1)} \subseteq X_j^{(2)}$. Similarly, we can prove $X_i^{(1)} \supseteq X_j^{(2)}$. Finally we have $X_i^{(1)} = X_j^{(2)}$. Since $x \in X_i^{(1)} \cap X_j^{(2)} \neq \varnothing$, hence $\exists m \in \tau_1(x, x) \subseteq Y_i^{(1)}$. Because of the definition of compatible AFS structures, $x \in \tau_2(m, m) \Rightarrow x \in \tau_2(m, m) \cap X_j^{(2)} \neq \varnothing$. Since

$$\left(X_j^{(2)}, f_{\tau_2}|_{Y_j^{(2)} \times Y_j^{(2)}}, Y_j^{(2)} \times Y_j^{(2)}\right)$$

is the maximum connected sub-system of the system $(X, f_{\tau_2}, Y \times Y)$, $m \in Y_j^{(2)} \Rightarrow m \in Y_i^{(1)} \cap Y_j^{(2)} \neq \varnothing$. Anyway, if $X_i^{(1)} \cap X_j^{(2)} \neq \varnothing$, then $X_i^{(1)} = X_j^{(2)}$, $Y_i^{(1)} \cap Y_j^{(2)} \neq \varnothing$ and the same as the arguments of $X_i^{(1)} \cap X_j^{(2)} \neq \varnothing$, we also can prove that if $Y_i^{(1)} \cap Y_j^{(2)} \neq \varnothing$, then $Y_i^{(1)} = Y_j^{(2)}$, $X_i^{(1)} \cap X_j^{(2)} \neq \varnothing$. This implies that

$$X_i^{(1)} \cap X_j^{(2)} \neq \varnothing \Leftrightarrow Y_i^{(1)} \cap Y_j^{(2)} \neq \varnothing.$$
$$X_i^{(1)} \cap X_j^{(2)} \neq \varnothing \Rightarrow X_i^{(1)} = X_j^{(2)}.$$
$$Y_i^{(1)} \cap Y_j^{(2)} \neq \varnothing \Rightarrow Y_i^{(1)} = Y_j^{(2)}.$$

Therefore we have

$$((Y, \tau_1, X), (X, \tau_2, Y)) = \oplus_{i \in I}((Y_i, \tau_1|_{X_i \times X_i}, X_i), (X_i, \tau_2|_{Y_i \times Y_i}, Y_i)).$$

In what follows, we prove (1), (2) and (3). $\forall x \in X_j$, and $\forall \zeta = \sum_{u \in U}(\prod_{m \in A_u} m) \in E(Y - Y_i), i, j \in I, i \neq j$, we have $\tau_1(x, x) \subseteq Y_i$ and $\tau_1(x, x) \cap (Y - Y_i) = \varnothing$. This implies that

$$A_i^{\tau_1}(x) = \{y \mid y \in X, \tau_1(x, y) \supseteq A_i\} = \varnothing.$$

It follows (1).

Since for any $v \in A_i$, if $\{x\} \subseteq \tau_2(v, u) \subseteq \tau_2(v, v)$, then $v \in \tau_1(x, x)$. It contradicts that $\tau_1(x, x) \cap (Y - Y_i) = \varnothing$. Hence

$$\{x\}^{\tau_2}(A_i) = \{u \mid u \in Y,\ \tau_2(v,u) \supseteq \{x\},\ \forall v \in A_i\} = \varnothing.$$

Then we have (2).

(3) can be directly proved by (1) and Definition 4.7.

Now, we have completed the proof. □

Theorem 5.31 ensures that any cognitive space can be decomposed into the direct sum of some independent and connected sub-cognitive spaces and the membership degree (coherence membership degree or AFS algebra represented membership degree) of any sample from one space belonging to any fuzzy concept from another space is always 0. In this way, the complexity of a complex system can be greatly decreased and the system structure can be easily comprehended.

Exercises

Exercise 5.1. Let $X_1,\ldots,X_n,M$ be n+1 non-empty sets. If $A_t \subseteq A_s$, $u_{rt} \supseteq u_{rs}$, $r = 1,2,\ldots,n$, $t,s \in I$, $t \neq s$, $\sum_{i\in I}(u_{1i}\ldots u_{ni}A_i) \in EX_1\ldots X_nM$. Show the following

$$\sum_{i\in I}(u_{1i}\ldots u_{ni}A_i) = \sum_{i\in I-\{s\}}(u_{1i}\ldots u_{ni}A_i).$$

Exercise 5.2. Let $X_1,\ldots X_n,M$ be $n+1$ non-empty sets. Prove that $(EX_1\ldots X_nM,\vee,\wedge)$ forms a completely distributive lattice under the binary compositions $\vee$ and $\wedge$ defined as follows: $\forall \sum_{i\in I}(u_{1i}\ldots u_{ni}A_i)$, $\sum_{j\in J}(v_{1j}\ldots v_{nj}B_j) \in EX_1\ldots X_nM$,

$$\begin{aligned}
\sum_{i\in I}(u_{1i}\ldots u_{ni}A_i) \vee \sum_{j\in J}(v_{1j}\ldots v_{nj}B_j) &= \sum_{k\in I\sqcup J}(w_{1k}\ldots w_{nk}C_k),\\
\sum_{i\in I}(u_{1i}\ldots u_{ni}A_i) \wedge \sum_{j\in J}(v_{1j}\ldots v_{nj}B_j) &= \sum_{i\in I,j\in J}\left[(u_{1i}\cap v_{1j}\ldots u_{ni}\cap v_{nj})(A_i\cup B_j)\right],
\end{aligned}$$

where $\forall k \in I\sqcup J$, $C_k = A_k, w_{rk} = u_{rk}$ if $k \in I$ and $C_k = B_k, w_{rk} = v_{rk}$ if $k \in J$, $r = 1,2,\ldots,n$.

Exercise 5.3. Let $X_1,X_2,\ldots,X_n$ be n non-empty sets. Prove that $(E^{\#}X_1X_2\ldots X_n,\ \vee,\ \wedge)$ forms a completely distributive lattice under the binary compositions $\vee$ and $\wedge$ defined as follows:

$$\begin{aligned}
\sum_{i\in I}a_{1i}a_{2i}\ldots a_{ni} \vee \sum_{j\in J}b_{1j}b_{2j}\ldots b_{nj} &= \sum_{k\in I\sqcup J}c_{1k}c_{2k}\ldots c_{nk}\\
\sum_{i\in I}a_{1i}a_{2i}\ldots a_{ni} \wedge \sum_{j\in J}b_{1j}b_{2j}\ldots b_{nj} &= \sum_{i\in I,j\in J}(a_{1i}\cap b_{1j})(a_{2i}\cap b_{2j})\ldots(a_{ni}\cap b_{nj})
\end{aligned}$$

where $\forall k \in I\sqcup J$, $c_{rk} = a_{rk}$, $r = 1,\ldots,n$, if $k \in I$ and $c_{rk} = b_{rk}$, $r = 1,\ldots,n$, if $k \in J$.

Exercise 5.4. Let X be a set and $E^{\#}X$ be the $E^{\#}I$ algebra a over X. Show the following assertions hold: for any $\sum_{i\in I}a_i \in E^{\#}X$,

1. $\varnothing \vee \sum_{i\in I} a_i = \sum_{i\in I} a_i, \quad \varnothing \wedge \sum_{i\in I} a_i = \varnothing$;
2. $X \vee \sum_{i\in I} a_i = X, \quad X \wedge \sum_{i\in I} a_i = \sum_{i\in I} a_i$.

Exercise 5.5. Let (M,τ,X) be an AFS structure. For any given $x\in X$, if we define a mapping $\phi_x : EM \to EXM$, $\forall \sum_{i\in I}\prod_{m\in A_i} m \in EM$,

$$\phi_x(\sum_{i\in I}\prod_{m\in A_i} m) = \sum_{i\in I} A_i^{\tau}(\{x\})A_i \in EXM.$$

Prove that ϕ_x is a homomorphism from lattice $(EM,\vee,\wedge)$ to lattice $(EXM,\vee,\wedge)$.

Exercise 5.6. Let X and M be sets. (M,τ_1,X) is compatible with (X,τ_2,M). $\forall x\in X$, for any $\sum_{i\in I}\prod_{m\in A_i} m \in EM$, if we define

$$\phi_x(\sum_{i\in I}\prod_{m\in A_i} m) = \sum_{i\in I} A_i^{\tau_1}(x)\{x\}^{\tau_2}(A_i)A_i \in EXMM,$$

prove that ϕ_x is a homomorphism from lattice $(EM,\vee,\wedge)$ to lattice $(EXMM,\vee,\wedge)$, where $A_i^{\tau_1}(x)$ and $\{x\}^{\tau_2}(A_i)$ are defined by (5.8).

Exercise 5.7. Let $X_1,\ldots,X_n,M$ be $n+1$ non-empty sets, $EX_1\ldots X_nM$ be EI^{n+1} algebra over $X_1,\ldots,X_n,M$ and M be a finite set of simple concepts, S_r be a σ-algebra over X_r, $r=1,2,\ldots,n$. Prove that $\xi\vee\eta$, $\xi\wedge\eta \in \sigma(EX_1\ldots X_nM)$ for all $\xi,\eta\in\sigma(EX_1\ldots X_nM)$, i.e., $(\sigma(EX_1\ldots X_nM),\vee,\wedge)$ is a sublattice of $(EX_1\ldots X_nM,\vee,\wedge)$. Here $\sigma(EX_1\ldots X_nM)$ is defined by (5.17).

Exercise 5.8. Let $X_1,\ldots,X_n,M$ be $n+1$ non-empty sets, $EX_1\ldots X_nM$ be EI^{n+1} algebra over $X_1,\ldots,X_n,M$ and M be a finite set of simple concepts, S_r be a σ-algebra over X_r, $r=1,2,\ldots,n$. For each simple concept $\zeta\in M$, let $\mathscr{M}_\zeta$ be the measure defined by Definition 5.6 for ρ_ζ. Prove that the map $||.|| : \sigma(EX_1\ldots X_nM)\to[0,1]$ defined as follows is a fuzzy norm of the lattice $(\sigma(EX_1\ldots X_nM),\vee,\wedge)$: for any $\sum_{i\in I}(u_{1i}\ldots u_{ni}A_i)\in\sigma(EX_1\ldots X_nM)$,

$$||\sum_{i\in I}(u_{1i}\ldots u_{ni}A_i)|| = \sup_{i\in I}\left(\prod_{m\in A_i, 1\le r\le n}\mathscr{M}_m(u_{ri})\right)\in[0,1].$$

Exercise 5.9. Let M be a set. For $\sum_{i\in I}A_i, \sum_{j\in J}B_j\in EM$, if $\sum_{i\in I}A_i=\sum_{j\in J}B_j$, $\sum_{i\in I}A_i$ and $\sum_{j\in J}B_j$ are both irreducible, show $\{B_j|j\in J\}=\{A_i|i\in I\}$.

Exercise 5.10. Let $M=\{m_1, m_2, \cdots, m_n\}$. Show $|C_1(M)| = l_1(n) = 2^n$.

Exercise 5.11. Prove that $(EM,\vee,\wedge)$ is a Boolean algebra if $|M|=1$.

Exercise 5.12. Let $(L,\vee,\wedge,\sigma)$ be a fuzzy lattice. Prove that the strong De Morgan Law holds, that is, for any $a_t\in L, t\in T$,

$$\left(\bigvee_{t\in T} a_t\right)^{\sigma} = \bigwedge_{t\in T} a_t^{\sigma}, \quad \left(\bigwedge_{t\in T} a_t\right)^{\sigma} = \bigvee_{t\in T} a_t^{\sigma}.$$

Exercise 5.13. Let M be a set and $(EM,\vee,\wedge)$ be the EI algebra over M. Prove that $\{A \mid A \in 2^M\}$ is the set of all strong $\vee$—irreducible elements in EM (refer to Definition 2.10).

Exercise 5.14. Let M be a non-empty set and $(EM,\vee,\wedge)$ be the EI algebra over M. Let

$$SEM = \left\{ \sum_{i\in I} A_i \mid A_i \in 2^M - \{\varnothing\}, i \in I, I \textit{ is anyindexing set} \right\} \subseteq EM.$$

Prove that $(SEM,\vee,\wedge)$ is a sublattice of $(EM,\vee,\wedge)$ with minimum element M and maximum element $\sum_{m\in M}\{m\}$. Furthermore $(SEM,\vee,\wedge)$ is a molecular lattice.

Exercise 5.15. Let M be a non-empty set and $(EM,\vee,\wedge)$ be the EI algebra over M. $(SEM,\vee,\wedge)$ is not a Boolean algebra if $|M| > 2$. Nevertheless, if $|M| = 2$, i.e., $M = \{m_1,m_2\}$, prove that $(SEM,\vee,\wedge)$ is a Boolean algebra.

Exercise 5.16. Let X_1, X_2,..., X_n, M be $n+1$-non-empty sets and $(EX_1X_2...X_nM, \vee,\wedge)$ be the EI^{n+1} algebra over X_1, X_2,..., X_n, M. Let

$$SEX_1X_2...X_nM = \Big\{ \sum_{i\in I}(u_{1i}...u_{ni}A_i) \mid A_i \in 2^M - \{\varnothing\}, \\ i \in I, u_{ri} \in 2^{X_r}, r = 1,2,...,n, i \in I, I \textit{ is anyindexing set} \Big\} \subseteq EM.$$

Prove that $(SEX_1X_2...X_nM,\vee,\wedge)$ is a sublattice of $(EX_1X_2...X_nM,\vee,\wedge)$ with minimum element $\varnothing...\varnothing M$ and maximum element $\sum_{m\in M} X_1...X_n\{m\}$. Moreover $(SEX_1X_2...X_nM, \vee, \wedge)$ is a molecular lattice.

Exercise 5.17. Let X_1, ..., X_n, M be n+1 non-empty sets and $(EX_1...X_nM,\vee,\wedge)$ be the EI^{n+1} algebra over X_1, ..., X_n, M. For any $\alpha = \sum_{i\in I}(u_{1i}...u_{ni}A_i), \beta = \sum_{j\in J}(v_{1i}...v_{ni}B_j) \in EX_1...X_nM$, show the following assertions hold:

1. $\alpha\wedge\beta = \beta\wedge\alpha,\ \alpha\vee\beta = \beta\vee\alpha$; (Commutativity)
2. $(\alpha\wedge\beta)\wedge\gamma = \alpha\wedge(\beta\wedge\gamma),\ (\alpha\vee\beta)\vee\gamma = \alpha\vee(\beta\vee\gamma)$; (Associativity)
3. $(\alpha\wedge\beta)\vee\alpha = \alpha,\ (\alpha\vee\beta)\wedge\alpha = \alpha$; (Absorbance)
4. $\alpha\wedge(\beta\vee\gamma) = (\alpha\wedge\beta)\vee(\alpha\wedge\gamma),\ \alpha\vee(\beta\wedge\gamma) = (\alpha\vee\beta)\wedge(\alpha\vee\gamma)$; (Distributivity)
5. $\alpha\wedge\alpha = \alpha,\ \alpha\vee\alpha = \alpha$. (Idempotence)

Exercise 5.18. Let $X_1, \ldots, X_n$, M be n+1 non-empty sets and $(EX_1...X_nM,\vee,\wedge)$ be the EI^{n+1} algebra over X_1, ..., X_n, M. Prove that

$$\{u_1...u_nA \mid A \in 2^M, u_r \in 2^{X_r},\ r = 1,2,...,n\}$$

is the set of all $\vee$—irreducible elements in $EX_1...X_nM$.

Exercise 5.19. Let $X_1, \ldots, X_n$, M be n+1 non-empty sets and $(EX_1...X_nM,\vee,\wedge)$ be the EI^{n+1} algebra over X_1, ..., X_n, M. For $\sum_{i\in I}(u_{1i}...u_{ni}A_i) \in EX_1...X_nM$, let

$$\mathscr{P}_i = \{(v_1...v_nB) \mid B \supseteq A_i, v_r \subseteq u_{ri}, r = 1,2,\ldots,n\} \subseteq EX_1...X_nM,$$

$S_i \subseteq \mathscr{P}_i (i \in I)$. Prove that $\mathscr{B} = \bigcup_{i\in I} \mathscr{P}_i$ is a standard minimal family of $\sum_{i\in I}(u_{1i}...u_{ni}A_i)$ in molecular lattice $(EX_1...X_nM, \vee, \wedge)$.

Exercise 5.20. Let M be a non-empty set. If we define binary operations $\vee, \wedge$ as follows. For $\sum_{i\in I} A_i$,$\sum_{j\in J} B_j \in EM$,

$$\sum_{i\in I} A_i \vee \sum_{j\in J} B_j = \sum_{i\in I, j\in J} (A_i \cup B_j),$$

$$\sum_{i\in I} A_i \wedge \sum_{j\in J} B_j = \sum_{k\in I\sqcup J} C_k = \sum_{i\in I} A_i + \sum_{j\in J} B_j.$$

Prove that $(^*EM, \vee, \wedge)$ is a molecular lattice, in which for $\sum_{i\in I} A_i, \sum_{j\in J} B_j \in EM$, $\sum_{i\in I} A_i \leq \sum_{j\in J} B_j$ if and only if for any B_j $(j \in J)$ there exists A_k $(i \in I)$ such that $B_j \supseteq A_i$.

Exercise 5.21. Let X be a non-empty set. Let $(^*EX, \vee, \wedge)$ and $(^*E^\# X, \vee, \wedge)$ be the $^*EI$ algebra and $^*E^\# I$ algebra over X. Show the following assertions hold.

(1) The sublattices $(SE^\# X, \vee, \wedge)$ and $(SEX, \vee, \wedge)$ are isomorphism;
(2) The sublattices $(^*SE^\# X, \vee, \wedge)$ and $(^*SEX, \vee, \wedge)$ are isomorphism;
(3) The lattices $(^*EX, \vee, \wedge)$ and $(^*E^\# X, \vee, \wedge)$ are isomorphism.

Exercise 5.22. Let $(L_1, \geq), (L_2, \geq)$ be two lattices and f be an isomorphism from L_1 to L_2. Prove that for $x \in L_1$, x is a $\vee-$irreducible element in L_1 if and only if $f(x)$ is a $\vee-$irreducible element in L_2; and x is a $\wedge-$irreducible element in L_1 if and only if $f(x)$ is a $\wedge-$irreducible element in L_2.

Exercise 5.23. Let $X_1, ..., X_n, Y_1, ..., Y_m, M$ be $n+m+1$ non-empty sets. Prove that there is no isomorphism from $(EY_1...Y_mM, \geq)$ to $(E^\# X_1...X_n, \geq)$.

Exercise 5.24. Let $X_1, ..., X_n$ be n non-empty sets and $(E^\# X_1...X_n, \vee, \wedge)$ be the $E^\# I^n$ algebra over $X_1, ..., X_n$. Prove that the following assertions hold.

(1) If there exists $i_0 \in \{1,2,\cdots,n\}$ such that $|X_{i_0}| > 1$, then $(E^\# X_1...X_n, \vee, \wedge)$ is not a fuzzy lattice.
(2) If for all $i \in \{1,2,\cdots,n\}$ such that $|X_i| = 1$, then $E^\# X_1...X_n$ is a boolean algebra.

Exercise 5.25. Let X be a set and M be a set of simple concepts on X. Let (M, τ, X) be an AFS structure and $||.||$ be a fuzzy norm of an AFS algebra. For any fuzzy concept $\xi \in EM$, let $\xi(x)$ is the AFS algebra representation membership degree by any one of (5.10), (5.12),(5.13) and (5.14). Prove that $\{\mu_\xi(x) \mid \xi \in EM\}$ is the set of coherence membership functions of the AFS fuzzy logic system $(EM, \vee, \wedge, ')$ and the AFS structure (M, τ, X).

Open problems

Problem 5.1. Let M be a finite set and $k \geq 3$. How many elements in $C_k(EM)$?

Problem 5.2. Let $X_1, \ldots, X_n, M$ be n+1 non-empty sets and $|M| > 1, n > 1$. Whether $(SEX_1 \ldots X_n M, \vee, \wedge)$ is a fuzzy lattice?

Problem 5.3. Let $X_1, X_2, ..., X_n$ be n non-empty sets and $(E^\# X_1 ... X_n, \vee, \wedge)$ be the $E^\# I^n$ algebra over $X_1, ..., X_n$. What are the algebraic properties of $(SE^\# X_1 ... X_n, \vee, \wedge)$ corresponding to those of the lattice $(SEX_1 ... X_n, \vee, \wedge)$? i.e., the following problems.

(1) What are the set of all $\vee$–irreducible elements and the set of all $\wedge$–irreducible elements in the lattice $(SEX_1 ... X_n, \vee, \wedge)$?
(2) What are the set of all atom elements and the set of all dual atom elements in the lattice $(SEX_1 ... X_n, \vee, \wedge)$?
(3) Is the lattice $(SEX_1 ... X_n, \vee, \wedge)$ a fuzzy lattice?

Problem 5.4. How to explore other combinatoric properties of an AFS structure of data considering combinatoric techniques? And is there any interpretation of these combinatoric properties when applied to data analysis?

References

1. Deogun, J.S., Saquer, J.: Monotone concepts for formal concept analysis. Discrete Applied Mathematics 144, 70–78 (2004)
2. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer, Berlin (1999)
3. Graver, J.E., Watkins, M.E.: Combinatorics with Emphasis on the Theory of Graphs. Springer, New York (1977)
4. Halmos, P.R.: Measure theory. Springer, New York (1974)
5. Jacobson, N.: Basic Algebra I, 2nd edn. W. H. Freeman and Company, New York (1985)
6. Kim, K.H.: Boolean matrix Theory and Applications. Marcel Dekker, Inc (1982)
7. Liu, X.D.: The Fuzzy Theory Based on AFS Algebras and AFS Structure. Journal of Mathematical Analysis and Applications 217, 459–478 (1998)
8. Liu, X.D.: The Topology on AFS Algebra and AFS Structure. Journal of Mathematical Analysis and Applications 217, 479–489 (1998)
9. Liu, X.D.: The Structure of Fuzzy Matrices. International Journal of Fuzzy Mathematics 2, 311–325 (1994)
10. Liu, X.D.: A New Mathematical Axiomatic System of Fuzzy Sets and Systems. International Journal of Fuzzy Mathematics 3, 559–560 (1995)
11. Liu, X.D.: The Fuzzy Sets and Systems Based on AFS Structure, EI Algebra and EII algebra. Fuzzy Sets and Systems 95, 179–188 (1998)
12. Liu, X.D.: Two Algebra Structures of AFS structure. Journal of Fuzzy Mathematics 3, 561–562 (1995)
13. Liu, X.D., Chai, T.Y., Wang, W., Liu, W.Q.: Approaches to the Representations and Logic Operations for Fuzzy Concepts in the Framework of Axiomatic Fuzzy Set Theory I. Information Sciences 177, 1007–1026 (2007)
14. Liu, X.D., Wang, W., Chai, T.Y., Liu, W.Q.: Approaches to the Representations and Logic Operations for Fuzzy Concepts in the Framework of Axiomatic Fuzzy Set Theory II. Information Sciences 177, 1027–1045 (2007)
15. Liu, X.D.: A New Fuzzy Model of Pattern Recognition and Hitch Diagnoses of Complex Systems. Fuzzy Sets and Systems 104, 289–297 (1999)

16. Liu, X.D., Zhang, Q.L.: AFS Fuzzy Logic and Its Applications to Fuzzy Information Processing. Dongbei Daxue Xuebao 23(4), 321–323 (2002) (in Chinese)
17. Liu, X.D., Pedrycz, W., Zhang, Q.L.: Axiomatic Fuzzy Sets Logic. IEEE International Conference on Fuzzy Systems 1, 55–60 (2003)
18. Liu, X.D., Zhang, Q.L.: The EI Algebra Representations of Fuzzy Concept. Fuzzy Systems and Mathematics 16(2), 27–35 (2002)
19. Liu, X.D., Zhu, K.J., Huang, H.Z.: The Representations of Fuzzy Concepts Based on the Fuzzy Matrix Theory and the AFS Theory. In: Proc. IEEE Int. Symp. Intelligent Control, Houston, TX, October 2003, pp. 1006–1011 (2003)
20. Liu, X.D., Zhang, Q.L.: The Fuzzy Cognitive Maps Based on AFS Fuzzy Logic, Dynamics of Continuous. Discrete and Impulsive Systems 11(5-6), 787–796 (2004)
21. Liu, X.D., Zhang, L.S., Zou, J., Zhu, K.J., Zhang, Q.L.: The structures of EI algebras generated by information attributes. International Journal Intelligent Systems Technologies and Applications 3(3/4), 341–355 (2007)
22. Stumme, G., Wille, R. (eds.): Begriffliche Wissensverarbeitung—-Methoden und Anwendungen. Springer, Heidelberg (2000)
23. Wang, G.J.: Theory of topological molecular lattices. Fuzzy Sets and Systems 47, 351–376 (1992)
24. Zadeh, L.A.: Fuzzy Sets. Information and Control 8, 338–353 (1965)
25. Zhang, Y.J., Liang, D.Q., Tong, S.C.: On AFS Algebra Part I. Information Sciences 167, 263–286 (2004)
26. Zhang, Y.J., Liang, D.Q., Tong, S.C.: On AFS Algebra Part II. Information Sciences 167, 287–303 (2004)

Part III
Applications of AFS Theory

Chapter 6
AFS Fuzzy Rough Sets

In this chapter, in order to describe the linguistically represented concepts coming from data available in a certain information system, the concept of fuzzy rough sets are redefined and further studied in the setting of the Axiomatic Fuzzy Set (AFS) theory. These concepts will be referred to as *AFS fuzzy rough sets* [32]. Compared with the "conventional" fuzzy rough sets, the advantages of AFS fuzzy rough sets are twofold. They can be directly applied to data analysis present in any information system without resorting to the details concerning the choice of the implication ϕ, t-norm and a similarity relation S. Furthermore such rough approximations of fuzzy concepts come with a well-defined semantics and therefore offer a sound interpretation.

The underlying objective of this chapter is to demonstrate that the AFS rough sets constructed for fuzzy sets form their meaningful approximations which are endowed by the underlying semantics. At the same time, the AFS rough sets become directly reflective of the available data.

6.1 Rough Sets and Fuzzy Rough Sets

Rough set theory, proposed by Pawlak in 1982 [38, 39] can be viewed as a new mathematical approach to represent and process vagueness. The rough set philosophy dwells on the assumption that with every object of the universe of discourse we associate some information (data, knowledge). Objects characterized by the same information are indiscernible (similar) in view of the available information about them. The indiscernibility relation generated in this way constitutes a sound mathematical basis of the theory of rough sets [41, 42, 44]. As such, it has underwent a number of extensions and generalizations since the original inception in 1982. Based on the notion of a relation of being a (proper) part was proposed by Lesniewski [33], Polkowski and Skowron (1996 [43]) extended it to the system of approximate mereological calculus called rough mereology. Dubois and Prade (1990 [6]) introduced fuzzy rough sets as a generalization of rough sets. Radzikowska and Kerre (2002 [46]) proposed (ϕ, t)-fuzzy rough sets as a broad family of fuzzy rough sets, which are determined by some implication operator (implicator) ϕ, and a certain t-norm.

X. Liu and W. Pedrycz: Axiomatic Fuzzy Set Theory and Its Applications, STUDFUZZ 244, pp. 227–267.
springerlink.com

6.1.1 Rough Sets and Their Equivalent Definition

We start with some preliminaries of rough set theory which are relevant to this study. For details, the reader may refer to [20, 38, 40, 41, 42]. Pawlak [39] derived the rough probabilities by defining the *approximation space* $\mathscr{A} = (U, R)$ where U is a finite nonempty set called the universe and $R \subseteq U \times U$ is an equivalence relation on U, i.e., R is reflexive, symmetric, and transitive. R partitions the set U into disjoint subsets. Elements in the same equivalence class are said to be indistinguishable. Equivalence classes of R are called elementary sets. Every union of elementary sets is called a definable set. The empty set is considered to be a definable set, thus all the definable sets form a Boolean algebra. Given an arbitrary set $X \subseteq U$, one can characterize X by a pair of lower and upper approximations. The *lower approximation* $A_*(X)$ is the greatest definable set contained in X, and the *upper approximation* $A^*(X)$ is the least definable set containing X. They can be computed in the following manner.

$$
\begin{aligned}
A_*(X) &= \{x \mid [x]_R \subseteq X\}, \\
A^*(X) &= \{x \mid [x]_R \cap X \neq \varnothing\},
\end{aligned} \tag{6.1}
$$

where, $[x]_R$ denotes the equivalence class of the relation R containing x.

Proposition 6.1. *Let U be a set and $\mathscr{A}$=(U, R) be an approximation space. Then the lower approximation $A_*(X)$ and upper approximation $A^*(X)$ for any $X \subseteq U$ satisfy the following properties:*

(1) $A_*(X) \subseteq X \subseteq A^*(X)$;
(2) $A_*(\varnothing) = A^*(\varnothing) = \varnothing,\ A_*(U) = A^*(U) = U$;
(3) $A_*(X \cap Y) = A_*(X) \cap A_*(Y),\quad A^*(X \cup Y) = A^*(X) \cup A^*(Y)$;
(4) *If* $X \subseteq Y$, *then* $A_*(X) \subseteq A_*(Y),\ A^*(X) \subseteq A^*(Y)$;
(5) $A^*(X \cap Y) \subseteq A^*(X) \cap A^*(Y),\quad A_*(X \cup Y) \supseteq A_*(X) \cup A_*(Y)$;
(6) $A^*(X')' = A^*(X),\quad A_*(X')' = A_*(X)$;
(7) $A_*(A_*(X))$=A^* $(A_*(X))$=$A_*(X)$;
(8) $A^*(A^*(X))$=A_* $(A^*(X))$=$A^*(X)$.

An *information system* [45] is viewed as a pair $\mathbb{S} =< U,\ A >$, or a function $f : U \times A \to V$, where U is a nonempty finite set of objects called the universe, A is a nonempty finite set of attributes, and V stands for a value set such that $a : U \to V_a$ for every $a \in A$. The set V_a is called the *value set of the attribute a*. An information (decision) system may be represented as an attribute value (decision) table, in which rows are labeled by objects of the universe and columns by the attributes. Any subset B of A determines a binary relation R_B on U, i.e., $R_B \subseteq U \times U$, called an *indiscernibility relation* [44], defined by $(x,\ y) \in R_B$ if and only if $a(x) = a(y)$ for every $a \in B$. Obviously, R_B is an equivalence relation. The block of the partition of R_B, containing x will be denoted by $[x]_B = \{y \in X \mid (x,y) \in R_B\}$ for $x \in X$ and $B \subseteq A$. Thus in view of the data we are unable, in general, to observe individual objects but we are forced to reason only about the accessible granules of knowledge. Equivalence classes of the relation R_B (or blocks of the partition) are referred to as *B-elementary*

Table 6.1 Hiring Process: An Example of a Decision Table

	Diploma(i)	Experience(e)	French(f)	Reference(r)	Decision
x_1	MBA	Medium	yes	Excellent	Accept
x_2	MBA	Low	yes	Neutral	Reject
x_3	MCE	Low	yes	Good	Reject
x_4	MSc	High	yes	Neutral	Accept
x_5	MSc	Medium	yes	Neutral	Reject
x_6	MSc	High	yes	Excellent	Reject
x_7	MBA	High	No	Good	Accept
x_8	MCE	Low	No	Excellent	Reject

sets or *B-elementary granules*. In the rough set approach the B-elementary sets are the basic building blocks (concepts) of our knowledge about reality. The unions of B-elementary sets are called *B-definable sets*.

In many situations, the result of classification is provided and represented in the form of some decision variable. Information systems of this type are called decision systems. A decision system is any information system of the form $\mathbb{D} = (U, A \cup \{d\})$, where $d \notin A$ is the decision attribute. The elements of A are called conditional attributes. For a decision system $(U, A \cup \{d\})$ we can induce the AFS structure $(M_A \cup M_d,\ \tau,\ U)$ where M_A and M_d are the simple concepts associated with the attributes in A and d, respectively.

In Table 6.1, a data table describes a set of applicants. The individuals $x_1, x_2, \ldots, x_8$ shown there are characterized by some attributes, i.e., A, a set of attributes, like *Diploma*, *Experience*, *French*, *Reference*, etc. With every attribute $a \in A$, a set of its values is associated, i.e., V_a, such as the values of the attribute Experience, V_e=$\{$*Low*, *Medium*, *High* $\}$. In data analysis the basic problem we are interested in is to find patterns in data, i.e., to find a relationship between some sets of attributes, e.g., we might be interested whether an application is accepted depends on *Diploma* and *Experience*. A decision system is an information system of the form $A = (U, A \cup \{d\})$, where $d \notin A$ is the decision attribute, e.g., attribute: *Decision* in Table 6.1. The elements of A are called conditional attributes. In the approximation space $(U,\ R_B)$, $B \subseteq A$, we can apply B-elementary sets to approximate a set of objects with the expected values of decision attributes, thus we can know that under which condition is described by the granules of knowledge in $(U,\ R_A)$, the expected result is lead according to the data in an information system. For instance, we might be interested how to describe or interpret the conditions leading an application to *accept* according to the data in Table 6.1. For X=$\{x_1$,x_4,$x_7\}$, the set of accepted people, since both $A_*(X)$ and $A^*(X)$ are the unions of the sets $[x]_B$ for $x \in U$ and every set $[x]_B$ has a definite interpretation with the condition attributes, e.g., $[x]_{\{i,e\}}$ is the set of objects having the same values of attributes *Diploma*, *Experience* as x, hence $A_*(X)$ and $A^*(X)$ can be applied to build the relation between the decision attributes and the condition attributes. It is clear that $[x]_B = \cap_{a \in B}[x]_{\{a\}}$ for any $B \subseteq A$. This implies that the lower approximation and upper approximation of any set $X \subseteq U$ are some unions or intersections of the sets in $\{[x]_{\{a\}} | a \in B, x \in U\}$, i.e., they are members of the Boolean algebra generated by the family of sets $\{[x]_{\{a\}} | a \in B, x \in U\}$.

Following the above discussion, we observe that the rough sets defined by an equivalence relation (6.1) have raised some difficulties when being directly applied to the above information systems. Thus, we give the following equivalent definition of rough sets which relies on a family of subsets of U. This definition is also helpful to expand the concept of rough sets to fuzzy rough sets.

Definition 6.1. Let U be a set and Λ be the set of some Boolean subsets of U. The *upper approximation* (denoted by $S^*(X)$) and *lower approximation* (denoted by $S_*(X)$) of $X \subseteq U$ in regard to Λ is defined by:

$$S^*(X) = \bigcup_{x \in X} \delta_x, \quad S_*(X) = \bigcup_{x \in U, \delta_x \subseteq X} \delta_x \tag{6.2}$$

where $x \in U$, $\delta_x \in \Lambda^-$, δ_x is the smallest set containing x, Λ^- be the set of all sets generated by the sets in Λ, using set intersection $\cap$ and complementation $'$.

Because δ_x is the smallest subset in Λ^-, δ_x is a description of x using Boolean concepts in Λ^- such that x can be distinguished among other elements in X at the maximum extent. If for any $x \in U$, $\psi(x)= \delta_x$, then ψ is a mapping from U into Λ^- and ψ determines a classification of U (i.e., x, $y \in U$, x, y in the same class if and only if $\psi(x) = \psi(y)$ or $\delta_x=\delta_y$). The equivalence relation which corresponds to the classification induced by Λ is denoted as R_Λ and $[x]_\Lambda=\{y \mid (x, y) \in R_\Lambda \}$. We can prove that for any $x \in X$, $[x]_\Lambda=\delta_x$ as follows. For any $y \in[x]_\Lambda$, since $\delta_x=\delta_y$, hence $y \in\delta_x$. This implies that $[x]_\Lambda \subseteq\delta_x$. Assume that there exists $z \in\delta_x$ such that $z \notin[x]_\Lambda$, i.e., $\delta_x \neq\delta_z$. Since δ_z is the smallest set containing z, hence for $z \in\delta_x\cap\delta_z$, we have $\delta_x\cap\delta_z=\delta_z$, i.e., $\delta_x \supseteq\delta_z$ and $x \notin\delta_z$. This implies that $x \in\delta_x\cap\delta_z' \in\Lambda^-$, contradicting that δ_x is the smallest set containing x. Thus, for any $x \in X$, $[x]_\Lambda=\delta_x$. Therefore the rough sets defined by (6.1) in the approximation space $\mathscr{A} = (U,R_\Lambda)$ are equivalent to that defined by (6.2).

In practice, the attributes or features of many information systems may be described by real numbers, Boolean variables, ordered relations or fuzzy (linguistic) labels. Thus in this chapter we expand Definition 6.1 to fuzzy sets, i.e., Λ is a set of fuzzy linguistic terms in the framework of AFS theory.

6.1.2 Fuzzy Rough Sets

Since the initial concept of rough sets theory, the extension of this fundamental idea to a fuzzy environment has been a topic of study. In real world applications, some attributes are often measured in a continuous domain and their values are described according to a partition of this domain, which is discretized making use of intervals, linguistic terms or ordered relations. The application of rough sets to data analysis will depend greatly on the values taken by the limits that define these intervals. Smoothing these limits through membership functions of fuzzy sets could be a viable alternative to improve the system's robustness to small variations in the collected data.

In practice, the fuzzy partition can be constructed using any clustering technique. Constructing fuzzy similarity relations for the most general type of fuzzy data described by [8] proceeds as follows: In a dataset of N samples, sample x_i ($i \in [1,N]$) is described by N_F features, while the value of each feature j is expressed by a set of N_{LL} grades of membership, μ_i^{jk} ($k = 1,...,N_{LL}$), to N_{LL} linguistic labels. Thus, sample x_i can be characterized with the aid of the following values:

$$x_i = [(\mu_i^{11},...,\mu_i^{1N_{LL}}),...,(\mu_i^{j1},...,\mu_i^{jN_{LL}}),...,(\mu_i^{N_F 1},...,\mu_i^{N_F N_{LL}})], \qquad (6.3)$$

where $i = 1,...,N; j = 1,...,N_F$.

All of "conventional" fuzzy rough sets are based on the same concept: the substitution, in Pawlak's original set approximation definitions of the Boolean equivalence relation R by a fuzzy relation S, on which several conditions are imposed. The most general approach, as provided by Greco et al. [17] only requires that S be a reflexive fuzzy relation ($S(x,x) = 1$). Further restrictions are imposed by Dubois and Prade [5, 6], who demand that S be a *T-similarity relation*, i.e., a *fuzzy relation*, $S : U \times U \to [0,1]$ for which the following conditions should hold:

(1) For any $x \in U$, $S(x,x) = 1$; (Reflexivity)
(2) For any $x,y \in U$, $S(x,y) = S(y,x)$; (Symmetry)
(3) For any $x,y,z \in U$, $S(x,z) \geq T(S(x,y),S(y,z))$. (T-transitivity)

Here, T is a t-norm, that is, a commutative, monotonic and associative aggregation operator, $T(x,y) : [0,1] \times [0,1] \to [0,1]$, that satisfies the boundary condition $T(a,1) = a$. If T is the minimum operator, then this definition coincides with Zadeh's original expression for similarity relations [53].

Instead of directly substituting R with a fuzzy relation, its set of equivalence classes (quotient set), U/R, can also be replaced by a family ϕ of fuzzy sets $F_1, F_2, ..., F_n$, for which it is usually required that they form a *weak fuzzy partition*, that is, the following conditions should be satisfied: for any $x \in U$,

(1) $\inf_{x \in U} \max_{1 \leq i \leq n} \mu_{F_i}(x) > 0$,
(2) For any i, j, $\sup_{x \in U} \min(\mu_{F_i}(x), \mu_{F_j}(x)) < 1$,

where $\mu_{F_i}(x)$ is the membership function of fuzzy sets F_i, $i = 1,2,...,n$. The first requirement ensures that ϕ covers all elements of U, while the second one imposes a disjointness condition to be satisfied between the elements of F_i.

Using either a fuzzy similarity relation, S, or a weak fuzzy partition, ϕ, the following three approaches have been proposed:

(1) Approach based on possibility theory (Dubois, Prade and Farinas del Cerro). This is probably the most cited approach to Fuzzy Rough Sets, which was introduced in the seminal papers [5, 6, 9]. According to this proposal, if F represents a fuzzy set with membership function μ_F and S is a fuzzy similarity relation (which here is assumed to be reflexive, symmetric and T-transitive) with membership degree $\mu_S(x,y)$, then the *upper approximation* and *lower approximation* of F in regard to S can be calculated as the degrees of necessity and possibility

of F (in the sense of Zadeh [4, 54]) taking as referential the equivalence classes of S. These approximations are

$$\mu_{S^*(F)}(x) = \sup_{y\in U}\min(\mu_F(y), \mu_S(x,y)), \tag{6.4}$$

$$\mu_{S_*(F)}(x) = \inf_{y\in U}\max(\mu_F(y), 1-\mu_S(x,y)). \tag{6.5}$$

If S is Boolean, the reflexivity, symmetry and T-transitivity requirements would make it a Boolean equivalence relation (Note that this does not depend on T). These equations would then be reduced to Pawlak's original definitions.

Greco et al. [17] presented a fuzzy extension of Slowinski and Vanderpooten's proposal [47] in which only Boolean reflexive relations were used (as opposed to Pawlak's equivalence relations). According to their proposal, if $T(x,y)$ represents a t-norm, $C(x,y)$ its associated t-conorm, $N(x)$ is a negation operator, $S(x,y)$ a fuzzy reflexive relation (which does not have to be symmetric or transitive), and F a fuzzy set with membership function $\mu_F(x,y)$, then the *upper approximation* and *lower approximation* of S can be defined as

$$\mu_{S^*(F)}(x) = C_{y\in U}T(\mu_F(y), \mu_S(x,y)), \tag{6.6}$$

$$\mu_{S_*(F)}(x) = T_{y\in U}C(\mu_F(y), N(\mu_S(x,y))). \tag{6.7}$$

We can see that these formulas become Dubois and Prade's expressions (Eqs. (6.4) and (6.5)) when T and C represent the standard intersection and union operators (minimum and maximum) and S is also symmetric and transitive for a certain t-norm, N, which does not necessarily have to coincide with T. Using logic transformations, these equations can also be expressed in terms of t-norms and related implication operators. Based on this, Radzikowska and Kerre have carried out an exhaustive formal study on the theoretical properties of these fuzzy rough sets [46].

Ziarko's Variable Precision Rough Set model [56], can also be introduced in Dubois and Prade's Fuzzy Rough Set framework. To do this, Eqs. (6.4) and (6.5) should be rewritten in the following form: for any $x \in U$,

$$\mu_{S^*(F)}(x) = \max(\mu_F(x), I_{S^*(F)}(x)), \tag{6.8}$$

$$\mu_{S_*(F)}(x) = \min(\mu_F(x), I_{S_*(F)}(x)), \tag{6.9}$$

where

$$I_{S^*(F)}(x) = \max_{y\in U(y\neq x)} \min(\mu_F(y), \mu_S(x,y)),$$

$$I_{S_*(F)}(x) = \min_{y\in U(y\neq x)} \max(\mu_F(y), 1-\mu_S(x,y))$$

$I_{S_*(F)}(x)$ is an index that expresses the degree of inclusion of all similar objects to x in the fuzzy set F. In the same fashion, $I_{S^*(F)}(x)$ expresses the degree of inclusion of at least one similar object to x in F.

It may be noted that the mere presence of only one sample that is very similar to x but has low degree of membership in F will force $I_{S_*(F)}(x)$ to be low and x will be considered to be excluded from $S_*(F)$. In the same way, only one sample that may be very similar to x but has a high degree of membership in F will cause a high value of $I_{S^*(F)}(x)$. If the cardinality of the dataset is high, this single sample may be the result of noise or an error in classification. In this case, the values of membership calculated for $I_{S_*(F)}(x)$ and $I_{S_*(F)}(x)$ may not be adequate for further decision making.

In order to deal with this limitation, Salido and Murakami in [8] have developed the concept β-precision aggregation operators. These aggregators allow for some tolerance to distorting values in the aggregation operands when the cardinality of the aggregated values is high. The properties associated to this concept were characterized in [10]. Its application to t-norms and t-conorms results in what are called β-precision quasi-t-norms and B-precision quasi-t-conorms, whose formal definition can be found in [10]. Therefore, Salido and Murakami proposed to extend Ziarko's Variable Precision Rough Set Model to the Dubois and Prade Fuzzy Rough Set framework by extending the maximum and minimum operators used to calculate the inclusion indexes of Eqs. (6.5) and (6.4) to their β-precision counterparts, max_β and min_β as follows:

$$I_{S^*(F)_\beta}(x) = max_{\beta_{y\in U(y\neq x)}} \min(\mu_F(y), \mu_S(x,y))$$
$$I_{S_*(F)_\beta}(x) = min_{\beta_{y\in U(y\neq x)}} \max(\mu_F(y), 1-\mu_S(x,y))$$

Dubois and Prade's lower and upper approximation equations can then be expressed in the β-precision context as follows: for any $x \in U$,

$$\mu_{S^*(F)_\beta}(x) = \max(\mu_F(x), I_{S^*(F)_\beta}(x)), \tag{6.10}$$
$$\mu_{S_*(F)_\beta}(x) = \min(\mu_F(x), I_{S_*(F)_\beta}(x)). \tag{6.11}$$

In the practical implementation of these formulas, adequate values for β could be around 0.98 or 0.99, which allow for a 1-2 % of noisy operands in the aggregation process. However, the optimal value of β will depend on the problem's domain and the accuracy of the description of the attributes. The maximum value to which β can be set (which should determine the generalization capability of this approach) will also depend on these circumstances.

(2) Approach based on fuzzy inclusions (Kuncheva and Bodjanova). Kuncheva [21] and Bodjanova [2] give new definitions of Fuzzy Rough Sets. Both approaches deal with the approximation of a fuzzy set in terms of a weak fuzzy partition, for which they use different measures of fuzzy set inclusion, many of which have been studied in [3]. As to the generation of the weak fuzzy partition from a set of fuzzy data, Kuncheva does not impose any restrictions on how this is to be done: she implies that this can be resolved through fuzzy clustering or any other technique like, for example, the generation of fuzzy equivalence classes from a fuzzy similarity relation. Bodjanova, on the other hand, generates fuzzy partitions using unions and intersections of the fuzzy features measured in the

data. This way of resolving a rough set analysis of fuzzy data through degrees of inclusions of fuzzy sets can also be considered to be an extension of Ziarko's Variable Precision Rough Set Model [56].

(3) Approach based on α-levels of fuzzy sets (Yao and Nakamura). First Nakamura [35, 36], and later on Yao [52] proposed a rough set analysis of fuzzy data through the application of Boolean rough set theory to the α-levels of a fuzzy similarity relation obtained from this data. The computational complexity of this approach increases with the level of resolution with which the α-levels of the fuzzy similarity relation are formed.

In practice, the implementation of all three approaches requires a prior determination of a fuzzy similarity relation or a partition of fuzzy similarity classes, and the choice of a certain implication operator ϕ operator and the t-norm. Those components are determined on a basis of some available experimental data. The fuzzy partition can be obtained using any clustering technique applied to these data. The construction of fuzzy similarity relations for some general type of fuzzy data has been described in [8]. The fuzzy similarity relation matrix S=(s_{ij}), s_{ij}=$\mu_S(x_i, x_j)$, formed on the data set X=$\{x_1$,x_2,..., $x_n\}$, can be derived or constructed by means of various optimization mechanisms such as, e.g., Fuzzy C-Means (FCM), k-NN fuzzy clustering algorithms or some aggregative algorithms such as the one presented by Salido and Murakami [8]. The similarity relation S obtained in this manner might not exhibit any semantics with well-defined linguistic labels formed for each feature (attribute) shown in (6.3). Thus the upper approximation $S^*(F)$ and lower approximation $S_*(F)$ of a fuzzy set F based on S are just the numerical membership functions, and the semantic relationships between the linguistic labels and F, which is acceptable by being comprehended by humans, might not be clearly expressed.

As discussed in the above section, let $\mathbb{D} = (U, A \cup \{d\})$ be a decision system, where $d \notin A$ is the decision attribute. The elements of A are conditional attributes. For a decision system (U, A∪$\{d\}$) we can induce the AFS structure ($M_A \cup M_d$, τ, U) where M_A and M_d are the simple concepts associated with the attributes in A and d, respectively. The AFS fuzzy rough sets support the determination of fuzzy sets in EM_A that are used to approximate a given fuzzy set γ∈EM_d by the AFS algebra $E(M_A \cup M_d)$. It is worth noting that in comparison with the forenamed conventional fuzzy rough sets, the AFS fuzzy rough sets can be directly applied to process data in the information systems without explicitly using the implicator ϕ, a t-norm and a similarity relation S. The upper and lower approximation $S^*(F)$ and $S_*(F)$ of a fuzzy set F in EM_d are fuzzy sets in EM_A which have well-defined semantics with the simple concepts on the conditional attributes. AFS fuzzy rough sets approximate a given fuzzy concept on the decision attributes using the fuzzy concepts on the condition attributes. Thus, adhering to existing data in an information system, AFS fuzzy rough sets can offer semantically meaningful interpretation for the conditions under which some expected result may lead to. They are essential in knowledge engineering, decision-making and intelligent systems, in general as pointed out in the context of computing with words [55].

6.2 Fuzzy Rough Sets under Framework of AFS Theory

In this section, we introduce AFS fuzzy rough sets by expanding Definition 6.1 to fuzzy sets, i.e., we apply a set of simple concepts associating to the condition attributes to approximate a fuzzy set, which associates to the decision attribute and is given in advance, to describe the decision result in a data under the framework of AFS theory.

6.2.1 AFS Structure of Information Systems

Many information systems involve a mixture of quantitative and qualitative feature variables like those shown in Table 6.1 and Table 4.1. Besides quantitative features, qualitative features, which could be further divided into nominal and ordinal features, are also commonly seen. The information systems described by information-based criteria such as human perception-based information, gain ratio, symmetric uncertainty, order are regularized to be the AFS structures by the two axioms in Definition 4.5. Thus the fuzzy concepts, membership functions and fuzzy logic on the raw data can be explored by the AFS theory using the AFS structure of the data.

Let $\mathbb{D} = (U, A \cup \{d\})$ be a decision system shown as Table 6.1 and M be a set of some fuzzy or Boolean concepts on U. Every $m \in M$ associates to an attribute $a \in A \cup \{d\}$ and by the values $a(x)$, $a(y) \in V_a$, one can compare the degrees of x, $y \in U$ belonging to m. For example, let m be the fuzzy concept "*Low Experience*" which associates to attribute $e \in A$, i.e., e: "*Experience*". By V_e={*Low*, *Medium*, *High* } and the attribute value $e(x)$ shown in Table 6.1, we can construct the binary relation R_m defined by Definition 4.2 as follows:

$(x, x) \in R_m$, if $e(x)$=*Low* or *Medium*;
$(x, y) \in R_m$, for any $y \in U$, if $e(x)$=*Low*;
$(x, y) \in R_m$, if $e(x)$=*Medium*, and $e(y)$=*Medium* or *High*.

For Boolean concept m: "*French-yes*" which associates to attribute $f \in A$, i.e., f: "*French*", R_m is constructed as follows: $(x, y) \in R_m$ for any $y \in U$, if $f(x)$=*yes*. Similarly, we can construct R_m for each concept $m \in M = \{m_1, \ldots, m_{13}\}$ according to the information system $f: U \times A \to V$, where m_1: MBA_i, m_2: MCE_i, m_3: MSc_i, m_4: Low_e, m_5: Medium_e, m_6: High_e, m_7: yes_f, m_8: No_f, m_9: Excellent_r, m_{10}: Neutral_r, m_{11}: Good_r, m_{12}: Accept_d, m_{13}: Reject_d. By Definition 4.3, one can verify that each concept in M is a simple concept. Thus the information system can be represented by the AFS structure (M, τ, U) using (4.26). Let M_A and M_d be the sets of simple concepts associating to the condition attributes in A and the decision attribute d, respectively. We apply fuzzy concepts in EM_A to approximate a given fuzzy set $\gamma \in EM_d$, in order to know that under what condition is described by the fuzzy concepts in EM_A, the expected result γ is lead according to the data in an information system.

6.2.2 Representations of Fuzzy Concepts in the Information Systems

In the decision system $\mathbb{D} = (U, A \cup \{d\})$, where $d \notin A$ is the decision attribute, we always need to learn the conditions which can lead to a decision (result) which may be represented by a fuzzy or Boolean set with given membership function μ : $X \rightarrow$[0,1]. For instance, in Table 6.1, the decision is "*accept*". In order to study the $E^{\#}I$ algebra represented fuzzy concept approximations of a given fuzzy or Boolean set representing a decision result, we transfer a fuzzy set with given membership degrees in the interval [0,1] into a fuzzy set with membership degrees in the $E^{\#}I$ algebra as follows. Let X be a set and $\mathscr{F}(X) = \{\eta | \mu_\eta : X \rightarrow [0,1]\}$. For $\theta \in \mathscr{F}(X)$, the *$E^{\#}I$ algebra represented membership function of the fuzzy set* θ is defined as follows: for any $x \in X$,

$$\theta(x) = \underline{\theta}(x) \in E^{\#}X, \tag{6.12}$$

where

$$\underline{\theta}(x) = \begin{cases} \{y \in X \mid \mu_\theta(y) \leq \mu_\theta(x)\}, & \mu_\theta(x) \neq 0 \\ \varnothing, & \mu_\theta(x) = 0 \end{cases}$$

Thus for any fuzzy set $\theta \in \mathscr{F}(X)$, the lower and upper approximations of θ under the meaning of the $E^{\#}I$ algebra represented fuzzy sets defined by (5.13) can be studied via the lattices EM and $E^{\#}X$.

Let $S \subseteq 2^X$ be an σ-algebra over X and m_ρ be a measure on S with $0 \leq m_\rho(A) \leq 1$ for any $A \in S$,

$$m_\rho(A) = \frac{\sum_{x \in A} \rho(x)}{\sum_{x \in X} \rho(x)}, \tag{6.13}$$

which is defined as Definition 5.6 for the map $\rho : X \rightarrow R^+ = [0, \infty)$. From Proposition 5.7, the fuzzy norm of the lattice $E^{\#}X$ can be constructed by measure m_ρ as follows. For $\sum_{i \in I} a_i \in E^{\#}X$,

$$||\sum_{i \in I} a_i|| = \sup_{i \in I}\{m_\rho(a_i)\}$$

and for any fuzzy concept $\zeta = \sum_{i \in I}(\prod_{m \in A_i} m) \in EM$, its membership function is defined as follows: for any $x \in X$,

$$\mu_\zeta(x) = ||\sum_{i \in I} A_i^\tau(x)|| = \sup_{i \in I}\{m_\rho(A_i^\tau(x))\}. \tag{6.14}$$

From Theorem 5.6, one knows that the membership function of fuzzy set θ defined by the formula (6.14) is the coherence membership function of the AFS fuzzy logic $(EM, \vee, \wedge, ')$ and the AFS structure (M, τ, X).

We can verify that the above m_ρ for the function ρ is a measure over X by Definition 1.40. $\rho(x)$ may have various interpretations depending on the specificity of the problem at hand. For instances one can allude to what we have discussed in

Chapter 4, 5. In general, $\rho(x)$ weights how essential is the relationship of the sample x to the category of concepts under consideration. Later on we show how to derive the weight function $\rho(x)$ from a given membership function of a fuzzy set in order to study the lower and upper approximations of this fuzzy set.

Given (4.27), we know that in (6.14), $A_i^\tau(x) \subseteq X$ is the set of all elements in X whose degrees belonging to $\prod_{m\in A_i} m$ are less than or equal to that of x. $A_i^\tau(x) = \cap_{m\in Ai} \{m\}^\tau(x)$ is determined by the semantic meanings of the simple concepts in A_i and the distribution of the original data. In general, the larger the set $A_i^\tau(x)$, the higher the degree x belonging to $\prod_{m\in A_i} m$ if all elements in X have the equally essential relation to the considering group of concepts, i.e., for any $x, y \in X$, $\rho(x)$=$\rho(y)$. For $x \in X$, if x has so limited relationship with the considering group of concepts that whether or not they are included in $A_i^\tau(x)$ has not significant influence on the evaluating the degree of x belongingness (membership) to $\prod_{m\in A_i} m$. Then $\rho(x)$ should be very small or practically be equal to 0. Since $A_i^\tau(x)$ and $A_i^\tau(y)$ are independent on $\rho(x)$ and $\rho(y)$, hence $\rho(x) > \rho(y)$ does not mean that $A_i^\tau(x) \supseteq A_i^\tau(y)$ or $\mathrm{m}(A_i^\tau(x)) > \mathrm{m}(A_i^\tau(y))$. In other words, the more essential a sample is does not entail that the degree of the membership to the fuzzy concept in EM is always higher. In other words, $\rho(.)$ weights the referring value of each sample in X for the determining of the membership functions of fuzzy concepts in EM.

In the decision system $\mathbb{D}$=$(U, A\cup\{d\})$, for a fuzzy concept ν with the membership function $\mu_\nu(x)$: $X \to$[0,1], which is given in advance to describe the decision results in a data, we find the interpretations or descriptions with the fuzzy concepts in EM_A for ν, where M_A is the set of simple concepts on the condition attributes. We can explore the lower and upper approximations of ν under the membership degrees defined by (5.13) and (6.12) via the AFS algebras EM and $E^\#X$. Furthermore, if we can construct a function $\rho : X \to [0,\infty)$ for the measure m_ρ such that the membership function of ν defined by (6.14) is equal the given $\mu_\nu(x)$, then the lower and upper approximations of ν under the membership degrees defined by (6.14) can be explored. In what follows, we find ρ_ν which is induced by $\mu_\nu(.)$ satisfying $\mu_\nu(x)$=$\mathrm{m}_\nu(\underline{\nu}(x))$, for any $x \in X$. Let X be a finite set and $\mu_\nu(X)$=$\{y$=$\mu_\nu(x) \mid x \in X\}$. Let

$$\mu_\nu(X) = \{y_1, y_2, \ldots, y_n\}$$

and $y_i < y_j$ for any $i < j$. For any $u \in X$, $\rho_\nu(u)$ can be obtained by solving the following equation:

$$\sum_{u\in\nu(x)} \rho_\nu(u) = \mu_\nu(x), \ \ x \in X,$$

Following this expression we derive

$$y_k - y_{k-1} = \sum_{u\in\mu_\nu^{-1}(y_k)} \rho_\nu(u), \ \ k = 2,3,\ldots,n,$$

where $\mu_\nu^{-1}(y) = \{x \in X \mid \mu_\nu(x) = y\}$. Considering for any $u \in \mu_\nu^{-1}(y_k)$, $\mu_\nu(u) = y_k$, let the weights of all $u \in \mu_\nu^{-1}(y_k)$ be equal. Therefore we have

$$\rho(x) = \begin{cases} \frac{y_1}{|\mu_v^{-1}(y_1)|}, & \mu_v(x) = y_1 \\ \frac{y_k - y_{k-1}}{|\mu_v^{-1}(y_k)|}, & \mu_v(x) = y_k, 2 \leq k \leq n \end{cases} \quad (6.15)$$

and $\sum_{u \in X} \rho_v(u) = y_n$. Thus the membership function $\mu_v(x)$ is represented by the measure m_v as follows: for any $x \in X$

$$\mu_v(x) = y_n m_v(\underline{v}(x)), \quad (6.16)$$

where for any $A \in 2^X$, $A \neq \varnothing$,

$$\mathrm{m}_v(A) = \frac{\sum_{u \in A} \rho_v(u)}{\sum_{u \in X} \rho_v(u)} = \frac{\sum_{u \in A} \rho_v(u)}{y_n}$$

is expressed by (6.13) for function ρ_v and

$$y_n = \max\{\mu_v(x) \mid x \in X\}. \quad (6.17)$$

Example 6.1. Let us consider the *Japanese Credit Screening* data [34]. Here the number of instances is equal to 125 where each of them is described by 17 features. The first 12 features are Boolean while the remaining ones assume real values. Let U={x_1, x_2, ..., x_{125}} be the set of the 125 instances, from the samples x_1 to x_{85} are examples of positive credit (positive credit decision) and x_{86} to x_{125} are negative credit samples. Let M={m_1, m_2, ..., m_{44}} be the concepts formulated on the attributes of the information system $< U,A >$.

On each of the first 12 features, two concepts are chosen and the semantic meanings of m_1 to m_{24} are shown as follows:

m_1: *positive credit*, m_2: *negative credit*; m_3: *jobless*, m_4: *no jobless*; m_5: *purchase pc*, m_6: *no purchase pc*; m_7: *purchase car*, m_8: *no car purchase* ; m_9: *stereo purchase*, m_{10}: *no stereo purchase* ; m_{11}: *purchase jewelery*, m_{12}: *no purchase jewelery*; m_{13}: *purchase medinstru*, m_{14}: *no purchase medinstru*; m_{15}: *purchase bike*, m_{16}: *no purchase bike*; m_{17}: *purchase furniture*, m_{18}: *no purchase furniture*; m_{19}: *male*, m_{20}: *female*; m_{21}: *unmarried*, m_{22}: *married*; m_{23}: *located in problematic region*, m_{24}: *located in non- problematic region*.

For each of the features (from 13^{th} to 17^{th}), in order to describe the concept "*positive credit*", we choose four fuzzy concepts with the following semantic meanings to express linguistic labels "*large*", "*not large*", "*middle*", "*not middle*".

For the 13th feature (age) we choose m_{25}: *old*, m_{26}: *not old*, m_{35}: *average age*, m_{36}: *not an average age*. On the 14th feature which is the amount of money in the bank we choose m_{27}: *more money on deposit in the bank*, m_{28}: *not more money in the bank*, m_{37}: *the amount of money in the bank about average*, m_{38}: *the amount of money in the bank not about average*. On the 15th feature which is "monthly loan payment amount" we choose m_{29}: *loan payment amount large*, m_{30}: *loan payment amount not large*, m_{39}: *loan payment about average*, m_{40}: *loan payment not about average*. For the 16th feature viz. "the number of months expected to pay off the loan" we choose m_{31}: *expected to pay off loan more*, m_{32}: *expected to pay off loan not more*, m_{41}: *expected to pay off loan about average*, m_{42}: *expected to pay off* *loan not about average*. For the feature 17th the number of years working at current company we choose m_{33}: *the number of years working more*, m_{34}: *the number of years working not more*, m_{43}: *the number of years working about average*, m_{44}: *the number of years working not about average*.

By Definition 4.2, we can obtain R_m, the binary relation for every $m \in M$ by comparing the degree of each pair of persons belonging to m according to its underlying semantics. To show that, as an example consider m_{37}: *"the amount of money on deposit in bank about average*". The average deposit in bank of all samples in U is 69.46. Let deposit$_x$ denote the amount of money deposited in bank of the given sample $x \in U$. Then $(x, y) \in R_m$ if and only if $|69.46 - deposit_x| \leq |69.46 - deposit_y|$. Making use of Definition 4.3 we can verify that each $m \in M$ is a simple concept. (M, τ, X) is an AFS structure if τ is defined by (4.26) as follows: for any $x_i, x_j \in U$,

$$\tau(x_i, x_j) = \{m \mid m \in M, (x_i, x_j) \in R_m\}. \tag{6.18}$$

Let S=2^U be the σ-algebra over U. Let ν be a given fuzzy set on U, for $x \in X$, $\mu_\nu(x)$=1, if sample x comes with a positive credit decision, otherwise $\mu_\nu(x)$=0.15 to stress that ν be a fuzzy concept here. Any small number can be associated with the negative samples. By (6.15), we have $\rho_\nu(x)$=0.85/85=0.01 if sample x comes with a positive credit decision, otherwise $\rho_\nu(x)$=0.15/40=0.0037. The positive credit samples have higher weight than the negative ones and our intent is to stress the importance of positive samples.

Thus we are able to construct ρ_ν for the measure m_{ρ_ν} defined by (6.13) and then apply (6.14) to determine the membership function for any fuzzy concept in *EM*. Since (*EM*, $\wedge$, $\vee$, $'$) is an algebra system, i. e., *EM* is closed under the fuzzy logic operations $\wedge$, $\vee$, $'$, hence for any fuzzy concepts α, $\beta \in EM$, the membership functions of their fuzzy logic operations, $\mu_{\alpha\wedge\beta}(x)$, $\mu_{\alpha\vee\beta}(x)$ and $\mu_{\alpha'}(x)$ are also well defined by (6.14). Figures 6.1 – 6.2 show the membership functions of fuzzy concepts $\alpha, \gamma', \alpha \wedge \gamma$, respectively, where α=$m_1m_3 + m_{16}m_{35}m_{42} + m_9m_{11}m_{43}$, γ=$m_{42}m_{43} + m_{16}m_{35}m_1 \in EM$. γ' is the negation of the fuzzy concept γ (refer to (4.19)). By inspecting the plots in Figure 6.1, one cannote that for most samples x, $\mu_{\alpha\wedge\beta}(x) = \min\{\mu_\alpha(x), \mu_\beta(x)\}$ however for few samples we have $\mu_{\alpha\wedge\beta}(x) < \min\{\mu_\alpha(x), \mu_\beta(x)\}$. By Theorem 5.6, we know that the membership functions defined by (6.14) are coherence membership functions, hence for any $x \in X$, $\mu_{\alpha\wedge\beta}(x) \leq \min\{\mu_\alpha(x), \mu_\beta(x)\}$ because of Proposition 4.2. This implies that in the AFS fuzzy logic $\mu_\alpha(x)$ and $\mu_\beta(x)$ are not sufficient to determine $\mu_{\alpha\wedge\beta}(x)$, which is the membership degree of x belongingness (membership) to the conjunction of the two fuzzy concepts α, β, and $\mu_{\alpha\wedge\beta}(x)$ is determined by the distributions of the original data and the semantics of fuzzy concepts themselves. This stands in contrast with the existing fuzzy logic systems equipped by some t-norm, in which $\mu_{\alpha\wedge\beta}(x) = T(\mu_\alpha(x), \mu_\beta(x))$ is fully determined by the membership degrees $\mu_\alpha(x)$ and $\mu_\beta(x)$ and is independent from the distribution of the original data.

Hence, the constructed membership functions and the logic operations in the AFS theory include more information about the distributions of the original data and

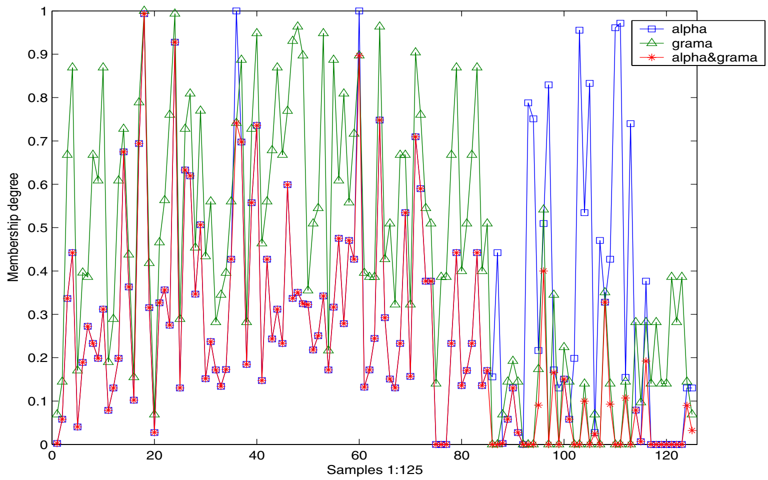


Fig. 6.1 Membership functions of fuzzy concept α, γ, and $\alpha \wedge \gamma$

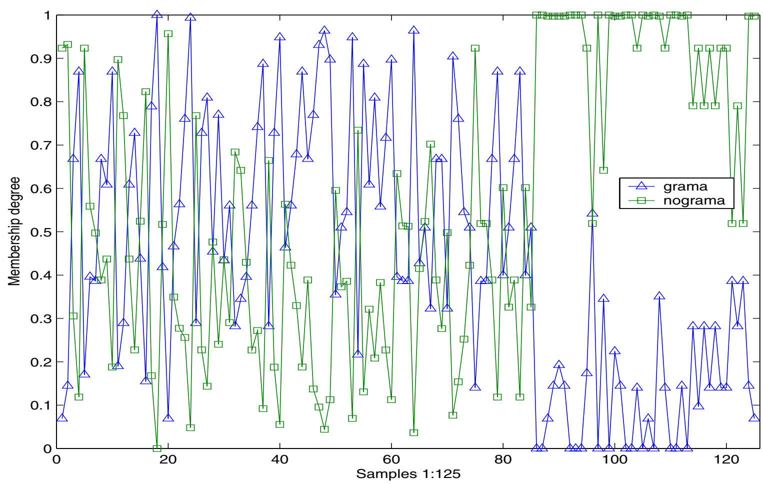


Fig. 6.2 Membership functions of γ, γ'

the underlying interpretations., i.e., They veritably reflect the logical relationships among the fuzzy concepts described by given data.

Again by inspecting Figure 6.2, we note that for any sample x, one of the following three situations may occur:

$$\mu_{\gamma}(x)+\mu_{\gamma'}(x)=1, \quad \mu_{\gamma}(x)+\mu_{\gamma'}(x) < 1, \quad \mu_{\gamma}(x)+\mu_{\gamma'}(x) > 1,$$

where γ' is the negation of γ. This situation emphasizes that in the AFS fuzzy logic the knowledge of the membership $\mu_{\gamma}(x)$ is not sufficient to determine the value of the negation, that is $\mu_{\gamma'}(x)$. As before the distribution of the original data influences the degree x belongs to γ' and γ. For example, if ζ is the fuzzy concept of the form: "*beautiful car*" and x is a person, then $\mu_{\zeta}(x)+\mu_{\zeta'}(x) < 1$ and the differences of the degrees of x belonging to ζ' and ζ may be not too much significant, or some x may belong both to the concept of "*beautiful car*" and the concept "*not beautiful car*" to a very low degree. In other ways, $\mu_{\zeta}(x)+\mu_{\zeta'}(x) < 1$ if x cannot be distinguished when considering the terms of a beautiful car and a car that is not beautiful. Thus in practice, the sum of the degrees of an element belonging to the fuzzy concept and its negation may not be equal to 1. In general, for a two-valued concept $\nu, \mu_{\nu}(x)+\mu_{\nu'}(x)$ always equals to 1 and for simple concept ξ the sum, $\mu_{\xi}(x)+\mu_{\xi'}(x)$ might be almost equal to 1, however for some complex concept η like γ, the $\mu_{\eta}(x)+\mu_{\eta'}(x)$ can assume values that could be larger or lower than 1. It is the vagueness of the concept η that breaks the fundamental law of excluded middle. A systematic comparison between AFS fuzzy logical systems and the conventional fuzzy logic equipped by some t-norm still remains as an open problem.

Again, let us emphasize that in fuzzy logic systems equipped by some negation operator N, we have $\mu_{\gamma'}(x) = N(\mu_{\gamma}(x))$ and the value of the complement is fully determined by the membership degree $\mu_{\gamma}(x)$ and becomes independent from the distribution of the original data and the relationship between x and the semantics of γ.

6.2.3 Definitions of AFS Fuzzy Rough Sets

Since AFS fuzzy rough sets as a generalization in the sense of Definition 6.1 are based on the set inclusion relation "$\subseteq$", hence in what follows, we define four types of set inclusions according to the representations of the fuzzy concepts in the framework of the AFS theory.

Definition 6.2. Let X be a set and M be a set of simple concepts on X. Let (M, τ, X) be an AFS structure, S be a σ-algebra over X and ρ be a weight function $X \to [0,\infty)$. The *EI algebra inclusion* "$\subseteq_{EI}$", the *EII algebra inclusion* "$\subseteq_{EII}$", the *$E^{\#}I$ algebra inclusion* "$\subseteq_{E^{\#}I}$" and the *inclusion with the weight function* ρ "$\subseteq_{\rho}$" as follows:

(1) $\alpha \subseteq_{EI} \beta \Leftrightarrow \alpha \leq \beta$ in lattice $(EM, \vee, \wedge)$, for $\alpha, \beta \in EM$;
(2) $\alpha \subseteq_{EII} \beta \Leftrightarrow \forall x \in X, \alpha(x) \leq \beta(x)$ in lattice $(EXM, \vee, \wedge)$, $\alpha(x), \beta(x)$ are the *EII* algebra represented membership function of fuzzy concepts $\alpha, \beta \in EM$ defined as (5.10);
(3) $\alpha \subseteq_{E^{\#}I} \beta \Leftrightarrow \forall x \in X$, $\alpha(x) \leq \beta(x)$ in lattice $(E^{\#}X, \vee, \wedge)$, $\alpha(x), \beta(x)$ are the $E^{\#}I$ algebra represented membership function of fuzzy concepts α, β defined as (5.13) or (6.12);
(4) $\alpha \subseteq_{\rho} \beta \Leftrightarrow \forall x \in X, \mu_{\alpha}(x) \leq \mu_{\beta}(x), \mu_{\alpha}(x)$, $\mu_{\beta}(x)$ are membership function defined of fuzzy concepts α, β by (6.14) using the measure m_{ρ} defined by (6.13) for weight function ρ or given beforehand as the fuzzy set on X (e.g., the fuzzy set associating to the decision attribute in a decision system).

By Theorem 5.3, Proposition 5.4 and what is defined by (6.12), we know that

$$\alpha \subseteq_{EI} \beta \Rightarrow \alpha \subseteq_{E^{\#}I} \beta \Rightarrow \alpha \subseteq_{EII} \beta \Rightarrow \alpha \subseteq_{\rho} \beta \tag{6.19}$$

for $\alpha, \beta \in EM$ and for any weight function ρ. This implies that the membership functions of the fuzzy concepts in EM defined by (5.10), (5.13) and (6.14) are consistent, and they reflect the semantics of the concepts and the distributions of the original data. We will define different AFS fuzzy rough sets based on these set inclusions.

Example 6.2. Let $X = \{x_1, x_2, ..., x_{10}\}$ be a set of 10 people and their features (attributes) which are described by real numbers (age, height, weight, salary, estate), Boolean values (gender) and the ordered relations (hair black, hair white, hair yellow), see Table 6.2; there the number i in the "*hair color*" columns which corresponds to some $x \in X$ implies that the hair color of x has ordered ith following our perception of the color by our intuitive perception. Let $M = \{m_1, m_2, ..., m_{10}\}$ be the set of fuzzy or Boolean concepts on X and each $m \in M$ associate to a single feature. Where m_1 : *"old people"*, m_2 : *"tall people"*, m_3 : *"heavy people"*, m_4 : *"high salary"*, m_5 : *"more estate"*, m_6 :*"male"*, m_7 : *"female"*, m_8 : *"black hair people"*, m_9 : *"white hair people"*, m_{10} : *"yellow hair people"*.

Let (M, τ, X) be the AFS structure of the data shown in Table 6.2. For simplicity, let S=2^X be the σ-algebra over X and m_ρ be the measure defined by (6.13) for $\rho(x)$=1, $\forall x \in X$. Let α=$m_2m_3m_5 + m_2m_6m_7 + m_2m_9 + m_2m_3m_4 + m_4m_5m_6\ m_7 + m_4m_5m_9 + m_3m_4m_5 + m_6m_8m_9$, β=$m_6m_8m_9 + m_4m_5 + m_2$. By Theorem 4.1, we can verify that $\alpha \le \beta$. Thus by Definition 6.2, we have $\alpha \subseteq_{EI} \beta$. This inclusion relation is determined in terms of the semantics of α and β. Let ν=$m_1m_7 + m_2m_7$, γ=$m_7m_9 + m_5m_7m_{10}$, ζ=$m_1m_6 + m_2m_7$, ξ=$m_2m_7m_9 + m_1\ m_6 + m_2m_3m_5m_7m_{10} \in EM$. Then by the AFS structure (M, τ, X), formula (5.13) and Theorem 5.24, we can verify that in lattice $(E^{\#}X, \vee, \wedge)$, for any $x \in X$, $\nu(x) \le \gamma(x)$, $\zeta(x) \le \xi(x)$. Thus by Definition 6.2, we have $\nu \subseteq_{E^{\#}I} \gamma$, $\zeta \subseteq_{E^{\#}I} \xi$. This inclusion relation is determined by both the semantic interpretations of α, β and the distribution of the original data shown in Table 6.2. By Theorem 4.1, we have $\xi \le \zeta$ in lattice $(EM, \vee, \wedge)$. This implies that $\xi \subseteq_{E^{\#}I} \zeta$ and for any $x \in X$, $\zeta(x)$=$\xi(x)$ in lattice $(E^{\#}X, \vee, \wedge)$ and $\mu_\zeta(x)$=$\mu_\xi(x)$ under meaning (6.14), in this case denoted as $\xi =_{E^{\#}I} \zeta$, $\xi =_\rho \zeta$. It is obvious that for

Table 6.2 Descriptions of features

		appearance		wealth		gender		hair color		
	age	*height*	*weigh*	*salary*	*estate*	*male*	*female*	*black*	*white*	*yellow*
x_1	20	1.9	90	1	0	1	0	6	1	4
x_2	13	1.2	32	0	0	0	1	4	3	1
x_3	50	1.7	67	140	34	0	1	6	1	4
x_4	80	1.8	73	20	80	1	0	3	4	2
x_5	34	1.4	54	15	2	1	0	5	2	2
x_6	37	1.6	80	80	28	0	1	6	1	4
x_7	45	1.7	78	268	90	1	0	1	6	4
x_8	70	1.65	70	30	45	1	0	3	4	2
x_9	60	1.82	83	25	98	0	1	4	3	1
x_{10}	3	1.1	21	0	0	0	1	2	5	3

$\eta \in EM$, if $\xi \leq \eta \leq \zeta$, then $\xi =_{E^{\#}I} \eta =_{E^{\#}I} \zeta$ and $\xi =_{\rho} \eta =_{\rho} \zeta$. The following Table 6.3 shows the membership functions of ν and γ defined by (6.14). By Table 6.3 and Definition 6.2, one knows that $\nu \subseteq_{\rho} \gamma$. This inclusion relation is determined by the semantics of ν and γ, the distribution of the original data and the function ρ which expresses how much each $x \in X$ contributes to the concepts under consideration.

Table 6.3 Membership functions defined by (6.14)

fuzzy concepts	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
$\mu_{\nu}(.)$	0	0.2	0.7	0	0	0.5	0	0	0.9	0.1
$\mu_{\gamma}(.)$	0	0.6	1.0	0	0	1.0	0	0	1.0	0.2

Proposition 6.2. *Let X be a set and M be a set of simple concepts on X. Let (M, τ, X) be an AFS structure. Then the following assertions hold: for $\alpha, \beta, \gamma, \eta \in EM$,*

(1) For the EI algebra inclusions

$$\gamma \subseteq_{EI} \alpha \Rightarrow \alpha' \subseteq_{EI} \gamma'$$
$$\gamma \subseteq_{EI} \alpha,\ \ \eta \subseteq_{EI} \beta \Rightarrow \gamma \wedge \eta \subseteq_{EI} \alpha \wedge \beta,\ \ \gamma \vee \eta \subseteq_{EI} \alpha \vee \beta$$
$$\alpha \subseteq_{EI} \gamma,\ \ \beta \leq \alpha \Rightarrow \beta \subseteq_{EI} \gamma$$
$$\gamma \subseteq_{EI} \alpha,\ \ \alpha \leq \beta \Rightarrow \gamma \subseteq_{EI} \beta;$$

(2) For the EII algebra inclusions

$$\gamma \subseteq_{EII} \alpha \Rightarrow \alpha' \subseteq_{EII} \gamma'$$
$$\gamma \subseteq_{EII} \alpha,\ \ \eta \subseteq_{EII} \beta \Rightarrow \gamma \wedge \eta \subseteq_{EII} \alpha \wedge \beta,\ \ \gamma \vee \eta \subseteq_{EII} \alpha \vee \beta$$
$$\alpha \subseteq_{EII} \gamma,\ \ \beta \leq \alpha \Rightarrow \beta \subseteq_{EII} \gamma$$
$$\gamma \subseteq_{EII} \alpha,\ \ \alpha \leq \beta \Rightarrow \gamma \subseteq_{EII} \beta;$$

(3) For the $E^{\#}I$ algebra inclusions

$$\gamma \subseteq_{E^{\#}I} \alpha,\ \ \eta \subseteq_{E^{\#}I} \beta \Rightarrow \gamma \wedge \eta \subseteq_{E^{\#}I} \alpha \wedge \beta,\ \ \gamma \vee \eta \subseteq_{E^{\#}I} \alpha \vee \beta$$
$$\alpha \subseteq_{E^{\#}I} \gamma,\ \ \beta \leq \alpha \Rightarrow \beta \subseteq_{E^{\#}I} \gamma$$
$$\gamma \subseteq_{E^{\#}I} \alpha,\ \ \alpha \leq \beta \Rightarrow \gamma \subseteq_{E^{\#}I} \beta;$$

For any $A, B \subseteq M$, if $\forall x \in X, A^{\tau}(x) \subseteq B^{\tau}(x) \Rightarrow \Sigma_{m \in A}\{m'\}^{\tau}(x) \geq \Sigma_{m \in B}\{m'\}^{\tau}(x)$, then

$$\alpha \subseteq_{E^{\#}I} \beta \Rightarrow \beta' \subseteq_{E^{\#}I} \alpha',$$

where m' is the negation of the simple concept $m \in M$.

Proof. (1) to (2) can be proved through the direct use of the definitions. Their proofs are left to the reader.

(3) For any $x \in X$, α=$\sum_{i\in I}(\Pi_{m\in Ai}m)$, β=$\sum_{j\in J}(\Pi_{m\in Bj}m)$ $\in EM$, if $\sum_{i\in I} A_i^{\tau}(x) \leq \sum_{j\in J} B_j^{\tau}(x)$ in lattice (EX, $\vee$, $\wedge$), then for any $i \in I$, $\exists k \in J$ such that $A_i^{\tau}(x) \subseteq B_k^{\tau}(x)$. By the assumption, we have $\sum_{m\in A_i}\{m'\}^{\tau}(x) \geq \sum_{m\in B_k}\{m'\}^{\tau}(x)$. This implies that for any $x \in X$, there exists a map $p_x : I \to J$, $\forall i \in I$, $p_x(i) = k \in J$ such that $\sum_{m\in A_i}\{m'\}^{\tau}(x) \geq \sum_{m\in B_k}\{m'\}^{\tau}(x)$ in lattice $(E^{\#}X, \vee, \wedge)$. Thus for any $x \in X$,

$$\begin{aligned}\left(\sum_{i\in I}(\Pi_{m\in A_i}m)\right)'(x) &= \wedge_{i\in I}(\sum_{m\in A_i}\{m'\}^{\tau}(x)) \\ &\geq \wedge_{i\in I}(\sum_{m\in B_{p_x(i)}}\{m'\}^{\tau}(x)) \\ &\geq \wedge_{j\in J}(\sum_{m\in B_j}\{m'\}^{\tau}(x)) \\ &= \left(\sum_{j\in J}(\Pi_{m\in B_j}m)\right)'(x)\end{aligned}$$

Therefore $(\sum_{i\in I}\Pi_{m\in A_i}m)' \subseteq_{E^{\#}I} (\sum_{j\in J}\Pi_{m\in B_j}m)'$. □

In what follows, by making use of the four types of fuzzy set inclusions defined in Definition 6.2, the four types of fuzzy rough sets (*AFS fuzzy rough sets*) are defined in Definition 6.3.

Definition 6.3. Let X be a set and M be a set of simple concepts on X. Let $\Lambda \subseteq EM$, $\sum_{m\in M} m \in \Lambda$, $\forall\alpha \in \Lambda$, $\alpha' \in \Lambda$. Let $(\Lambda)_{EI}$ be the sub EI algebra generated by the fuzzy concepts in Λ and $\rho : X \to [0,\infty)$. Let the fuzzy set inclusions "$\subseteq_{EI}$", "$\subseteq_{EII}$", "$\subseteq_{E^{\#}I}$" and "$\subseteq_{\rho}$" be denoted by "$\subseteq_i$" where i denotes EI, EII, $E^{\#}I$ or ρ (in virtue of Definition 6.2, ρ induces a measure for the membership function defined by (6.14)). For any fuzzy concept $\gamma \in EM$ or any fuzzy set γ which is given in advance to describe the decision result in an information system (under "$\subseteq_i$", i=EI, EII, $E^{\#}I$ or ρ), the upper approximation and lower approximation (denoted as $S^*(\gamma)$ and $S_*(\gamma)$ respectively) of γ are called *AFS rough sets with regard to the set of fuzzy concepts* Λ and defined as follows:

$$S_*(\gamma) = \bigvee_{\beta\in(\Lambda)_{EI},\ \beta\subseteq_i\gamma} \beta, \quad S^*(\gamma) = \bigwedge_{\beta\in(\Lambda)_{EI},\ \gamma\subseteq_i\beta} \beta, \tag{6.20}$$

Where $(\Lambda)_{EI}$ is the sub EI algebra generated by the fuzzy concepts in Λ. The *AFS fuzzy approximation spaces* are denoted by $\mathbb{A}^{EI} = (M,\Lambda)$, $\mathbb{A}^{EII} = (M,\Lambda,X)$, $\mathbb{A}^{E\#I} = (M,\Lambda,X)$ and $\mathbb{A}^{\rho} = (M,\Lambda,\rho,X)$, respectively.

For the fuzzy concept γ, the AFS fuzzy rough sets $S^*(\gamma)$, $S_*(\gamma) \in EM$ with well-defined semantic interpretations are the approximate descriptions of fuzzy concept γ using the given fuzzy concepts in Λ.

6.2.4 Some Properties of AFS Fuzzy Rough Sets

In this section, we will prove that AFS fuzzy rough sets not only have all the properties of the "conventional" rough sets defined by (6.1), but also are equivalent to the rough sets defined (6.1) if every concept in Λ is a Boolean one.

Theorem 6.1. *Let X be a non-empty set and M be a set of Boolean concepts on X. Then the rough sets in $\mathbb{A}^{EII}=(M,\Lambda,X)$, $\mathbb{A}^{E\#I}=(M,\Lambda,X)$ or $\mathbb{A}^{\rho}=(M,\Lambda,\rho,X)$ defined by (6.20) is equivalent to the rough sets defined by (6.2) in Definition 6.1 (i.e., rough sets in the classic sense defined by (6.1)).*

Proof. We just prove the theorem for $\mathbb{A}^{E^{\#}I}$ and the others remain as exercises. Since every concept in M is a Boolean concept, hence from Definition 4.3 and formula (5.13) and (6.12), either $\gamma(x)=X$ or $\gamma(x)=\varnothing$ in the lattice $E^{\#}X$, for any $\gamma\in EM$ and any $x\in X$. It can be verified that $\{X,\varnothing\}\subseteq E^{\#}X$ is a sub $E^{\#}I$ algebra. If p: $\{X,\varnothing\}\rightarrow\{1,0\}$ defined by $p(X)$=1, $p(\varnothing)$=0, then p is an $E^{\#}I$ algebra isomorphism from $\{X,\varnothing\}$ to the Boolean algebra. Thus each fuzzy set in EM is degenerated to a Boolean subset in X. Furthermore the fuzzy logic operations $\vee,\wedge,'$ in $(EM,\vee,\wedge,')$ are degenerated to the set operations $\cap,\cup,'$ and the inclusion "$\subseteq_{E^{\#}I}$" is degenerated to set inclusion "$\subseteq$". $(\Lambda)_{EI}$ is the set of sets generated by sets in Λ, using Boolean set operations $\cap,\cup,'$. This implies that $\forall x\in X$, $\delta_x\in\Lambda^{-}\subseteq(\Lambda)_{EI}$ (δ_x and Λ^{-} refer to Definition 6.1) and for any $d\in EM$,

$$\bigvee_{\beta\in(\Lambda)_{EI},\beta\subseteq d}\beta\supseteq\bigcup_{x\in X,\delta_x\subseteq d}\delta_x. \tag{6.21}$$

Since $d\subseteq\bigcup_{x\in d}\delta_x\in(\Lambda)_{EI}$. Hence

$$\bigwedge_{\beta\in(\Lambda)_{EI},d\subseteq\beta}\beta\subseteq\bigcup_{x\in d}\delta_x. \tag{6.22}$$

For any $x\in\bigvee_{\beta\in(\Lambda)_{EI},\beta\subseteq d}\beta$, $\exists\beta\in\Lambda$ such that $x\in\beta\subseteq d$. This implies that $\delta_x\subseteq d$ and $x\in\bigcup_{y\in X,\delta_y\subseteq d}\delta_y$. Therefore

$$\bigvee_{\beta\in(\Lambda)_{EI},\beta\subseteq d}\beta=\bigcup_{x\in X,\delta_x\subseteq d}\delta_x \tag{6.23}$$

If there exists $x\in d$ such that,

$$\delta_x\nsubseteq\bigwedge_{\beta\in(\Lambda)_{EI},d\subseteq\beta}\beta,$$

then $\exists\beta\in\Lambda$ such that $\beta\supset d$ and $\delta_x\nsubseteq\beta$. One can verify that $x\in\delta_x\cap\beta$ and $\delta_x\cap\beta\in\Lambda^{-}$. This implies that $\delta_x\cap\beta$ is a proper subset of δ_x, contradicting the fact that $\delta_x\in\Lambda^{-}$ and δ_x is the smallest set containing x. Therefore

$$\bigwedge_{\beta\in(\Lambda)_{EI},d\subseteq\beta}\beta=\bigcup_{x\in d}\delta_x \tag{6.24}$$

□

The following Proposition 6.3 prove that all properties (1)-(8) of the conventional rough sets listed in Proposition 6.1 hold for the AFS fuzzy rough sets defined by Definition 6.3.

Proposition 6.3. *Let X be a set and M be a set of simple concepts on X. Let (M,τ,X) be an AFS structure and $\mathbb{A}^i$ be the AFS fuzzy approximation spaces defined by (6.20) for $i = EI, EII, E^{\#}I, \rho$. Then for any fuzzy sets α, $\beta \in EM$ or whose membership functions are given in advance, the following assertions of AFS fuzzy rough sets $S^*(\alpha)$, $S_*(\beta)$ hold:*

(1) $S_*(\alpha) \subseteq_i \alpha \subseteq_i S^*\alpha)$ *for* $i = EI, EII, E^{\#}I, \rho$;
(2) $S_*(M) = M = S^*(M)$ *for* $i = EI, EII, E^{\#}I, \rho$;
(3) $S_*(\sum_{m\in M} m) = \sum_{m\in M} m = S^*(\sum_{m\in M} m)$ *for* $i = EI, EII, E^{\#}I, \rho$;
(4) If $\alpha \subseteq_i \beta$, *then* $S_*(\alpha) \subseteq_i S_*(\beta)$ *and* $S^*(\alpha) \subseteq_i S^*(\beta)$ *for* $i = EI, EII, E^{\#}I, \rho$;
(5) $S_*(S_*(\alpha)) = S_*(\alpha)$, $S^*(S^*(\alpha)) = S^*(\alpha)$ *for* $i = EI, EII, E^{\#}I, \rho$;
(6) $S^*(S_*(\alpha)) = S_*(\alpha)$, $S_*(S^*(\alpha)) = S^*(\alpha)$ *for* $i = EI, EII, E^{\#}I, \rho$;
(7) $(S^*(\alpha'))' = S_*(\alpha)$, $(S_*(\alpha'))' = S^*(\alpha)$ *for* $i = EI, EII$ *and* $\alpha \in EM$;
$(S^*(\alpha'))' = S_*(\alpha)$, $(S_*(\alpha'))' = S^*(\alpha)$ *for* $i = E^{\#}I$ *and* $\alpha \in EM$ *provided that for* A, $B \subseteq M$, $\forall x \in X$, $A^\tau(x) \subseteq B^\tau(x) \Rightarrow \sum_{m\in A}\{m'\}^\tau(x) \geq \sum_{m\in B}\{m'\}^\tau(x)$;
(8) $S_*(\alpha\wedge\beta) = S_*(\alpha)\wedge S_*(\beta)$, $S^*(\alpha\wedge\beta) \subseteq_i S^*(\alpha)\wedge S^*(\beta)$ *for* $i = EI, EII, E^{\#}I, \rho$;
(9) $S^*(\alpha\vee\beta) = S^*(\alpha)\vee S^*(\beta)$, $S_*(\alpha)\vee S_*(\beta) \subseteq_i S_*(\alpha\vee\beta)$ *for* $i = EI, EII, E^{\#}I, \rho$.

Proof. The proofs of (1)-(6) which remain as exercises can be proved directly by the definitions and the theorems. The proof of (7) is completed as follows. First we prove it hold for *i*=*EI*. Using formula (4.19), for every fuzzy set $\alpha \in EM$, we have

$$\begin{aligned}(S^*(\alpha'))' &= \left(\bigwedge_{\beta\in(\Lambda)_{EI},\alpha'\subseteq_{EI}\beta} \beta\right)' \\ &= \bigwedge_{\beta\in(\Lambda)_{EI},\alpha'\subseteq_{EI}\beta} \beta' \\ &= \bigwedge_{\beta\in(\Lambda)_{EI},\beta'\subseteq_{EI}\alpha} \beta' \\ &= \bigwedge_{\gamma\in(\Lambda)_{EI},\gamma\subseteq_{EI}\alpha} \gamma \\ &= S_*(\alpha)\end{aligned}$$

Similarly, one can prove that $(S_*(\alpha'))' = S^*(\alpha)$. In a similar fashion, one can prove (7) for $i = EII, E^{\#}I$.

Let us present a proof of (8). For any fuzzy sets α, $\beta \in EM$ and $i = EI$, we have

$$S_*(\alpha\wedge\beta) = \bigvee_{\gamma\in(\Lambda)_{EI},\gamma\subseteq_{EI}(\alpha\wedge\beta)} \gamma \leq \bigvee_{\gamma\in(\Lambda)_{EI},\gamma\subseteq_{EI}\alpha} \gamma = S_*(\alpha).$$

Similarly, one can prove that $S_*(\alpha\wedge\beta) \leq S_*(\beta)$. This implies that

$$S_*(\alpha \wedge \beta) \leq S_*(\alpha) \wedge S_*(\beta).$$

$$\begin{aligned} S_*(\alpha) \wedge S_*(\beta) &= \left(\bigvee_{\gamma \in (\Lambda)_{EI},\ \gamma \subseteq_{EI} \alpha} \gamma \right) \wedge \left(\bigvee_{\eta \in (\Lambda)_{EI},\ \eta \subseteq_{EI} \beta} \eta \right) \\ &= \left(\bigvee_{\gamma,\ \eta \in (\Lambda)_{EI},\ \gamma \subseteq_{EI} \alpha,\ \eta \subseteq_{EI} \beta} \gamma \wedge \eta \right) \leq \left(\bigvee_{\zeta \in (\Lambda)_{EI},\ \zeta \subseteq_{EI} \alpha \wedge \beta} \zeta \right) \\ &= S_*(\alpha \wedge \beta) \end{aligned}$$

For any fuzzy sets α, $\beta \in EM$, we have

$$S^*(\alpha \wedge \beta) = \bigwedge_{\gamma \in (\Lambda)_{EI},\ (\alpha \wedge \beta) \subseteq_{EI} \gamma} \gamma \leq \bigwedge_{\gamma \in (\Lambda)_{EI},\ \alpha \subseteq_{EI} \gamma} \gamma = S^*(\alpha).$$

Similarly, we can prove that $S^*(\alpha \wedge \beta) \leq S^*(\beta)$. Therefore $S^*(\alpha \wedge \beta) \leq S^*(\alpha) \wedge S^*(\beta)$ and $S^*(\alpha \wedge \beta) \subseteq_{EI} S^*(\alpha) \wedge S^*(\beta)$.

For $i = E^{\#}I$, α, $\beta \in EM$ or whose membership functions are given in advance. We have $\alpha \wedge \beta \subseteq_i \alpha$ and $\alpha \wedge \beta \subseteq_i \beta$. It follows by (4), $S_*(\alpha \wedge \beta) \subseteq_i S_*(\alpha) \wedge S_*(\beta)$.

$$\begin{aligned} S_*(\alpha) \wedge S_*(\beta) &= \left(\bigvee_{\zeta \in (\Lambda)_{EI},\ \zeta \subseteq_{E^{\#}I} \alpha} \zeta \right) \wedge \left(\bigvee_{\eta \in (\Lambda)_{EI},\ \eta \subseteq_{E^{\#}I} \beta} \eta \right) \\ &= \left(\bigvee_{\zeta,\ \eta \in (\Lambda)_{EI},\ \zeta \subseteq_{E^{\#}I} \alpha,\ \eta \subseteq_{E^{\#}I} \beta} \zeta \wedge \eta \right) \\ &\subseteq_i \left(\bigvee_{\zeta \in (\Lambda)_{EI},\ \zeta \subseteq_{E^{\#}I} \alpha \wedge \beta} \zeta \right) \\ &= S_*(\alpha \wedge \beta) \end{aligned}$$

Thus $S_*(\alpha \wedge \beta) = S_*(\alpha) \wedge S_*(\beta)$. By the similar method, we can prove that $S^*(\alpha \wedge \beta) \subseteq_i S^*(\alpha) \wedge S^*(\beta)$ and $S^*(\alpha \wedge \beta) \subseteq_i S^*(\alpha) \wedge S^*(\beta)$. For $i = EII, \rho$, we also can prove them in the same way as $i = E^{\#}I$. Therefore (8) holds.

The proof of (9). For $i = EI$ and any fuzzy sets $\alpha, \beta \in EM$, we have

$$\begin{aligned} S_*(\alpha \vee \beta) &= \bigvee_{\gamma \in (\Lambda)_{EI},\ \gamma \subseteq_{EI} (\alpha \vee \beta)} \gamma \\ &\geq \bigvee_{\gamma \in (\Lambda)_{EI},\ \gamma \subseteq_{EI} \alpha} \gamma \\ &= S_*(\alpha). \end{aligned}$$

Similarly, one can prove $S_*(\alpha\vee\beta)\geq S_*(\beta)$. Therefore $S_*(\alpha\vee\beta)\geq S_*(\alpha)\vee S_*(\beta)$ and $S_*(\alpha)\vee S_*(\beta)\subseteq_{EI} S_*(\alpha\vee\beta)$.

For any fuzzy sets α, $\beta\in EM$, we have

$$\begin{aligned} S^*(\alpha\vee\beta) &= \bigwedge_{\gamma\in(\Lambda)_{EI},\ (\alpha\vee\beta)\subseteq_{EI}\gamma} \gamma \\ &\geq \bigwedge_{\gamma\in(\Lambda)_{EI},\ \alpha\subseteq_{EI}\gamma} \gamma \\ &= S^*(\alpha). \end{aligned}$$

Similarly, one can prove $S^*(\alpha\vee\beta)\geq S^*(\beta)$. This implies that $S^*(\alpha\vee\beta)\geq S^*(\alpha)\vee S^*(\beta)$.

$$\begin{aligned} S^*(\alpha)\vee S^*(\beta) &= \left(\bigwedge_{\gamma\in(\Lambda)_{EI},\ \alpha\subseteq_{EI}\gamma} \gamma\right)\vee\left(\bigwedge_{\eta\in(\Lambda)_{EI},\ \beta\subseteq_{EI}\eta} \eta\right) \\ &= \left(\bigwedge_{\gamma,\ \eta\in(\Lambda)_{EI},\ \alpha\subseteq_{EI}\gamma,\ \beta\subseteq_{EI}\eta} \gamma\vee\eta\right) \\ &\leq \left(\bigwedge_{\zeta\in(\Lambda)_{EI},\ (\alpha\vee\beta)\subseteq_{EI}\zeta} \zeta\right) \\ &= S^*(\alpha\vee\beta) \end{aligned}$$

Similarly, for $i=EII, E^{\#}I, \rho$, we can prove (9). □

Proposition 6.4. *Let X be a set and M be a set of simple concepts on X. Let (M, τ, X) be an AFS structure and S be a σ-algebra over X and ρ: X →[0,∞). Let* $\Lambda\subseteq EM$ *and M,* $\sum_{m\in M} m\in\Lambda$, $\forall\alpha\in\Lambda$, $\alpha'\in\Lambda$. *Let* $\mathbb{A}^{EI}=(M,\Lambda)$, $\mathbb{A}^{EII}=(M,\Lambda,X)$, $\mathbb{A}^{E^{\#}I}=(M,\Lambda,X)$, $\mathbb{A}^{\rho}=(M,\Lambda,\rho,X)$ *be the AFS fuzzy approximation spaces defined by (6.20). For any* $\gamma\in EM$ *or whose membership function is given in advance, let*

$$S^*_{EI}(\gamma),\ S_*^{EI}(\gamma);\ \ S^*_{EII}(\gamma),\ S_*^{EII}(\gamma);\ \ S^*_{E^{\#}I}(\gamma),\ S_*^{E^{\#}I}(\gamma);\ \ S^*_{\rho}(\gamma),\ S_*^{\rho}(\gamma)$$

be the upper and lower approximations for γ *in* $\mathbb{A}^{EI}$, $\mathbb{A}^{EII}$, $\mathbb{A}^{E^{\#}I}$, $\mathbb{A}^{\rho}$, *respectively. Then in lattice (EM, ∧,∨) the following assertions hold*

$$S_*^{EI}(\gamma)\leq S_*^{EII}(\gamma)\leq S_*^{E^{\#}I}(\gamma)\leq S_*^{\rho}(\gamma)\leq\gamma\leq\ S^*_{\rho}(\gamma)\leq\ S^*_{E^{\#}I}(\gamma)\leq S^*_{EII}(\gamma)\leq S^*_{EI}(\gamma) \tag{6.25}$$

Proof. Since for α, $\beta\in EM$, $\alpha\subseteq_{EI}\beta\Rightarrow\alpha\subseteq_{EII}\beta\Rightarrow\alpha\subseteq_{E_I^{\#}}\beta\Rightarrow\alpha\subseteq_{\rho}\beta$, hence (6.25) holds. □

6.2.5 Algorithm of Approximating Upper and Lower Approximations of Fuzzy Concepts

If there are a few fuzzy concepts in Λ, then $S_*(\gamma), S^*(\gamma)$, defined by (6.20), the upper and lower approximations for $\gamma \in EM$ or whose membership function is given in advance, can be easily determined. To determine $S_*(\gamma)$, $S^*(\gamma)$ given a large set Λ becomes a challenging problem and quite often we have to resort ourselves to some approximate solutions. An algorithm to determine $\overline{S}_*(\gamma)$, $\overline{S}^*(\gamma)$ forming the approximate solutions to (6.20) is outlined below.

Let (M,τ,X) be an AFS structure of a given information system. In AFS fuzzy approximation space $\mathbb{A}^{EII} = (M,\Lambda,X)$, $\mathbb{A}^{E^\#I} = (M,\Lambda,X)$ or $\mathbb{A}^{\rho} = (M,\Lambda,\rho,X)$. Let $\gamma \in EM$ or whose membership function is given in advance. For $\mathbb{A}^{EII}$ or $\mathbb{A}^{E^\#I}$, suppose that $\forall x \in X$, $\exists \alpha \in \Lambda$, such that $\alpha(x) \geq \gamma(x)$ in the lattice EXM or $E^\#X$, where $\alpha(x)$ and $\gamma(x)$ are the AFS algebra membership degrees of x belonging to the fuzzy concepts α, γ defined by (5.10), (5.13) or (6.12). For $\mathbb{A}^{\rho}$, suppose that $\forall x \in X$, $\exists \alpha \in \Lambda$, such that $\mu_\alpha(x) \geq \mu_\gamma(x)$, where $\mu_\alpha(x)$ and $\mu_\gamma(x)$ are the membership degrees of x belonging to the fuzzy concepts α, γ defined by (6.14).

- **STEP 1:** For each $x \in X$, find the set $B_x \subseteq \Lambda$ defined as follows

$$B_x^{EII} = \{\alpha \in \Lambda \mid \alpha(x) \geq \gamma(x)\},\ in\ \mathbb{A}^{EII}, \tag{6.26}$$

$$B_x^{E^\#I} = \{\alpha \in \Lambda \mid \alpha(x) \geq \gamma(x)\},\ in\ \mathbb{A}^{E^\#I}, \tag{6.27}$$

$$B_x^{\rho} = \{\alpha \in \Lambda \mid \mu_\alpha(x) \geq \mu_\gamma(x)\},\ in\ \mathbb{A}^{\rho}. \tag{6.28}$$

  Since for any $x \in X$, $\exists \alpha \in \Lambda$ such that or $\mu_\alpha(x) \geq \mu_\gamma(x)$, hence for any $x \in X$, $B_x \neq \varnothing$. It is clear that

$$B_x^{EII} \subseteq B_x^{E^\#I} \subseteq B_x^{\rho} \subseteq \Lambda \subseteq EM.$$

  In practice, each of B_x^{EII}, $B_x^{E^\#I}$ and B_x^{ρ} has a much lower number of elements than the Λ.
- **STEP 2:** The approximate solutions to the upper approximations defined by (6.20) in $\mathbb{A}^{EII}$, $\mathbb{A}^{E^\#I}$ and $\mathbb{A}^{\rho}$ are listed as follows.

$$\overline{S}^*_{EII}(\gamma) = \bigvee_{x \in X} \left(\bigwedge_{\eta \in B_x^{EII}} \eta \right) \in (\Lambda)_{EI}, \tag{6.29}$$

$$\overline{S}^*_{E^\#I}(\gamma) = \bigvee_{x \in X} \left(\bigwedge_{\eta \in B_x^{E^\#I}} \eta \right) \in (\Lambda)_{EI}, \tag{6.30}$$

$$\overline{S}^*_{\rho}(\gamma) = \bigvee_{x \in X} \arg \min_{\eta \in (B_x^{\rho})_{EI}} \{\eta \mid \mu_\eta(x) \geq \mu_\gamma(x)\} \in (\Lambda)_{EI}. \tag{6.31}$$

In what follows, we will prove that

$$S_{\rho}^{*}(\gamma)\leq\overline{S}_{\rho}^{*}(\gamma),\ \ S_{E^{\#}I}^{*}(\gamma)\leq\overline{S}_{E^{\#}I}^{*}(\gamma),\ \ S_{EII}^{*}(\gamma)\leq\overline{S}_{EII}^{*}(\gamma),$$
$$S_{\rho}^{*}(\gamma)\leq\overline{S}_{\rho}^{*}(\gamma)\leq\overline{S}_{E^{\#}I}^{*}(\gamma)\leq\overline{S}_{EII}^{*}(\gamma).$$

That is, what are given in (6.29), (6.30) and (6.31) are the approximate solutions to the upper approximations of γ defined by (6.20). Since $B_x^{EII}\subseteq B_x^{E^{\#}I}\subseteq B_x^{\rho}$, hence $\overline{S}_{\rho}^{*}(\gamma)\leq\overline{S}_{E^{\#}I}^{*}(\gamma)\leq\overline{S}_{EII}^{*}(\gamma)$. From Theorem 5.3 and Proposition 5.4, we know that over the AFS structure (M,τ,X), for any given $x\in X$, the maps $p_x : EM\rightarrow EXM$, $q_x : EM\rightarrow E^{\#}X$ defined as follows are homomorphisms from the lattice $(EM,\wedge,\vee)$ to the lattice $(EXM,\wedge,\vee)$ and from the lattice $(EM,\wedge,\vee)$ to the lattice $(E^{\#}X,\wedge,\vee)$, respectively. For any $\prod_{m\in Ai}m\in EM$,

$$p_x\left[\sum_{i\in I}(\prod_{m\in Ai}m)\right]=\sum_{i\in I}A_i^{\tau}(x)A_i\in EXM,$$
$$q_x\left[\sum_{i\in I}(\prod_{m\in Ai}m)\right]=\sum_{i\in I}A_i^{\tau}(x)\in E^{\#}X.$$

Hence by (6.26) and (6.27), for any $x\in X$ we have

$$\left(\bigwedge_{\eta\in B_x^{EII}}\eta\right)(x)=\bigwedge_{\eta\in B_x^{EII}}\eta(x)\geq\gamma(x), \tag{6.32}$$
$$\left(\bigwedge_{\eta\in B_x^{E^{\#}I}}\eta\right)(x)=\bigwedge_{\eta\in B_x^{E^{\#}I}}\eta(x)\geq\gamma(x). \tag{6.33}$$

This implies that for any $x\in X$,

$$\left[\bigvee_{x\in X}\left(\bigwedge_{\eta\in B_x^{EII}}\eta\right)\right](x)=\bigvee_{x\in X}\left(\bigwedge_{\eta\in B_x^{EII}}\eta(x)\right)\geq\gamma(x),$$
$$\left[\bigvee_{x\in X}\left(\bigwedge_{\eta\in B_x^{E^{\#}I}}\eta\right)\right](x)=\bigvee_{x\in X}\left(\bigwedge_{\eta\in B_x^{E^{\#}I}}\eta(x)\right)(x)\geq\gamma(x).$$

Furthermore, by Definition 6.2 we have

$$\gamma\subseteq_{EII}\bigvee_{x\in X}\left(\bigwedge_{\eta\in B_x^{EII}}\eta\right)\in(\Lambda)_{EI},$$
$$\gamma\subseteq_{E^{\#}I}\bigvee_{x\in X}\left(\bigwedge_{\eta\in B_x^{E^{\#}I}}\eta\right)\in(\Lambda)_{EI}.$$

It follows by Definition 6.3, we have $S^*_{E^\# I}(\gamma) \le \overline{S}^*_{E^\# I}(\gamma),\ S^*_{EII}(\gamma) \le \overline{S}^*_{EII}(\gamma)$. Since for any $x \in X$, $B^\rho_x \subseteq \Lambda$, hence $(B^\rho_x)_{EI} \subseteq (\Lambda)_{EI}$ and

$$\xi_x \triangleq \arg \min_{\eta \in (B^\rho_x)_{EI}} \{\eta \mid \mu_\eta(x) \ge \mu_\gamma(x)\} \in (\Lambda)_{EI} \tag{6.34}$$

For any $x \in X$, we have

$$\mu_{\overline{S}^*_\rho(\gamma)}(x) = \mu_{\bigvee_{x\in X}\xi_x}(x) \ge \mu_\gamma(x) \Rightarrow \gamma \subseteq_\rho \overline{S}^*_\rho(\gamma).$$

By Definition 6.3, we have $S^*_\rho(\gamma) \le \overline{S}^*_\rho(\gamma)$. Thus we have

$$S^*_\rho(\gamma) \le \overline{S}^*_\rho(\gamma) \le \overline{S}^*_{E^\# I}(\gamma) \le \overline{S}^*_{EII}(\gamma).$$

Next, we find the approximate solutions to the lower approximations of γ defined by (6.20). For each $x \in X$, since (6.32), (6.33) and (6.34), the lower approximations of γ in $\mathbb{A}^{EII}$, $\mathbb{A}^{E^\# I}$, $\mathbb{A}^{E^\# I}$ should be searched in the following sets: D^{EII}_x, $D^{E^\# I}_x$ and D^ρ_x.

$$\begin{aligned} D^{EII}_x &= \left\{ \xi \in (\Lambda)_{EI} \mid \xi \le \bigwedge_{\eta \in B^{EII}_x} \eta \right\} \subseteq (\Lambda)_{EI}, \\ D^{E^\# I}_x &= \left\{ \xi \in (\Lambda)_{EI} \mid \xi \le \bigwedge_{\eta \in B^{E^\# I}_x} \eta \right\} \subseteq (\Lambda)_{EI}, \\ D^\rho_x &= \{\xi \in (\Lambda)_{EI} \mid \xi \le \xi_x\} \subseteq (\Lambda)_{EI}. \end{aligned}$$

Since $\bigwedge_{\eta \in B^{EII}_x} \eta \ge \bigwedge_{\eta \in B^{E^\# I}_x} \eta \ge \xi_x$, hence $D^{EII}_x \supseteq D^{E^\# I}_x \supseteq D^\rho_x$. In practice, D^{EII}_x, $D^{E^\# I}_x$ and D^ρ_x have a much lower number of elements than $(\Lambda)_{EI}$. Thus we have the approximate solutions to the lower approximations defined by (6.20) in $\mathbb{A}^{EII}$, $\mathbb{A}^{E^\# I}$ and $\mathbb{A}^\rho$ listed as follows.

$$\overline{S}^i_*(\gamma) = \bigvee_{x \in X} \left(\bigvee_{\eta \in D^i_x,\ \eta \subseteq_i \gamma} \eta \right) \in (\Lambda)_{EI}, \tag{6.35}$$

for $i = EII, E^\# I, \rho$. It is clear that $\overline{S}^i_*(\gamma) \le S^i_*(\gamma)$. We should note that for each $x \in X$, if $\eta \subseteq_i \gamma$, then for any $\varsigma \in D^i_x$, $\varsigma \le \eta$ in lattice EM,

$$\varsigma \vee \left(\bigvee_{\eta \in D^i_x,\ \eta \subseteq_i \gamma} \eta \right) = \left(\bigvee_{\eta \in D^i_x,\ \eta \subseteq_i \gamma} \eta \right).$$

By noting this, we can reduce computing overhead when looking for $\overline{S}^i_*(\gamma)$. Furthermore $\bigvee_{\eta \in D^i_x,\ \eta \subseteq_i \gamma} \eta$, $x \in X$ can be computed independently which brings some thoughts about its potential parallel realization of the overall process. By Definition 6.2, we know that for any weight function ρ

$$\overline{S}_{*}^{\rho}(\gamma) \subseteq_{\rho} S_{*}^{\rho}(\gamma) \subseteq_{\rho} \gamma \subseteq_{\rho} S_{\rho}^{*}(\gamma) \subseteq_{\rho} \overline{S}_{\rho}^{*}(\gamma).$$

Thus we can apply this algorithm to find $\overline{S}_{*}^{i}(\gamma), \overline{S}_{i}^{*}(\gamma)$ being the approximate solutions to (6.20) in AFS fuzzy approximation space $\mathbb{A}^{i}$ for $i = EII, E^{\#}I, \rho$.

In the following examples, $\overline{S}_{*}^{i}(\gamma)$ and $\overline{S}_{i}^{*}(\gamma)$ are found with the use of this algorithm for $i = EII, E^{\#}I, \rho$.

Example 6.3. Let M=$\{m_1, m_2, \ldots, m_{18}\}$ where m_1 to m_{10} are the same as presented Example 6.2 while m_{18}= m'_3, $m_{17} = m'_4$, $m_{16} = m'_5$, $m_{15} = m'_6$, $m_{13} = m'_8$, m_{12}= m'_9, $m_{11} = m'_{10}$. Here for $m \in M$, m' is the negation of concept m. According to Table 6.2, the AFS structure (M, τ, X) is established by means of (4.26). Let σ-algebra on X be S=2^X and the measure m be defined as $\text{m}(A) = |A|/|X|$, i.e., the weight function $\rho(x)$=1, $\forall x \in X$.

Suppose that γ=$m_1m_7 + m_2m_7 + m_1m_2 \in EM$ to describe some decision result in a data shown as Table 6.2. Then the set of simple concepts associating to condition features is Λ=$M - \{m_1, m_2, m_7\}$. Running the algorithm presented above we obtain

$$\begin{aligned}
\overline{S}_{*}^{E^{\#}I}(\gamma) &= m_3m_4m_6m_8m_9m_{11}m_{13}m_{16} + m_3m_4m_6m_9m_{10}m_{11}m_{13}m_{16} \\
&+ m_3m_4m_6m_9m_{11}m_{13}m_{16}m_{18} + m_3m_4m_6m_9m_{11}m_{13}m_{16}m_{17} + m_3m_4m_6m_9m_{11}m_{12}m_{13}m_{16} \\
&+ m_3m_4m_9m_{10}m_{13}m_{15}m_{16}m_{17} + m_3m_9m_{10}m_{13}m_{15}m_{16}m_{17}m_{18} \\
&+ m_3m_9m_{10}m_{12}m_{13}m_{15}m_{16}m_{17} + m_5m_9m_{11}m_{13}m_{15} + m_9m_{11}m_{12}m_{13}m_{15} \\
&+ m_6m_{15} + m_4m_5m_6m_8m_{11}m_{12}m_{17} + m_5m_6m_9 + m_5m_6m_{16} + m_5m_6m_{13} \\
&+ m_3m_5m_{10}m_{15} + m_5m_8m_{10}m_{15} + m_5m_9m_{10}m_{15} + m_5m_{10}m_{15}m_{18} \\
&+ m_5m_{10}m_{15}m_{17} + m_5m_{10}m_{15}m_{16} + m_5m_{10}m_{13}m_{15} + m_5m_{10}m_{12}m_{15} + m_5m_{10}m_{11}m_{15},
\end{aligned}$$

$$\overline{S}_{E^{\#}I}^{*}(\gamma) = m_3m_9m_{10}m_{13}m_{15}m_{16}m_{17} + m_9m_{11}m_{13}m_{15} + m_6 + m_5m_{10}m_{15}.$$

The membership functions of fuzzy concepts $\overline{S}_{*}^{E^{\#}I}(\gamma), \overline{S}_{E^{\#}I}^{*}(\gamma) \in EM$ are shown in Table 6.4. Since they are approximate solutions to (6.20) in AFS fuzzy approximation space $\mathbb{A}^{E^{\#}I}$, hence the results of the real solution of (6.20) should be better than this.

Table 6.4 The membership functions of $\overline{S}_{*}^{E^{\#}I}(\gamma)$ and $\overline{S}_{E^{\#}I}^{*}(\gamma)$ defined by (6.14)

fuzzy concepts	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
$\mu_{\overline{S}_{E^{\#}I}^{*}(\gamma)}(.)$	1.0	0.2	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.1
$\mu_{\gamma}(.)$	0.3	0.2	0.7	0.8	0.3	0.5	0.5	0.5	0.9	0.1
$\mu_{\overline{S}_{*}^{E^{\#}I}(\gamma)}(.)$	0.1	0.1	0.6	0.3	0.3	0.5	0.1	0.2	0.6	0.1

6.2.6 Experiments

In this section, some experiments involving the *Credit-Screening*, *Iris Data*, the *Wine Classification Data*, and *Wisconsin Breast Cancer* coming from the Machine-learning database at University of California, Irvine [34] are presented to clarify the concepts and properties of AFS fuzzy rough sets and demonstrate the computational details of the approach.

A. **Credit-Screening Data:** Let $(M,\ \tau,\ X)$ be the AFS structure of the credit-screening data established in Example 6.1 and the σ-algebra over X be $S=2^X$. Let us study the credit assignment problems using the data presented in Example 6.1, by viewing positive (samples 1 to 85) and negative (samples 86 to 150) instances (that is people who were and were not granted credit). Let ζ be a fuzzy set defined as follows. For $x \in X$, $\mu_\zeta(x)=1$, if sample x comes with a positive credit decision, otherwise $\mu_\zeta(x)=0.15$ to indicate the membership degrees of negative samples belonging to fuzzy concept "*credit*". Any small number can be arranged to each negative sample to show its extent belonging to "*credit*". The differences of membership degrees of positive samples and negative samples reflect the grades we distinguish positive instances from negative ones. However, it is clear that the larger the difference, the more difficult is to find the lower and upper approximation. Here if $\mu_\zeta(x)=0.1$ for the negative samples, then its lower approximation will be difficult to determine following the algorithm presented here. By formula (6.15), we have $\rho_\zeta(x)=0.85/85=0.01$ if sample x comes with a positive credit decision, otherwise $\rho_\zeta\ (x)=0.15/40=0.0037$. Let m_ζ be the measure over σ-algebra S for ρ_ζ defined as (6.13). If we view this as an information system, $M_A = M-\{m_1,$ $m_2\}$, $M_d=\{m_1,\ m_2\}$, where m_1: *"positive credit"* and m_2: *"negative credit"*. Let $\Lambda=M-\{m_1, m_2\}$, then $\mathbb{A}^{\rho_\zeta} = (M,\Lambda,\rho_\zeta,X)$ becomes an AFS fuzzy approximation space in sense of definition given by (6.20). With the help of (6.31) and (6.35), we obtain $\overline{S}_*^{\rho_\zeta}(\zeta)$, $\overline{S}^*_{\rho_\zeta}(\zeta)$, which are approximate solutions to (6.20) shown in Figure 6.3. $\overline{S}_*^{\rho_\zeta}(\zeta),\overline{S}^*_{\rho_\zeta}(\zeta) \in EM$ have 490 terms and 50 terms, respectively.

$$
\begin{aligned}
\overline{S}^*_{\rho_\zeta}(\zeta) &= m_4m_5m_8m_{10}m_{12}m_{14}m_{16}m_{18}m_{20}m_{21}m_{24} \\
&\quad +m_4m_5m_8m_{10}m_{12}m_{14}m_{16}m_{18}m_{20}m_{22}m_{24} \\
&\quad +m_4m_5m_8m_{10}m_{12}m_{14}m_{16}m_{18}m_{20}m_{21}m_{23}+\ldots\ldots. \\
\overline{S}_*^{\rho_\zeta}(\zeta) &= m_4m_9m_{18}m_{20}m_{21}m_{24}m_{31}+m_4m_6m_8m_9m_{14}m_{19}m_{22}m_{31} \\
&\quad +m_4m_6m_8m_9m_{16}m_{19}m_{22}m_{31}+m_4m_6m_8m_9m_{18}m_{19}m_{22}m_{31}+\ldots.
\end{aligned}
$$

As illustrated in Figure 6.3, although $\overline{S}_*^{\rho_\zeta}(\zeta),\overline{S}^*_{\rho_\zeta}(\zeta) \in EM$ have a large number of terms (490 and 50, respectively), they come with a well defined semantics expressed by the simple concepts in M shown in Example 6.1. For instance, the first term in $\overline{S}_*^{\rho_\zeta}(\zeta)$ "$m_4m_9m_{18}m_{20}m_{21}m_{24}m_{31}$" states that **IF a person is** characterized as the one with "*Job*, *purchase stereo*, *purchase furniture*, *female*, *unmarried*, *located in no- problematic region*, *expected to pay off loan more*" , **THEN the person is characterized by a positive credit score**. The first term in $\overline{S}^*_{\rho_\zeta}(\zeta)$

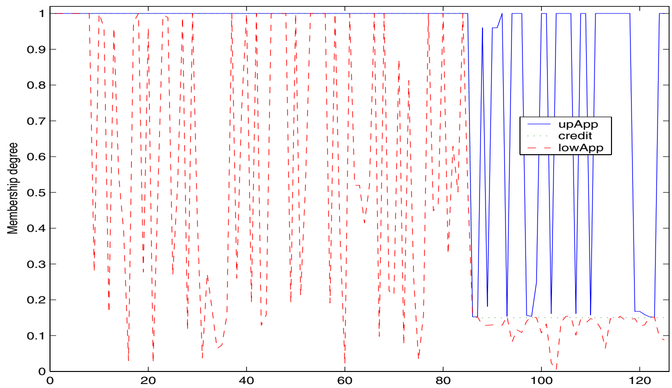


Fig. 6.3 Membership functions of $\overline{S}_*^{\rho_\zeta}(\zeta), \zeta = credit, \overline{S}^*_{\rho_\zeta}(\zeta)$

"$m_4m_5m_8m_{10}m_{12}m_{14}m_{16}m_{18}m_{20}m_{21}m_{24}$" states that **IF a person is described as** "*job*, *purchase pc*, *no purchase car*, *no purchase stereo*, *no purchase jewel*, *no purchase medinstru*, *no purchase bike*, *no purchase furniture*, *female*, *unmarried*, *located in no- problematic region*", **THEN the person has positive credit score**. Similarly, one can find the semantics of other terms in them. In the following Example 6.4, we will show the semantics of the lower and upper approximations using the concepts with few terms. Thus through the data given in Example 6.1, the concept "*positive credit*" formed on the decision attributes and which is modeled by some fuzzy set is approximated by its upper and lower bounds $\overline{S}_*^{\rho_\zeta}(\zeta), \overline{S}^*_{\rho_\zeta}(\zeta) \in E(M - \{m_1, m_2\})$.

B. **Iris Data:** The iris data set is one of the most popular data sets to examine the performance of novel methods in pattern recognition and machine learning. Here the number of instances is equal to 150, evenly distributed in three classes:1:50 iris-setosa, 51:100 iris-versicolor, and 101:150 iris-virginica, where each of the samples is described by 4 features: sepal length, sepal width, petal length, petal width. Λ=M=$\{m_1, m_2, \ldots, m_{16}\}$ is the concepts formulated on the features of the information system $< X, A >$, where $m_{4(i-1)+1}, m_{4(i-1)+2}, m_{4(i-1)+3}, m_{4(i-1)+4}$, i=1, 2, 3, 4 are the simple concepts "*large*", "*medium*", "*not medium*", "*small*" associating to the ith feature respectively and the weight function of them $\rho(x)$ is defined by $\rho(x)$=1 for all $x \in X$. $\mathbb{A}^\rho = (M, \Lambda, \rho, X)$ becomes an AFS fuzzy approximation space in sense of definition given by (6.20). Let ζ be a fuzzy set defined as follows. For $x \in X$, $\mu_\zeta(x)$=0.7, if sample x in the class iris-setosa, otherwise $\mu_\zeta(x)$=0.1 to indicate the membership degrees of negative samples belonging to "*iris-setosa*". With the help of (6.31) and (6.35), we obtain $\overline{S}_*^{\rho}(\zeta)$, $\overline{S}^*_{\rho}(\zeta)$ listed as follows, which are approximate solutions to (6.20) shown in Figure 6.4.

$$\begin{aligned}\overline{S}^*_{\rho}(\zeta) &= m_{12} + m_4 + m_{16}m_{18} + m_4m_7m_{10} \\ &\quad + m_2m_6m_{11}m_{14}m_{15}m_{18} + m_3m_6m_7m_{10}m_{11}m_{14}m_{15}, \\ \overline{S}_*^{\rho}(\zeta) &= m_6m_7m_{16}m_{18} + m_4m_7m_{10}m_{16}m_{18} + m_4m_7m_9m_{12}m_{16}m_{18}.\end{aligned}$$

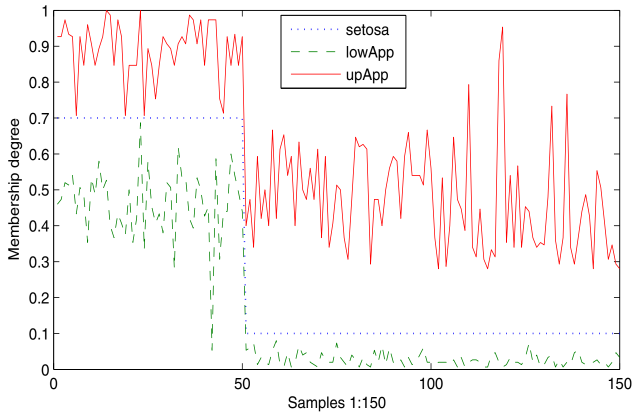


Fig. 6.4 Membership functions of $\overline{S}_{*}^{\rho}(\zeta)$, ζ:"iris-setosa", $\overline{S}_{\rho}^{*}(\zeta)$ for Iris data

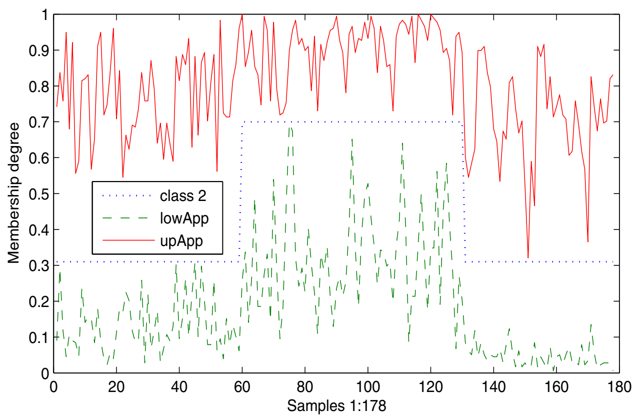


Fig. 6.5 Membership functions of $\overline{S}_{*}^{\rho}(\zeta)$, ζ:"class–2", $\overline{S}_{\rho}^{*}(\zeta)$ for Wine data

C. Wine Classification Data: Chemical analysis of wines grown in the same region in Italy, but derived from three different cultivars, should be sufficient to recognize the source of the wine. Here the number of instances is equal to 178, evenly distributed in three classes:1:59 class-1, 60:130 class-2, and 131:178 class-3, where each of the samples is described by 13 features, including alcohol content, hue, color intensity, and content 9 chemical compounds. Λ=M=$\{m_1, m_2, \ldots, m_{26}\}$ is the concepts formulated on the features of the information system $<X,A>$, where $m_{2(i-1)+1}, m_{2(i-1)+2}$, i=1, 2, ..., 13 are the simple concepts "*large*", "*small*" associating to the ith feature respectively and the weight function of them $\rho(x)$ is defined by $\rho(x)$=1 for all $x \in X$. $\mathbb{A}^{\rho} = (M, \Lambda, \rho, X)$ becomes an AFS fuzzy approximation space in sense of definition given by (6.20). Let ζ be a fuzzy set defined as follows. For $x \in X$, $\mu_{\zeta}(x)$=0.7, if sample x in the class-2, otherwise $\mu_{\zeta}(x)$=0.31 to indicate the membership degrees of negative samples belonging to "class-2". With the help of (6.31) and (6.35), we obtain $\overline{S}^{\rho}_{*}(\zeta), \overline{S}^{*}_{\rho}(\zeta)$, which are approximate solutions to (6.20) shown in Figure 6.5. $\overline{S}^{\rho}_{*}(\zeta), \overline{S}^{*}_{\rho}(\zeta) \in EM$, respectively are described as follows.

$$\begin{aligned}
\overline{S}^{*}_{\rho}(\zeta) &= m_2 + m_6 + m_{20} + m_{15}m_{24} + m_{11}m_{13}, \\
\overline{S}^{\rho}_{*}(\zeta) &= m_2m_6m_{12}m_{21} + m_2m_{11}m_{16}m_{20} + m_2m_{15}m_{19}m_{23} \\
&\quad + m_2m_6m_{13}m_{23} + m_6m_{11}m_{13}m_{20} + m_2m_6m_{16}m_{20}m_{23} \\
&\quad + m_{15}m_{20}m_{21}m_{23}m_{24} + m_{14}m_{15}m_{20}m_{23}m_{24} \\
&\quad + m_2m_6m_{20}m_{23}m_{24} + m_2m_5m_{20}m_{22}m_{23} + m_{15}m_{20}m_{22}m_{23}m_{24} \\
&\quad + m_2m_{16}m_{20}m_{23}m_{24} + m_2m_{12}m_{16}m_{20}m_{23} + m_2m_{16}m_{20}m_{22}m_{23} \\
&\quad + m_2m_5m_6m_{13}m_{20}m_{21} + m_2m_6m_{14}m_{20}m_{21}m_{23} \\
&\quad + m_2m_6m_{14}m_{15}m_{20}m_{23} + m_2m_6m_{12}m_{15}m_{20}m_{23} \\
&\quad + m_2m_5m_6m_{12}m_{15}m_{23} + m_2m_5m_6m_{15}m_{20}m_{23}.
\end{aligned}$$

We can observe that the number of items of the $\overline{S}^{\rho_\zeta}_{*}(\zeta), \overline{S}^{*}_{\rho_\zeta}(\zeta)$, which are approximate solutions to (6.20) for *Credit-screening* are much larger than that of *Iris data* and the *Wine classification data*. The main reason is that 12/17 features of *Credit-screening* are described by Boolean value while all features of *Iris data* and the *Wine classification data* are numerical and the associating simple concepts are all fuzzy. This implies that the AFS fuzzy rough sets have some advantages to deal with fuzzy information systems.

D. Wisconsin Breast Cancer Data: The Wisconsin Breast Cancer Diagnostic data set contains 699 patterns distributed into two output classes, "*benign*" and "*malignant*". Each pattern consists of nine input features. There are 16 patterns with incomplete feature descriptions. We use 683 patterns 1:444 class benign, 445:683 class

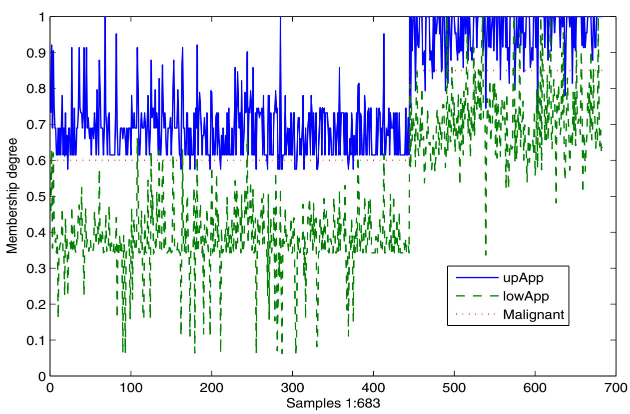


Fig. 6.6 Membership functions of $\overline{S}^{\rho}_{*}(\zeta)$, ζ:"malignant", $\overline{S}^{*}_{\rho}(\zeta)$ for Wisconsin Breast Cancer data

malignant, to evaluate the performance of the proposed fuzzy algorithm. Λ=M=$\{m_1, m_2, \ldots, m_{18}\}$ is the concepts formulated on the features of the information system $< X, A >$, where $m_{2(i-1)+1}$, $m_{2(i-1)+2}$, i=1, 2, 3, 4, 5, 6, 7, 8 9, are the simple concepts "*large*", "*small*" associating to the ith feature respectively and the weight function of them $\rho(x)$ is defined by $\rho(x)$=1 for all $x \in X$. $\mathbb{A}^\rho = (M, \Lambda, \rho, X)$ becomes the fuzzy approximation space in sense of definition given by (6.20). Let ζ be a fuzzy set defined as follows. For $x \in X$, $\mu_\zeta(x)$=0.85, if sample x in the class malignant, otherwise $\mu_\zeta(x)$=0.6 to indicate the membership degrees of negative samples belonging to "malignant". With the help of (6.31) and (6.35), we obtain $\overline{S}_*^\rho(\zeta), \overline{S}_\rho^*(\zeta)$, which are approximate solutions to (6.20) shown in Figure 6.6. $\overline{S}_*^\rho(\zeta)$, $\overline{S}_\rho^*(\zeta) \in EM$ are shown as follows.

$$
\begin{aligned}
\overline{S}_\rho^*(\zeta) &= m_1 + m_3 + m_5 + m_7 + m_9 + m_{13}, \\
\overline{S}_*^\rho(\zeta) &= m_1 m_3 m_7 m_9 + m_1 m_7 m_{11} m_{13} m_{17} + m_1 m_3 m_7 m_{11} m_{13} \\
&\quad + m_1 m_5 m_9 m_{13} m_{17} + m_1 m_3 m_5 m_9 m_{15} m_{17} + m_1 m_3 m_5 m_{13} m_{15} m_{17} \\
&\quad + m_3 m_5 m_7 m_9 m_{11} m_{15} + m_1 m_5 m_7 m_9 m_{11} m_{15} m_{17} \\
&\quad + m_5 m_7 m_9 m_{11} m_{13} m_{15} m_{17} + m_1 m_5 m_7 m_9 m_{11} m_{13} m_{15}.
\end{aligned}
$$

If the data includes noisy samples or the knowledge in Λ is insufficient, then for some sample x, there may not exist δ, $\beta \in (\Lambda)_{EI}$ such that $\mu_\delta(x) \leq \mu_\zeta(x)$ or $\mu_\beta(x) \geq \mu_\zeta(x)$ and δ, β satisfying (6.20). Therefore in Figure 6.6, one can observe that the degrees of some samples belonging to $\overline{S}_\rho^*(\zeta)$ are less than that of ζ or the degree of some samples belonging to $\overline{S}_*^\rho(\zeta)$ are larger than that of ζ. This implies that these samples cannot be distinguished by the given simple concepts in Λ, may be noisy samples or our algorithm fails to determine δ, $\beta \in (\Lambda)_{EI}$ satisfying (6.20) for these samples.

6.3 Comparisons with Fuzzy Rough Sets and Other Constructs of Rough Sets

Here we provide some comparative analysis of the AFS fuzzy rough sets with previously developed forms of rough sets. This will cast the investigations in some broader perspective.

6.3.1 Comparisons with Fuzzy Rough Sets

The main question addressed by previous fuzzy rough sets can be formulated as follows: how to represent a fuzzy set $\gamma \in \mathscr{F}(X) = \{\beta : X \to [0,1]\}$ by some fuzzy sets $S_*(\gamma)$, $S^*(\gamma) \in \mathscr{F}(X)$, according to a border implicator ϕ: [0,1]×[0,1]→[0,1] and a similarity relation R, i.e., a fuzzy relation, R: $X \times X$ →[0,1] for which the following conditions should hold:

(1) $\forall\, x \in X, R(x, x)$=1; (Reflexivity)
(2) $\forall\, x, y \in X, R(x, y) = R(y, x)$; (Symmetry)
(3) $\forall\, x, y, z \in X, R(x, z) \geq T(R(x, y), R(y, z))$. (T-transitivity)

Here, T is a t-norm. In [8], the authors defined $S_*(\gamma)$, $S^*(\gamma)$ the lower and upper fuzzy approximations of $\gamma \in \mathscr{F}(X)$ as follows: for $x \in X$,

$$\mu_{S^*(\gamma)}(x) = \sup_{y\in X} \min\{\mu_\gamma(y), R(x,y)\}, \tag{6.36}$$

$$\mu_{S_*(\gamma)}(x) = \inf_{y\in X} \max\{\mu_\gamma(y), 1 - R(x,y)\}. \tag{6.37}$$

Which are the same as what are described by (6.4) and (6.5). In order to compare AFS fuzzy rough sets with the fuzzy rough sets, we study the following example.

Example 6.4. Let X=$\{x_1, x_2, \ldots, x_{10}\}$ and R=(r_{ij}), $r_{ij} = R(x_i, x_j)$, where R is a similarity relation

$$R = \begin{pmatrix} 1 & .56 & 1 & .50 & .67 & 1 & .37 & .50 & .54 & .11 \\ & 1 & .56 & .50 & .56 & .56 & .37 & .50 & .54 & .11 \\ & & 1 & .50 & .67 & 1 & .37 & .50 & .54 & .11 \\ & & & 1 & .50 & .50 & .37 & .81 & .50 & .11 \\ & & & & 1 & .67 & .37 & .50 & .54 & .11 \\ & & & & & 1 & .37 & .50 & .54 & .11 \\ & & & & & & 1 & .37 & .37 & .11 \\ & & & & & & & 1 & .50 & .11 \\ & & & & & & & & 1 & .11 \\ & & & & & & & & & \end{pmatrix} \tag{6.38}$$

One can verify that R is a T-similarity relation [8] on X if t-norm is the minimum and t-conorm is the maximum operator. For the fuzzy set γ with $\mu_\gamma(x_1)$=1, $\mu_\gamma(x_2)$=.50, $\mu_\gamma(x_3)$=1, $\mu_\gamma(x_4)$=.70, $\mu_\gamma(x_5)$=.50, $\mu_\gamma(x_6)$=1, $\mu_\gamma(x_7)$=.20, $\mu_\gamma(x_8)$=.60, $\mu_\gamma(x_9)$=.50, $\mu_\gamma(x_{10})$=.10. The lower and upper fuzzy approximations of γ defined by (6.36) and (6.37) are shown as Table 6.5.

Table 6.5 The fuzzy rough sets based on fuzzy relation Rdefined by (6.36) and (6.37)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
$\mu_{S^*(\gamma)}(x)$	1.0	0.56	1.0	.70	.67	1.0	.37	.70	.54	.11
$\mu_\gamma(x)$	1.0	.50	1.0	.70	.50	1.0	.20	.60	.50	.10
$\mu_{S_*(\gamma)}(x)$	.50	.50	.50	.50	.50	.50	.20	.50	.50	.10

Let M=$\{m_1, m_2, \ldots, m_{20}\}$ be the set of some fuzzy concepts on X. Where the fuzzy concept m_{2k+1}: *"similar to x_{k+1}"*, k=0, 1, …, 9 and m_{2k+2}: *"not similar to x_{k+1}"* is the negation of m_{2k+1}. For the fuzzy similarity relation R=(r_{ij}), $r_{ki} \geq r_{kj}$ implies that the degree of x_i being similar to x_k is larger than or equal to that of x_j. Thus each concept in M is a simple concept on X. According to the information provided by the fuzzy similarity relation R=(r_{ij}) shown as (6.38), we can establish the AFS structure (M, τ, X) as follows. For any x_i, $x_j \in X$, introduce

$\tau(x_i,x_j)$={ m_{2k+1} | $r_{(k+1)i} \geq r_{(k+1)j}$, $0\leq k \leq 9$ }∪{ m_{2l+2} | $1-r_{(l+1)i} \geq 1-r_{(l+1)i}$, $0\leq l \leq 9$ }.

Let σ-algebra over X be S=2^X and m be the measure over S defined by (6.13) for ρ: $X \to [0, \infty)$, $\rho(x)$=1 for any $x \in X$. Thus for any fuzzy concept $\eta \in EM$, we can obtain the membership function $\mu_\eta(x)$ by (6.14). Let $\Lambda=\{m_i \mid i=1, 2, \ldots, 20\} \subseteq EM$. There exist the lower and upper fuzzy sets $S_*(\eta)$ and $S^*(\eta)$ in $\mathbb{A}^\rho = (M,\Lambda,\rho,X)$ defined by (6.20) for any fuzzy set $\eta \in \mathscr{F}(X)$. Based on (6.31) and (6.35) which are the algorithm of finding the approximate solutions of (6.20), we have the following $\overline{S}_*^\rho(\gamma)$ and $\overline{S}_\rho^*(\gamma)$ of the above fuzzy set γ and their membership functions are shown in Table 6.6.

$$\overline{S}_\rho^*(\gamma) = m_1m_5m_{11}m_{20} + m_2m_4m_6m_7m_{10}m_{12}m_{13}m_{11}m_{15}m_{18}m_{19}m_{20}$$
$$+m_1m_2m_3m_4m_5m_6m_7m_8m_{10}m_{11}m_{12}m_{13}m_{14}m_{15}m_{16}m_{17}m_{18}m_{19}m_{20},$$
$$\overline{S}_*^\rho(\gamma) = m_1m_7m_{20} + m_5m_7m_{20} + m_7m_{11}m_{20}.$$

Table 6.6 The approximate solutions to (6.20) for AFS upper and lower approximations of fuzzy set γ

fuzzy concepts	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
$\mu_{\overline{S}_\rho^*(\gamma)}(.)$	1.0	.60	1.0	.70	.70	1.0	.20	.70	.50	.10
$\mu_\gamma(.)$	1.0	.50	1.0	.70	.50	1.0	.20	.60	.50	.10
$\mu_{\overline{S}_*^\rho(\gamma)}(.)$	.80	.40	.80	.40	.50	.80	.20	.30	.30	0

Although the fuzzy similarity relation R is not required for AFS fuzzy rough sets, this example shows that we also can find AFS rough sets based on the same information provided by R as for other fuzzy rough sets. Thus Example 6.4 is the comparison under the same conditions. The MSEs of the upper and lower approximations in Table 6.5 which are 0.00731, 0.08, respectively are larger than the MSEs reported in Table 6.6 which are 0.006, 0.036, respectively. This implies that the upper and lower approximation in Table 6.5 provide less information than those in Table 6.6. In that way, the approximate AFS rough sets $\overline{S}_*^\rho(\gamma)$ and $\overline{S}_\rho^*(\gamma)$ obtained with the help of (6.31) and (6.35) shown in Table 6.6 provide a better interpretation than that fuzzy rough sets shown in Table 6.5. This raises an open problem: Is this result universal? Considering

$$\overline{S}_*^\rho(\gamma) \subseteq_\rho S_*^\rho(\gamma) \subseteq_\rho \gamma \subseteq_\rho \ S_\rho^*(\gamma) \subseteq_\rho \ \overline{S}_\rho^*(\gamma),$$

$S_*^\rho(\gamma), S_\rho^*(\gamma)$ which are defined by (6.20), can provide more accurate approximations of γ than $\overline{S}_*^\rho(\gamma)$ and $\overline{S}_\rho^*(\gamma)$.

In light of the simple concepts in M shown above, $S_*^\rho(\gamma) = m_1m_7m_{20} + m_5m_7m_{20} + m_7m_{11}m_{20}$ states that "*similar to* x_1,x_4 *and not similar to* x_{10}" or "*similar to* x_3,x_4 *and not similar to* x_{10}" or "*similar to* x_4,x_6 *and not similar to* x_{10}".

As a lower approximation of the given fuzzy set γ, $S_*^{\rho}(\gamma)$ describes the approximate semantics of γ with the simple concepts in M, according to the given membership function $\mu_\gamma(.)$ and the similar relation R shown in (6.38). By $\mu_\gamma(.)$ shown in Table 6.6, we can observe that the semantic interpretation of $S_*^{\rho}(\gamma)$ is very close to the fuzzy concept represented by $\mu_\gamma(.)$. For instance, the degrees of x_1 belonging to γ and $S_*^{\rho}(\gamma)$ are 1 and 0.8, respectively. Taking into the consideration the similarity relation R, we know x_1 is similar to x_3, x_4 and x_{10} at the degrees 1, 0.67, 0.11, respectively. This implies that x_1 should belong to the fuzzy concept $m_5m_7m_{20}$ with the semantics "*similar to* x_3,x_4 *and not similar to* x_{10}" at high degree. In Table 6.6, x_1 belongs to the lower approximation of γ at the degree 0.8. Similarly, other samples in Table 6.6 also show that the lower and upper approximation of γ approximately interpret γ by both the semantic meanings and the membership degrees based on the data shown in (6.38). The membership function of the fuzzy concept $S_*^{\rho}(\gamma) \in EM$ defined by (6.14), which is determined by its semantic meaning and the similar relations of the 10 elements in X described by the fuzzy relation R (6.38), gives the degree of each element in X approximately belonging to γ. The similar phenomena can be observed for $\overline{S}_{\rho}^{*}(\gamma)$, that is the upper approximation of γ.

6.3.2 Comparisons with Other Constructs of Rough Sets

Boolean reasoning has been used for many years and was helpful in solving many problems relative to rough sets such as those reported in [1, 37, 42]. It is well-known [37] that any Boolean function f: $\{0, 1\}^n \rightarrow \{0, 1\}$ can be presented in its canonical form, particularly in a so-called Conjunctive Normal Form (CNF) and Disjunctive Normal Form (DNF). Boolean reasoning is based on encoding the investigated optimization representations of a Boolean function f. In AFS fuzzy rough sets, we find the lower, upper approximations, which are some fuzzy sets in $(\Lambda)_{EI}$, $\Lambda \subseteq EM$ and approximately represent fuzzy set ξ on universe X with membership degrees in the interval [0, 1] or AFS algebras.

In Boolean reasoning, the representations of Boolean functions can be found by searching in the lattice of all subsets of attributes which is a Boolean algebra. In AFS fuzzy rough sets, the lower and upper approximations can be found by searching in the lattice *EM*. In [23], the author has proved that AFS algebra is a more general algebraic structure than Boolean algebra. Therefore Boolean reasoning and AFS reasoning revolve around search processes realized in different lattice structures.

In [23], the authors have proved that *EI* algebra *EM* is degenerated to Boolean algebra if each concept in M is a Boolean concept. Thus many Boolean reasoning ideas and techniques can be applied to study AFS fuzzy rough sets.

There are a great deal of papers on rough sets and multi-criteria decision making, in particular related to preference relations such as those discussed in [48, 50]. Sub-preference relations (refer to Definition 4.3), which are applied to define simple concepts in AFS theory, are more general than preference relations, i.e., any preference relation is a sub-preference relation. In [14], the author pointed that the main difficulty with application of many existing multiple-criteria decision aiding (MCDA) methods lays in acquisition of the decision maker's (DM's) preferential information. Very often, this information has to be given in terms of preference model parameters, like importance weights, substitution rates and various thresholds. Formally, for each $q \in C$ being a criterion there exists an outranking relation S_q on the set of actions U such that $(x, y) \in S_q$ means "x is at least as good as y with respect to criterion q". S_q is a total pre-order, i.e., a strongly complete and transitive binary relation defined on U on the basis of evaluations $f(.,q)$. In [11], the author described the problems as follows. Let $X = \Pi_{i=1}^{n} X_i$ be a product space including a set of actions, where X_i is a set of evaluations of actions with respect to criterion i=1, ..., n; each action x is thus seen as a vector $[x_1, x_2, \ldots, x_n]$ of evaluations on n criteria. A comprehensive weak preference relation $\geq$ needed to define on X such that for each x, $y \in X$, $x \geq y$ means "x is at least as good as y". The symmetric part of $\geq$ is the indifference relation, denoted by $\sim$, while the asymmetric part of $\geq$ is the preference relation, denoted by $\geq$.

These approaches to study how to establish evaluations $f(., q)$ or a comprehensive weak preference relation $\geq$ by the pair wise comparison tables (PCT) which represent preferential information provided by the decision makers in form of a pair wise comparison of reference actions. In general, evaluations $f(., q)$ or a comprehensive weak preference relation $\geq$ are defined by the rules defined in advance and for the same PCT, different $f(., q)$ or $\geq$ may be defined because of different rule sets. These researches mainly focus on the comparison of each pair of actions for decision and rarely apply preference relations to study the fuzzy sets. In AFS theory, the sub-preference relation on X of each simple concept in M is determined by the given data sets and each fuzzy set in EM is represented by the AFS fuzzy logic combination of the simple concepts in M. The membership functions and the fuzzy logic operations of fuzzy sets in EM are determined by the AFS structure (M, τ, X), a special family of combinatorial systems [15], which is directly established according to the distributions of the original data. And in AFS framework we do not need to define evaluations $f(., q)$ or a comprehensive weak preference relation $\geq$ in advance. In [13], by regarding each preference relation on X as a sub set of the product set $X \times X$, the lower, upper approximations of a preference relation are the Pawlak's rough sets, i.e., represented by some given preference relations as sub sets of $X \times X$. In the framework of AFS theory, first, we apply sub-preference relations, AFS logic and the norm on AFS algebras to obtain membership functions and fuzzy logic operations of fuzzy sets, then the AFS fuzzy rough sets are based on the established AFS fuzzy logic. The upper, lower approximations of each fuzzy set on the universe X can be represented by the fuzzy sets in EM which have well-defined semantics meaning.

There is a significant deal of study on rough sets based on the concept of tolerance (similarity) such as e.g., [48, 19]. The tolerant rough set extends the existing equivalent rough sets, i.e., Pawlak's rough sets. Let us recall that the binary relation R defined on $U \times U$ is a tolerance (similarity) relation if and only if $(a, a) \in R$ and $(a, b) \in R \Rightarrow (b, a) \in R$, where a, $b \in U$. The similarity class of x, denoted by $[x]_R$, the set of objects that are similar to $x \in U$, is defined as $[x]_R = \{y \in U \mid (y, x) \in R\} \subseteq U$. The rough approximation of a set $X \subseteq U$ is a pair of sets called lower and upper approximations of X, denoted by $R_*(X)$ and $R^*(X)$, respectively, where

$R^*(X) = \cup_{x\in X}[x]_R$, $R_*(X)$=$\{x \in X \mid [x]_R \subseteq X\}$. Compared with the AFS fuzzy rough sets, the tolerant relation R has to be given in advance, $R_*(X)$, $R^*(X)$ and X are all Boolean sets and the tolerant rough sets do not have a direct semantics.

In order to apply the equivalent rough sets or the tolerant rough sets to study the data shown in Example 6.2, the continuous valued attributes such as age, height, weigh, salary ... etc, are often discretized as intervals of real values taken by the Boolean attributes [57], or define a similarity measure that quantifies the closeness between attribute values of objects to construct a tolerance relation among the data [19]. The quality of rules discovered by the equivalent or tolerant rough sets is strongly affected by the result of the discretization or the similarity measure. For AFS fuzzy rough sets, the discretization of continuous valued attributes and the similarity measure are not required.

Following the investigations in the study one can observe that the proposed three types of AFS fuzzy rough sets are more practical and efficient when dealing with information systems in comparison with some other generalizations of rough fuzzy sets. It can be directly applied to handling data without a need to deal with an implicator ϕ, a t-norm and a similarity relation R that are required to be provided in advance. The AFS rough approximations for the fuzzy concepts have well-defined semantics with the given simple concepts formed for the individual features.

Exercises

Exercise 6.1. Let U be a set and $\mathscr{A}$=(U, R) be an approximation space. Prove that the lower approximation $A_*(X)$ and upper approximation $A^*(X)$ for any $X \subseteq U$ satisfy the following properties:

(1) $A_*(X) \subseteq X \subseteq A^*(X)$;
(2) $A_*(\varnothing) = A^*(\varnothing) = \varnothing, A_*(U) = A^*(U) = U$;
(3) $A_*(X \cap Y) = A_*(X) \cap A_*(Y), A^*(X \cup Y) = A^*(X) \cup A^*(Y)$;
(4) If $X \subseteq Y$, then $A_*(X) \subseteq A_*(Y), A^*(X) \subseteq A^*(Y)$;
(5) $A^*(X \cap Y) \subseteq A^*(X) \cap A^*(Y), A_*(X \cup Y) \supseteq A_*(X) \cup A_*(Y)$;
(6) $A^*(X')' = A^*(X), A_*(X')' = A_*(X)$;
(7) $A_*(A_*(X))$=$A^*\ (A_*(X))$=$A_*(X)$;
(8) $A^*(A^*(X))$=$A_*\ (A^*(X))$=$A^*(X)$.

Exercise 6.2. Let X be a set and M be a set of simple concepts on X. Let (M, τ, X) be an AFS structure. Show the following assertions hold: for $\alpha, \beta, \gamma, \eta \in EM$,

(1) For the EI algebra inclusions

$$\gamma \subseteq_{EI} \alpha \Rightarrow \alpha' \subseteq_{EI} \gamma'$$
$$\gamma \subseteq_{EI} \alpha,\ \ \eta \subseteq_{EI} \beta \Rightarrow \gamma \wedge \eta \subseteq_{EI} \alpha \wedge \beta,\ \ \gamma \vee \eta \subseteq_{EI} \alpha \vee \beta$$
$$\alpha \subseteq_{EI} \gamma,\ \ \beta \leq \alpha \Rightarrow \beta \subseteq_{EI} \gamma$$
$$\gamma \subseteq_{EI} \alpha,\ \ \alpha \leq \beta \Rightarrow \gamma \subseteq_{EI} \beta;$$

(2) For the EII algebra inclusions

$$\begin{aligned}
&\gamma \subseteq_{EII} \alpha \Rightarrow \alpha' \subseteq_{EII} \gamma' \\
&\gamma \subseteq_{EII} \alpha,\ \ \eta \subseteq_{EII} \beta \Rightarrow \gamma \wedge \eta \subseteq_{EII} \alpha \wedge \beta,\ \ \gamma \vee \eta \subseteq_{EII} \alpha \vee \beta \\
&\alpha \subseteq_{EII} \gamma,\ \ \beta \le \alpha \Rightarrow \beta \subseteq_{EII} \gamma \\
&\gamma \subseteq_{EII} \alpha,\ \ \alpha \le \beta \Rightarrow \gamma \subseteq_{EII} \beta.
\end{aligned}$$

Exercise 6.3. Let X be a non-empty set and M be a set of Boolean concepts on X. Prove that the rough sets in $\mathbb{A}^{EII} = (M, \Lambda, X)$ or $\mathbb{A}^{\rho} = (M, \Lambda, \rho, X)$ defined by (6.20) is equivalent to the rough sets defined by (6.2) in Definition 6.1 (i.e., rough sets in the classic sense defined by (6.1)).

Exercise 6.4. Let X be a set and M be a set of simple concepts on X. Let (M, τ, X) be an AFS structure and $\mathbb{A}^{i}$ be the AFS fuzzy approximation spaces defined by (6.20) for $i = EI, EII, E^{\#}I, \rho$. Prove that for any fuzzy sets α, $\beta \in EM$ or whose membership functions are given in advance, the following assertions of AFS fuzzy rough sets $S^{*}(\alpha)$, $S_{*}(\beta)$ hold:

(1) $S_{*}(\alpha) \subseteq_i \alpha \subseteq_i S^{*}\alpha)$ for $i = EI, EII, E^{\#}I, \rho$;
(2) $S_{*}(M) = M = S^{*}(M)$ for $i = EI, EII, E^{\#}I, \rho$;
(3) $S_{*}(\sum_{m \in M} m) = \sum_{m \in M} m = S^{*}(\sum_{m \in M} m)$ for $i = EI, EII, E^{\#}I, \rho$;
(4) If $\alpha \subseteq_i \beta$, then $S_{*}(\alpha) \subseteq_i S_{*}(\beta)$ and $S^{*}(\alpha) \subseteq_i S^{*}(\beta)$ for $i = EI, EII, E^{\#}I, \rho$;
(5) $S_{*}(S_{*}(\alpha)) = S_{*}(\alpha)$, $S^{*}(S^{*}(\alpha)) = S^{*}(\alpha)$ for $i = EI, EII, E^{\#}I, \rho$;
(6) $S^{*}(S_{*}(\alpha)) = S_{*}(\alpha)$, $S_{*}(S^{*}(\alpha)) = S^{*}(\alpha)$ for $i = EI, EII, E^{\#}I, \rho$;

Open problems

Problem 6.1. A systematic comparison between the AFS fuzzy logic systems and the conventional fuzzy logic system equipped by some t-norm.

Problem 6.2. Let X be a set and M be a set of simple concepts on X. Let (M, τ, X) be an AFS structure. What are the necessary and sufficient conditions that $\alpha \subseteq_{E^{\#}I} \beta \Rightarrow \beta' \subseteq_{E^{\#}I} \alpha'$.

Problem 6.3. Let (M, τ, X) be an AFS structure of a given information system. In AFS fuzzy approximation space $\mathbb{A}^{EII} = (M, \Lambda, X)$, $\mathbb{A}^{E^{\#}I} = (M, \Lambda, X)$ or $\mathbb{A}^{\rho} = (M, \Lambda, \rho, X)$. Let $\gamma \in EM$ or whose membership function is given in advance. For $\mathbb{A}^{EII}$ or $\mathbb{A}^{E^{\#}I}$, suppose that $\forall x \in X$, $\exists \alpha \in \Lambda$, such that $\alpha(x) \ge \gamma(x)$ in the lattice EXM or $E^{\#}X$, where $\alpha(x)$ and $\gamma(x)$ are the AFS algebra membership degrees of x belonging to the fuzzy concepts α, γ defined by (5.10), (5.13) or (6.12). For $\mathbb{A}^{\rho}$, suppose that $\forall x \in X$, $\exists \alpha \in \Lambda$, such that $\mu_{\alpha}(x) \ge \mu_{\gamma}(x)$, where $\mu_{\alpha}(x)$ and $\mu_{\gamma}(x)$ are the membership degrees of x belonging to the fuzzy concepts α, γ defined by (6.14). $\overline{S}^{*}_{EII}(\gamma)$, $\overline{S}^{EII}_{*}(\gamma)$ (refer to (6.29) and (6.35)), $\overline{S}^{*}_{E^{\#}I}(\gamma)$, $\overline{S}^{E^{\#}I}_{*}(\gamma)$ (refer to (6.30) and (6.35))and $\overline{S}^{*}_{\rho}(\gamma)$, $\overline{S}^{\rho}_{*}(\gamma)$ (refer to (6.31) and (6.35)) are the approximate solutions to the upper and lower approximations defined by (6.20) in $\mathbb{A}^{EII}$, $\mathbb{A}^{E^{\#}I}$ and $\mathbb{A}^{\rho}$. What are the sufficient and necessary conditions of the following assertions.

(1) $S^*_\rho(\gamma) = \overline{S}^*_\rho(\gamma)$, $S^*_{E^\# I}(\gamma) = \overline{S}^*_{E^\# I}(\gamma)$, $S^*_{EII}(\gamma) = \overline{S}^*_{EII}(\gamma)$;

(2) $S^\rho_*(\gamma) = \overline{S}^\rho_*(\gamma)$, $S^{E^\# I}_*(\gamma) = \overline{S}^{E^\# I}_*(\gamma)$, $S^{EII}_*(\gamma) = \overline{S}^{EII}_*(\gamma)$.

Problem 6.4. Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set and $R = (r_{ij}), r_{ij} = R(x_i, x_j)$ be a similarity relation which is a T-similarity relation [8] on X. For any fuzzy set $\gamma \in \mathscr{F}(X)$, let $S^*(\gamma)$ and $S_*(\gamma)$ be the lower and upper fuzzy approximations of fuzzy set γ defined by (6.36) and (6.37). Let $M = \{m_1, m_2, \ldots, m_{2n}\}$ be the set of some fuzzy concepts on X. Where the fuzzy concept m_{2k+1}: *"similar to* x_{k+1}", k=0, 1, $\ldots$, $n-1$ and m_{2k+2}: *"not similar to* x_{k+1}" is the negation of m_{2k+1}. For the fuzzy similarity relation R=(r_{ij}), $r_{ki} \geq r_{kj}$ implies that the degree of x_i being similar to x_k is larger than or equal to that of x_j. Thus each concept in M is a simple concept on X. According to the information provided by the fuzzy similarity relation R=(r_{ij}) like (6.38) shown in Example 6.4, we can establish AFS structure (M, τ, X) as follows. For any x_i, $x_j \in X$, introduce

$\tau(x_i, x_j)$=$\{ m_{2k+1} \mid r_{(k+1)i} \geq r_{(k+1)j}, 0 \leq k \leq n-1\} \cup \{ m_{2l+2} \mid 1 - r_{(l+1)i} \geq 1 - r_{(l+1)i}, 0 \leq l \leq n-1 \}$.

Let σ-algebra over X be S=2^X and m be the measure over S defined by (6.13) for ρ: $X \rightarrow [0, \infty)$, $\rho(x)$=1 for any $x \in X$. There exist the lower and upper fuzzy sets $S_*(\eta)$ and $S^*(\eta)$ in $\mathbb{A}^\rho = (M, \Lambda, \rho, X)$ defined by (6.20) for any fuzzy set $\eta \in \mathscr{F}(X)$. Based on (6.31) and (6.35) which are the algorithm of finding the approximate solutions of (6.20), one can obtain $\overline{S}^\rho_*(\gamma)$ and $\overline{S}^*_\rho(\gamma)$ for any $\gamma \in \mathscr{F}(X)$. Do the following assertions hold? For any $\gamma \in \mathscr{F}(X)$,

$$\sum_{x \in X} \left(\mu_{\overline{S}^*_\rho(\gamma)}(x) - \mu_\gamma(x)\right)^2 \leq \sum_{x \in X} \left(\mu_{S^*(\gamma)}(x) - \mu_\gamma(x)\right)^2,$$

$$\sum_{x \in X} \left(\mu_{\overline{S}^\rho_*(\gamma)}(x) - \mu_\gamma(x)\right)^2 \leq \sum_{x \in X} \left(\mu_{S_*(\gamma)}(x) - \mu_\gamma(x)\right)^2.$$

Are there similar results for the other conventional fuzzy rough sets?

Problem 6.5. The design of feasible algorithms for solving (6.20) by making use of the properties of the AFS algebras and the combinatorial properties of the AFS structures.

Problem 6.6. The lower and upper approximations of an fuzzy concept in an AFS fuzzy approximation space $\mathbb{A}^i$ may have a significant number of terms as shown in some examples. This calls for the development of algorithms leading to the reduction of terms forming the lower and upper approximation.

Problem 6.7. Do any properties of conventional fuzzy rough sets and rough sets hold in the AFS fuzzy approximation space $\mathbb{A}^i$?

Problem 6.8. How does the weight function $\rho(x)$ influence the AFS rough sets?

References

1. Bonikowski, Z.: Algebraic Structures of Rough Sets in Representative Approximation Spaces. Electronic Notes in Theoretical Computer Science 82(4), 1–12 (2003), http://www.elsevier.nl/locate/entcs/volume82.html
2. Bodjanova, S.: Approximation of fuzzy concepts in decision making. Fuzzy Sets and Systems 85, 23–29 (1997)
3. Dubois, D., Prade, H.: Fuzzy Sets and Systems: Theory and Applications. Academic Press, New York (1980)
4. Dubois, D., Prade, H.: Possibility Theory. Plenum Press, New York (1988)
5. Dubois, D., Prade, H.: Putting rough sets and fuzzy sets together. In: Slowinski, R. (ed.) Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory, pp. 203–222. Kluwer, Dordrecht (1992)
6. Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets. Internat. J. General Systems 17(2-3), 191–209 (1990)
7. Dembczynski, K., Pindur, R., Susmag, R.: Generation of Exhaustive Set of Rules within Dominance-based Rough Set Approach. Electronic Notes in Theoretical Computer Science 82(4), 1–12 (2003), http://www.elsevier.nl/locate/entcs/volume82.html
8. Fernandez Salido, J.M., Murakami, S.: Rough Set Analysis of a General Type of Fuzzy Data Using Transitive Aggregations of Fuzzy Similarity Relations. Fuzzy Sets and Systems 139, 635–660 (2003)
9. Farinas del Cerro, L., Prade, H.: Rough sets, twofold fuzzy sets and modal logic-Fuzziness in indiscernibility and partial information. In: di Nola, A., Ventre, A.G.S. (eds.) The Mathematics of Fuzzy Systems, pp. 103–120. Verlag TU V, Rheinland (1986)
10. Fern'andez Salido, J.M., Murakami, S.: On β-precision aggregation. Fuzzy Sets and Systems 139, 15–25 (2003)
11. Greco, S., Matarazzo, B., Slowinski, R.: Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough-set decision rules. European Journal of Operational Research 158, 271–292 (2004)
12. Greco, S., Matarazzo, B., Slowinski, R.: Rough sets theory for multicriteria decision analysis. European Journal of Operational Research 129, 1–47 (2001)
13. Greco, S., Matarazzo, B., Slowinski, R.: Rough approximation of a preference relation by dominance relations. European Journal of Operational Research 117, 63–83 (1999)
14. Greco, S., Matarazzo, B., Slowinski, R.: Rough sets methodology for sorting problems in presence of multiple attributes and criteria. European Journal of Operational Research 138, 247–259 (2002)
15. Graver, J.E., Watkins, M.E.: Combinatorics with Emphasis on the Theory of Graphs. Springer, New York (1977)
16. Geng, L., Chan, C.W.: An Algorithm for Case Generation from a Database. Applied Mathematics Letters 17, 269–274 (2004)
17. Greco, S., Matarazzo, B., Slowinski, R.: Fuzzy similarity relations as a basis for rough approximations. In: Polkowski, L., Skowron, A. (eds.) Proc. 1st Int. Conf. on Rough Sets and Current Trends in Computing, RSTC, Warsaw, Poland, pp. 283–289 (1998)
18. Kim, D., Bang, S.Y.: A Handwritten Numeral Character Classification Using Tolerant Rough Set. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(9), 923–937 (2000)
19. Kim, D.: Data classification based on tolerant rough set. Pattern Recognition 34, 1613–1624 (2001)

20. Kim, K.H.: Boolean Matrix Theory and Applications. Marcel Dekker, New York (1982)
21. Kuncheva, L.I.: Fuzzy rough sets: application to feature extraction. Fuzzy Sets and Systems 51, 147–153 (1992)
22. Halmos, P.R.: Measure theory. Springer, New York (1974)
23. Liu, X.D.: The Fuzzy Theory Based on AFS Algebras and AFS Structure. Journal of Mathematical Analysis and Applications 217, 459–478 (1998)
24. Liu, X.D.: The Topology on AFS Algebra and AFS Structure. Journal of Mathematical Analysis and Applications 217, 479–489 (1998)
25. Liu, X.D.: The Fuzzy Sets and Systems Based on AFS Structure, EI Algebra and EII algebra. Fuzzy Sets and Systems 95, 179–188 (1998)
26. Liu, X.D., Pedrycz, W., Zhang, Q.L.: Axiomatics Fuzzy sets logic. In: The Proceedings of 2003 IEEE International Conference on Fuzzy Systems, vol. 1, pp. 55–60 (2003)
27. Liu, X.D., Chai, T.Y., Wang, W.: Approaches to the Representations and Logic Operations for Fuzzy Concepts in the Framework of Axiomatic Fuzzy Set Theory I, II. Information Sciences 177, 1007–1026, 1027-1045 (2007)
28. Liu, X.D., Wang, W., Chai, T.Y.: The Fuzzy Clustering Analysis Based on AFS Theory. IEEE Transactions on Systems, Man and Cybernetics 35(5), 1013–1027 (2005)
29. Liu, X.D., Pedrycz, W.: The Development of Fuzzy Decision Trees in the Framework of Axiomatic Fuzzy Set Logic. Applied Soft Computing 7, 325–342 (2007)
30. Liu, X.D., Zhu, K.J., Huang, H.Z.: The Representations of Fuzzy Concepts Based on the Fuzzy Matrix Theory and the AFS Theory. In: IEEE International Symposium on Intelligent Control, Houston, Texas, October 5-8, 2003, pp. 1006–1011 (2003)
31. Liu, X.D.: The Structure of Fuzzy Matrices. Journal of Fuzzy Mathematics 2, 311–325 (1994)
32. Liu, X.D., Pedrycz, W., Chai, T.Y., Song, M.L.: The Development of Fuzzy Rough Sets with the Use of Structures and Algebras of Axiomatic Fuzzy Sets. IEEE Transactions on Knowledge and Data Engineering 21(3), 443–462 (2009)
33. Lesniewski, S.: Foundations of the general theory of sets. Polish Scientific Circle, Moscow (1916)
34. Merz, C.J., Murphy, P.M.: UCI Repository for Machine Learning Data-Bases. Dept. of Information and Computer Science, University of California, Irvine, CA (1996), http://www.ics.uci.edu/~mlearn/MLRepository.html
35. Nakamura, A.: Application of fuzzy-rough classi/cations to logics. In: Slowinski, R. (ed.) Intelligent Decision Support: Handbook ofApplications and Advances ofthe Rough Sets Theory, pp. 233–250. Kluwer Academic Publishers, Dordrecht (1992)
36. Nakamura, A., Gao, J.M.: A logic for fuzzy data analysis. Fuzzy Sets and Systems 39, 127–132 (1992)
37. Nguyen, H.S.: On the Decision Table with Maximal Number of Reducts. Electronic Notes in Theoretical Computer Science 82(4), 8 pages (2003), http://www.elsevier.nl/locate/entcs/volume82.html
38. Pawlak, Z.: Rough sets. Internat. J. Comput. Inform. Sci. 11(5), 341–356 (1982)
39. Pawlak, Z.: Rough Probability. Bull. Polish Acad. Sci., Math. 32, 607–612 (1984)
40. Pawlak, Z.: Hard and soft sets. In: Ziarko, W.P. (ed.) Rough Sets, Fuzzy Sets and Knowledge Discovery, pp. 130–135. Springer, London (1994)
41. Pawlak, Z., Skowron, A.: Rough sets: Some Extensions. Information Sciences 177, 28–40 (2007)
42. Pawlak, Z., Skowron, A.: Rough Sets and Boolean Reasoning. Information Sciences 177, 41–70 (2007)
43. Polkowski, L., Skowron, A.: Rough Mereology: A New Paradigm for Approximate Reasoning. International Journal of Approximate Reasoning 15, 333–365 (1996)

44. Pawlak, Z., Skowron, A.: Rudiments of Rough Sets. Information Sciences 177, 3–27 (2007)
45. Pal, S.K., Mitra, P.: Case Generation Using Rough Sets with Fuzzy Representation. IEEE Transactions on Knowledge and Data Engineering 16(3), 292–300 (2004)
46. Radzikowska, A.M., Kerre, E.E.: A comparative study of fuzzy rough sets. Fuzzy Sets and Systems 126, 137–155 (2002)
47. Slowinski, R., Vanderpooten, D.: Similarity relations as a basis for rough approximations. In: Wang, P.P. (ed.) Advances in Machine Intelligence and Soft Computing, Bookwrights, Raleigh, NC, pp. 17–33 (1997)
48. Slowinski, R., Vanderpooten, D.: A Generalized Definition of Rough Approximations Based on Similarity. IEEE Transactions on Knowledge and Data Engineering 12(2), 331–336 (2000)
49. Skowron, A., Rauszer, C.: The Discernibility Matrices and Functions in Information Systems. In: Slowinski, R. (ed.) Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory, pp. 331–362. Kluwer Academic, Dordrecht (1992)
50. Wang, Q.H., Li, J.R.: A rough set-based fault ranking prototype system for fault diagnosis. Engineering Applications of Artificial Intelligence 17, 909–917 (2004)
51. Wang, G.J.: Theory of topological molecular lattices. Fuzzy Sets and Systems 47, 351–376 (1992)
52. Yao, Y.: Combination ofrough and fuzzy sets based on α-level sets. In: Lin, T.Y., Cercone, N. (eds.) Rough sets and Data Mining, pp. 301–321. Kluwer, Dordrecht (1997)
53. Zadeh, L.A.: Similarity relations and fuzzy orderings. Inform. Sci. 3, 177–200 (1971)
54. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1, 3–28 (1978)
55. Zadeh, L.A.: Fuzzy logic=computing with words. IEEE Transactions on Fuzzy Systems 4(2), 103–111 (1996)
56. Ziarko, W.: Variable precision rough set model. J. Comput. System Sci. 46, 39–59 (1993)
57. Zhong, N., Dong, J.Z., Ohsuga, S.: Meningitis data mining by cooperatively using GDT-RS and RSBR. Pattern Recognition Letters 24, 887–894 (2003)
58. Zaras, K.: Rough approximation of a preference relation by a multi-attribute stochastic dominance for determinist and stochastic evaluation problems. European Journal of Operational Research 130, 305–314 (2001)

Chapter 7
AFS Topology and Its Applications

In this chapter, first we construct some topologies on the AFS structures, discuss the topological molecular lattice structures on EI, $^*EI$, EII, $^*EII$ algebras, and elaborate on the main relations between these topological structures. Second, we apply the topology derived by a family of fuzzy concepts in EM, where M is a set of simple concepts, to analyze the relations among the fuzzy concepts. Thirdly, we propose the differential degrees and fuzzy similarity relations based on the topological molecular lattices generated by the fuzzy concepts on some features. Furthermore, the fuzzy clustering problems are explored using the proposed differential degrees and fuzzy similarity relations. Compared with other fuzzy clustering algorithms such as the Fuzzy C-Means and k-nearest-neighbor fuzzy clustering algorithms, the proposed fuzzy clustering algorithm can be applied to data sets with mixed feature variables such as numeric, Boolean, linguistic rating scale, sub-preference relations, and even descriptors associated with human intuition. Finally, some illustrative examples show that the proposed differential degrees are very effective in pattern recognition problems whose data sets do not form a subset of a metric space such as the Eculidean one. This approach offers a promising avenue that could be helpful in understanding mechanisms of human recognition.

7.1 Topology on AFS Structures and Topological Molecular Lattice on $^*EI^n$ Algebras

In this section, we first study the topological molecular lattice on the $^*EI$ algebra over a set M, i.e.,($^*EM,\vee,\wedge$), in which the lattice operators $\vee$, $\wedge$ are defined as follows: for any $\sum_{i\in I}A_i$, $\sum_{j\in J}B_j \in EM$,

$$\sum_{i\in I}A_i \vee \sum_{j\in J}B_j = \sum_{i\in I, j\in J} A_i \cup B_j, \tag{7.1}$$

$$\sum_{i\in I}A_i \wedge \sum_{j\in J}B_j = \sum_{i\in I}A_i + \sum_{j\in J}B_j. \tag{7.2}$$

M is the maximum element of the lattice $^*EM$ and $\varnothing$ is the minimum element of this lattice. That is, the above lattice $^*EM$ is a dual lattice of EM. In the lattice $^*EM$,

X. Liu and W. Pedrycz: Axiomatic Fuzzy Set Theory and Its Applications, STUDFUZZ 244, pp. 269–301.
springerlink.com

for $\sum_{i\in I}A_i, \sum_{j\in J}B_j \in EM$, $\sum_{i\in I}A_i \leq \sum_{j\in J}B_j$ if and only if for any B_j $(j \in J)$ there exists A_k $(i \in I)$ such that $B_j \supseteq A_k$ (refer to Theorem 5.24). Secondly, we study the topology on the universe of discourse X induced by the topological molecular lattice of some fuzzy concepts in EM. Finally the topological molecular lattice on the $^*EI^2$ algebra over the sets X, M, i.e.,($^*EXM, \vee, \wedge$), in which the lattice operators $\vee$, $\wedge$ are defined as follows: for any $\sum_{i\in I}a_iA_i$, $\sum_{j\in J}b_jB_j \in EXM$,

$$\sum_{i\in I}a_iA_i \vee \sum_{j\in J}b_jB_j = \sum_{i\in I, j\in J}a_i\cap b_jA_i\cup B_j, \tag{7.3}$$

$$\sum_{i\in I}a_iA_i \wedge \sum_{j\in J}b_jB_j = \sum_{i\in I}a_iA_i + \sum_{j\in J}b_jB_j. \tag{7.4}$$

$\varnothing M$ is the maximum element of the lattice $^*EM$ and $X\varnothing$ is the minimum element of the lattice $^*EM$. That is, the lattice $^*EXM$ is a dual lattice of EXM. In the lattice $^*EXM$, for $\sum_{i\in I}a_iA_i, \sum_{j\in J}b_jB_j \in EM$, $\sum_{i\in I}a_iA_i \leq \sum_{j\in J}b_jB_j$ if and only if for any b_jB_j $(j \in J)$ there exists a_kA_k $(i \in I)$ such that $B_j \supseteq A_k$ and $a_k \supseteq b_j$ (refer to Theorem 5.1).

Lemma 7.1. *Let M be a set and EM be the $^*EI$ algebra over M. For $A \subseteq M$, $\sum_{i\in I}A_i$, $\sum_{j\in J}B_j \in EM$, the following assertions hold:*

(1) $A \geq \sum_{i\in I}A_i$ and $A \geq \sum_{j\in J}B_j \Leftrightarrow A \geq \sum_{i\in I}A_i \vee \sum_{j\in J}B_j$;
(2) $A \geq \sum_{i\in I}A_i$ or $A \geq \sum_{j\in J}B_j \Leftrightarrow A \geq \sum_{i\in I}A_i \wedge \sum_{j\in J}B_j$.

Its proof is left as an exercise.

Definition 7.1. Let M be a set and $(^*EM, \vee, \wedge)$ be the $^*EI$ algebra over M defined by (7.1) and (7.2). Let $\eta \subseteq {^*EM}$. If $\varnothing, M \in \eta$ and η is closed under finite unions (i.e., $\vee$) and arbitrary intersections (i.e., $\wedge$), then η is called a *topological molecular lattice on the lattice* $^*EM$, denoted as $(^*EM, \eta)$. Let η be a topological molecular lattice on the lattice $^*EM$. If for any $\sum_{i\in I}A_i \in \eta$, $A_i \in \eta$ for any $i \in I$, then η is called an *elementary topological molecular lattice on the lattice* $^*EM$.

It is easy proved that if η is a topological molecular on the lattice $^*EM$ and η is a dual idea of the lattice $^*EM$, then η is an elementary topological molecular lattice on the lattice $^*EM$. In what follows, we apply the elementary topological molecular lattice on the lattice $^*EM$ to induce some topological structures on X via the AFS structure (M, τ, X) of a data. Thus the pattern recognition problem can by explored in the setting of these topological structures on X.

Definition 7.2. Let X and M be sets and (M, τ, X) be an AFS structure. Let $(^*EM, \eta)$ be a topological molecular lattice on $^*EI$ algebra over M. For any $x \in X$, $\sum_{i\in I}A_i \in \eta \subseteq {^*EM}$, the set $N^{\tau}_{\sum_{i\in I}A_i}(x) \subseteq X$ is defined as follows.

$$N^{\tau}_{\sum_{i\in I}A_i}(x) = \left\{ y \in X \ \mid \ \tau(x,y) \geq \sum_{i\in I}A_i \right\}, \tag{7.5}$$

and it is called the *neighborhood of x induced by the fuzzy concept* $\sum_{i\in I} A_i$ in the AFS structure (M,τ,X). The set $N_\eta^\tau(x) \subseteq 2^X$ is defined as follows.

$$N_\eta^\tau(x) = \left\{ N_{\sum_{i\in I} A_i}(x) \;\middle|\; \sum_{i\in I} A_i \in \eta \right\}, \tag{7.6}$$

and it is called the *neighborhood of x induced by the topological molecular lattice* η in the AFS structure (M,τ,X).

Since $\tau(x,y) \subseteq M$, hence $\tau(x,y)$ is an element in EM and $\tau(x,y) \geq \sum_{i\in I} A_i$ in (7.5) is well-defined.

Definition 7.3. Let X and M be sets and (M,τ,X) be an AFS structure. (M,τ,X) is called a *strong relative AFS structure* if $\forall (x,y) \in X\times X, \tau(x,y)\cup\tau(y,x) = M$.

Since in a strong relative AFS structure (M,τ,X), $\forall x \in X$, $\tau(x,x) = M$, hence $\forall x \in X, \forall m \in M$, x belongs to the simple concept m to some extent.

Proposition 7.1. *Let X and M be sets and (M,τ,X) be a strong relative AFS structure. Let η be a topological molecular lattice on $^*EI$ algebra over M. For any $\sum_{i\in I} A_i$, $\sum_{j\in J} B_j \in EM$, the following assertions hold: for any $x \in X$*

(1) If $\sum_{i\in I} A_i \geq \sum_{j\in J} B_j$ in $^*EM$, then $N^\tau_{\sum_{i\in I} A_i}(x) \subseteq N^\tau_{\sum_{j\in J} B_j}(x)$;
(2) $N^\tau_{\sum_{i\in I} A_i}(x) \cap N^\tau_{\sum_{j\in J} B_j}(x) = N^\tau_{\sum_{i\in I} A_i \vee \sum_{j\in J} B_j}(x)$;
(3) $N^\tau_{\sum_{i\in I} A_i}(x) \cup N^\tau_{\sum_{j\in J} B_j}(x) = N^\tau_{\sum_{i\in I} A_i \wedge \sum_{j\in J} B_j}(x)$.

Proof. (1) Let $y \in N^\tau_{\sum_{i\in I} A_i}(x)$. Then there exists $A_k, k \in I$ such that $\tau(x,y) \supseteq A_k$. On the other hand, since $\sum_{i\in I} A_i \geq \sum_{j\in J} B_j$, hence for A_k there exists $B_j, j \in J$ such that $\tau(x,y) \supseteq A_k \supseteq B_j$. This implies that $y \in N^\tau_{\sum_{j\in J} B_j}(x)$. It follows that $N^\tau_{\sum_{i\in I} A_i}(x) \subseteq N^\tau_{\sum_{j\in J} B_j}(x)$.

(2) For any $y \in N^\tau_{\sum_{i\in I} A_i}(x) \cap N^\tau_{\sum_{j\in J} B_j}(x)$, in virtue of Lemma 7.1, we have

$$\begin{aligned}
y \in N^\tau_{\sum_{i\in I} A_i}(x) \cap N^\tau_{\sum_{j\in J} B_j}(x) &\Leftrightarrow y \in N^\tau_{\sum_{i\in I} A_i}(x) \;\; and \;\; y \in N^\tau_{\sum_{j\in J} B_j}(x) \\
&\Leftrightarrow \tau(x,y) \geq \sum_{i\in I} A_i \;\; and \;\; \tau(x,y) \geq \sum_{j\in J} B_j \\
&\Leftrightarrow \tau(x,y) \geq \sum_{i\in I, j\in J} A_i \cup B_j \\
&\Leftrightarrow \tau(x,y) \geq \left(\sum_{i\in I} A_i\right) \vee \left(\sum_{j\in J} B_j\right) \\
&\Leftrightarrow y \in N^\tau_{\sum_{i\in I} A_i \vee \sum_{j\in J} B_j}(x).
\end{aligned}$$

So we have showed that (2) holds.

(3) For any $y \in N^{\tau}_{\sum_{i\in I} A_i}(x) \cup N^{\tau}_{\sum_{j\in J} B_j}(x)$, from Lemma 7.1, we have

$$\begin{aligned} y \in N^{\tau}_{\sum_{i\in I} A_i}(x) \cup N^{\tau}_{\sum_{j\in J} B_j}(x) &\Leftrightarrow y \in N^{\tau}_{\sum_{i\in I} A_i}(x) \;\; or \;\; y \in N^{\tau}_{\sum_{j\in J} B_j}(x) \\ &\Leftrightarrow \tau(x,y) \geq \sum_{i\in I} A_i \;\; or \;\; \tau(x,y) \geq \sum_{j\in J} B_j \\ &\Leftrightarrow \tau(x,y) \geq \sum_{i\in I} A_i + \sum_{j\in J} B_j \\ &\Leftrightarrow \tau(x,y) \geq \left(\sum_{i\in I} A_i\right) \wedge \left(\sum_{j\in J} B_j\right) \\ &\Leftrightarrow y \in N^{\tau}_{\sum_{i\in I} A_i \wedge \sum_{j\in J} B_j}(x). \end{aligned}$$

This implies that (3) is satisfied. □

Theorem 7.1. *Let X and M be sets and (M,τ,X) be a strong relative AFS structure. Let η be a topological molecular lattice on the lattice $^*EM$. If η is an elementary topological molecular lattice on the lattice $^*EM$ and we define*

$$\mathscr{B}_\eta = \left\{ N^{\tau}_{\sum_{i\in I} A_i}(x) \;\middle|\; x \in X, \; \sum_{i\in I} A_i \in \eta \right\},$$

then $\mathscr{B}_\eta$ is a base for some topology of X.

Proof. Firstly, because (M,τ,X) is a strong relative AFS structure, for any $x \in X$, $\tau(x,x) = M$. M is the maximum element of the lattice $^*EM$. This implies that for any $\sum_{i\in I} A_i \in \eta$, $\tau(x,x) \geq \sum_{i\in I} A_i$ so that $x \in N^{\tau}_{\sum_{i\in I} A_i}(x)$ and $X = \bigcup_{\beta\in\mathscr{B}_\eta} \beta$.

Secondly, suppose $x \in X, U, V \in \mathscr{B}_\eta$, and $x \in U \cap V$. We will prove there exists $W \in \mathscr{B}_\eta$ such that $x \in W \subseteq U \cap V$. By the hypothesis, we know there exists $\sum_{i\in I} A_i, \sum_{j\in J} B_j \in \eta$ such that $U = N^{\tau}_{\sum_{i\in I} A_i}(u)$, $V = N^{\tau}_{\sum_{j\in J} B_j}(v)$ for some $u,v \in X$ and $\exists l \in I, \exists k \in J$, $\tau(u,x) \supseteq A_l$ and $\tau(v,x) \supseteq B_k$. Since (M,τ,X) is a strong relative, hence $x \in N^{\tau}_{A_l}(x)$ and $x \in N^{\tau}_{B_k}(x)$. For any $y \in N^{\tau}_{A_l}(x)$, i.e., $\tau(x,y) \supseteq A_l$, by Definition 4.5, we have $\tau(u,y) \supseteq \tau(u,x) \cap \tau(x,y) \supseteq A_l$, that is $y \in U$. It follows $N^{\tau}_{A_l}(x) \subseteq U$. For the same reason, $N^{\tau}_{B_k}(x) \subseteq V$. By Proposition 7.1, we have

$$x \in N^{\tau}_{A_l}(x) \cap N^{\tau}_{B_k}(x) = N^{\tau}_{A_l \vee B_k}(x) \subseteq U \cap V.$$

Since η is an elementary topological molecular lattice on the lattice $^*EM$, hence $A_l, B_k \in \eta$ and we have $x \in W = N^{\tau}_{A_l \vee B_k}(x) \in \mathscr{B}_\eta$ such that $W \subseteq U \cap V$. Now by Theorem 1.21, $\mathscr{B}_\eta$ is a base for some topology on X. □

The topological space $(X, \mathscr{T}_\eta)$, in which $\mathscr{B}_\eta$ is the base for $\mathscr{T}_\eta$, is called the *topology of X induced by the topological molecular lattice η*.

Theorem 7.2. *Let X and M be sets and (M,τ,X) be a strong relative AFS structure. Let η be a topological molecular lattice on the lattice $^*EM$ and*

$$\mathscr{L}_\eta = \left\{ \sum_{i\in I} a_i A_i \in EXM \mid \sum_{i\in I} A_i \in \eta, a_i \in \mathscr{T}_\eta \ for\ any\ i \in I, \right.$$
$$\left. I\ is\ any\ non-empty\ indexing\ set \right\}. \tag{7.7}$$

Then $\mathscr{L}_\eta$ is a topological molecular lattice on the lattice ${}^*EXM$. It is called the ${}^*EI^2$ topological molecular lattice induced by the ${}^*EI$ the topological molecular lattice η.

Proof. For any finite integer n, let $\lambda_j = \sum_{i\in I_j} a_{ij} A_{ij} \in EXM$, $j = 1,2,...,n$. Because for any $f \in \prod_{1\le j\le n} I_j$, $\bigcap_{1\le j\le n} a_{f(j)j} \in \mathscr{T}_\eta$ and

$$\bigvee_{1\le j\le n} \left(\sum_{i\in I_j} A_{ij} \right) = \sum_{f\in\prod_{1\le j\le n} I_j} \bigcup_{1\le j\le n} A_{f(j)j} \in \eta,$$

then we have

$$\bigvee_{1\le j\le n} \lambda_j = \sum_{f\in\prod_{1\le j\le n} I_j} \left(\bigcap_{1\le j\le n} a_{f(j)j} \bigcup_{1\le j\le n} A_{f(j)j} \right) \in \mathscr{L}_\eta.$$

This implies that $\mathscr{L}_\eta$ is closed under finite unions (i.e.,$\vee$). It is obvious that $\wedge$ is closed under arbitrary intersection. Therefore $({}^*EXM, \mathscr{L}_\eta)$ is a ${}^*EI^2$ topological molecular lattice on the lattice ${}^*EXM$. □

It is clear that $\mathscr{T}_\eta$ the topology on X is determined based on the distribution of raw data and the chosen set of fuzzy concepts $\eta \subseteq EM$ and it is an abstract geometry relation among the objects in X under the fuzzy concepts under consideration, i.e., η. What are the interpretations of the special topological structures on X obtained from given database? What are the topological structures associated with the essential nature of database? All these questions are related to the metric space of the topology. With a metric in the topological space on X, it will be possible to handle pattern recognition problems for the databases with various data types.

Let X be a set and M be a set of simple concepts on X. Let (M,τ,X) be an AFS structure and S be the σ-algebra over X. In real world applications, it is obvious that only some fuzzy concepts in EM are related with the problem under consideration. Let these fuzzy concepts form the set $\Lambda \subseteq EM$. Let η be the topological molecular lattice generated by Λ and $(X,\mathscr{T}_\eta)$ be the topology induced by η. Let S be the σ-algebra generated by $\mathscr{T}_\eta$, i.e., the Borel set corresponding to the topological space $(X,\mathscr{T}_\eta)$ and (S,m) be a measure space. For the fuzzy concept $\sum_{i\in I} A_i \in EM$, if for any $x \in X$, any $i \in I$, $A_i^\tau(x) \in S$, then $\sum_{i\in I} A_i$ is called a *measurable fuzzy concept under the σ-algebra S*. Thus the membership function of each measurable fuzzy concept in EM can be obtained by the norm of $E^\# I$ algebra via (5.13), (5.24) and (S,m).

Theorem 7.3. *Let X and M be sets. Let (M,τ,X) be a strong relative AFS structure and η be an elementary topological molecular lattice on the lattice ${}^*EM$. Let η be a topological molecular lattice on the lattice ${}^*EM$ and the topological space $(X,\mathscr{T}_\eta)$*

be the topology induced by η. *Let* S *be the* σ*-algebra generated by* $\mathscr{T}_\eta$ *and* $\mathscr{L}_\eta$ *be the* $^*EI^2$ *topological molecular lattice on* $^*EXM$ *induced by* η. *Then the following assertions hold.*

(1) For any fuzzy concept $\sum_{i\in I} A_i \in \eta$, $\sum_{i\in I} A_i$ *is a measurable concept under* S *;*
(2) For each fuzzy concept $\gamma = \sum_{i\in I} A_i \in \eta$, *let* $\gamma : X \to EXM$ *be the* EI^2 *algebra representation membership degrees defined by (5.10) as follows: for any* $x \in X$,

$$\gamma(x) = \sum_{i\in I} A_i^\tau(x) A_i \in EXM. \tag{7.8}$$

Let D *be a directed set and* $\delta : D \to X$ *be a net (i.e.,* $\{\delta(d) \mid d \in D\}$*). If* δ *is converged to* $x_0 \in X$ *under topology* $\mathscr{T}_\eta$, *then the net of the composition* $\gamma \cdot \delta : D \to EXM$ *(i.e.,*$\{\gamma(\delta(d)) \mid d \in D\}$*) converges to* $\gamma(x_0) = \sum_{i\in I} A_i^\tau(x_0) A_i$ *under the topological molecular lattice* $\mathscr{L}_\eta$. *That is, the membership function of any fuzzy concept in EM defined by (7.8) is a continuous function from the topological space* $(X, \mathscr{T}_\eta)$ *to the topological molecular lattice* $(^*EXM, \mathscr{L}_\eta)$.

Proof. (1) For any $\sum_{i\in I} A_i \in \eta$, since $A_i \geq \sum_{i\in I} A_i$ for all $i \in I$ and η is an elementary topological molecular lattice on the lattice $^*EM$, hence $A_i \in \eta$ for all $i \in I$ and

$$A_i^\tau(x) = N_{A_i}^\tau(x) \in \mathscr{T}_\eta \Rightarrow A_i^\tau(x) \in S, \quad for\ any\ x \in X\ and\ any\ i \in I.$$

Therefore $\sum_{i\in I} A_i$ is a measurable concept under S.

(2) Suppose $\sum_{j\in J} p_j P_j \in \mathscr{L}_\eta$ and $\sum_{j\in J} p_j P_j$ is a R-neighborhood of $\sum_{i\in I} A_i^\tau(x_0) A_i$, i.e., $\sum_{i\in I} A_i^\tau(x_0) A_i \nleq \sum_{j\in J} p_j P_j$. This implies that there exists $p_l P_l$ $(l \in J)$ such that for any $i \in I$, either $A_i^\tau(x_0) \not\supseteq p_l$ or $P_l \not\supseteq A_i$. First, assume $\forall k \in I$, $P_l \not\supseteq A_k$. It follows, for any $d \in D$, $\sum_{i\in I} A_i^\tau(\delta(d)) A_i \nleq \sum_{j\in J} p_j P_j$.

Second, assume that $k \in I$, $A_k^\tau(x_0) \not\supseteq p_l$. Since $x_0 \in A_k^\tau(x_0) \in \mathscr{T}_\eta$ and δ is converged to $x_0 \in X$ under $\mathscr{T}_\eta$, hence the exists $N \in D$ such that for any $d \in D$, $d \geq N$, $\delta(d) \in A_k^\tau(x_0) \not\supseteq p_l$. For any $y \in A_k^\tau(\delta(d))$, i.e., $\tau(\delta(d), y) \supseteq A_k$, since $\delta(d) \in A_k^\tau(x_0)$, i.e., $\tau(x_0, \delta(d)) \supseteq A_k$ and τ is an AFS structure, hence we have

$$\tau(x_0, y) \supseteq \tau(x_0, \delta(d)) \cap \tau(\delta(d), y) \supseteq A_k \Rightarrow y \in A_k^\tau(x_0) \Rightarrow A_k^\tau(x_0) \supseteq A_k^\tau(\delta(d)).$$

This implies that for $i \in I$ if $A_i^\tau(x_0) \not\supseteq p_l$, then exists $N \in D$ such that for any $d \in D, d \geq N$, $A_i^\tau(\delta(d)) \not\supseteq p_l$. Thus for any R-neighborhood of $\sum_{i\in I} A_i^\tau(x_0) A_i$, $\upsilon \in \mathscr{L}_\eta$, there exists $N \in D$ such that for any $d \in D, d \geq N$,

$$\sum_{i\in I} A_i^\tau(\delta(d)) A_i \nleq \upsilon.$$

Therefore the net $\gamma \cdot \delta$ is converged to $\sum_{i\in I} A_i^\tau(x_0) A_i$ under the topological lattice $\mathscr{L}_\eta$. □

In a strong relative AFS structure (M, τ, X), $\forall x \in X$, $\tau(x, x) = M$, i.e. $\forall x \in X, \forall m \in M$, x belongs to the simple concept m at some extent. It is too strict to be exploited in the setting of real world applications. In order to offer an abstract description of the similar relation between the objects in X concerning some given concepts, Definition 7.2 should be modified as follows.

Definition 7.4. Let X and M be sets and (M,τ,X) be an AFS structure. Let η be a topological molecular lattice on the lattice $^*EM$. For any $x \in X$, $\sum_{i\in I} A_i \in \eta$, the set $N^{\triangle\tau}_{\sum_{i\in I} A_i}(x) \subseteq X$ is defined as follows.

$$N^{\triangle\tau}_{\sum_{i\in I} A_i}(x) = \left\{ y \in X \mid \tau(x,y) \cap \tau(y,y) \geq \sum_{i\in I} A_i \right\}, \tag{7.9}$$

and it is called the *limited neighborhood of x induced by the fuzzy concept* $\sum_{i\in I} A_i \in \eta$, if $N^{\triangle\tau}_{\sum_{i\in I} A_i}(x) \neq \varnothing$. The set $N^{\triangle\tau}_{\eta}(x) \subseteq 2^X$ is defined as follows.

$$N^{\triangle\tau}_{\eta}(x) = \left\{ N_{\sum_{i\in I} A_i}(x) \neq \varnothing \mid \sum_{i\in I} A_i \in \eta \right\},$$

and it is called the *limited neighborhood of x induced by the topological molecular lattice η.*

By the definition of the AFS structure (refer to Definition 4.5), we know that for any $x,y \in X$,

$$\tau(x,x) \supseteq \tau(x,y) \supseteq \tau(x,y) \cap \tau(y,y).$$

Therefore $N^{\triangle\tau}_{\sum_{i\in I} A_i}(x) \subseteq N^{\tau}_{\sum_{i\in I} A_i}(x)$ for any $x \in X$, any $\sum_{i\in I} A_i \in \eta$ and

$$N^{\triangle\tau}_{\sum_{i\in I} A_i}(x) \neq \varnothing \Leftrightarrow x \in N^{\tau}_{\sum_{i\in I} A_i}(x).$$

Proposition 7.2. *Let X and M be sets and (M,τ,X) be an AFS structure. Let η be a topological molecular lattice on the lattice $^*EM$. For any $x \in X$, $\sum_{i\in I} A_i, \sum_{j\in J} B_j \in EM$, the following assertions hold.*

(1) If $\sum_{i\in I} A_i \geq \sum_{j\in J} B_j$ in the lattice $^*EM$, then $N^{\triangle\tau}_{\sum_{i\in I} A_i}(x) \subseteq N^{\triangle\tau}_{\sum_{j\in J} B_j}(x)$ for any $x \in X$;
(2) $N^{\triangle\tau}_{\sum_{i\in I} A_i}(x) \cap N^{\triangle\tau}_{\sum_{j\in J} B_j}(x) = N^{\triangle\tau}_{\sum_{i\in I} A_i \vee \sum_{j\in J} B_j}(x)$ for any $x \in X$;
(3) $N^{\triangle\tau}_{\sum_{i\in I} A_i}(x) \cup N^{\triangle\tau}_{\sum_{j\in J} B_j}(x) = N^{\triangle\tau}_{\sum_{i\in I} A_i \wedge \sum_{j\in J} B_j}(x)$ for any $x \in X$.

Proof. (1) Suppose $y \in N^{\triangle\tau}_{\sum_{i\in I} A_i}(x)$, $x \in X$. By (7.9), we know that there exists A_k, $k \in I$ such that $\tau(x,y) \cap \tau(y,y) \supseteq A_k$. Since $\sum_{i\in I} A_i \geq \sum_{j\in J} B_j$, then for A_k, there exists B_l, $l \in J$ such that

$$\tau(x,y) \cap \tau(y,y) \supseteq A_k \supseteq B_l \Rightarrow \tau(x,y) \cap \tau(y,y) \geq \textstyle\sum_{j\in J} B_j.$$

This implies that $y \in N^{\triangle\tau}_{\sum_{j\in J} B_j}(x)$. It follows $N^{\triangle\tau}_{\sum_{i\in I} A_i}(x) \subseteq N^{\triangle\tau}_{\sum_{j\in J} B_j}(x)$.

(2) For any $y \in N^{\triangle\tau}_{\sum_{i\in I} A_i}(x) \cap N^{\triangle\tau}_{\sum_{j\in J} B_j}(x)$,

$$\begin{aligned} y \in N^{\triangle\tau}_{\sum_{i\in I} A_i}(x) \cap N^{\triangle\tau}_{\sum_{j\in J} B_j}(x) &\Leftrightarrow y \in N^{\triangle\tau}_{\sum_{i\in I} A_i}(x) \ \ and \ \ y \in N^{\triangle\tau}_{\sum_{j\in J} B_j}(x) \\ &\Leftrightarrow \tau(x,y)\cap\tau(y,y) \geq \sum_{i\in I} A_i \ and \ \tau(x,y)\cap\tau(y,y) \geq \sum_{j\in J} B_j \\ &\Leftrightarrow \ \tau(x,y)\cap\tau(y,y) \geq \sum_{i\in I} A_i \vee \sum_{j\in J} B_j \quad (by\ Lemma\ 7.1) \\ &\Leftrightarrow y \in N^{\triangle\tau}_{\sum_{i\in I} A_i \vee \sum_{j\in J} B_j}(x). \end{aligned}$$

Therefore $N^{\triangle\tau}_{\sum_{i\in I} A_i}(x) \cap N^{\triangle\tau}_{\sum_{j\in J} B_j}(x) = N^{\triangle\tau}_{\sum_{i\in I} A_i \vee \sum_{j\in J} B_j}(x)$.

(3) For any $y \in N^{\triangle\tau}_{\sum_{i\in I} A_i}(x) \cup N^{\triangle\tau}_{\sum_{j\in J} B_j}(x)$,

$$\begin{aligned} y \in N^{\triangle\tau}_{\sum_{i\in I} A_i}(x) \cup N^{\triangle\tau}_{\sum_{j\in J} B_j}(x) &\Leftrightarrow y \in N^{\triangle\tau}_{\sum_{i\in I} A_i}(x) \ \ or \ \ y \in N^{\triangle\tau}_{\sum_{j\in J} B_j}(x) \\ &\Leftrightarrow \tau(x,y)\cap\tau(y,y) \geq \sum_{i\in I} A_i \ \ or \ \ \tau(x,y)\cap\tau(y,y) \geq \sum_{j\in J} B_j \\ &\Leftrightarrow \tau(x,y)\cap\tau(y,y) \geq \sum_{i\in I} A_i \wedge \sum_{j\in J} B_j \quad (by\ Lemma\ 7.1) \\ &\Leftrightarrow y \in N^{\triangle\tau}_{\sum_{i\in I} A_i \wedge \sum_{j\in J} B_j}(x). \end{aligned}$$

Subsequently (3) is satisfied. □

Theorem 7.4. *Let X and M be sets and (M,τ,X) be an AFS structure. Let η be a topological molecular lattice on the lattice $^*EM$. If η is an elementary topological molecular lattice on the lattice $^*EM$ and $\mathscr{B}^{\triangle}_{\eta}$ is defined as follows*

$$\mathscr{B}^{\triangle}_{\eta} = \{N_{\sum_{i\in I} A_i}(x) \ | \ x \in X, \sum_{i\in I} A_i \in \eta\}, \tag{7.10}$$

then $\mathscr{B}^{\triangle}_{\eta}$ is a base for some topology of X.

Proof. Firstly, for any $x \in X$, since $\varnothing \in \eta$, hence $\tau(x,x) \geq \varnothing$, i.e., $x \in N^{\triangle\tau}_{\varnothing}(x)$. This implies that $X = \bigcup_{N\in\mathscr{B}^{\triangle}_{\eta}} N$. Secondly, suppose $x \in X$, $U, V \in \mathscr{B}^{\triangle}_{\eta}$, and $x \in U \cap V$. We will prove that there exists $W \in \mathscr{B}^{\triangle}_{\eta}$ such that $x \in W \subseteq U \cap V$. By (7.10), we know that there exists $\sum_{i\in I} A$, $\sum_{j\in J} B_j \in \eta$, $u, v \in X$ such that there $U = N^{\triangle\tau}_{\sum_{i\in I} A_i}(u)$, $V = N^{\triangle\tau}_{\sum_{j\in J} B_j}(v)$. That is, $\exists l \in I$, $\exists k \in J$, $\tau(u,x) \cap \tau(x,x) \supseteq A_l$ and $\tau(v,x) \cap \tau(x,x) \supseteq B_k$. By $\tau(u,x) \cap \tau(x,x) \subseteq \tau(x,x)$ and $\tau(v,x) \cap \tau(x,x) \subseteq \tau(x,x)$, we have $x \in N^{\triangle\tau}_{A_l}(x)$ and $x \in N^{\triangle\tau}_{B_k}(x)$. For any $y \in N^{\triangle\tau}_{A_l}(x)$, i.e., $\tau(x,y) \cap \tau(y,y) \supseteq A_l$, by AX1 and AX2 in Definition 4.5, we have $\tau(x,y) \subseteq \tau(x,x)$ and $\tau(u,x) \cap \tau(x,y) \subseteq \tau(u,y)$. It follows

$$\tau(u,y) \cap \tau(y,y) \supseteq \tau(u,x) \cap \tau(x,x) \cap \tau(x,y) \cap \tau(y,y) \supseteq A_l.$$

This fact implies that $\tau(u,y) \cap \tau(y,y) \geq \sum_{i \in I} A$ and $y \in N^{\triangle\tau}_{\sum_{i\in I} A_i}(u)$. Thus we have $N^{\triangle\tau}_{A_l}(x) \subseteq U$. Similarly, we can prove $N^{\triangle\tau}_{B_k}(x) \subseteq V$. Since η is an elementary topological molecular lattice on the lattice ${}^{*}EM$, hence A_l, $B_k \in \eta$, and $A_l \cup B_k = A_l \vee B_k \in \eta$. In virtue of Proposition 7.2, one has $W = N^{\triangle\tau}_{A_l}(x) \cap N^{\triangle\tau}_{B_k}(x) = N^{\triangle\tau}_{A_l \vee B_k}(x) \in \mathscr{B}^{\triangle}_{\eta}$ such that $x \in W \subseteq U \cap V$. Therefore by Theorem 1.21 $\mathscr{B}^{\triangle}_{\eta}$ is a base for some topology on X. □

The topological space $(X, \mathscr{T}^{\triangle}_{\eta})$, in which $\mathscr{B}^{\triangle}_{\eta}$ is a base for $\mathscr{T}^{\triangle}_{\eta}$, is called the *limited topology of X induced by the topological molecular lattice* η.

In what follows, we look more carefully at these topological structures by discussing the following illustrative examples.

Example 7.1. Let $X = \{x_1, x_2, ..., x_5\}$ be a set of 5 persons. M = {old, heavy, tall, high salary, more estate, male, female } be a set of simple concepts on the attributes which are shown as Table 7.1.

Table 7.1 Description of attributes

	age	*heigh*	*weigh*	*salary*	*estate*	*male*	*female*
x_1	21	1.69	50	0	0	1	0
x_2	30	1.62	52	120	200,000	0	1
x_3	27	1.80	65	100	40,000	1	0
x_4	60	1.50	63	80	324,000	0	1
x_5	45	1.71	54	140	940,000	1	0

We can construct the AFS structure τ according to the data shown in Table 7.1 and the semantics of the simple concepts in M. τ is shown as the following Table 7.2. Here A: *old*, M: *male*, W: *female*, H: *tall*, We: *heavey*, S: *high salary*, Q: *more estate*.

Table 7.2 The AFS structure (M, τ, X) of data shown in Table 7.1

$\tau(.,.)$	x_1	x_2	x_3	x_4	x_5
x_1	{A,M,H,We,S }	{M,H }	{M }	{M,H }	{M }
x_2	{A,W,We,S,Q }	{A,W,H,We,S,Q }	{A,W,S,Q}	{W,H,S }	{W }
x_3	{A,M,H,We,S,Q }	{M,H,We }	{A,M,H,We,S,Q}	{M,H,We,S}	{M,H,We}
x_4	{A,W,We,Q}	{A,W,We,Q }	{A,W,Q }	{A,W,We,H,S,Q}	{A,W,We }
x_5	{A,M,S,Q}	{A,M,H,We,S,Q }	{A,M,S,Q }	{M,H,S,Q }	{A,M,H,S,We,Q }

We can verify that τ satisfies Definition 4.5 and (M, τ, X) is an AFS structure. Since for any $x \in X$, $\tau(x,x) \neq M$, hence (M, τ, X) is not an strong relative AFS structure.

If we consider some health problem and suppose the problem just involves the attributes *age, high and weight*. Thus we just consider simple concepts $A, H, We \in M$ and let $M_1 = \{A, H, We\}$. (M_1, τ_{M_1}, X) is an AFS structure if the map $\tau_{M_1} : X \times X \to 2^{M_1}$ is defined as follows: for any $x, y \in X$, $\tau_{M_1}(x,y) = \tau(x,y) \cap M_1$. Obviously,

(M_1, τ_{M_1}, X) is a strong relative AFS structure. Let $\eta \subseteq EM$ be the topological molecular lattice generated by the fuzzy concepts $\{A\}, \{H\}, \{We\} \in EM$ on the lattice $^*EM$. η consists of the following elements.

$\varnothing$, M, $\{A\}+\{H\}+\{We\}$, $\{A\}+\{H\}$, $\{A\}+\{We\}$, $\{H\}+\{We\}$, $\{A\}$, $\{H\}$, $\{We\}$, $\{A,H\}+\{A,We\}+\{We,H\}$, $\{A,H\}+\{A,We\}$, $\{A,H\}+\{We,H\}$, $\{A,We\}+\{We,H\}$, $\{A,H\}$, $\{A,We\}$, $\{We,H\}$, $\{A\}+\{We,H\}$, $\{H\}+\{A,We\}$, $\{We\}+\{A,H\}$, $\{A,H,We\}$.

It could be easily verified that η is an elementary topological molecular lattice on the lattice $^*EM$. Now we study $\mathscr{T}_\eta$ the topology on X induced by η via the AFS structure (M_1, τ_{M_1}, X). The neighborhood of x_1 induced by the fuzzy concepts in η, which is obtained by Definition 7.2, are listed as follows.

$N^\tau_{\{A\}+\{H\}+\{We\}}(x_1)=\{x_1,x_2,x_4\}$, $N^\tau_{\{A\}+\{H\}}(x_1)=\{x_1,x_2,x_4\}$, $N^\eta_{\{A\}+\{We\}}(x_1)=\{x_1\}$,
$N^\tau_{\{H\}+\{We\}}(x_1)=\{x_1,x_2,x_4\}$, $N^\tau_{\{A\}}(x_1)=\{x_1\}$, $N^\tau_{\{H\}}(x_1)=\{x_1,x_2,x_4\}$,
$N^\eta_{\{We\}}(x_1)=\{x_1\}$, $N^\tau_{\{A,H\}+\{A,We\}+\{We,H\}}(x_1)=\{x_1\}$, $N^\tau_{\{A,H\}+\{A,We\}}(x_1)=\{x_1\}$,
$N^\tau_{\{A,H\}+\{We,H\}}(x_1)=\{x1\}$, $N^\tau_{\{A,We\}+\{We,H\}}(x_1)=\{x_1\}$, $N^\tau_{\{A,H\}}(x_1)=\{x_1\}$,
$N^\tau_{\{A,We\}}(x_1)=\{x_1\}$, $N^\tau_{\{We,H\}}(x_1)=\{x_1\}$, $N^\tau_{\{A\}+\{We,H\}}(x_1)=\{x_1\}$, $N^\tau_\varnothing=X$
$N^\tau_{\{H\}+\{A,We\}}(x_1)=\{x_1,x_2,x_4\}$, $N^\tau_{\{We\}+\{A,H\}}(x_1)=\{x_1\}$, $N^\tau_{\{A,H,We\}}(x_1)=\{x_1\}$.

Therefore the neighborhood of x_1 induced by the fuzzy concepts in η comes as

$$N^\tau_\eta(x_1)=\{X,\ \{x_1,x_2,x_4\},\ \{x_1\}\}.$$

Similarly, we have the neighborhood of other elements in X as follows.

$$N^\tau_\eta(x_2)=\{X,\ \{x_1,x_2,x_3,x_4\},\ \{x_1,x_2,x_4\},\ \{x_1,x_2,x_3\},\ \{x_1,x_2\},\ \{x_2,x_4\},\ \{x_2\}\},$$
$$N^\tau_\eta(x_3)=\{X,\ \{x_1,x_3\}\},$$
$$N^\tau_\eta(x_4)=\{X,\ \{x_1,x_2,x_4,x_5\},\ \{x_4\}\},$$
$$N^\tau_\eta(x_5)=\{X,\ \{x_1,x_2,x_3,x_5\},\ \{x_2,x_4,x_5\},\ \{x_2,x_5\}\}.$$

What is the interpretations of the above topological structure on X obtained from the given data shown as Table 7.1? This remains an open problem. How to establish a distance function according the above topology on X for a pattern recognition problem will be explored in Section 7.3. Here we just simply analyze it alluding to intuition. One can observe that x_1,x_2,x_4 are discrete points for the topology $\mathscr{T}_\eta$. Coincidentally, their membership degrees to the fuzzy concepts $\{A\},\{H\},\{We\} \in EM$ taken on minimal values, respectively. For any $U \in \mathscr{T}_\eta$, we can prove that if $x_5 \in U$ then $x_2 \in U$. This implies that the degree of x_5 belonging to any concept in EM_1 is always larger than or equal to that of x_2. Since $x_5 \notin \{x_1,x_3\} \in \mathscr{T}_\eta$, $x_3 \notin \{x_2,x_5\} \in \mathscr{T}_\eta$, i.e., the separation property of topology $\mathscr{T}_\eta$. This implies that there exist two fuzzy concepts in η such that x_5,x_3 can be distinguished by them.

7.2 Topology on AFS Structures and Topological Molecular Lattice on EI^n Algebras

Most of the results in this section can be proved by using the similar methods to those we exercised in the previous section, since the lattice of EI^n algebras is the dual lattice of $^*EI^n$ algebras. We first list the corresponding results of the topological molecular lattice on the EI algebra over a set M, in which the lattice operators $\vee$, $\wedge$ are defined as follows: for any $\sum_{i\in I} A_i,\ \sum_{j\in J} B_j \in EM$,

$$\sum_{i\in I} A_i \wedge \sum_{j\in J} B_j = \sum_{i\in I, j\in J} A_i \cup B_j,$$
$$\sum_{i\in I} A_i \vee \sum_{j\in J} B_j = \sum_{i\in I} A_i + \sum_{j\in J} B_j.$$

$\varnothing$ is the maximum element of the lattice EM and M is the minimum element of the lattice EM. That is, the above lattice EM is a dual lattice of $^*EM$. In the lattice EM, for $\sum_{i\in I} A_i, \sum_{j\in J} B_j \in EM$, $\sum_{i\in I} A_i \geq \sum_{j\in J} B_j$ if and only if for any $B_j\ (j\in J)$ there exists $A_k\ (i\in I)$ such that $B_j \supseteq A_k$ (refer to Theorem 4.1). Secondly, we list the results for the topology on the universe of discourse X induced by the topological molecular lattice of some fuzzy concepts in EM. Finally, we present the results of the topological molecular lattice on the EI^2 algebra over the sets X, M, i.e.,$(EXM, \vee, \wedge)$, in which the lattice operators $\vee$, $\wedge$ are defined as follows: for any $\sum_{i\in I} a_i A_i,\ \sum_{j\in J} b_j B_j \in EXM$,

$$\sum_{i\in I} a_i A_i \wedge \sum_{j\in J} b_j B_j = \sum_{i\in I, j\in J} a_i \cap b_j A_i \cup B_j,$$
$$\sum_{i\in I} a_i A_i \vee \sum_{j\in J} b_j B_j = \sum_{i\in I} a_i A_i + \sum_{j\in J} b_j B_j.$$

$X\varnothing$ is the maximum element of the lattice EM and $\varnothing M$ is the minimum element of the lattice EM. That is, the lattice EXM is a dual lattice of $^*EXM$. In the lattice EXM, for $\sum_{i\in I} a_i A_i, \sum_{j\in J} b_j B_j \in EM$, $\sum_{i\in I} a_i A_i \geq \sum_{j\in J} b_j B_j$ if and only if for any $b_j B_j\ (j\in J)$ there exists $a_k A_k\ (i\in I)$ such that $B_j \supseteq A_k$ and $a_k \supseteq b_j$ (refer to Theorem 5.1).

Definition 7.5. Let M be set and $(EM, \vee, \wedge)$ be the EI algebra over M. Let $\eta \subseteq EM$. If $\varnothing, M \in \eta$ and η is closed under finite unions (i.e., $\vee$) and arbitrary intersections (i.e., $\wedge$), then η is called a *topological molecular lattice on the lattice EM*, denoted as (EM, η). Let η be a topological molecular lattice on the lattice EM. If for any $\sum_{i\in I} A_i \in \eta$, $A_i \in \eta$ for any $i \in I$, then η is called an *elementary topological molecular lattice on the lattice EM*.

In what follows, we apply the elementary topological molecular lattice on the lattice EM to induce some topological structures on X via the AFS structure (M, τ, X) of a data. Thus the pattern recognition problem can by explored under these topological structures on X.

Definition 7.6. Let X and M be sets and (M,τ,X) be an AFS structure. Let (EM,η) be a topological molecular lattice on EI algebra over M. For any $x\in X$, $\sum_{i\in I}A_i\in\eta\subseteq EM$, the set $N^{\tau}_{\sum_{i\in I}A_i}(x)\subseteq X$ is defined as follows.

$$N^{\tau}_{\sum_{i\in I}A_i}(x)=\left\{y\in X\mid \tau(x,y)\leq\sum_{i\in I}A_i\right\}, \tag{7.11}$$

and it is called the *neighborhood of x induced by the fuzzy concept* $\sum_{i\in I}A_i$ in the AFS structure (M,τ,X). The set $N^{\tau}_{\eta}(x)\subseteq 2^X$ is defined as follows.

$$N^{\tau}_{\eta}(x)=\left\{N_{\sum_{i\in I}A_i}(x)|\sum_{i\in I}A_i\in\eta\right\}, \tag{7.12}$$

and it is called the *neighborhood of x induced by the topological molecular lattice* η in the AFS structure (M,τ,X).

Since $\tau(x,y)\subseteq M$, hence $\tau(x,y)$ is an element in EM and $\tau(x,y)\leq\sum_{i\in I}A_i$ in (7.11) is well-defined.

Proposition 7.3. *Let X and M be sets and (M,τ,X) be an strong relative AFS structure. Let η be a topological molecular lattice on EI algebra over M. For any $x\in X$, $\sum_{i\in I}A_i$, $\sum_{j\in J}B_j\in EM$, the following assertions hold: for any $x\in X$*

(1) If $\sum_{i\in I}A_i\geq\sum_{j\in J}B_j$ in EM, then $N^{\tau}_{\sum_{i\in I}A_i}(x)\supseteq N^{\tau}_{\sum_{j\in J}B_j}(x)$;
(2) $N^{\tau}_{\sum_{i\in I}A_i}(x)\cap N^{\tau}_{\sum_{j\in J}B_j}(x)=N^{\tau}_{\sum_{i\in I}A_i\wedge\sum_{j\in J}B_j}(x)$;
(3) $N^{\tau}_{\sum_{i\in I}A_i}(x)\cup N^{\tau}_{\sum_{j\in J}B_j}(x)=N^{\tau}_{\sum_{i\in I}A_i\vee\sum_{j\in J}B_j}(x)$.

Its proof, which is similar to the proof of Proposition 7.1, remains as an exercise.

Theorem 7.5. *Let X and M be sets and (M,τ,X) be a strong relative AFS structure. Let η be a topological molecular lattice on the lattice EM. If η is an elementary topological molecular lattice on the lattice EM and we define*

$$\mathscr{B}_{\eta}=\left\{N^{\tau}_{\sum_{i\in I}A_i}(x)\mid x\in X,\ \sum_{i\in I}A_i\in\eta\right\},$$

then $\mathscr{B}_{\eta}$ is a base for some topology of X.

Its proof, which is similar to the proof of Theorem 7.1, is left to the reader.

Theorem 7.6. *Let X and M be sets, (M,τ,X) be a strong relative AFS structure. Let η be a topological molecular lattice on the lattice EM and*

$$\mathscr{L}_{\eta}=\left\{\sum_{i\in I}a_iA_i\in EXM\ \mid\sum_{i\in I}A_i\in\eta,\ a_i\in\mathscr{T}_{\eta}\ for\ any\ i\in I\right\}. \tag{7.13}$$

Then $\mathscr{L}_\eta$ *is a topological molecular lattice on the lattice EXM. It is called the* EI^2 topological molecular lattice induced by the EI topological molecular lattice η.

Its proof (similar to the proof of Theorem 7.2) remains as an exercise.

Theorem 7.7. *Let X and M be sets. Let* (M,τ,X) *be a strong relative AFS structure and* η *be an elementary topological molecular lattice on the lattice EM. Let* η *be a topological molecular lattice on the lattice EM and the topological space* $(X,\mathscr{T}_\eta)$ *be the topology induced by* η*. Let S be the* σ*-algebra generated by* $\mathscr{T}_\eta$ *and* $\mathscr{L}_\eta$ *be the* EI^2 *topological molecular lattice on EXM induced by* η*. Then the following assertions hold.*

(1) For any fuzzy concept $\sum_{i\in I} A_i \in \eta$, $\sum_{i\in I} A_i$ *is a measurable concept under S ;*
(2) For each fuzzy concept $\gamma = \sum_{i\in I} A_i \in \eta$*, let* $\gamma : X \to EXM$ *be the* EI^2 *algebra representation membership degrees defined by (5.10) as follows: for any* $x \in X$,

$$\gamma(x) = \sum_{i\in I} A_i^\tau(x) A_i \in EXM. \tag{7.14}$$

Let D be a directed set and $\delta : D \to X$ *be a net (i.e.,* $\{\delta(d) \mid d \in D\}$*). If* δ *is converged to* $x_0 \in X$ *under topology* $\mathscr{T}_\eta$*, then the net of the composition* $\gamma\cdot\delta : D \to EXM$ *(i.e.,*$\{\gamma(\delta(d)) \mid d \in D\}$*) converges to* $\gamma(x_0) = \sum_{i\in I} A_i^\tau(x_0) A_i$ *under the topological molecular lattice* $\mathscr{L}_\eta$*. That is the membership function defined by (7.14) is a continuous function from the topological space* $(X,\mathscr{T}_\eta)$ *to the topological molecular lattice* $(EXM,\mathscr{L}_\eta)$.

Its proof, which is similar to the proof of Theorem 7.3, can be treated as an exercise.

Example 7.2. Let us study the topological structures on the same AFS structure (M_1,τ_{M_1},X) of the same data we used in Example 7.1.

Let $\eta \subseteq EM$ be the topological molecular lattice generated by the fuzzy concepts $\{A\},\{H\},\{We\} \in EM$ on the lattice EM. η consists of the following elements which are the same as for η in Example 7.1.

$\varnothing$, M, $\{A\}+\{H\}+\{We\}$, $\{A\}+\{H\}$, $\{A\}+\{We\}$, $\{H\}+\{We\}$, $\{A\}$, $\{H\}$, $\{We\}$, $\{A,H\}+\{A,We\}+\{We,H\}$, $\{A,H\}+\{A,We\}$, $\{A,H\}+\{We,H\}$, $\{A,We\}+\{We,H\}$, $\{A,H\}$, $\{A,We\}$, $\{We,H\}$, $\{A\}+\{We,H\}$, $\{H\}+\{A,We\}$, $\{We\}+\{A,H\}$, $\{A,H,We\}$.

It can be easily to verify that η is an elementary topological molecular lattice on the lattice EM. Now we study $\mathscr{T}_\eta$-the topology on X induced by η via the AFS structure (M_1,τ_{M_1},X). The neighborhood of x_1 induced by the fuzzy concepts in η, which is obtained by Definition 7.6, is listed as follows.

$N^\tau_{\{A\}+\{H\}+\{We\}}(x_1) = \{x_1,x_2,x_3,x_4,x_5\}$, $N^\tau_{\{A\}+\{H\}}(x_1) = \{x_1,x_2,x_3,x_4,x_5\}$,
$N^\tau_{\{A\}+\{We\}}(x_1) = \{x_1,x_2,x_3,x_4,x_5\}$, $N^\tau_{\{H\}+\{We\}}(x_1) = \{x_1,x_2,x_3,x_4\}$,
$N^\tau_{\{A\}}(x_1) = \{x_1,x_2,x_3,x_4,x_5\}$, $N^\tau_{\{H\}}(x_1) = \{x_1,x_3\}$, $N^\tau_{\{We\}}(x_1) = \{x_1,x_2,x_3,x_4\}$,
$N^\tau_{\{A,H\}+\{A,We\}+\{We,H\}}(x_1) = \{x_1,x_2,x_3,x_4\}$, $N^\tau_{\{A,H\}+\{A,We\}}(x_1) = \{x_1,x_2,x_3,x_4\}$,

$N^{\tau}_{\{A,H\}+\{We,H\}}(x_1)=\{x_1,x_3\}$, $N^{\tau}_{\{A,We\}+\{We,H\}}(x_1)=\{x_1,x_2,x_3,x_4\}$,
$N^{\tau}_{\{A,H\}}(x_1)=\{x_1,x_3\}$, $N^{\tau}_{\{A,We\}}(x_1)=\{x_1,x_2,x_3,x_4\}$, $N^{\tau}_{\{We,H\}}(x_1)=\{x_1,x_3\}$,
$N^{\tau}_{\{A\}+\{We,H\}}(x_1)=\{x_1,x_2,x_3,x_4,x_5\}$, $N^{\tau}_{\{H\}+\{A,We\}}(x_1)=\{x_1,x_2,x_3,x_4,x_5\}$,
$N^{\tau}_{\{We\}+\{A,H\}}(x_1)=\{x_1,x_2,x_3,x_4\}$, $N^{\tau}_{\{A,H,We\}}(x_1)=\{x_1,x_3,x_5\}$, $N^{\tau}_{M}(x_1)=X$.

Therefore the neighborhood of x_1 induced by the fuzzy concepts in η is

$$N^{\tau}_{\eta}(x_1)=\{X,\ \{x_1,x_2,x_3,x_4\},\ \{x_1,x_3,x_5\},\ \{x_1,x_3\}\}.$$

Similarly, we have the neighborhood of other elements in X as follows.

$$\begin{aligned}
N^{\tau}_{\eta}(x_2) &= \{X,\ \{x_2,x_3,x_4,x_5\},\ \{x_1,x_2,x_3,x_4\},\ \{x_2,x_4,x_5\},\ \{x_2,x_3,x_5\},\{x_2,x_5\}\},\\
N^{\tau}_{\eta}(x_3) &= \{X,\ \{x_2,x_3,x_4,x_5\},\ \{x_3\}\},\\
N^{\tau}_{\eta}(x_4) &= \{X,\ \{x_3,x_4\},\ \{x_4\}\},\\
N^{\tau}_{\eta}(x_5) &= \{X,\ \{x_3,x_4,x_5\},\ \{x_4,x_5\},\ \{x_3,x_5\},\ \{x_5\}\}.
\end{aligned}$$

Here we just simply analyze the topology on X resorting ourselves to intuition. One can observe that x_3, x_4, x_5 are discrete points for the topology $\mathscr{T}_\eta$. Coincidentally, their membership degrees to the fuzzy concepts $\{A\},\{H\},\{We\}$ taken on the maximal values, respectively. For any $U\in\mathscr{T}_\eta$, we can prove that if $x_1\in U$ then $x_3\in U$. This implies that the degree of x_3 belonging to any fuzzy concept in EM_1 is always larger than or equal that of x_1. Since $x_2\in\{x_1,x_3\}\in\mathscr{T}_\eta$, $x_1\in\{x_2,x_5\}\in\mathscr{T}_\eta$ i.e., the separation property of topology $\mathscr{T}_\eta$, hence there exist two fuzzy concepts in η such that x_2,x_1 can be distinguished by them. Compared with the topological structure on X induced by the topological molecular lattice on $^*EI^n$ algebra, the above topological structure has many differences. What are the relationship between these topological structures still remains as an open problem.

7.3 Fuzzy Similarity Relations Based on Topological Molecular Lattices

In this section, by considering the AFS structure (M,τ,X) of a data, we apply $\mathscr{T}_\eta$ the topology on X induced by the topological molecular lattice η of some fuzzy concepts on $^*EM$ to study the fuzzy similarity relations on X for problems of pattern recognition. The topology $\mathscr{T}_\eta$ on X is determined by the original data and some selected fuzzy concepts in EM. It represents the abstract geometry relations among the objects in X. We study the interpretations of the induced topological structures on the AFS structures directly obtained by a given data set through the differential degrees between objects in X and the fuzzy similarity relations on X in the topological space $(X,\mathscr{T}_\eta)$. We know that human can classify, cluster and recognize the objects in the set X without any metric in Euclidean space. What is human recognition based on if X is not a subset of some metric space in Euclidean space? For example, if you want to classify all your friends into two classes $\{close\,friends\}$ and $\{common\,friends\}$. The criteria/metric you are using in the process is very important though it may not be based on the Euclidean metric. By the fuzzy clustering analysis based on the topological spaces induced by the fuzzy concepts in EM, we hope find some clues for these challenge problems.

Theorem 7.8. *The following three conditions on a topological space are equivalent.*

(1) The space is metrizable;
(2) The space is T_1 and regular, and the topology has a σ-locally finite base;
(3) The space is T_1 and regular, and the topology has a σ-discrete base.

Here a topological space is a T_1 space if and only if each set which consists of a single point is closed, a topological space is regular if and only if for each point x and each neighborhood U of x there is a closed neighborhood V of x such that $V \subseteq U$, and a family is σ-locally finite (σ-discrete) if and only if it is the union of a countable number of locally finite (respectively, discrete) subfamilies.

Its proof (refer to Theorem 1.43) is left to the reader.

The topology $\mathscr{T}_\eta$ on X induced by the topological molecular lattice η of some fuzzy concepts in EM is a description of the abstract geometry relations among the objects determined by the semantic interpretations of the fuzzy concepts in η and the distributions of the original data. We can state the problem in mathematical ways as follows: Let X be a set of some objects and $\mathcal{F}$ be the set of all features, including features which are independent or irrelated to the problems under considering. M is the set of simple concepts on the features in $\mathcal{F}$. $\Lambda \subseteq EM$, Λ is the set of fuzzy concepts an individual considers crucial to his problem. η is the topological molecular lattice generated by Λ. If the topology $\mathscr{T}_\eta$ satisfies (2) or (3) in Theorem 7.8, then the topology space $(X, \mathscr{T}_\eta)$ is metrizable. Thus we can study the clustering and recognition problems by the metric induced by topology $\mathscr{T}_\eta$, i.e., the distance function d on the cartesian product $X \times X$ to the non-negative reals defined by Definition 1.33 as follows: for all points x, y, and z of X,

1. $d(x,y) = d(y,x)$,
2. $d(x,y) + d(y,z) \geq d(x,z)$, (*triangle inequality*)
3. $d(x,y) = 0$ if $x = y$, and
4. if $d(x,y) = 0$, then $x = y$.

However, for a real world applications, it is very difficult to satisfy the conditions of Theorem 7.8. In other words, this theorem cannot be directly applied to real world classification scenarios. By the analysis of the definition of metric in metrizable topology space $(X, \mathscr{T}_\eta)$ in mathematics (refer to Urysohn Lemma Lemma 1.1), we know that the more fuzzy concepts distinguish x from y are there in η, the larger the distance of x and y, i.e., $d(x,y)$. In practice, for X a set of objects and $\Lambda \subseteq EM$ a set of selected fuzzy or Boolean concepts, although $\mathscr{T}_\eta$ the topology induced by the topological molecular lattice η seldom satisfies (2) or (3) in Theorem 7.8, $\mathscr{T}_\eta$ also can reflect the similar relations between the objects in X determined by the concepts in Λ and the distributions of the original data. Thus we define the differential degree and the similarity degree of x, $y \in X$ based on the topology $\mathscr{T}_\eta$ as follows.

Definition 7.7. Let X and M be finite sets and (M,τ,X) be an AFS structure. Let η be a topological molecular lattice on the lattice $^*EM$ and $(X,\mathscr{T}_\eta)$ be the topology space on X induced by η. We define the *partial distance function* $D(x,y)$, the *differential degree* $d(x,y)$ and the *similarity degree* $s(x,y)$ in the topological space $(X,\mathscr{T}_\eta)$ as follows: for $x,y \in X$,

$$D(x,y) = \sum_{\delta\in\mathscr{T}_\eta, x\in\delta, y\notin\delta} |\delta|; \tag{7.15}$$

$$d(x,y) = D(x,y)+D(y,x); \tag{7.16}$$

$$s(x,y) = 1-\frac{d(x,y)}{\max_{z\in X}\{d(z,y)\}}. \tag{7.17}$$

Because there are too many fuzzy concepts in η, in practice, it is difficult or impossible to calculate $d(x,y)$ by Definition 7.7 for the topological molecular lattice η generated by Λ, if $|\Lambda|>4$. The following Definition 7.8 and Definition 7.9 introduce the differential degrees of x,y, $d(x,y)$ which are more expedient to compute than that in Definition 7.7, although they may loose some information compared with the concept captured by Definition 7.7. Definition 7.8 and Definition 7.9 are applicable to discuss real world problems while Definition 7.7 is more appealing from the theoretical perspective.

Definition 7.8. Let X and M be finite sets and (M,τ,X) be an AFS structure. Let η be an elementary topological molecular lattice on the lattice $^*EM$ and $(X,\mathscr{T}_\eta)$ be the topology space on X induced by η. We define $D_A(x,y)$, the *distance function on the molecular* A; $d_M(x,y)$, the *molecular differential degree*; and $s_M(x,y)$, the *molecular similarity degree* in the topological space $(X,\mathscr{T}_\eta)$ as follows: for $x,y\in X$, $A\subseteq M$, $A\in\eta$,

$$D_A(x,y) = \sum_{u\in X, x\in N_A^{\triangle\tau}(u), y\notin N_A^{\triangle\tau}(u)} |\, N_A^{\triangle\tau}(u)|; \tag{7.18}$$

$$d_M(x,y) = \sum_{A\subseteq M, A\in\eta} (D_A(x,y)+D_A(y,x)); \tag{7.19}$$

$$s_M(x,y) = 1-\frac{d_M(x,y)}{\max_{z\in X}\{d_M(z,y)\}}. \tag{7.20}$$

$D_A(x,y)$ in Definition 7.8 is considered under the fuzzy molecular concept $A\in\eta$ and $d_M(x,y)$, the molecular differential degree of x,y is the sum of the distances of x,y under all fuzzy molecular concepts in η.

Definition 7.9. Let X and M be finite sets and (M,τ,X) be an AFS structure. Let η be an elementary topological molecular lattice on the lattice $^*EM$. Let $(X,\mathscr{T}_\eta)$ be the topology space on X induced by η. We define the *elementary partial distance function* $D^e(x,y)$, the *elementary differential degree* $d^e(x,y)$ and the *elementary similarity degree* $s_e(x,y)$ in the topological space $(X,\mathscr{T}_\eta)$ as follows: for any $x,y\in X$,

$$D_e(x,y) = \sum_{\delta \in \mathscr{B}_M^{\triangle\tau}, x \in \delta, y \notin \delta} |\delta|; \tag{7.21}$$

$$d_e(x,y) = D_e(x,y) + D_e(y,x); \tag{7.22}$$

$$s_e(x,y) = 1 - \frac{d_e(x,y)}{\max_{z \in X}\{d_e(z,y)\}}. \tag{7.23}$$

Here

$$\mathscr{B}_M^{\triangle\tau} = \left\{ N_A^{\triangle\tau}(x) \mid A \subseteq M, A \in \eta, x \in X \right\}.$$

It is clear that $\mathscr{B}_M^{\triangle\tau} \subseteq \mathscr{T}_\eta$ is the set of all neighborhoods induced by the fuzzy molecular concepts in η which determine the distances and similarity degrees defined by Definition 7.9. However, in Definition 7.7, they are determined by all neighborhoods in $\mathscr{T}_\eta$. Since the number of the elements of $\mathscr{T}_\eta$ is much larger than that of the set $\mathscr{B}_M^{\triangle\tau}$, hence much time will save if Definition 7.9 or Definition 7.8 is applied to a pattern recognition problem. The problem is still open: are the similarity degrees defined by Definition 7.7, Definition 7.8 and Definition 7.9 equivalent?

Proposition 7.4. *Let X and M be finite sets and (M,τ,X) be an AFS structure. Let η be an elementary topological molecular lattice on the lattice $^*EM$ and $(X,\mathscr{T}_\eta)$ be the topology space on X induced by η. Then for any $x,y \in X$ the following assertions hold.*

(1) $d(x,x) = 0$, $d(x,y) = d(y,x)$ *and* $s(x,y) = s(y,x) \le s(x,x)$;
(2) $d_M(x,x) = 0$, $d_M(x,y) = d_M(y,x)$ *and* $s_M(x,y) = s_M(y,x) \le s_M(x,x)$;
(3) $d_e(x,x) = 0$, $d_e(x,y) = d_e(y,x)$ *and* $s_e(x,y) = s_e(y,x) \le s_e(x,x)$.

Its proof is left to the reader.

7.4 Fuzzy Clustering Algorithms Based on Topological Molecular Lattices

Numerous mathematical tools, investigated for clustering, have been considered to detect similarities between objects inside a cluster. The two-valued clustering is described by a characteristic function. This function assigns each object to one and only one of the clusters, with a degree of membership equal to one. However, the boundaries between the clusters are not often well-defined and this description does not fully reflect the reality. The fuzzy clustering, founded upon fuzzy set theory [35], is meant to deal with not well-defined boundaries between clusters. Thus, in fuzzy clustering, the membership function is represented by grades located anywhere in-between zero and one. Therefore, this membership degree indicates how the object is classified (allocated) to each cluster. This can be advantageous for patterns located in the boundary region which may not be precisely defined. In particular, we could flag some patterns that are difficult to assign to a single cluster as being inherently positioned somewhere at the boundary of the clusters.

Many fuzzy clustering algorithms have been developed, but the most widely used is the Fuzzy C-Means algorithm (FCM) along with a significant number of their variants. Conceived by Dunn [2] and generalized by Bezdek [1], this family of algorithms is based on iterative optimization of a fuzzy objective function. The convergence of the algorithm, proved by Bezdek, shows that the method converges to some local minima [4]. Nevertheless, the results produced by these algorithms depend on some predefined distance formulated in a metric space, for instance Euclidean space R^n. However, in this section we will cluster the objects in ordinary data set $X \nsubseteq R^{p\times n}$ according to the fuzzy concepts or attributes on the features without using any kind of distance functions expressed in the Euclidean space.

In general, FCM is an objective function optimization approach to solve the following problem [1, 4]:

$$minimize: \ \ J_m(U,V) = \sum_i \sum_k u_{ik}^m d^2(x_k, v_i)$$

with respect to $U = [u_{ik}] \in R^{c\times n}$, a fuzzy c-partition of n data set $X = \{x_1,...,x_n\} \in R^{p\times n}$ and V, a set of c cluster centers $V = \{v_1,...,v_c\} \in R^{p\times c}$. The parameter $m > 1$ is a fuzziness coefficient. $d(x_k, v_i)$ is a distance from x_k to the ith cluster center v_i. The performance of FCM is affected by different distances $d(.,.)$. In general, the distance is expressed in some metric space [4, 34], if data set X is a subset of a metric space. FCM fuzzy clustering algorithms are very efficient if the data set $X \subset R^{p\times n}$, as in this case there exists a distance function. Let c be a positive integer greater than one. $\mu = \{\mu_1,...,\mu_c\}$ is called a fuzzy c-partition of X, if $\mu_i(x)$ is the membership functions in fuzzy sets μ_i on X assuming values in the $[0,1]$ such that $\sum_{i=1}^c \mu_i(x) = 1$ for all x in X. Thus, the Fuzzy C-Mean (FCM) objective function $J(\mu,V)$ is also defined as

$$J(\mu,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} \mu_i^m(x_j)||x_j - v_i||^2, \tag{7.24}$$

where $\mu_i(x_j) = u_{ij} = \mu_{ij}$ and $d(x_k, v_i) = \|x_k - v_i\|$. The FCM clustering is an iterative algorithm where the update formulas for the prototypes and the partition matrix read as follows:

$$v_i = \frac{\sum_{j=1}^n \mu_{ij}^m x_j}{\sum_{j=1}^n \mu_{ij}^m}, i = 1,...,c \tag{7.25}$$

and

$$\mu_{ij} = \mu_i(x_j) = (\sum_{k=1}^{c} \frac{||x_j - v_i||^{2/(m-1)}}{||x_j - v_k||^{2/(m-1)}})^{-1}, i = 1,...,c, j = 1,...,n. \tag{7.26}$$

If the feature vectors are numeric data in R^d, the FCM clustering algorithm is a suitable optimization tool. However, when applying the FCM to data set with mixed features such as Boolean, partial order and linguistic rating scale, we encounter some problems, because the conventional distance $||.||$ is not suitable any longer.

To overcome these problems, the differential degrees $d(x,y)$ (or $d_e(x,y)$, $d_M(x,y)$) defined in the above section can substitute the Euclidean distance $||.||$ and the FCM can be modified as follows.

$$\underset{\{v_1,...,v_c\}\subseteq X}{min:}\ J(\mu,V)=\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_i^m(x_j)d(x_j,v_i)^2 \tag{7.27}$$

subject to

$$\mu_i(x_j)=(\textstyle\sum_{k=1}^{c}\frac{d(x_j,v_i)^{2/(m-1)}}{d(x_j,v_k)^{2/(m-1)}})^{-1}, i=1,...,c, j=1,...,n.$$

This algorithm is called the *AFS fuzzy c-mean algorithm* (AFS_FCM).

In order to compare the differential functions defined in the above section with the Euclidean distance function, we directly apply the similarity matrix derived by the differential function and Euclidean distance function to the clustering problem. Let $X=\{x_1,x_2,...,x_n\}$ and the similarity matrix $S=(s_{ij})_{n\times n}$, where $s_{ij}=s_e(x_i,x_j)$ is elementary similarity degree of x,y defined by Definition 7.8. For the similarity matrix S, we know $s_{ij}=s_{ji}$ and $s_{ij}\le s_{ii}$, $1\le i,j\le n$ from Proposition 7.4, hence there exists an integer r such that $S\le S^2\le ...\le S^r=S^{r+1}$, where $S^2=(r_{ij})=SS$ is the fuzzy matrix product, i.e., $r_{ij}=\max_{1\le k\le n}\min\{s_{ik},s_{kj}\}$. Thus, $(S^r)^2=S^r$ (S^r is the transitive closure matrix of S) and the fuzzy equivalence relation matrix $Q=(q_{ij})=S^r$ can yield a partition tree with equivalence classes in which x_i and x_j are in the same cluster (i.e., in the same equivalence classe) under some threshold $\alpha\in[0,1]$ if and only if $q_{ij}\ge\alpha$.

7.5 Empirical Studies

In this section, we apply the similarity relations and the differential functions defined by Definition 7.8 to the conventional FCM and compare the elementary differential function with the Euclidean distance function in the clustering analysis of the Iris data. Furthermore, they are also applied to Taiwan airfreight forwarder data which is just described by means of linguistic terms. These examples show that the topology $\mathscr{T}_\eta$ on a universe of discourse X induced by the topological molecular lattice η of some fuzzy concepts in EM can be applied to the real world pattern recognition problems for the data set with mixed features on which the classical distance functions could not be defined.

7.5.1 Empirical Examples of Taiwan Airfreight Forwarder

In what follows, we apply the elementary differential degree and elementary similarity degree defined by Definition 7.8 to empirical examples of Taiwan airfreight forwarder for the clustering and analyzing current operation strategies in [27]. In [27], the authors gathered 28 strategic criteria from scholars, experts and proprietors. They select 30 companies of airfreight forwarder in Taiwan by random selection. Using Statistical Analysis System (SAS), they obtain seven factors: Factor1: *Core* *ability*, Factor2: *Organization management*, Factor3: *Pricing*, Factor4: *Competitive forces*, Factor5: *Finance*, Factor6: *Different advantage*, Factor7: *Information technology*. The decision-makers may tackle preference rating system by adopting one of various rating scales assumed in the literature [8, 28, 29] or may develop their own rating scales system by using trapezoidal fuzzy number to show the individual conception of the linguistic variable "*attention degree*". According to the preference ratings proposed by Liang and Wang [28], it is suggested that the decision-makers utilize the linguistic rating set

$$W = \{VL, B.VL\&L, L, B.L\&M, M, B.M\&H, H, B.H\&VH, VH\},$$

where VL: *Very Low*, $B.VL\&L$: *Between Very Low and Low*, L: *Low*, $B.L\&M$: *Between Low and Medium*, M: *Medium*, $B.M\&H$: *Between Medium and High*, H:*High*, $B.H\&VH$:*Between High and Very High*, VH:*Very High*, to assess the attention degree of subjects of companies under each of the management strategies. The decision-makers utilize the linguistic rating as above and obtain the evaluation results as Table 7.3. Let $X = \{C_1, ..., C_5\}$ and $M = \{m_1, m_2, ..., m_7\}$ be the set of simple concepts on the features *Factor1* to *Factor7*. Where m_i: *great attention degree of Factor i*, $i = 1, 2, ..., 7$. The following order relation of the elements in the linguistic rating set W is determined by their linguistic rating scales:

$$VL < B.VL\&L < L < B.L\&M < M < B.M\&H < H < B.H\&VH < VH \quad (7.28)$$

For each $m_i \in M$, we can define a binary relation R_{m_i} on X by Table 7.3 and the order relation shown as (7.28): $(C_k, C_k) \in R_i$, for any $k = 1, ..., 5$ and for any $k \neq l$, $(C_k, C_l) \in R_i \Leftrightarrow C_k$(Factor i)$\geq C_l$(Factor i), where C_j(Factor i) is the linguistic rating scale of C_j for Factor i. By Definition 4.3, one can verify that for each $m_i \in M$, R_i is a simple concept. (X, τ, M) is an AFS structure if τ is defined as follows: For any C_i, $C_j \in X$, $\tau(C_i, C_j) = \{m_k \in M | (C_i, C_j) \in R_k\}$ (refer to (4.26)). Let $\Lambda = \{\{m_1\}, ..., \{m_7\}\} \subseteq EM$ and η be the topological molecular lattice generated by Λ. Let $(X, \mathscr{T}_\eta)$ be the topology space on X induced by η. Let $d_e(C_i, C_j)$ be the elementary differential degree of C_i, C_j and $s_e(C_i, C_j)$ be the elementary similarity degree of x, y defined by Definition 7.8. We obtain the following fuzzy similar matrix $S = (s_{ij})_{n\times n}$, $s_{ij} = s_e(C_i, C_j)$ and the following elementary differential matrix $T = (t_{ij})_{n\times n}$, $t_{ij} = d_e(C_i, C_j)$.

Table 7.3 The evaluation results of five companies

Company	Factor						
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
C1	M	H	H	B.H & VH	VH	L	B.M & H
C2	H	B.L& M	M	B.M & H	H	B.M & H	VL
C3	H	H	B.M & H	H	H	VH	B.M & H
C4	VL	M	H	B.VL&L	H	B.L& M	M
C5	L	M	B.H & VH	H	B.H & VH	B.VL&L	B.M & H

$$T=\begin{bmatrix} 0 & 1513 & 1175 & 1112 & 666 \\ 1513 & 0 & 638 & 1067 & 1391 \\ 1175 & 638 & 0 & 1161 & 1263 \\ 1112 & 1067 & 1161 & 0 & 918 \\ 666 & 1391 & 1263 & 918 & 0 \end{bmatrix}$$

$$S=\begin{bmatrix} 1.0 & 0 & 0.2234 & 0.2650 & 0.5598 \\ 0 & 1.0 & 0.5783 & 0.2948 & 0.0806 \\ 0.2234 & 0.5783 & 1.0 & 0.2327 & 0.1652 \\ 0.2650 & 0.2948 & 0.2327 & 1.0 & 0.3933 \\ 0.5598 & 0.0806 & 0.1652 & 0.3933 & 1.0 \end{bmatrix}$$

Then, the transitive closure of similar matrix S is S^4, i.e.,

$$(S^4)^2=S^4=\begin{bmatrix} 1.0 & 0.2948 & 0.2948 & 0.3933 & 0.5598 \\ 0.2948 & 1.0 & 0.5783 & 0.2948 & 0.2948 \\ 0.2948 & 0.5783 & 1.0 & 0.2948 & 0.2948 \\ 0.3933 & 0.2948 & 0.2948 & 1.0 & 0.3933 \\ 0.5598 & 0.2948 & 0.2948 & 0.3933 & 1.0 \end{bmatrix}$$

Let the threshold $\alpha = 0.5$. Then the clusters are {C1, C5}, {C2, C3} and {C4}. In [27], the transitive closure of the compatibility relation R_T of Table 7.3 is obtained as follows:

$$R_T=\begin{bmatrix} 1 & 0.389 & 0.415 & 0.590 & 0.679 \\ 0.389 & 1 & 0.389 & 0.389 & 0.389 \\ 0.415 & 0.389 & 1 & 0.415 & 0.415 \\ 0.590 & 0.389 & 0.415 & 1 & 0.590 \\ 0.679 & 0.389 & 0.415 & 0.590 & 1 \end{bmatrix}.$$

By taking $\lambda \in (0.590, 0.679]$, the authors in [27] obtained the clusters: {C1,C5}, {C2}, {C3} and {C4}.

By the application of the AFS-FCM algorithm described by (7.27) with the elementary differential degree defined by Definition 7.8 to the data of the 30 companies shown in **Appendix A**, let the cluster number c be equal to 5, we obtain the clustering results

cluster1={C2, C3, C6,C7}, cluster2={C1, C4, C5, C10, C16, C21, C23, C25, C28},
cluster3={C9, C11, C13, C17, C19, C27},
cluster4={C8, C18, C20, C24, C26, C29}, cluster5={C12, C14, C15, C22, C30}.

Figure 7.1, in which the x-axis is the re-order of the C1,..., C30 by the order cluster 1,...,cluster 5, i.e., 1:4 cluster 1; 5:13 cluster 2; 14:19 cluster 3; 20:25 cluster 4; 26:30 cluster 5, shows the membership functions of the fuzzy partition matrix of X, $\mu = \{\mu_1, ..., \mu_5\}$.

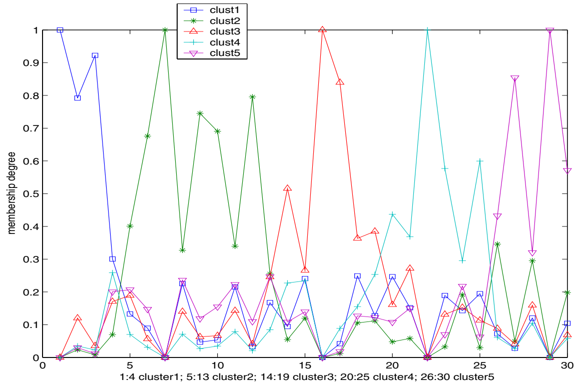


Fig. 7.1 The membership functions μ_i of the fuzzy 5-partition of X, $\mu = \{\mu_1, ..., \mu_5\}$

7.5.2 Experimental Studies on the Iris Data Set

The Iris data [30] have 150×4 matrix $W = (w_{ij})_{150\times 4}$ evenly distributed in three classes: iris-setosa, iris-versicolor, and iris-virginica. Vector of sample i, $(w_{i1}$, w_{i2}, w_{i3}, $w_{i4})$ has four features: sepal length and width, and petal length and width (all given in centimeters). Let $X = \{x_1, x_2, ..., x_{150}\}$ be the set of the 150 samples, where $x_i = (w_{i1}, w_{i2}, w_{i3}, w_{i4})$. Let $M = \{m_1, m_2, ..., m_8\}$ be the set of simple concepts on the features, where

m_1: *the sepal is long*, m_2: *sepal is wide*, m_3: *petal is long*, m_4: *petal is wide*; $m_5 = m_4'$: *petal is not wide*, $m_6 = m_3'$: *petal is not long*, $m_7 = m_2'$: *sepal is not wide*, $m_8 = m_1'$: *the sepal is not long.*

Given the original Iris data, we can verify that each concept $m \in M$ is a simple concept and (M, τ, X) is an AFS structure if for any $x, y \in X$, we define $\tau(x,y) = \{m | m \in M, (x,y) \in R_m\}$ (refer to (4.26)). For example, $\tau(x_1, x_1) = \{m_1, m_2, m_3, m_4, m_5, m_6, m_7, m_8\}$, since the sample x_1 has sepal length and width, and petal length and width. Similarly we can get $\tau(x_i, x_i)$, $i = 2, ..., 150$. For sample $x_4 = (4.6, 3.1, 1.5, 0.2)$ and sample $x_7 = (4.6, 3.4, 1.4, 0.3)$, we have $\tau(x_4, x_7) = \{m_1, m_3, m_5, m_7, m_8\}$, since the degrees of x_4 belonging to simple concepts *long sepal*, *long petal*, *not wide petal*, *not wide sepal*, *not long sepal* are larger than or equal to that of x_7. Similarly, we can determine $\tau(x_i, x_j)$ for any i, j according to the given feature values of the samples or the binary relation R_m of the simple concepts $m \in M$.

Let (M, τ, X) be the AFS structure of the Iris data set and η be the topological molecular lattice on the lattice $^*EM$ generated by all simple concepts in M,i.e., $\Lambda = \{\{m_1\}, ..., \{m_8\}\} \subseteq EM$. Let $(X, \mathscr{T}_\eta)$ be the topology space of X induced by the topological molecular lattice η. In order to compare the elementary differential degree of x, y in topology $\mathscr{T}_\eta$ with Euclidean distance in R^4. Let R_η be the fuzzy relation matrix derived by topology $\mathscr{T}_\eta$, where $R_\eta = S_\eta^r$, $(S_\eta^r)^2 = S_\eta^r$, $S_\eta = (s_{ij})$, $s_{ij} = s_e(x_i, x_j)$, the elementary similarity degree is defined by Definition 7.8. Let

R_E be the fuzzy relation matrix derived by the Euclidean distance where $R_E = S_E^k$, $(S_E^k)^2 = S_E^k$, $S_E = (e_{ij})$, $e_{ij} = 1 - (\sum_{1\leq k\leq 4}(x_{ik} - x_{jk})^2)^{\frac{1}{2}}$. Let fuzzy equivalence relation matrix $Q = (q_{ij}) = R_\eta$ or R_E, and for each threshold $\alpha \in [0,1]$, let Boolean matrix $Q_\alpha = (q_{ij}^a)$, $q_{ij}^a = 1 \Leftrightarrow q_{ij} \geq \alpha$. Since $R_\eta^2 = R_\eta$, $R_E^2 = R_E$, hence for each threshold $\alpha \in [0,1]$, Q_α is an equivalence relation Boolean matrix and it can yield a partition on X (refer to [23]). The following Figure 7.2 shows the clustering accuracy rates of fuzzy equivalence relation matrices R_η and R_E for threshold $\alpha \in [0.1,1]$. The accuracy is determined as follows: Suppose that the clusters C_1, C_2, ...,C_l are obtained by the fuzzy equivalence relation matrices R_η or R_E for some specific threshold α. Let $N_1 = \{1,2,...,50\}, N_2 = \{51,52,...,100\}, N_3 = \{101,102,...,150\}$. For $l \geq 3$, the clustering accuracy rate r is

$$r = \max_{1\leq i,j,k\leq l, i\neq j, i\neq k, j\neq i}\{\frac{|N_1\cap C_i|+|N_2\cap C_j|+|N_3\cap C_k|}{150}\};$$

For $l = 2$, let

$$|N_k \cap C_1| = \max_{1\leq u\leq 3}\{|N_u \cap C_1|\},\ 1 \leq k \leq 3,$$
$$|N_l \cap C_2| = \max_{1\leq u\leq 3, u\neq k}\{|N_u \cap C_2|\},\ 1 \leq l \leq 3,\ l \neq k,$$

$r = \frac{|N_k\cap C_1|+|N_l\cap C_2|}{150}$. For $l = 1$, let $r = \frac{1}{3}$. When threshold $\alpha = 0.8409$, the clustering accuracy rate of R_η is 90.67% (the best one), 9 clusters are obtained, the error clustering samples are $x_{23}, x_{42}, x_{69}, x_{71}, x_{73}, x_{78}, x_{84}, x_{88}, x_{107}, x_{109}, x_{110}, x_{118}, x_{132}, x_{135}$. When threshold $\alpha = 0.8905$, the clustering accuracy rate of R_E is 72.67% (the best one), and 29 clusters have been obtained. In Figures 7.2, we can observe that the elementary differential degrees defined by Definition 7.8 are better than those obtained for the Eculidean distance when it comes to the description of the difference of objects for this cluster analysis.

In order to compare the fuzzy equivalence relation matrices R_η with R_E, we show that the similar relation degrees of x_k to $\forall x \in X$,i.e., $R_\eta(x_k,x)$ and $R_E(x_k,x)$, $k = 71,72,...,130$ in Figures 7.3- 7.14 in Appendix B as examples. Since for Iris-data,

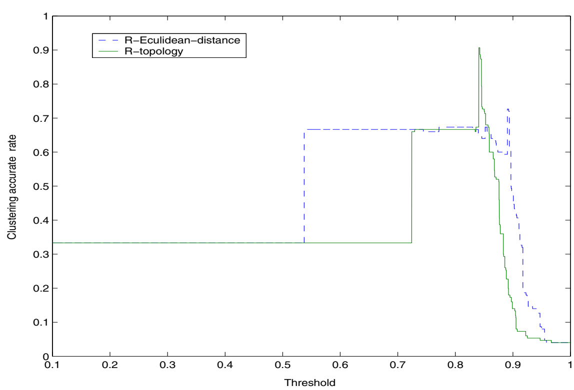


Fig. 7.2 The clustering accuracy rates of fuzzy equivalence relation matrices R_η and R_E for threshold $\alpha \in [0.1,1]$

the samples $x_{71},...,x_{130}$ are most difficult to be clustered, hence we show $R_\eta(x_k,x)$ and $R_E(x_k,x)$, $k=71,72,...,130$ in the figures. For Iris-data, samples 1:50 are cluster 1, i.e., iris-setosa; samples 51:100 are cluster 2, i.e., iris-versicolor; samples 101:150 are cluster 3, i.e., iris-virginica. In Figures 7.3 and Figures 7.4, for $x_{71},x_{72},...,x_{80}$ which are cluster 2, $R_\eta(x_k,x)$ and $R_E(x_k,x)$, $x\in X$ are shown. Compared with Figure 7.4, we can observe that in Figure 7.3, the similarity degrees of x_k to most samples in cluster 2 are larger than that of x_k to the samples in cluster 1,3. This implies that x_k are more similar to the samples in cluster 2 and $R_\eta(x_k,x)$, $k=71,...,80$, in Figure 7.3 are more clearly distinguish x_k from the samples in cluster 1, 3 than $R_E(x_k,x)$, $k=71,...,80$, in Figure 7.4. Similar phenomenon can be observed in Figures 7.4-7.14 and the others for $k=1,...,70,131,...,150$ which are not shown here. These examples show that the fuzzy equivalence relation matrix based on the topology is obviously better than that based on Eculidean distance for clustering of Iris data.

By the application of the AFS-FCM algorithm shown in (7.27) to the distance matrix $T=(t_{ij})_{150\times150}$, $t_{ij}=d_e(x_i,x_j)$ defined by Definition 7.8, the clustering accuracy rate is 86.67%. Using the function k means in MATLAB toolbox, which is based on the well known k-mean clustering algorithm [32], the clustering accuracy rate is 89.33%. And using the function FCM in MATLAB toolbox, which is based on the FCM clustering algorithm [1], the clustering accuracy rate is also 89.33%. Considering that the cluster centers of AFS-FCM must be the samples, i.e., $\{v_1,...,v_c\}\subseteq X$, while the cluster centers of FCM can be any vectors, i.e., $\{v_1,...,v_c\}\subseteq R^n$, the clustering accuracy rate of AFS-FCM is acceptable.

In some situations, it is difficult or impossible to describe some features of objects using real numbers, considering some inevitable errors and noise. For example, we do not describe a degree "*white hair*" of a person by counting the number of white hair on his head. But the order relations can be easily and accurately established by the simple comparisons of each pair of person's hair. In the framework of AFS theory, (M,τ,X) is determined by the binary relations R_m, $m\in M$ and the order relations are enough to establish the AFS structure of a data system. The membership functions and their logic operations of the fuzzy concepts in EM can be obtained by the AFS fuzzy logic system $(EM,\vee,\wedge,')$ and the AFS structure (M,τ,X). Therefore the AFS-FCM can be applied to the data set with the attributes described by mixed features such as numeric data, Boolean, order, even descriptors of human intuition, but FCM and k-mean can only be applied to the data set with the attributes described by numeric data.

The differential degrees and similarity degrees based on the topology induced by some fuzzy concepts are the criteria/metric human are using in their recognition process. This criteria/metric may not be the metric in the Euclidean space. The illustrative examples give some interpretations of the special topological structures on the AFS structures directly obtained by a given data set. Thus this approach also offers a new idea to data mining, artificial intelligence, pattern recognition,..., etc. Furthermore the real world examples demonstrate that this approach is promising.

Exercises

Exercise 7.1. Let M be a set and EM be the $^*EI$ algebra over M. For $A \subseteq M$, $\sum_{i\in I} A_i$, $\sum_{j\in J} B_j \in EM$, show the following assertions hold:

(1) $A \geq \sum_{i\in I} A_i$ and $A \geq \sum_{j\in J} B_j \Leftrightarrow A \geq \sum_{i\in I} A_i \vee \sum_{j\in J} B_j$.
(2) $A \geq \sum_{i\in I} A_i$ or $A \geq \sum_{j\in J} B_j \Leftrightarrow A \geq \sum_{i\in I} A_i \wedge \sum_{j\in J} B_j$.

Exercise 7.2. Let X and M be sets and (M,τ,X) be an strong relative AFS structure. Let η be a topological molecular lattice on EI algebra over M. For any $x \in X$, $\sum_{i\in I} A_i$, $\sum_{j\in J} B_j \in EM$, show the following assertions hold: for any $x \in X$

(1) If $\sum_{i\in I} A_i \geq \sum_{j\in J} B_j$ in EM, then $N^{\tau}_{\sum_{i\in I} A_i}(x) \supseteq N^{\tau}_{\sum_{j\in J} B_j}(x)$;
(2) $N^{\tau}_{\sum_{i\in I} A_i}(x) \cap N^{\tau}_{\sum_{j\in J} B_j}(x) = N^{\tau}_{\sum_{i\in I} A_i \wedge \sum_{j\in J} B_j}(x)$;
(3) $N^{\tau}_{\sum_{i\in I} A_i}(x) \cup N^{\tau}_{\sum_{j\in J} B_j}(x) = N^{\tau}_{\sum_{i\in I} A_i \vee \sum_{j\in J} B_j}(x)$.

Exercise 7.3. Proved that if η is a topological molecular on the lattice $^*EM$ and η is a dual idea of the lattice $^*EM$, then η is an elementary topological molecular lattice on the lattice $^*EM$.

Exercise 7.4. ([13]) Let X and M be sets and (M,τ,X) be a strong relative AFS structure. Let η be a topological molecular lattice on the lattice EM. If η is an elementary topological molecular lattice on the lattice EM and we define

$$\mathscr{B}_\eta = \left\{ N^{\tau}_{\sum_{i\in I} A_i}(x) \mid x \in X,\ \sum_{i\in I} A_i \in \eta \right\},$$

prove that $\mathscr{B}_\eta$ is a base for some topology of X.

Exercise 7.5. Let X and M be sets, (M,τ,X) be a strong relative AFS structure. Let η be a topological molecular lattice on the lattice EM and

$$\mathscr{L}_\eta = \left\{ \sum_{i\in I} a_i A_i \in EXM \mid \sum_{i\in I} A_i \in \eta, a_i \in \mathscr{T}_\eta \ for\ any\ i \in I \right\}. \tag{7.29}$$

Prove that $\mathscr{L}_\eta$ is a topological molecular lattice on the lattice EXM.

Exercise 7.6. Let X and M be sets. Let (M,τ,X) be a strong relative AFS structure and η be an elementary topological molecular lattice on the lattice EM. Let η be a topological molecular lattice on the lattice EM and the topological space $(X,\mathscr{T}_\eta)$ be the topology induced by η. Let S be the σ-algebra generated by $\mathscr{T}_\eta$ and $\mathscr{L}_\eta$ be the EI^2 topological molecular lattice on EXM induced by η. Show the following assertions hold.

(1) For any fuzzy concept $\sum_{i\in I} A_i \in \eta$, $\sum_{i\in I} A_i$ is a measurable concept under S ;
(2) The membership function defined by (7.14) is a continuous function from the topological space $(X,\mathscr{T}_\eta)$ to the topological molecular lattice $(EXM,\mathscr{L}_\eta)$.

Exercise 7.7. Prove that the following three conditions on a topological space are equivalent.

(1) The space is metrizable;
(2) The space is T_1 and regular, and the topology has a σ-locally finite base;
(3) The space is T_1 and regular, and the topology has a σ-discrete base.

Exercise 7.8. Let X and M be finite sets and (M,τ,X) be an AFS structure. Let η be an elementary topological molecular lattice on the lattice $^*EM$ and $(X,\mathscr{T}_\eta)$ be the topology space on X induced by η. Show for any $x,y\in X$ the following assertions hold.

(1) $d(x,x)=0$, $d(x,y)=d(y,x)$ and $s(x,y)=s(y,x)\leq s(x,x)$;
(2) $d_M(x,x)=0$, $d_M(x,y)=d_M(y,x)$ and $s_M(x,y)=s_M(y,x)\leq s_M(x,x)$;
(3) $d_e(x,x)=0$, $d_e(x,y)=d_e(y,x)$ and $s_e(x,y)=s_e(y,x)\leq s_e(x,x)$.

Open problems

Problem 7.1. Let X be a set and M be the set of simple concepts on X. Let (M,τ,X) be an AFS structure. If M is a finite set, then for any topological molecular lattice η on the EI algebra EM is also a topological molecular lattice on the $^*EI$ algebra $^*EM$. What are the relationships between the topological structures on X induced by η as a topological molecular on EM and that induced by η as a topological molecular on $^*EM$?

Problem 7.2. It is clear that $\mathscr{T}_\eta$ the topology on X is determined based on the distribution of raw data and the chosen set of fuzzy concepts $\eta\subseteq EM$ and it is an abstract geometry relation among the objects in X under the considering fuzzy concepts, i.e., η.

1. What are the interpretations of the special topological structures on X obtained from given database?
2. What are the topological structures associating with the essential nature of database?

Problem 7.3. Let X and M be sets. Let (M,τ,X) be a strong relative AFS structure and η be an elementary topological molecular lattice on the lattice EM. Let η be a topological molecular lattice on the lattice EM and the topological space $(X,\mathscr{T}_\eta)$ be the topology induced by η.

(1) How to induce a topological molecular lattice $\mathscr{L}_\eta^2$ on the lattices $^*EXMM$, $EXMM$ and a topological molecular lattice $\mathscr{L}_\eta^1$ on the lattices $^*E^\#X$, $E^\#X$?
(2) Are the membership functions defined on the lattices $EXMM$ by (5.12) and $E^\#X$ by (5.13) continuous from the topological space $(X,\mathscr{T}_\eta)$ to the topological molecular lattices $(EXMM,\mathscr{L}_\eta^2)$, $(^*EXMM,\mathscr{L}_\eta^2)$, $(^*E^\#X,\mathscr{L}_\eta^1)$, $(E^\#X,\mathscr{L}_\eta^1)$?

Problem 7.4. With a metric in the topological space on X, solving the pattern recognition problems will be possible for the database with various data types. Though we can have different choices from the topological theory for the metrics, what is suitable metric for this data of the pattern recognition problem?

Problem 7.5. Are the similarity degrees defined by Definition 7.7, Definition 7.8 and Definition 7.9 equivalent?

Problem 7.6. Let $(X, \mathscr{T}_\eta)$ be the topological space induced by η. Where η is the topological molecular lattice generated by some fuzzy concepts in EM. So far, we cannot obtain the differential degree and the similarity degree if η is the topological molecular lattice generated by more than 12 fuzzy concepts in EM. The more effective algorithm for the computation of the differential degree and the similarity degree in $(X, \mathscr{T}_\eta)$ are the most required.

Appendix A

Table 7.4 Evaluate results of 30 companies

Company	Factor						
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
C1	H	H	H	B.H & VH	VH	B.M & H	B.M & H
C2	H	H	B.M & H	B.M & H	H	B.M & H	B.M & H
C3	H	H	B.M & H	H	H	B.M & H	B.M & H
C4	H	B.H & VH	H	B.H & VH	H	B.H & VH	B.M & H
C5	H	B.H & VH	H	B.H & VH	B.H & VH	B.H & VH	B.M & H
C6	H	H	B.M & H	M	B.M & H	B.M & H	B.M & H
C7	H	H	M	B.H & VH	B.M & H	B.M & H	M
C8	B.M & H	H	M	B.M & H	B.M & H	H	B.L& M
C9	B.M & H	H	H	H	B.H & VH	B.M & H	M
C10	H	VH	H	B.M & H	B.H & VH	H	B.M & H
C11	M	H	M	H	B.M & H	B.H & VH	B.H & VH
C12	VH	VH	H	B.H & VH	VH	H	M
C13	B.M & H	H	B.M & H	H	B.H & VH	B.M & H	B.M & H
C14	H	H	B.M & H	B.H & VH	B.H & VH	H	M
C15	H	H	H	H	B.H & VH	VH	M
C16	H	H	H	B.H & VH	B.H & VH	B.H & VH	B.H & VH
C17	B.M & H	H	M	H	B.H & VH	B.M & H	B.M & H
C18	M	H	B.M & H	B.H & VH	B.H & VH	M	M
C19	B.M & H	H	B.M & H	B.M & H	VH	B.M & H	B.H & VH
C20	B.M & H	M	M	H	B.H & VH	B.M & H	B.L& M
C21	B.H & VH	VH	B.H & VH	B.H & VH	B.H & VH	B.H & VH	VH
C22	H	B.H & VH	B.M & H	B.H & VH	B.H & VH	H	M
C23	B.H & VH	VH	H	B.H & VH	H	B.M & H	B.M & H
C24	H	B.M & H	M	M	B.H & VH	M	M
C25	VH	VH	H	B.H & VH	B.H & VH	B.H & VH	B.M & H
C26	H	M	H	B.H & VH	B.H & VH	B.H & VH	L
C27	B.M & H	B.M & H	H	H	B.H & VH	H	B.M & H
C28	B.H & VH	B.H & VH	B.M & H	H	B.H & VH	B.M & H	B.H & VH
C29	H	B.M & H	M	H	B.M & H	B.M & H	L
C30	H	B.H & VH	H	B.H & VH	H	H	M

Appendix B

The following figures show the plots of $R_\eta(x_k,x)$ and $R_E(x_k,x), \forall x \in X. k=71,\ldots,130$.

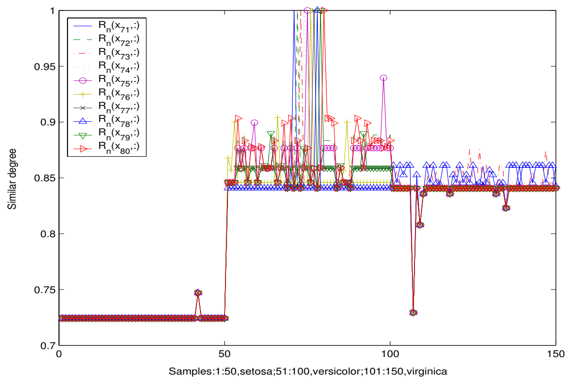


Fig. 7.3 The degrees of similarity relation of x_k to $x \in X$ based on topology, i.e., $R_\eta(x_k,x)$, $k = 71,72,\ldots,80$

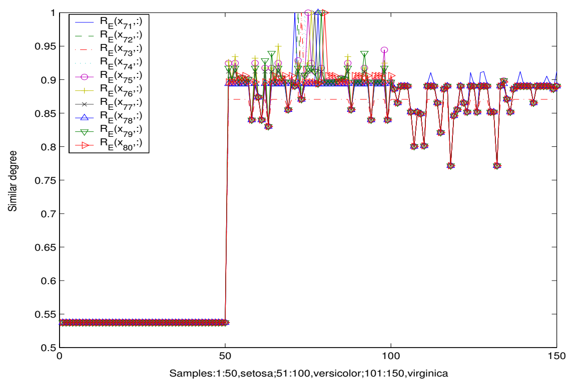


Fig. 7.4 The degrees of similarity relation of x_k to $x \in X$ based Euclidean distance, i. e., $R_E(x_k,x)$, $k = 71,72,\ldots,80$

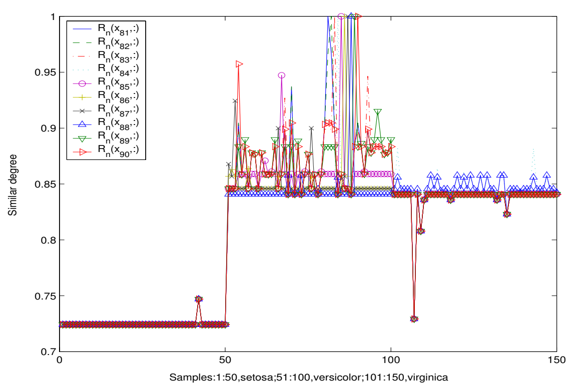


Fig. 7.5 The degrees of similarity relation of x_k to $x \in X$ based on topology, i.e., $R_\eta(x_k,x)$, $k = 81,82,\ldots,90$

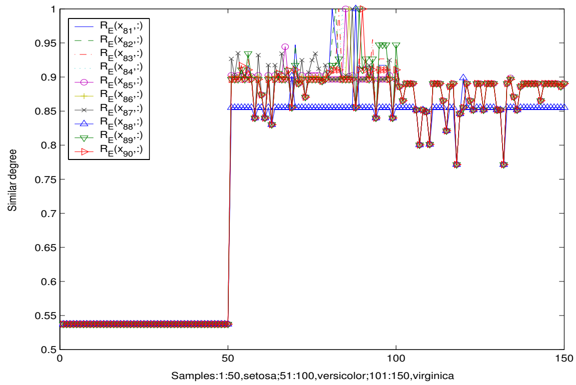


Fig. 7.6 The degrees of similarity relation of x_k to $x \in X$ based Euclidean distance, i. e., $R_E(x_k, x)$, $k = 81, 82, ..., 90$

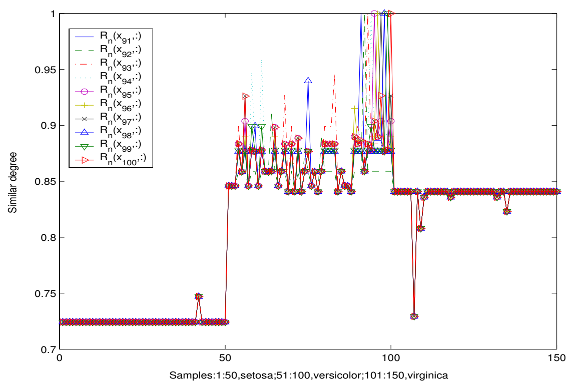


Fig. 7.7 The degrees of similarity relation of x_k to $x \in X$ based on topology, i.e., $R_\eta(x_k, x)$, $k = 91, 92, ..., 100$

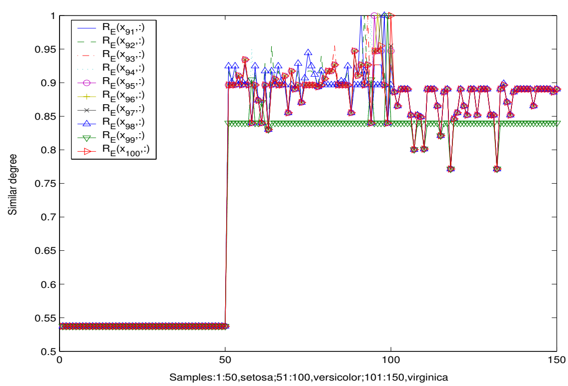


Fig. 7.8 The degrees of similarity relation of x_k to $x \in X$ based Euclidean distance, i. e., $R_E(x_k, x)$, $k = 91, 92, ..., 100$

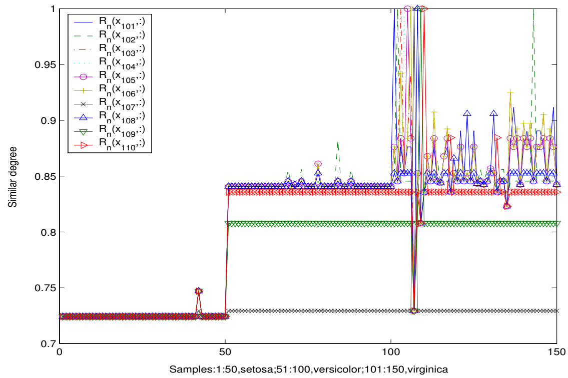


Fig. 7.9 The degrees of similarity relation of x_k to $x \in X$ based on topology, i.e., $R_\eta(x_k,x)$, $k = 101, 102, \ldots, 110$

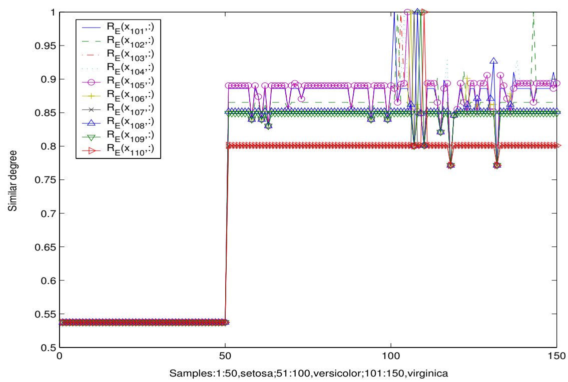


Fig. 7.10 The degrees of similarity relation of x_k to $x \in X$ based Euclidean distance, i. e., $R_E(x_k,x)$, $k = 101, 102, \ldots, 110$

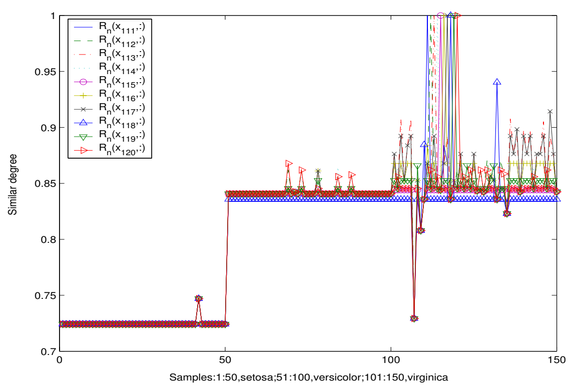


Fig. 7.11 The degrees of similarity relation of x_k to $x \in X$ based on topology, i.e., $R_\eta(x_k,x)$, $k = 111, 112, \ldots, 120$

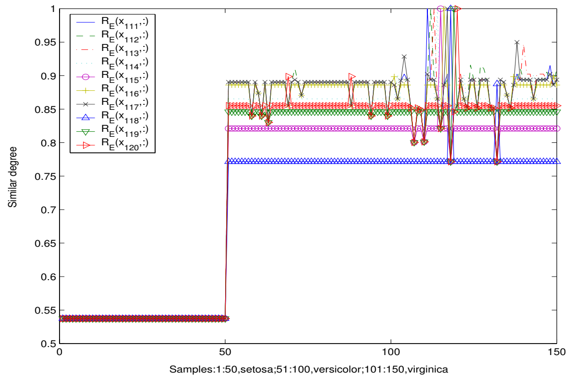


Fig. 7.12 The degrees of similarity relation of x_k to $x \in X$ based Euclidean distance, i. e., $R_E(x_k, x)$, $k = 111, 112, \ldots, 120$

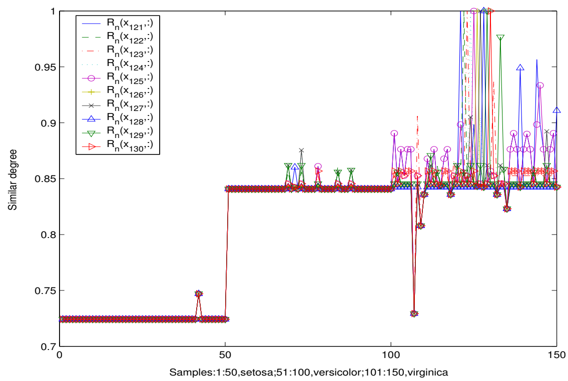


Fig. 7.13 The degrees of similarity relation of x_k to $x \in X$ based on topology, i.e., $R_\eta(x_k, x)$, $k = 121, 122, \ldots, 130$

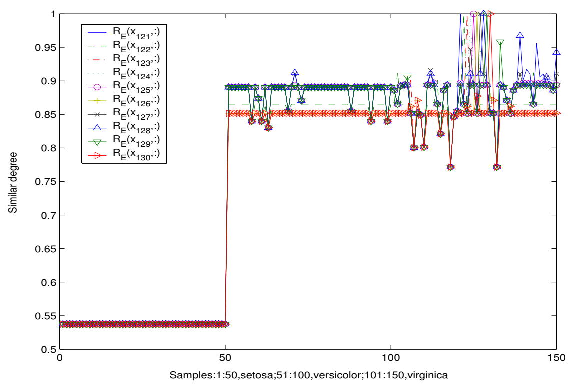


Fig. 7.14 The degrees of similarity relation of x_k to $x \in X$ based Euclidean distance, i. e., $R_E(x_k, x)$, $k = 121, 122, \ldots, 130$

References

1. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum, New York (1981)
2. Dunn, J.C.: A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J. Cybernet 3(3), 32–57 (1974)

3. Dubois, D., Prade, H.: The three semantics of fuzzy sets. Fuzzy Sets and Systems 90, 141–150 (1997)
4. Bezdek, J.C.: A convergence theorem for the fuzzy ISODATA clustering algorithms. IEEE Trans. Pattern Anal. Machine Intell. PAMI 2(1), 1–8 (1980)
5. Baum, J.D.: Elements of Point Set Topology. Prentice-Hall, Englewood Cliffs (1964)
6. Booker, J.M., Mayer, M.A.: Common problem in the elicitation and analysis of expert opinion affecting probabilistic safety assessments. In: CSNI Workshop on PSA Application and Limitation, Santa Fe, NM, September 4-6, 1990, pp. 353–368 (1990)
7. Graver, J.E., Watkins, M.E.: Combinatorics with Emphasis on the Theory of Graphs. Springer, New York (1977)
8. Gbyym, S.H.: A semi-linguistic fuzzy approach to multifactor decision-making: Application to aggregation of experts judgments. Annals of Nuclear Energy 26, 1097–1112 (1999)
9. Kelley, J.L.: General Topology. Springer, New York (1955)
10. Kim, K.H.: Boolean matrix Theory and Applications. Marcel Dekker, New York (1982)
11. Keller, J.M., Gray, M.R., Givens Jr., J.A.: A fuzzy K-nearest neighbors algorithm. IEEE Trans. Systems Man, Cybernet. SMC 15(4), 580–585 (1985)
12. Liu, X.D.: The Fuzzy Theory Based on AFS Algebras and AFS Structure. Journal of Mathematical Analysis and Applications 217, 459–478 (1998)
13. Liu, X.D.: The Topology on AFS Algebra and AFS Structure. Journal of Mathematical Analysis and Applications 217, 479–489 (1998)
14. Liu, X.D.: A new fuzzy model of pattern recognition and hitch diagnoses of complex systems. Fuzzy Sets and Systems 104, 289–297 (1999)
15. Liu, X.D.: A New Mathematical Axiomatic System of Fuzzy Sets and Systems. Journal of Fuzzy Mathematics 3, 559–560 (1995)
16. Liu, X.D.: The Fuzzy Sets and Systems Based on AFS Structure, EI Algebra and EII algebra. Fuzzy Sets and Systems 95, 179–188 (1998)
17. Liu, X.D.: Two Algebra Structures of AFS structure. Journal of Fuzzy Mathematics 3, 561–562 (1995)
18. Liu, X.D., Wang, W., Chai, T.Y.: The Fuzzy Clustering Analysis Based on AFS Theory. IEEE Transactions on Systems, Man and Cybernetics Part B 35(5), 1013–1027 (2005)
19. Liu, X.D., Zhang, Q.L.: The Fuzzy Cognitive Maps Based on AFS Fuzzy Logic. Dynamics of Continuous, Discrete and Impulsive Systems 11(5-6), 787–796 (2004)
20. Liu, X.D., Zhu, K.J., Huang, H.Z.: The Representations of Fuzzy Concepts Based on the Fuzzy Matrix Theory and the AFS Theory. In: IEEE International Symposium on Intelligent Control, Texas, USA, October 5-8, pp. 1006–1011 (2003)
21. Liu, X.D., Pedrycz, W., Zhang, Q.L.: Axiomatic Fuzzy Set Logic. In: The Proceedings of 2003 IEEE International Conference on Fuzzy Systems, St. May Louis Missouri, USA, vol. 1, pp. 55–60 (2003)
22. Liu, X.D., Pedrycz, W.: The Development of Fuzzy Decision Trees in the Framework of Axiomatic Fuzzy Set Logic. Applied Soft Computing 7, 325–342 (2007)
23. Liu, X.D.: The Structure of Fuzzy Matrices. Journal of Fuzzy Mathematics 2, 311–325 (1994)
24. Liu, X.D., Chai, T.Y., Wang, W.: Approaches to the Representations and Logic Operations for Fuzzy Concepts in the Framework of Axiomatic Fuzzy Set Theory I. Information Sciences 177, 1007–1026 (2007)
25. Liu, X.D., Wang, W., Chai, T.Y.: Approaches to the Representations and Logic Operations for Fuzzy Concepts in the Framework of Axiomatic Fuzzy Set Theory II. Information Sciences 177, 1027–1045 (2007)

26. Liu, X.D., Zhang, L.S., Chai, T.Y.: The Fuzzy Clustering Analysis Based on Topology Space (manuscript)
27. Liang, G.S., Chou, T.Y., Han, T.C.: Cluster analysis based on fuzzy equivalence relation. European Journal of Operational Research 166, 160–171 (2005)
28. Liang, G.S., Wang, M.J.: A fuzzy multi-criteria decision-making method for facility site selection. International Journal of Product Research 29(11), 2313–2330 (1991)
29. Liou, T.S., Wang, M.J.J.: Subjective assessment of mental workload-a fuzzy linguistic multi-criteria approach. Fuzzy Sets and Systems 62, 155–165 (1994)
30. Mertz, J., Murphy, P.M.: UCI repository of machine learning databases, http://www.ics.uci.edu/pub/machinelearning-data-bases
31. Pu, B.M., Liu, Y.M.: Fuzzy Topology I, Neighborhood structure of Fuzzy Point and Moore-Smith Convergence. Journal of Mathematical Analysis and Applications 76, 571–599 (1980)
32. Seber, G.A.F.: Multivariate Observations. Wiley, New York (1984)
33. Wang, G.J.: Theory of topological molecular lattices. Fuzzy Sets and Systems 47, 351–376 (1992)
34. Wu, K.L., Yang, M.S.: Alternative c-means clustering algorithms. Pattern Recognition 35, 2267–2278 (2002)
35. Zadeh, L.A.: Fuzzy Sets. Information and Control 8, 338–353 (1965)
36. Zhang, Y.J., Liang, D.Q., Tong, S.C.: On AFS Algebra Part I, II. Information Sciences 167, 263–286, 287-303 (2004)

Chapter 8
AFS Formal Concept and AFS Fuzzy Formal Concept Analysis

In this chapter, based on the original idea of Wille of formal concept analysis and the AFS (Axiomatic Fuzzy Set) theory, we presents a rigorous mathematical treatment of fuzzy formal concept analysis referred to as an AFS Formal Concept Analysis (AFSFCA). It naturally augments the existing formal concepts to fuzzy formal concepts, with the aim of deriving their mathematical properties and applying them in the exploration and development of knowledge representation. Compared with other fuzzy formal concept approaches such as the L-concept [1, 2] and the fuzzy concept [48], the main advantages of AFSFCA are twofold. One is that the original data and facts are the only ones required to generate AFSFCA lattices thus human interpretation is not required to define the fuzzy relation or the fuzzy set on $G \times M$ to describe the uncertainty dependencies between the objects in G and the attributes in M. Another advantage comes with the fact that is that AFSFCA is more expedient and practical to be directly applied to real world applications.

FCA(Formal Concept Analysis) was introduced by Rudolf Wille in 1980s [10]. In the past two decades, FCA has been a topic of interest both from the conceptual as well as applied perspective. In artificial intelligence community, FCA is used as a knowledge representation mechanism [15, 50, 51] as well as it can support the ideas of a conceptual clustering [4, 40] for Boolean concepts. Traditional FCA-based approaches are hardly able to represent vague information. To tackle with this problem, fuzzy logic can be incorporated into FCA to facilitate handling uncertainty information for conceptual clustering and concept hierarchy generation. Pollandt [42], Burusco and Fuentes-Gonza lez [3], Huynh and Nakamori [16], and Belohlavek [1, 2] have proposed the use of the L-Fuzzy context as an attempt to combine fuzzy logic with the FCA. The primary notion in this investigation is that of a fuzzy context (L-context): it comes as a triple $(G,M,\mathbb{I})$, where G and M are sets interpreted as the set of objects (G) and the set of attributes (M), and $\mathbb{I} \in L^{G\times M}$ is a fuzzy relation between G and M. The value $\mathbb{I}(g,m) \in L$ (L is a lattice) is interpreted as the truth value of the fact "the object $g \in G$ has the attribute $m \in M$". In accordance with the Port-Royal definition, a (formal) fuzzy concept (L-concept) is a pair (A,B), $A \in L^G, B \in L^M$, A plays the role of the extent (fuzzy set of objects which determine the concept), B plays the role of the intent (fuzzy set of attributes

X. Liu and W. Pedrycz: Axiomatic Fuzzy Set Theory and Its Applications, STUDFUZZ 244, pp. 303–349.
springerlink.com

which determine the concept). The L-Fuzzy context uses linguistic variables, which are linguistic terms associated with fuzzy sets, to represent uncertainty in the context. However, human interpretation is required to define the linguistic variables and the fuzzy relation between G and M (i.e., $\mathbb{I} \in L^{G\times M}$). Moreover, the fuzzy concept lattice generated from the L-fuzzy context usually causes a combinatorial explosion of concepts as compared to the traditional concept lattice.

Tho, Hui, Fong, and Cao [48] proposed a technique that combines fuzzy logic and FCA giving rise to the idea of the Fuzzy Formal Concept Analysis (FFCA), in which the uncertainty information is directly represented by membership grades. The primary notion is that of a fuzzy context: it is a triple $(G,M,\mathbb{I})$, where G is a set of objects, M is a set of attributes, and $\mathbb{I}$ is a fuzzy set on domain $G\times M$. Each relation $(g,m)\in\mathbb{I}$ has a membership value $\mu_{\mathbb{I}}(g,m)$ in $[0,1]$. Compared to the fuzzy concept lattice generated from the L-fuzzy context, the fuzzy concept lattice generated by using FFCA is simpler in terms of the number of formal concepts. However, human interpretation is still referring to it as the required to define the membership function of the fuzzy set $\mathbb{I}$ for FFCA. In real world applications, just based on human interpretation, it is very difficult to properly define the fuzzy set $\mathbb{I}$ to describe the uncertainty relations between the objects and the attributes.

In order to cope with the above problems, we propose a new framework of fuzzy formal concept analysis based on the AFS (Axiomatic Fuzzy Set) theory [18, 54] referring to it as the AFS Formal Concept Analysis (AFSFCA,for brief). In the proposed AFSFCA, each fuzzy complex attribute in EM, which plays the role of the intent of an AFS formal concept, corresponds to a fuzzy set, which is automatically determined by the AFS structure and the AFS algebra via what we have discussed in Chapter 4, 5, and plays the role of the extent of the AFS formal concept. Thus the original data and facts are only required to generate AFSFCA lattices and human interpretation is not required to define the fuzzy relation or the fuzzy set $\mathbb{I}$ on $G\times M$ to describe the uncertainty relations between the objects and the attributes. Compared with the fuzzy concept lattices based on L-fuzzy context, the fuzzy concept lattice generated using AFSFCA will be simpler in terms of the number of formal concepts. Compared with FFCA, the fuzzy concept lattice generated using AFSFCA will be richer in expression, more relevant and practical.

8.1 Concept Lattices and AFS Algebras

In Chapter 4, 5, various kinds of representations and logic operations for fuzzy concepts in EM have been extensively discussed in the framework of AFS theory, in which the membership functions and their logic operations are automatically determined in an algorithmic fashion by taking advantage of the existing distribution of the original data. The purpose of this section is to extend these approaches by combining the AFS and FCA theories.

Let us briefly recall the Wille's notion of formal concept [57]: The basic notions of FCA are those of a formal context and a formal concept. A *formal context* is a triple (G,M,I) where G is a set of objects, M is a set of features or attributes, and I is a binary relation from G to M, i.e., $I\subseteq G\times M$. gIm, which is also written as

$(g,m) \in I$, denotes that the object g possesses the feature m. An example of a context (G,M,I) is shown in Table 8.1, where $G = \{g_1, g_2, ..., g_6\}$ and $M = \{m_1, m_2, ..., m_5\}$. An "×" is placed in the ith row and jth column to indicate that $(g_i, m_j) \in I$. For a set of objects $A \subseteq G$, $\beta(A)$ is defined as the set of features shared by all the objects in A, that is,

$$\beta(A) = \{m \in M | (g,m) \in I, \forall g \in A\}. \tag{8.1}$$

Similarly, for $B \subseteq M$, $\alpha(B)$ is defined as the set of objects that possesses all the features in B, that is,

$$\alpha(B) = \{g \in G | (g,m) \in I, \forall m \in B\}. \tag{8.2}$$

The pair (β, α) is a *Galois connection* between the power sets of G and M. For more information on Galois connections, interested readers are referred to [57]. In this chapter, the symbols α, β always denote the Galois connection defined by (8.1) and (8.2). In the FCA, concept lattice, or Galois lattice is the core of its mathematical theory and can be used as an effective tool for symbolic data analysis and knowledge acquisition.

Table 8.1 Example of a context

	m_1	m_2	m_3	m_4	m_5
g_1	×		×	×	×
g_2	×	×	×	×	
g_3	×	×	×	×	
g_4	×				×
g_5	×				×
g_6	×				×

Lemma 8.1. *Let (G,M,I) be a context. Then the following assertions hold:*

(1) for $A_1, A_2 \subseteq G$, $A_1 \subseteq A_2$ implies $\beta(A_1) \supseteq \beta(A_2)$ and for $B_1, B_2 \subseteq M$, $B_1 \subseteq B_2$ implies $\alpha(B_1) \supseteq \alpha(B_2)$;

(2) $A \subseteq \alpha(\beta(A))$ and $\beta(A) = \beta(\alpha(\beta(A)))$ for all $A \subseteq G$, and $B \subseteq \beta(\alpha(B))$ and $\alpha(B) = \alpha(\beta(\alpha(B)))$ for all $B \subseteq M$.

Its proof is left to the reader.

Definition 8.1. ([51]) A *formal concept* in the context (G,M,I) is a pair (A,B) such that $\beta(A) = B$ and $\alpha(B) = A$, where $A \subseteq G$ and $B \subseteq M$.

In other words, a formal concept is a pair (A,B) of two sets $A \subseteq G$ and $B \subseteq M$, where A is the set of objects that possesses all the features in B and B is the set of features common to all the objects in A. In what follows, a formal concept (A,B) in (G,M,I) briefly noticed as $(A,B) \in (G,M,I)$. The set A is called the *extent of the concept* and B is called its *intent*. If we review $B \subseteq M$ as a new attribute generated by the "*and*" of all attributes in B like that in [28], then A is the set of objects that possess the new attribute B. The adjective "*formal*" in formal concept means that the concept is a rigorously defined mathematical object [8]. From the point of view of logic, the intent of a formal concept can be seen as a conjunct of features that each object of the extent must possess. For any given context (G,M,I), neither every subset of G nor every subset of M corresponds to a concept.

Definition 8.2. ([51]) A set $B \subseteq M$ is called a *feasible intent* if set B is the intent of the unique formal concept $(\alpha(B),B)$. Similarly, a set $A \subseteq G$ is called a *feasible extent* if A is the extent of the unique formal concept $(A,\beta(A))$. A set X is called a *feasible set* if it is either a feasible extent or a feasible intent. Otherwise, X is called *non-feasible*.

An important notion in FCA is that of a concept lattice, which makes it possible to depict the information represented in a context as a complete lattice. Let $\mathscr{L}(G,M,I)$ denote the set of all formal concepts of the context (G,M,I). An order relation on $\mathscr{L}(G,M,I)$ is defined as follows [51]. Let (A_1,B_1) and (A_2,B_2) be two concepts in $\mathscr{L}(G,M,I)$, then $(A_1,B_1) \leq (A_2,B_2)$ if and only if $A_1 \subseteq A_2$ (or equivalently $B_1 \supseteq B_2$). The formal concept (A_1,B_1) is called a *sub formal concept of the formal concept* (A_2,B_2) and (A_2,B_2) is called a *super formal concept* of (A_1,B_1). The fundamental theorem of Wille about concept lattices, states that $(\mathscr{L}(G,M,I),\vee,\wedge)$ is a complete lattice called the *concept lattice of the context* (G,M,I).

Lemma 8.2. *(Will's Lemma) Let (G,M,I) be a context and $\mathscr{L}(G,M,I)$ denote the set of all formal concepts of the context (G,M,I). Then*

$$\mathscr{L}(G,M,I) = \{(\alpha(B),\beta(\alpha(B))) \mid B \subseteq M\}. \tag{8.3}$$

Proposition 8.1. *Let (G,M,I) be a context. Then for any $A_i \subseteq G, i \in I$, $B_j \subseteq M$, $j \in J$,*

$$\alpha\left(\bigcup_{j\in J} B_j\right) = \bigcap_{j\in J} \alpha(B_j),$$
$$\beta\left(\bigcup_{i\in I} A_i\right) = \bigcap_{i\in I} \beta(A_i).$$

Proof. By the definitions, for any $g \in \alpha(\bigcup_{j\in J} B_j)$, we have

$$\begin{aligned} g \in \alpha\left(\bigcup_{j\in J} B_j\right) &\Leftrightarrow \forall m \in \bigcup_{j\in J} B_j,\ (g,m) \in I \\ &\Leftrightarrow \forall j \in J,\ \forall m \in B_j,\ (g,m) \in I \\ &\Leftrightarrow \forall j \in J,\ g \in \alpha(B_j) \\ &\Leftrightarrow g \in \bigcap_{j\in J} \alpha(B_j). \end{aligned}$$

Similarly, we can prove that $\beta\left(\bigcup_{i\in I} A_i\right) = \bigcap_{i\in I} \beta(A_i)$. □

Theorem 8.1. *(Fundamental Theorem of FCA) Let (G,M,I) be a context. Then $(\mathscr{L}(G,M,I),\vee,\wedge)$ is a complete lattice in which suprema and infima are given as follows: for any formal concepts $(A_j,B_j)\in\mathscr{L}(G,M,I)$, $j\in J$,*

$$\bigvee_{j\in J}(A_j,B_j)=\left(\gamma_G\left(\bigcup_{j\in J}A_j\right),\ \bigcap_{j\in J}B_j\right),\tag{8.4}$$

$$\bigwedge_{j\in J}(A_j,B_j)=\left(\bigcap_{j\in J}A_j,\ \gamma_M\left(\bigcup_{j\in J}B_j\right)\right),\tag{8.5}$$

where $\gamma_G=\alpha\cdot\beta$, $\gamma_M=\beta\cdot\alpha$.

Proof. First, let us explain the formula for the infimum. Since $A_j=\alpha(B_j)$, for each $j\in J$,

$$\left(\bigcap_{j\in J}A_j,\gamma_M\left(\bigcup_{j\in J}B_j\right)\right)$$

by Proposition 8.1 it can be transformed into

$$\left(\alpha\left(\bigcup_{j\in J}B_j\right),\gamma_M\left(\bigcup_{j\in J}B_j\right)\right),$$

i.e., it has the form $(\alpha(X),\gamma_M(X))$ and is therefore a concept. That this can only be the infimum, i.e., the largest common subconcept of the concepts (A_j,B_j), follows immediately from the fact that the extent of this concept is exactly the intersection of the extents of (A_j,B_j). The formula for the supremum is substantiated correspondingly. Thus, we have proven that $(\mathscr{L}(G,M,I),\vee,\wedge)$ is a complete lattice. □

In what follows, we denote the subsets of G with small letters and the subsets of M with capital letters in order to distinguish subsets of objects in G from subsets of attributes in M.

By sets G,M, we can establish the *EII* algebra over G,M and $(EGM,\vee,\wedge)$ is a completely distributivity lattice. Now, we study the relationship between the lattice $(\mathscr{L}(G,M,I),\vee,\wedge)$ and the lattice $(EGM,\vee,\wedge)$. We define $\alpha(EM)$ a sub sets of EGM as follows

$$\alpha(EM)=\left\{\gamma\in EGM\mid \gamma=\sum_{i\in I}b_iB_i,\ \forall i\in I,\ b_i=\alpha(B_i)\right\}.\tag{8.6}$$

Lemma 8.3. *Let (G,M,I) be a context. Then $\alpha(EM)$ is a sub EII algebra of EGM, i.e. $k\in K$, $\zeta_k=\sum_{i\in I_k}b_{ki}B_{ki}\in\alpha(EM)$, $\bigvee_{k\in K}\zeta_k,\bigwedge_{k\in K}\zeta_k\in\alpha(EM)$, and $(\alpha(EM),\vee,\wedge)$ is also a completely distributivity lattice.*

Proof. It could be easily verified that $\bigvee_{k\in K}\zeta_k\in\alpha(EM)$. Since EGM is a completely distributivity lattice, hence

$$\bigwedge_{k\in K}\zeta_k=\sum_{f\in\Pi_{k\in K}I_k}(\bigcap_{k\in K}b_{kf(k)}\bigcup_{k\in K}B_{kf(k)}).$$

By Proposition 8.1 and $\alpha(B_{kj})=b_{kj}$, for any $k\in K, j\in I_j$, we have

$$\alpha\left(\bigcup_{k\in K}B_{kf(k)}\right)=\bigcap_{k\in K}\alpha(B_{kf(k)})=\bigcap_{k\in K}b_{kf(k)}.$$

Therefore $\bigwedge_{k\in K}\zeta_k\in\alpha(EM)$. Because $(EGM,\vee,\wedge)$ is a completely distributivity lattice, $(\alpha(EM),\vee,\wedge)$ is also a completely distributivity lattice. □

Theorem 8.2. *Let (G,M,I) be a context. p_I is a homomorphism from lattice $(EM,\vee,\wedge)$ to lattice $(\mathscr{L}(G,M,I),\vee,\wedge)$ provided p_I is defined as follows: for any $\sum_{i\in I}B_i\in EM$,*

$$p_I\left(\sum_{i\in I}B_i\right)=\bigvee_{i\in I}(\alpha(B_i),\beta\cdot\alpha(B_i))=\left(\alpha\cdot\beta(\bigcup_{i\in I}\alpha(B_i)),\ \bigcap_{i\in I}\beta\cdot\alpha(B_i)\right). \quad (8.7)$$

Proof. By Lemma 8.2, for any $\sum_{i\in I}B_i\in EM$, one knows that $\forall i\in I,(\alpha(B_i),\beta\cdot\alpha(B_i))\in\mathscr{L}(G,M,I)$. Since lattice $(\mathscr{L}(G,M,I),\leq)$ is a complete lattice, hence $\forall\sum_{i\in I}B_i\in EM$,

$$\begin{aligned}p_I\left(\sum_{i\in I}B_i\right)&=\left(\alpha\cdot\beta(\bigcup_{i\in I}\alpha(B_i)),\ \bigcap_{i\in I}\beta(\alpha(B_i))\right)\\&=\bigvee_{i\in I}(\alpha(B_i),\ \beta\cdot\alpha(B_i))\in\mathscr{L}(G,M,I).\end{aligned}$$

Next, we prove that p_I is a map from EM to $\mathscr{L}(G,M,I)$. Suppose $\sum_{i\in I_1}B_{1i}=\sum_{i\in I_2}B_{2i}\in EM$. By Lemma 8.1, one has $\forall i\in I_1,\exists k\in I_2$ such that $B_{1i}\supseteq B_{2k}\Rightarrow\alpha(B_{1i})\subseteq\alpha(B_{2k})$ and $\forall j\in I_2,\exists l\in I_1$ such that $B_{2j}\supseteq B_{1l}\Rightarrow\alpha(B_{2j})\subseteq\alpha(B_{1l})$. Therefore $\cup_{i\in I_1}\alpha(B_{1i})=\cup_{j\in I_2}\alpha(B_{2i})$ and

$$\alpha\cdot\beta\left(\bigcup_{i\in I_1}\alpha(B_{1i})\right)=\alpha\cdot\beta\left(\bigcup_{j\in I_1}\alpha(B_{2j})\right).$$

Since both

$$\left((\alpha\cdot\beta\bigcup_{i\in I_1}\alpha(B_{1i})),\bigcap_{i\in I_1}\beta\cdot\alpha(B_{1i})\right)$$

and

$$\left(\alpha\cdot\beta(\bigcup_{i\in I_2}\alpha(B_{2i})),\bigcap_{i\in I_2}\beta\cdot\alpha(B_{2i})\right)$$

are formal concepts in (G,M,I), hence

$$(\alpha\cdot\beta(\cup_{i\in I_1}\alpha(B_{1i})),\cap_{i\in I_1}\beta\cdot\alpha(B_{1i})) = (\alpha\cdot\beta(\cup_{i\in I_2}\alpha(B_{2i})),\cap_{i\in I_2}\beta\cdot\alpha(B_{2i})),$$
$$p_I\left(\sum_{i\in I_1}B_{1i}\right) = p_I\left(\sum_{i\in I_2}B_{2i}\right).$$

For any $\zeta=\sum_{i\in I}A_i, \eta=\sum_{j\in J}B_j\in EM$, by (8.7), (8.4) and Proposition 8.1, we have

$$\begin{aligned}&p_I(\zeta\vee\eta)\\ &=(\alpha\cdot\beta[(\cup_{i\in I}\alpha(A_i))\cup(\cup_{j\in J}\alpha(B_j))],[(\cap_{i\in I}\beta\cdot\alpha(A_i))\cap(\cap_{j\in J}\beta\cdot\alpha(B_j))]).\\ &p_I(\zeta)\vee p_I(\eta)\\ &=(\alpha\cdot\beta(\cup_{i\in I}\alpha(A_i)),\cap_{i\in I}\beta\cdot\alpha(A_i))\vee(\alpha\cdot\beta(\cup_{j\in J}\alpha(B_j)),\cap_{j\in J}\beta\cdot\alpha(B_j))\\ &=(\alpha\cdot\beta[\alpha\cdot\beta(\cup_{i\in I}\alpha(A_i))\cup\alpha\cdot\beta(\cup_{j\in J}\alpha(B_j))],[(\cap_{i\in I}\beta\cdot\alpha(A_i))\cap(\cap_{j\in J}\beta\cdot\alpha(B_j))])\end{aligned}$$

Since both $p_I(\zeta\vee\eta)$ and $p_I(\zeta)\vee p_I(\eta)$ are formal concepts of the context (G,M,I), hence $p_I(\zeta\vee\eta)=p_I(\zeta)\vee p_I(\eta)$. By (8.7), we have

$$p_I(\zeta\wedge\eta)=p_I\left(\sum_{i\in I,j\in J}A_i\cup B_j\right)=\bigvee_{i\in I,j\in J}(\alpha(A_i\cup B_j),\beta\cdot\alpha(A_i\cup B_j)).$$

In addition, for any $i\in I, j\in J$, it follows by (8.5)

$$(\alpha(A_i),\beta\cdot\alpha(A_i))\wedge(\alpha(B_i),\beta\cdot\alpha(B_i))=(\alpha(A_i)\cap\alpha(B_j),\beta\cdot\alpha[\beta\cdot\alpha(A_i)\cup\beta\cdot\alpha(B_j)]).$$

By Proposition 8.1, we have $\alpha(A_i)\cap\alpha(B_j)=\alpha(A_i\cup B_j)$ and

$$\begin{aligned}\beta\cdot\alpha[\beta\cdot\alpha(A_i)\cup\beta\cdot\alpha(B_j)] &= \beta\cdot\alpha(\beta(\alpha(A_i))\cup\beta(\alpha(B_j)))\\ &=\beta(\alpha(\beta(\alpha(A_i)))\cap\alpha(\beta(\alpha(B_j))))\\ &=\beta(\alpha(A_i)\cap\alpha(B_j))\\ &=\beta\cdot\alpha(A_i\cup B_j)).\end{aligned}$$

Therefore for any $i\in I, j\in J$,

$$(\alpha(A_i\cup B_j),\beta\cdot\alpha(A_i\cup B_j))=(\alpha(A_i),\beta\cdot\alpha(A_i))\wedge(\alpha(B_i),\beta\cdot\alpha(B_i)).$$

and

$$\begin{aligned}p_I(\zeta\wedge\eta) &= \bigvee_{i\in I,j\in J}(\alpha(A_i\cup B_j),\beta\cdot\alpha(A_i\cup B_j))\\ &=\bigvee_{i\in I,j\in J}[(\alpha(A_i),\beta\cdot\alpha(A_i))\wedge(\alpha(B_i),\beta\cdot\alpha(B_i))]\\ &=\left[\bigvee_{i\in I}(\alpha(A_i),\beta\cdot\alpha(A_i))\right]\wedge\left[\bigvee_{j\in J}(\alpha(B_i),\beta\cdot\alpha(B_i))\right]\\ &=p_I(\zeta)\wedge p_I(\eta).\end{aligned}$$

This demonstrates that p_I is homomorphism. □

Theorem 8.3. *Let (G,M,I) be a context. p_I is homomorphism from lattice $(\alpha(EM),\vee,\wedge)$ to lattice $(\mathscr{L}(G,M,I),\vee,\wedge)$, if for any $\sum_{i\in I} b_iB_i \in \alpha(EM)$, p_I is defined as*

$$p_I\left(\sum_{i\in I} b_iB_i\right) = \bigvee_{i\in I}(b_i,\beta(b_i)) = \left(\alpha\cdot\beta(\bigvee_{i\in I} b_i),\ \bigcap_{i\in I}\beta(b_i)\right). \tag{8.8}$$

Proof. By Lemma 8.2 and (8.6), for any $\sum_{i\in I} b_iB_i \in \alpha(EM)$, one knows that $\forall i\in I$, $(b_i,\beta(b_i)) = (\alpha(B_i),\beta(\alpha(B_i))) \in \mathscr{L}(G,M,I)$. This implies that

$$(\alpha\cdot\beta(\cup_{i\in I}b_i),\cap_{i\in I}\beta(b_i)) = \bigvee_{i\in I}(b_i,\beta(b_i)) \in \mathscr{L}(G,M,I).$$

Now, we prove that p_I is a map from $\alpha(EM)$ to $\mathscr{L}(G,M,I)$. Suppose $\sum_{i\in I_1} b_{1i}B_{1i} = \sum_{i\in I_2} b_{2i}B_{2i} \in \alpha(EM)$, i.e., $\forall i\in I_1, \exists k\in I_2$ such that $B_{1i}\supseteq B_{2k}$, $b_{2k}\supseteq b_{1i} \Rightarrow \beta(b_{2k})\subseteq \beta(b_{1i})$ and $\forall j\in I_2, \exists l\in I_1$ such that $B_{2j}\supseteq B_{1l}$, $b_{2j}\subseteq b_{1l}$, $\beta(b_{2j})\supseteq\beta(b_{1l})$. This implies that

$$\cup_{i\in I_1} b_{1i} = \cup_{j\in I_2} b_{2j},\ \ \cap_{i\in I_1}\beta(b_{1i}) = \cap_{i\in I_2}\beta(b_{2i}).$$

Therefore $p_I(\sum_{i\in I_1} b_{1i}B_{1i}) = p_I(\sum_{i\in I_2} b_{2i}B_{2i})$, i.e., p_I is a map. Then for any $\zeta = \sum_{i\in I} a_iA_i, \eta = \sum_{j\in J} b_jB_j \in \alpha(EM)$, by (8.4) and (8.8), we have

$$\begin{aligned} p_I(\zeta\vee\eta) &= (\alpha\cdot\beta[(\cup_{i\in I}a_i)\cup(\cup_{j\in J}b_j)],\ \ [(\cap_{i\in I}\beta(a_i))\cap(\cap_{j\in J}\beta(b_j))]) \\ p_I(\zeta)\vee p_I(\eta) &= (\alpha\cdot\beta(\cup_{i\in I}a_i),\ \ \cap_{i\in I}\beta(a_i))\vee(\alpha\cdot\beta(\cup_{j\in J}b_j),\ \ \cap_{j\in J}\beta(b_j)) \\ &= (\alpha\cdot\beta[\alpha\cdot\beta(\cup_{i\in I}a_i)\cup\alpha\cdot\beta(\cup_{j\in J}b_j)],[(\cap_{i\in I}\beta(a_i))\cap(\cap_{j\in J}\beta(b_j))]) \end{aligned}$$

Since both $p_I(\zeta\vee\eta)$ and $p_I(\zeta)\vee p_I(\eta)$ are formal concepts of the context (G,M,I), hence $p_I(\zeta\vee\eta) = p_I(\zeta)\vee p_I(\eta)$. By (8.5) and (8.8), we have

$$\begin{aligned} p_I(\zeta\wedge\eta) &= p_I\left(\sum_{i\in I, j\in J} a_i\cap b_jA_i\cup B_j\right) \\ &= \left(\alpha\cdot\beta\left(\bigcup_{i\in I, j\in J} a_i\bigcap b_j\right),\ \bigcap_{i\in I, j\in J}\beta(a_i\cap b_i)\right) \\ &= \bigvee_{i\in I, j\in J}(a_i\cap b_j,\beta(a_i\cap b_i)). \end{aligned}$$

In addition, for any $i\in I, j\in J$, it follows by (8.5)

$$(a_i,\beta(a_i))\wedge(b_j,\beta(b_j)) = (a_i\cap b_j,\beta\cdot\alpha[\beta(a_i)\cup\beta(b_j)]).$$

By Proposition 8.1 and Lemma 8.2, for any $i\in I, j\in J$, we have

$$\begin{aligned} \beta\cdot\alpha[\beta(a_i)\cup\beta(b_j)] &= \beta\cdot\alpha(\beta(a_i)\cup\beta(b_j)) \\ &= \beta(\alpha(\beta(a_i))\cap\alpha(\beta(b_j))) \end{aligned}$$

$$
\begin{aligned}
&= \beta(\alpha(\beta(\alpha(A_i))) \cap \alpha(\beta(\alpha(B_j)))) \\
&= \beta(\alpha(A_i) \cap \alpha(B_j)) \\
&= \beta(a_i \cap b_j).
\end{aligned}
$$

Therefore

$$(a_i, \beta(a_i)) \wedge (b_j, \beta(b_j)) = (a_i \cap b_j, \beta(a_i \cap b_i))$$

and

$$
\begin{aligned}
p_I(\zeta \wedge \eta) &= \bigvee_{i \in I, j \in J} (a_i \cap b_j, \beta(a_i \cap b_i)) \\
&= \bigvee_{i \in I, j \in J} [(a_i, \beta(a_i)) \wedge (b_j, \beta(b_j))] \\
&= \left[\bigvee_{i \in I} (a_i, \beta(a_i))\right] \wedge \left[\bigvee_{j \in J} (b_j, \beta(b_j))\right] \\
&= p_I(\zeta) \wedge p_I(\eta).
\end{aligned}
$$

Therefore p_I is homomorphism. □

By Theorem 8.2, 8.3, we know that concept lattice $\mathscr{L}(G,M,I)$ has similar algebraic properties to EI algebra and EII algebra. $\mathscr{L}(G,M,I)$ as a lattice is finer than the lattices $\alpha(EM)$ and $\mathscr{L}(G,M,I)$ as an algebra structure is more rigorous than EI, EII algebras. EI, EII algebras can be applied to study fuzzy attributes while $\mathscr{L}(G,M,I)$ can only be applied to Boolean attributes.

Theorem 8.4. *Let (G,M,I) be a context and $\mathscr{L}(G,M,I)$ be a concept lattice of the context (G,M,I). Let EGM be the EI^2 algebra over the sets G,M. If the map $h : \mathscr{L}(G,M,I) \rightarrow EGM$ is defined as follows: for any formal concept $(b,B) \in \mathscr{L}(G,M,I)$, $h(b,B) = bB \in EGM$, then the following assertions hold.*

(1) If $(a,A),(b,B) \in \mathscr{L}(G,M,I)$, $(a,A) \le (b,B)$, then $h(a,A) \le h(b,B)$;
(2) For $(a,A),(b,B) \in \mathscr{L}(G,M,I)$,

$$
\begin{aligned}
h((a,A) \vee (b,B)) &\ge h(a,A) \vee h(b,B), \\
h((a,A) \wedge (b,B)) &\le h(a,A) \wedge h(b,B).
\end{aligned}
$$

Proof. (1) $(a,A) \le (b,B) \Rightarrow a \subseteq b$, $A \supseteq B$. By Definition 5.2 and Theorem 5.1, one has

$$h(a,A) \vee h(b,B) = aA + bB = bB = h(b,B).$$

This implies that $h(a,A) \le h(b,B)$ in the lattice EGM.

(2) By the definition of the map h and (8.4), (8.5), we have

$$h((a,A) \vee (b,B)) = h(\alpha \cdot \beta(a \cup b), A \cap B) = \alpha \cdot \beta(a \cup b) A \cap B.$$

By Proposition 8.1 and Lemma 8.2, we have

$$\alpha \cdot \beta(a \cup b) = \alpha \cdot \beta(\alpha(A) \cup \alpha(B)) = \alpha \cdot \beta(\alpha(A \cap B))$$
$$= \alpha(\beta(\alpha(A \cap B))) = \alpha(A \cap B).$$

Thus

$$\alpha \cdot \beta(a \cup b) = \alpha(A \cap B) \supseteq \alpha(A) = a,$$
$$\alpha \cdot \beta(a \cup b) = \alpha(A \cap B) \supseteq \alpha(B) = b.$$

Similarly, we can prove $\beta \cdot \alpha(A \cup B) \supseteq A \cup B$. Then by Theorem 5.1, we have

$$h((a,A) \vee (b,B)) = \alpha \cdot \beta(a \cup b)A \cap B \geq aA + bB = h(a,A) \vee h(b,B).$$
$$h((a,A) \wedge (b,B)) = a \cap b[\beta \cdot \alpha(A \cup B)] \leq aA \wedge bB = h(a,A) \wedge h(b,B).$$ □

8.2 Some AFS Algebraic Properties of Formal Concept Lattices

In order to explore some algebraic properties of formal concept lattices, we define a new algebra class $E^C II$ for a context (G,M,I), which is a new family of AFS algebra different from the AFS algebras discussed in some other chapters.

Definition 8.3. Let G and M be sets and (G,M,I) be a context, EGM^I is a set defined as follows:

$$EGM^I = \left\{ \sum_{u \in U} a_u A_u \mid A_u \subseteq M, a_u = \alpha(A_u),\ u \in U,\ U \ is\ a\ non-empty\ indexing\ set \right\}.$$

Where each $\sum_{u \in U} a_u A_u$ as an element of EGM^I is the "formal sum" of terms $a_u A_u$. $\sum_{u \in U} a_u A_u$ and $\sum_{u \in U} a_{p(u)} A_{p(u)}$ are the same elements of EGM^I if p is a bijection from I to I. R is a binary relation on EGM^I defined as follows: $\sum_{u \in U} a_u A_u$, $\sum_{v \in V} b_v B_v \in EGM^I$, $(\sum_{u \in U} a_u A_u, \sum_{v \in V} b_v B_v) \in R \Leftrightarrow$ (i) $\forall a_u A_u\ (u \in U)\ \exists b_k B_k\ (k \in V)$ such that $a_u \subseteq b_k$, $A_u \subseteq B_k$, (ii) $\forall b_v B_v\ (v \in V)\ \exists a_l A_l\ (l \in U)$ such that $b_v \subseteq a_l$, $B_v \subseteq A_l$.

It is obvious that R is an equivalence relation on EGM^I. The quotient set EGM^I/R is denoted as $E^I GM$. $\sum_{u \in U} a_u A_u = \sum_{v \in V} b_v B_v$ means that $\sum_{u \in U} a_u A_u$ and $\sum_{v \in V} b_v B_v$ are equivalent under the equivalence relation R.

Proposition 8.2. *Let G and M be sets, (G,M,I) be a context and $E^I GM$ be defined as Definition 8.3. For $\sum_{u \in U} a_u A_u \in E^I GM$, if $a_q \subseteq a_w$, $A_q \subseteq A_w$, $w,q \in U$, $w \neq q$, then*

$$\sum_{u \in U} a_u A_u = \sum_{u \in U, u \neq q} a_u A_u.$$

Its proof remains as an exercise.

Definition 8.4. Let G and M be sets, (G,M,I) be a context and $E^I GM$ be the set defined as Definition 8.3. We introduce the following definitions.

(1) For $\sum_{u\in U} a_u A_u \in E^I GM$, $\sum_{u\in U} a_u A_u$ is called *$E^C II$ irreducible* if $\forall w \in U$, $\sum_{u\in U} a_u A_u \neq \sum_{u\in U, u\neq w} a_u A_u$.

(2) For any $\sum_{u\in U} a_u A_u \in E^I GM$, $|\sum_{u\in U} a_u A_u|$, the *set of all $E^C II$ irreducible items in $\sum_{u\in U} a_u A_u$*, is defined as follows.

$$|\sum_{u\in U} a_u A_u| \triangleq \left\{a_u A_u \ \mid \ u\in U, a_u \nsubseteq a_j, A_u \nsubseteq A_j \ for\ any\ j\in U\right\}.$$

$||\sum_{u\in U} a_u A_u||$, the length of $\sum_{u\in U} a_u A_u$, is defined as follows

$$||\sum_{u\in U} a_u A_u|| \triangleq |\left\{a_u A_u \ \mid \ u\in U, a_u \nsubseteq a_j, A_u \nsubseteq A_j \ for\ any\ j\in U\right\}|.$$

Proposition 8.3. *Let G and M be sets, (G,M,I) be a context and $E^I GM$ be the set defined as Definition 8.3. The binary relation $\leq$ is a partial order relation if $\sum_{u\in U} a_u A_u$, $\sum_{v\in V} b_v B_v \in E^I GM$, $\sum_{u\in U} a_u A_u \leq \sum_{v\in V} b_v B_v \Leftrightarrow \forall a_u A_u$ $(u\in U)$ $\exists b_k B_k$ $(k\in V)$ such that $a_u \subseteq b_k$, $A_u \subseteq B_k$.*

Its proof remains as an exercise.

Proposition 8.4. *Let G and M be sets, (G,M,I) be a context and $E^I GM$ be defined as Definition 8.3. Then for any $\Gamma \subseteq \{A\in 2^M \mid A = \beta\cdot\alpha(A)\} \subseteq M$, $\varnothing \neq \Gamma$, $\sum_{B\in\Gamma} \alpha(B)B$ is $E^C II$ irreducible.*

Proof. Suppose there exists $A\in\Gamma$ such that $\sum_{B\in\Gamma} \alpha(B)B = \sum_{B\in\Gamma, B\neq A} \alpha(B)B$. By Definition 8.3, for $\alpha(A)A$ standing on the left side of the equation, we know that $\exists E\in\Gamma$, $E\neq A$ such that $\alpha(A)\subseteq\alpha(E), A\subseteq E$. By the properties of the Galois connection α,β shown in Lemma 8.1 and $A\subseteq E$, we have $\alpha(A)\supseteq\alpha(E)$. This implies that $\alpha(A)=\alpha(E)$ and $A=\beta\cdot\alpha(A)=\beta\cdot\alpha(E)=E$. It contradicts that $E\neq A$. Therefore $\sum_{B\in\Gamma}\alpha(B)B$ is $E^C II$ irreducible. □

Proposition 8.5. *Let (G,M,I) be a context and $E^I GM$ be defined as Definition 8.3. If for any $\sum_{u\in U} a_u A_u$, $\sum_{v\in V} b_v B_v \in E^I GM$, we define*

$$(\sum_{u\in U} a_u A_u) * (\sum_{v\in V} b_v B_v) = \sum_{u\in U, v\in V} a_u \cap b_v A_u \cup B_v, \tag{8.9}$$

$$(\sum_{u\in U} a_u A_u) + (\sum_{v\in V} b_v B_v) = \sum_{u\in U\sqcup V} c_u C_u, \tag{8.10}$$

where $u\in U\sqcup V$ (the disjoint union of indexing sets U, V), $c_u = a_u$, $C_u = A_u$, if $u\in U$; $c_u = b_u$, $C_u = B_u$, if $u\in U$. Then "+" and "$*$" are binary compositions on $E^I GM$.

Its proof remains as an exercise.

The algebra system $(E^IGM, *, +, \leq)$ is called the *E^CII algebra of context* (G,M,I) and denoted as E^IGM, where $*$ and $+$ are defined by (8.9) and (8.10), and $\leq$ is defined by Proposition 8.3. For $\sum_{u\in U} a_uA_u \in E^IGM$, let

$$\left(\sum_{u\in U} a_uA_u\right)^h = \overbrace{\sum_{u\in U} a_uA_u * \ldots * \sum_{u\in U} a_uA_u}^{h}.$$

The algebra system $(E^IGM, *, +, \leq)$ has the following properties which can be further applied to study the formal concept lattice.

Proposition 8.6. *Let G and M be finite sets, (G,M,I) be a context and $(E^IGM, *, +, \leq)$ be the E^CII algebra of context (G,M,I). Then the following assertions hold. For any ψ, ϑ, γ, $\eta \in E^IGM$,*

(1) $\psi + \vartheta = \vartheta + \psi, \quad \psi * \vartheta = \vartheta * \psi;$
(2) $(\psi + \vartheta) + \gamma = \psi + (\vartheta + \gamma), \quad (\psi * \vartheta) * \gamma = \psi * (\vartheta * \gamma);$
(3) $(\psi + \vartheta) * \gamma = (\psi * \gamma) + (\vartheta * \gamma), \quad \psi * (\varnothing M) = (\varnothing M), \quad \psi * (X\varnothing) = \psi;$
(4) If $\psi \leq \vartheta$, $\gamma \leq \eta$, then $\psi + \gamma \leq \vartheta + \gamma, \quad \psi * \gamma \leq \vartheta * \gamma;$
(5) For any $\zeta \in E^IGM$, any positive integer n,

$$\zeta \leq \zeta^n, \ \ (\zeta + \varnothing M)^n = \zeta^n + \varnothing M.$$

(6) Let $A_j \subseteq M, j \in J$, J be any non-empty indexing set. For any $A \subseteq M$, $U(A)$ the set of all intents containing A is defined as follows.

$$U(A) = \{B \mid A \subseteq B \subseteq M, B = \beta \cdot \alpha(B)\}.$$

Then the following assertions hold.

$$\left(\sum_{j\in J}\sum_{A\in U(A_j)} \alpha(A)A\right)^2 = \sum_{j\in J}\sum_{A\in U(A_j)} \alpha(A)A.$$

(7) For any $\sum_{j\in J} a_jA_j \in E^IGM$ and any positive integer l,

$$\left(\sum_{j\in J} a_jA_j\right)^l \leq \sum_{j\in J}\sum_{A\in U(A_j)} \alpha(A)A.$$

(8) If $\gamma = \sum_{m\in M} \alpha(\{m\})\{m\}$, then there exists an positive integer h such that $(\gamma^h)^2 = \gamma^h$, $|\gamma^h|$ is the set of all concepts of context (G,M,I) except $(X,\varnothing)$ ($|\gamma^h|$ defined by Definition 8.4).

Proof. (1), (2), (3) and (4) can be directly proved by using the definitions.

Now we prove (5). Let $\zeta = \sum_{u\in U} a_u A_u \in E^I GM$.

$$\begin{aligned}\left(\sum_{u\in U} a_u A_u\right) * \left(\sum_{u\in U} a_u A_u\right) &= \sum_{u,v\in U} a_u \cap a_v A_u \cup A_v \\ &= \sum_{u\in U} a_u A_u + \sum_{u,v\in U, u\neq v} a_u \cap a_v A_u \cup A_v \\ &\geq \sum_{u\in U} a_u A_u.\end{aligned}$$

Thus $\zeta \leq \zeta^2$. By (4), one has $\zeta \leq \zeta^2 \leq \zeta^3 \Rightarrow \zeta \leq \zeta^2 \leq \zeta^3 \leq \zeta^4 \Rightarrow ... \Rightarrow \zeta \leq \zeta^2 ... \leq \zeta^n$. From (1), (2), (3), we have

$(\sum_{u\in U} a_u A_u + \varnothing M) * (\sum_{u\in U} a_u A_u + \varnothing M)$
$= \sum_{u\in U} a_u A_u * \sum_{u\in U} a_u A_u + \varnothing M * \sum_{u\in U} a_u A_u + \sum_{u\in U} a_u A_u * \varnothing M + \varnothing M * \varnothing M$
$= (\sum_{u\in U} a_u A_u)^2 + \varnothing M.$

Now we prove it by induction with respect on n. Suppose

$$(\sum_{u\in U} a_u A_u + \varnothing M)^{n-1} = (\sum_{u\in U} a_u A_u)^{n-1} + \varnothing M.$$

We have

$(\sum_{u\in U} a_u A_u + \varnothing M)^n$
$= ((\sum_{u\in U} a_u A_u)^{n-1} + \varnothing M) * (\sum_{u\in U} a_u A_u + \varnothing M)$
$= (\sum_{u\in U} a_u A_u)^n + (\sum_{u\in U} a_u A_u)^{n-1} * \varnothing M + \varnothing M * \sum_{u\in U} a_u A_u + \varnothing M$
$= (\sum_{u\in U} a_u A_u)^n + \varnothing M.$

Therefore the assertion holds.

(6) Let $\sum_{j\in J} a_j A_j \in E^I GM$. For any $u, v \in J$,

$$\left(\sum_{A\in U(A_u)} \alpha(A)A\right) * \left(\sum_{A\in U(A_v)} \alpha(A)A\right) = \sum_{A\in U(A_u), B\in U(A_v)} \alpha(A)\cap\alpha(B) A\cup B.$$

For any $A \in U(A_u), B \in U(A_v)$, if $A\cup B$ is an intent of a concept of context (G,M,I), then $A\cup B \in U(A_u)\cap U(A_v)$. If $A\cup B$ is not an intent of a concept of context (G,M,I), then $A\cup B \subset \beta\cdot\alpha(A\cup B) \in U(A_u)\cap U(A_v)$ and $\alpha\cdot\beta\cdot\alpha(A\cup B) = \alpha(A\cup B) = \alpha(A)\cap\alpha(B)$. Thus in any case, for any $A \in U(A_u), B \in U(A_v)$ there exists $D \in U(A_v)$ such that $\alpha(A)\cap\alpha(B) \subseteq \alpha(D)$ and $A\cup B \subseteq D$ (e.g., $D = \beta\cdot\alpha(A\cup B)$). By Proposition 8.3, one has

$$\left(\sum_{A\in U(A_u)} \alpha(A)A\right) * \left(\sum_{A\in U(A_u)} \alpha(A)A\right) \leq \sum_{A\in U(A_u)} \alpha(A)A$$

From (5), we have

$$\left(\sum_{A\in U(A_u)} \alpha(A)A\right) * \left(\sum_{A\in U(A_u)} \alpha(A)A\right) = \sum_{A\in U(A_u)} \alpha(A)A$$

It follows from (1), (2) and (3),

$$\begin{aligned}\left(\sum_{u\in U}\sum_{A\in U(A_u)}\alpha(A)A\right)^2 &= \sum_{u,v\in U}\left(\sum_{A\in U(A_u)}\alpha(A)A*\sum_{A\in U(A_u)}\alpha(A)A\right)\\ &= \sum_{u,v\in U}\left(\sum_{A\in U(A_u)}\alpha(A)A\right)\\ &= \sum_{u\in U}\sum_{A\in U(A_u)}\alpha(A)A\end{aligned}$$

(7) Let $\sum_{j\in J}a_jA_j \in E^IGM$. It is obvious that $\forall u\in J, A_u\subseteq \beta\cdot\alpha(A_u)\in U(A_u)$ and $a_u=\alpha(A_u)=\alpha\cdot\beta\cdot\alpha(A_u)$. This implies that $\sum_{j\in J}a_jA_j\leq\sum_{j\in J}\sum_{A\in U(A_j)}\alpha(A)A$. By (4), (5) and (6), for any integer l, we have

$$\left(\sum_{j\in J}a_jA_j\right)^l \leq \left(\sum_{j\in J}\sum_{A\in U(A_j)}\alpha(A)A\right)^l = \sum_{j\in J}\sum_{A\in U(A_j)}\alpha(A)A.$$

(8) By (5), we know that

$$\sum_{m\in M}\alpha(\{m\})\{m\} \leq \left(\sum_{m\in M}\alpha(\{m\})\{m\}\right)^2 \leq \ldots \leq \left(\sum_{m\in M}\alpha(\{m\})\{m\}\right)^r.$$

Since both G and M are finite sets, hence there are finite number of elements in E^IGM and there exists an integer h such that

$$\left(\sum_{m\in M}\alpha(\{m\})\{m\}\right)^h = \left(\sum_{m\in M}\alpha(\{m\})\{m\}\right)^{2h}.$$

From (7), we know that for any integer r,

$$\left(\sum_{m\in M}\alpha(\{m\})\{m\}\right)^r \leq \left(\sum_{m\in M}\alpha(\{m\})\{m\}\right)^h \leq \sum_{m\in M}\sum_{A\in U(\{m\})}\alpha(A)A.$$

Then for any $m\in M, A\in U(m)$, there exists an item $\alpha(B)B$ in $(\sum_{m\in M}\alpha(\{m\})\{m\})^{|M|}$ such that $\alpha(B)=\cap_{m\in A}\alpha(\{m\})\supseteq\alpha(A), B=\cup_{m\in A}\{m\}\supseteq A$. This implies that

$$\left(\sum_{m\in M}\alpha(\{m\})\{m\}\right)^{|M|} \geq \sum_{m\in M}\sum_{A\in U(\{m\})}\alpha(A)A$$

Therefore

$$\left(\sum_{m\in M}\alpha(\{m\})\{m\}\right)^{|M|}=\left(\sum_{m\in M}\alpha(\{m\})\{m\}\right)^{h}=\sum_{m\in M}\sum_{A\in U(\{m\})}\alpha(A)A.$$

Let

$$\left(\sum_{m\in M}\alpha(\{m\})\{m\}\right)^{h}=\sum_{m\in M}\sum_{A\in U(\{m\})}\alpha(A)A=\sum_{j\in J}a_jA_j,$$

and $\sum_{j\in J}a_jA_j$ is E^CII irreducible. Since for any concept $(\alpha(A),A)$ of the context (G,M,I), $\alpha(A)A$ is an item in $\sum_{m\in M}\sum_{A\in U(\{m\})}\alpha(A)A$ and $\sum_{m\in M}\sum_{A\in U(\{m\})}\alpha(A)A=\sum_{j\in J}a_jA_j$. Then for any concept $(\alpha(A),A)$ there exists $j\in J$, such that $\alpha(A)\subseteq a_j=\alpha(A_j)$, $A\subseteq A_j$. By the properties of Galois connection α,β in Lemma 8.1 and $A\subseteq A_j$, we have $\alpha(A)\supseteq\alpha(A_j)=a_j$, $\alpha(A)=a_j=\alpha(A_j)$, $A=\beta\cdot\alpha(A)=\beta\cdot\alpha(A_j)=A_j$. Thus $(\alpha(A),A)\in\{(a_j,A_j)|j\in J\}$. For any $w\in J$, if (a_w,A_w) is not a concept of the context (G,M,I), then A_w is a proper subset of $\beta(a_w)=\beta\cdot\alpha(A_w)$ and $(a_w,\beta(a_w))$ is a concept of the context (G,M,I). This implies that $a_w\beta(a_w)$ is an item in $\sum_{m\in M}\sum_{A\in U(\{m\})}\alpha(A)A$. By Proposition 8.2, we know that item a_wA_w will be reduced and it cannot appear in $\sum_{j\in J}a_jA_j$. It is a contradiction. Therefore $\{(a_j,A_j)|j\in J\}$ is the set of all concepts of context (G,M,I) except $(X,\varnothing)$. □

Theorem 8.5. *Let G and M be finite sets, (G,M,I) be a context and $(E^IGM,*,+,\leq)$ be the E^CII algebra of context (G,M,I). Let $\gamma=\sum_{m\in M}\alpha(\{m\})\{m\}$. For an item $aA\in|\gamma^k|$ ($|\gamma^k|$ defined by Definition 8.4), if $|A|<k$, then (a,A) is a formal concept of the context (G,M,I), i.e., $\beta(a)=A$, $\alpha(A)=a$, where k is any positive integer.*

Proof. Assume that there exists an item $aA\in|\gamma^k|$ with $|A|<k$ in $|\gamma^k|$ such that (a,A) isn't a formal concept of the context (G,M,I). This implies that there exist $B\subseteq M$, $A\subseteq B$ and $|B|=|A|+1$ such that $a=\alpha(B)$. It is obvious that aB is an item in $\gamma^{|A|+1}$. By 5 of Proposition 8.6, one knows that $\gamma^{|A|+1}\leq\gamma^k$. From Proposition 8.3, we know there exist an item cC in γ^k such that $a=\alpha(B)\subseteq c,\ A\subseteq B\subseteq C$. By Proposition 8.2, item aA can be reduced and it contradicts $aA\in|\gamma^k|$. Therefore (a,A) is a concept of the context (G,M,I). □

In what follows, we discuss how to find concepts of a context using the above results. The following theorem gives a very simple way to compute the power of an E^CII element.

Theorem 8.6. *Let G and M be finite sets, (G,M,I) be a context and $(E^IGM,*,+,\leq)$ be the E^CII algebra of the context (G,M,I). For any nonempty set $C\subseteq M$, let $\gamma=\sum_{m\in C}\alpha(\{m\})\{m\}$ and $\gamma^k=\sum_{i\in I}\alpha(A_i)A_i$, where k is a positive integer, $k\leq|C|$. Then*

$$\gamma^{2k}=\sum_{i\in I,|A_i|<k}\alpha(A_i)A_i+\left(\sum_{i\in I,|A_i|=k}\alpha(A_i)A_i\right)^2,$$

where $\sum_{i\in I,|A_i|=k}\alpha(A_i)A_i\triangleq\varnothing C$, if there does not exist $i\in I$ such that $|A_i|=k$.

Proof. Let $\gamma^k = \sum_{i\in I}\alpha(A_i)A_i = \sum_{i\in I,|A_i|<k}\alpha(A_i)A_i + \sum_{i\in I,|A_i|=k}\alpha(A_i)A_i$. Then

$$\begin{aligned}\gamma^{2k} &= \left(\sum_{i\in I,|A_i|<k}\alpha(A_i)A_i + \sum_{i\in I,|A_i|=k}\alpha(A_i)A_i\right)^2\\ &= \left(\sum_{i\in I,|A_i|<k}\alpha(A_i)A_i\right)^2 + \left(\sum_{i\in I,|A_i|<k}\alpha(A_i)A_i\right) * \left(\sum_{i\in I,|A_i|=k}\alpha(A_i)A_i\right)\\ &\quad + \left(\sum_{i\in I,|A_i|=k}\alpha(A_i)A_i\right)^2 \end{aligned} \tag{8.11}$$

Because $\gamma = \sum_{m\in C}\alpha(\{m\})\{m\}$. For any set $B\subseteq C$, $|B|<k$, $\alpha(B)B$ is an item in $\sum_{i\in I,|A_i|<k}\alpha(A_i)A_i$. Although $\alpha(B)B$ may be reduced by other items in $\sum_{i\in I,|A_i|<k}\alpha(A_i)A_i$, we always have $\alpha(B)B \le \sum_{i\in I,|A_i|<k}\alpha(A_i)A_i$ and

$$\alpha(B)B + \sum_{i\in I,|A_i|<k}\alpha(A_i)A_i = \sum_{i\in I,|A_i|<k}\alpha(A_i)A_i. \tag{8.12}$$

Similarly, since for every set $B\subseteq C$, $|B|=k$, $B'B \le \sum_{i\in I,|A_i|=k}A_i'A_i$ and for any set $E\subseteq C$, $k\le |E|\le 2k$, there exist $F,H\subseteq C$, $|F|=|H|=k$ such that $E=F\cup H$. By 4 of Proposition 8.3 and the facts $\alpha(F)F \le \sum_{i\in I,|A_i|=k}\alpha(A_i)A_i$ and $\alpha(H)H \le \sum_{i\in I,|A_i|=k}\alpha(A_i)A_i$, we have

$$\alpha(E)E = \alpha(F)\cap\alpha(H)F\cup H \le \left(\sum_{i\in I,|A_i|=k}\alpha(A_i)A_i\right)^2,$$

$$\alpha(E)E + \left(\sum_{i\in I,|A_i|=k}\alpha(A_i)A_i\right)^2 = \left(\sum_{i\in I,|A_i|=k}\alpha(A_i)A_i\right)^2. \tag{8.13}$$

According to (8.11), (8.12) and (8.13), we have

$$\begin{aligned}\gamma^{2k} &= \left(\sum_{i\in I,|A_i|<k}\alpha(A_i)A_i\right)^2 + \left(\sum_{i\in I,|A_i|=k}\alpha(A_i)A_i\right)^2\\ &= \sum_{i\in I,|A_i|<k}\alpha(A_i)A_i + \left(\sum_{i\in I,|A_i|=k}A_i'A_i\right)^2. \end{aligned}$$ □

The above discussion shows that the algebra characteristics of the formal concepts of a context can be explored by the $E^C II$ algebra of the context. For example, Theorem 8.5, Theorem 8.6 and Proposition 8.6 can be applied to identify all formal concepts of a context. For any context (G,M,I), let $\gamma = \sum_{m\in M}\alpha(\{m\})\{m\}$. By 8 of Proposition 8.6, we know that $|\gamma^{|M|}|$ defined by Definition 8.4 is the set of all formal concepts of the context (G,M,I). Notice that there is only one formal concept whose extent is $\varnothing$, i.e., $\varnothing M$ for (G,M,I). In order to simplify the computation of $\gamma^{|M|}$, we compute $(\gamma+\varnothing M)^2$, $(\gamma+\varnothing M)^4$, $(\gamma+\varnothing M)^8$,...,$(\gamma+\varnothing M)^{2^k}$, until $2^k \geq |M|$. By 5 of Proposition 8.6, one knows that $(\gamma+\varnothing M)^n = \gamma^n+\varnothing M$ for any positive integer n. So each item $\varnothing A$ in γ^n can be reduced by $\varnothing M$ and the number of items of $(\gamma+\varnothing M)^n$ is much lower than γ^n. Let $\gamma^k = \sum_{i\in I}\alpha(A_i)A_i = \sum_{i\in I,|A_i|<k}\alpha(A_i)A_i + \sum_{i\in I,|A_i|=k}\alpha(A_i)A_i$. According to Theorem 8.5 and Theorem 8.6, we know that for any formal concept (a,A), aA is in $|\sum_{i\in I,|A_i|<k}\alpha(A_i)A_i|$ if $|A|<k$ and aA is in $|(\sum_{i\in I,|A_i|=k}\alpha(A_i)A_i)^2|$ if $k\leq|A|\leq 2k$, where k is a positive integer. This fact and the following equation can further facilitate the computing,

$$\gamma^{2k} = \sum_{i\in I,|A_i|<k}\alpha(A_i)A_i + \left(\sum_{i\in I,|A_i|=k}\alpha(A_i)A_i\right)^2.$$

Example 8.1 demonstrates how the detailed calculations are carried out.

Table 8.2 The Reduced Mushroom

	m_1	m_2	m_3	m_4	m_5
Mushroom 1	×		×		
Mushroom 2	×		×		×
Mushroom 3	×			×	×
Mushroom 4	×			×	×
Mushroom 5	×		×		×
Mushroom 6	×		×		
Mushroom 7		×		×	×
Mushroom 8		×		×	×
Mushroom 9		×		×	×
Mushroom 10		×		×	

Example 8.1. The Table 8.2 shows the reduced mushroom example database from the UCI KDD Archive (http://kdd.ics.uci.edu) in [43]. Where m_1: *edible*, m_2: *poisonous*, m_3: *cap_shape:convex*, m_4: *cap-shape: flat*, m_5: *cap-surface:fibrous*. Let (G,M,I) be the context of Table 8.2, $G=\{1,2,...,10\}$, $M=\{m_1, m_2, m_3, m_4, m_5\}$. Let us find all formal concepts of the context (G,M,I) by $E^C II$ algebra via the computing on the power of the following γ.

$$\begin{aligned}\gamma = &\{1,2,3,4,5,6\}\{m_1\}+\{7,8,9,10\}\{m_2\}+\{1,2,5,6\}\{m_3\}\\ &+\{3,4,7,8,9,10\}\{m_4\}+\{2,3,4,5,7,8,9\}\{m_5\}\in E^I GM.\end{aligned}$$

For any positive integer $k>1$, let $\gamma^k = \sum_{i\in I}\alpha(A_i)A_i = \underline{\gamma}^k + \bar{\gamma}^k$. Where $\underline{\gamma}^k = \sum_{i\in I|A_i|<k}\alpha(A_i)A_i$, $\bar{\gamma}^k = \sum_{i\in I|A_i|=k}\alpha(A_i)A_i$.

$$
\begin{aligned}
(\gamma+\varnothing M)^2 = & \{1,2,3,4,5,6\}\{m_1\}+\{3,4,7,8,9,10\}\{m_4\}+\{2,3,4,5,7,8,9\}\{m_5\} \\
& +\{1,2,5,6\}\{m_1m_3\}+\{2,3,4,5\}\{m_1m_5\}+\{3,4\}\{m_1m_4\} \\
& +\{7,8,9,10\}\{m_2m_4\}+\{7,8,9\}\{m_2m_5\}+\{2,5\}\{m_3m_5\} \\
& +\{3,4,7,8,9\}\{m_4m_5\}+\varnothing M.
\end{aligned}
$$

$$
\begin{aligned}
(\gamma+\varnothing M)^4 = & \gamma^4+\varnothing M=\underline{\gamma}^2+(\bar{\gamma}^2)^2+\varnothing M \\
= & \{1,2,3,4,5,6\}\{m_1\}+\{3,4,7,8,9,10\}\{m_4\}+\{2,3,4,5,7,8,9\}\{m_5\} \\
& +\{1,2,5,6\}\{m_1m_3\}+\{2,3,4,5\}\{m_1m_5\}+\{7,8,9,10\}\{m_2m_4\}+ \\
& \{3,4,7,8,9\}\{m_4m_5\}+\{2,5\}\{m_1m_3m_5\}+\{3,4\}\{m_1m_4m_5\}+ \\
& \{7,8,9\}\{m_2m_4m_5\}+\varnothing M
\end{aligned}
$$

Since there does not exist item aA in γ^4 such that $|A|=4$, hence $\overline{\gamma}^4=\varnothing M$ and

$$(\gamma+\varnothing M)^8=\gamma^8+\varnothing M=\underline{\gamma}^4+(\overline{\gamma}^4)^2+\varnothing M=\underline{\gamma}^2+(\bar{\gamma}^2)^2+\varnothing M=(\gamma+\varnothing M)^4.$$

According to Theorem 8.5 and 8 of Proposition 8.6, all concepts of the context for Table 8.2 are the items shown in the above $(\gamma+\varnothing M)^4$ except $(X,\varnothing)$. It is the same result as what has been obtained by the TITANIC algorithm presented in [43].

8.3 Concept Analysis via Rough Set and AFS Algebra

In this section, combining formal concept analysis (FCA) and AFS algebra, we propose AFS formal concept, which can be viewed as the generalization and development of monotone concept proposed by Deogun and Saquer (2003) [8]. Moreover, we show that the set of all AFS formal concepts forms a complete lattice. AFS formal concept can be applied to represent the logic operations of queries in information retrieval. Furthermore, we present an approach to find the AFS formal concepts whose intents (extents) approximate any fuzzy concepts in EM by virtue of rough set theory.

The characteristic of concept lattice theory lies in reasoning on the possible attributes of data sets [66]. Currently, FCA has been extended to other types for requirements of real word applications, such as fuzzy concept lattice [2, 46], triadic concept [57], monotone concept [8], variable threshold concept lattice [65], rough formal concept [66], etc.

Rough set and FCA are related and complementary. In recent years, many efforts have been made to compare and combine these two theories [61, 62, 64, 65]. The combination of FCA and rough set theory provides some new approaches for data analysis and knowledge discovery [44, 45, 55, 66].

In [8], Deogun and Saquer discussed some of limitations of Wille's formal concept [10] and proposed monotone concept. In Wille's notation of concepts, only one set is allowed as extent (intent). For many applications, it is necessary to allow intents to be disjunction expression. Monotone concept is a generalization of Wille's notion of concept where disjunctions are allowed in the intent and set unions are allowed in the extent. This generalization allows an information retrieval query containing disjunctions to be understood as the intent of a monotone concept whose answer is the extent of that concept. In [44], by using rough set theory, Saquer and Deogun formulated a general solution to find monotone concepts whose intents are close to the query, and show how to find monotone concepts whose extents approximate any given set of objects.

In this section, we propose AFS formal concept, which extend the Galois connection α, β of a context (X,M,I) to the connection between two AFS algebra systems $(EM,\vee,\wedge)$ and $(E^{\#}X,\vee,\wedge)$. The intent of an AFS formal concept is an element of the *EI* algebra $(EM,\vee,\wedge)$—a kind of AFS algebra over M; correspondingly, the extent of the AFS formal concept is an element of the $E^{\#}I$ algebra $(E^{\#}X,\vee,\wedge)$—another kind of AFS algebra over X. Where M is a set of elementary attributes on X, EM is the set of attributes logically compounded by some elementary attributes in M under logic operations $\vee$ and $\wedge$ (i.e., "*and*" and "*or*"). Each element of EM is called a complex attribute (or a fuzzy concept), and has definitely semantic interpretation. The extent and intent of an AFS formal concept can uniquely determine each other. Thus, the intent of an AFS formal concept not only generalizes that of the formal concept, but also has a well-defined semantic interpretation.

In an information retrieval system, the logic relationships between queries are usually expressed by logic connectives such as "*and*" and "*or*". AFS formal concepts can be used to represent the query with complex logic operations. When using the information retrieval system, we often find that not all queries are exactly contained in database, but some items close to those are enough to satisfy user's need. Thus, it is necessary to investigate how to approximate a complex attribute by AFS formal concepts such that the intents of lower and upper approximating concept are closely to the complex attribute underlying semantics.

In this section, first, FCA and rough set are briefly summarized. Monotone concept is also introduced and studied. Second, AFS formal concept is proposed and the mathematical properties of AFS formal concepts are discussed. Third, we show that the set of all AFS formal concepts forms a complete lattice. Fourth, an approach to approximate the element of the EM $(E^{\#}X)$ is proposed.

8.3.1 Monotone Concept

Let us first recall monotone concept [8] and study the aspects which should be improved in concept representation and approximation. In [8], Deogun and Saquer introduced some notations as follows: (X,M,I) is a context. Associating with every set $B \subseteq M$, a Boolean conjunctive expression $\hat{B}$ is the conjunction of the elements of B. For example, if $B = \{a,b,c\}$, then the associated Boolean conjunctive expression is $\hat{B} = a \wedge b \wedge c$. A disjunction of Boolean conjunctive expressions is referred to as a *monotone Boolean formula*. If $\hat{B}_1, \hat{B}_2, \ldots, \hat{B}_n$ are Boolean conjunctive expressions, then $F = \hat{B}_1 \vee \hat{B}_2 \vee \ldots \vee \hat{B}_n = \bigvee_{i=1}^{n} \hat{B}_i$ is *monotone formula*. For example, let $B_1 = \{a,b,c\}$, $B_2 = \{a,d\}$, then $\hat{B}_1 = a \wedge b \wedge c$, $\hat{B}_2 = a \wedge d$, $F = \hat{B}_1 \vee \hat{B}_2 = (a \wedge b \wedge c) \vee (a \wedge d)$ is monotone formula. For simplicity, F would be written as $abc \vee ad$ [44].

Definition 8.5. ([8]) Let (X,M,I) be a context. For the monotone formula $F = \vee_{i=1}^{n}\hat{B}_i$, $B_i \subseteq M$, $\delta(F)$ is defined as the set of all objects that satisfy F, that is, $\delta(F) = \cup_{i=1}^{n}\delta(\hat{B}_i) \subseteq X$, where $\delta(\hat{B}_i)$ is the *set of all objects that satisfy* $\hat{B}_i$. For $A = \cup_{j=1}^{n}A_j$, define $\gamma(A)$ to be $\vee_{j=1}^{n}\gamma(A_j) \subseteq M$, where $A_j \subseteq X$, $\gamma(A_j)$ is defined to be the *Boolean conjunctive expression associated with* $\beta(A_j)$, "β" is Galois connection.

Example 8.2. Let $X = \{1,2,...,13\}$ and $M = \{a,b,c,d,e,f,h,i,j,l,x\}$ be the set of attributes on X. The context (X,M,I) is shown as Table 8.3. Assume that the monotone formula $F = \hat{B}_1 \vee \hat{B}_2 = abc \vee lx$, $A = \cup_{i=1}^{3}A_i = \{4,6\} \cup \{6,7\} \cup \{5\}$. By Definition 8.5, we have the following $\delta(F)$ and $\gamma(A)$.

$$\delta(F) = \delta(\hat{B}_1) \cup \delta(\hat{B}_2) = \{4\} \cup \{6\} = \{4,6\},$$
$$\gamma(A) = \vee_{i=1}^{3}\gamma(A_i) = efhl \vee fhijx \vee cdefhix.$$

Table 8.3 Relationship between objects and attributes [44]

	a	b	c	d	e	f	h	i	j	k	l	x
1				×								
2				×		×	×					
3						×	×					
4	×	×	×	×	×	×	×				×	
5			×	×	×	×	×	×				×
6					×	×	×	×	×	×	×	×
7						×	×	×	×			×
8			×	×	×	×	×					
9			×	×		×					×	
10			×	×	×	×					×	
11						×	×					
12			×	×	×	×	×	×				×
13			×	×	×	×	×	×				×

Definition 8.6. ([8]) Let (X,M,I) be a context, $A_i \subseteq X, B_j \subseteq M$, $1 \le i,j \le n$. A pair (A,F) where $A = \cup_{i=1}^{n}A_i$, $F = \vee_{j=1}^{n}\hat{B}_j$ is *monotone concept* if $\delta(F) = A$, $\gamma(A) = F$. A is called its *extent of the monotone concept* (A,F), F its *intent of the monotone concept* (A,F). Where B_j is the set of features associated with $\hat{B}_j$, and for each A_i, there exists a B_j such that (A_i,B_j) is a formal concept.

A monotone formula F is called *feasible* if it is the intent of a monotone concept; otherwise, F is called *non-feasible*. Similarly, $A \subseteq X$ is called *feasible* if it is the extent of a monotone concept; otherwise, A is called *non-feasible*. For instance, assume $F = ef \vee fhix$, $A = \{4,5,6,8,10,12,13\} \cup \{5,6,7,12,13\}$ in Table 8.3 of Example 8.2. One can verify that $\delta(F) = \{4,5,6,8,10,12,13\} \cup \{5,6,7,12,13\} = A$, and $\gamma(A) = ef \vee fhix = F$. Hence (A,F) is a monotone concept, and A, F are feasible. If $F = \hat{B}_1 \vee \hat{B}_2 = abc \vee lx$, $A = \{4,6,7\}$. According to Table 8.3, we have

$\delta(F) = \delta(\hat{B_1}) \cup \delta(\hat{B_2}) = \{4,6\} \neq A$, $\gamma(\delta(F)) \neq F$, $\delta(\gamma(A)) \neq A$. (A,F) is not a monotone concept, and A, and F are non-feasible.

Although the monotone concept overcomes some limitations of the Wille's formal concept [10], there remain two aspects that could be improved:

1) In monotone concept, intent and extent may not uniquely determine each other. In Example 8.2, according to Table 8.3 and Definition 8.6, we know that $(\{4,5,6,7,8,10,12,13\}, abcdefhl \vee ef \vee fhix)$ is a monotone concept which is different from $(\{4,5,6,7,8,10,12,13\}, ef \vee fhix)$, but their extents are identical.

2) Consider $\{4\}, \{4,5,6,8,10,12,13\}, \{5,6,7,12,13\}$ and $abcdefhl, fhix, ef$ in Table 8.3. It is easy to verify that $\{4\} \cup \{4,5,6,8,10,12,13\} \cup \{5,6,7,12,13\} = \{4,5,6,7,8,10,12,13\}$, and $(\{4,5,6,7,8,10,12,13\}, abcdefhl \vee ef \vee fhix)$ is a monotone concept. If considering $abcdefhl$, ef and $fhix$ as query words in an information retrieval system, we can find that $abcdefhl \vee ef \vee fhix$ represents the logical relations "*or*" among them. Notice $\{e,f\} \subset \{a,b,c,d,e,f,h,l\}$. Thus, if one object satisfies the condition expressed by $abcdefhl$, then it must satisfy that expressed by ef, i.e., $abcdefhl$ is redundant when $abcdefhl \vee ef \vee fhix$ forms a query. In other words, the queries $abcdefhl \vee ef \vee fhix$ and $ef \vee fhix$ are equivalent in semantics. However, they are intents of different monotone concepts defined by Definition 8.6.

In [44], Saquer and Deogun gave a general solution to find monotone concepts whose intents are close to the queries, and show how to find monotone concepts whose extents approximate any given set of objects. However, it seems that the following aspects of extents and intents approximations could be developed.

i) Let D be a set of objects. In [44], D is written as the union of the maximal extents of formal concepts that are contained in D and, possibly, a subset containing whatever elements remain in D. For example, the non-feasible object set $\{4,5,6,7\}$ is written as $\{4,6\} \cup \{6,7\} \cup \{5\}$. But it is also reasonable in practice to write it down as $\{4\} \cup \{4,6\} \cup \{6,7\} \cup \{5\}$. For instance, similar expressions have existed in [44] (see $L(\psi)$ in Example 8.3). Accordingly, both $(\{4,5,6,7,12,13\}, efhl \vee fhijx \vee cdefhix)$ and $(\{4,5,6,7,12,13\}, abcdefhl \vee efhl \vee fhijx \vee cdefhix)$ could be the upper approximation monotone concepts of $\{4,5,6,7\}$ in Table 8.3.

ii) When approximating $D = \{2,3,4,5,6,7,8,9,10,11,12,13\}$ in Table 8.3, one can get an approximation of the monotone concept (D, f) by using the approximation method presented in [44]. However, we can verify that $(D, cdf \vee ef \vee fh)$ is also a monotone concept which is another approximation monotone concept of D.

In order to deal with these problems, in the sequel we propose the AFS formal concept.

8.3.2 AFS Formal Concept

In this section, we propose AFS formal concept in which the Galois connection "α, β" of context (X, M, I) [10] can be extended to the connection between the EI algebra $(EM,\vee,\wedge)$ and the $E^{\#}I$ algebra $(E^{\#}X,\vee,\wedge)$ as follows: for any $\sum_{i\in I}(\prod_{m\in A_i} m)\in EM$, $\sum_{j\in J} a_j \in E^{\#}X$,

$$\alpha\left(\sum_{i\in I}\left(\prod_{m\in A_i} m\right)\right)=\sum_{i\in I}\alpha(A_i)\in E^{\#}X, \tag{8.14}$$

$$\beta\left(\sum_{j\in J} a_j\right)=\sum_{j\in J}\left(\prod_{m\in\beta(a_j)} m\right)\in EM. \tag{8.15}$$

For any $A\subseteq M, a\subseteq X$, we notice that $\alpha(\prod_{m\in A} m)=\alpha(A), \beta(a)=\prod_{m\in\beta(a)} m$, which are the same as the Galois connection "α,β" defined by (8.1) and (8.2). Thus the conventional formal concept lattice [10] can be explored in a more general mathematical framework—the AFS formal concept lattice.

In what follows, we denote the subsets of X with the lower case letters and the subsets of M with the capital letters, in order to distinguish the subsets of X from those of M.

Theorem 8.7. *Let X be a set and M be a set of attributes on X. Let (X,M,I) be a context, EM be the EI algebra over M and $E^{\#}X$ be the $E^{\#}I$ algebra over X. Then the following assertions hold:*

(1) α, β are maps, where α,β are defined by (8.14) and (8.15).

(2) For any ζ, $\eta\in EM$, ν, $\varsigma\in E^{\#}X$,

$$\alpha(\zeta\vee\eta)=\alpha(\zeta)\vee\alpha(\eta),\quad \alpha(\zeta\wedge\eta)=\alpha(\zeta)\wedge\alpha(\eta),$$
$$\beta(\nu\vee\varsigma)=\beta(\nu)\vee\beta(\varsigma),\quad \beta(\nu\wedge\varsigma)\le\beta(\nu)\wedge\beta(\varsigma).$$

(3) For any ζ, $\eta\in EM$, ν, $\varsigma\in E^{\#}X$,

$$\zeta\le\eta\Rightarrow\alpha(\zeta)\le\alpha(\eta),$$
$$\nu\le\varsigma\Rightarrow\beta(\nu)\le\beta(\varsigma).$$

(4) For any $\zeta\in EM$, $\varsigma\in E^{\#}X$,

$$\zeta\ge\beta(\alpha(\zeta)),\quad \alpha(\zeta)=\alpha(\beta(\alpha(\zeta))),$$
$$\varsigma\le\alpha(\beta(\varsigma)),\quad \beta(\varsigma)=\beta(\alpha(\beta(\varsigma))).$$

Proof. (1) Suppose $\zeta=\sum_{i\in I}(\prod_{m\in A_i} m)$, $\eta=\sum_{j\in J}(\prod_{m\in B_j} m)\in EM$, $\zeta=\eta$. That is, $\forall A_i$ $(i\in I)$, $\exists B_k$ $(k\in J)$ such that $A_i\supseteq B_k$ and $\forall B_j$ $(j\in J)$, $\exists A_l$ $(l\in I)$ such that $B_j\supseteq A_l$. This implies that $\forall\alpha(A_i)$ $(i\in I)$, $\exists\alpha(B_k)$ $(k\in J)$ such that $\alpha(A_i)\subseteq\alpha(B_k)$ and $\forall\alpha(B_j)$ $(j\in J)$, $\exists\alpha(A_l)$ $(l\in I)$ such that $\alpha(B_j)\subseteq\alpha(A_l)$. Therefore

$$\alpha\left(\sum_{i\in I}\left(\prod_{m\in A_i} m\right)\right)=\sum_{i\in I}\alpha(A_i)=\sum_{j\in J}\alpha(B_j)=\alpha\left(\sum_{j\in J}\left(\prod_{m\in B_j} m\right)\right)$$

and α is a map. Similarly, we can prove that β is also a map.

(2) $\alpha(\zeta \vee \eta) = \alpha(\zeta) \vee \alpha(\eta)$ and $\beta(\nu \vee \varsigma) = \beta(\nu) \vee \beta(\varsigma)$ can be directly verified by (8.14) and (8.15). Let $\zeta = \sum_{i\in I}(\prod_{m\in A_i} m)$, $\eta = \sum_{j\in J}(\prod_{m\in B_j} m) \in EM$, $\nu = \sum_{i\in I} a_i$, $\varsigma = \sum_{j\in J} b_j \in E^{\#}X$.

$$\begin{aligned}\alpha(\zeta \wedge \eta) = \alpha\left(\sum_{i\in I, j\in J}\left(\prod_{m\in A_i\cup B_j} m\right)\right) &= \sum_{i\in I, j\in J}\alpha(A_i \cup B_j)\\ &= \sum_{i\in I, j\in J}\alpha(A_i)\cap\alpha(B_j) = \alpha(\zeta)\wedge\alpha(\eta).\end{aligned}$$

$$\beta(\nu \wedge \varsigma) = \beta\left(\sum_{i\in I, j\in J} a_i \cap b_j\right) = \sum_{i\in I, j\in J}\beta(a_i \cap b_j).$$

For any $i \in I$, $j \in J$, since $\beta(a_i \cap b_j) \supseteq \beta(a_i)$, $\beta(a_i \cap b_j) \supseteq \beta(b_j)$, hence $\beta(a_i \cap b_j) \supseteq \beta(a_i) \cup \beta(b_j)$. This implies that

$$\begin{aligned}\beta(\nu \wedge \varsigma) = &\sum_{i\in I, j\in J}\Big(\prod_{m\in\beta(a_i\cap b_j)} m\Big) \leq \sum_{i\in I, j\in J}\Big(\prod_{m\in\beta(a_i)\cup\beta(b_j)} m\Big)\\ &= \left(\sum_{i\in I}\Big(\prod_{m\in\beta(a_i)} m\Big)\right)\wedge\left(\sum_{j\in J}\Big(\prod_{m\in\beta(b_j)} m\Big)\right) = \beta\left(\sum_{i\in I} a_i\right)\wedge\beta\left(\sum_{j\in J} b_j\right).\end{aligned}$$

(3) It can be directly verified by Theorem 4.1 and Theorem 5.24 and the properties of the Galois connection in Proposition 8.1.

(4) For $\zeta = \sum_{i\in I}(\prod_{m\in A_i} m) \in EM$, since for any $i \in I$, $A_i \subseteq \beta\cdot\alpha(A_i)$, $\alpha(A_i) = \alpha\cdot\beta\cdot\alpha(A_i)$, hence

$$\begin{aligned}&\beta(\alpha(\zeta)) = \beta\left(\sum_{i\in I}\alpha(A_i)\right) = \sum_{i\in I}\Big(\prod_{m\in\beta\cdot\alpha(A_i)} m\Big) \leq \sum_{i\in I}\Big(\prod_{m\in A_i} m\Big),\\ &\alpha(\beta(\alpha(\zeta))) = \sum_{i\in I}\alpha\cdot\beta\cdot\alpha(A_i) = \sum_{i\in I}\alpha(A_i) = \alpha(\zeta).\end{aligned}$$

For $\nu = \sum_{i\in I} a_i \in E^{\#}X$, since for any $i \in I$, $a_i \subseteq \alpha\cdot\beta(a_i)$, $\beta(a_i) = \beta\cdot\alpha\cdot\beta(a_i)$, hence

$$\alpha(\beta(\nu)) = \alpha\left(\sum_{i\in I}\Big(\prod_{m\in\beta(a_i)} m\Big)\right) = \sum_{i\in I}\alpha\cdot\beta(a_i) \geq \sum_{i\in I} a_i,$$

$$\beta(\alpha(\beta(\nu))) = \sum_{i\in I}\beta\cdot\alpha\cdot\beta(a_i) = \sum_{i\in I}\beta(a_i) = \beta(\nu).$$

The proof is complete. □

Definition 8.7. Let X be a set and M be a set of attributes on X. Let (X, M, I) be a context, EM be the EI algebra over M and $E^{\#}X$ be the $E^{\#}I$ algebra over X. Let $\zeta = \sum_{i\in I}(\prod_{m\in A_i} m) \in EM$, $\nu \in \sum_{j\in J} a_j \in E^{\#}X$. (ν, ζ) is called an *AFS formal concept of the context* (X, M, I), if $\alpha(\zeta) = \nu$, $\beta(\nu) = \zeta$. ν is called the *extent of the AFS* *formal concept* (ν, ζ) and ζ is called the *intent of the AFS formal concept* (ν, ζ). $\mathscr{L}(E^{\#}X, EM, I)$ is the *set of all AFS formal concepts of the context* (X, M, I).

In virtue of the semantics of each element in EM demonstrated in the previous chapters, we know the complex attributes in EM are much richer in expressions than the attributes in 2^M. In real world situations, many phenomena can be described by AFS formal concepts. For example, it is necessary to allow a query containing few search conditions when we use an information retrieval system. The relationships among the search conditions are usually "*or*" and "*and*" logic expression. Thus the query can be represented by the intent of an AFS formal concept. For example, $ab + bcd + e + hi$ in Table 8.3 can be used to represent the query "*ab* OR *bcd* OR *e* OR *hi*". The answer to query can be represented by the extent of an AFS formal concept.

Definition 8.8. Let X be a set and M be a set of attributes on X. Let (X, M, I) be a context, EM be the EI algebra over M and $E^{\#}X$ be the $E^{\#}I$ algebra over X. $\zeta = \sum_{i\in I}(\prod_{m\in A_i} m) \in EM$, $\nu = \sum_{j\in J} a_j \in E^{\#}X$, if $\beta(\alpha(\zeta)) \neq \zeta$, ζ is called a *non-feasible fuzzy concept*. If $\alpha(\beta(\nu)) \neq \nu$, ν is called a *non-feasible $E^{\#}I$ element*.

For example, let $\zeta = ab + f$ in Table 8.3. Due to $\beta \cdot \alpha(\{a, b\}) = \beta(\{4\}) = \{a, b, c, d, e, f, h, l\} \neq \{a, b\}$, $\beta \cdot \alpha(\{f\}) = f$. Then, $\beta(\alpha(\zeta)) \neq \zeta$, ζ is non-feasible.

Lemma 8.4. *Let X be a set and M be a set of attributes on X. Let (X, M, I) be a context. Then the following assertions hold:*

(1) For any $(\nu, \zeta) \in \mathscr{L}(E^{\#}X, EM, I)$, let $\nu = \sum_{i\in I} a_i$, $\zeta = \sum_{j\in J}(\prod_{m\in A_j} m)$. If $\sum_{j\in J}(\prod_{m\in A_j} m)$ and $\sum_{i\in I} a_i$ are irreducible, then $|I| = |J|$ ($|I|$ denotes the cardinality of I) and for any $i \in I$, $j \in J$, A_j is the intent of a formal concept of (X, M, I), a_i is the extent of a formal concept of (X, M, I).

(2) Let $\nu = \sum_{i\in I} a_i \in E^{\#}X$, $\zeta = \sum_{j\in J}(\prod_{m\in A_j} m) \in EM$, and $\sum_{i\in I} a_i$, $\sum_{j\in J}(\prod_{m\in A_j} m)$ be irreducible. If for any $j \in J$, A_j is the intent of a formal concept of context (X, M, I), then $(\alpha(\zeta), \zeta) \in \mathscr{L}(E^{\#}X, EM, I)$. If for any $i \in I$, a_i is the extent of a formal concept of context (X, M, I), then $(\nu, \beta(\nu)) \in \mathscr{L}(E^{\#}X, EM, I)$.

Proof. (1) Assume $|I| \neq |J|$. Without loss of generality, let $|I| < |J|$. By the fact that (ν, ζ) is an AFS formal concept (Definition 8.7), we know that $\beta(\nu) = \zeta$ and the cardinality of $\beta(\nu)$ is $|I|$. Since $|I| < |J|$, hence $\sum_{j\in J}(\prod_{m\in A_j} m)$ is not irreducible, which contradicts the fact that $\sum_{j\in J}(\prod_{m\in A_j} m)$ is irreducible.

Next, we will prove for any $j \in J$, A_j is the intent of a formal concept of (X, M, I) and a_i is an extent of some formal concept with an intent in $\{A_j \mid j \in J\}$. By Definition 8.7, we have $\beta(\alpha(\zeta)) = \zeta$, $\alpha(\beta(\nu)) = \nu$. This implies that there exists a bijection p from I to J such that for any $i \in I$, $\beta(a_i) = A_{p(i)}$. Since $\alpha(\beta(\nu)) = \nu$, then

$$\alpha(\beta(\nu)) = \alpha\left(\sum_{i\in I}(\prod_{m\in A_{p(i)}} m)\right) = \sum_{i\in I}\alpha(A_{p(i)}) = \sum_{i\in I} a_i.$$

If there exists $i \in I$ such that $\alpha(A_{p(i)}) = \alpha(\beta(a_i)) \neq a_i$, which means there exists a_k, $i \neq k$, such that $\alpha(A_{p(i)}) = \alpha(\beta(a_i)) = a_k$. By the properties of Galois connection "α, β", we have $a_k = \alpha(\beta(a_i)) \supseteq a_i$. It contradicts the fact that $\sum_{i\in I} a_i$ is irreducible. Thus, $\alpha(A_{p(i)}) = \alpha(\beta(a_i)) = a_i$ and $(a_i, A_{p(i)})$ is a concept of context (X,M,I).

(2) One can directly verify that $(\alpha(\zeta), \zeta)$ is an AFS formal concept of the context (X,M,I) by Definition 8.7, (8.14) and (8.15). Similarly, the second conclusion holds as well. □

Theorem 8.8. *Let X be a set and M be a set of attributes on X. Let (X,M,I) be a context and $\mathscr{L}(E^{\#}X, EM, I)$ be the set of all AFS formal concepts of the context (X,M,I). Then, for any $(\nu, \zeta) \in \mathscr{L}(E^{\#}X, EM, I)$, ν and ζ are uniquely determined by each other.*

Proof. Let $\nu = \sum_{i\in I} a_i \in E^{\#}X$, $\zeta = \sum_{j\in J}(\prod_{m\in A_j} m) \in EM$. Without loss of generality, let $\sum_{i\in I} a_i$ and $\sum_{j\in J}(\prod_{m\in A_j} m)$ be irreducible. By the Lemma 8.4, we get $|I| = |J|$. For simplicity, let $I = J$. Assume that ν and ζ are not uniquely determined by each other. Then, for ν, there exists $\rho = \sum_{k\in I}(\prod_{m\in B_k} m) \in EM$ ($\rho \neq \zeta$) such that $(\nu, \rho) \in \mathscr{L}(E^{\#}X, EM, I)$. Thus, there is at least one $i_0 \in I$ such that $A_{i_0} \neq B_i$ for any $i \in I$. From the Lemma 8.4 and Definition 8.7, we get that there exist $k \in I$, $j \in I$ such that (a_k, A_{i_0}), (a_k, B_j) are formal concepts of the context (X,M,I), then a_{i_0} is not an extent of a formal concept, which contradicts to (ν, ζ) is an AFS formal concept (by Lemma 8.4). Similarly, for ζ, there exists unique ν such that $(\nu, \zeta) \in \mathscr{L}(E^{\#}X, EM, I)$. □

Definition 8.9. Let X be a set and M be a set of attributes on X. Let (X,M,I) be a context and $\mathscr{L}(E^{\#}X, EM, I)$ be the set of all AFS formal concepts of the context (X,M,I). Let $(\nu_1, \zeta_1), (\nu_2, \zeta_2) \in \mathscr{L}(E^{\#}X, EM, I)$. Define $(\nu_1, \zeta_1) \leq (\nu_2, \zeta_2)$ if and only if $\nu_1 \leq \nu_2$ in lattice $E^{\#}X$ (or equivalently $\zeta_1 \leq \zeta_2$ in lattice EM).

It is obvious that $\leq$ defined by Definition 8.9 is a partial order on $\mathscr{L}(E^{\#}X, EM, I)$. The following theorem shows that the set $\mathscr{L}(E^{\#}X, EM, I)$ forms a complete lattice.

Theorem 8.9. *Let X be a set and M be a set of attributes on X. Let (X,M,I) be a context and $\mathscr{L}(E^{\#}X, EM, I)$ be the set of all AFS formal concepts of the context (X,M,I). Then $\mathscr{L}(E^{\#}X, EM, I, \leq)$ is a complete lattice in which suprema and infima are given as follows: for any $(\nu_k, \zeta_k) \in \mathscr{L}(E^{\#}X, EM, I)$,*

$$\bigvee_{k\in K} (\nu_k, \zeta_k) = \left(\bigvee_{k\in K} \alpha(\zeta_k),\ \beta\left(\bigvee_{k\in K} \alpha(\zeta_k) \right) \right), \tag{8.16}$$

$$\bigwedge_{k\in K} (\nu_k, \zeta_k) = \left(\bigwedge_{k\in K} \alpha(\zeta_k),\ \beta\left(\bigwedge_{k\in K} \alpha(\zeta_k) \right) \right), \tag{8.17}$$

where $k \in K$, K is any non-empty indexing set.

Proof. In order to show that $\mathscr{L}(E^{\#}X, EM, I, \leq)$ is a complete lattice, we need to show that any subset of $\mathscr{L}(E^{\#}X, EM, I)$ has a least upper bound (suprema) and a greatest lower bound (infima). Let $S=\{(\nu_k,\zeta_k)\mid k\in K\}$ be any subset of $\mathscr{L}(E^{\#}X,EM,I)$. Let $\zeta_k=\sum_{s_k\in J_k}(\prod_{m\in A_{ks_k}}m), k\in K$, J_k be the indexing set associating to ζ_k. We claim,

$$suprema=\left(\bigvee_{k\in K}\alpha(\zeta_k),\ \ \beta\left(\bigvee_{k\in K}\alpha(\zeta_k)\right)\right),$$

$$infima=\left(\bigwedge_{k\in K}\alpha(\zeta_k),\ \ \beta\left(\bigwedge_{k\in K}\alpha(\zeta_k)\right)\right).$$

First, we show that $suprema=(\bigvee_{k\in K}\alpha(\zeta_k),\ \ \beta(\bigvee_{k\in K}\alpha(\zeta_k)))$. By Theorem 8.7, we have

$$\begin{aligned}
&\alpha\left(\beta\left(\bigvee_{k\in K}\alpha(\zeta_k)\right)\right)\\
&=\alpha\left(\beta\left(\sum_{k\in K}\alpha\left(\sum_{s_k\in J_k}(\prod_{m\in A_{ks_k}}m)\right)\right)\right)=\alpha\left(\sum_{k\in K}\beta\left(\alpha\left(\sum_{s_k\in J_k}(\prod_{m\in A_{ks_k}}m)\right)\right)\right)\\
&=\sum_{k\in K}\sum_{s_k\in J_k}\alpha\left(\beta\left(\alpha\left(\prod_{m\in A_{ks_k}}m\right)\right)\right)=\sum_{k\in K}\sum_{s_k\in J_k}\alpha\left(\prod_{m\in A_{ks_k}}m\right)\\
&=\sum_{k\in K}\alpha\left(\sum_{s_k\in J_k}\prod_{m\in A_{ks_k}}m\right)=\bigvee_{k\in K}\alpha(\zeta_k).
\end{aligned}$$

This implies $(\bigvee_{k\in K}\alpha(\zeta_k),\beta(\bigvee_{k\in K}\alpha(\zeta_k)))\in\mathscr{L}(E^{\#}X,EM,I)$, i.e., it is an AFS formal concept. Moreover, for any $k\in K$, $\nu_k=\alpha(\zeta_k)\le\bigvee_{k\in K}\alpha(\zeta_k)$ holds. Furthermore, $(\bigvee_{k\in K}\alpha(\zeta_k),\ \ \beta(\bigvee_{k\in K}\alpha(\zeta_k)))$ is an upper bound for S. Let $(\nu,\zeta)\in\mathscr{L}(E^{\#}X,EM,I)$ and for any $k\in K$, $(\nu_k,\zeta_k)\le(\nu,\zeta)$, i.e., (ν,ζ) is another upper bound for S. It is easy to get $\nu_k=\alpha(\zeta_k)\le\nu$ for any $k\in K$. Therefore, $\vee_{k\in K}\alpha(\zeta_k)\le\nu$ and $(\bigvee_{k\in K}\alpha(\zeta_k),\ \ \beta(\bigvee_{k\in K}\alpha(\zeta_k)))\le(\nu,\zeta)$, i.e.,

$$suprema=\left(\bigvee_{k\in K}\alpha(\zeta_k),\ \ \beta\left(\bigvee_{k\in K}\alpha(\zeta_k)\right)\right).$$

Next, we show that $infima=(\bigwedge_{k\in K}\alpha(\zeta_k),\ \ \beta(\bigwedge_{k\in K}\alpha(\zeta_k)))$. Since $E^{\#}X$ is a complete distributive lattice according to Theorem 5.2. Hence for any $k\in K$, one has

$$\bigwedge_{k\in K}\alpha(\zeta_k)=\sum_{f\in\Theta}\bigcap_{k\in K}\alpha(A_{kf(k)})$$

where $\Theta=\{f|f:K\to\bigcup_{k\in K}J_k\ s.t.\ f(k)\in J_k\}$. Thus by the definitions of α,β (i.e., (8.14) and (8.15)), we have

$$\begin{aligned}
&\alpha\left(\beta\left(\bigwedge_{k\in K}\alpha(\zeta_k)\right)\right)\\
&=\alpha\left(\beta\left(\sum_{f\in\Theta}\bigcap_{k\in K}\alpha(A_{kf(k)})\right)\right)=\alpha\left(\beta\left(\sum_{f\in\Theta}\alpha\left(\bigcup_{k\in K}A_{kf(k)}\right)\right)\right)\\
&=\alpha\left(\sum_{f\in\Theta}\prod_{m\in\beta\cdot\alpha\left(\bigcup_{k\in K}A_{kf(k)}\right)}m\right)=\sum_{f\in\Theta}\alpha\cdot\beta\cdot\alpha\left(\bigcup_{k\in K}A_{kf(k)}\right)\\
&=\sum_{f\in\Theta}\alpha\left(\bigcup_{k\in K}A_{kf(k)}\right)=\sum_{f\in\Theta}\bigcap_{k\in K}\alpha(A_{kf(k)})=\bigwedge_{k\in K}\alpha(\zeta_k).
\end{aligned}$$

This shows $(\bigwedge_{k\in K}\alpha(\zeta_k),\ \beta(\bigwedge_{k\in K}\alpha(\zeta_k)))\in\mathscr{L}(E^{\#}X,EM,I)$, i.e., it is an AFS formal concept. Moreover, for any $k\in K$, $\bigwedge_{k\in K}\alpha(\zeta_k)\le\nu_k=\alpha(\zeta_k)$ holds, so $(\bigwedge_{k\in K}\alpha(\zeta_k),\beta(\bigwedge_{k\in K}\alpha(\zeta_k)))$ is a lower bound for S. Let $(\nu,\zeta)\in\mathscr{L}(E^{\#}X,EM,I)$ and for any $k\in K$, $(\nu_k,\zeta_k)\ge(\nu,\zeta)$, i.e., (ν,ζ) is another lower bound for S. This implies that for any $k\in K$, $\nu_k=\alpha(\zeta_k)\ge\nu$. Therefore, both $\bigwedge_{k\in K}\alpha(\zeta_k)\ge\nu$ and

$$\left(\bigwedge_{k\in K}\alpha(\zeta_k),\ \beta\left(\bigwedge_{k\in K}\alpha(\zeta_k)\right)\right)\ge(\nu,\zeta)$$

hold, i.e., $infima=(\bigwedge_{k\in K}\alpha(\zeta_k),\beta(\bigwedge_{k\in K}\alpha(\zeta_k)))$. □

In an AFS formal concept, its intent is a complex attribute of EM; correspondingly, its extent is an element of $E^{\#}X$. The extent and intent of an AFS formal concept can uniquely determine each other. Given some extents $a_1,a_2,\ldots,a_n$ of formal concept [10], we can find a unique $\zeta\in EM$, as the intent of the AFS concept with extent $\sum_{i=1}^{n}a_i$. ζ is a semantic description of $\sum_{i=1}^{n}a_i$. On the contrary, given some intents $A_1,A_2,\ldots,A_n$ of formal concept [10], we can find a unique $\nu\in E^{\#}X$, as the extent of the AFS concept with intent $\sum_{i=1}^{n}(\prod_{m\in A_i}m)$. ν is uniquely suitable to the description of $\sum_{i=1}^{n}(\prod_{m\in A_i}m)$.

Remark 8.1. By using AFS formal concepts, we can avoid the following two issues discussed above.

1. In the AFS formal concept, intent and extent can be uniquely determined by each other (Theorem 8.8), and there exists a bijection between each item of intent and each item of extent (Lemma 8.4).
2. AFS formal concept is based on the EI algebra $(EM,\vee,\wedge)$ and the $E^{\#}I$ algebra $(E^{\#}X,\vee,\wedge)$. In $(EM,\vee,\wedge)$, we can consider whether two complex attributes are equivalent or not under the semantics (Definition 4.1)existing in an information table. Thus we can filter some complex attributes without loss of main information. For instance, let us continue discussing items $\{4\}$, $\{4,5,6,8,10,12,13\}$, $\{5,6,7,12,13\}$ and $abcdefhl,ef,fhix$ in Table 8.3. In terms of the AFS algebra,

$abcdefhl + ef + fhix = ef + fhix$ (Definition 4.1). Thus items $\{4\}$, $\{4, 5, 6, 8, 10, 12, 13\}$, $\{5,6,7,12,13\}$ and $abcdefhl, ef, fhix$ can consist of an AFS formal concept ($\{4, 5, 6, 8, 10, 12, 13\}$+ $\{5,6,7,12,13\}$, $ef + fhix$). Moreover, $\{5,6,7,12,13\} \cup \{4,5,6,8,10,12,13\}$ is just identical with extent of ($\{4, 5, 6, 7, 8, 10, 12, 13\}$, $abcdefhl \vee ef \vee fhix$). Then, AFS formal concept have not lost a crucial original information, although the intents of AFS formal concepts are usually simpler than those of monotone concepts. Thus AFS formal concept constitutes an improvement of the monotone concept.

In general, not all queries are exactly contained in an information system, but there exist many words (or phrases) close to those. For example, in Example 8.3, there does not exist an AFS formal concept with intent f+cd, but AFS formal concepts with intent $ef + cdf$ and $fh + cdf$ exist in information Table 8.3. Accordingly, we study how to approximate a complex attribute in EM (or an element in $E^{\#}X$) by AFS formal concepts. In next section, we will investigate this issue in terms of rough set theory.

8.3.3 Rough Set Theory Approach to Concept Approximation

Let (X,M,I) be a context. Inspired by [44], for each $m \in M$, denote set

$$Im = \{x \in X \mid (x,m) \in I\}$$

represent all objects that possess the attribute m. Define a binary relation R_I over M as follows, for any $m_i, m_j \in M$,

$$(m_i, m_j) \in R_I \Leftrightarrow Im_i = Im_j. \tag{8.18}$$

That is to say, two attributes are related under R_I if and only if they are possessed by the same object set. It is easy to demonstrate that R_I is an equivalence relation over M. Denote M/R_I to be the set of all equivalence classes deduced by R_I over M, i.e., $M/R_I = \{[m_i] \mid m_i \in M\}$, where $[m_i] = \{m_j \mid (m_i, m_j) \in R_I\} = \{m_j \mid Im_i = Im_j\}$.

Similarity, we can define an equivalence relation T_I over X:

$$(x_i, x_j) \in T_I \Leftrightarrow x_iI = x_jI, \tag{8.19}$$

where $x_i, x_j \in X$, $x_iI = \{m \in M \mid (x_i, m) \in I\}$ represent all attributes which are possessed by the object x_i. X/T_I be the set of all equivalence classes deduced by T_I over X, i.e., $X/T_I = \{[x_i] \mid x_i \in X\}$, where $[x_i] = \{x_j \mid (x_i, x_j) \in T_I\} = \{x_j \mid x_iI = x_jI\}$.

The lower and upper approximations of subset B of M in the approximation space $\mathscr{A} = (M, R_I)$ defined by (6.1) are listed as follows:

$$A_*(B) = \{m \in M \mid [m] \subseteq B\} = \bigcup\{Y \in M/R_I \mid Y \subseteq B\}, \tag{8.20}$$

$$A^*(B) = \{m \in M \mid [m] \cap B \neq \varnothing\} = \bigcup\{Y \in M/R_I \mid Y \cap B \neq \varnothing\}. \tag{8.21}$$

Similarity, the lower and upper approximations of subset a of X in the approximation space $\mathscr{A} = (X, T_I)$ defined by (6.1) are listed as follows:

$$A_*(a) = \{x \in X \mid [x] \subseteq X\} = \bigcup\{z \in X/T_I \mid z \subseteq a\}, \tag{8.22}$$

$$A^*(a) = \{x \in X \mid [x] \cap a \neq \varnothing\} = \bigcup\{z \in X/T_I \mid z \cap a \neq \varnothing\}. \tag{8.23}$$

Definition 8.10. Let X be a set and M be a set of attributes on X. Let (X,M,I) be a context, EM be the EI algebra over the set X and $E^\#X$ be the $E^\#I$ algebra over the set X. For any $\psi = \sum_{i\in I}(\prod_{m\in B_i} m) \in EM$, $\underline{\psi}$ the *lower approximation* and $\overline{\psi}$ the *upper approximation of the fuzzy concept* ψ are given in the form:

$$\underline{\psi} = \sum_{i\in I}(\prod_{m\in A^*(B_i)} m) \in EM, \quad \overline{\psi} = \sum_{i\in I}(\prod_{m\in A_*(B_i)} m) \in EM. \tag{8.24}$$

For any $\theta = \sum_{i\in I} a_i \in E^\#M$, $\underline{\theta}$ the *lower approximation* and $\overline{\theta}$ the *upper approximation of the $E^\#I$ algebra element* θ are defined as follows.

$$\underline{\theta} = \sum_{i\in I} A_*(a_i) \in E^\#M, \quad \overline{\theta} = \sum_{i\in I} A^*(a_i)) \in E^\#M. \tag{8.25}$$

Proposition 8.7. *Let X be a set and M be a set of attributes on X. Let (X,M,I) be a context, EM be the EI algebra over the set X and $E^\#X$ be the $E^\#I$ algebra over the set X. Then the following assertions hold.*

(1) for any $\psi_1, \psi_2, \gamma \in EM$,

$$\underline{\gamma} \leq \gamma \leq \overline{\gamma},$$
$$\overline{(\psi_1 \vee \psi_2)} = \overline{(\psi_1)} \vee \overline{(\psi_2)}\ ,\ \underline{(\psi_1 \vee \psi_2)} = \underline{(\psi_1)} \vee \underline{(\psi_2)},$$
$$\overline{(\psi_1 \wedge \psi_2)} \leq \overline{(\psi_1)} \wedge \overline{(\psi_2)}\ ,\ \underline{(\psi_1 \wedge \psi_2)} = \underline{(\psi_1)} \wedge \underline{(\psi_2)}.$$

(2) for any $\theta_1, \theta_2, \vartheta \in E^\#X$,

$$\underline{\vartheta} \leq \vartheta \leq \overline{\vartheta},$$
$$\overline{(\theta_1 \vee \theta_2)} = \overline{(\theta_1)} \vee \overline{(\theta_2)}\ ,\ \underline{(\theta_1 \vee \theta_2)} = \underline{(\theta_1)} \vee \underline{(\theta_2)},$$
$$\overline{(\theta_1 \wedge \theta_2)} \leq \overline{(\theta_1)} \wedge \overline{(\theta_2)}\ ,\ \underline{(\theta_1 \wedge \theta_2)} = \underline{(\theta_1)} \wedge \underline{(\theta_2)}.$$

Its proof is left to the reader. Whether the upper and lower approximations defined by (8.24) and (8.25) have the same properties as the upper and lower approximation defined by (6.1) remains an open problem.

Let (X,M,I) be a context, M be set of elementary attributes, $B_i \subseteq M$, $\psi = \sum_{i\in I}(\prod_{m\in B_i} m) \in EM$ be non-feasible, i.e., $\beta(\alpha(\psi)) \neq \psi$ (Definition 8.8). We are interested in finding AFS formal concepts whose intents approximate ψ. Let $L(\psi)$ and $U(\psi)$ be two AFS formal concepts, whose intents are the *lower and upper approximations of* ψ respectively, as follows:

$$L(\psi)=\left(\sum_{i\in I}\alpha(A^*(B_i)),\ \sum_{i\in I}\prod_{m\in\beta\cdot\alpha(A^*(B_i))} m\right)\in\mathscr{L}(E^{\#}X,EM,I),\quad (8.26)$$

$$U(\psi)=\left(\sum_{i\in I}\alpha(A_*(B_i)),\ \sum_{i\in I}\prod_{m\in\beta\cdot\alpha(A_*(B_i))} m\right)\in\mathscr{L}(E^{\#}X,EM,I).\quad (8.27)$$

where $A_*(B_i)$, $A^*(B_i)$ defined by (8.20) and (8.21), respectively. "α,β" is Galois connection defined by (8.1) and (8.2). The following Proposition 8.8 shows that $L(\psi)$ and $U(\psi)$ are AFS formal concepts of the context (X,M,I). $L(\psi)$ is called the *lower AFS formal concept approximation of the fuzzy concept* ψ and $U(\psi)$ is called the *upper AFS formal concept approximation of the fuzzy concept* ψ.

Proposition 8.8. *Let X be a set and M be a set of attributes on X. Let (X,M,I) be a context, EM be the EI algebra over the set X and $E^{\#}X$ be the $E^{\#}I$ algebra over the set X. Then for any $\psi=\sum_{i\in I}(\prod_{m\in B_i} m)\in EM$, the following assertions hold for the lower and upper AFS formal concept approximations of the fuzzy concept ψ:*

$$L(\psi)=\left(\sum_{i\in I}\alpha(A^*(B_i))),\ \sum_{i\in I}\prod_{m\in\beta\cdot\alpha(A^*(B_i))} m\right)=(\alpha(\underline{\psi}),\beta\cdot\alpha(\underline{\psi})),$$

$$U(\psi)=\left(\sum_{i\in I}\alpha(A_*(B_i)),\ \sum_{i\in I}\prod_{m\in\beta\cdot\alpha(A_*(B_i))} m\right)=(\alpha(\overline{\psi}),\beta\cdot\alpha(\overline{\psi})).$$

where α and β defined by (8.14) and (8.15), respectively. $\underline{\psi}$ and $\overline{\psi}$ defined by (8.24).

The proof of this proposition remains as an exercise. By Proposition 8.8, Definition 8.7 and Theorem 8.7, we know that both $L(\psi)$ and $U(\psi)$ are AFS formal concepts of the context (X,M,I).

Proposition 8.9. *Let X be a set and M be a set of attributes on X. Let (X,M,I) be a context, EM be the EI algebra over the set X and $E^{\#}X$ be the $E^{\#}I$ algebra over the set X. Then the following assertions hold:*

(1) For any $\psi\in EM$, $L(\psi)\le(\alpha(\psi),\beta(\alpha(\psi)))\le U(\psi)$, where α,β defined by (8.14) and (8.15);
(2) For $\psi_1,\psi_2\in EM$, $\psi_1\le\psi_2\Rightarrow L(\psi_1)\le L(\psi_2)$, $U(\psi_1)\le U(\psi_2)$,

where $L(.)$ and $U(.)$ are defined by (8.26) and (8.27), respectively.

Proof. (1) Let $\psi=\sum_{i\in I}(\prod_{m\in A_i} m)\in EM$. For any $i\in I$, we can get $A_*(A_i)\subseteq A_i\subseteq A^*(A_i)$ from the formulas (8.20) and (8.21). By using properties of the Galois connection "α,β" Proposition 8.1, we have $\alpha(A_*(A_i))\supseteq\alpha(A_i)\supseteq\alpha(A^*(A_i))$. From the definition of AFS formal concept (Definition 8.7) and the formulas (8.26)–(8.27), we get $L(\psi)\le(\alpha(\psi),\beta(\alpha(\psi)))\le U(\psi)$.

(2) Let $\psi_1=\sum_{j\in J}(\prod_{m\in B_j} m)$, $\psi_2=\sum_{i\in I}(\prod_{m\in A_i} m)\in EM$. Since $\psi_1\le\psi_2$, hence for any $j\in J$, there exists an $i\in I$ such that $A_i\subseteq B_j$. By proposition 6.1, $A_*(A_i)\subseteq$

$A_*(B_j)$. From the definition of AFS formal concept (Definition 8.7) and the formulas (8.26)–(8.27), we get $L(\psi_2) \leq L(\psi_1)$. Similarly, we obtain $U(\psi_2) \leq U(\psi_1)$. □

In Example 8.3, we compare AFS formal concept approximations with results in [44].

Example 8.3. Let X be a set and M be a set of attributes on X. Consider the context (X,M,I) given in Table 8.3. An "×" is placed in the p-th row and q-th column to indicate that object p has attribute q. Let $B=\{f,h,i\}$, from (8.18), (8.20) and (8.21), one can get that $M/R_I=\{ab,c,d,e,f,h,ix,j,k,l\}$. In the approximation space $\mathscr{A}=(X,R_I)$, $A_*(B)=\{f,h\}$, $A^*(B)=\{f,h,i,x\}$. Let $\varphi = fhi \in EM$. Then owing to formulas (8.26)–(8.27), we have

$$\begin{aligned} L(\varphi) &= (\{5,6,7,12,13\}, fhix), \\ U(\varphi) &= (\{2,3,4,5,6,7,8,11,12,13\}, fh). \end{aligned}$$

The authors in [44] gave an example on approximating a non-feasible monotone formula in which $\psi = ab \vee bcd \vee e \vee hi \vee fhi$. Due to $\alpha \cdot \beta(\{a,b\}) = \alpha(\{4\}) = \{a,b,c,d,e,f,h,l\} \neq \{a,b\}$, ψ is non-feasible. $L(\psi)$ and $U(\psi)$ are computed as illustrated in Table 8.4.

Table 8.4 The lower and upper approximation of ψ [44]

i	B_i	$A^*(B_i)$	$A_*(B_i)$	$L(B_i)$	$U(B_i)$
1	$\{a,b\}$	$\{a,b\}$	$\{a,b\}$	({4}, $abcdefhl$)	({4}, $abcdefhl$)
2	$\{b,c,d\}$	$\{a,b,c,d\}$	$\{c,d\}$	({4}, $abcdefhl$)	({4,5,8,9,10,12,13}, cdf)
3	$\{e\}$	$\{e\}$	$\{e\}$	({4,5,6,8,10,12,13}, ef)	({4,5,6,8,10,12,13}, ef)
4	$\{h,i\}$	$\{h,i,x\}$	$\{h\}$	({5,6,7,12,13}, $fhix$)	({2,3,4,5,6,7,8,11,12,13}, fh)
5	$\{f,h,i\}$	$\{f,h,i,x\}$	$\{f,h\}$	({5,6,7,12,13}, $fhix$)	({2,3,4,5,6,7,8,11,12,13}, fh)

The authors concluded that

$$\begin{aligned} L(\psi) &= (\{4\} \cup \{4,5,6,8,10,12,13\} \cup \{5,6,7,12,13\}, \\ &\qquad abcdefhl \vee ef \vee fhix) \\ &= (\{4,5,6,7,8,10,12,13\},\ abcdefhl \vee ef \vee fhix), \\ U(\psi) &= (\{4\} \cup \{4,5,8,9,10,12,13\} \cup \{4,5,6,8,10,12,13\} \\ &\qquad \cup \{2,3,4,5,6,7,8,11,12,13\},\ abcdefhl \vee cdf \vee ef \vee fh) \\ &= (\{2,3,4,5,6,7,8,9,10,11,12,13\},\ abcdefhl \vee cdf \vee ef \vee fh). \end{aligned}$$

However, by using Definition 4.1, we find that in EM

$$\begin{aligned} abcdefhl + ef + fhix &= ef + fhix, \\ abcdefhl + cdf + ef + fh &= cdf + ef + fh. \end{aligned}$$

By Definition 5.3, we find that in $E^{\#}X$

$$\begin{aligned}&\{4\}+\{4,5,6,8,10,12,13\}+\{5,6,7,12,13\}\\&\quad=\{4,5,6,8,10,12,13\}+\{5,6,7,12,13\},\\&\{4\}+\{4,5,8,9,10,12,13\}+\{4,5,6,8,10,12,13\}+\{2,3,4,5,6,7,8,11,12,13\}\\&\quad=\{4,5,8,9,10,12,13\}+\{4,5,6,8,10,12,13\}+\{2,3,4,5,6,7,8,11,12,13\}.\end{aligned}$$

By the formulas $L(\psi)$ and $U(\psi)$, we can get that

$$\begin{aligned}L(\psi)&=(\{4\}+\{4,5,6,8,10,12,13\}+\{5,6,7,12,13\},\\&\qquad abcdefhl+ef+fhix)\\&=(\{4,5,6,8,10,12,13\}+\{5,6,7,12,13\},ef+fhix),\end{aligned}$$

and

$$\begin{aligned}U(\psi)&=(\{4\}+\{4,5,8,9,10,12,13\}+\{4,5,6,8,10,12,13\}\\&\qquad+\{2,3,4,5,6,7,8,11,12,13\},abcdefhl+cdf+ef+fh)\\&=(\{4,5,8,9,10,12,13\}+\{4,5,6,8,10,12,13\}\\&\qquad+\{2,3,4,5,6,7,8,11,12,13\},cdf+ef+fh).\end{aligned}$$

It is easy to verify that ψ, $L(\psi)$, $U(\psi)$ satisfy (1) of Proposition 8.9.

Remark 8.2. From Example 8.3, one can observe that the semantics of the intents of the lower and upper approximations of ψ by AFS formal concepts are equivalent to those of ψ by monotone concepts. However, the extents of them are different, and the extents of AFS formal concepts preserve more information than those of monotone concepts. In addition, the semantic equivalence and logic operations are introduced in AFS formal concepts. These are more conveniently to represent the logic operations of queries in information retrieval.

Let (X,M,I) be a context, $a_i \subseteq X$, $\theta=\sum_{i\in I} a_i \in E^{\#}X$. We are interested in finding AFS formal concepts whose extents approximate θ. Let $L(\theta)$ and $U(\theta)$ be two AFS formal concepts, whose extents represent the lower and upper approximations of θ, respectively, as follows

$$L(\theta)=\left(\sum_{i\in I}\alpha\cdot\beta(A_*(a_i)),\ \sum_{i\in I}\prod_{m\in\beta(A_*(a_i))} m\right)\in\mathscr{L}(E^{\#}X,EM,I), \quad (8.28)$$

$$U(\theta)=\left(\sum_{i\in I}\alpha\cdot\beta(A^*(a_i)),\ \sum_{i\in I}\prod_{m\in\beta(A^*(a_i))} m\right)\in\mathscr{L}(E^{\#}X,EM,I). \quad (8.29)$$

where $A_*(a_i)$ and $A^*(a_i)$ defined by (8.22) and (8.23), respectively. "α,β" are Galois connection defined by (8.14) and (8.15). The following Proposition 8.10 shows that

$L(\theta)$ and $U(\theta)$ are AFS formal concepts of the context (X,M,I). $L(\theta)$ is called the *lower AFS formal concept approximation of the $E^{\#}I$ algebra element θ* and $U(\psi)$ is called the *upper AFS formal concept approximation of the $E^{\#}I$ algebra element θ.*

Proposition 8.10. *Let X be a set and M be a set of attributes on X. Let (X,M,I) be a context, EM be the EI algebra over the set X and $E^{\#}X$ be the $E^{\#}I$ algebra over the set X. Then for any $\theta = \sum_{i\in I} a_i \in E^{\#}X$, the following assertions hold for the lower and upper AFS formal concept approximations of θ:*

$$L(\psi) = \left(\sum_{i\in I} \alpha\cdot\beta(A_*(a_i)),\ \sum_{i\in I}\prod_{m\in\beta(A_*(a_i))} m\right) = (\alpha\cdot\beta(\underline{\theta}),\ \beta(\underline{\theta})),$$

$$U(\psi) = \left(\sum_{i\in I} \alpha\cdot\beta(A^*(a_i)),\ \sum_{i\in I}\prod_{m\in\beta(A^*(a_i))} m\right) = (\alpha\cdot\beta(\overline{\theta}),\ \beta(\overline{\theta})).$$

where α and β defined by (8.14) and (8.15), respectively. $\underline{\theta}$ and $\overline{\theta}$ defined by (8.25).

The proof of this proposition remains as an exercise. By Proposition 8.10, Definition 8.7 and Theorem 8.7, we know that both $L(\theta)$ and $U(\theta)$ are AFS formal concepts of the context (X,M,I).

Proposition 8.11. *Let X be a set and M be a set of attributes on X. Let (X,M,I) be a context, EM be the EI algebra over the set X and $E^{\#}X$ be the $E^{\#}I$ algebra over the set X. Then the following assertions hold:*

(1) For any $\theta \in E^{\#}X$, $L(\theta) \le (\alpha\cdot\beta(\theta),\ \ \beta(\theta)) \le U(\theta)$, where α, β defined by (8.14) and (8.15);
(2) For $\theta_1, \theta_2 \in EM$, $\theta_1 \le \theta_2 \Rightarrow L(\theta_1) \le L(\theta_2)$, $U(\theta_1) \le U(\theta_2)$,

where $L(.)$ and $U(.)$ defined by (8.28) and (8.29), respectively.

Example 8.4. Let X be a set and M be a set of attributes on X. Consider the context (X,M,I) given in Table 8.3. From formula (8.19), one obtains

$$X/T_I = \{\{1\},\{2\},\{3,11\},\{4\},\{5,12,13\},\{6\},\{7\},\{8\},\{9\},\{10\}\}.$$

Let $\theta = \sum_{i\in I} a_i = \{2,3\} + \{4\} + \{5,6,7\} \in E^{\#}X$. From formulas (8.22) and (8.23), $L(\theta)$ and $U(\theta)$ are computed as presented in Table 8.5.

Table 8.5 The lower and upper approximation of θ

i	a_i	$A_*(a_i)$	$A^*(a_i)$	$L(a_i)$	$U(a_i)$
1	{2,3}	{2}	{2,3,11}	({2,4,5,8,12,13}, dfh)	({2,3,4,5,6,7,8,11,12,13}, fh)
2	{4}	{4}	{4}	({4}, $abcdefhl$)	({4}, $abcdefhl$)
3	{5,6,7}	{6,7}	{5,6,7,12,13}	({6,7}, $fhijx$)	({5,6,7,12,13}, $fhix$)

Therefore, from properties of EI, $E^{\#}I$ algebra and the formulas (8.28)-(8.29), we get

$$\begin{aligned} L(\theta) &= (\{2,4,5,8,12,13\}+\{4\}+\{6,7\}, dfh+abcdefhl+fhijx) \\ &= (\{2,4,5,8,12,13\}+\{6,7\}, dfh+fhijx), \end{aligned}$$

and

$$\begin{aligned} U(\theta) = (&\{2,3,4,5,6,7,8,11,12,13\}+\{4\}+\{5,6,7,12,13\}, \\ &fh+abcdefhl+fhix) \\ = (&\{2,3,4,5,6,7,8,11,12,13\}, fh). \end{aligned}$$

Remark 8.3. The extent and intent of AFS formal concept can uniquely determine each other. Thus, concept approximation by AFS formal concepts can avoid the issues i) and ii) stated in the abvoe section of monotone concepts and is more conveniently for query. When approximating $\{2,3,4,5,6,7,8,11,12,13\}+\{4,5,8,9,10,12,13\}+\{4,5,6,8,10,12,13\}$ by AFS formal concepts, we get that $(\{4,5,8,9,10,12,13\}+\{4,5,6,8,10,12,13\}+\{2,3,4,5,6,7,8,11,12,13\}, cdf+ef+fh)$ instead of (D,f), where $D=\{2,3,4,5,6,7,8,9,10,11,12,13\}$. When approximating D by AFS formal concepts, one also can obtain that (D,f) by union all of the items of extent of the AFS formal concept, which is the same as the approximation realized by monotone concepts.

In this section, the AFS formal concept is proposed, which can be more conveniently applied to represent query in information retrieval systems than both the monotone concept and the formal concept. The set of all AFS formal concepts forms a complete lattice. Furthermore, by virtue of rough set theory, we discuss how to find AFS formal concepts whose intents (extents) approximate a fuzzy concept in EM (or an element of $E^{\#}X$). The examples and remarks demonstrate that not only the forms of approximation results by using AFS formal concepts may be concise, but they do not lead to any loss of crucial information. In this way, the AFS formal concepts can be viewed as the generalization of the monotone concept and the formal concept.

8.4 AFS Fuzzy Formal Concept Analysis

In the above sections, the set M in any context (G,M,I) is a set of Boolean attributes on X. However, in the real world applications the set M often represents a set of fuzzy or Boolean attributes. Given this, in the this section, we show that any context $(G,M,\mathbb{I})$ with fuzzy attributes in M, where $\mathbb{I}$ stresses that there are fuzzy attributes in M, can be described by an AFS structure. Let G be a set of objects and M be a set of fuzzy or Boolean attributes. $\forall g_1,g_2 \in G$, τ is defined by

$$\tau(g_1,g_2)=\{m|m\in M,(g_1,g_2)\in R_m\},$$

where $(g_1,g_2)\in R_m$ (refer to Definition 4.2)$\Leftrightarrow$ g_1 belongs to attribute m at some degree and the degree of g_1 belonging to m is larger than or equal to that of g_2, or

g_1 belongs to m at some degree and g_2 does not at all. For a given context (G,M,I), we can establish an AFS structure (M,τ_I,G) according to (G,M,I) in the following manner.

$$\tau_I(g_1,g_2) = \{m \in M | (g_1,g_2) \in R_m\},$$

where for $m \in M$ and binary relation $I \subseteq G \times M$, g_1 belongs to attribute m at some degree which means that $(g_1,m) \in I$. Since each $m \in M$, m is a Boolean attribute, hence $(g_1,m) \in I$ implies that the degree of g_1 belonging to m is larger than or equal to that of g_2 for any $g_2 \in G$. Therefore

$$\tau_I(g_1,g_2) = \{m \in M | (g_1,g_2) \in R_m\} = \{m \in M | (g_1,m) \in I\}.$$

Now, we discuss the AFS formal concept analysis, in which M is a set of fuzzy or Boolean attributes on X.

Definition 8.11. Let X, M be sets and (M,τ,X) be an AFS structure. A binary relation $\mathbb{I}_\tau$ from $X \times X$ to M is defined as follows: for $(x,y) \in X \times X, m \in M$,

$$((x,y),m) \in \mathbb{I}_\tau \Leftrightarrow m \in \tau(x,y). \tag{8.30}$$

It is clear that $(X \times X,\ M,\ \mathbb{I}_\tau)$ is a formal context defined by [10]. The formal context $(X \times X, M, \mathbb{I}_\tau)$ is called the *fuzzy context associating with the AFS structure* (M,τ,X).

Definition 8.12. Let X be a set and $E^\#(X \times X)$ be the $E^\#I$ algebra on $X \times X$. For any $a \subseteq X \times X$, any $x \in X$, we define

$$a^R(x) = \{y \in X \mid (x,y) \in a\} \subseteq X. \tag{8.31}$$

For any $\gamma = \sum_{i \in I} a_i \in E^\#(X \times X)$, the $E^\#I$ algebra valued membership function $\gamma^R : X \to E^\#X$ is defined as follows: for any $x \in X$,

$$\gamma^R(x) = \sum_{i \in I} a_i^R(x) \in E^\#X. \tag{8.32}$$

By the fuzzy norm (5.24) with $\mathscr{M}_\rho$ the measure shown as (5.16) for the function $\rho : X \to [0,+\infty)$, the membership function $\mu_{\gamma^R}(x)$ of γ^R is defined as follows: for any $x \in X$,

$$\mu_{\gamma^R}(x) = ||\gamma^R(x)||_\rho = \sup_{i \in I}\{\mathscr{M}_\rho(a_i^R(x))\} \in [0,1]. \tag{8.33}$$

Thus every $\gamma \in E^\#(X \times X)$ can be regarded as a fuzzy set on X whose membership functions are defined by (8.32) or (8.33).

Since $E^\#X$ is a lattice, hence for each $\gamma \in E^\#(X \times X)$, $\gamma^R : X \to E^\#X$ defined by the formula (8.32) is a lattice valued fuzzy set. One can verify that for $\gamma, \eta \in E^\#(X \times X)$, if $\gamma \leq \eta$ in lattice $E^\#(X \times X)$, then for any $x \in X$, $\gamma^R(x) \leq \eta^R(x)$ in lattice $E^\#X$. Thus in $(X \times X,\ M,\ \mathbb{I}_\tau)$, the fuzzy context associated with the AFS structure (M,τ,X), for each attribute $\eta \in EM$, $\alpha(\eta)$ is a fuzzy set on X with the membership functions defined by (8.32) or (8.33), where "α" is the Galois connection defined by (8.15).

Contrastively, for any $\gamma \in E^{\#}(X \times X)$ as a fuzzy set defined by (8.32) or (8.33), $\beta(\gamma)$ is an attribute in EM, where "β" is the Galois connection defined by (8.14). If (γ, η) is an AFS formal concept defined by Definition 8.7, then the fuzzy set γ is the extent of (γ, η) and the attribute η, which is the AFS logic combination of the simple attributes in M and has a definitely semantic interpretation, is the intent of (γ, η).

Theorem 8.10. *Let X be a set and M be a set of simple attributes on X. Let (M, τ, X) be an AFS structure in which for any $x, y \in X, \tau(x, y) = \{m \in M \mid (x, y) \in R_m\}$ (refer to (4.26)) and $(X \times X,\ M,\ \mathbb{I}_\tau)$ be the fuzzy context associating with (M, τ, X). Then for $\zeta, \varsigma \in EM$, if $\beta(\alpha(\zeta)) = \beta(\alpha(\varsigma))$, i.e., both $\beta(\alpha(\zeta))$ and $\beta(\alpha(\varsigma))$ are the intent of an AFS formal concept, then $\forall x \in X$, $\zeta(x) = \varsigma(x)$ and $\mu_\zeta(x) = \mu_\varsigma(x)$, where "$\alpha, \beta$" are the Galois connections defined by (8.14) and (8.15); for any fuzzy attribute $\gamma = \sum_{u \in U}(\prod_{m \in C_u} m) \in EM, \gamma(x) = \sum_{u \in U} C_u^\tau(x) \in E^{\#}X$ is the $E^{\#}I$ valued membership function of γ defined by (5.13) and $\mu_\gamma(x) = ||\sum_{u \in U} C_u^\tau(x)||_\rho \in [0,1]$ is the membership function of γ defined by (5.25) for the fuzzy norm (5.24) with $\mathscr{M}_\rho$ the measure shown as (5.16) for the function $\rho : X \to [0, +\infty)$.*

Proof. According to the definitions of $(X \times X, M, \mathbb{I}_\tau)$ and the Galois connection α, for any $m \in M$, we have

$$\alpha(\{m\}) = \{(x, y) \in X \times X \mid m \in \tau(x, y)\}.$$

By Proposition 8.1 and (8.32), we can verify that for any $A \subseteq M$, any $x \in X$,

$$\begin{aligned} \alpha(A)^R(x) &= \left(\bigcap_{m \in A} \alpha(\{m\})\right)^R (x) \\ &= \left(\bigcap_{m \in A} \{(x, y) \in X \times X \mid m \in \tau(x, y)\}\right)^R (x) \\ &= (\{(x, y) \in X \times X \mid A \subseteq \tau(x, y)\})^R(x). \end{aligned} \tag{8.34}$$

By (4.27) and (8.31), we have

$$(\{(x, y) \in X \times X \mid A \subseteq \tau(x, y)\})^R(x) = A^\tau(x). \tag{8.35}$$

Furthermore for any $\gamma = \sum_{u \in U}(\prod_{m \in C_u} m) \in EM$ and any $x \in X$, from (8.35) and (8.32), one has

$$\alpha(\gamma)^R(x) = \sum_{u \in U} \alpha(C_u)^R(x) = \sum_{u \in U} C_u^\tau(x) = \gamma(x) \tag{8.36}$$

That is the $E^{\#}I$ valued membership function of the fuzzy attribute γ defined by (5.13). For $\zeta, \varsigma \in EM$, if $\beta(\alpha(\zeta)) = \beta(\alpha(\varsigma))$, then

$$\alpha(\zeta) = \alpha(\beta(\alpha(\zeta))) = \alpha(\beta(\alpha(\varsigma))) = \alpha(\varsigma).$$

It follows from (8.36) that for any $x \in X$,

$$\zeta(x) = \alpha(\zeta)^R(x) = \alpha(\varsigma)^R(x) = \varsigma(x),$$

Since $||.||$ is a fuzzy norm on the lattice $E^{\#}X$, we have

$$\mu_\zeta(x) = ||\zeta(x)||_\rho = ||\varsigma(x)||_\rho = \mu_\varsigma(x).$$ □

Assume that each attribute in M is a Boolean attribute on X. For any $m \in M$, let R_m be the binary relation of m defined by Definition 4.2. Since $m \in M$ is Boolean concept, hence for any $x \in X$, either $(x,y) \in R_m$ for any $y \in X$ or $(x,y) \notin R_m$ for any $y \in X$. By (8.31), one has that for any $m \in M$, any $x \in X$, either $\{m\}^R(x) = X$ or $\{m\}^R(x) = \varnothing$. This implies that for any $A \subseteq M$, any $x \in X$, either $A^R(x) = X$ or $A^R(x) = \varnothing$. Further, by (8.32) and (8.33), for any $\zeta \in E^{\#}(X \times X)$, any $x \in X$, either $\zeta^R(x) = X$ or $\zeta^R(x) = \varnothing$, and either $\mu_{\zeta^R}(x) = 1$ or $\mu_{\zeta^R}(x) = 0$, i.e., $\mu_{\zeta^R}(x)$ is the characteristic function of a Boolean set $C_\zeta \subseteq X$. Proposition 4.3 has showed that the AFS logic system $(EM, \vee, \wedge, ')$ will degenerate into Boolean logic system $(2^X, \cup, \cap, ')$ if every attribute in M is a Boolean attribute. Therefore if each $m \in M$ is a Boolean attribute, then the AFS formal concept lattice of an AFS structure (M, τ, X) will degenerate into the formal concept lattice of context (X, M, I), where for $x \in X$ and $m \in M$, $(x,m) \in I \Leftrightarrow ((x,y),m) \in \mathbb{I}_\tau$ for any $y \in X \Leftrightarrow (x,y) \in R_m$ for any $y \in X \Leftrightarrow x$ has attribute m (refer to Definition 4.2). For each AFS formal concept (γ, η), the intent $\eta = \sum_{i \in I}(\prod_{m \in A_i} m) \in EM$ corresponds to the disjunctive normal form of a monotone Boolean formula $\bigvee_{i \in I} A_i$, where each $\prod_{m \in A_i} m$ is a Boolean conjunctive expression $\bigwedge_{a \in A_i} a$, and the extent $\gamma \subseteq X$ is

$$\gamma = \alpha(\eta) = \bigcup_{i \in I} \bigcap_{a \in A_i} \alpha(a).$$

For instance, in Example 8.1, for instance, the attribute $\xi = m_1 + m_2m_4 + m_4m_5 \in EM$ read as "*edible*" or "*poisonous and cap-shape*" or "*cap-shape and cap-surface: fibrous*". According to Table 8.2, we know that $\{m_1\}$, $\{m_2, m_4\}$ and $\{m_4, m_5\}$ are all intents of some concepts of the context (G, M, I). From Lemma 8.4, one has that

$$\left(\alpha(\{m_1\}) + \left(\alpha(\{m_2\}) \bigcap \alpha(\{m_4\})\right) + \left(\alpha(\{m_4\}) \bigcap \alpha(\{m_5\})\right),\ \ \xi\right)$$

is an AFS formal concept of the context (X, M, I). The following Example 8.5 demonstrates how to implement AFS fuzzy formal concept analysis for a data with both fuzzy and Boolean attributes.

Example 8.5. Let $X = \{x_1, x_2, ..., x_{10}\}$ be a set of 10 people and their features (attributes) which are described by real numbers (age, height, weight, salary, estate), Boolean values (gender) and the ordered relations (hair black, hair white, hair yellow), see Table 8.6; there the number i in the "*hair color*" columns which corresponds to some $x \in X$ implies that the hair color of x has ordered ith following our perception of the color by our intuitive perception. Let $M = \{m_1, m_2, ..., m_{10}\}$ be

Table 8.6 Descriptions of features

	appearance			wealth		gender		hair color		
	age	*height*	*weigh*	*salary*	*estate*	*male*	*female*	*black*	*white*	*yellow*
x_1	20	1.9	90	1	0	1	0	6	1	4
x_2	13	1.2	32	0	0	0	1	4	3	1
x_3	50	1.7	67	140	34	0	1	6	1	4
x_4	80	1.8	73	20	80	1	0	3	4	2
x_5	34	1.4	54	15	2	1	0	5	2	2
x_6	37	1.6	80	80	28	0	1	6	1	4
x_7	45	1.7	78	268	90	1	0	1	6	4
x_8	70	1.65	70	30	45	1	0	3	4	2
x_9	60	1.82	83	25	98	0	1	4	3	1
x_{10}	3	1.1	21	0	0	0	1	2	5	3

the set of fuzzy or Boolean concepts on X and each $m \in M$ associate to a single feature. Where m_1 : *"old people"*, m_2 : *"tall people"*, m_3 : *"heavy people"*, m_4 : *"high salary"*, m_5 : *"more estate"*, m_6 :*"male"*, m_7 : *"female"*, m_8 : *"black hair people"*, m_9 : *"white hair people"*, m_{10} : *"yellow hair people"*.

Let (M,τ,X) be the AFS structure of the data shown in Table 8.6. For simplicity, let S=2^X be the σ-algebra over X and m_ρ be the measure defined by (5.16) for the a weight function $\rho(x)$=1, $\forall x \in X$. Let

$$\zeta = m_1m_3m_4 + m_1m_3m_7,\ \xi = m_1m_2m_3m_4 + m_1m_2m_3m_7$$

be two fuzzy attributes in EM. It is obvious that $\zeta \geq \xi$ and $\zeta \neq \xi$ in lattice EM. One can verify that

$$\beta(\alpha(\prod_{m\in\{m_1,m_3,m_4\}} m)) = \prod_{m\in\{m_1,m_2,m_3,m_4\}} m,$$
$$\beta(\alpha(\prod_{m\in\{m_1,m_3,m_7\}} m)) = \prod_{m\in\{m_1,m_2,m_3,m_7\}} m.$$

Although ζ and ξ are different attributes in EM, i.e., ζ and ξ capture different semantics, the fuzzy sets defined by (5.13) or the norm of the lattice $E^{\#}X$ defined by (5.24) are identical, i.e., their extents are equal as shown in Table 8.7.

Table 8.7 Membership functions of ζ and ξ defined by (8.33)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
$\mu_\zeta(.) = \mu_\xi(.)$	0.3	0.2	0.4	0.4	0.3	0.4	0.4	0.4	0.7	0.1

Let $N = \{m_1, m_2, m_3, m_6, m_7\} \subseteq M$. Here, we study the AFS fuzzy formal concept lattice $\mathscr{L}(E^{\#}(X\times X), EN, \mathbb{I}_\tau)$. According to Lemma 8.4, we know that for any AFS formal concept $(\nu,\eta) \in \mathscr{L}(E^{\#}(X\times X), EN, \mathbb{I}_\tau)$ there exist $A_i \subseteq N, i \in I$, A_i is the intent of a formal concept of context $(X\times X, N, \mathbb{I}_\tau)$ which is the fuzzy context associating with the AFS structure (N,τ,X) (refer to Definition 8.11) such that

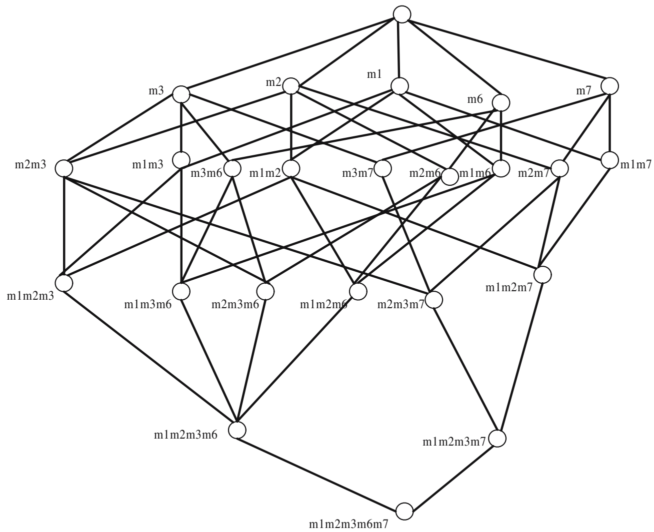


Fig. 8.1 Concept lattice of context $(X \times X, N, \mathbb{I}_\tau)$

Table 8.8 Membership functions of the extents of the formal concepts shown in Figure 8.1

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
$\mu_{m_1}(\cdot)$	0.3	0.2	0.7	1	0.4	0.5	0.6	0.9	0.8	0.1
$\mu_{m_1m_2}(\cdot)$	0.3	0.2	0.6	0.8	0.3	0.4	0.5	0.5	0.7	0.1
$\mu_{m_1m_3}(\cdot)$	0.3	0.2	0.4	0.6	0.3	0.4	0.4	0.5	0.7	0.1
$\mu_{m_1m_6}(\cdot)$	0.3	0	0	1	0.4	0	0.6	0.9	0	0
$\mu_{m_1m_7}(\cdot)$	0	0.2	0.7	0	0	0.5	0	0	0.8	0.1
$\mu_{m_1m_2m_3}(\cdot)$	0.3	0.2	0.4	0.6	0.3	0.4	0.4	0.4	0.7	0.1
$\mu_{m_1m_2m_6}(\cdot)$	0.3	0	0	0.8	0.3	0	0.5	0.5	0	0
$\mu_{m_1m_2m_7}(\cdot)$	0	0.2	0.6	0	0	0.4	0	0	0.7	0.1
$\mu_{m_1m_3m_6}(\cdot)$	0.3	0	0	0.6	0.3	0	0.4	0.5	0	0
$\mu_{m_1m_2m_3m_6}(\cdot)$	0.3	0	0	0.6	0.3	0	0.4	0.4	0	0
$\mu_{m_1m_2m_3m_7}(\cdot)$	0	0.2	0.4	0	0	0.4	0	0	0.7	0.1
$\mu_{m_2}(\cdot)$	1	0.2	0.7	0.8	0.3	0.4	0.7	0.5	0.9	0.1
$\mu_{m_2m_3}(\cdot)$	1	0.2	0.4	0.6	0.3	0.4	0.6	0.4	0.9	0.1
$\mu_{m_2m_6}(\cdot)$	1	0	0	0.8	0.3	0	0.7	0.5	0	0
$\mu_{m_2m_7}(\cdot)$	0	0.2	0.7	0	0	0.4	0	0	0.9	0.1
$\mu_{m_2m_3m_6}(\cdot)$	1	0	0	0.6	0.3	0	0.6	0.4	0	0

$\eta = \sum_{i \in I}(\prod_{m \in A_i} m)$. So we show the concept lattice generated by the fuzzy context $(X \times X, N, \mathbb{I}_\tau)$ in Figure 8.1 and the membership functions of the extents are shown in Table 8.8. Notice that although both the extents and the intents of the formal concepts in $(X \times X, N, \mathbb{I}_\tau)$ may be fuzzy sets and fuzzy attributes, $(X \times X, N, \mathbb{I}_\tau)$ is a traditional context [10]. This implies that its complexity is the same as a traditional context with $|X|^2$ objects and $|N|$ attributes.

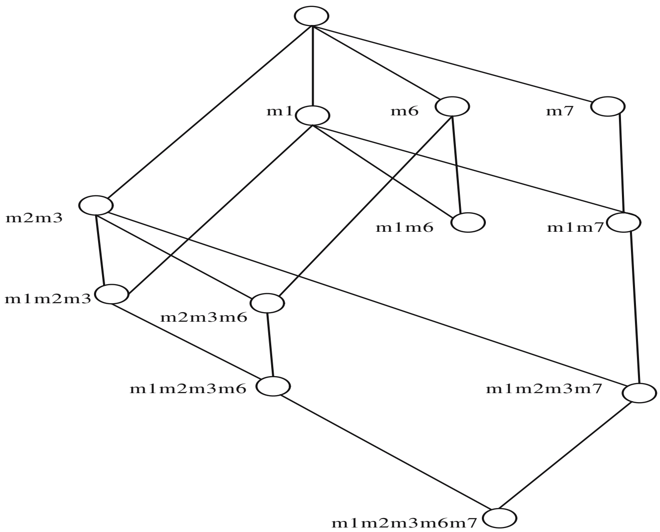


Fig. 8.2 Concept lattice of context $(X_1 \times X_1, N, \mathbb{I}_\tau)$

Table 8.9 Membership functions of the extents of the formal concept shown in Figure 8.2

	μ_{m_1}	$\mu_{m_1m_6}$	$\mu_{m_1m_7}$	$\mu_{m_1m_2m_3}$	$\mu_{m_1m_2m_3m_6}$	$\mu_{m_1m_2m_3m_7}$	$\mu_{m_2m_3}$	$\mu_{m_2m_3m_6}$	μ_{m_6}	μ_{m_7}
x_1	0.6	0.6	0	0.6	0.6	0	1	1	1	0
x_2	0.4	0	0.4	0.4	0	0.4	0.4	0	0	1
x_5	0.8	0.8	0	0.6	0.6	0	0.6	0.6	1	0
x_6	1	0	1	0.8	0	0.8	0.8	0	0	1
x_{10}	0.2	0	0.2	0.2	0	0.2	0.2	0	0	1

Let $X_1 = \{x_1, x_2, x_5, x_6, x_{10}\} \subseteq X$. Figure 8.2 shows the concept lattice generated by $(X_1 \times X_1, N, \mathbb{I}_\tau)$ and the membership functions of the extents are shown in Table 8.9. Although the intent and extent of an AFS formal concept are a fuzzy attribute in EM and a fuzzy set on X respectively, the context $(X \times X, M, \mathbb{I}_\tau)$ associating with an AFS structure (M, τ, X) is a traditional context which can be directly established by the original data without the use of the fuzzy set $\mathbb{I}$ to describe the uncertainty between the objects and the attributes. Thus the AFS formal concept lattices preserve more information contained in original data than the other fuzzy formal concept lattices. This observation stresses that the AFS formal concept analysis naturally extends the traditional formal concepts to the fuzzy formal concepts.

In order to cope with the data with various data types such as real numbers, Boolean value and even the human intuition description with sub-preferences, the AFS fuzzy formal concept analysis, which intuitively augments the traditional formal concepts to fuzzy formal concepts and overcomes the difficulties of other fuzzy formal concepts to define the fuzzy binary relation by human interpretations, is proposed and developed. The examples demonstrate that the AFS fuzzy formal concept analysis can be directly applied to the original data with both fuzzy and Boolean attributes and preserve more information contained in the original data than other fuzzy formal concepts. In the framework of AFS fuzzy formal concept analysis, the original data is only required to generate AFSFFCA lattices, human interpretation is not required to define the fuzzy binary relations and the fuzzy sets corresponding to all attributes in EM are automatically determined by a consistent algorithm according to the AFS structure and the AFS algebra. So AFSFFCA lattices are more objective and comprehensive representations of the knowledge contained in the original data than traditional and other fuzzy formal concepts. The theorems prove that AFS fuzzy formal concept lattices are more general mathematization of the traditional formal concept lattices. Many already existing mathematical tools such as topology, measure theory, combinatorics and algebras can be applied to the research of the AFS theory. These facts encourage us to derive mathematical properties of AFSFFCA and apply them to future research and development of knowledge representation schemes.

Exercises

Exercise 8.1. Let (G,M,I) be a context. Show that the following assertions hold:

(1) for $A_1,A_2 \subseteq G$, $A_1 \subseteq A_2$ implies $\beta(A_1) \supseteq \beta(A_2)$ and
for $B_1,B_2 \subseteq M$, $B_1 \subseteq B_2$ implies $\alpha(B_1) \supseteq \alpha(B_2)$;
(2) $A \subseteq \alpha(\beta(A))$ and $\beta(A) = \beta(\alpha(\beta(A)))$ for all $A \subseteq G$, and
$B \subseteq \beta(\alpha(B))$ and $\alpha(B) = \alpha(\beta(\alpha(B)))$ for all $B \subseteq M$.

Exercise 8.2. (Wille's Lemma) Let (G,M,I) be a context and $\mathscr{L}(G,M,I)$ denote the set of all formal concepts of the context (G,M,I). Show that

$$\mathscr{L}(G,M,I) = \{(\alpha(B),\beta(\alpha(B))) \mid B \subseteq M\}.$$

Exercise 8.3. (Fundamental Theorem of FCA) Let (G,M,I) be a context. Prove that $(\mathscr{L}(G,M,I),\vee,\wedge)$ is a complete lattice in which suprema and infima are given as follows: for any formal concepts $(A_j,B_j) \in \mathscr{L}(G,M,I)$, $j \in J$,

$$\bigvee_{j\in J}(A_j,B_j) = \left(\gamma_G\left(\bigcup_{j\in J}A_j\right),\ \bigcap_{j\in J}B_j\right),$$

$$\bigwedge_{j\in J}(A_j,B_j) = \left(\bigcap_{j\in J}A_j,\ \gamma_M\left(\bigcup_{j\in J}B_j\right)\right),$$

where $\gamma_G = \alpha\cdot\beta$, $\gamma_M = \beta\cdot\alpha$.

Exercise 8.4. Let X and M be sets, (G,M,I) be a context and E^IXM be defined as Definition 8.3. For $\sum_{u\in U} a_uA_u \in E^IXM$, if $a_q \subseteq a_w$, $A_q \subseteq A_w$, $w,q \in U$, $w \neq q$, prove that

$$\sum_{u\in U} a_u A_u = \sum_{u\in U, u\neq q} a_u A_u.$$

Exercise 8.5. Let X and M be sets, (G,M,I) be a context and $E^I XM$ be the set defined as Definition 8.3. Prove that the binary relation $\leq$ is a partial order relation if $\sum_{u\in U} a_u A_u$, $\sum_{v\in V} b_v B_v \in E^I XM$, $\sum_{u\in U} a_u A_u \leq \sum_{v\in V} b_v B_v \Leftrightarrow \forall a_u A_u$ $(u \in U)$ $\exists b_k B_k$ $(k \in V)$ such that $a_u \subseteq b_k$, $A_u \subseteq B_k$.

Exercise 8.6. Let (G,M,I) be a context and $E^I GM$ be defined as Definition 8.3. If for any $\sum_{u\in U} a_u A_u$, $\sum_{v\in V} b_v B_v \in E^I GM$, we define

$$\left(\sum_{u\in U} a_u A_u\right) * \left(\sum_{v\in V} b_v B_v\right) = \sum_{u\in U, v\in V} a_u \cap b_v A_u \cup B_v,$$
$$\left(\sum_{u\in U} a_u A_u\right) + \left(\sum_{v\in V} b_v B_v\right) = \sum_{u\in U\sqcup V} c_u C_u,$$

where $u \in U \sqcup V$ (the disjoint union of indexing sets U, V), $c_u = a_u$, $C_u = A_u$,if $u \in U$; $c_u = b_u$, $C_u = B_u$,if $u \in U$. Prove that "+" and "$*$" are binary compositions on $E^I GM$.

Exercise 8.7. Let G and M be finite sets, (G,M,I) be a context and $(E^I GM, *, +, \leq)$ be the $E^C II$ algebra of context (G,M,I). Show that the following assertions hold. For any ψ, ϑ, γ, $\eta \in E^I GM$,

(1) $\psi + \vartheta = \vartheta + \psi, \quad \psi * \vartheta = \vartheta * \psi;$
(2) $(\psi + \vartheta) + \gamma = \psi + (\vartheta + \gamma), \quad (\psi * \vartheta) * \gamma = \psi * (\vartheta * \gamma);$
(3) $(\psi + \vartheta) * \gamma = (\psi * \gamma) + (\vartheta * \gamma), \quad \psi * (\varnothing M) = (\varnothing M), \psi * (X\varnothing) = \psi;$
(4) If $\psi \leq \vartheta$, $\gamma \leq \eta$, then $\psi + \gamma \leq \vartheta + \gamma, \quad \psi * \gamma \leq \vartheta * \gamma;$

Exercise 8.8. Let X be a set and M be a set of attributes on X. Let (X,M,I) be a context, EM be the EI algebra over the set X and $E^\# X$ be the $E^\# I$ algebra over the set X. Show the validity of the following assertions hold.

(1) for any $\psi_1, \psi_2, \gamma \in EM$,

$$\underline{\gamma} \leq \gamma \leq \overline{\gamma},$$
$$\overline{(\psi_1 \vee \psi_2)} = \overline{(\psi_1)} \vee \overline{(\psi_2)}\ ,\ \underline{(\psi_1 \vee \psi_2)} = \underline{(\psi_1)} \vee \underline{(\psi_2)},$$
$$\overline{(\psi_1 \wedge \psi_2)} \leq \overline{(\psi_1)} \wedge \overline{(\psi_2)}\ ,\ \underline{(\psi_1 \wedge \psi_2)} = \underline{(\psi_1)} \wedge \underline{(\psi_2)}.$$

(2) for any $\theta_1, \theta_2, \vartheta \in E^\# X$,

$$\underline{\vartheta} \leq \vartheta \leq \overline{\vartheta},$$
$$\overline{(\theta_1 \vee \theta_2)} = \overline{(\theta_1)} \vee \overline{(\theta_2)}\ ,\ \underline{(\theta_1 \vee \theta_2)} = \underline{(\theta_1)} \vee \underline{(\theta_2)},$$
$$\overline{(\theta_1 \wedge \theta_2)} \leq \overline{(\theta_1)} \wedge \overline{(\theta_2)}\ ,\ \underline{(\theta_1 \wedge \theta_2)} = \underline{(\theta_1)} \wedge \underline{(\theta_2)}.$$

Exercise 8.9. Let X be a set and M be a set of attributes on X. Let (X,M,I) be a context, EM be the EI algebra over the set X and $E^{\#}X$ be the $E^{\#}I$ algebra over the set X. Let (X,M,I) be a context, $B_i \subseteq M$, $\psi = \sum_{i\in I}(\prod_{m\in B_i} m) \in EM$ be a complex attribute. Prove that for any $\psi \in EM$, the lower and upper AFS formal concept approximations of the fuzzy concept ψ satisfy the relationships

$$L(\psi) = (\sum_{i\in I} \alpha(A^*(B_i))), \sum_{i\in I} \prod_{m\in\beta\cdot\alpha(A^*(B_i))} m) = (\alpha(\underline{\psi}), \beta(\alpha(\underline{\psi})))$$
$$U(\psi) = (\sum_{i\in I} \alpha(A_*(B_i)), \sum_{i\in I} \prod_{m\in\beta\cdot\alpha(A_*(B_i))} m) = (\alpha(\overline{\psi}), \beta(\alpha(\overline{\psi})))$$

where α and β are defined by (8.14) and (8.15), respectively. $\underline{\psi}$ and $\overline{\psi}$ are defined by (8.24).

Exercise 8.10. Let X be a set and M be a set of attributes on X. Let (X,M,I) be a context, EM be the EI algebra over the set X and $E^{\#}X$ be the $E^{\#}I$ algebra over the set X. For any $\theta = \sum_{i\in I} a_i \in E^{\#}X$, show the following assertions hold for the lower and upper AFS formal concept approximations of θ:

$$L(\psi) = \left(\sum_{i\in I} \alpha\cdot\beta(A_*(a_i)),\ \sum_{i\in I} \prod_{m\in\beta(A_*(a_i))} m \right) = (\alpha\cdot\beta(\underline{\theta}),\ \beta(\underline{\theta})),$$
$$U(\psi) = \left(\sum_{i\in I} \alpha\cdot\beta(A^*(a_i)),\ \sum_{i\in I} \prod_{m\in\beta(A^*(a_i))} m \right) = (\alpha\cdot\beta(\overline{\theta}),\ \beta(\overline{\theta})).$$

where α and β are defined by (8.14) and (8.15), respectively. $\underline{\theta}$ and $\overline{\theta}$ are defined by (8.25).

Exercise 8.11. Let X be a set and M be a set of attributes on X. Let (X,M,I) be a context, EM be the EI algebra over the set X and $E^{\#}X$ be the $E^{\#}I$ algebra over the set X. Show the following assertions hold:

(1) For any $\theta \in E^{\#}X$, $L(\theta) \le (\alpha\cdot\beta(\theta),\ \ \beta(\theta)) \le U(\theta)$, where α, β defined by (8.14) and (8.15);
(2) For $\theta_1, \theta_2 \in EM$, $\theta_1 \le \theta_2 \Rightarrow L(\theta_1) \le L(\theta_2)$, $U(\theta_1) \le U(\theta_2)$,

where $L(.)$ and $U(.)$ are defined by (8.28) and (8.29), respectively.

Open problems

Problem 8.1. Let X and M be sets, (G,M,I) be a context and $(E^I XM, \le)$ be the partially ordered set defined as Definition 8.3. Show whether $(E^I XM, \le)$ is a lattice. What are the lattice operations $\vee$ and $\wedge$?

Problem 8.2. Demonstrate whether the upper and lower approximations defined by (8.24) and (8.25) have the same properties as the upper and lower approximation defined by (6.1).

Problem 8.3. Discuss whether the upper and lower AFS formal concept approximations defined by (8.26) and (8.27) (or (8.28) and (8.29)) have the same properties as the upper and lower approximation defined by (6.1).

Problem 8.4. Let X be a set and M be a set of attributes on X. Let (X,M,I) be a context, EM be the EI algebra over the set X and $E^{\#}X$ be the $E^{\#}I$ algebra over the set X. For any $\psi \in EM$, whether $L(\psi)$ is the maximal formal concept smaller than $(\alpha(\psi),\beta(\alpha(\psi)))$ and $U(\psi)$ is the minimal formal concept larger than $(\alpha(\psi),\beta(\alpha(\psi)))$? Here α,β are defined by (8.14) and (8.15), $L(.)$ and $U(.)$ are defined by (8.26) and (8.27), respectively.

Problem 8.5. Let X be a set and M be a set of attributes on X. Let (X,M,I) be a context, EM be the EI algebra over the set X and $E^{\#}X$ be the $E^{\#}I$ algebra over the set X. For any $\psi \in EM$ and $\theta \in E^{\#}X$, what are the relationships between the following pairs?

$$L(\psi) \ \ and \ \ L(\alpha(\psi)), \ \ U(\psi) \ \ and \ \ U(\alpha(\psi)),$$
$$L(\theta) \ \ and \ \ L(\beta(\theta)), \ \ U(\theta) \ \ and \ \ U(\beta(\theta)).$$

Here $L(.)$ is defined by (8.26) or (8.28), and $U(.)$ is defined by (8.27) or (8.29), respectively.

References

1. Belohlavek, R.: Lattices generated by binary fuzzy relations. Tatra Mountains Mathematical Publications 16, 11–19 (1999); special issue of fuzzy sets and Systems
2. Belohlavek, R.: Fuzzy Galois connections. Mathematical Logic Quarterly 45, 497–504 (1999)
3. Burusco, A., Gonzalez, R.F.: The Study of the L-Fuzzy Concept Lattice. Mathware and Soft Computing 1(3), 209–218 (1994)
4. Carpineto, C., Romano, G.: A lattice conceptual clustering system and its application to browsing retrieval. Mach. Learning 10, 95–122 (1996)
5. Deogun, J., Saquer, J.: Monotone concepts for formal concept analysis. Discrete Applied Mathematics 144, 70–78 (2004)
6. Diaz-Agudo, B., Gonzalez-Calero, P.A.: Formal concept analysis as a support technique for CBR. Knowledge-Based Systems 14, 163–171 (2001)
7. Ding, R., Liu, X., Chen, Y.: The Fuzzy Clustering Algorithm Based on AFS Topology. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS (LNAI), vol. 4223, pp. 89–98. Springer, Heidelberg (2006)
8. Deogun, J., Saquer, J.: Monotone concepts for formal concept analysis. Discrete Applied Mathematics 144, 70–78 (2004)
9. du Boucher-Ryana, P., Bridge, D.: Collaborative recommending using formal concept analysis. Knowledge-Based Systems 19, 309–315 (2006)
10. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer, Berlin (1999)
11. Graver, J.E., Watkins, M.E.: Combinatorics with Emphasis on the Theory of Graphs. Springer, New York (1977)

12. Hu, Q., Liu, J., Yu, D.: Mixed feature selection based on granulation and approximation. Knowledge-Based Systems 21, 294–304 (2008)
13. Godin, R., Mili, H., Mineau, G., Missaoui, R., Arfi, A., Chau, T.: Design of class hierarchies based on concept Galois lattices. TAPOS 4(2), 117–134 (1998)
14. Halmos, P.R.: Measure theory. Springer, New York (1974)
15. Harms, S.K., Deogun, J.S.: Sequential association rule mining with time lags. J. Intell. Inform. Systems 22(1), 7–22 (2004)
16. Huynh, V.N., Nakamori, Y.: Fuzzy Concept Formation Based on Context Model. Knowledge-Based Intelligent Information Eng. Systems and Allied Technologies, 687–691 (2001)
17. Liu, X.D., Pedrycz, W.: The development of fuzzy decision trees in the framework of axiomatic fuzzy set logic. Applied Soft Computing 7, 325–342 (2007)
18. Liu, X.D., Chai, T.Y., Wang, W.: Approaches to the Representations and Logic Operations for Fuzzy Concepts in the Framework of Axiomatic Fuzzy Set Theory I, II. Information Sciences 177, 1007–1026, 1027-1045 (2007)
19. Liu, X.D.: The Fuzzy Theory Based on AFS Algebras and AFS Structure. Journal of Mathematical Analysis and Applications 217, 459–478 (1998)
20. Liu, X.D.: The Topology on AFS Algebra and AFS Structure. Journal of Mathematical Analysis and Applications 217, 479–489 (1998)
21. Liu, X.D.: The Fuzzy Sets and Systems Based on AFS Structure, EI Algebra and EII algebra. Fuzzy Sets and Systems 95, 179–188 (1998)
22. Liu, X.D.: A new fuzzy model of pattern recognition and hitch diagnoses of complex systems. Fuzzy Sets and Systems 104, 289–297 (1999)
23. Liu, X.D., Pedrycz, W., Zhang, Q.L.: Axiomatics Fuzzy sets logic. In: IEEE International Conference on Fuzzy Systems, vol. 1, pp. 55–60 (2003)
24. Liu, X.D., Wang, W., Chai, T.Y.: The Fuzzy Clustering Analysis Based on AFS Theory. IEEE Transactions on Systems, Man and Cybernetics 35(5), 1013–1027 (2005)
25. Liu, X.D., Chai, T.Y., Wang, W.: AFS Fuzzy Logic Systems and Its Applicationsto Model and Control. International Journal of Information and Systems Sciences 2(3), 1–21 (2006)
26. Liu, X.D., Zhang, Q.L., Jun, Z., Zhu, K.J., Zhang, Q.L.: The Structures of EI Algebras Generated by Information Attributes. Int. J. Intelligent Systems Technologies and Applications 3(3/4), 341–355 (2007)
27. Liu, X.D., Zhang, Q.L.: The EI Algebra Representations of Fuzzy Concept. Fuzzy Systems and Mathematics 16(2), 27–35 (2002)
28. Liu, X.D., Zhu, K.J., Huang, H.Z.: The Representations of Fuzzy Concepts Based on the Fuzzy Matrix Theory and the AFS Theory. In: IEEE International Symposium on Intelligent Control, October 5-8, Texas, USA (2003)
29. Liu, X.D.: The Structure of Fuzzy Matrices. Journal of Fuzzy Mathematics 2, 311–325 (1994)
30. Liu, X.D.: A New Mathematical Axiomatic System of Fuzzy Sets and Systems. Journal of Fuzzy Mathematics 3, 559–560 (1995)
31. Liu, X.D.: Two Algebra Structures of AFS structure. Journal of Fuzzy Mathematics 3, 561–562 (1995)
32. Liu, X.D., Zhang, Q.L.: The Fuzzy Cognitive Maps Based on AFS Fuzzy Logic. Dynamics of Continuous, Discrete and Impulsive Systems 11(5-6), 787–796 (2004)
33. Liu, X., Liu, W.: Credit rating analysis with AFS fuzzy logic. In: Wang, L., Chen, K., S. Ong, Y. (eds.) ICNC 2005. LNCS, vol. 3612, pp. 1198–1204. Springer, Heidelberg (2005)

34. Mineau, G., Godin, R.: Automatic structuring of knowledge bases by conceptual clustering. IEEE Transactions on Knowledge and Data Engineering 7, 824–829 (1995)
35. Pawlak, Z.: Rough sets. International Journal of Computer and Information Science 11, 341–356 (1982)
36. Pawlak, Z., Skowron, A.: Rough sets: Some extensions. Information Sciences 177, 28–40 (2007)
37. Pawlak, Z., Skowron, A.: Rough sets and Boolean reasoning. Information Sciences 177, 41–73 (2007)
38. Quinlan, J.R.: Decision trees and decision making. IEEE Transactions on Systems, Man and Cybernetics Part C 20, 339–346 (1990)
39. Ren, Y., Song, M.L., Liu, X.D.: New approaches to the fuzzy clustering via AFS theory. International Journal of Information and Systems Science 3, 307–325 (2007)
40. Mineau, G., Godin, R.: Automatic structuring of knowledge bases by conceptual clustering. IEEE Transactions on Knowledge and Data Engineering 7(5), 824–829 (1995)
41. Missikoff, M., Scholl, M.: An algorithm for insertion into a lattice: application to type classification. In: Litwin, W., Schek, H.-J. (eds.) FODO 1989. LNCS, vol. 367, pp. 64–82. Springer, Heidelberg (1989)
42. Pollandt, S.: Fuzzy-Begriffe: Formale Begriffsanalyse Unscharfer Daten. Springer, Heidelberg (1996)
43. Stumme, G., Taouil, R., Bastide, Y., Pasquier, N., Lakhal, L.: Computing iceberg concept lattices with TITANIC. Data & Knowledge Engineering 42, 189–222 (2002)
44. Saquer, J., Deogun, J.: Approximating monotone concepts. Design and Application of Hybrid Intelligent System, 605–613 (2003)
45. Saquer, J., Deogun, J.: Concept approximations based on rough sets and similarity measure. International Journal of Appllied Mathematics and Compututer Science 11, 655–674 (2001)
46. Shao, M., Liu, M., Zhang, W.: Set approximations in fuzzy formal concept analysis. Fuzzy Sets and Systems 23, 2627–2640 (2007)
47. Stumme, G., Wille, R. (eds.): Begriffliche Wissensverarbeitung—-Methoden und Anwendungen. Springer, Heidelberg (2000)
48. Tho, Q.T., Hui, S.C., Fong, A.C.M., Cao, T.H.: Automatic Fuzzy Ontology Generation for Semantic Web. IEEE Transactions on Knowledge and Data Engineering 18(6), 842–856 (2006)
49. Vogt, F., Wille, R.: TOSCANA—-A graphical tool for analyzing and exploring data. In: Tamassia, R., Tollis, I(Y.) G. (eds.) GD 1994. LNCS, vol. 894, pp. 226–233. Springer, Heidelberg (1995)
50. Wille, R.: Restructuring lattice theory: an approach based on hierarchies of concepts. In: Ivan Rival, R. (ed.) Ordered Sets, pp. 445–470. Reidel, Dordecht (1982)
51. Wille, R.: Knowledge acquisition by methods of formal concept analysis. In: Diday, E. (ed.) Data Analysis, Learning Symbolic and Numeric Knowledge, pp. 365–380. Nova Science, NewYork (1989)
52. Wang, G.: Theory of topological molecular lattices. Fuzzy Sets and Systems 47, 351–376 (1992)
53. Wang, C., Wu, C., Chen, D.: A systematic study on attribute reduction with rough sets based on general binary relations. Information Sciences 178, 2237–2261 (2008)
54. Wang, L.D., Liu, X.D.: Concept analysis via rough set and AFS algebra. Information Sciences 178, 4125–4137 (2008)
55. Wang, L.D., Liu, X.D.: A new model of evaluating concept similarity. Knowledge-Based Systems 21, 842–846 (2008)

56. Wang, L.D., Liu, X.D.: AFS algebra properties of formal concepts. Information Sciences (submitted)
57. Wille, R.: The basic theorem of triadic concept analysis. Order 12, 149–158 (1995)
58. Zhang, L., Liu, X.: Concept lattice and AFS algebra. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS (LNAI), vol. 4223, pp. 290–299. Springer, Heidelberg (2006)
59. Zhang, Y.J., Liang, D.Q., Tong, S.C.: On AFS algebra part I. Information Sciences 167, 263–286 (2004)
60. Zhang, Y.J., Liang, D.Q., Tong, S.C.: On AFS algebra part II. Information Sciences 167, 287–303 (2004)
61. Yao, Y.: A comparative study of fuzzy sets and rough sets. Information Sciences 109, 227–242 (1998)
62. Yao, Y.: Concept lattices in rough set theory. In: Proceedings of, Annual Meeting Meeting of the North American Fuzzy Information Processing Society, pp. 796–801 (2004)
63. Zhang, L.S., Liu, X.D.: On $E^{\#} I^n$ algebra (manuscript)
64. Zhang, J., Wang, Y.: A rough margin based support vector machine. Information Sciences 178, 2204–2214 (2008)
65. Zhang, W., Ma, J., Fan, S.: Variable threshold concept lattices. Information Sciences 177, 4883–4892 (2007)
66. Zhao, Y., Halang, W.A., Wang, X.: Rough ontology mapping in E-Business integration. Studies in Computational Intelligence 37, 75–93 (2007)

Chapter 9
AFS Fuzzy Clustering Analysis

In this chapter, we apply the AFS theory to propose an elementary algorithm of fuzzy clustering. In the proposed approach, each cluster is interpreted by taking advantage of the semantics captured by the AFS logic. Within the framework of AFS theory, we develop new techniques of feature selection, concept categorization and characteristic description (i.e.,the characteristic description of an object or a group of objects using the fuzzy concepts) which are often encountered in tasks of machine learning and pattern recognition. The elementary fuzzy clustering algorithm is evolved to three more elaborate fuzzy clustering techniques by incorporating new techniques of feature selection, concept categorization and characteristic description. We show that they are simpler and produce more interpretable results when contrasted with some existing techniques. Several benchmark data and the evaluation data of 30 companies are considered to evaluate the effectiveness of the proposed AFS fuzzy clustering algorithms. We provide a detailed comparative analysis in which we compare the obtained results with those produced by some "conventional" methods such as FCM, *k*-means, and some newer algorithms including a two-level SOM-based clustering algorithm. The proposed algorithms can be applied to the data sets with mixed features such as sub-preference relations and even those including descriptions of human intuitive judgment. We show that the flexibility of the approach comes from the fact that the distance function and the class number need not be given beforehand. These two facets offers a far more higher flexible and contribute to a powerful framework for representing human knowledge and studying intelligent systems encountered in real world applications.

Clustering algorithms are mainly based on partitioning a set of objects into "natural" clusters. Numerous mathematical tools, investigated for clustering, have been considered to detect similarities between objects within a cluster. The two-valued clustering is described by characteristic functions. This function assigns each object to one and only one of the clusters with a degree of membership equal to one. However, the boundaries between the clusters might not be well-defined and this Boolean description may not fully reflect the reality. The fuzzy clustering, founded upon fuzzy set theory [81], is intended to deal with ill-defined boundaries between clusters. Membership degrees captured by membership functions indicates how much

X. Liu and W. Pedrycz: Axiomatic Fuzzy Set Theory and Its Applications, STUDFUZZ 244, pp. 351–421.
springerlink.com

the object is assigned to (belongs to) a certain cluster. This quantification can be advantageous in case of the the boundary region which may not be precisely defined.

Many fuzzy clustering algorithms have been developed, but the most widely used technique is the Fuzzy C-Means (FCM). Proposed by Dunn [12] and generalized by Bezdek [1], this family of algorithms is based on an iterative optimization of a certain objective function. The objective function produces a local minima or partial optimal points [2]. The algorithms of this family depend on initial guesses (cluster number, clusters centers,...). These prior arrangements are necessary but they do not guarantee that the method may reach the global minimum. In general, the objective function-based optimization is concerned with the following problem cf. [1, 2]: minimize $J_m(U,V) = \sum_i \sum_k u_{ik}^m d^2(x_k, V_i)$ with respect to $U = [u_{ik}] \in R^{c\times n}$, a fuzzy c-partition of n unlabeled data set $X = \{x_1,...,x_n\} \in R^{p\times n}$ and to V, a set of c fuzzy cluster centers $V = (V_1,...,V_c) \in R^{p\times c}$. The parameter $m > 1$ is referred to as a fuzziness index (fuzziness factor). $d(x_k,V_i)$ is a distance from x_k to the center (prototype) of ith cluster V_i. The performance of the FCM is significantly affected by the choice of distance $d(.,.)$. In general, the distance is expressed in some metric space [2, 29, 74], if data set X itself is a subset of this metric space. Another fuzzy clustering algorithm–the fuzzy k-nearest neighbor algorithm, k-NN algorithm [29], produces the membership degree for each sample j belonging to class i. More specifically, in [29] proposed was the following class assignment: $\mu_{ij} = 0.51 + 0.49(n_{ij}/k)$, when the class label of sample j is i; and $\mu_{ij} = 0.49(n_{ij}/k)$, otherwise, where n_{ij} is the number of the neighbors of sample j belonging to the ith class.

FCM and k-NN fuzzy clustering algorithms are efficient if for the data set $X \subset R^{p\times n}$, there exists a distance function, and the number of classes has been properly specified beforehand. For the FCM, we should notice that the objective function optimization problem is very difficult to solve for any ordinary data $X \nsubseteq R^{p\times n}$. For k-NN, it is too strict to solve real world problem that the class label of each sample has to be given in advance. The partition matrix $U = (u_{ij})_{c\times n}$ obtained by FCM and $U = (\mu_{ij})_{c\times n}$ obtained by k-NN do not show how the fuzzy concepts or attributes formed for each feature influence the clustering results. In contrast, humans can cluster the objects in ordinary data set $X \nsubseteq R^{p\times n}$ according to the fuzzy concepts or attributes on the features and give some linguistic descriptions to each class by the fuzzy concepts on some feature without using distance function.

In this chapter, we propose a new methodology of fuzzy clustering. Compared with the current state of development the art in the area, the fuzzy clustering to be discussed exhibits several essential advantages:

- The attributes of objects can involve various data types or sub-preference relations, even descriptors that are reflective of a human intuition.
- The distance function and the objective function are not required, and the cluster number or the class label need not to be given beforehand.
- Each class is described by a fuzzy set in EM, which is the AFS fuzzy logic compound of the simple attributes formed on some features with well-structured semantics which determines the degree of each pattern belongs to this class.

Given massive data available, knowledge acquisition and representation constitute a major bottleneck. There are various approaches aimed at alleviating this problem. The incorporation of fuzzy sets into the representation of fuzzy concepts makes it possible to combine the capabilities of uncertainty handling and approximate reasoning with comprehensibility of description of the phenomenon.

In many pattern recognition and decision making tasks, there is often a very limited prior information available about the data. Thus data preprocessing becomes an indispensable step. It is a genuine prerequisite in data mining and machine learning, which aims to turn data into business intelligence or knowledge. Feature selection is a preprocessing technique commonly used for high dimensional data. Feature selection focuses on how to select a subset of features that are to be effectively used to construct models describing data. The purpose of feature selection is to reduce dimensionality by removing irrelevant and redundant features, reducing the amount of data needed for learning, and enhance the comprehensibility of the constructed models. Feature selection has been widely studied in the context of supervised learning (see [8, 27, 34, 35] and references therein). However, feature selection has received comparatively little attention in unsupervised learning. In the theory of fuzzy sets we see a limited number of studies focused on this subject. One important reason is that it is not at all clear at all on how to assess the relevance of a subset of features without resorting to class labels. The problem becomes even more challenging when the number of clusters is unknown, since the optimal number of clusters and the optimal feature subset are interrelated. In this chapter, we propose a new method to deal with the fuzzy feature selection problem, which is expressed in terms of unsupervised learning within the framework of the AFS theory. Actually, the feature selection is carried out by making use of the idea of the fuzzy similarities occurring among features determined by the AFS fuzzy logic.

In general, the concepts exhibiting a significant level of correlation are often placed in the same category when carrying out data analysis. For instance, height and weight seem to be highly correlated, i.e., in general, the higher the person is, the heavier the person is. So in practice, height and weight are placed in the same category which describes human appearance. Usually, a group of the highly correlated concepts is always related to a particular characteristic of the objects in the clusters. The concepts height and weight are related to the appearance of human and they are salient concepts if we cluster the set of people according to their appearance. The aim of studying the concept categorization problem is to find different categories (i.e., the clusters of fuzzy concepts) so as to properly describe different characteristics of objects. In other words, a particular characteristic will be described by a corresponding category of concepts. Based on the AFS theory, an algorithm of clustering concepts into categories is proposed which could exhibit significant relevance to ideas of pattern recognition and decision-making.

Let us briefly analyze the human recognition process. For a sample, human always apply the selected collection of simple concepts to form complex fuzzy concepts which serve as the description of this sample. Other people can find the sample from all the samples according to the given description. Take clustering as an example: the samples with the same or similar descriptions form the same cluster. In this chapter, we study how to describe a sample in the framework of the AFS fuzzy logic, and argue that the fuzzy descriptions could be similar to the descriptions generated by humans.

In order to illustrate the feasibility, applicability and effectiveness of the feature selection, the concept categorization and the characteristic description, these techniques are applied to fuzzy clustering. Thus, based on the feature selection, the concept categorization and the characteristic description, some new fuzzy clustering algorithms in the framework of AFS theory are to be developed. Compared with other fuzzy clustering algorithms such as the conventional algorithms FCM [4] and k-means [73], SOM-based clustering algorithm [79], the proposed algorithms come with the main advantage except the aforementioned ones: they do not require training which are typically needed when dealing with other clustering techniques.

9.1 Elementary Fuzzy Clustering Method via AFS Fuzzy Logic

In this section, we apply the AFS theory to study the essence of fuzzy clustering. Clustering realized by humans is a procedure whose results are determined by the objectivity of the original data and the subjectivity of individual point of view, i.e., from the total attributes of original data and facts, an individual subjectively chooses some attributes he regards to be important within the setting of the clustering problem. The individual clusters (set of objects) follow the procedure shown in Figure 9.1. For each object, we always find a description (in general, a fuzzy description) of the object using some chosen attributes. Next given the description of the object, we can find similar objects in the entire collection of objects. Thus the similar objects will be viewed as a cluster. Second, one evaluates the similarity between the objects according to the descriptions of the objects. Third the clusters are formed by looking at the similarity degrees which are determined by the similarity between the objects. Finally, we select the most visible clustering result from all feasible clustering outcomes coming with the similarity degrees. The descriptions always involve some fuzzy or Boolean concepts.

In what follows, we study how to describe an object for the clustering in the framework of AFS fuzzy logic. The level of correlations between the two objects is determined by their fuzzy descriptions. We can state the problem more formally as follows: Let X be a set of objects and $\mathcal{F}$ be the set of all attributes or concepts (including attributes which are independent or unrelated to the clustering problem) involved the objects in X. For each attribute $m \in \mathcal{F}$, m is a Boolean or fuzzy attribute on X. $\Lambda \subseteq \mathcal{F}$, Λ is the set of attributes (features), which are some relatively important attributes subjectively chosen from an individual point of view. For instance, suppose that each $\alpha \in \Lambda$, α is simply a Boolean set, i.e., $\alpha \subseteq X$. For $x, y \in X, x, y$ belong to the same cluster if and only if there does not exist $\alpha \in \Lambda$ such that $x \in \alpha, y \notin \alpha$ or $x \notin \alpha, y \in \alpha$. In other words, x, y cannot be distinguished by any attributes in Λ. It is obvious that for a different choice of Λ, different clustering results may be produced. Therefore human clustering comes as a procedure whose results are determined by the objectivity of the original data and the subjectivity of an individual opinions, i.e., Λ, which is a set of the fuzzy attributes subjectively

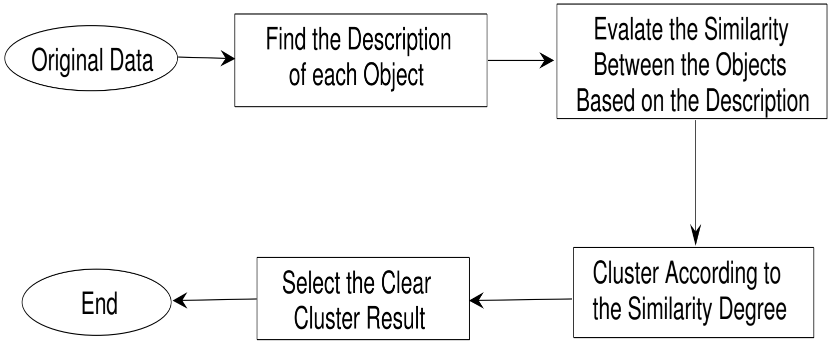


Fig. 9.1 Clustering procedure realized by human beings

chosen from F by an individual. Using the following simple example, we first explain the idea in the situation when all attributes in Λ are Boolean. Then we expand the idea to situations in which we are faced with fuzzy attributes.

Example 9.1. Let $X = \{x_1, x_2, ..., x_{10}\}$ be a set of 10 persons, $F = \{male, engineer, lawyer, female, male, weight, age, salary$ high, ..., $etc\}$. For each attribute $m \in F$, m is a fuzzy or Boolean attribute which is objectively dependent on the original data and facts. We study the clustering under $\Lambda = \{male, engineer, lawyer, female, no\text{-}engineer, no\text{-}lawyer\} \subseteq F$ and suppose we are provided with Table 9.1 for the attributes in Λ. Where A_1, A_2, A_3 are attributes such as male, engineer and lawyer, respectively. Let $\overline{\Lambda}$ be the set generated by sets A_1, A_2, A_3, A_1', A_2', A_3', using logic operators $\cap$. $\delta_{x_i} \in \overline{\Lambda}$, δ_{x_i} is the smallest set which contains x_i. Given Table 9.1, one has:

$$
\begin{aligned}
\delta_{x_1} &= A_1 \cap A_3 \cap A_2';\\
\delta_{x2} &= \delta_{x8} = \delta_{x10} = A_1 \cap A_2 \cap A_3';\\
\delta_{x3} &= \delta_{x5} = A_2 \cap A_1' \cap A_3';\\
\delta_{x4} &= A_1 \cap A_2' \cap A_3';\\
\delta_{x6} &= \delta_{x7} = A_3 \cap A_2' \cap A_1'.
\end{aligned}
$$

In this way we obtain a mapping $\Psi : X \to \overline{\Lambda}$, for any $x \in X$, $\Psi(x) = \delta_x$, which is a description of x using attributes in Λ such that x can be distinguished among other elements in X to the maximum extent. Since Ψ is a mapping, hence Ψ determines a

Table 9.1 Boolean description of attributes

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
A_1	1	1	0	1	0	0	0	1	0	1
A_2	0	1	1	0	1	0	0	1	0	1
A_3	1	0	0	0	0	1	1	0	0	0

classification in X (i.e., $x,y \in X$, x,y in the same cluster if and only if $\Psi(x) = \Psi(y)$ or $\delta_x = \delta_y$). Since δ_x, δ_y form the smallest set containing x,y, hence $\delta_x = \delta_y \Leftrightarrow x,y \in \delta_x \cap \delta_y$. Ten persons are then clustered into six classes (the finest classification obtained when using attributes $A_1, A_2, A_3, A_1', A_2', A_3'$):

Class1: $\{x_1\} = A_1 \cap A_2' \cap A_3$, describing as "*male, not-engineer, lawyer*".
Class2: $\{x_2, x_8, x_{10}\} = A_1 \cap A_2 \cap A_3'$, describing as "*male, not-lawyer, engineer*".
Class3: $\{x_3, x_5\} = A_1' \cap A_2 \cap A_3'$, describing as "*female, not-lawyer, engineer*".
Class4: $\{x_4\} = A_1 \cap A_2' \cap A_3'$, describing as "*not-lawyer, not-engineer, male*".
Class5: $\{x_6, x_7\} = A_1' \cap A_2' \cap A_3$, describing as "*female, lawyer, not-engineer*".
Class6: $\{x_9\} = A_1' \cap A_2' \cap A_3'$,describing as "*female, not-lawyer, not-engineer*".

Next, we summarize the clustering algorithm by introducing the Boolean attributes $A_1, A_2, \ldots, A_k$ on X as follows.

- **Step 1:** Generate the Boolean algebra $\overline{\Lambda}$ by Boolean attributes in Λ, using logic operator $\cap$.
- **Step 2:** For each $x \in S$, find δ_x which is the smallest set $\delta_x \in \overline{\Lambda}$ such that $x \in \delta_x$.
- **Step 3:** For $x,y \in S$, x,y are in the same cluster if and only if $x,y \in \delta_x \cap \delta_y$.

Now, we expand the above algorithm to the case of fuzzy attributes by using the AFS fuzzy logic. This will give rise to the *elementary fuzzy clustering method via AFS fuzzy logic*. Let X be the universe of discourse, M be a set of simple attributes on X, (M, τ, X) be an AFS structure of the original data and facts. Let $\{\mu_\xi(x) \mid \xi \in EM\}$ be a set of coherence membership functions of the AFS fuzzy logic system $(EM, \vee, \wedge, ')$ and the AFS structure (M, τ, X). (refer to Definition 4.7). Assume that $\Lambda \subseteq EM$, Λ is a family of fuzzy sets which are selected to cluster the objects in X. The *elementary fuzzy clustering method via AFS fuzzy logic* can be outlined as follows.

The elementary fuzzy clustering method realized via AFS fuzzy logic

- **Step 1:** Find fuzzy set $\vartheta = \bigvee_{b \in \Lambda} b$, $x \in X$, $\mu_{\bigvee_{b \in \Lambda} b}(x)$ is the highest degree of x belonging to any cluster, due to ϑ being the maximum element in $(\Lambda)_{EI}$. In order to produce a well-defined clustering result, each x should belong to ϑ to the highest extent. Proposition 9.1 outlines the properties of the fuzzy set ϑ.
- **Step 2:** Find the fuzzy description of each object: for each $x \in X$, find the *fuzzy description* ζ_x of x, which is δ_x for the Boolean case. For fuzzy set $\zeta_x \in (\Lambda)_{EI}$, where $(\Lambda)_{EI}$ is the sub EI algebra generated by Λ, not only is $\mu_{\zeta_x}(x)$ approaching $\mu_{\bigvee_{b \in \Lambda} b}(x)$, but also $\mu_{\zeta_x}(y)$ is as small as possible for $y \in X$, $y \neq x$. In other words, x can be distinguished by ζ_x from other objects in X at the highest extent.
- **Step 3:** Evaluating the similarity between objects based on the fuzzy descriptions: apply ζ_x the fuzzy description of each $x \in X$ to establish the *fuzzy matrix* $M = (m_{ij})$ on $X = \{x_1, x_2, \ldots, x_n\}$, where m_{ij} the *similarity degree between* x_i and x_j which is defined as follows: for any $x_i, x_j \in X$,

$$m_{ij} = \min\{\mu_{\zeta_{x_i} \wedge \zeta_{x_j}}(x_i), \mu_{\zeta_{x_i} \wedge \zeta_{x_j}}(x_j)\}. \tag{9.1}$$

Theorem 9.2 demonstrates that there exists an integer r such that $(M^r)^2 = M^r$, i.e., fuzzy matrix $Q = M^r$ can yield a partition tree with equivalence classes.

- **Step 4:** Cluster according to the determined similarity degrees: let $Q = M^r = (q_{ij})$ and the Boolean matrix $Q_\alpha = (q_{ij}^\alpha)$, where $q_{ij}^\alpha = 1 \Leftrightarrow q_{ij} \geq \alpha$, the threshold $\alpha \in [0,1]$. For $\alpha \in [0,1]$, $x_i, x_j \in X, x_i, x_j$ are in the same cluster for given threshold α if and only if $q_{ij}^\alpha = 1$. For some $x_i \in X$, if $q_{ii}^\alpha = 0$, then the clustering label of x_i cannot be determined for fuzzy attributes in Λ under the threshold α.
- **Step 5:** Select the well-delineated clustering results: for each cluster $C \subseteq X$ under the threshold α, the *fuzzy description of* C, ζ_C is defined as follows.

$$\zeta_C = \bigvee_{x \in C} \zeta_x, \tag{9.2}$$

the fuzzy description ζ_C of class C whose membership degree $\mu_{\zeta_C}(x)$ is not only the most approachable $\mu_{\bigvee_{b \in \Lambda} b}(x)$, for each $x \in C$, but also $\mu_{\zeta_C}(y)$ is as small as possible for $y \in X$, $y \notin C$. In other words, the objects in cluster C can be distinguished from other objects in X to the highest possible extent. Theorem 9.1 shows how to obtain ζ_x for each $x \in X$ while Proposition 9.3 shows that the fuzzy descriptions of two different clusters do not have common molecular elements of the lattice EM (i.e., fuzzy point). The *fuzzy description of the boundary among the clusters* $C_1, C_2, ..., C_l$ is a fuzzy set $\zeta_{bou} \in EM$,

$$\zeta_{bou} = \bigvee_{1 \leq i,j \leq l, i \neq j} (\zeta_{C_i} \wedge \zeta_{C_j}), \tag{9.3}$$

where $\zeta_{C_i}, i = 1, 2, ..., l$ is the fuzzy description for the ith cluster C_i. The clarity of the fuzzy clustering for some the threshold α can be evaluated by I_α a *fuzzy cluster validity index* defined as follows. For any threshold $\alpha \in [0,1]$,

$$I_\alpha = \frac{\sum_{x \in \bigcup_{1 \leq i \leq l} C_i} \mu_{\zeta_{bou}}(x)}{\sum_{x \in \bigcup_{1 \leq i \leq l} C_i} \mu_{\zeta_{Tatal}}(x)}, \tag{9.4}$$

where $\zeta_{Total} = \bigvee_{1 \leq i \leq l} \zeta_{C_i} \in EM$, $l \geq 2$. The ith cluster is described by a fuzzy set $\zeta_{C_i} \in EM$ which determines the degree each object belong to the ith cluster. Fuzzy set ζ_{bou} which describes the boundary between the clusters implies the maximal degree of each object belonging to two different clusters. It is clear that the lower I_α is, the clearer and the better the clustering result under threshold α is. Thus the best clustering outcome can be selected from all clustering results under threshold $\alpha \in [0,1]$.

This fuzzy clustering algorithm can apply to the data sets with mixed features taking on values for integers, real numbers, Boolean values, and sub-preference relations. Likewise the distance function and the number of cluster number are not required to be supplied in advance. Since each cluster is described by ξ_C a fuzzy set in EM with well-formed semantics, hence the clustering results are more interpretable than those produced by some conventional fuzzy clustering.

In what follows, we give some proofs and include pertinent analysis to ensure that the proposed clustering algorithm is feasible. Let X be a finite set and M be a set of simple concepts on X. Assume $\Lambda \subseteq EM$. Then $(\Lambda)_{EI}$ the sub algebra of EM generated by the fuzzy concepts in Λ is shown as follows.

$$(\Lambda)_{EI}=\left\{\bigvee_{i\in I}(\bigwedge_{j\in J_i} a_{ij}) \mid a_{ij}\in\Lambda, i\in I, j\in J_i, I \text{ and } J_i \text{ are any indexing sets}\right\}. \tag{9.5}$$

Its proof remains as an exercise. It is the smallest sub EI algebra of EM which contains Λ.

Proposition 9.1. *Let X be a universe of discourse and M be a finite set of simple concepts. Let $\{\mu_\xi(x) \mid \xi\in EM\}$ be a set of coherence membership functions of the AFS fuzzy logic system $(EM,\vee,\wedge,')$ and the AFS structure (M,τ,X). (seeing Definition 4.7). Let $\Lambda\subseteq EM$. Then for any $\beta\in(\Lambda)_{EI}$, for any $x\in X$, $\mu_\beta(x)\le \mu_{\bigvee_{b\in\Lambda} b}(x)$.*

Proposition 9.1 implies that for each $x\in X$, the degree of x belonging to fuzzy set $\bigvee_{b\in\Lambda} b$ is the largest of other fuzzy sets in $(\Lambda)_{EI}$. But $\bigvee_{b\in\Lambda} b$ is not the fuzzy description of x, because $\bigvee_{b\in\Lambda} b$ is the maximum fuzzy set in lattice $(\Lambda)_{EI}$ and for each $y\in X$, $y\ne x$ the degree of y belonging to fuzzy set $\bigvee_{b\in\Lambda} b$ is also the largest of other fuzzy sets in $(\Lambda)_{EI}$. Therefore for a given $x\in X$, we should find the fuzzy set ζ_x in $(\Lambda)_{EI}$ such that not only is $\mu_{\zeta_x}(x)$ approaches $\mu_{\bigvee_{b\in\Lambda} b}(x)$, but also $\mu_{\zeta_x}(y)$ is as small as possible for each $y\in X$ and $y\ne x$. In what follows, we find the fuzzy set ζ_x in $(\Lambda)_{EI}$ for each given x. For $\varepsilon\ge 0$ (in general, ε is very small), we define

$$B_x^\varepsilon = \left\{A_k \mid \mu_{A_k}(x)\ge \mu_{\bigvee_{b\in\Lambda} b}(x)-\varepsilon,\ k\in I,\ a=\sum_{i\in I}A_i\in\Lambda\right\}, \tag{9.6}$$

$$\overline{B}_x^\varepsilon = \left\{\bigwedge_{\beta\in H}\beta \mid H\subseteq B_x^\varepsilon,\ \ \mu_{\bigwedge_{\beta\in H}\beta}(x)\ge \mu_{\bigvee_{b\in\Lambda} b}(x)-\varepsilon\right\}. \tag{9.7}$$

$$\Lambda_x^\varepsilon = \left\{\gamma \mid \gamma \ \textit{is a minimal element in}\ \overline{B}_x^\varepsilon\right\}. \tag{9.8}$$

Since

$$\mu_{\bigvee_{\beta\in\Lambda}\beta}(x)=\sup_{\beta\in\Lambda}\mu_\beta(x)=\sup_{\beta\in\Lambda}\sup_{k\in I_\beta}\mu_{A_k^\beta}(x),$$

where $\beta=\sum_{k\in I_\beta}A_k^\beta\in\Lambda$, hence for any $\varepsilon>0$, $B_x^\varepsilon\ne\varnothing$, $\Lambda_x^\varepsilon\ne\varnothing$ and $\Lambda_x^{\varepsilon_1}\supseteq\Lambda_x^{\varepsilon_2}$ if $\varepsilon_1>\varepsilon_2$. Next, we analyze the composition of the set B_x^ε. For each fuzzy set $\sum_{i\in I}A_i$, each $A_i, i\in I$, is a molecular element which play a role similar to a "*point*" of the fuzzy set. B_x^ε is the set of molecular elements (in other words, "fuzzy point") of the fuzzy sets in $\Lambda, \forall A\in B_x^\varepsilon$, $\mu_A(x)\ge\mu_{\bigvee_{b\in\Lambda} b}(x)-\varepsilon$. In some cases, there may be $A,B\in B_x^\varepsilon, A\ne B$, such that $\mu_{A\wedge B}(x)\ge\mu_{\bigvee_{b\in\Lambda} b}(x)-\varepsilon$. We know that the molecule $A\wedge B$ produces a more accurate description of x than A or B. Λ_x^ε is the set of the minimal molecules $\bigwedge_{\beta\in H}\beta, H\subseteq B_x^\varepsilon$, such that $\mu_{\bigwedge_{\beta\in H}\beta}(x)\ge\mu_{\bigvee_{b\in\Lambda} b}(x)-\varepsilon$.

Theorem 9.1. *Let X be a universe of discourse and M be a finite set of simple concepts, (M,τ,X) be an AFS structure. Let $\{\mu_\xi(x) \mid \xi \in EM\}$ be a set of coherence membership functions of the AFS fuzzy logic system $(EM,\vee,\wedge,')$ and the AFS structure (M,τ,X). (seeing Definition 4.7). Let $\Lambda \subseteq EM$. For a given $x \in X$ and a given $\varepsilon > 0$, $\alpha \in \Lambda_x^\varepsilon$, let $\vartheta_\alpha^x = \{\beta \in (\Lambda)_{EI} \mid \beta \geq \alpha\}$. Where Λ_x^ε is defined by (9.8). Then the following observations hold:*

(1) ϑ_α^x is a sub-EI algebra of $(\Lambda)_{EI}$.
(2) $\mu_{\bigwedge_{b\in\vartheta_\alpha^x} b}(x) \geq \mu_\alpha(x) \geq \mu_{\bigvee_{b\in\Lambda} b}(x) - \varepsilon$.
(3) For $\eta \in (\Lambda)_{EI}$, if $\mu_\eta(x) > \mu_{\bigvee_{b\in\Lambda} b}(x) - \varepsilon$, then $\exists \alpha \in \Lambda_x^\varepsilon$, for any $y \in X$, $y \neq x$,

$$\mu_\eta(y) \geq \mu_{\bigwedge_{b\in\vartheta_\alpha^x} b}(y) \geq \mu_\alpha(y).$$

(4) $\overline{\zeta}_x = \bigvee_{\alpha\in\Lambda_x^\varepsilon}(\bigwedge_{b\in\vartheta_\alpha^x} b) \geq \zeta_x = \bigvee_{\alpha\in\Lambda_x^\varepsilon} \alpha$.

Proof. (1) and (2) can be proved directly by taking into account the corresponding definitions.

(3) For given $\varepsilon > 0, x \in X$, suppose $\eta = \bigvee_{i\in I}\bigwedge_{j\in J_i} a_{ij} \in (\Lambda)_{EI}$, $a_{ij} = \sum_{k\in K_j} A_k^{ij} \in \Lambda$, $j \in J_i$, $i \in I$, K_j, and $\mu_\eta(x) > \mu_{\bigvee_{b\in\Lambda} b}(x) - \varepsilon$. Since $(\Lambda)_{EI}$ is a molecular lattice, hence

$$\begin{aligned}\eta &= \bigvee_{i\in I}\bigwedge_{j\in J_i} a_{ij} = \bigvee_{i\in I}\bigwedge_{j\in J_i}\bigvee_{k\in K_j} A_k^{ij} \\ &= \bigvee_{i\in I}\bigvee_{f\in\prod_{j\in J_i} K_j}(\bigwedge_{j\in J_i} A_{f(j)}^{ij}).\end{aligned}$$

Since

$$\mu_\eta(x) \geq \mu_{\bigvee_{b\in\Lambda} b}(x) - \varepsilon$$

hence $\exists l \in I$, $\exists g \in \prod_{j\in J_i} K_j$ such that

$$\mu_{\bigwedge_{j\in J_l} A_{g(j)}^{ij}}(x) \geq \mu_{\bigvee_{b\in\Lambda} b}(x) - \varepsilon.$$

Therefore for any $j \in J_l$,

$$\mu_{A_{g(j)}^{ij}}(x) \geq \mu_{\bigwedge_{j\in J_l} A_{g(j)}^{ij}}(x) \geq \mu_{\bigvee_{b\in\Lambda} b}(x) - \varepsilon.$$

This implies that$\forall j \in J_l, A_{g(j)}^{ij} \in B_x^\varepsilon$ and $\exists \alpha \in \Lambda_x^\varepsilon$ such that $\bigwedge_{j\in J_l} A_{g(j)}^{ij} \geq \alpha$ and

$$\eta \geq \bigwedge_{j\in J_l} A_{g(j)}^{ij} \geq \alpha.$$

Since ϑ_α^x is an upper set of $(\Lambda)_{EI}$, hence $\eta \in \vartheta_\alpha^x$ and $\mu_\eta(y) \geq \mu_{\bigwedge_{b\in\vartheta_\alpha^x} b}(y) \geq \mu_\alpha(y)$ for any $y \in X$, $y \neq x$.

(4) They can be proved by (1). □

Remark 9.1. Since ϑ_α^x is a sub-*EI* algebra of $(\Lambda)_{EI}$ for each $\alpha \in \Lambda_x^\varepsilon$, hence $\overline{\zeta}_x = \bigvee_{\alpha\in\Lambda_x^\varepsilon}(\bigwedge_{b\in\vartheta_\alpha^x}) \in (\Lambda)_{EI}$. If $\overline{\zeta}_x$ and ζ_x are applied to describe x, then $\overline{\zeta}_x$ is a fuzzy description of x by the fuzzy sets in $(\Lambda)_{EI}$ and ζ_x is a fuzzy description of x by the molecular elements of the fuzzy sets in $(\Lambda)_{EI}$. $\overline{\zeta}_x$ is a rougher description of x than ζ_x. But we should notice that in some case, $\zeta_x \notin (\Lambda)_{EI}$. By 3 of Theorem 9.1, one knows that for $\alpha \in \Lambda_x^\varepsilon$, fuzzy set $\bigwedge_{b\in\vartheta_\alpha^x} b$ ensures that any $y \neq x, y$ belongs to $\bigwedge_{b\in\vartheta_\alpha^x} b$ at low degree while x belongs $\bigwedge_{b\in\vartheta_\alpha^x} b$ at high degree. It could be easily proved that $\overline{\zeta}_x = \zeta_x$, if any selected attribute $\gamma \in \Lambda$, γ is a molecular element, i.e., $\gamma = A, A \subseteq M$.

Therefore both $\overline{\zeta}_x$ and ζ_x shown below can serve as the fuzzy description of $x \in X$.

$$\overline{\zeta}_x = \bigvee_{\alpha\in\Lambda_x^\varepsilon}(\bigwedge_{b\in\vartheta_\alpha^x} b), \tag{9.9}$$

$$\zeta_x = \bigvee_{\alpha\in\Lambda_x^\varepsilon} \alpha. \tag{9.10}$$

In this section, we employ ζ_x as the fuzzy description of x. $\{\zeta_x \mid x \in X\}$ is called a *fuzzy description of X under* Λ and ε. Now we get the fuzzy description ζ_x for each $x \in X$ (step 2). Next, we study step 3 and step 4 of the elementary fuzzy clustering method via the AFS fuzzy logic.

Definition 9.1. Let M be a set. Let $A = (a_{ij})_{m\times k}, B = (b_{ij})_{l\times n}$ be the matrices over the *EI* algebra *EM* (called *EI matrices*), where $a_{ij}, b_{ij}, c \in EM$. Then the matrix operations are defined as follows

(1) $A + B = (a_{ij} \vee b_{ij})$, if $m = l, k = n$.
(2) $AB = (\bigvee_{u=1}^k (a_{iu} \wedge b_{uj}))$, if $k = l$.
(3) $cA = Ac = (c \wedge a_{ij})$.

Proposition 9.2. *Let M be a set and EM be the EI algebra over M. Let A,B,C be any EI matrices with appropriate dimensions. Then the following assertions hold.*

(1) $A(B+C) = AB + AC$;
(2) $A(BC) = (AB)C$.

Let $X = \{x_1, x_2,..., x_n\}$ and M be a set of simple concepts on X. Let (M, τ, X) be an AFS structure and $\{\mu_\xi(x) \mid \xi \in EM\}$ be a set of coherence membership functions of the AFS fuzzy logic system $(EM, \vee, \wedge, ')$ and the AFS structure (M, τ, X). Let $\{\zeta_x \mid x \in X\}$ be the fuzzy description of X under Λ and ε. The *EI* matrix $B = (\zeta_{x_i} \wedge \zeta_{x_j})_{n\times n}$ is the *EI algebra relation matrix* which determines a fuzzy relation matrix $M = (m_{ij})$ on X,, i.e., the degree of x_i, x_j satisfying the fuzzy relation is

$$m_{ij} = \min\{\mu_{\zeta_{x_i}\wedge\zeta_{x_j}}(x_i), \mu_{\zeta_{x_i}\wedge\zeta_{x_j}}(x_j)\}.$$

Theorem 9.2. *Let* $X = \{x_1, x_2,..., x_n\}$ *and M be a set of simple concepts on X. Let* (M, τ, X) *be an AFS structure and* $\{\mu_\xi(x) \mid \xi \in EM\}$ *be a set of coherence membership functions of the AFS fuzzy logic system* $(EM, \vee, \wedge, ')$ *and the AFS structure*

(M,τ,X). Let $\{\zeta_x \mid x\in X\}$ be the fuzzy description of X under Λ and ε. Let $B=(\zeta_{x_i}\wedge \zeta_{x_j})_{n\times n}$ be the EI algebra relation matrix and $M=(m_{ij})_{n\times n}$ be the fuzzy relation matrix, where $m_{ij}=\min\{\mu_{\zeta_{x_i}\wedge\zeta_{x_j}}(x_i),\ \mu_{\zeta_{x_i}\wedge\zeta_{x_j}}(x_j)\}$. Then the following assertions hold.

(1) $B^2=B$;
(2) $M^2\geq M$;
(3) There exists an integer r such that $(M^r)^2=M^r$.

Here for fuzzy relation matrices $A=(a_{ij})_{n\times n}, B=(b_{ij})_{n\times n}$, $AB=(c_{ij})_{n\times n}$, $c_{ij}=\max_{1\leq k\leq n}\{\min\{a_{ik},b_{kj}\}\}$.

Proof. (1) By Definition 9.1, we have

$$B=(\zeta_{x_i}\wedge\zeta_{x_j})_{n\times n}=\begin{pmatrix}\zeta_{x_1}\\ \zeta_{x_2}\\ \vdots\\ \zeta_{x_n}\end{pmatrix}\left(\,\zeta_{x_1}\ \ \zeta_{x_2}\ \cdots\ \zeta_{x_n}\,\right),$$

hence

$$\begin{aligned}B^2&=\begin{pmatrix}\zeta_{x_1}\\ \vdots\\ \zeta_{x_n}\end{pmatrix}\left(\bigvee_{1\leq i\leq n}(\zeta_{x_i}\wedge\zeta_{x_i})\right)_{1\times 1}\left(\,\zeta_{x_1}\ \cdots\ \zeta_{x_n}\,\right)\\ &=\begin{pmatrix}\zeta_{x_1}\\ \vdots\\ \zeta_{x_n}\end{pmatrix}\left(\bigvee_{1\leq k\leq n}\zeta_{x_k}\right)_{1\times 1}\left(\,\zeta_{x_1}\ \cdots\ \zeta_{x_n}\,\right)\\ &=\left((\bigvee_{1\leq k\leq n}\zeta_{x_k})\wedge(\zeta_{x_i}\wedge\zeta_{x_j})\right)_{n\times n}\\ &=\left(\zeta_{x_i}\wedge\zeta_{x_j}\right)_{n\times n}=B.\end{aligned}$$

(2) Since the EI matrix $B=(\zeta_{x_i}\wedge\zeta_{x_j})_{n\times n}=(b_{ij})_{n\times n}$, $b_{ii}=\zeta_{x_i}\geq\zeta_{x_i}\wedge\zeta_{x_j}=b_{ij}$ in EM, hence

$$\begin{aligned}M^2&=\left(\bigvee_{1\leq k\leq n}(\mu_{\zeta_{x_i}\wedge\zeta_{x_k}}(x_i)\wedge\mu_{\zeta_{x_i}\wedge\zeta_{x_k}}(x_k))\wedge(\mu_{\zeta_{x_k}\wedge\zeta_{x_j}}(x_k)\wedge\mu_{\zeta_{x_k}\wedge\zeta_{x_j}}(x_j)))\right)_{n\times n}\\ &\geq\left(\mu_{\zeta_{x_i}\wedge\zeta_{x_i}}(x_i)\wedge\mu_{\zeta_{x_i}\wedge\zeta_{x_i}}(x_i)\wedge\mu_{\zeta_{x_i}\wedge\zeta_{x_j}}(x_i)\wedge\mu_{\zeta_{x_i}\wedge\zeta_{x_j}}(x_j)\right)_{n\times n}\\ &=\left(\min\{\mu_{\zeta_{x_i}\wedge\zeta_{x_j}}(x_i),\ \ \mu_{\zeta_{x_i}\wedge\zeta_{x_j}}(x_j)\}\right)_{n\times n}\\ &=M.\end{aligned}$$

(3) Since $m_{ij} = \min\{\mu_{\zeta_{x_i}\wedge\zeta_{x_j}}(x_i), \mu_{\zeta_{x_i}\wedge\zeta_{x_j}}(x_j)\} = m_{ji}$, $1 \le i, j \le n$, hence $M^T = M$. Because the fuzzy matrix M has a finite numbers of different elements and $M^2 \ge M$, there exists an integer r such that $M^r = M^{r+1} = ... = M^{2r}$. □

We should notice that Theorem 9.2 ensures that EI algebra relation matrix B and fuzzy relation matrix $Q = M^r$ can yield a partition tree with equivalence classes. This implies that for any $\alpha \in [0,1]$, $Q_\alpha = (q_{ij}^\alpha)$ is an equivalence relation (Boolean matrix) and it can yield a partition on X. Next, we study the step 5.

Proposition 9.3. *Let $C_1, C_2, ..., C_l$ be the clusters determined by equivalence relation Boolean matrix $Q_\alpha = (q_{ij}^\alpha)$ obtained in step 4 for the threshold $\alpha \in [0,1]$. For each $i = 1, 2, ..., l$, let*

$$\Lambda_{C_i} = \bigcup_{x \in C_i} \{A \in \Lambda_x^\varepsilon \mid \mu_A(x) \ge \alpha\}.$$

Where Λ_x^ε is defined by (9.8). Then $\Lambda_{C_i} \cap \Lambda_{C_j} = \varnothing, i \ne j$, for any $i, j = 1, 2, ..., l$.

Proof. Suppose that for some $i, j = 1, 2, ..., l, i \ne j, A \in \Lambda_{C_i} \cap \Lambda_{C_j} \ne \varnothing$. By the definition of Λ_x^ε given in (9.8), we know there exist $x \in C_i$ and $y \in C_j$ such that $A \in \Lambda_x^\varepsilon, \mu_A(x) \ge \alpha$ and $A \in \Lambda_y^\varepsilon, \mu_A(y) \ge \alpha$. This implies that

$$\zeta_x = \bigvee_{B \in \Lambda_x^\varepsilon} B \ge A,$$

$$\zeta_y = \bigvee_{B \in \Lambda_y^\varepsilon} B \ge A.$$

Therefore $\zeta_x \wedge \zeta_y \ge A$ and the degree of relationship between x and y comes as

$$\min\{\mu_{\zeta_x\wedge\zeta_y}(x), \mu_{\zeta_x\wedge\zeta_y}(y)\} \ge \min\{\mu_A(x), \mu_A(y)\} \ge \alpha.$$

By $Q = (q_{ij}) \ge M = (m_{ij})$, we know that $q_{xy}^\alpha = 1$ and x and y are in the same class. It contradicts that $x \in C_i$ and $y \in C_j, i \ne j$. □

Example 9.2. Let $X = \{x_1, x_2, ..., x_{10}\}$ be a set of 10 persons. $M = \{m_1, m_2, ..., m_{10}\}$, where m_1 : "*old*", m_2 : "*height high*", m_3 : "*weigh*, m_4 : "*salary high*, m_5 : "*larger fortune*", m_6 : "*male*", m_7 : "*female*, m_8 : "*black hair* ", m_9 : "*white hair* ", m_{10} : "*yellow hair*". About the universe of discourse X and the attribute set M, the original data and facts are shown as Table 9.2 and sub-preference relations expressed by the chains. $x = y$ in the chain means the degrees of x and y belonging to the attribute are equal, instead of x and y being the same element in X.

$$m_8 : x_7 > x_{10} > x_4 = x_8 > x_2 = x_9 > x_5 > x_6 = x_3 = x_1;$$
$$m_9 : x_6 = x_3 = x_1 > x_5 > x_2 = x_9 > x_4 = x_8 > x_{10} > x_7;$$
$$m_{10} : x_2 = x_9 > x_4 = x_8 = x_5 > x_{10} > x_6 = x_3 = x_1 = x_7.$$

Table 9.2 Description of Attributes

	m_1	m_2	m_3	m_4	m_5	m_6	m_7
x_1	20	1.9	90	1	0	1	0
x_2	13	1.2	32	0	0	0	1
x_3	50	1.7	67	140	34	0	1
x_4	80	1.8	73	20	80	1	0
x_5	34	1.4	54	15	2	1	0
x_6	37	1.6	80	80	28	0	1
x_7	45	1.7	78	268	90	1	0
x_8	70	1.65	70	30	45	1	0
x_9	60	1.82	83	25	98	0	1
x_{10}	3	1.1	21	0	0	0	1

First, we determine the weight function $\rho_m(x)$ for each simple attribute $m \in M$ according to available data and facts: $\rho_{m_i}(x)$, for $i = 1,2,...,7$, is the value of x_j for attribute m_i in Table 9.2, for example, $\rho_{m_2}(x_1) = 1.9$, $\rho_{m_4}(x_2) = 0$, $\rho_{m_6}(x_2) = 0$, $\rho_{m_7}(x_2) = 1$; $\rho_{m_i}(x_j) = 1$, for $i = 8,9,10$, if x_j belongs to simple concept m_i, otherwise $\rho_{m_i}(x_j) = 0$, for example, $\rho_{m_8}(x_7) = 1, \rho_{m_9}(x_7) = 0$, according to the order relations m_8, m_9 given above. For each $m \in M$, let

$$\rho_{m'}(x) = \max_{y \in X}\{\rho_m(y)\} - \rho_m(x), \quad x \in X,$$

where m' is the negation of the simple concept m. By Definition 4.8, we can verify that each ρ_m is the weight function of concept m. τ is defined according to Table 9.2 and the meaning of the simple concepts in M quantified by formula (4.26). Thus the set of coherence membership functions $\{\mu_\xi(x) \mid \xi \in EM\}$ can be obtained by formula (5.18) in Proposition 5.6.

In what follows, we apply the elementary fuzzy clustering method via AFS fuzzy logic to study the fuzzy clustering problems involving the data and facts shown in Example 9.2. Let the fuzzy concept "*high credit*" be expressed by the fuzzy set $m_6m_4m_1 + m_6m_5m_1 + m_7m_4 + m_7m_5 \in EM$ with the semantic interpretation: "*high salary old male*" or "*more fortune old male*" or "*high salary female*" or "*more fortune female*" (refer to Table 9.2). In the following examples, we apply the AFS structure (M, τ, X) and the weight functions given in Example 9.2 to establish the set of coherence membership functions $\{\mu_\xi(x) \mid \xi \in EM\}$ and the study the clustering problems of the 10 persons by the different selected attribute set $\Lambda \subseteq EM$.

Example 9.3. Let us consider the clustering based on the following attributes "*gender, age*", "*gender, credit*", "*gender, hair white*" , i.e., $\Lambda = \{\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1, \gamma_2\}$, where $\alpha_1 = m_1m_6$, $\alpha_2 = m_1m_7$, $\gamma_1 = m_9m_6$, $\gamma_2 = m_9m_7$, $credit = m_6m_4m_1 + m_6m_5m_1 + m_7m_4 + m_7m_5$,

$$\beta_1 = credit \wedge m_6 = m_6m_4m_1 + m_6m_5m_1 + m_7m_4m_6 + m_7m_5m_6,$$
$$\beta_2 = credit \wedge m_7 = m_6m_4m_1m_7 + m_6m_5m_1m_7 + m_7m_4 + m_7m_5 = m_7m_4 + m_7m_5.$$

Step 1: Find the fuzzy set ϑ. Owing to (4.19), we form ϑ' which is the negation of the fuzzy concept ϑ.

$$\begin{aligned}\vartheta &= \bigvee_{b\in\Lambda} b = \alpha_1 \vee \alpha_2 \vee \beta_1 \vee \beta_2 \vee \gamma_1 \vee \gamma_2\\ &= m_1m_6 + m_1m_7 + m_7m_4 + m_7m_5 + m_9m_6 + m_9m_7\\ \vartheta' &= m_9'm_5'm_4'm_1' + m_9'm_7'm1' + m_7'm_6'.\end{aligned}$$

Table 9.3 Membership degrees of belongingness to the fuzzy concepts ϑ and ϑ'

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
$\mu_\vartheta(.)$	1	.56	1	1	.67	1	.37	.80	1	.11
$\mu_{\vartheta'}(.)$	0	.43	0	0	.14	0	.24	.03	0	.86

The resulting membership functions are shown in Table 9.3. We notice that according to Λ and the original data and facts in Example 9.2, the highest degree of each x_i belonging to any cluster is $\mu_\vartheta(x_i)$.

Step 2: Using (4) of Theorem 9.1, for each $x \in X$, find the fuzzy description ζ_x of x:

$$\zeta_{x_1} = \zeta_{x_5} = m_6m_9,\quad \zeta_{x_2} = \zeta_{x_3} = \zeta_{x_6} = \zeta_{x_{10}} = m_7m_9,$$
$$\zeta_{x_4} = \zeta_{x_8} = m_6m_1,\quad \zeta_{x_7} = m_1m_4m_6 + m_1m_5m_6,\quad \zeta_{x_9} = m_5m_7.$$

Step 3: Apply ζ_x the fuzzy description of each $x \in X$ to establish the fuzzy relation matrix $m_{ij} = \min\{\mu_{\zeta_{x_i}\wedge\zeta_{x_j}}(x_i),\ \mu_{\zeta_{x_i}\wedge\zeta_{x_j}}(x_j)\}$,

$$M = (m_{ij}) = \begin{bmatrix} 1 & 0 & 0 & .09 & .67 & 0 & 0 & .09 & 0 & 0\\ & .56 & .56 & 0 & 0 & .56 & 0 & 0 & 0 & .11\\ & & 1 & 0 & 0 & 1 & 0 & 0 & .17 & .11\\ & & & 1 & .12 & 0 & .37 & .80 & 0 & 0\\ & & & & .67 & 0 & 0 & .12 & 0 & 0\\ & & & & & 1 & 0 & 0 & .08 & .11\\ & & & & & & .37 & .29 & 0 & 0\\ & & & & & & & .80 & 0 & 0\\ & & & & & & & & 1 & 0\\ & & & & & & & & & .11 \end{bmatrix}.$$

One can check that for each $x_i \in X$, occurring in relation matrix M, the similarity degree of x_i with other persons are less than $\mu_\vartheta(x_i)$ (refer to Table 9.3). $Q^2 = Q$, if $Q = M^3, Q$ can yield a partition tree with equivalence classes.

$$Q = \begin{bmatrix} 1 & 0 & 0 & .12 & .67 & 0 & .12 & .12 & 0 & 0 \\ & .56 & .56 & 0 & 0 & .56 & 0 & 0 & .17 & .11 \\ & & 1 & 0 & 0 & 1 & 0 & 0 & .17 & .11 \\ & & & 1 & .12 & 0 & .37 & .80 & 0 & 0 \\ & & & & .67 & 0 & .12 & .12 & 0 & 0 \\ & & & & & 1 & 0 & 0 & .17 & .11 \\ & & & & & & .37 & .37 & 0 & 0 \\ & & & & & & & .80 & 0 & 0 \\ & & & & & & & & 1 & .11 \\ & & & & & & & & & .11 \end{bmatrix}.$$

Step 4: When threshold $\alpha = 1$, then $I_\alpha = 0.2451$.
$C_1 = \{x_1\}$, $\zeta_{C_1} = m_6 m_9$ states "*person who is a white hair male*";
$C_2 = \{x_3, x_6\}$, $\zeta_{C_2} = m_7 m_9$ states "*person who is a white hair female*";
$C_3 = \{x_4\}$, $\zeta_{C_3} = m_6 m_1$ states "*person who is an age male*";
$C_4 = \{x_9\}$, $\zeta_{C_4} = m_5 m_7$ states "*person who is a credit female*".

Similarly, when threshold $\alpha = 0.8$, $I_\alpha = .2496$.

$C_1 = \{x_1\}$, $C_2 = \{x_3, x_6\}$, $C_3 = \{x_4, x_8\}$, $C_4 = \{x_9\}$;
$\zeta_{C_1} = m_6 m_9$, $\zeta_{C_2} = m_7 m_9$, $\zeta_{C_3} = m_6 m_1$, $\zeta_{C_4} = m_5 m_7$.
$\zeta_{boun} = m_6 m_7 m_9 + m_1 m_6 m_9 + m_5 m_7 m_9 + m_1 m_5 m_6 m_7$.

The membership functions are shown in Table 9.4. Compared with Table 9.3, for each x_i, the degree of belonginess (membership) of x_i to any cluster is less than $\mu_\vartheta(x_i)$, where ϑ is the sum of all selected attributes. This implies that the selected attribute set Λ not only determines the cluster results, but also implies the degree each object belongs to every cluster. When we consider the threshold value $\alpha = 0.5$, then $I_\alpha = 0.2235, C_1 = \{x_1, x_5\}$, $C_2 = \{x_2, x_3, x_6\}, C_3 = \{x_4, x_8\}, C_4 = \{x_9\}$; $\zeta_{C_1} = m_6 m_9$, $\zeta_{C_2} = m_7 m_9$, $\zeta_{C_3} = m_6 m_1$, $\zeta_{C_4} = m_5 m_7$. $\zeta_{boun} = m_6 m_7 m_9 + m_1 m_6 m_9 + m_5 m_7 m_9 + m_1 m_5 m_6 m_7$.

In the above example, since we have considered the Boolean attribute "*gender*" in the clustering process, hence the persons in each class are characterized by the same gender. Compared with the above example, in the next example, we consider the attributes "*age*", "*credit*", "*hair white*" without the attribute of "*gender*".

Table 9.4 Membership degrees to each cluster and the boundary

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
$\mu_{\zeta_{C_1}}$	1	0	0	.33	.67	0	0	.33	0	0
$\mu_{\zeta_{C_2}}$	0	.56	1	0	0	1	0	0	.56	.11
$\mu_{\zeta_{C_3}}$	.09	0	0	1	.17	0	.37	.80	0	0
$\mu_{\zeta_{C_4}}$	0	0	.17	0	0	.08	0	0	1	0
ζ_{boun}	.09	0	.17	.33	.12	.08	0	.22	.56	0

Example 9.4. Let us now consider the clustering realized according to attributes "*age*", "*credit*", "*hair white*", i.e., $\Lambda = \{\alpha, \beta, \gamma\}$,where $\alpha = m_1$, $\gamma = m_9$, $\beta = credit = m_6m_4m_1 + m_6m_5m_1 + m_7m_4 + m_7m_5$. By repeating steps 1-3, we obtain fuzzy descriptions for each $x \in X$,

$$\zeta_{x_1} = \zeta_{x_2} = \zeta_{x_3} = \zeta_{x_5} = \zeta_{x_6} = \zeta_{x_{10}} = m_9;$$
$$\zeta_{x_4} = \zeta_{x_8} = m_1, \quad \zeta_{x_7} = m_1m_4m_6 + m_1m_5m_6, \quad \zeta_9 = m_5m_7.$$

The fuzzy equivalence matrix Q yields a partition tree with equivalence classes.

$$\begin{bmatrix} 1 & .56 & 1 & .33 & .67 & 1 & .33 & .33 & .17 & .11 \\ & .56 & .56 & .33 & .56 & .56 & .33 & .33 & .17 & .11 \\ & & 1 & .33 & .67 & 1 & .33 & .33 & .17 & .11 \\ & & & 1 & .33 & .33 & .37 & .81 & .17 & .11 \\ & & & & .67 & .67 & .33 & .33 & .17 & .11 \\ & & & & & 1 & .37 & .33 & .17 & .11 \\ & & & & & & .37 & .37 & .17 & .11 \\ & & & & & & & .81 & .17 & .11 \\ & & & & & & & & 1 & .11 \\ & & & & & & & & & .11 \end{bmatrix}.$$

Considering the threshold $\alpha = 1, I_\alpha = 0.3613$ one has

$C_1 = \{x_1, x_3, x_6\}$, $\zeta_{C_1} = m_9$ reads "*person who is white hair*";
$C_2 = \{x_4\}$, $\zeta_{C_2} = m_1$ reads "*person who is age*";
$C_3 = \{x_9\}$, $\zeta_{C_4} = m_5m_7$ reads "*person who is credit*".
$\zeta_{bou} = m_1m_9 + m_5m_7m_9 + m_1m_5m_7.$

Similarly, when threshold $\alpha = 0.8, I_\alpha = .3495$, we obtain

$C_1 = \{x_1, x_3, x_6\}$, $C_2 = \{x_4, x_8\}$, $C_3 = \{x_9\}$;
$\zeta_{C_1} = m_9$, $\zeta_{C_2} = m_1$, $\zeta_{C_3} = m_5m_7$.
$\zeta_{bou} = m_1m_9 + m_5m_7m_9 + m_1m_5m_7.$

The membership functions are shown in Table 9.5. When the value of the threshold $\alpha = 0.5$, $C_1 = \{x_1, x_2, x_3, x_4, x_5, x_6, x_8, x_9\}$, $\zeta_{C_1} = m_1 + m_9 + m_4m_7 + m_5m_7$.

Table 9.5 Membership degrees pertaining to each cluster and the boundary

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
$\mu_{\zeta_{C_1}}$	1	.56	1	.33	.67	1	0	.33	.56	.11
$\mu_{\zeta_{C_2}}$	.09	.04	.50	1	.17	.26	.37	.81	.64	.01
$\mu_{\zeta_{C_3}}$	.00	0	.54	.50	.03	.30	.37	.29	1	0
$\mu_{\zeta_{bou}}$	.09	.04	.49	.33	.12	.26	0	.22	.64	.01

Since in this example we did not choose attribute of male and female, hence the males and females can be located in a same cluster. In the above examples, we note that $X \neq \bigcup_{1\leq i\leq l} C_i$. This implies that since the membership degrees of x $(x \in X - \bigcup_{1\leq i\leq l} C_i)$ belonging to the fuzzy set ϑ the sum of all selected attributes is too small, hence there is not enough information to determine its cluster. In the real-world problem we cannot ignore the existence of the elements in $X - \bigcup_{1\leq i\leq l} C_i$, we should consider the clustering under the fuzzy attribute set $\overline{\Lambda} = \{\alpha' | \alpha \in \Lambda\} \cup \Lambda$, or let $X - \bigcup_{1\leq i\leq l} C_i$ be single one cluster where its fuzzy description becomes $(\vee_{1\leq i\leq l} \zeta_{C_i})'$.

Example 9.5. In this example, let us consider the clustering problem with the attributes "*old*", "*credit*", "*hair white*" and "*not old*", "*not credit*", "*not hair white*" i.e., $\Lambda = \{\alpha, \beta, \gamma, \alpha', \beta', \gamma'\}$, where $\alpha = m_1$, $\gamma = m_9$, $\beta = credit = m_6m_4m_1 + m_6m_5m_1 + m_7m_4 + m_7m_5$, $\alpha' = m_1'$, $\gamma' = m_9'$, $\beta' = m_6'm_7' + m_4'm_5' + m_7'm_1'$. By running steps 1-3, we form the fuzzy descriptions for each $x \in X$,

$$\begin{aligned}
&\zeta_{x_1} = \zeta_{x_3} = \zeta_{x_5} = \zeta_{x_6} = m_9;\ \ \zeta_{x_2} = m_5'm_4',\\
&\zeta_{x_4} = m_1,\ \ \zeta_{x_7} = m_9',\ \ \zeta_{x_8} = m_1 + m_9',\\
&\zeta_{x_9} = m_5m_7,\ \ \zeta_{x_{10}} = m_1' + m_5'm_4',
\end{aligned}$$

The fuzzy equivalence matrix Q gives rise to the partition tree with the following equivalence classes.

$$\begin{bmatrix}
1 & .46 & 1 & .33 & .67 & 1 & .33 & .33 & .17 & .46\\
 & 1 & .46 & .33 & .46 & .46 & .33 & .33 & .17 & 1\\
 & & 1 & .33 & .67 & 1 & .33 & .33 & .17 & .46\\
 & & & 1 & .33 & .33 & .71 & .81 & .17 & .17\\
 & & & & .67 & .67 & .33 & .33 & .17 & .46\\
 & & & & & 1 & .33 & .33 & .17 & .46\\
 & & & & & & 1 & .71 & .17 & .33\\
 & & & & & & & .81 & .17 & .33\\
 & & & & & & & & 1 & .17\\
 & & & & & & & & & 1
\end{bmatrix}.$$

When threshold $\alpha = 1$, then $I_\alpha = 0.5692$.

$C_1 = \{x_1, x_3, x_6\}$, $\zeta_{C_1} = m_9$ whose interpretation comes as "*person who has white hair*";

$C_2 = \{x_2, x_{10}\}$, $\zeta_{C_2} = m_5'm_4' + m_1'$ with the meaning "*person who is not aged, or not high salary-not more fortune*";

$C_3 = \{x_4\}$, $\zeta_{C_3} = m_1$ with sematic interpretation "*person who is aged*";

$C_4 = \{x_7\}$, $\zeta_{C_4} = m_9'$ with sematic interpretation "*person who does not have white hair*";

$C_5 = \{x_9\}$, $\zeta_{C_5} = m_5m_7$ with sematic interpretation "*person who has credit*".

$\zeta_{bou} = m_9m_1' + m_9m_5'm_4' + m_1m_9 + m_9m_9' + m_5m_7m_9 + m_1m_1' + m_1m_5'm_4' + m_9'm_1' + m_9'm_5'm_4' + m_5m_7m_1' + m_5m_7m_5'm_4' + m_1m_9' + m_1m_5m_7 + m_5m_7m_9'$.

Table 9.6 Membership degrees of the concepts, see the details above

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
$\mu_{\zeta_{C_1}}$	1	.56	1	.33	.67	1	0	.33	.56	.11
$\mu_{\zeta_{C_2}}$	.68	1	.15	.04	.51	.36	.24	.10	.08	1
$\mu_{\zeta_{C_3}}$	.09	.43	.49	1	.17	.26	1	.81	.64	.86
$\mu_{\zeta_{C_4}}$	0	0	.17	0	0	.08	0	0	1	0
$\mu_{\zeta_{bou}}$	.68	.46	.49	.33	.29	.36	.24	.22	.64	.86

Similarly, when threshold $\alpha = 0.6$, $I_\alpha = .4815$.

$C_1 = \{x_1, x_3, x_5, x_6\}$, $C_2 = \{x_2, x_{10}\}$, $C_3 = \{x_4, x_7, x_8\}$, $C_4 = \{x_9\}$, $\zeta_{C_1} = m_9$, $\zeta_{C_2} = m_1' + m_5'm_4'$, $\zeta_{C_3} = m_1 + m_9'$, $\zeta_{C_4} = m_5m_7$.

$\zeta_{bou} = m_9m_1' + m_9m_5'm_4' + m_1m_9 + m_9m_9' + m_5m_7m_9 + m_1m_1' + m_9'm_1' + m_1m_5'm_4' + m_9'm_5'm_4' + m_5m_7m_1' + m_5m_7m_5'm_4' + m_1m_5m_7 + m_5m_7m_9'$.

Compared with the two previous examples, since in this example we consider the negation of each selected attributes, hence any person belongs to at least one cluster at a high degree of membership. The fuzzy set ζ_{bou} indicates that a certain person may belong to more than one cluster at high degree; for example this happens in case of x_{10}. Making use of these examples, we can conclude that the clustering results are appealing from an intuitive point of view.

Example 9.6. The data set [15, 24] used for this problem is shown in Table 9.7. It consists of data of fats and oils having four quantitative features of interval type and one qualitative feature. First, for the attributes in Table 9.7 represented by interval $[a, b]$, the interval $[a, b]$ is normalized (represented) as two numbers $(a+b)/2, b-a$. For example, the original data for sample Linseed oil are normalized as 0.9325, 0.0050, -17.5000, 19.0000, 187.0000, 34.0000, 157.0000, 78.0000. Let sample No. i be normalized as $S_i = (s_{i1}, s_{i2}, ..., s_{i8}), i = 0, 1, ..., 7$. Let m_1 be the simple concept: "*the average Gravity of the sample is high*", m_2 be the negation of simple concept m_1, i.e., $m_2 = m_1'$; m_3 be the simple concept: "*the difference between highest and lowest values of the Gravity of sample is high*", m_4 be the negation of simple concept m_3; m_5 be the simple concept: "*the average Freezing point of the sample is*

Table 9.7 Fat-Oil data

No.	Sample name	Gravity(g/cm^3)	Freezing point	io.value	sa.value	m.f.acids
0	Linseed oil	0.930-0.935	-27 to -8	170-204	118-196	L,Ln,O,P,M
1	Perilla oil	0.930-0.937	-5 to -4	192-208	188-197	L.Ln,O,P,S
2	Cotton-seed	0.916-0.918	-6 to -1	99-113	189-198	L,O,P,M,S
3	Seaame oil	0.920-0.926	-6 to -4	104-116	187-193	L,O,P,S,A
4	Camellia	0.916-0.917	-21 to -15	80-82	189-193	L,O
5	Olive-oil	0.914-0.919	0 to 6	79-90	187-196	L,O,P,S
6	beef-tallow	0.860-0.870	30 to 38	40-48	190-199	O,P,M,S,C
7	Lard	0.858-0.864	22 to 32	53-77	190-202	L,O,P,M,S,Lu

high", m_6 be the negation of simple concept m_5; m_7 be the simple concept: "*the difference between highest and lowest values of the Freezing point of sample is high*", m_8 be the negation of simple concept m_7; m_9 be the simple concept: "*the average io.value of the sample is high*", m_{10} be the negation of simple concept m_9; m_{11} be the simple concept: "*the difference between highest and lowest values of io.value of sample is high*", m_{12} be the negation of simple concept m_{11}; m_{13} be the simple concept: "*the average sa.value of the sample is high*", m_{14} be the negation of simple concept m_{13}; m_{15} be the simple concept: "*the difference between highest and lowest values of sa.value of sample is high*", m_{16} be the negation of simple concept m_{15}; m_{17} be the simple concept: "*symbol likes 'L,Ln,O,P,M'* ", m_{18} is the negation of simple concept m_{17}; m_{19} be the simple concept: "*symbol likes 'L.Ln,O,P,S'*", m_{20} be the negation of simple concept m_{19}; m_{21} be the simple concept: "*symbol likes 'L,O,P,M,S'*", m_{22} be the negation of simple concept m_{21}; m_{23} be the simple concept: "*symbol likes 'L,O,P,S,A'*", m_{24} be the negation of simple concept m_{23}; m_{25} be the simple concept: "*symbol likes 'L,O'* ", m_{26} be the negation of simple concept m_{25}; m_{27} be the simple concept: "*symbol likes 'L,O,P,S'*", m_{28} be the negation of simple concept m_{27}; m_{29} be the simple concept "*symbol likes 'O,P,M,S,C'*", m_{30} be the negation of simple concept m_{29}; m_{31} be the simple concept: "*symbol likes 'L,O,P,M,S,Lu'*", m_{32} be the negation of simple concept m_{31}. We obtain the weight function ρ_{m_i} for the simple concept m_i in the following way: For $q = 2k+1, 0 \le k \le 7$, let $L_k = \min\{s_{0k}, s_{1k}, ..., s_{7k}\}$, $H_k = \max\{s_{0k}, s_{1k}, ..., s_{7k}\}$.

$$\rho_{m_q}(S_i) = \begin{cases} \frac{s_{ik}-L_k}{H_k-L_k} H_k, & \text{when } L_k < s_{ik} \le H_k, \\ 0, & \text{when } s_{ik} \le L_k, \\ 1, & \text{when } H_k < s_{ik}, \end{cases}$$
$$\rho_{m'_q}(S_i) = 1 - \rho_{m_q}(S_i).$$

For the fuzzy concept on words, $q = 2k+1, 8 \le k \le 15$, we have

$$\rho_{m_q}(S_i) = 1 - \frac{|w_{k-8}| + |w_i| - 2|w_{k-8} \cap w_i|}{|w_{k-8} \cup w_i|},$$

where w_i is the set of words representing attribute m.f.acids, for sample, S_i. For example $k = 8$, m_{17} be the simple concept "*symbol likes 'L,Ln,O,P,M'*", $\rho_{m_{17}}(S_0) = 1$, $\rho_{m_{17}}(S_1) = 0.8$, $\rho_{m_{17}}(S_2) = 0.8$, $\rho_{m_{17}}(S_3) = 0.6$, $\rho_{m_{17}}(S_4) = 0.57$, $\rho_{m_{17}}(S_5) = 0.67$, $\rho_{m_{17}}(S_6) = 0.6$, $\rho_{m_{17}}(S_7) = 0.73$. In what follows, we set up the AFS structure (M, τ, X) according to the data and facts in Table 9.7, where $X = \{S_0, S_1, ..., S_7\}$ and $M = \{m_1, m_2, ..., m_{32}\}$.

$$\tau(S_i, S_i) = \{m_l \mid \rho_{m_l}(S_i) > 0\}, \ \ \tau(S_i, S_j) = \{m_l \mid \rho_{m_l}(S_i) \ge \rho_{m_l}(S_j)\}, \ for\ i \ne j.$$

One can verify that (M, τ, X) is an AFS structure. Thus the set of coherence membership functions $\{\mu_\xi(x) \mid \xi \in EM\}$ can be obtained by formula (5.18) given in Proposition 5.6. In this example, let us $\Lambda = M$. Through step 2, we get the fuzzy description ζ_{S_i} for each sample S_i as follows:

$$\zeta_{S_0} = m_1m_7m_9m_{11}m_{14}m_{15}m_{17},\ \zeta_{S_1} = m_1m_8m_9m_{19},\ \zeta_{S_2} = m_4m_{13}m_{21} + m_4m_{16}m_{21},$$
$$\zeta_{S_3} = m_8m_{16}m_{23}, \zeta_{S_4} = m_4m_6m_{12}m_{16}m_{25},\ \zeta_{S_5} = m_{12}m_{27} + m_{16}m_{27},$$
$$\zeta_{S_6} = m_3m_5m_{10}m_{29},\ \zeta_{S_7} = m_2m_{13}m_{31}.$$

By carrying out steps 3-4, we obtain

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & .28 & .28 & .15 & .28 & .14 & .14 \\ 0 & .28 & .66 & .34 & .15 & .34 & .14 & .14 \\ 0 & .28 & .34 & .67 & .15 & .35 & .14 & .14 \\ 0 & .15 & .15 & .15 & 1 & .15 & .14 & .14 \\ 0 & .28 & .34 & .35 & .15 & .70 & .14 & .14 \\ 0 & .14 & .14 & .14 & .14 & .14 & 1 & .48 \\ 0 & .14 & .14 & .14 & .14 & .14 & .48 & 1 \end{bmatrix}$$

If the threshold α is set to 0.15, then $C_1 = \{S_0\}, C_2 = \{S_1, S_2, S_3, S_4, S_5\}, C_3 = \{S_6, S_7\}$ which are the same as the result in [23]. The fuzzy descriptions of the classes are shown as follows.

$$\zeta_{C_1} = m_1m_7m_9m_{11}m_{14}m_{15}m_{17},\ \ \zeta_{C_3} = m_3m_5m_{10}m_{29} + m_2m_{13}m_{31},$$
$$\zeta_{C_2} = m_1m_8m_9m_{19} + m_4m_{13}m_{21} + m_4m_{16}m_{21} + m_8m_{16}m_{23} + m_4m_6m_{12}m_{16}m_{25} + m_{12}m_{27} + m_{16}m_{27},$$

Their membership functions are included in Table 9.8. If more detailed clustering is required, then a higher value of the threshold α can be selected.

Table 9.8 Membership degrees belonging to each cluster and the boundary

	S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7
$\mu_{\zeta_{C_1}}$	1	0	.04	.02	0	.05	0	0
$\mu_{\zeta_{C_2}}$	0	1	.66	.71	1	.70	.17	.29
$\mu_{\zeta_{C_3}}$	0	0	.13	.08	.08	.22	1	1
$\mu_{\zeta_{bou}}$	0	0	.13	.08	.08	.14	.17	.29

9.2 Applications of the Elementary Fuzzy Clustering for Management Strategic Analysis

Cluster analysis has been used frequently in product position, strategy formulation, market segmentation studies and business system planning. In addition, we could discriminate one or more strategies from airfreight industry and to comprehend the competitive situation in more detail.

In this section, first the elementary fuzzy clustering method via AFS fuzzy logic is investigated further by amending the algorithm to be more applicable to management strategic analysis. Next it will be used to analyze the evaluation results of 30 companies which have been studied and analyzed by G.-S. Liang et al. [56]. Compared with the Liang's algorithm, the elementary fuzzy clustering method is more transparent, understandable and the results are easy to interpret. The method can be applied to the management strategic analysis based on the data sets described by mixed features such as real numbers, Boolean logical values, and linguistic descriptions. The illustrative examples show that the interpretations of the clustering results of the 30 companies are highly consistent with the expert's intuition.

In [56], the authors defined the linguistic values by the trapezoidal fuzzy numbers which are represented as four-dimension vectors. For example, they defined "*Very Low*"as $(0,0,0,0.2)$. They proposed a clustering which is based on the following idea.

First, the distance function between two trapezoidal fuzzy numbers is used to aggregate the linguistic values about attribute ratings to obtain the compatibility relation. Then a fuzzy equivalence relation based on the fuzzy compatibility relation is constructed. Finally, they determined the best number of clusters using a cluster validity index which also depends on the distances computed for the trapezoidal fuzzy numbers.

In this section, instead of subjectively defining the linguistic values as done in [56], the membership functions are determined by the AFS algorithm according to the ordered relations formed for the attributes and the semantics of the fuzzy concepts. Compared with the Liang's method, the clustering algorithm has the following advantages:

- The features of the data sets can be mixed.
- The trapezoidal fuzzy numbers and their distance function are not required. This can help avoid inconsistent results of Liang's algorithm due to different choices of the fuzzy numbers and the distance functions. (This shortcoming will be shown in Section 9.2.2 by running some experiments).
- The clustering results are easy to interpret.

The experimental study on the evaluation results of 30 companies shows that some aspects of the elementary fuzzy clustering method via AFS fuzzy logic need to be improved. Thus, the algorithm has been enhanced in the following manner:

1. The description of each object is optimized by a selecting method.
2. The fuzzy cluster validity index is improved by adding the rate of the number of clusters and the threshold α.

The applications of the improved clustering algorithm to the evaluation results of the data sets in [56] show that the interpretation of each cluster is consistent with experts' intuitions, and the algorithm can be applied to the management strategic analysis for the data sets with mixed features.

Example 9.7 serves as an introductory illustration to show how to undertake strategic analysis using the elementary fuzzy clustering method.

Example 9.7. Let $\{c_1, c_2, ..., c_5\}$ be the set of five companies. Factor1,..., Factor7 are seven factors (attributes or features) obtained from experts, where Factor1: "*Core ability*", Factor2: "*Organization management*", Factor3: "*Pricing*", Factor4:

Table 9.9 Evaluation Results of the Five Companies

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
c_1	M	H	H	B.H&VH	VH	L	B.M &H
c_2	H	B.L& M	M	B.M & H	H	B.M &H	VL
c_3	H	H	B.M &H	H	H	VH	B.M & H
c_4	VL	M	H	B.VL & L	H	B.L &M	M
c_5	L	M	B.H &VH	H	B.H &VH	B.VL &L	B.M& H

"*Competitive forces*", Factor5: "*Finance*", Factor6: "*Different advantage*", Factor7: "*Information technology*". The evaluation results of the five companies are shown in Table 9.9 which is taken from [56]. Where $VL = Very\ Low$, $B.VL\&L = Between\ Very\ Low\ and\ Low$, $L = Low$, $B.L\&M = Between\ Low\ and\ Medium$, $M = Medium$, $B.M\&H = Between\ Medium\ and\ High$, $H = High$, $B.H\&VH = Beween\ High\ and\ Very\ High$, $VH = Very\ High$.

Let $X = \{c_1, c_2, ..., c_5\}$ be a set of the five companies, $M = \{m_1, m_2, ..., m_7\}$ be a set of fuzzy attributes on X. Where m_1: "*Factor1 is strong*", m_2: "*Factor2 is strong*",..., m_7: "*Factor7 is strong*". Each fuzzy concept in EM represents a definitely sematic interpretation. For instance, we may have γ: $m_2m_3 + m_2m_4$ which translates as "*Organization management and Pricing are strong "or "Organization management and Competitive force are strong* ".

Next, we demonstrate how to establish an AFS structure according to the original data in Example 9.7. Let $X = \{c_1, c_2, ..., c_5\}$, $M = \{m_1, m_1', m_2, m_2', ..., m_7, m_7'\}$, where m_1': "*Factor1 is not strong*", m_2': "*Factor2 is not strong*"; ..., m_7': "*Factor7 is not strong*". For the semantic meanings of the linguistic values, we have the following order relation:

"VH"> "B.H&VH"> "H"> "B.M&H"> "M"> "B.L&M"> "L"> "B.VL&L"> "VL".

Using Table 9.9 and taking into account the semantics the attributes in M, we have the following order relations of the simple concepts in M:

m_1: $c_4 <_{m_1} c_5 <_{m_1} c_1 <_{m_1} c_3 =_{m_1} c_2$, m_1': $c_4 >_{m_1'} c_5 >_{m_1'} c_1 >_{m_1'} c_3 =_{m_1'} c_2$.
m_2: $c_2 <_{m_2} c_4 =_{m_2} c_5 <_{m_2} c_3 =_{m_2} c_1$, m_2': $c_2 >_{m_2'} c_4 =_{m_2'} c_5 >_{m_2'} c_3 =_{m_2'} c_1$.
m_3: $c_2 <_{m_3} c_3 <_{m_3} c_4 =_{m_3} c_1 <_{m_3} c_5$, m_3': $c_2 >_{m_3'} c_3 >_{m_3'} c_4 =_{m_3'} c_1 >_{m_3'} c_5$.
m_4: $c_4 <_{m_4} c_2 <_{m_4} c_5 =_{m_4} c_3 <_{m_4} c_1$, m_4': $c_4 <_{m_4'} c_2 <_{m_4'} c_5 =_{m_4'} c_3 <_{m_4'} c_1$.
m_5: $c_2 =_{m_5} c_3 =_{m_5} c_4 <_{m_5} c_5 <_{m_5} c_1$, m_5': $c_2 =_{m_5'} c_3 =_{m_5'} c_4 >_{m_5'} c_5 >_{m_5'} c_1$
m_6: $c_5 <_{m_6} c_1 <_{m_6} c_4 <_{m_6} c_2 <_{m_6} c_3$, m_6': $c_5 >_{m_6'} c_1 >_{m_6'} c_4 >_{m_6'} c_2 >_{m_6'} c_3$
m_7: $c_2 <_{m_7} c_4 <_{m_7} c_5 =_{m_7} c_3 =_{m_7} c_1$, m_7': $c_2 >_{m_7'} c_4 >_{m_7'} c_5 =_{m_7'} c_3 =_{m_7'} c_1$.

Thus by (4.26), the AFS structure (M, τ, X) of Table 9.9 is well-defined. For the weight function $\rho : X \to [0,1]$, $\rho(x) = 1$ for any $x \in X$. Then the set of coherence membership functions $\{\mu_\xi(x) \mid \xi \in EM\}$ can be obtained by (5.24) in Proposition 5.7.

$$\mu_\eta(x) = \sup_{i \in I} \frac{|A_i^\tau(x)|}{|X|}, \tag{9.11}$$

The esxpression $A_i^\tau(x)$ is calculated by (4.27). If $\sigma = 2^X$, for $W \in 2^X$, $\mathscr{M}(W) = |W|$ ($|W|$ is the cardinal number of the set W, i.e., the number of elements in W) in Proposition 5.7. In Example 9.7, let $\eta_1 = m_1$, $\eta_2 = m_2$, $\eta_3 = m_1 m_2$, $\eta_4 = m_1 + m_2 \in EM$. By the formula (9.11), we get:

For $\eta_1, A = \{m_1\}, A^\tau(c_1) = \{c_1, c_4, c_5\}$, $\mu_{\eta_1}(c_1) = \frac{|A^\tau(c_1)|}{|X|} = 3/5 = 0.6$.
For $\eta_2, A = \{m_2\}, A^\tau(c_1) = \{c_1, c_2, c_3, c_4, c_5\}$ $\mu_{\eta_2}(c_1) = \frac{|A^\tau(c_1)|}{|X|} = 5/5 = 1.0$.
For $\eta_3, A = \{m_1, m_2\}, A^\tau(c_1) = \{c_1, c_4, c_5\}$, $\mu_{\eta_3}(c_1) = \frac{|A^\tau(c_1)|}{|X|} = 3/5 = 0.6$.
For $\eta_4, A_1 = \{m_1\}, A_2 = \{m_2\}$, $\mu_{\eta_4}(c_1) = \sup_{i=1,2}(\frac{|A_i^\tau(c_1)|}{|X|}) = \sup\{3/5, 5/5\} = 1.0$.

Remark 9.2. Compared with the Liang's method, the membership function defined by (9.11) depends on the ordered relations on the attributes and the AFS structure of the data without relying on the subjectively defined membership functions of the trapezoidal fuzzy numbers.

9.2.1 Improvements of the Elementary Fuzzy Clustering Method

By immediate applications of the elementary fuzzy clustering method to the data of [56] shown as Table 7.4 in the Appendix A in Chapter 7, it is sometimes difficult to obtain satisfactory results. Through a careful analysis of the algorithm, we find that the following two issues (i.e., step 2 and step 5 in the algorithm) contribute to the lower performance:

a) The description of each object cannot characterize it well enough. In the algorithm, $\zeta_{c_i} = \bigvee_{\alpha \in \Lambda_{c_i}^\varepsilon} \alpha$, the final description of an object c_i is the sum (i.e., the *EI* algebra operation "$\vee$", refer to (9.10)) of all fuzzy concepts in $\Lambda_{c_i}^\varepsilon$ defined by (9.8), where $\Lambda_{c_i}^\varepsilon$ is the set of all feasible fuzzy descriptions of the object. Because $\Lambda_{c_i}^\varepsilon$ often includes the descriptions of both essential and redundant characteristics of the object, the final description of the object may be too "*rough*" so that it may include the "*improper*" descriptions in $\Lambda_{c_i}^\varepsilon$ which describe the redundant nature of the object. The improper description always lowers the clustering accuracy. In the improved algorithm, we just choose the best description from all the feasible descriptions in $\Lambda_{c_i}^\varepsilon$ as the final description of the object through running a selection method.

b) The fuzzy cluster validity index (9.4) which is used to select the best clustering result considers only the clarity of the boundary among the clusters. However, the number of the clusters is also an important factor which influences the quality of the clustering results. Thus, in the new fuzzy cluster validity index, both the clarity of the boundary and the number of the clusters have been considered.

Improvements of the elementary fuzzy clustering method via the AFS fuzzy logic

Let $X = \{c_1, c_2, ..., c_n\}$ be the set of the objects, $M = \{m_1, m_2, ..., m_p\}$ be a set of attributes on X. In what follows, the AFS structure is constructed by (4.26) and the membership functions of the fuzzy concepts in EM are defined by (9.11). Let $\vartheta = \sum_{m_i \in M} m_i$.

STEP 1: Find the fuzzy concept $\zeta_{c_i} \in EM$ to describe the object $c_i \in X$, which satisfies that not only the membership degree of c_i belonging to ζ_{c_i} (i.e., $\mu_{\zeta_{c_i}}(c_i)$) is the most approach to the membership degree of c_i belonging to ϑ (i.e., $\mu_\vartheta(c_i)$), but also $\mu_{\zeta_{c_i}}(c_j)$ is as small as possible for $c_j \in X, i \neq j$. In other words, c_i can be distinguished by ζ_{c_i} from other objects in X to a maximal extent.

The best fuzzy description ζ_{c_i} for each object c_i is determined by running the following procedure:

– Let $\varepsilon \geq 0$ (in the examples in this section $\varepsilon = 0$). Find the set $B^\varepsilon_{c_i}$ defined as follows:

$$B^\varepsilon_{c_i} = \left\{ m_k \in M \mid \mu_{m_k}(c_i) \geq \mu_\vartheta(c_i) - \varepsilon \right\} \tag{9.12}$$

$B^\varepsilon_{c_i}$ is the set of the fuzzy attributes in M the degrees of c_i belonging them are larger than or equal to $\mu_\vartheta(c_i) - \varepsilon$.

– Find the set $\overline{B}^\varepsilon_{c_i}$ defined as follows:

$$\overline{B}^\varepsilon_{c_i} = \left\{ \prod_{m \in A} m \mid \mu_{\prod_{m \in A} m}(c_i) \geq \mu_\vartheta(c_i) - \varepsilon, \ A \subseteq B^\varepsilon_{c_i} \right\} \tag{9.13}$$

$\overline{B}^\varepsilon_{c_i}$ is the set of the conjunctions of the attributes in $B^\varepsilon_{c_i}$ such that the degrees of c_i belonging to the conjunctions are larger than or equal to $\mu_\vartheta(c_i) - \varepsilon$.

– Select the best fuzzy description $\zeta_{c_i} \in \overline{B}^\varepsilon_{c_i}$ for the object c_i:

$$\zeta_{c_i} = \arg \min_{\zeta \in \overline{B}^\varepsilon_{c_i}} \left\{ \sum_{c \in X, c \neq c_i} \mu_\zeta(c) \right\} \tag{9.14}$$

Thus c_i can be distinguished by ζ_{c_i} from other objects in X to the highest degree.

Remark 9.3. For $\alpha, \beta \in \overline{B}^\varepsilon_{c_i}$, if $\alpha \leq \beta$ in the lattice $(EM, \wedge, \vee)$, then for any $c \in X$, $\mu_\alpha(c) \leq \mu_\beta(c)$. By (9.14), the description ζ_{c_i} can be simply found just by checking the membership degrees of c_i belonging to the minimal elements in $\overline{B}^\varepsilon_{c_i}$. In general, there may be many elements in $\overline{B}^\varepsilon_{c_i}$, but there are often just a few of the minimal elements.

STEP 2: Apply the fuzzy description ζ_{c_i} of each $c_i \in X$ to establish the fuzzy relation matrix $F = (f_{ij})_{n \times n}$ on $X = \{c_1, c_2, ..., c_n\}$, where

$$f_{ij} = \min \left\{ \mu_{\zeta_{c_i} \wedge \zeta_{c_j}}(c_i), \ \mu_{\zeta_{c_i} \wedge \zeta_{c_j}}(c_j) \right\}.$$

Remark 9.4. In [56], the authors used the distance function $d_p(N_i, N_k)$ between two fuzzy numbers N_i, N_k to define the fuzzy compatibility relation. Different distance functions may lead to different fuzzy compatibility relations. Thus the inconsistent clustering results may be obtained for the same data set. However, in STEP 2, the fuzzy relation matrix F is uniquely determined by the AFS structure (M, τ, X) of the data. This implies that the results of the proposed clustering algorithm are objectively determined by the original data.

STEP 3: Find the feasible clusters corresponding to the threshold $\alpha \in [0,1]$. Let $Q = F^r = (q_{ij})_{n\times n}$ and the Boolean matrix $Q_\alpha = (q_{ij}^\alpha)_{n\times n}$, where $q_{ij}^\alpha = 1 \Leftrightarrow q_{ij} \geq \alpha$, $\alpha \in [0,1]$. For $\alpha \in [0,1]$, $c_i, c_j \in X$, c_i, c_j are in the same cluster under threshold α if and only if $q_{ij}^\alpha = 1$. For some $c_i \in X$, if $q_{ii}^\alpha = 0$ then the threshold α cannot be determined which cluster the object c_i belongs to. For each object $c_i \in X$, if $q_{ii}^\alpha = 1$, $q_{ij}^\alpha = 0$, for any j, $i \neq j$, i.e., the object c_i itself is a cluster, then this clustering result under the threshold α is considered to be invalid.

Remark 9.5. In practice, the values of the thresholds α in the range $\{q_{ij} | 1 \leq i, j \leq n\}$ are just considered.

STEP 4: Determine the best clustering result out of all the results under the threshold $\alpha \in \{q_{ij} | 1 \leq i, j \leq n\}$ using the fuzzy cluster validity index I_α defined by (9.15) and compute the fuzzy description ζ_{C_k} of the cluster C_k via the AFS logic operation $\vee$. For each cluster $C_k \subseteq X$, where C_k is a cluster under the threshold α, $\zeta_{C_k} = \bigvee_{c \in C_k} \zeta_c$ is the fuzzy description of the cluster C_k. It is clear that for each $c \in C_k$, not only the membership degree of c belonging to the cluster C_k, which is $\mu_{\zeta_{C_k}}(c)$, is as large as possible, but also $\mu_{\zeta_{C_k}}(y)$ is as small as possible for $y \in X, y \notin C_k$. In other words, the objects in the cluster C_k can be distinguished by ζ_{C_k} from the objects outside C_k to the maximal extent.

The fuzzy concept $\zeta_{bou} = \bigvee_{1 \leq k_1, k_2 \leq l, k_1 \neq k_2} (\zeta_{C_{k_1}} \wedge \zeta_{C_{k_2}}) \in EM$ describes the boundary among the feasible clusters $C = \{C_1, C_2, ..., C_l\}$. Since $\mu_{\zeta_{bou}}(c)$ produces the membership degree of each object in X belonging to the boundary ζ_{bou}, hence it can be used to evaluate the clarity of the boundary among the clusters. Thus the new fuzzy cluster validity index I_α is defined as follows:

$$I_\alpha = \frac{1}{\alpha^2} \times \frac{\sum_{c \in \bigcup_{1 \leq k \leq l} C_k} \mu_{\zeta_{bou}(c)}}{\sum_{c \in \bigcup_{1 \leq k \leq l} C_k} \mu_{\zeta_{Total}(c)}} + \frac{|C|}{|X|}. \tag{9.15}$$

Where $\zeta_{Total} = \bigvee_{1 \leq k \leq l} \zeta_{C_k}$ for $l \geq 2$. $|C|$ is the number of the clusters, $|X|$ is the number of the objects. It is obvious that the lower the value of I_α is, the clearer and the better the clustering result under threshold α is. Thus the best clustering result can be selected by looking at the value of I_α.

9.2.2 Experimental Study of the Liang's Algorithm

In [56], the authors proposed a cluster analysis method based on fuzzy equivalence relation by the distance function between two trapezoidal fuzzy numbers. The method comprises of 5 steps:

- Express the original attributes in terms of predefined trapezoidal fuzzy numbers and normalize the original attribute preference rating.
- Using the distance function between two trapezoidal fuzzy numbers to aggregate the linguistic values to obtain the compatibility relation matrix.
- Find the fuzzy equivalence relation matrix based on the fuzzy compatibility relation matrix.
- Find all feasible clusters induced by the fuzzy equivalence relation matrix.
- Using a cluster validity index determine the best number of clusters.

By the application of Liang's algorithm, we find that the different choices of the trapezoidal fuzzy numbers and their distance functions can lead to different clustering results. For example, two different selections of fuzzy numbers and distance function are shown as follows:

Selection 1: "VL": $(0,0,0,0)$, "B.VL&L": $(0,0,0.1,0.2)$, "L": $(0,0.2,0.2,0.2)$, "B.L&M": $(0,0.2,0.4,0.5)$, "M": $(0,0.3,0.6,0.7)$, "B.M&H": $(0.3,0.5,0.8,1)$, "H": $(0.6,0.8,0.8,1)$, "B.H&VH": $(0.6,0.8,0.9,1)$, "VH":$(1,1,1,1)$, and $p=3$ for the distance function $d_p(\cdot\,,\,\cdot)$.
Selection 2: "VL": $(0,0,0,0.2)$, "B.VL&L": $(0,0,0.2,0.4)$, "L": $(0,0.2,0.2,0.4)$, "B.L&M": $(0,0.2,0.5,0.7)$, "M": $(0.3,0.5,0.5,0.7)$, "B.M&H": $(0.3,0.5,0.8,1)$, "H": $(0.6,0.8,0.8,1)$, "B.H&VH": $(0.6,0.8,1,1)$, "VH":$(0.8,1,1,1)$, and $p=2$ for the distance function $d_p(\cdot\,,\,\cdot)$. This selection is the same as Liang's in [56].

Applying Liang's algorithm to the same data as shown in Table 9.9, different results from [56] are obtained by Selection 1. More details are shown in Table 9.10.

Table 9.10 The clustering results for Selection 1

λ interval	Number of clusters	L value	clusters
(0.3788,0.3842)	2	0.3497	$\{c_2,c_3\}$, $\{c_1,c_4,c_5\}$
(0.3842,0.5308)	3	0.1925	$\{c_2\}$, $\{c_3\}$, $\{c_1,c_4,c_5\}$
(0.5308,0.5543)	4*	0.0936	$\{c_1\}$,$\{c_2\}$, $\{c_3\}$ $\{c_4,c_5\}$

Thus, the best fuzzy clustering result for Selection 1 is $\{c_1\}$, $\{c_2\}$, $\{c_3\}$, $\{c_4,c_5\}$. However, in [56] the best clustering result for Selection 2 is $\{c_2\}$, $\{c_3\}$, $\{c_4\}$ and $\{c_1,c_5\}$.

It is clear that the different selections of the fuzzy numbers and the distance function for Liang's algorithm may lead to inconsistent results. Furthermore the method to select the suitable fuzzy numbers and distance function has not been established yet. Thus it is very hard to objectively analyze management strategy formed on a base of this data.

Applying the Liang's algorithm with the same fuzzy numbers and the distance function in [56] (i.e., Selection 2) to the evaluation results of 30 companies, we obtain the following clustering result: $C_1=\{c_8\}$, $C_2=\{c_{11}\}$, $C_3=\{c_{24}\}$, $C_4=\{c_{26}\}$, $C_5=\{\text{The rest of the objects}\}$ through the same cluster validity index L given in [56]. However, it is difficult to explain and interpret the clustering result.

9.2.3 Applications of the Improved Algorithm

In this section, we apply the improved algorithm to the example which has been studied in [56]. The data set contains 30 objects and seven factors (attributes) shown in Table 9.9 of Example 9.7. Table 9.9 lists the data for 5 companies. For simplicity, we first use the small data with 5 companies to illustrate the performance of the proposed algorithm. Next we proceed with the entire data set.

9.2.3.1 Application of the Improved Algorithm to Data of 5 Companies

We continue to study the data shown in Example 9.7. Let $X = \{c_1, c_2, ..., c_5\}$, $M = \{m_1, m'_1, ..., m_7, m'_7\}$, $\vartheta = m_1 + m'_1 + ... + m_7 + m'_7$. By (9.11), we can get: $\mu_\vartheta(c_1) = \mu_\vartheta(c_2) = \mu_\vartheta(c_3) = \mu_\vartheta(c_4) = \mu_\vartheta(c_5) = 1.0$.

STEP 1. Let $\varepsilon = 0$, for c_1: $\mu_{m_2}(c_1) = \mu_{m_4}(c_1) = \mu_{m_5}(c_1) = \mu_{m_7}(c_1) = 1 = \mu_\vartheta(c_1)$. By (9.12), we have $B^0_{c1} = \{m_2, m_4, m_5, m_7\}$.

Since $\mu_{m_2m_4m_5m_7}(c_1) = 1 = \mu_\vartheta(c_1)$, by (9.13) and Remark 9.3, we know that $m_2m_4m_5m_7$ is the minimal element in $\Lambda^0_{c_1} = \{m_2, m_4, m_5, m_7, m_2m_4, m_2m_5, m_2m_7, m_4m_5, m_4m_7, m_5m_7, m_2m_4m_5, m_2m_4m_7, m_4m_5m_7, m_2m_5m_7, m_2m_4m_5m_7\}$. So,

$\zeta_{c_1} = m_2m_4m_5m_7$. In the same way we obtain the others:

$B^0_{c_2} = \{m_1, m'_2, m'_3, m'_5, m'_7\}$, $\zeta_{c_2} = m_1m'_2m'_3m'_5m'_7$.
$B^0_{c_3} = \{m_1, m_2, m'_5, m_6, m_7\}$, $\zeta_{c_3} = m_1m_2m'_5m_6m_7$.
$B^0_{c_4} = \{m'_1, m'_4, m'_5\}$, $\zeta_{c_4} = m'_1m'_4m'_5$.
$B^0_{c_5} = \{m_3, m'_6, m_7\}$, $\zeta_{c_5} = m_3m'_6m_7$.

STEP 2. The fuzzy relation matrix F and F^2 are shown as follows.

$$F = \begin{bmatrix} 1.0 & 0.2 & 0.2 & 0.2 & 0.6 \\ & 1.0 & 0.2 & 0.2 & 0.2 \\ & & 1.0 & 0.2 & 0.2 \\ & & & 1.0 & 0.4 \\ & & & & 1 \end{bmatrix}, \qquad F^2 = \begin{bmatrix} 1.0 & 0.2 & 0.2 & 0.4 & 0.6 \\ & 1.0 & 0.2 & 0.2 & 0.2 \\ & & 1.0 & 0.2 & 0.2 \\ & & & 1.0 & 0.4 \\ & & & & 1.0 \end{bmatrix}.$$

Since $F^2 = (F^2)^2$, hence $Q = (q_{ij}) = F^2$ can yield a partition tree with equivalence classes. By Remark 9.5, the threshold α can be chosen as 0.2, 0.4, 0.6, 1.0.

STEP 3-4. When threshold $\alpha = 0.2$, there is only one cluster:

$$C_1 = \{c_1, c_2, c_3, c_4, c_5\},\ I_{0.2} = 25.2000.$$

When threshold $\alpha = 0.4$, we encounter the following three clusters:

$$C_1 = \{c_2\}, C_2 = \{c_3\}, C_3 = \{c_1, c_4, c_5\}, I_{0.4} = 2.6000.$$

When threshold $\alpha = 0.6$, there are four clusters shown as follows:

$$C_1 = \{c_2\}, C_2 = \{c_3\}, C_3 = \{c_4\}, C_4 = \{c_1, c_5\}.$$

The descriptions $\zeta_{C_i},\ i=1,2,...5$ of the clusters C_1-C_5 come in the form:

$\zeta_{C_1}=\zeta_{c_2}=m_1m_2'm_3'm_5'm_7',\quad \zeta_{C_2}=\zeta_{c_3}=m_1m_2m_5'm_6m_7,$
$\zeta_{C_3}=\zeta_{c_4}=m_1'm_4'm_5',\quad \zeta_{C_4}=\zeta_{c_1}\vee\zeta_{c_5}=m_2m_4m_5m_7+m_3m_6'm_7,$
$I_{0.6}=1.9111.$

When threshold $\alpha=1$, there are five clusters:

$$C_1=\{c_1\},C_2=\{c_2\},C_3=\{c_3\},C_4=\{c_4\},C_5=\{c_5\}.$$

This clustering result is invalid.

In virtue of the above facts, we know that $I_{0.6}$ is the smallest of I_α for all $\alpha\in\{q_{ij}\mid 1\le i,j\le 5\}=\{0.2,0.4,0.6,1.0\}$. Thus the best clustering result is $C_1=\{c_2\}$, $C_2=\{c_3\},C_3=\{c_4\},C_4=\{c_1,c_5\}$. The clustering results are the same as the result given in [56]. Also, we can obtain the description of each cluster as follows:

The fuzzy description ζ_{C_1} of Cluster C_1 is $m_1m_2'm_3'm_5'm_7'$ with the following interpretation: "*Core ability is strong but Organization management, Pricing, Finance and Information technology are not strong* ".

The fuzzy description ζ_{C_2} of Cluster C_2, $m_1m_2m_5'm_6m_7$, states: "*Core ability, Organization management, Different advantage and Information technology are strong but Finance is not strong* ".

The fuzzy description ζ_{C_3} of Cluster C_3, $m_1'm_4'm_5'$, reads as follows: "*Core ability, Competitive forces and Finance are not strong* ".

The fuzzy description ζ_{C_4} of Cluster C_4 is $m_2m_4m_5m_7+m_3m_6'm_7$ with the sematic interpretation: "*Organization management, Competitive force, Finance and Information technology are strong* "*or* "*Pricing and Information technology are strong but Different advantage is not strong*".

9.2.3.2 Evaluation Results of the 30 Companies via the Improved Algorithm

In what follows, we apply the proposed clustering algorithm to the data of the 30 companies [56] shown as Table 7.4 in the Appendix A in Chapter 7. Let $X=\{c_1,c_2,...,c_{30}\},M=\{m_1,m_1',...,m_7,m_7'\}$, $\vartheta=m_1+m_1'+...+m_7+m_7'$.

STEP 1. Let $\varepsilon=0$. We just show c_1, c_{13} as an example:

$B^0_{c_1}=\{m_4,m_5\}$, $\zeta_{c1}=m_4m_5$.
$B^0_{c_{13}}=\{m_1',m_6'\}$, $\Lambda^0_{c_{13}}=\{m_1',m_6'\}$. By considering method (9.14), we obtain

$$\sum_{c_j\in X,j\neq 13}\mu_{m_1'}(c_j)=19.4333,\qquad \sum_{c_j\in X,j\neq 2}\mu_{m_6'}(c_j)=18.600.$$

Therefore $\zeta_{c_{13}}=m_6'$ is selected as the best fuzzy description of the object c_{13}.

STEP 2. The fuzzy relation matrix $(F^4)^2=F^4$ and $Q=F^4$.

STEP 3-4. When the value of the threshold $\alpha = 0.5667$, there are two clusters:

$C_1 = \{c_2, c_3, c_6, c_8, c_9, c_{10}, c_{11}, c_{13}, c_{15}, c_{17}, c_{19}, c_{20}, c_{24}, c_{27}, c_{28}, c_{29}\}$,
$C_2 = \{c_1, c_4, c_5, c_7, c_{12}, c_{14}, c_{16}, c_{18}, c_{21}, c_{22}, c_{23}, c_{25}, c_{26}, c_{30}\}$.

The description ζ_{C_i} of each cluster:

$\zeta_{C_1} = m_2 + m_3 + m_3' + m_4'm_5' + m_5 + m_6 + m_6' + m_7$, $\zeta_{C_2} = m_4$.
$I_{0.5667} = 2.1425$.

When threshold $\alpha = 0.6667$, there are three clusters:

$C_1 = \{c_7\}$, $C_2 = \{c_2, c_3, c_6, c_8, c_9, c_{10}, c_{11}, c_{13}, c_{15}, c_{17}, c_{19}, c_{20}, c_{24}, c_{27}, c_{28}, c_{29}\}$,
$C_3 = \{c_1, c_4, c_5, c_{12}, c_{14}, c_{16}, c_{18}, c_{21}, c_{22}, c_{23}, c_{25}, c_{26}, c_{30}\}$.

The description ζ_{C_i} of each cluster :

$\zeta_{C_1} = m_3'm_4m_5'$, $\zeta_{C_2} = m_2 + m_3 + m_3' + m_4'm_5' + m_5 + m_6 + m_6' + m_7$,
$\zeta_{C_3} = m_4$. $I_{0.6667} = 1.5999$.

When threshold $\alpha = 0.7333$, there are five clusters:

$C_1 = \{c_7\}$, $C_2 = \{c_{18}\}$, $C_3 = \{c_9, c_{15}, c_{27}\}$,
$C_4 = \{c_2, c_3, c_6, c_8, c_{10}, c_{11}, c_{13}, c_{17}, c_{19}, c_{20}, c_{24}, c_{28}, c_{29}\}$,
$C_5 = \{c_1, c_4, c_5, c_{12}, c_{14}, c_{16}, c_{21}, c_{22}, c_{23}, c_{25}, c_{26}, c_{30}\}$.

The description ζ_{C_i} of each cluster :

$\zeta_{C_1} = m_3'm_4m_5'$, $\zeta_{C_2} = m_1'm_4m_6'$, $\zeta_{C_3} = m_3 + m_6$,
$\zeta_{C_4} = m_2 + m_3' + m_4'm_5' + m_5 + m_6' + m_7$, $\zeta_{C_5} = m_4$.
$I_{0.7333} = 1.6611$.

When threshold $\alpha = 0.7667$, we obtain six clusters:

$C_1 = \{c_7\}$, $C_2 = \{c_{15}\}$, $C_3 = \{c_{18}\}$, $C_4 = \{c_9, c_{27}\}$,
$C_5 = \{c_2, c_3, c_6, c_8, c_{10}, c_{11}, c_{13}, c_{17}, c_{19}, c_{20}, c_{24}, c_{28}, c_{29}\}$,
$C_6 = \{c_1, c_4, c_5, c_{12}, c_{14}, c_{16}, c_{21}, c_{22}, c_{23}, c_{25}, c_{26}, c_{30}\}$.

The description ζ_{C_i} of each cluster is the following:

$\zeta_{C_1} = m_3'm_4m_5'$, $\zeta_{C_2} = m_6$, $\zeta_{C_3} = m_1'm_4m_6'$, $\zeta_{C_4} = m_3$,
$\zeta_{C_5} = m_2 + m_3' + m_4'm_5' + m_5 + m_6' + m_7$, $\zeta_{C_6} = m_4$.
$I_{0.7667} = 1.5709$.

Remark 9.6. For the above results, the clustering result obtained under $\alpha = 0.7667$ is different from that under $\alpha = 0.7333$ by just one object c_{15}. From the original data in [56], we can observe that the data of c_{15} is quite dissimilar to those of others. Thus the difference of the clustering for c_{15} may cause great differences in the values of the invalidity index I_α.

All clustering results and associated values of I_α for the threshold $\alpha \in \{q_{ij} \mid 1 \le i, j \le 30\}$ are shown in Table 9.11. The smallest value of I_α

Table 9.11 The clustering results of the evaluation results of 30 companies

α	0.5667	0.6667	0.7333	0.7667	0.8	0.8333	0.8667	0.9	0.9333	0.9667
Num.of Cluster	2	3	5	6*	10	11	16	17	19	21
I_α	2.1425	1.5999	1.6611	1.5709	1.7994	1.7195	1.8078	1.7500	1.7337	1.7303

is 1.5709. Therefore, the "best" number of fuzzy clusters is six. Here each cluster can be described as follows.

- *The description of Cluster 1:* The fuzzy description of cluster C_1, $\zeta_{C_1} = m_3'm_4m_5'$ states: "*Competitive force is strong but Pricing and Finance are not strong*". It is consistent with the experts' intuitive description of Group 4 shown in [56] as follows: "The character of this group is that there is not a particular strategy of operation. The managers think that all of the strategic items are equally important. The financial performance of this group is not successful in the airfreight forwarder industry".
- *The description of Cluster 2:* The fuzzy description of Cluster C_2, $\zeta_{C_2} = m_6$ states: "*Different advantage is strong* ". It is consistent with the experts' intuitive description of Group 5 presented in [56] as follows: "Differential advantage is the principal operating strategy for this group. The main differential strategy is risk reduction for customers. Based on the research of Global Trade the safety of cargo is the minimum requirement of customers. This group is focused on cargo tracking and security of consignment. Using this strategy, a young company will make the profit ny keep growing".
- *The description of Cluster 3:* The fuzzy description of Cluster C_3, $\zeta_{C_3} = m_1'm_4m_6'$ states: "*Core ability and Different advantage are not strong but Competitive forces is strong* ".
- *The description of Cluster 4:* The fuzzy description of Cluster C_4, $\zeta_{C_4} = m_3$ states: "*Pricing is strong* ". It is consistent with the experts' intuitive description of Group 3 shown in [56]: "In this group, pricing is the main strategy. The strategies of "Competitive weapon"and "Innovation/ development"are not important for them. Financial performance in this group is not perfect. Thus, we can suggest that the strategy of price war is not a good policy in the Taiwan's airfreight market."
- *The description of Cluster 5:* The fuzzy description of Cluster C_5, $\zeta_{C_5} = m_2 + m_3' + m_4'm_5' + m_5 + m_6' + m_7$ states: "*Competitive forces and Finance are not strong "or "Different advantage is not strong "or "Pricing is not strong "or "Organization management is strong "or "Finance is strong "or "Information technology is strong* ". It is consistent with the experts' intuitive description of Group 1 presented in [56]: "Compared with other groups, the objects in group 1 pay more attention to each strategic item compared to the others. Especially, the strategies of organization management and information technology are the most important for them. Regarding the financial performance of group 1, we get the trend of profit is on the downside. To judge the reason, we find that the main line of the objects of group 1 is Japan and Korea. After Asia finance storm, the import

cargo quantity of Japan and Korea area is withered. And it causes the financial performance inferior to other groups".

- *The description of Cluster 6*: The fuzzy description of Cluster C_6, $\zeta_{C_6} = m_4$ states: "*Competitive forces is strong*". It means that compared with other factors, "Competitive forces "is the strongest. According to the companies of this cluster in the original data, the factor "Core ability","Differential advantagement"are also strong. This result does not conflict with the experts' intuitive description of Group 2 as shown in [56]: "Many objects belong to group 2. We can judge that the operation strategy of airfreight forwarders in Taiwan is learning mutually from the phenomenon. In this group, the main strategies are core ability and differential advantage. Pricing and information technology are less important for them".

Remark 9.7. The expert's intuition groups considered the respectable degree of airfreight forwarder which could not be considered in our current study. This may be the reason that the clustering result C_3 does not fully coincide with experts' intuition.

As the examples showed, the results of the improved algorithm are almost consistent with the experts' assessment. Compared with the Liang's algorithm, our algorithm is more transparent, understandable, and interpretable. It can be applied to the data with the mixed features and linguistic descriptions. The results obtained so far indicate that the proposed fuzzy clustering method are practical and useful.

9.3 Feature Selection, Concept Categorization and Characteristic Description via AFS Theory

The AFS theory will be applied to study some new techniques of feature selection, concept categorization and characteristic description; those problems are often encounter in machine learning, pattern recognition and data mining. These techniques developed under the framework of AFS theory are simpler and more interpretable than those supplied by the conventional methods. In order to evaluate the effectiveness of the feature selection, the concept categorization and the characteristic description, these new techniques are applied to fuzzy clustering. Several benchmark data sets are used for this purpose. Accuracy of clustering is comparable with or superior to the results produced by the conventional algorithms such as FCM, *k*-means, and some newer algorithms such as e.g., two-level SOM-based clustering etc.

In this section, for an AFS structure (M,τ,X) of some data, we always select the weight function $\rho : X \to [0,1]$, $\rho(x) = 1$ for any $x \in X$. Then the set of coherence membership functions $\{\mu_\xi(x) \mid \xi \in EM\}$ can be obtained by (5.24) presented in Proposition 5.7, which is defined as follows.

$$\mu_\eta(x) = \sup_{i\in I} \frac{|A_i^\tau(x)|}{|X|}, \tag{9.16}$$

Here $A_i^\tau(x)$ is calculated by (4.27). If $\sigma = 2^X$, for $W \in 2^X$, $\mathscr{M}(W) = |W|$ ($|W|$ is the cardinal number of the set W, i.e., the number of elements in W) in Proposition 5.7.

9.3.1 Feature Selection

In this section, we present a technique to select salient features by the similarities among features and the entropy of the features in the framework of AFS theory.

9.3.1.1 Similarity and Entropy of Features

Let $X = \{x_1, x_2, \cdots, x_n\}$ be a set of samples and $F = \{f_1, f_2, \cdots, f_s\}$ be a set of the features on X, $x_i = (w_{i1}, w_{i2}, \cdots, w_{is}) \in R^s$, $1 \le i \le n$, where $w_{ij} = f_j(x_i)$ is the value of x_i on the feature $f_j \in F$. (F, τ_F, X) is an AFS structure in which τ_F is defined as follows: for any x_i, x_j in X,

$$\tau_F(x_i, x_j) = \{f | f(x_i) \ge f(x_j), f \in F\}. \tag{9.17}$$

Definition 9.2. Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of samples, $F = \{f_1, f_2, \cdots, f_s\}$ be a set of features. Let (F, τ_F, X) be an AFS structures defined as (9.17). For $\alpha, \beta \in F$, the *similarity between the features* α, β is defined as follows:

$$SI(\alpha, \beta) = \frac{\sum_{x \in X} \mu_{\alpha \wedge \beta}(x)}{\sum_{x \in X} \mu_{\alpha \vee \beta}(x)} \tag{9.18}$$

where the membership functions $\mu_{\alpha \wedge \beta}(x)$ and $\mu_{\alpha \vee \beta}(x)$ are given by (9.16).

The similarity defined by (9.18) shows that the fuzzy similarity degree between the features $\alpha, \beta \in F$, is determined by $\mu_{\alpha \wedge \beta}(x)$ and $\mu_{\alpha \vee \beta}(x)$. The larger $SI(\alpha, \beta)$ is, the higher similarity degree between the features α and β is.

Definition 9.3. Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of samples, F be a set of features. The entropy $E(f_j)$ of a feature $f_j \in F$ is defined as follows:

$$E(f_j) = -\sum_{k=1}^{n} p_j(x_k) log_2 p_j(x_k), \tag{9.19}$$

where

$$p_j(x_i) = \frac{\sum_{x \in \{f_j\}^{\tau}(x_i)} f_j(x)}{\sum_{k=1}^{n} f_j(x_k)} \in [0, 1],$$
$$\{f_j\}^{\tau}(x_i) = \{x \in X | f_j(x_i) \ge f_j(x)\}.$$

Entropy characterizes the (im)purity of an arbitrary collection of examples. Here, we just take into account entropy of a single feature. The selected features are the ones coming with small values of entropies.

9.3.1.2 Selecting Features

The feature selection strategy we propose for selecting the salient ones involves two stages: First, a fuzzy equivalence similarity matrix is established according to the similarity provided by Definition 9.2. Next the features are clustered assuming a certain value of threshold p. Second, the clusters in which the features have large similarities will be maintained, and the cluster which has a single feature and the feature exhibits large entropy will be discarded. The optimal clustering of the features, i.e., the optimal threshold p, is determined by the feature selecting validation index (9.20) based on similarity and entropy of the features. The detailed scheme of the feature selection is outlined as follows:

a. Establish a fuzzy equivalence similarity matrix G of features:
$H = (y_{ij})_{s\times s}$, where $y_{ij} = SI(f_i, f_j)$, $f_i, f_j \in F$. There exists an integer r such that $(H^r)^2 = H^r$ and $G = H^r$ is a fuzzy equivalence similarity matrix. $H^r = \overbrace{H \cdot H \cdot \ldots \cdot H}^{r}$ is the fuzzy matrix product of the r fuzzy matrix H (refer to Theorem 9.2), i.e., let $H^2 = (o_{ij}), o_{ij} = \max\limits_{1\le k\le s}\{\min\{y_{ik}, y_{kj}\}\}$.

b. Determine the initial clusters $U_1^{\alpha}, U_2^{\alpha}, \ldots, U_d^{\alpha}$:
The fuzzy equivalence matrix $G = H^r = (q_{ij})$ can yield a partition tree with equivalence classes. If $q_{ij} \ge \alpha$, then f_i, f_j are in the same cluster under the threshold $\alpha \in [0,1]$. The cluster which just has a single element will be discarded, then we can obtain the clusters $U_1^{\alpha}, U_2^{\alpha}, \ldots, U_d^{\alpha}$. Thus, each cluster is a group of features with relatively large level of similarity.

c. Determine optimal clusters $U_1^p, U_2^p, \ldots, U_d^p$ by index $\mathscr{V}(p)$:
Let $P = \{p_1, p_2, \ldots, p_l\} = \{q_{ij} |\ 1 \le i \le s, 1 \le j \le s\}$ be the set of all the entries in the fuzzy matrix G and $p_1 < p_2 < \ldots < p_l$. The best clustering result can be obtained by searching for the largest *feature selecting validation index* $\mathscr{V}(p), p \in P$ defined as follows:

$$\mathscr{V}(p) = \mathscr{L}(\triangle_p)\frac{\sum_{f_j\in B} E(f_j)}{|B|}, \tag{9.20}$$

where $B = F - \bigcup_{1\le i\le d} U_i^p$, and

$$\mathscr{L}(\triangle_p) = \begin{cases} 1, & if\ \ \triangle_p \le \lceil s/10\rceil, \\ \frac{s+1-\triangle_p}{s+1-\lceil s/10\rceil}, & otherwise, \end{cases} \tag{9.21}$$

and

$$\triangle_p = \max_{1\le j\le d}\{|U_j^p|\} - \min_{1\le j\le d}\{|U_j^p|\}\ \ (\triangle_p = |U_1^p|\ \ if\ \ d = 1).$$

$E(f_j)$ is the entropy of the feature f_j defined by formula (9.19). Each feature in $B = F - \bigcup_{1\le i\le d} U_i^p$ becomes a cluster by itself, i.e., it has too low similarity with other features to be clustered with other features into one group. $\frac{\sum_{f_j\in B} E(f_j)}{|B|}$ is the average entropy of the features in B. A larger value of the average entropy implies that the features in B contain the less valuable information about the data. Thus the larger value of the average entropy, the more valuable features in

$\bigcup_{1\leq i\leq d} U_i^p$. On the other hand, $\mathscr{L}(\triangle_p)$ in (9.21) normalizes $\triangle_p$ into [0,1]. The smaller $\triangle_p$ is, the larger $\mathscr{L}(\triangle_p)$ is. Intuitively, smaller value of $\triangle_p$ implies that the features in each cluster have large similarities, and the numbers of features in different clusters are evenly distributed. In summary, the larger the $\mathscr{V}(p)$, the better the features are clustered under threshold p. If $p_0 = \arg\max_{p\in P}\{\mathscr{V}(p)\}$, the final selected features set is

$$F^{select} = \bigcup_{1\leq i\leq d} U_i^{p_0}. \tag{9.22}$$

F^{select} is the set of the selected features which are considered as the salient features for ensuing learning. The experimental study of the fuzzy clustering on benchmark data sets will demonstrate and quantify its effectiveness.

This feature selection procedure is performed in unsupervised mode, i.e., any prior information about data is unknown. It is a challenging task yet a practical tool for exploratory data analysis. Compared with other algorithms such as MFCMS [65] and EM algorithm [59], our feature selection can be directly applied to the data. However, for MFCMS, its final membership degrees, weights and prototypes depend upon initialization, and it may return different results for different trials; For the EM algorithm, data set was first randomly divided into two parts, that is a training set and a testing one.

9.3.2 Principal Concept Selection and Concept Categorization

A critical issue in fuzzy pattern analysis is to select the relative concepts to describe the objects. Instead of using all available concepts in the data, one selects a subset of concepts to describe the characteristics of the system under consideration. In general, for different aims, different experiences, different perception, different knowledge background, etc., different people may select different concepts to study the system. Therefore, before further studies, humans often select the important (or relative) simple concepts from the predetermined concept set. In the following study, a principal concept selection method is proposed to achieve this goal.

Furthermore, the concepts which exhibit strong correlations are often placed in a single category. For instance, height and weight, hair black and hair melanin show strong correlations, i.e., in general, the higher the person is, the heavier the person is, and the more the hair melanin, the blacker the hair color. So in practice, height and weight are located in a single category associated with the appearance of human. Hair melanin and hair black are the same category being associated with the hair of human. Thus, we will present the simple concept categorization and show how to describe different characteristics of objects with different categories below.

9.3.2.1 Principal Concepts Selection by PCA

PCA (Principal Component Analysis)[17] constructs eigenvectors and eigenvalues from a covariance matrix constructed from the input data. The first orthogonal dimension of this eigenspace captures the greatest amount of variance in the database whilst the last dimension captures the least amount of variance. This allows for a suitable dimension of eigenspace to be chosen. Then the data are projected into the eigenspace, thus creating a lower dimensional space. The simple concept set M_0 will be selected from the total simple concepts M by using PCA.

Let $X = \{x_1, x_2, \cdots, x_n\}$ be a data set with r features and $M = \{m_{j,k} \mid 1 \leq j \leq r, 1 \leq k \leq l\}$ be a set of simple concepts on X. Here m_{jk} be the kth simple concept on the jth feature and there are l simple concepts on each feature. For any $x, y \in X$, τ is defined as (4.26) according to the given data and the semantics of the simple concepts in M. Then (M, τ, X) is an AFS structure. For each simple concept $m \in M$, its membership function $\mu_m(x)$ is defined as (9.16). Let the matrix $Y = (y_{u,v})_{n \times lr}$, where

$$y_{u,v} = \mu_{m_{j,k}}(x_i), u = i, v = l(j-1)+k, \ \ 1 \leq i \leq n, \ \ 1 \leq j \leq r, \ \ 1 \leq k \leq l \quad (9.23)$$

Let the correlation matrix be denoted as $C = (Y - \overline{Y})^T (Y - \overline{Y})$, where

$$\overline{Y} = J_{n \times lr} diag(\frac{1}{n} \sum_{u=1}^{n} y_{u,1}, \frac{1}{n} \sum_{u=1}^{n} y_{u,2}, \ldots, \frac{1}{n} \sum_{u=1}^{n} y_{u,lr}),$$

where $J_{n \times lr}$ is the n rows and lr columns matrix whose elements are all 1 (i.e., the universal matrix). By SVD decomposition, we can obtained the eigenvalues of C: $\lambda_1, \lambda_2, \ldots, \lambda_{lr}$ ($\lambda_1 > \lambda_2 > \ldots > \lambda_{lr}$), and the corresponding eigenvectors of C: $v_1, v_2, \ldots, v_{lr}$, $v_i = (a^i_{1,1}, \ldots, a^i_{1,l}, \ldots, a^i_{j,k}, \ldots, a^i_{r,1}, \ldots, a^i_{r,l})^T$. Usually, the normalization constraint $v_i^T v_i = 1$ on v_i has been adopted. Notice that $a^i_{j,k}$ associates with the simple concept $m_{j,k}$, and $a^i_{j,k}$ indicates the contribution of $m_{j,k}$ in the ith axe v_i. The larger $|a^i_{j,k}|$ is, the larger the projection of $m_{j,k}$ on the axe v_i is, i.e., the more principal contribution of $m_{j,k}$ for v_i is.

The information retention ratio e_i [33] is defined as:

$$e_i = \frac{\lambda_i}{\lambda_1}. \quad (9.24)$$

Here, we just take some principal trends into consideration, and analysis the effects (proportions) of each simple concept on such principal trends, finally the concepts whose effects are weak will be discarded. In detail, if

$$e_i > \delta, \quad (9.25)$$

the v_i will be retained. Given C, $\{v_i | i = 1, \ldots, h\}$ are retained. $a^i_{1,1}$, $a^i_{1,2}, \ldots$, $a^i_{1,l}$ $\ldots$, $a^i_{j,k}$, $\ldots$, $a^i_{r,1}$, $a^i_{r,2}, \ldots$, $a^i_{r,l}$ can be considered as the projections of lr simple concepts on v_i, and if $(a^i_{j,k})^2$ is larger than $1/(4 \times lr)$, then $m_{j,k}$ can be considered as an important simple concept for partitioning samples on the ith trend.

$$M_0 = \{m_{j,k} \mid (a^i_{j,k})^2 > 1/(4 \times lr), 1 \leq i \leq h\} \quad (9.26)$$

is the final *set of principal simple concepts* for next stage.

Notice that we did not select the new (extracted) concepts created by PCA that are linear combinations of the initial simple concepts which do not have clear interpretation, but select the initial principal concepts $m_{j,k}$ from M which retain the original semantics. Thus the descriptions of the characteristic of the systems or objects with the principal concepts exhibit sound semantics.

9.3.2.2 Concept Categorization

We introduce the concept of correlation between two concepts.

Definition 9.4. Let $X = \{x_1, x_2, \cdots, x_n\}$ be a data set with r features and $M = \{m_{j,k} \mid 1 \leq j \leq r, 1 \leq k \leq l\}$ be a set of simple concepts on X. Here m_{jk} be the kth simple concept on the jth feature and there are l simple concepts on each feature. Let (M, τ, X) be an AFS structure. For $\alpha, \beta \in M$, the *correlation between the simple concepts* α, β is calculated as follows:

$$Co(\alpha,\beta) = \frac{\sum_{i=1}^{n} \mu_\alpha(x_i)\mu_\beta(x_i)}{\sqrt{\sum_{i=1}^{n} \mu_\alpha^2(x_i)}\sqrt{\sum_{i=1}^{n} \mu_\beta^2(x_i)}}, \tag{9.27}$$

where the membership functions $\mu_\alpha(x)$ and $\mu_\beta(x)$ are given by (9.16).

Example 9.8. Let $X = \{x_1, x_2, ..., x_{10}\}$ be a set of 10 persons and their features (attributes) which are described by real numbers (age, height, weight), Boolean values (gender), nominal values (melanin) and the order relations (hair black, hair white); see Table 9.12. Where H : *High*, i.e., the content of hair melanin is high, M : *Medium*, L : *Low*. Here the number i in the "*hair black or hair white*" columns which corresponds to some $x \in X$ implies that the hair color of x has been ordered following our perception of the color ith. For example, the numbers in the column *"hair black"* imply a certain order (>)

$$x_7 > x_{10} > x_4 = x_8 > x_2 = x_9 > x_5 > x_6 = x_3 = x_1$$

When moving from the right to the left, the relationship states how strongly the hair color resembles black color. In this order, $x_i > x_j$ (e.g., $x_7 > x_{10}$) states that the hair of x_i is closer to the black color than the color of hair the individual x_j. The relationship $x_i = x_j$ (e.g., $x_4 = x_8$) means that the hair of x_i looks as black as the one of x_j. Let $M = \{m_1, m_2, ..., m_{11}\}$ be the set of fuzzy or Boolean concepts on X and each $m \in M$ associate with a single feature. Where m_1 : *"old persons"*, m_2 : *"tall persons"*, m_3 : *"heavy persons"*, m_4 : *"more hair melanin"*, m_5 :*"male"*, m_6 : *"female"*, m_7 : *"black hair persons"*, m_8 : *"white hair persons"*, m_9 : *"yellow hair persons"*, m_{10} : "*young persons*", m_{11} : "*the persons about 40 years old*".

In Example 9.8, given (9.27), we obtain

$$Co(m_2, m_3) = 0.9912, \;\; Co(m_4, m_7) = 0.9924,$$
$$Co(m_2, m_7) = 0.7887, \;\; Co(m_1, m_5) = 0.7293,$$

Table 9.12 Descriptions of features

Sample	Age	Appearance		Gender		Hair		
		height	weight	male	female	melanin	black	white
x_1	20	1.9	90	1	0	L	6	1
x_2	13	1.2	32	0	1	M	4	3
x_3	50	1.7	67	0	1	L	6	1
x_4	80	1.8	73	1	0	M	3	4
x_5	34	1.4	54	1	0	L	5	2
x_6	37	1.6	65	0	1	L	6	1
x_7	45	1.7	78	1	0	H	1	6
x_8	70	1.65	70	1	0	M	3	4
x_9	60	1.82	83	0	1	M	4	3
x_{10}	3	1.1	21	0	1	H	2	5

$Co(m_2,m_3) = 0.9912$ and $Co(m_4,m_7) = 0.9924$ implies "*height*" has strong correlation with "*weight*" and implies "*hair melanin*" has strong correlation with "*hair black*". While $Co(m_2,m_7) = 0.7887$ implies "*height*" has low correlation with "*hair black*", $Co(m_1,m_5) = 0.7293$ implies "*age*" has low correlation with "*gender*".

Based on the correlations between the concepts, concepts categorization method is summarized as follows:

Let $X = \{x_1,x_2,\ldots,x_n\}$ be a set of objects, $M = \{m_1,m_2,\ldots,m_z\}$ be a set of simple concepts.

a. Establish a fuzzy equivalence correlation matrix B on the set of simple concepts M:
$M_{Co} = (l_{ij})_{z\times z}$, where $l_{ij} = Co(m_i,m_j)$, m_i, m_j are simple concepts in M. There exists an integer t such that $(M^t_{Co})^2 = M^t_{Co}$, $B = M^t_{Co} = (b_{ij})_{z\times z}$ is a fuzzy equivalence correlation matrix.

b. Determine the initial clusters $M^\alpha_1, M^\alpha_2, \ldots, M^\alpha_d$:
The fuzzy equivalence correlation matrix $B = M^t_{Co}$ can yield a partition tree with equivalence classes. If $q_{ij} \geq \alpha$, then n_i, n_j are in the same cluster assuming a certain level of the threshold α. The cluster which just has a single element will be discarded, then the clusters $M^\alpha_1, M^\alpha_2, \ldots, M^\alpha_d$ can be obtained.

c. Determine the final clusters $M_1, M_2, \ldots, M_t$:
Let $\{\alpha_1, \alpha_2, ..., \alpha_v\} = \{b_{ij} | \ 1 \leq i, j \leq z\}$ be the set of all the entries in the fuzzy equivalence correlation matrix $B = (b_{ij})_{z\times z}$ and $\alpha_1 < \alpha_2 < ... < \alpha_v$. For u from 1 to v, check $M^{\alpha_u}_1, M^{\alpha_u}_2, \ldots, M^{\alpha_u}_t$ ($|M^{\alpha_u}_i| > 1, t \geq 2$), once the following condition is satisfied

$$\max_{1\leq i\leq t}\{|M^{\alpha_u}_i|\} \leq 2 \min_{1\leq i\leq t}\{|M^{\alpha_u}_i|\}, \tag{9.28}$$

then we let $M_i = M^{\alpha_u}_i$ $(i = 1,2,\ldots,t)$.

The concepts in each M_i possess strong correlations, so M_i is regarded as a *simple concept category* to describe some characteristics of the objects in X. In the above algorithm, M_i is just a set of fuzzy concepts which is selected to describe a particular characteristic of the objects in X. In practice, such cluster can be regarded as a category and a particular characteristic of objects is usually described by the fuzzy concepts in a category.

9.3.2.3 Concept Selection for the Description of a Sample via $1/k-A$-Nearest Neighbors

In human clustering and recognition procedures shown in Figure 9.1 and following the elementary fuzzy clustering method, humans first choose some features or some simple concepts formed on the features which they regard important to his clustering or recognition problems. These are the following procedures: feature selection, concept selection, and concept categorization as we discussed in the previous sections. Then for each object or sample, they choose some concepts from the above selected set of concepts to describe the sample. In what follows, we discuss how to select some simple concepts for the description of an object. That is, for the sample x, by the $1/k-A$-nearest neighbors of x (defined by Definition 9.6), a special set Λ_x of simple concepts is selected to form the fuzzy description of x.

In Chapter 4, the weight function ρ_ν of a simple concept ν (seeing Definition 4.8) associating to a feature, which can be defined according to the data and the semantic meaning of ν, are applied to derive the coherence membership functions of an AFS structure and an AFS fuzzy logic system. In this section, the weight functions of simple concepts are applied to define the distances associating to a set of simple concepts.

Let $X=\{x_1,x_2,\cdots,x_n\}$ be a set of some samples and $F=\{f_1,f_2,\cdots,f_r\}$ be the set of the features on X, $x_i=(w_{i1},w_{i2},\cdots,w_{ir})$, $1\leq i\leq r$, where $w_{ij}=f_j(x_i)$ is the value of x_i on the feature f_j. For the sake of simplicity, let $M=\{m_{ij} \mid 1\leq i\leq r,\ 1\leq j\leq 4\}$ be a set of simple concepts on X. Where $m_{i1},m_{i2},m_{i3},m_{i4}$ are the simple concepts, *"small" , "medium", "no-medium", "large"* associating to the feature f_i respectively. We discuss the weight functions of simple concepts with the following semantics:

$$\rho_{m_{j1}}(x_i) = \frac{h_{j1}-f_j(x_i)}{h_{j1}-h_{j2}}, \tag{9.29}$$

$$\rho_{m_{j2}}(x_i) = \frac{h_{j4}-|f_j(x_i)-h_{j3}|}{h_{j4}-h_{j5}}, \tag{9.30}$$

$$\rho_{m_{j3}}(x_i) = \frac{|f_j(x_i)-h_{j3}|-h_{j5}}{h_{j4}-h_{j5}}, \tag{9.31}$$

$$\rho_{m_{j4}}(x_i) = \frac{f_j(x_i)-h_{j2}}{h_{j1}-h_{j2}}, \tag{9.32}$$

with the semantics of terms of "*small*, "*medium*", "*not medium*", and "*large*", respectively. Here $j=1,2,...,s$,

$$\begin{aligned} h_{j1} &= \max\{f_j(x_1), f_j(x_2), \cdots, f_j(x_n)\}, \\ h_{j2} &= \min\{f_j(x_1), f_j(x_2), \cdots, f_j(x_n)\}, \\ h_{j3} &= \frac{f_j(x_1) + f_j(x_2) + \cdots + f_j(x_n)}{n}, \\ h_{j4} &= \max\{|f_j(x_k) - h_{j3}| \mid k = 1, 2, \cdots, n\}, \\ h_{j5} &= \min\{|f_j(x_k) - h_{j3}| \mid k = 1, 2, \cdots, n\}. \end{aligned}$$

By Definition 4.8, ones can verify that for every $m_{ij} \in M$, $\rho_{m_{ij}}(x)$ defined by (9.29)–(9.32) is a weighting function of the simple concept m_{ij}. For example, let X be a set of ten people and their ages are 19, 8, 65, 80, 20, 23, 59, 70, 68, and 6, respectively. Let $m_{11}, m_{12}, m_{13}, m_{14}$ be the simple concept on "*age* "and their semantic meanings are *"age is small ", "age is medium ", "age is not-medium ", "age is large, "* respectively. We have

$h_{11} = \max\{19, 8, 65, 80, 20, 23, 59, 70, 68, 6\} = 80,$
$h_{12} = \min\{19, 8, 65, 80, 20, 23, 59, 70, 68, 6\} = 6,$
$h_{13} = (19+8+65+80+20+23+59+70+68+6)/10 = 41.8,$
$h_{14} = \max\{|19-41.8|, |8-41.8|, |65-41.8|, |80-41.8|, |20-41.8|, |23-41.8|, |59-41.8|, |70-41.8|, |68-41.8|, |6-41.8|\} = 38.2,$
$h_{15} = \min\{|19-41.8|, |8-41.8|, |65-41.8|, |80-41.8|, |20-41.8|, |23-41.8|, |59-41.8|, |70-41.8|, |68-41.8|, |6-41.8|\} = 17.2.$
$\rho_{m_{11}}(x_1) = \frac{h_{11}-w_{11}}{h_{11}-h_{12}} = \frac{80-19}{80-6} = 0.82, \quad \rho_{m_{11}}(x_2) = \frac{h_{11}-w_{21}}{h_{11}-h_{12}} = \frac{80-8}{80-6} = 0.97,$
$\rho_{m_{11}}(x_3) = \frac{h_{11}-w_{31}}{h_{11}-h_{12}} = \frac{80-65}{80-6} = 0.20, \quad \rho_{m_{12}}(x_1) = \frac{h_{14}-|w_{11}-h_{13}|}{h_{14}-h_{15}} = 0.73,$
$\rho_{m_{13}}(x_1) = \frac{|w_{11}-h_{13}|-h_{15}}{h_{14}-h_{15}} = 0.27, \ \rho_{m_{14}}(x_1) = \frac{w_{11}-h_{12}}{h_{11}-h_{12}} = \frac{19-6}{80-6} = 0.18.$

Definition 9.5. Let X be a set of samples and M be the set of simple concepts on X. Let X and M be finite sets. For $A \subseteq M$, and $x, y \in X$, we define $d_A(x,y)$ the *distance of* x, y *associating to the simple concepts in* A as follows:

$$d_A(x,y) = \left(\sum_{m \in A} (\rho_m(x) - \rho_m(y))^2 \right)^{\frac{1}{2}} \tag{9.33}$$

where $\rho_m(x)$ is a weight function of m defined by Definition 4.8.

For $x \in X$, and an positive integer k, $A \subseteq M$, the following Definition 9.6 defines the $1/k - A$-nearest neighbor $D^A_{1/k-x}$ of x in order to select simple concepts for the descriptions of the sample x, i.e., the set of $|X|/k$ samples in X whose distances associating to the simple concepts in A with x are the nearest of the others $|X| - |X|/k$ samples in X.

Definition 9.6. Let X be a set of samples and M be the set of simple concepts on X. Let X and M be finite sets. For $A \subseteq M$, and $x \in X$, k is a positive integer, a subset of X is called a $1/k - A$*-nearest neighbor of* x, denoted as $D^A_{1/k-x}$, if for any $y \in D^A_{1/k-x}$ and any $z \in X - D^A_{1/k-x}$, $d_A(x,y) < d_A(x,z)$, and $|D^A_{1/k-x}|/|X| = 1/k$.

9.3.3 Characteristic Description of Samples

Let us recall the essence of the recognition process as it is typically carried out by humans. Human always form a complex fuzzy concepts from the selected (given) simple concepts, in order to describe the object (sample). The complex fuzzy concepts can be served as the fuzzy characteristic description of the object. This description can represent the characteristic of the sample, and can distinguish this sample from other objects. In what follows, we study how to describe a sample in the AFS framework. There are many alternative methods [44, 70, 71] to determine a fuzzy set $\zeta_x \in EM$ to describe each x ($x \in X$). Fuzzy set ζ_x describes the prototype of x with the simple concepts in M in order to distinguish x among other samples in X to the highest extent. However fuzzy descriptions ζ_x of sample x given in [44, 70, 71] are too particular to appropriately to represent its cluster. In this section, we present two kinds of the fuzzy descriptions of a sample based on the Occam's razor principle [64] (we always prefer the simplest hypothesis that fits the data) and the *Minimum Description Length Principle*. We describe a sample by using a minimal number of possible simple concepts. Sometimes we can describe some samples just by using one simple concept, which is the most salient characteristic of the sample.

For example, we want to find a people in the crowd with a description and there are two descriptions of the people: "*high height and large eyes and lameness*" or "*female with long hair and yellow skin*" or "*wearing a red dust coat and a white cap*", another is "*lameness*" which is the most primary characteristic of the people. It is clear that we can quickly and accurately find the people using the latter one. In what follows, we provide two methods to obtain the descriptions which will be applied to two different fuzzy clustering algorithms.

Description Method A: *Description based on principal concepts or a category of concepts*

Let X be the universe of discourse, $x \in X$, N be a set of principal concepts selected by PCA or a category of concepts described in the above section. We establish an AFS structure (N, τ, X) on a basis on the original data and facts. The membership functions of the fuzzy concepts in EN are defined by (9.16).

$$\begin{aligned} B_x &= \{m \in N \mid \mu_m(x) = \mu_{\vee_{b\in N} b}(x)\}, \\ \zeta_x &= \wedge_{\beta \in B_x} \beta \in EN, \end{aligned} \tag{9.34}$$

ζ_x is the fuzzy description of x under the concept set N, which can represent the most salient characteristics.

Description Method B: *Description based on the* $1/k-$*A-Nearest Neighbors*

Now, we describe how to select simple concepts for the fuzzy description of each sample. For $x \in X$, and $m \in M$, we calculate the $1/k-M-$nearest neighbor $D^M_{1/k-x}$ of x and the $1/k-\{m\}-$nearest neighbor $D^m_{1/k-x}$ of x. The larger $|D^m_{1/k-x} \cap D^M_{1/k-x}|$ is, the greater the representative ability of the simple concept m is. Therefore, the importance of simple concept m to describe x is determined by $|D^m_{1/k-x} \cap D^M_{1/k-x}|$. We choose $|M|/k$ simple concepts in M whose $|D^m_{1/k-x} \cap D^M_{1/k-x}|$ are the largest of the other $|M|-|M|/k$ simple concepts in M as the selected simple concepts and denote the *set of the selected* $|M|/k$ *simple concepts* as Λ_x.

For every $x \in X$, let $\vartheta = \sum_{m_i \in \Lambda_x} m_i$. Then ϑ is the maximum element of the completely distributive lattice $(E\Lambda_x, \vee, \wedge)$ (refer to Theorem 4.1). Therefore $\mu_\vartheta(x)$ is the maximum membership degree of x belonging to the concepts in $E\Lambda_x$ (i.e., $(\Lambda_x)_{EI}$) which is the set of all concepts generated by the simple concepts in Λ_x. In order to distinguish x among other samples in X at the maximal degree, we choose a fuzzy concept $\zeta_x \in E\Lambda_x$ which satisfies that not only $\mu_{\zeta_x}(x)$ approaches $\mu_\vartheta(x)$ to the highest extent, but also $\mu_{\zeta_x}(y)$ becomes as low as possible for $y \in X, x \neq y$. In other words, x can be distinguished by ζ_x among other samples in X to the highest degree. The details of this procedure are presented as follows.

Find-Fuzzy-Description-Algorithm:

1 Let $\varepsilon \geq 0$. Find set B^ε_x defined as follows:

$$B^\varepsilon_x = \{m \in \Lambda_x \mid \mu_m(x) \geq \mu_\vartheta(x) - \varepsilon\} \tag{9.35}$$

B^ε_x is the set of the simple concepts in Λ_x such that the degrees of x belonging to them are larger than or equal $\mu_\vartheta(x) - \varepsilon$. Where ε controls the "*roughness*" extent of the description of x. In our algorithm, ε is a certain predetermined constant for all samples.

2 Find the set A^ε_x defined as follows:

$$A^\varepsilon_x = \left\{ \prod_{m \in B} m \mid \mu_{\prod_{m \in B} m}(x) \geq \mu_\vartheta(x) - \varepsilon,\ B \subseteq B^\varepsilon_x \right\} \tag{9.36}$$

A^ε_x is the set of the conjunctions of the simple concepts in B^ε_x such that the degrees of x belonging to them are larger than or equal to $\mu_\vartheta(x) - \varepsilon$.

3 Choose the best description of x from $\zeta_x \in A^\varepsilon_x$ as follows:

$$\zeta_x = \arg\min_{\varphi \in A^\varepsilon_x} |\{y \in X \mid \mu_\varphi(y) \geq \mu_\varphi(x)\}|\,. \tag{9.37}$$

Thus x can be distinguished by ζ_x among other samples in X at maximal extent.

In the following sections two AFS fuzzy clustering algorithms are developed via the "**Description Method A**" and "**Description Method B**"as described above.

9.3.4 AFS Fuzzy Clustering Algorithm Based on the $1/k-A$ *Nearest Neighbors*

The AFS clustering scheme based on the $1/k-A$-nearest neighbors shown as Figure 9.2 consists of the following design phases: A. Select simple concepts for the fuzzy description of each sample; B. Find the description of every sample using

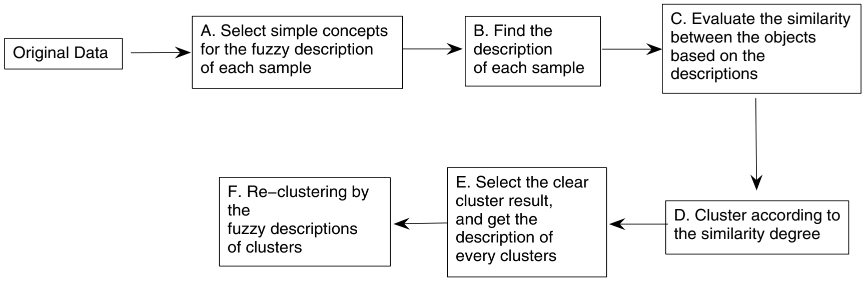


Fig. 9.2 The procedure of our clustering algorithm

the selected simple concepts; C. Evaluate the similarity between the samples based on their descriptions; D. Cluster the samples according to the similarity degrees deriving from their descriptions; E. Select the clear clustering result via the fuzzy cluster validity index defined by (9.40), and obtain the descriptions of the clusters; F. All samples are re-clustered according to the descriptions of clusters. In what follows, we introduce each phase in detail.

A. Evaluate the similarity between the samples

For every $x \in X$, the fuzzy description ζ_x of x is obtained in the above Description Method B: description based on the $1/k-A$-Nearest Neighbors. The fuzzy descriptions ζ_x of $x \in X$ are applied to establish the fuzzy relation matrix $F_\zeta = (f_{ij})$ on $X = \{x_1, x_2, ..., x_n\}$, where

$$f_{ij} = \min\left\{\mu_{\zeta_{x_i}\wedge\zeta_{x_j}}(x_i),\ \mu_{\zeta_{x_i}\wedge\zeta_{x_j}}(x_j)\right\}. \tag{9.38}$$

where f_{ij} is the similarity degree between the sample x_i and x_j determined by their descriptions. Theorem 9.2 ensures us that there exists an integer h such that $(F_\zeta^h)^2 = F_\zeta^h$, i.e., the fuzzy relation $Q = F_\zeta^h$ can yield a partition tree with equivalence classes.

B. Cluster the samples, and get the description of every cluster

Fuzzy relation matrix $Q = (q_{ij})$ is obtained by the above *Procedure A*. For the threshold $\alpha \in [0,1]$, we have the Boolean matrix $Q_\alpha = (q_{ij}^\alpha)$, $q_{ij}^\alpha = 1 \Leftrightarrow q_{ij} \geq \alpha$. x_i, x_j are clustered in the same cluster under the threshold α if and only if $q_{ij} \geq \alpha$ (i.e., $q_{ij}^\alpha = 1$). Thus, some clusters $C_1, C_2, \ldots, C_r$ are obtained for the threshold α. In general, $X \neq C_1 \cup C_2 \cup \ldots \cup C_r$, because of the following reasons: 1) if $q_{ii} < \alpha$ then x_i cannot be determined by Q under the threshold α; 2) a cluster C contains too few samples to be considered as a cluster. In this clustering algorithm, the cluster C is considered to be invalid if $|C| < |X|/30$. The cluster labels of the samples in $X - C_1 \cup C_2 \cup \ldots \cup C_r$ can be determined by the following *Procedure D. Re-cluster the samples*.

For the cluster C_i, let $\mathcal{T}_{C_i} = \{\zeta_x \mid x \in C_i\}$ (the descriptions of all samples in C_i), $i = 1, 2, ..., r$. The set of the typical descriptions selected from $\mathcal{T}_{C_i}$, denoted as $\mathcal{D}_{C_i}$, is defined as follows

$$\mathcal{D}_{C_i} = \left\{ \zeta_x \mid x \in C_i,\ \max_{u \in C_i}\{\mu_{\zeta_x}(u)\} > \max_{v \in X - C_i}\{\mu_{\zeta_x}(v)\} + 0.1 \right\}. \tag{9.39}$$

Thus, the cluster C_i is described by the fuzzy concept $\zeta_{C_i} = \bigvee_{\zeta_x \in \mathcal{D}_{C_i}} \zeta_x \in EM$.

C. Select the best clustering result

For different threshold values $\alpha \in [0,1]$, the different clustering results: C_1, C_2, $\cdots$, C_r and their fuzzy descriptions ζ_{C_1}, ζ_{C_2}, $\cdots$, ζ_{C_r} may be obtained by running the *Procedure B* discussed above. The threshold α controls level of roughness of the clustering results. We can determine the optimal fuzzy clustering result C_1, C_2, $\cdots$, C_r and their fuzzy descriptions ζ_{C_1}, ζ_{C_2}, $\cdots$, ζ_{C_r} by I_α the *fuzzy cluster validity index* defined as follows:

$$I_\alpha = \frac{1}{r|X|}\left(\sum_{x \in X}\sum_{i=1}^{r} \mu_{\zeta_{C_i}}(x) - \sum_{x \in X}\sum_{1 \le i < j \le r} \mu_{\zeta_{C_i}}(x)\mu_{\zeta_{C_j}}(x)\mu_{\zeta_{C_i} \wedge \zeta_{C_j}}(x) \right). \tag{9.40}$$

Here $\alpha \in [0,1]$, r is the number of clusters. The fuzzy cluster validity index I_α evaluates the clear and discriminate extent between the clusters obtained by the algorithm under the threshold $\alpha \in (0,1)$. The larger the value of I_α, the clearer and more discriminative the clustering result becomes.

D. Re-cluster the samples

Let the clusters C_1, C_2, $\cdots$, C_r and their fuzzy describe ζ_{C_1}, ζ_{C_2}, $\cdots$, ζ_{C_r} be the optimal clustering result determined by the above *Procedure C*. All samples in X are re-clustered to the clusters $\overline{C}_1$, $\overline{C}_2$, $\cdots$, $\overline{C}_r$ by the fuzzy descriptions ζ_{C_1}, ζ_{C_2}, $\cdots$, ζ_{C_r} as follows: for each sample $x \in X$, $x \in \overline{C}_q$, if $q = \arg\max_{1 \le i \le r}\{\mu_{\zeta_{C_i}}(x)\}$. Thus, the cluster labels of all sample in X including those samples in $X - C_1 \cup C_2 \cup \ldots \cup C_r$ can be determined.

9.3.5 AFS Fuzzy Clustering Algorithm Based on Principal Concepts and the Categories of Concepts

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of objects and $F = \{f_1, f_2, \ldots, f_s\}$ be a set of features.

- F^{select} is the set of the selected features shown as (9.22) via the feature selecting algorithm described in Section 9.3.1.2. The selecting features, $F^{select} \subseteq F$.
- $M = \{m_{j,k} \mid 1 \le j \le r,\ \ 1 \le k \le l\}$, where $m_{j,1}, m_{j,2}, ..., m_{j,l}$ are the simple concepts associating with the features f_j in F^{select}, respectively.
- M_0 is the set of principal concepts which is selected from M by PCA shown as (9.26) according to the method described in Section 9.3.2.1 Principal Concepts Selection by PCA.

- $M_i, i = 1, 2, \ldots, t$ are the concept categories on M_0 according to the method described in Section 9.3.2.2. Concept Categorization $M_i, i = 1, 2, \ldots, t$ are shown as given by (9.28).
- For $x_j \in X$, $\zeta^i_{x_j}$ is the description of x_j under the concept set (concept category) M_i shown as (9.34) according to the method described by Description Method A in Section 9.3.3 Characteristic Description of Each Sample. The final fuzzy description of x_j is expressed as:

$$\zeta_{x_j} = \zeta^k_{x_j}, if\ k = \arg\max_{1 \le i \le t} \{\mu_{\zeta^i_{x_j}}(x_j)\}. \tag{9.41}$$

Once the fuzzy characteristic description of each sample has been obtained, the set of the samples is clustered by the AFS clustering algorithm presented in Table 9.13 for each threshold $\alpha \in [0, 1]$. The best clustering result is obtained by I_α, the fuzzy cluster validation index shown as (9.4) in the elementary fuzzy clustering method.

In what follows, we offer some explanation for the procedure *Find the fuzzy description* $\zeta_{\overline{C}_i}$ *for each cluster* $\overline{C}_i \subseteq X$ in Table 9.13. $\zeta_{\overline{C}_i} = \wedge_{\gamma \in \Gamma_i} \gamma$, and the elements in Γ_i are the fuzzy descriptions of some samples in the ith clusters $\overline{C}_i$. Obviously, we do not consider the fuzzy descriptions of all the samples in the ith cluster, because in our opinion, the fuzzy descriptions of some samples are typical, and others may be not typical enough to represent its cluster or be noise. Thus, we want to choose the fuzzy descriptions of some typical samples from the ith clusters. On the other hand, we want to present the algorithm of finding the fuzzy description ζ_{C_i} based on the

Table 9.13 AFS Fuzzy Clustering Algorithm Based on Principal Concepts and the Categories of Concepts

AFS-Clustering-Algorithm($X, \alpha, \zeta_{x_i}, i = 1, 2, \ldots, n$)
    $X = \{x_1, x_2, \ldots, x_n\}$, *$\alpha$ is a threshold, ζ_{x_i} is the fuzzy description of x_i, $x_i \in X$.*

- *Establish a fuzzy equivalence relation matrix Q*
    - $R \longleftarrow (r_{ij})$, where $r_{ij} = \min\{\mu_{\zeta_{x_i} \wedge \zeta_{x_j}}(x_i), \mu_{\zeta_{x_i} \wedge \zeta_{x_j}}(x_j)\}$, $1 \le i, j \le n$
    - Find an integer t such that $(R^t)^2 = R^t$, R^t is the fuzzy equivalence relation matrix
    - $Q \longleftarrow R^t$
- *Determine the initial clusters $\overline{C}_1, \overline{C}_2, \ldots, \overline{C}_l$ for $Q = (q_{ij})$ under a given threshold $\alpha \in [0, 1]$*
    - If $q_{ij} \ge \alpha$, then x_i, x_j are in the same cluster under threshold α. The clusters $\overline{C}_1, \overline{C}_2, \ldots, \overline{C}_l$, which have more than one samples, can be obtained (i.e., the cluster with one single sample is discarded)
- *Find the fuzzy description $\zeta_{\overline{C}_i}$ for each cluster $\overline{C}_i \subseteq X$*
    - $\Gamma_i \longleftarrow \{\zeta_{x_j} \mid \frac{|\{y | y \in \overline{C}_i, \mu_{\zeta_{x_j}}(y) \ge \lambda\}|}{|\overline{C}_i|} \ge \omega, x_j \in \overline{C}_i\}$, $i = 1, \ldots, l$
    - $\zeta_{\overline{C}_i} \longleftarrow \bigwedge_{\gamma \in \Gamma_i} \gamma$, if $\Gamma_i \ne \varnothing$
- Return $C_1, C_2, \ldots, C_l$, $\zeta_{C_1}, \zeta_{C_2}, \ldots, \zeta_{C_l}$

Occam's razor principle [64] in order to describe one cluster by using the concepts as simple as possible.

The fuzzy description ζ_{x_k}, for which the proportion between the samples in cluster $\overline{C}_i$ whose degrees belonging to ζ_{x_k} are higher than λ and the samples in $\overline{C}_i$ with levels higher than ω, will be placed into Γ_i. Each description ζ_{x_j} of x_j in $\overline{C}_i$ represents one of the different characters of the cluster $\overline{C}_i$. It is obvious that some characters may be more universal, e.g., $\zeta_{x_j}(x_j \in \overline{C}_i)$, more samples in $\overline{C}_i$ belong to ζ_{x_j} at large degree, than other characters which are more particular. By this observation, the appropriate characteristics, i.e., appropriate $\zeta_{x_j}(x_j \in \overline{C}_i)$ should be selected to describe the prototype of $\overline{C}_i$. The universality and particularity of the fuzzy description of $\overline{C}_i$ can be controlled by the values of ω and λ.

Remark 9.8. In the following experimental study, we choose the same numeric values for the parameters ω and λ to be used the three benchmark data sets. The experimental study shows that the effect of varying the values of the parameters ω and λ on clustering results of *AFS Fuzzy Clustering Algorithm Based on Principal Concepts or the Categories of Concepts* is not very significant.

9.4 Experimental Evaluation of the Algorithms

We consider the application of the clustering algorithms studied so far, that is "*AFS Fuzzy Clustering Algorithm Based on the* $1/k-A$ *Nearest Neighbors*" and "*AFS Fuzzy Clustering Algorithm Based on Principal Concepts and the Categories of Concepts*" to three benchmark data sets coming from the UCI Repository of machine learning databases obtained via an anonymous ftp server (ftp://ftp.ics.uci.edu/ pub/ machine-learning-databases/).

9.4.1 Experimental Studies of AFS Fuzzy Clustering Algorithm Based on Principal Concepts and the Categories of Concepts

In order to show the efficiency of the proposed approach, we present experimental results of the applications of the feature selection, the principal concept selection, concept categorization, and the characteristic description to the AFS fuzzy clustering problem.

9.4.1.1 Wine Data

Chemical analysis of wines grown in the same region in Italy, but derived from three different cultivars, should be sufficient to recognize the source of the wine. The analysis determined 13 quantities, including alcohol content, hue, color intensity, and content of 9 chemical compounds. The data is stored in UC Irvine repository of Machine Learning [3], where more details about it may be found as well. The number of data samples coming Classes 1, 2, and 3 is 59, 71 and 48, respectively.

We illustrate the use of the feature selection, the principal concept selection, the concept categorization and the characteristic description via the AFS clustering.

Feature selection

Let $X = \{x_1, x_2, ..., x_{178}\}$ be the set of 178 samples, and $F = \{f_1, f_2, \ldots, f_{13}\}$ be the set of features on X, $x_i = (w_{i,1}, w_{i,2}, \ldots, w_{i,13})$, $i = 1, 2, \ldots, 178$, where $w_{i,1} = f_1(x_i)$ is alcohol content (AL) of x_i, $w_{i,2} = f_2(x_i)$ is malic acid content (MAC) of x_i, $w_{i,3} = f_3(x_i)$ is ash content of x_i, $w_{i,4} = f_4(x_i)$ is alcalinity of ash (AA) of x_i, $w_{i,5} = f_5(x_i)$ is magnesium content (MA) of x_i, $w_{i,6} = f_6(x_i)$ is total phenols (TP) of x_i, $w_{i,7} = f_7(x_i)$ is flavanoids (FL) of x_i, $w_{i,8} = f_8(x_i)$ is nonflavanoids phenols (NFP) of x_i, $w_{i,9} = f_9(x_i)$ is proanthocyaninsm (PR) of x_i, $w_{i,10} = f_{10}(x_i)$ is color intensity (CI) of x_i, $w_{i,11} = f_{11}(x_i)$ is hue of x_i, $w_{i,12} = f_{12}(x_i)$ is OD280/OD315 (O) of diluted wines of x_i, $w_{i,13} = f_{13}(x_i)$ is praline (P) of x_i. For each simple concept $f_i \in F$, the structure τ_F of AFS structure (F, τ_F, X) is well defined by (9.17).

When $p = 0.56$, $\mathscr{V}(0.56)$ shown in (9.20) is maximum value and the final selected sets of features shown as (9.22) are listed as follows.

$$\begin{aligned} U_1^{0.56} &= \{f_6, f_7, f_9, f_{11}, f_{12}\}, \\ U_2^{0.56} &= \{f_1, f_{10}, f_{13}\}, \\ F^{select} &= U_1^{0.56} \cup U_2^{0.56} = \{f_1, f_6, f_7, f_9, f_{10}, f_{11}, f_{12}, f_{13}\}. \end{aligned}$$

In order to demonstrate the effectiveness of the proposed Feature Selection described in Section 9.3.1.2, we conducted several experiments with the FCM [4] clustering algorithm and experimental results of three real world data sets are shown in Table 9.14. When all the features are applied, the obtained precision is 94.9%. While 8 of 13 features $f_1, f_6, f_7, f_9, f_{10}, f_{11}, f_{12}, f_{13}$ have been selected to be applied to the FCM, a little bit lower accuracy of 92.7% has been reported. However, feature selection has resulted in the lower computing cost. For Iris data and WDBC data, when all 4 features of Iris and 30 features of WDBC are used to run the FCM, we obtained accuracies of 89.3% (Iris) and 92.8% (WDBC). But for Iris data, even when only 50% features f_3, f_4 have been selected, accuracy of 96.7% are obtained. And for WDBC data, when only 46.7% features $f_1, f_3, f_4, f_7, f_8, f_{11}, f_{13}, f_{14}, f_{21}, f_{23}, f_{24}, f_{26}, f_{27}, f_{28}$ are selected, accuracies of 93.2% is reported. Thus, the results obtained using the proposed selected feature method are comparable. In fact, this shows that the feature selection method is able to effectively extract significant features for running the FCM method.

Table 9.14 Accuracy obtained by the FCM clustering for the data based on the selected features and the complete set of features

Data	Selected features		All features	
	Accuracy	Number of features	Accuracy	Number of features
Wine	92.7%	8	94.9%	13
Iris	96.7%	2	89.3%	4
WDBC	93.2%	14	92.8%	30

Principal concept selection and concept categorization

Let M be the set of simple concepts on the selected features in F^{select} (i.e., f_1, f_6, f_7, f_9, f_{10}, f_{11}, f_{12}, f_{13}) shown as follows.

$$\begin{aligned} M &= \{m_{j,k} | j = 1,6,7,9,10,11,12,13, 1 \leq k \leq 3\} \\ &= \{m_{1,1}, m_{1,2}, m_{1,3}, m_{6,1}, m_{6,2}, m_{6,3}, m_{7,1}, m_{7,2}, m_{7,3}, m_{9,1}, m_{9,2}, m_{9,3}, m_{10,1}, m_{10,2}, \\ &\quad m_{10,3}, m_{11,1}, m_{11,2}, m_{11,3}, m_{12,1}, m_{12,2}, m_{12,3}, m_{13,1}, m_{13,2}, m_{13,3}\}, \end{aligned}$$

where $m_{j,1}, m_{j,2}, m_{j,3}$ are the simple concepts, " *large*", "*small*" , "*medium*" associating with the feature $f_j \in F^{select}$, respectively. For any $x, y \in X$, τ is defined by (4.26), then (M, τ, X) is an AFS structure. Let $\sigma = 2^X$. For each simple concept $m \in M$, its membership function $\mu_m(x)$ is defined as (9.16). Let $Y = (y_{u,v})_{178\times 24}$. We have $y_{u,v}$ being defined by (9.23) as follows.

$$\begin{aligned} y_{u,1} &= \mu_{m_{1,1}}(x_u), y_{u,2} = \mu_{m_{1,2}}(x_u), y_{u,3} = \mu_{m_{1,3}}(x_u), \\ y_{u,4} &= \mu_{m_{6,1}}(x_u), y_{u,5} = \mu_{m_{6,2}}(x_u), y_{u,6} = \mu_{m_{6,3}}(x_u), \\ y_{u,7} &= \mu_{m_{7,1}}(x_u), y_{u,8} = \mu_{m_{7,2}}(x_u), y_{u,9} = \mu_{m_{7,3}}(x_u), \\ y_{u,10} &= \mu_{m_{9,1}}(x_u), y_{u,11} = \mu_{m_{9,2}}(x_u), y_{u,12} = \mu_{m_{9,3}}(x_u), \\ y_{u,13} &= \mu_{m_{10,1}}(x_u), y_{u,14} = \mu_{m_{10,2}}(x_u), y_{u,15} = \mu_{m_{10,3}}(x_u), \\ y_{u,16} &= \mu_{m_{11,1}}(x_u), y_{u,17} = \mu_{m_{11,2}}(x_u), y_{u,18} = \mu_{m_{11,3}}(x_u), \\ y_{u,19} &= \mu_{m_{12,1}}(x_u), y_{u,20} = \mu_{m_{12,2}}(x_u), y_{u,21} = \mu_{m_{12,3}}(x_u), \\ y_{u,22} &= \mu_{m_{13,1}}(x_u), y_{u,23} = \mu_{m_{13,2}}(x_u), y_{u,24} = \mu_{m_{13,3}}(x_u) \end{aligned}$$

and $u = 1, \ldots, 178$. Considering the SVD decomposition, we obtained the eigenvalues of $(Y - \overline{Y})^T (Y - \overline{Y})$, that are equal to $\lambda_1 = 0.6974, \lambda_2 = 0.4552, \lambda_3 = 0.1483, \ldots, \lambda_{23} = 0.0003, \lambda_{24} = 0.0002$. According to (9.24), we have $e_1 = 1, e_2 = 0.6527, e_3 = 0.2127, \ldots, e_{23} = 0.0004, e_{24} = 0.0003$. In this section, we always let $\delta = 0.4$ for all the data sets according to our experiences. Because $e_1, e_2 > \delta$ (refer to (9.25)), v_1, v_2 are retained which are considered as some meaningful trend, where

$$\begin{aligned} v_1 &= (a^1_{1,1}, a^1_{1,2}, a^1_{1,3}, a^1_{6,1}, \ldots, a^1_{12,3}, a^1_{13,1}, a^1_{13,2}, a^1_{13,3})^T \\ &= (-0.1135, 0.0845, 0.1263, -0.3089, \ldots, -0.1240, -0.1911, 0.1590, 0.1716)^T, \\ v_2 &= (a^2_{1,1}, a^2_{1,2}, a^2_{1,3}, a^2_{6,1}, \ldots, a^2_{12,3}, a^2_{13,1}, a^2_{13,2}, a^2_{13,3})^T \\ &= (0.3390, -0.3558, 0.0737, 0.0763, \ldots, -0.1720, 0.2538, 0.2898, 0.0444)^T. \end{aligned}$$

The number of concepts in M is $3r = 24$ (where r is the number of features in the selected feature set and here $r = 8$). We have

$$\begin{aligned} &(a^1_{1,1})^2, (a^1_{1,3})^2, (a^1_{6,1})^2, (a^1_{6,2})^2, (a^1_{7,1})^2, (a^1_{7,2})^2, (a^1_{9,1})^2, (a^1_{9,2})^2, (a^1_{10,3})^2, (a^1_{11,1})^2, \\ &(a^1_{11,2})^2, (a^1_{11,3})^2, (a^1_{12,1})^2, (a^1_{12,2})^2, (a^1_{12,3})^2, (a^1_{13,1})^2, (a^1_{13,2})^2, (a^1_{13,3})^2 > 1/(4\times 24); \end{aligned}$$

$(a_{1,1}^2)^2, (a_{1,2}^2)^2, (a_{6,3}^2)^2, (a_{7,3}^2)^2, (a_{9,3}^2)^2, (a_{10,1}^2)^2, (a_{10,2}^2)^2, (a_{11,1}^2)^2, (a_{11,2}^2)^2, (a_{12,2}^2)^2,$
$(a_{12,3}^2)^2, (a_{13,1}^2)^2, (a_{13,2}^2)^2 > 1/(4 \times 24),$

By formula (9.26), M_0 the set of the principle simple concepts is shown as follows.

$$M_0 = \{m_{1,1}, m_{1,2}, m_{1,3}, m_{6,1}, m_{6,2}, m_{6,3}, m_{7,1}, m_{7,2}, m_{7,3}, m_{9,1}, m_{9,2}, m_{9,3}, m_{10,1}, m_{10,2}, m_{10,3}, m_{11,1}, m_{11,2}, m_{11,3}, m_{12,1}, m_{12,2}, m_{12,3}, m_{13,1}, m_{13,2}, m_{13,3}\}.$$

When applying the Concept Categorization method described in Section 9.3.2.2 to the principal simple concepts in M_0, we obtain two concept categories M_1 and M_2 shown as follows.

$$M_1 = \{m_{1,1}, m_{6,1}, m_{7,1}, m_{9,1}, m_{10,1}, m_{11,1}, m_{12,1}, m_{12,3}, m_{13,1}\},$$
$$M_2 = \{m_{1,2}, m_{6,2}, m_{6,3}, m_{7,2}, m_{7,3}, m_{9,2}, m_{10,2}, m_{11,2}, m_{12,2}, m_{13,2}, m_{13,3}\}.$$

Find the fuzzy characteristic description ζ_x of x:

Using Description Method A: Description based on principal concepts or a category of concepts shown in formula (9.34), we study the fuzzy characteristic descriptions of 178 samples by the concept categories M_1 and M_2. As an example, the fuzzy characteristic descriptions of x_1, x_{27}, x_{65}, x_{99}, x_{147}, x_{178} are listed as follows:

The fuzzy characteristic descriptions under the simple concepts in category M_1:

$$\zeta_{x_1}^1 = m_{12,1},\ \zeta_{x_{27}}^1 = m_{13,1},\ \zeta_{x_{65}}^1 = m_{11,1},\ \zeta_{x_{99}}^1 = m_{6,1}m_{7,1},\ \zeta_{x_{147}}^1 = m_{1,1},\ \zeta_{x_{178}}^1 = m_{10,1}.$$

The fuzzy characteristic descriptions under the simple concepts in category M_2:

$$\zeta_{x_1}^2 = m_{6,3},\ \zeta_{x_{27}}^2 = m_{11,2},\ \zeta_{x_{65}}^2 = m_{13,2},\ \zeta_{x_{99}}^2 = m_{13,3},\ \zeta_{x_{147}}^2 = m_{6,2}m_{7,2},\ \zeta_{x_{178}}^2 = m_{11,2}.$$

For each $x_i \in X$, the best fuzzy description of $\zeta_{x_i}^1$ and $\zeta_{x_i}^2$ is selected by (9.37) according to their membership degrees $\mu_{\zeta_{x_i}^1}(x_i)$ and $\mu_{\zeta_{x_i}^2}(x_i)$. For instance,

$$1 = \arg\max_{1 \le i \le 2} \{\mu_{\zeta_{x_1}^i}(x_1)\},$$

so the final fuzzy characteristic description of x_1 is $\zeta_{x_1} = \zeta_{x_1}^1 = m_{12,1}$ with the semantic interpretation: "*the value of OD280/OD315 (O) of diluted wines is large*". Similarly, $\zeta_{x_{27}} = m_{13,1}$, $\zeta_{x_{65}} = m_{11,1}$, $\zeta_{x_{99}} = m_{6,1}m_{7,1}$, $\zeta_{x_{147}} = m_{6,2}m_{7,2}$, $\zeta_{x_{178}} = m_{10,1}$.

AFS Clustering

Input ζ_x, the fuzzy characteristic description of each $x \in X$ into the "*AFS Fuzzy Clustering Algorithm Based on Principal Concepts or the Categories of Concepts*" shown in Table 9.13, we get the clusters $C_1, C_2, ..., C_l$ and the fuzzy description ζ_{C_i} of each cluster C_i. Figure 9.3 shows when $\alpha \in (0.9045, 0.9157]$, $I_\alpha = 0.1626$ (the fuzzy cluster validation index shown as (9.4) in the elementary fuzzy clustering

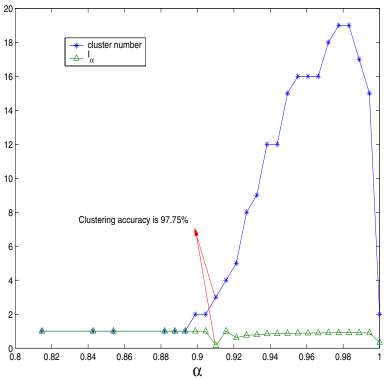


Fig. 9.3 Number of clusters and fuzzy cluster validity index I_α for Wine data

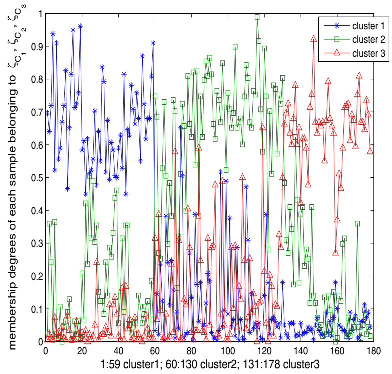


Fig. 9.4 Membership functions of ζ_{C_1}, ζ_{C_2}, ζ_{C_3}, the fuzzy descriptions of three clusters for Wine data

method) is the smallest, i.e., the clustering is the most distinguishable. When $\alpha = 0.9101$, we have the fuzzy descriptions of the clusters listed as follows.

The fuzzy description of cluster 1:

$\zeta_{C_1} = m_{6,1}m_{7,1}m_{13,1}$, with the semantic interpretation: "*the value of total phenols, flavanoids and praline are large*".

The fuzzy description of cluster 2:

$\zeta_{C_2} = m_{1,2}m_{10,2}$, i.e., with the semantic interpretation: "*the value of alcohol content and color intensity are small*".

The fuzzy description of cluster 3:

$\zeta_{C_3} = m_{6,2}m_{7,2}m_{11,2}m_{12,2}$, i.e., with the semantics: "*the value of total phenols, flavanoids, hue and OD280/OD315 (O) of diluted wines are small*".

All samples in X are re-clustered into three clusters C_1, C_2, C_3 by the fuzzy descriptions $\zeta_{C_1}, \zeta_{C_2}, \zeta_{C_3}$ of 3 clusters. In detail, if $q = \arg\max_{1\le i\le 3}\{\mu_{\zeta_{C_i}}(x)\}$, $x \in C_q$. 4 samples $x_{62}, x_{74}, x_{84}, x_{119}$ are incorrectly clustered comparing with the expected clusters: the samples $x_i, 1 \le i \le 59$ are in C_1; $x_i, 60 \le i \le 130$ are in C_2; $x_i, 131 \le i \le 178$ are in C_3. The clustering accuracy is 97.8%. I_α achieves the lowest value equal to 0.1626, i.e., C_1, C_2, C_3 are the most compact (clearest) clustering. Figure 9.4 shows the membership degrees of every sample in X belonging to the fuzzy descriptions of the three clusters, $\zeta_{C_1}, \zeta_{C_2}, \zeta_{C_3} \in EM$. It is evident that the fuzzy descriptions of the samples can be applied to other recognition problems. The proposed clustering algorithm is comparable with other methods in terms of the achieved clustering accuracy, which is reported in Table 9.15. The reason is that the algorithm can select approximately optimal (suboptimal) feature subset and in this way mimic the human reasoning processes. In this example, we encountered four errors, corresponding to

Table 9.15 Clustering accuracy compared with other pattern clustering algorithms for the Wine data set

Algorithm	Accuracy (%)
Proposed algorithm	97.8
New two-level SOM-based clustering algorithm [79]	98.3
Iterative Fuzzy Clustering Algorithm [57]	95.4
Localized feature selection-*k*-means [58]	97.7
Global feature selection-*k*-means [58]	96.1
Robust deterministic annealing (RDA) algorithm [80]	97.2
Information cut clustering algorithm [28]	97.2
Robust clustering based on Laplace mixture [11]	94.2-96.2
Mixture-based clustering [59]	93.4
Extended SOM (minimum variance) [79]	93.3
The Gath–Geva (GG) algorithm [22]	95.5
Maximum certainty data partitioning [69]	97.8
Mercer kernel based clustering algorithm [21]	97.8
Neural network for cluster-detection-and-labelling [16]	91.57
General fuzzy min-max neural network [20]	88.64-100
FCM [4]	94.9
Direct *k*-means [79]	97.8

an equivalent clustering performance of 97.8%. A new two-level SOM-based clustering algorithm can achieve the best clustering result with a highest clustering accuracy 98.3%. But the main drawback of this SOM clustering algorithm is a lack of transparent interpretation. Furthermore the method itself, is dependent on the kernel function, training data and the techniques of filtering. The three clusters are of near spherical shape, but exhibit some level of noise which places some points in-between the clusters. Therefore, the direct *k*-means algorithm comes with a sound clustering accuracy of 97.8%. Although the clustering accuracy achieved by the proposed algorithm is lower than a new two-level SOM-based clustering algorithm, the proposed algorithm exhibits some advantages such as higher interpretability and independence from the predetermined distance function and the number of clusters.

9.4.1.2 Iris Data

The Iris data can be represented in the form of a 150×4 matrix $W = (w_{ij})_{150\times 4}$ with patterns evenly distributed into three classes: Iris-setosa, Iris-versicolor, and Iris-virginica. Let $X = \{x_1, x_2, ..., x_{150}\}$ be the set of 150 samples, and $F = \{f_1, f_2, f_3, f_4\}$ be the set of features on X. Here we just list the results in each step of the algorithm. *The set of the selected salient features* is $F^{select} = \{f_3, f_4\} \subset F$.

$M = \{m_{j,k} | j = 3,4, 1 \leq k \leq 3\}$, where $m_{j,1}, m_{j,2}, m_{j,3}$ are the simple concepts, "*large*", "*small*", "*medium*" associating with the feature f_j in F^{select}, respectively.

The set of the selected principle concepts is $M_0 = \{m_{3,1}, m_{3,2}, m_{3,3}, m_{4,1}, m_{4,2}, m_{4,3}\}$.

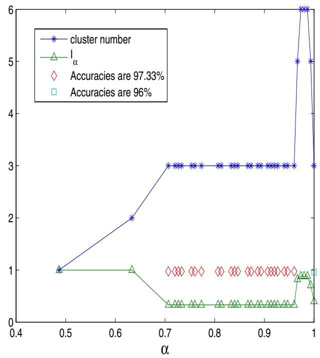


Fig. 9.5 Accuracy, number of clusters and fuzzy cluster validity index I_α for Iris data

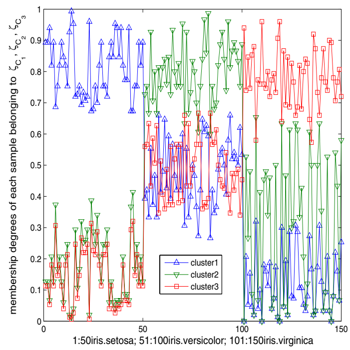


Fig. 9.6 Membership functions of ζ_{C_1}, ζ_{C_2}, ζ_{C_3}, the fuzzy descriptions of three clusters for Iris data

Table 9.16 Clustering accuracy compared with other pattern clustering algorithms for Iris data

Algorithm	Accuracy (%)
Proposed algorithm	97.3
New two-level SOM-based clustering algorithm [79]	96
Extended SOM (minimum variance) [79]	90.3
Iterative Fuzzy Clustering Algorithm [57]	95.1
Localized feature selection-*k*-means [58]	96
Global feature selection-*k*-means [58]	95.3
Robust deterministic annealing (RDA) algorithm [80]	92.7
Gath–Geva (GG) algorithm [22]	93.3
Mercer kernel based clustering algorithm [21]	98
Neural network for cluster-detection-and-labelling [16]	96
General fuzzy min-max neural network [20]	92–100
Support Vector Clustering [7]	97.3
A Novel Kernel Method for Clustering [9]	94.7
Gustafson–Kessel (GK) algorithm [19]	90
Neural Gas Network for Vector Quantization [63]	91.7
The Ng-Jordan algorithm [66]	84.3
Generalization of learning vector quantization [68]	91.3
FCM [4]	89.3
k-means [73]	89.3

Two concept categories formed on the set of principal simple concepts M_0 are $M_1 = \{m_{3,1}, m_{3,3}, m_{4,1}, m_{4,3}\}, M_2 = \{m_{3,2}, m_{4,2}\}$.

Figure 9.5 shows that when $\alpha \in (0.63333, 0.96]$, I_α returns the most discriminant result, i.e., the clustering is clearest. When threshold $\alpha = 0.70667$, the algorithm clusters the samples into three clusters and gives the fuzzy descriptions of the three clusters as follows: $\zeta_{C_1} = m_{3,2}m_{4,2}$, $\zeta_{C_2} = m_{3,3}m_{4,3}$, $\zeta_{C_3} = m_{3,1}m_{4,1}$.

4 samples $x_{71}, x_{78}, x_{84}, x_{107}$ are incorrectly clustered. The clustering accuracy is 97.3%. Figure 9.6 shows the membership functions of the fuzzy descriptions $\zeta_{C_1}, \zeta_{C_2}, \zeta_{C_3}$ of the three clusters. Results shown in Table 9.16 help complete some comparative analysis.

9.4.1.3 Wisconsin Diagnostic Breast Cancer Data (WDBC)

Wisconsin Diagnostic Breast Cancer data (WDBC) has 569 patients, each of which is characterized by 30 features. These 30 features are the mean, the standard error, and the largest error of radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry and fractal dimension. Three values are calculated on these ten features, average, standard error and the "*worst*" or the largest. There are 357 benign patients and 212 malignant patients. Let $X = \{x_1, x_2, \ldots, x_{569}\}$ be the set of 569 samples, $F = \{f_1, f_2, \ldots, f_{30}\}$ be the set of features.

The set of the selected salient features is $F^{select} = \{f_1, f_3, f_4, f_7, f_8, f_{11}, f_{13}, f_{14}, f_{21}, f_{23}, f_{24}, f_{26}, f_{27}, f_{28}\} \subset F$.

$M = \{m_{j,k} | j = 1,3,4,7,8,11,13,14,21,23,24,26,27,28, 1 \leq k \leq 3\}$, where $m_{j,1}, m_{j,2}, m_{j,3}$ are the simple concepts, " *large*", " *small*", " *medium*" associated with the feature f_j in F^{select}, respectively.

The set of the selected principle concepts is $M_0 = \{m_{1,1}, m_{1,2}, m_{3,1}, m_{3,2}, m_{4,1}, m_{4,2}, m_{7,1}, m_{7,2}, m_{8,1}, m_{8,2}, m_{11,1}, m_{11,2}, m_{13,1}, m_{13,2}, m_{14,1}, m_{14,2}, m_{21,1}, m_{21,2}, m_{23,1}, m_{23,2}, m_{24,1}, m_{24,2}, m_{26,1}, m_{26,2}, m_{27,1}, m_{27,2}, m_{28,1}, m_{28,2}\}$.

The concept categories on the set of principal simple concepts M_0 are $M_1 = \{m_{1,1}, m_{3,1}, m_{4,1}, m_{7,1}, m_{8,1}, m_{11,1}, m_{13,1}, m_{14,1}, m_{21,1}, m_{23,1}, m_{24,1}, m_{26,1}, m_{27,1}, m_{28,1}\}$ and $M_2 = \{m_{1,2}, m_{3,2}, m_{4,2}, m_{7,2}, m_{8,2}, m_{11,2}, m_{13,2}, m_{14,2}, m_{21,2}, m_{23,2}, m_{24,2}, m_{26,2}, m_{27,2}, m_{28,2}\}$.

Figure 9.7 shows that when $\alpha \in (0.8770, 0.8858]$, the clustering returns the most compact structure. When threshold $\alpha = 0.8805$, this method clusters the samples into two clusters and produces the fuzzy descriptions of the two clusters:

$$\zeta_{C_1} = m_{1,1}m_{3,1}m_{4,1}m_{7,1}m_{8,1}m_{11,1}m_{13,1}m_{14,1}m_{21,1}m_{23,1}m_{24,1}m_{26,1}m_{27,1}m_{28,1},$$
$$\zeta_{C_2} = m_{7,2}m_{8,2}m_{21,2}m_{23,2}m_{24,2}m_{28,2}.$$

31 samples $x_{11}, x_{35}, x_{37}, x_{38}, x_{41}, x_{53}, x_{71}, x_{74}, x_{80}, x_{90}, x_{96}, x_{135}, x_{137}, x_{142}, x_{146}, x_{170}, x_{176}, x_{192}, x_{206}, x_{241}, x_{257}, x_{266}, x_{277}, x_{282}, x_{358}, x_{396}, x_{451}, x_{457}, x_{500}, x_{508}, x_{518}$ are incorrectly clustered comparing with the expected clusters: the samples $x_i, 1 \leq i \leq 212$ are in C_1; $x_i, 213 \leq i \leq 569$ are in C_2. The clustering accuracy is 94.6%. Figure 9.8 shows the membership functions of the fuzzy descriptions ζ_{C_1}, ζ_{C_2} of the two clusters. The clustering accuracy is now 94.6%.

Table 9.17 offers a comparison between the proposed method and other published models. The proposed algorithm achieved the best clustering result with the highest clustering accuracy of 94.6%. The FCM and the *k*-means come with the same

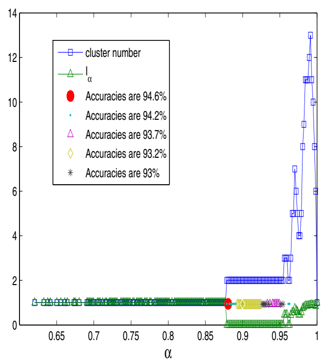


Fig. 9.7 Accuracy, number of clusters and fuzzy cluster validity index I_α for WDBC data

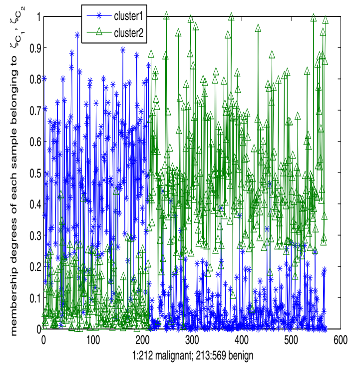


Fig. 9.8 Membership functions of ζ_{C_1}, ζ_{C_2}, the fuzzy descriptions of two clusters for WDBC data

Table 9.17 Clustering accuracy obtained for different clustering methods for WDBC data

Algorithm	Accuracy (%)
Proposed algorithm	94.6
FCM [4]	92.8
k-means [73]	92.8
Mixture-based clustering [59]	90.7
Localized feature selection-*k*-means [58]	90
Global feature selection-*k*-means [58]	91
Robust clustering based on Laplace mixture [11]	91.2-93.8

clustering accuracy of 92.8% which is lower than obtained here. The Mixture-based clustering [59] achieves accuracy of 90.7%. It can only be applied to the data sets with numeric valued features.

9.4.1.4 Parameters Analysis

The "*AFS Fuzzy Clustering Algorithm Based on Principal Concepts and the Categories of Concepts*", comes with two parameters, that is λ and ω. Section 9.3.5 elaborates on their intuitive meaning and offers some insights into their nature. In what follows, we analyze the effects of varying their values on the quality of the obtained results. Figure 9.9, Figure 9.10 and Figure 9.11 show the clustering accuracies are stable when $\lambda \in [0.5, 0.9]$ and $\omega \in [0.5, 0.9]$. For these three data sets, we see that the accuracy level lower than 90% occur when the parameters λ and ω are set unreasonably, i.e., the choice of the parameters are not consistent with the intuitive meaning of λ and ω presented in Section 9.3.5, such as $(\lambda, \omega) = (0.9, 0.9), (\lambda, \omega) = (0.5, 0.5)$. Thus, the underlying experience tells that the clustering algorithm is not overly sensitive to the numeric setting of these parameters assuming that these values come from "sound" intervals as described above. In this section, we let $\lambda = 0.6$ and $\omega = 0.76$ to be applied to the three data sets.

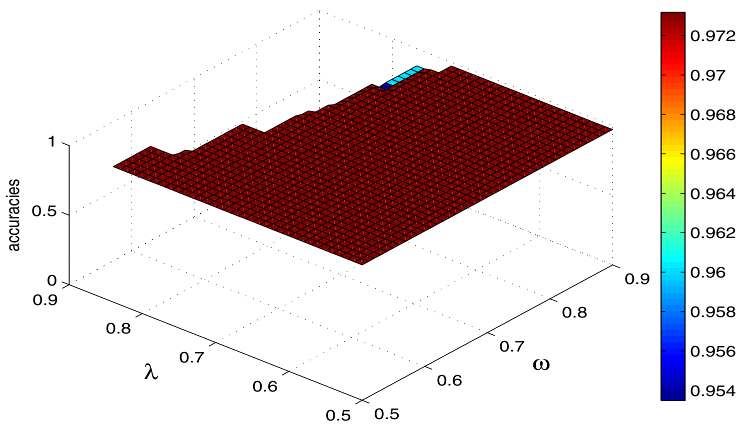


Fig. 9.9 Clustering accuracies when $\lambda \in [0.5, 0.9]$ and $\omega \in [0.5, 0.9]$ for Iris data

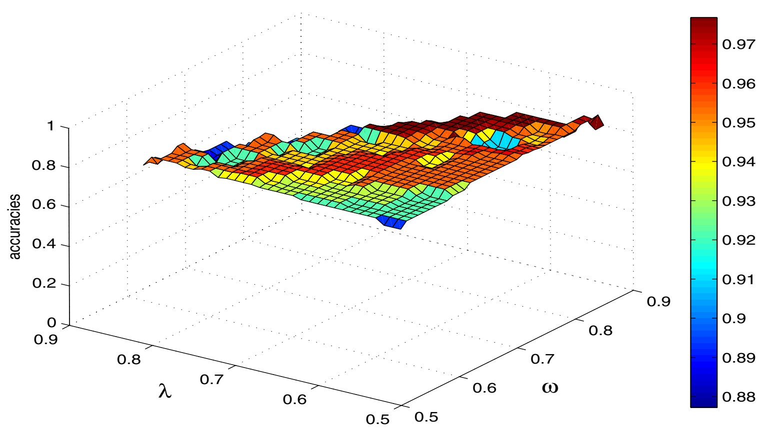


Fig. 9.10 Clustering accuracies when $\lambda \in [0.5, 0.9]$ and $\omega \in [0.5, 0.9]$ for Wine data

The experimental results show that, the proposed technique of the feature selection, the principal concept selection and concept categorization, the characteristic description are effective in supporting the clustering process. High accuracy of clustering may be due to the fact that the salient features and the principal simple concepts are selected, the optimal concept categories are obtained, and the fuzzy descriptions are similar to the descriptions consistent with a human intuition.

The feature selection is the task of selection the "*best*" feature subset. The concept categorization we are interested in clustering, the fuzzy concepts of high correlations. The characteristic description, concepts are built from simple concepts using the logic connectives, and thus can present the fuzzy characteristic of the object.

Many other methods do need initial structure learning to determine the initial structure of the fuzzy system, including the rules, the shape of membership

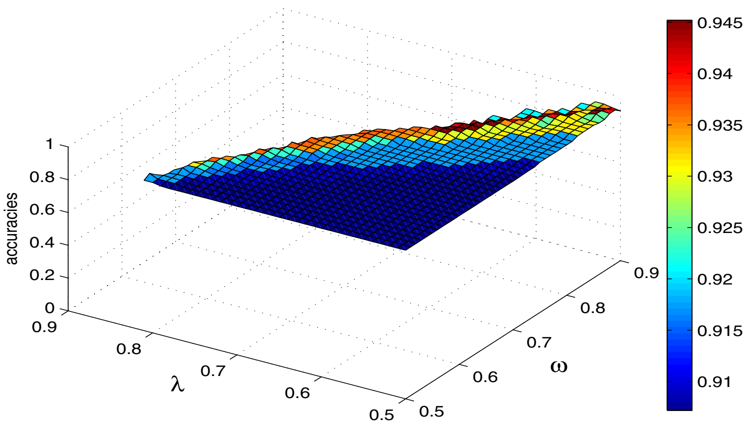


Fig. 9.11 Clustering accuracies when $\lambda \in [0.5, 0.9]$ and $\omega \in [0.5, 0.9]$ for WDBC data

functions, the number of rules and the number of membership functions associated with each feature. The clustering result is also interpretable, as each cluster have a fuzzy set in *EM* with a definite linguistic interpretation.

9.4.2 Experimental Results of AFS Fuzzy Clustering Algorithm Based on the $1/k-A$ Nearest Neighbors

We evaluate the performance of the AFS Fuzzy Clustering Algorithm Based on the $1/k-A$ Nearest Neighbors and using the same three benchmark data sets as in the previous experiment. In all experiments presented in this section, the values of the parameters are $k = 3$, ε is about 0.4 and the weight functions of concepts *"small"* , *"medium"*, *"no-medium"*, *"large"* associated with the features are defined by (9.29)-(9.32). Then the set of coherence membership functions $\{\mu_\xi(x) \mid \xi \in EM\}$ can be obtained by (5.24) in Proposition 5.7, which is defined as (9.16). We will show the details of the experiment when using the Iris data.

9.4.2.1 Iris Data

In the Iris data we encounter four features: sepal length and width, and petal length and width (all given in centimeters). Let $X = \{x_1, x_2, ..., x_{150}\}$ be the set of 150 samples, and $F = \{f_1, f_2, f_3, f_4\}$ be the set of the features on X. $x_i = (w_{i1}, w_{i2}, w_{i3}, w_{i4})$, $i = 1, 2, ..., 150$, where $w_{i1} = f_1(x_i)$ is the sepal-length of x_i, $w_{i2} = f_2(x_i)$ is the sepal-width of x_i, $w_{i3} = f_3(x_i)$ is the petal-length of x_i, $w_{i4} = f_4(x_i)$ is the petal-width of x_i. Let $M = \{m_{ij} \mid 1 \le i \le 4,\ 1 \le j \le 4\}$ be the set of the simple concepts on X. Where $m_{i1}, m_{i2}, m_{i3}, m_{i4}$ are the fuzzy concepts, *"small"*, *"medium"*, *"not medium"*, *"large"* associating to the feature f_i respectively and the weight functions of them are defined by (9.29)-(9.32). The semantic meanings of the simple concepts in M are shown as follows: $m_{1,1}$: *"short sepal length"*, $m_{1,2}$: *"mid sepal length"*, $m_{1,3}$: *"not mid sepal length"*, $m_{1,4}$: *"long sepal length"*; $m_{2,1}$: *"narrow sepal width"*, $m_{2,2}$:

"mid sepal width", $m_{2,3}$: *"not mid sepal width"*, $m_{2,4}$: *"wide sepal width"*; $m_{3,1}$: *"short petal length"*, $m_{3,2}$: *"mid petal length"*, $m_{3,3}$: *"not mid petal length"*, $m_{3,4}$: *"long petal length"*; $m_{4,1}$: *"narrow petal width"*, $m_{4,2}$: *"mid petal width"*, $m_{4,3}$: *"not mid petal width"*, $m_{4,4}$: *"wide petal width"*. One can verify that every $m \in M$ is a simple concept by Definition 4.3. For any $x, y \in X$, τ is defined by (4.26), then (M, τ, X) is an AFS structure. In what follows, we illustrate the *Description Method B: Description based on the* $1/k - A$*-Nearest* in Section 9.3.3 and *Procedure A-D* of the algorithm in Section 9.3.4 by its applications to cluster the samples of Iris data in a detailed fashion.

Select simple concepts for the fuzzy description of each sample

Using the $1/k$-A-nearest neighbor described in Section 9.3.2.3, for each $x \in X$, the set of simple concepts $\Lambda_x \subseteq M$ is selected by the *Description Method B* presented in Section 9.3.3. Here, as an example, we just show the detailed results of its every step to select simple concepts for the fuzzy description of the sample x_1. For each simple concept $m_{ij} \in M$, the number of the intersection of the $1/k$-M-nearest neighbor of x_1 and the $1/k$-$\{m_{ij}\}$-nearest neighbor of x_1 defined by Definition 9.6 is shown as follows:

$$
\begin{aligned}
&|D^M_{1/3-x_1} \cap D^{m_{1,1}}_{1/3-x_1}| = 38,\ |D^M_{1/3-x_1} \cap D^{m_{1,2}}_{1/3-x_1}| = 23,\ |D^M_{1/3-x_1} \cap D^{m_{1,3}}_{1/3-x_1}| = 23,\\
&|D^M_{1/3-x_1} \cap D^{m_{1,4}}_{1/3-x_1}| = 38,\ |D^M_{1/3-x_1} \cap D^{m_{2,1}}_{1/3-x_1}| = 32,\ |D^M_{1/3-x_1} \cap D^{m_{2,2}}_{1/3-x_1}| = 17,\\
&|D^M_{1/3-x_1} \cap D^{m_{2,3}}_{1/3-x_1}| = 17,\ |D^M_{1/3-x_1} \cap D^{m_{2,4}}_{1/3-x_1}| = 32,\ |D^M_{1/3-x_1} \cap D^{m_{3,1}}_{1/3-x_1}| = 50,\\
&|D^M_{1/3-x_1} \cap D^{m_{3,2}}_{1/3-x_1}| = 42,\ |D^M_{1/3-x_1} \cap D^{m_{3,3}}_{1/3-x_1}| = 42,\ |D^M_{1/3-x_1} \cap D^{m_{3,4}}_{1/3-x_1}| = 50,\\
&|D^M_{1/3-x_1} \cap D^{m_{4,1}}_{1/3-x_1}| = 50,\ |D^M_{1/3-x_1} \cap D^{m_{4,2}}_{1/3-x_1}| = 41,\ |D^M_{1/3-x_1} \cap D^{m_{4,3}}_{1/3-x_1}| = 41,\\
&|D^M_{1/3-x_1} \cap D^{m_{4,4}}_{1/3-x_1}| = 50 \qquad .
\end{aligned}
$$

The $|M|/3 = 16/3$ simple concepts (i.e., 5 simple concepts) in M corresponding to the largest $|D^m_{1/k-x_1} \cap D^M_{1/k-x_1}|$ (i.e., 50, 42) of the others in M are selected. Since there are 4 simple concepts $m_{4,1}, m_{4,4}, m_{3,1}, m_{3,4}$ corresponding to 50 and 2 simple concepts $m_{3,2}, m_{3,3}$ corresponding to 42, hence one of $m_{3,2}$ and $m_{3,3}$ can be randomly selected as the fifth one. Finally, the 5 simple concepts in $\Lambda_{x_1} = \{m_{4,1}, m_{4,4}, m_{3,1}, m_{3,4}, m_{3,3}\}$ are selected to describe the sample x_1. The selected concepts to describe x_{21}, x_{66}, x_{108} are listed below:

$$
\begin{aligned}
\Lambda_{x_{21}} &= \{m_{4,1}, m_{4,4}, m_{3,1}, m_{3,4}, m_{4,3}\},\\
\Lambda_{x_{66}} &= \{m_{3,3}, m_{3,1}, m_{3,4}, m_{2,2}, m_{4,1}\},\\
\Lambda_{x_{108}} &= \{m_{1,1}, m_{1,4}, m_{3,1}, m_{3,4}, m_{4,1}\}.
\end{aligned}
$$

Find the description of each sample

From the *Description Method B: Description based on the* $1/k - A$*-Nearest* in Section 9.3.3, for every $x \in X$, the set Λ_x of the selected simple concepts to describe x is obtained by the above procedure. The fuzzy description ζ_x of x based on the selected simple concepts in Λ_x can be formed using the formulas (9.35), (9.36), (9.37) in the *Find-Fuzzy-Description-Algorithm* of the *Description Method B* in which $\varepsilon = 0.34$ is a predetermined constant for all the samples. We just show the detailed results of its every step to determine the descriptions of the samples x_1, x_{21} and x_{66} taken here as selected examples. Let $\vartheta = \sum_{m\in\Lambda_{x_1}} m = m_{4,1} + m_{4,4} + m_{3,1} + m_{3,4} + m_{3,3}$. Then for any fuzzy concept $\eta \in E\Lambda_{x_1}$, the set of all fuzzy concepts generated by the simple concepts, where

$$\Lambda_{x_1} = \{m_{4,1}, m_{4,4}, m_{3,1}, m_{3,4}, m_{3,3}\},$$

$\mu_\eta(x) \leq \mu_\vartheta(x)$ for any $x \in X$. $\mu_\vartheta(x_1) - \varepsilon = 0.62$. By

$$\begin{aligned}
&\mu_{m_{4,1}}(x_1) = 0.9600 \geq 0.62, \quad \mu_{m_{4,4}}(x_1) = 0.2267 \leq 0.62,\\
&\mu_{m_{3,1}}(x_1) = 0.9267 \geq 0.62, \quad \mu_{m_{3,4}}(x_1) = 0.1533 \leq 0.62,\\
&\mu_{m_{3,3}}(x_1) = 0.8867 \geq 0.62,
\end{aligned}$$

we have $B^{0.34}_{x_1} = \{m_{4,1}, m_{3,1}, m_{3,3}\}$ defined by (9.35). From $\mu_{m_{3,1}m_{3,3}m_{4,1}}(x_1) = 0.8533 \geq 0.62$, we have $A^{0.34}_{x_1} = \{m_{4,1}m_{3,1}m_{3,3}\}$ defined by (9.36). $\zeta_{x_1} = m_{4,1}m_{3,1}m_{3,3}$ is selected to describe x_1 by (9.37) due to $A^{0.34}_{x_1}$ just having one element.

For $x_{21} \in X$,

$$\begin{aligned}
\Lambda_{x_{21}} &= \{m_{4,1}, m_{4,4}, m_{3,1}, m_{3,4}, m_{4,3}\},\\
\vartheta &= m_{4,1} + m_{4,4} + m_{3,1} + m_{3,4} + m_{4,3},
\end{aligned}$$

$\mu_\vartheta(x_{21}) - \varepsilon = 0.62$. By

$$\begin{aligned}
&\mu_{m_{4,1}}(x_{21}) = 0.9600 \geq 0.62, \quad \mu_{m_{4,4}}(x_{21}) = 0.2267 \leq 0.62,\\
&\mu_{m_{3,1}}(x_{21}) = 0.7067 \geq 0.62, \quad \mu_{m_{3,4}}(x_{21}) = 0.3200 \leq 0.62,\\
&\mu_{m_{4,3}}(x_{21}) = 0.8467 \geq 0.62,
\end{aligned}$$

we have $B^{0.34}_{x_{21}} = \{m_{4,1}, m_{3,1}, m_{4,3}\}$. From

$$\begin{aligned}
&\mu_{m_{4,1}m_{3,1}}(x_{21}) = 0.7067 \geq 0.62, \quad \mu_{m_{4,1}m_{4,3}}(x_{21}) = 0.8467 \geq 0.62,\\
&\mu_{m_{4,3}m_{3,1}}(x_{21}) = 0.5933 < 0.62,
\end{aligned}$$

we have $A^{0.34}_{x_{21}} = \{m_{4,1}m_{3,1},\ m_{4,1}m_{4,3}\}$. By (9.37), $\zeta_{x_{21}} = m_{4,1}m_{4,3}$ is selected to describe x_{21} according to

$$|\{x \in X | \mu_{m_{4,1}m_{3,1}}(x) \geq \mu_{m_{4,1}m_{3,1}}(x_{21})\}| = 43,$$

$$|\{x \in X | \mu_{m_{4,1}m_{4,3}}(x) \geq \mu_{m_{4,1}m_{4,3}}(x_{21})\}| = 34.$$

Similarly, for $x_{66} \in X$,

$$\Lambda_{x_{66}} = \{m_{3,3}, m_{3,1}, m_{3,4}, m_{3,2}, m_{4,1}\}, \;\; B_{x_{66}}^{0.34} = \{m_{3,4}, m_{3,1}, m_{3,2}\},$$
$$A_{x_{66}}^{0.34} = \{m_{3,1}m_{3,2},\; m_{3,4}\}, \;\; \zeta_{x_{66}} = m_{3,1}m_{3,2}.$$

For $x_{108} \in X$,

$$\Lambda_{x_{108}} = \{m_{1,1}, m_{1,4}, m_{3,1}, m_{3,4}, m_{4,1}\}, \;\; B_{x_{108}}^{0.34} = \{m_{1,4}, m_{3,4}\},$$
$$A_{x_{108}}^{0.34} = m_{1,4}m_{3,4}, \;\; \zeta_{x_{108}} = m_{1,4}m_{3,4}.$$

Figures 9.12 to 9.15 present the membership functions of ζ_x the fuzzy descriptions of the 4 samples x_1, x_{21} in cluster Iris-setosa, x_{66} in cluster Iris-versicolor, x_{108} in Iris-virginica. One can observe that most of the samples in the same cluster as x belong to ζ_x the fuzzy description of x at relatively high degrees.

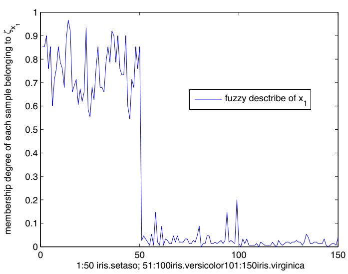


Fig. 9.12 Membership function of ζ_{x_1}

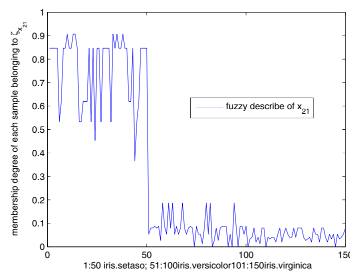


Fig. 9.13 Membership function of $\zeta_{x_{21}}$

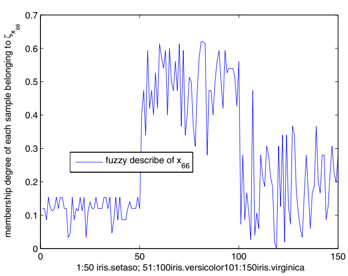


Fig. 9.14 Membership function of $\zeta_{x_{66}}$

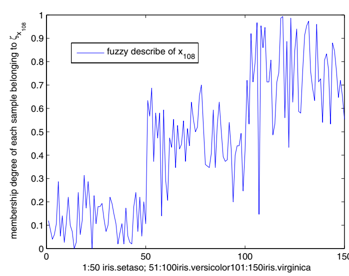


Fig. 9.15 Membership function of $\zeta_{x_{108}}$

Evaluate the similarity between the samples

The above fuzzy descriptions ζ_{x_i} of $x_i \in X$ are applied to establish the fuzzy relation matrix $F_\zeta = (f_{ij})$ in which f_{ij} is the similarity degree between x_i and x_j defined by (9.38)(i.e., the procedure A in Section 9.3.4). The similarity degrees between some samples are listed here as examples: $f_{1,1} = 0.8533, f_{1,21} = 0.5467, f_{1,66} = 0.0133, f_{1,108} = 0.0067, f_{21,21} = 0.8467, f_{21,66} = 0.0467, f_{21,108} = 0.08, f_{66,66} =$

$0.5, f_{66,108} = 0.0067, f_{108,108} = 0.9533$. For the fuzzy relation matrix $Q = F_{\zeta}^{4}$, $Q^2 = Q$.

Cluster the samples, and get the description of every cluster

In what follows, we demonstrate the procedure B is described in Section 9.3.4. As an example, here we just show the results of the algorithm for threshold $\alpha = 0.55$ under which the best clustering result is obtained (i.e., I_α the fuzzy cluster validity index defined by (9.40) reaches the maximum at $\alpha = 0.55$). The valid clusters C_1, C_2, C_3 are obtained by the Boolean matrix $Q_{0.55}$ and the set of the descriptions of all samples in C_i, $\mathscr{T}_{C_i} = \{\zeta_x \mid x \in C_i\}$, $i = 1,2,3$ are shown as follows:

$$\mathscr{T}_{C_1} = \{m_{3,1}m_{3,3}m_{4,1},\ m_{1,1}m_{3,1}m_{4,1},\ m_{3,1}m_{4,1},\ m_{3,1}m_{3,1}m_{4,3},\ m_{4,1}m_{4,3}\},$$
$$\mathscr{T}_{C_2} = \{m_{3,2},\ m_{4,2},\ m_{3,2}m_{4,2},\ m_{2,1}m_{4,2},\ m_{2,1}m_{3,2}\},$$
$$\mathscr{T}_{C_3} = \{m_{4,4},\ m_{1,4}m_{3,4},\ m_{1,4}m_{4,4},\ m_{3,4}m_{4,4}\}.$$

$\mathscr{D}_{C_i}$, the sets of the typical descriptions selected by formula (9.39) from $\mathscr{T}_{C_i}$, are shown as follows:

$$\mathscr{D}_{C_1} = \{m_{3,1}m_{4,1},\ m_{4,3}m_{4,1}\},\ \ \mathscr{D}_{C_2} = \{m_{3,2}m_{4,2}\},\ \ \mathscr{D}_{C_3} = \{m_{1,4}m_{3,4},\ m_{4,4}\}.$$

Thus, the cluster C_i is described by the fuzzy concept $\zeta_{C_i} = \bigvee_{\zeta_x \in \mathscr{D}_{C_i}} \zeta_x \in EM$ in the following manner:

$$\zeta_{C_1} = m_{3,1}m_{4,1} + m_{4,3}m_{4,1},\ \ \zeta_{C_2} = m_{3,2}m_{4,2},\ \ \zeta_{C_3} = m_{1,4}m_{3,4} + m_{4,4}.$$

Select the best result

Using the values of the I_α, the fuzzy cluster validity index defined by (9.40) and exploited in the procedure C in Section 9.3.4, for $\alpha \in [0.52, 0.54]$, $\alpha \in [0.5229, 0.5352]$ is the largest of the other $\alpha \in [0,1]$ and this clustering algorithm clusters the samples into two clusters: all samples Iris-setosa are placed in one cluster and all samples in Iris-versicolor or Iris-virginica are positioned in another cluster. According to the distribution of the Iris data, indeed this is the clearest clustering of the data, although the expected number of clusters equal to three. For $\alpha \in [0.5667, 0.5901]$, $I_\alpha = 0.4291$ is the largest of all $\alpha \in [0.54, 0.9]$, under which the samples are clustered into three clusters and the clustering accuracy is the highest, and the fuzzy descriptions of the three clusters:

$\zeta_{C_1} = m_{3,1}m_{4,1} + m_{4,3}m_{4,1}$ states *"short petal length and narrow petal width"* or *"not mid petal width and narrow petal width"*;

$\zeta_{C_2} = m_{3,2}m_{4,2}$ reads *"mid petal length and mid petal width"*;

$\zeta_{C_3} = m_{1,4}m_{3,4} + m_{4,4}$ with the semantic interpretation: *"long sepal length and long petal length" or "wide petal width"*.

Figure 9.16 shows the values of the fuzzy cluster validity index I_α defined by (9.40), the clustering accuracy and the number of clusters of this algorithm versus the values of the threshold α.

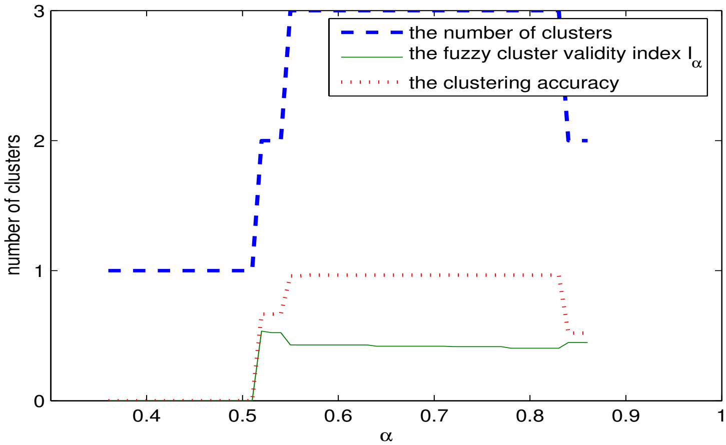


Fig. 9.16 The number of clusters, the clustering accuracy and the fuzzy cluster validity index I_α of the algorithm applying to Iris data versus the threshold α; the values of the threshold α are shown here in the [0.4, 0.9] interval

Re-cluster the samples

According to *Procedure D* in Section 9.3.4, all samples in X are re-clustered into three clusters $\overline{C}_1$, $\overline{C}_2$, $\overline{C}_3$ by using ζ_{C_1}, ζ_{C_2}, ζ_{C_3}, the fuzzy descriptions of the clusters which are the best clustering result selected by the fuzzy cluster validity index defined by (9.40). In detail, if $q = \arg\max_{1\leq i\leq 3}\{\mu_{\zeta_{C_i}}(x)\}$ then $x \in \overline{C}_q$. 6 samples x_{53}, x_{57}, x_{71}, x_{78}, x_{84}, x_{86} are incorrectly clustered compared with the expected clusters: the samples $x_i, 1 \leq i \leq 50$ are Iris-setosa, $x_i, 51 \leq i \leq 100$ are iris-versicolor, $x_i, 101 \leq i \leq 150$ are iris-virginica. The clustering accuracy is 96%. Figure 9.17 shows the membership degrees of every sample in X belonging to the fuzzy descriptions of the three clusters, ζ_{C_1}, ζ_{C_2}, $\zeta_{C_3} \in EM$. The fuzzy descriptions of the samples can be applied to other recognition problems.

Given Figure 9.16, one can observe that the clustering accuracies of the clustering results determined by the thresholds $\alpha \in [0.5667, 0.8333]$ fall in-between 96% and 96.67%. Different fuzzy descriptions of the clusters may be obtained by choosing different threshold values. Table 9.18 shows the different fuzzy descriptions of the clusters for threshold $\alpha \in [0.5667, 0.8333]$. However, for all threshold levels $\alpha \in [0.5667, 0.8333]$, only four different fuzzy descriptions of C_1 Iris-setosa are found and they have the following relationships in the lattice EM,

$$\begin{aligned} m_{1,1}m_{4,1}m_{3,1} &\leq m_{4,1}m_{3,1}m_{3,3} + m_{1,1}m_{4,1}m_{3,1} + m_{4,1}m_{3,1}m_{4,3} \\ &\leq m_{4,1}m_{3,1} \leq m_{4,1}m_{3,1} + m_{4,1}m_{4,3}; \end{aligned}$$

One fuzzy description of C_2 Iris-versicolor is $m_{3,2}m_{4,2}$; two different fuzzy descriptions of C_3 are found and they exhibit the following relationships in the lattice EM,

$$m_{1,4}m_{3,4} + m_{4,4}m_{3,4} + m_{1,4}m_{4,4} \leq m_{1,4}m_{3,4} + m_{4,4}.$$

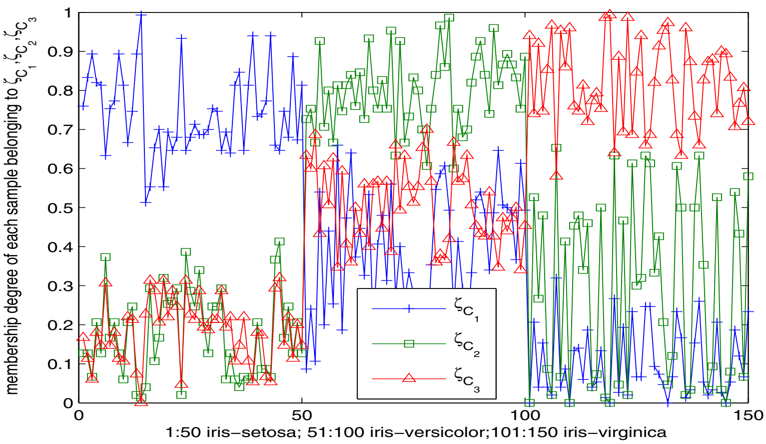


Fig. 9.17 Membership functions of the fuzzy descriptions of C_1, C_2, C_3 for iris data

Table 9.18 Different fuzzy descriptions of the clusters obtained for different threshold α

$\alpha \in [0.55, 0.83]$	ζ_{C_1}	ζ_{C_2}	ζ_{C_3}
$\alpha = [0.55, 0.59]$	$m_{4,1}m_{3,1} + m_{4,1}m_{4,3}$	$m_{3,2}m_{4,2}$	$m_{1,4}m_{3,4} + m_{4,4}$
$\alpha = [0.6, 0.63]$	$m_{4,1}m_{3,1}$	$m_{3,2}m_{4,2}$	$m_{1,4}m_{3,4} + m_{4,4}$
$\alpha = [0.64, 0.71]$	$m_{4,1}m_{3,1}m_{3,3}$ $+m_{1,1}m_{4,1}m_{3,1}$ $+m_{3,1}m_{4,1}m_{4,3}$	$m_{3,2}m_{4,2}$	$m_{1,4}m_{3,4} + m_{4,4}$
$\alpha = [0.72, 0.77]$	$m_{1,1}m_{4,1}m_{3,1}$	$m_{3,2}m_{4,2}$	$m_{1,4}m_{3,4} + m_{4,4}$
$\alpha = [0.78, 0.83]$	$m_{1,1}m_{4,1}m_{3,1}$	$m_{3,2}m_{4,2}$	$m_{1,4}m_{3,4}$ $+m_{4,4}m_{3,4}$ $+m_{1,4}m_{4,4}$

This implies that both the clustering accuracy and the fuzzy descriptions of the algorithm are very stable.

9.4.2.2 Wine Classification Data

The wine data set contains 178 wines that are brewed in the same region of Italy but derived from three different cultivars. Thirteen continuous features are measured on each wine: f_1 alcohol content (AL),f_2 malic acid content (MA), f_3 ash content (AS), f_4 alcalinity of ash (AA), f_5 magnesium content (MAG), f_6 total phenols (TP), f_7 flavanoids (FL), f_8 nonflavanoids phenols (NP), f_9 proanthocyaninsm (PR), f_{10} color intensity (CI), f_{11} hue (HU), f_{12} OD280/OD315 (OD) of diluted wines, and f_{13} praline (PRO). The numbers of patterns in the three classes are $|C_1| = 59$, $|C_2| = 71$ and $|C_3| = 48$, respectively. Let $X = \{x_1, x_2, \cdots, x_{178}\}$ and $F = \{f_1, f_2, \ldots, f_{13}\}$

be a set of features on X. Let $M = \{m_{ij} \mid 1 \le i \le 13,\ 1 \le j \le 2\}$. Where m_{i1}, m_{i2} are the fuzzy concepts *"low", "high"* associating to the feature f_i respectively and the weight functions of them is defined by (9.29) and (9.32). By same procedures as what we applied to Iris Data with the same parameters setting, for $\alpha = 0.5021$, at which the fuzzy cluster validity index $I_\alpha = 0.4040$ (defined by (9.40)) reaches the maximum value and the clustering accuracy is the highest (reference to Figure 9.18), the clustering algorithm places the samples into three clusters and gives the fuzzy descriptions of the three clusters in the following manner:

$$\begin{aligned}
\zeta_{C_1} &= m_{1,2}m_{7,2}m_{13,2} + m_{4,1}m_{8,1}m_{13,2} + m_{1,2}m_{4,1}m_{13,2} + m_{6,2}m_{7,2}m_{13,2} \\
&\quad + m_{3,2}m_{7,2}m_{12,2} + m_{7,2}m_{11,2}m_{13,2} + m_{4,1}m_{7,2}m_{13,2} + m_{1,2}m_{6,2}m_{7,2} \\
&\quad + m_{4,1}m_{11,2}m_{13,2} + m_{6,2}m_{7,2}m_{9,2} + m_{1,1}m_{7,2}m_{12,2}, \\
\zeta_{C_2} &= m_{3,1}m_{4,1}m_{8,1}m_{10,1} + m_{2,1}m_{3,1}m_{13,1} + m_{1,1}m_{10,1} + m_{2,1}m_{10,1} \\
&\quad + m_{7,1}m_{10,1} + m_{8,2}m_{13,1} + m_{4,2}m_{10,1}m_{13,1} + m_{4,2}m_{5,1}m_{10,1},
\end{aligned}$$

$$\begin{aligned}
\zeta_{C_3} &= m_{2,2}m_{6,1}m_{8,2}m_{12,1} + m_{5,1}m_{7,1}m_{12,1} + m_{6,1}m_{9,1}m_{11,1}m_{12,1} + m_{6,1}m_{7,1}m_{9,1}m_{12,1} \\
&\quad + m_{7,1}m_{8,2}m_{12,1} + m_{2,2}m_{6,1}m_{7,1}m_{12,1} + m_{6,1}m_{7,1}m_{8,2}m_{9,1} + m_{6,1}m_{7,1}m_{9,1}m_{11,1} \\
&\quad + m_{2,2}m_{7,1}m_{9,1}m_{12,1} + m_{2,2}m_{6,1}m_{7,1}m_{11,1} + m_{6,1}m_{7,1}m_{11,1}m_{12,1}m_{2,2}m_{7,1}m_{11,1}m_{12,1} \\
&\quad + m_{2,2}m_{6,1}m_{11,1}m_{12,1} + m_{6,1}m_{10,2}m_{11,1}m_{12,1} + m_{4,2}m_{7,1}m_{11,1}m_{12,1} \\
&\quad + m_{8,2}m_{10,2}m_{11,1}m_{12,1} + m_{7,1}m_{10,2}m_{11,1}m_{12,1}.
\end{aligned}$$

6 samples $x_{66}, x_{67}, x_{74}, x_{99}, x_{122}, x_{127}$ are incorrectly clustered in comparison with the expected clusters: the samples $x_i, 1 \le i \le 59$ are in C_1, $x_i, 60 \le i \le 130$ are in C_2, $x_i, 131 \le i \le 178$ are in C_3. The clustering accuracy is 96.63%. Figure 9.19 shows the membership functions of the fuzzy descriptions of the three clusters $\zeta_{C_1}, \zeta_{C_2}, \zeta_{C_3}$.

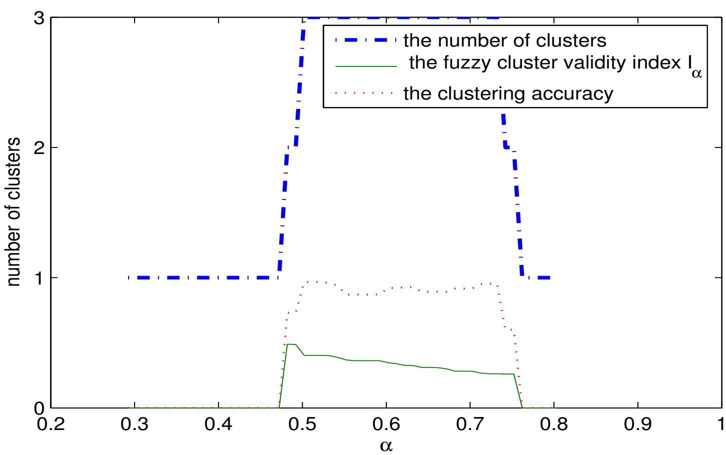


Fig. 9.18 The number of clusters, the clustering accuracy and the fuzzy cluster validity index I_α of the algorithm applying to wine data versus the threshold α; the values of the threshold α are shown here in the [0.3, 0.8] interval

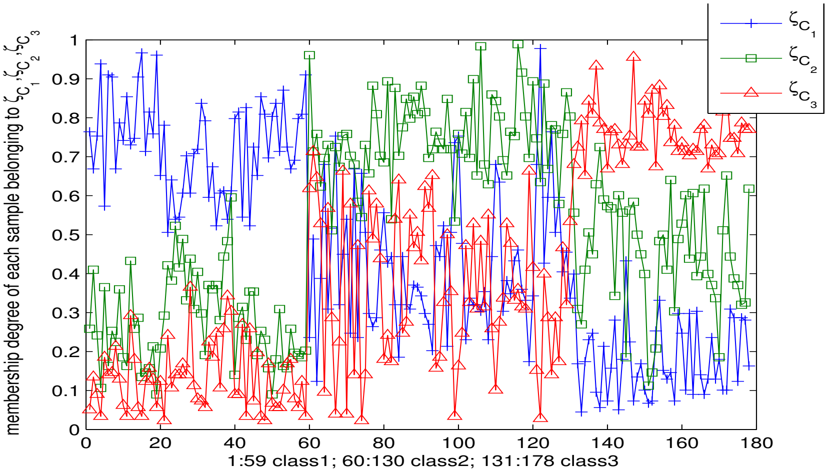


Fig. 9.19 Membership functions of the fuzzy descriptions of C_1, C_2, C_3 for wine data

9.4.2.3 Wisconsin Breast Cancer Diagnostic Data

The Wisconsin Breast Cancer Diagnostic data set contains 699 patterns distributed into two output classes, namely *"benign"* and *"malignant"*. Each pattern consists of nine features: f_1 clump thickness (CT), f_2 uniformity of cell size (UC), f_3 uniformity of cell shape (UCS), f_4 marginal adhesion (MA), f_5 single epithelial cell size (SECS), f_6 bare nuclei (BN), f_7 bland chromatin (BC), f_8 normal nuclei (NN), and f_9 mitoses (MI). The nine features have integer values in the range of 1—10 that describe visually assessed characteristics of fine needle aspiration (FNA) samples. There are 458 patterns for *"benign"* (labeling as "2" in the data set) and 241 patterns for *"malignant"* (labeling as "4"). There are 16 patterns with incomplete feature descriptions marked as "?". We used 683 patterns to evaluate the performance of the proposed fuzzy clustering algorithm, $X = \{x_1, x_2, \cdots, x_{683}\}$. The samples x_i, $1 \le i \le 444$ are *"benign"* and the samples x_i, $445 \le i \le 683$ are *"malignant"*. Let $M = \{m_{ij} \mid 1 \le i \le 9,\ 1 \le j \le 2\}$. Here m_{i1}, m_{i2} are the fuzzy concepts *"low", "high"* associating to the feature f_i, respectively and the weight functions of them is defined by (9.29) and (9.32). Let k also be 3 and ε be 0.3. The maximum fuzzy cluster validity index I_α is 0.5384 at $\alpha = 0.66$. Considering Figure 9.20, one can also observe that the clustering result of this algorithm under threshold $\alpha = 0.66$ reaches the highest clustering accuracy. For $\alpha = 0.66$, the method clusters the 683 samples into two clusters and produces the fuzzy descriptions of C_1 the class *"benign"* and C_2 the class *"malignant"* as follows:

$$\begin{aligned}\zeta_{C_1} = {} & m_{5,1}m_{7,1} + m_{2,1}m_{5,1}m_{6,1} + m_{1,1}m_{3,1} + m_{1,1}m_{6,1}m_{7,1} + m_{1,1}m_{5,1} + m_{1,1}m_{6,1}m_{9,1} \\ & + m_{3,1}m_{8,1}m_{9,1} + m_{1,1}m_{2,1}m_{8,1} + m_{1,1}m_{7,1}m_{8,1} + m_{1,1}m_{8,1}m_{9,1} + m_{1,1}m_{2,1}m_{6,1} \\ & + m_{1,1}m_{6,1}m_{8,1} + m_{4,1}m_{5,1}m_{6,1} + m_{1,2}m_{7,1} + m_{3,1}m_{7,1} + m_{1,1}m_{4,1} + m_{1,2}m_{2,1}m_{6,1},\end{aligned}$$

$$\zeta_{C_2} = m_{2,2}m_{6,2} + m_{2,2}m_{3,2}m_{8,2} + m_{6,2}m_{7,2}m_{8,2} + m_{2,2}m_{6,1} + m_{4,2}m_{6,2}m_{8,2} + m_{2,2}m_{3,2}m_{9,1} + m_{4,2}m_{5,2}m_{6,2} + m_{2,2}m_{3,2}m_{7,2} + m_{2,2}m_{5,2}m_{8,2} + m_{1,2}m_{2,2} + m_{4,2}m_{6,2}m_{7,2} + m_{5,2}m_{6,2}m_{8,2} + m_{1,2}m_{3,2} + m_{1,2}m_{5,2}m_{8,2} + m_{2,2}m_{3,2}m_{5,2} + m_{2,2}m_{4,2}m_{5,2} + m_{3,2}m_{6,2} + m_{3,2}m_{7,2}m_{8,2} + m_{1,2}m_{7,2}m_{8,2} + m_{1,2}m_{6,2} + m_{3,2}m_{4,2} + m_{2,2}m_{7,2}m_{8,2}.$$

25 samples $x_2, x_4, x_{61}, x_{73}, x_{108}, x_{125}, x_{135}, x_{139}, x_{152}, x_{164}, x_{182}, x_{191}, x_{200}, x_{206}, x_{229}, x_{231}, x_{244}, x_{285}, x_{381}, x_{413}, x_{489}, x_{504}, x_{507}, x_{539}, x_{626}$ are incorrectly clustered comparing with the expected clusters: the samples $x_i, 1 \le i \le 444$ are in C_1, $x_i, 445 \le i \le 685$ are in C_2. The clustering accuracy is 96.34%. Figure 9.21 shows the membership functions of the fuzzy descriptions of the two clusters ζ_{C_1}, ζ_{C_2}.

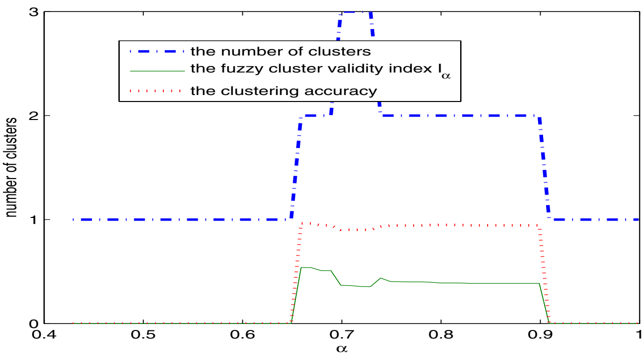


Fig. 9.20 The number of clusters, the clustering accuracy and the fuzzy cluster validity index I_α of our algorithm applying to breast cancer data versus the threshold α; the values of the threshold α are shown here in the [0.45, 0.95] interval

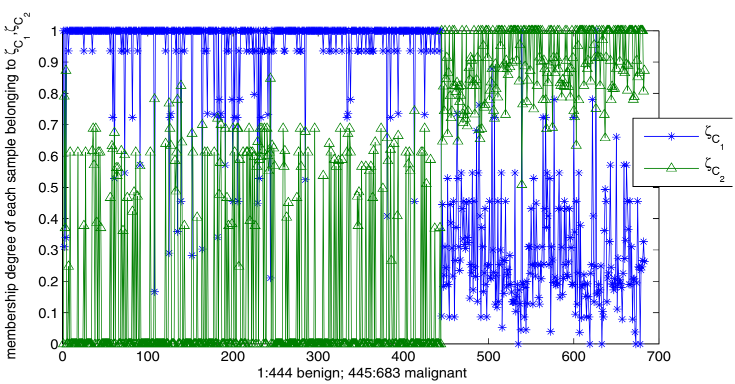


Fig. 9.21 Membership functions of the fuzzy descriptions of C_1, C_2 for breast cancer data

9.4.2.4 Comparative Analysis

In this section, the AFS fuzzy logic clustering algorithm is compared with the conventional algorithms: FCM [25], K-means [73], and some recently developed clustering algorithms: A Novel Kernel Method for Clustering [9], Mercer kernel based clustering algorithm [21], Support Vector Clustering [7], The Gath–Geva (GG) algorithm [22], the newer methods, two-level SOM-based clustering algorithm [79], Relative approach to hierarchical clustering [62], neural network for cluster-detection-and-labeling [16], Extended SOM (minimum variance) [30], the Gustafson–Kessel (GK) algorithm [19], Neural-Gas Network for Vector Quantization [63], the Ng-Jordan algorithm [66], generalization of learning vector quantization [68], General fuzzy min-max neural network [20], Mixture-based clustering [60], Maximum certainty data partitioning [69], General fuzzy min-max neural network [20], Self-Organized Formation of Feature Maps [31], Neural-Gas Network for Vector Quantization [63]. In Table 9.19, Table 9.20 and Table 9.21, the clustering accuracies of our algorithm reported for the three data sets already used in the previous experiments. From the comparative analysis, we conclude that the performance of the proposed algorithm is comparable with many other clustering algorithms. The interpretation facet of the algorithm is worth underlying here.

Through a thorough comparative analysis, one can observe that the AFS Fuzzy Clustering Algorithm Based on the $1/k-A$ Nearest Neighbors is comparable with the others in terms of the produced performance. Moreover, the fuzzy description

Table 9.19 A comparative analysis for some clustering algorithms – Iris data

Algorithm	Clustering accuracy (%)
Proposed fuzzy clustering algorithm	96–96.67
SOM-based clustering algorithm [79]	96
Mercer kernel based clustering algorithm [1]	98
Support Vector Clustering [7]	97.8
Relative approach to hierarchical clustering [62]	96–96.7
neural network for cluster-detection-and-labeling [16]	96
The Gath–Geva (GG) algorithm [22]	93.3
A Novel Kernel Method for Clustering [9]	94.7
Extended SOM (minimum variance) [30]	90.3
FCM [25]	89.3
K-Means [73]	89.3
the Gustafson–Kessel (GK) algorithm [19]	90
Neural-Gas Network for Vector Quantization [63]	91.7
the Ng-Jordan algorithm [66]	84.3
generalization of learning vector quantization. [68]	91.3
General fuzzy min-max neural network [20]	92–100
A relative approach to hierarchical clustering [62]	88.7–91.3
Maximum certainty data partitioning [69]	98
fully self-organizing simplified adaptive resonance theory [5]	69.1–95.5

Table 9.20 A comparative analysis for some clustering algorithms – Wine data

Algorithm	Clustering accuracy (%)
Proposed fuzzy clustering algorithm	96.63
SOM-based clustering algorithm [79]	98.3
Mixture-based clustering [60]	96.09
Extended SOM (minimum variance) [30, 79]	93.3
Maximum certainty data partitioning [69]	97.8
FCM [25]	94.9
neural network for cluster-detection-and-labeling [16]	91.57
Mercer kernel based clustering algorithm [21]	97.75
General fuzzy min-max neural network [20]	88.64–100

Table 9.21 A comparative analysis for some clustering algorithms – Wisconsin Breast Cancer Data

Algorithm	Clustering accuracy (%)
Proposed fuzzy clustering algorithm	96.34%
FCM [25]	95.6%
K-Means [73]	96.1%
Self-Organized Formation of Feature Maps [31, 9]	96.7%
Neural-Gas Network for Vector Quantization [63, 9]	96.1%
the Ng-Jordan algorithm [66, 9]	95.5%
A Novel Kernel Method for Clustering [9]	97%

$\zeta_C \in EM$ for each cluster C, which determines degree of each sample belong to the cluster C, has a sound interpretation with the simple concepts formed for the features.

Exercises

Exercise 9.1. Let X be a finite set and M be a set of simple concepts on X. Assume that $\Lambda \subseteq EM$. Prove that the following $(\Lambda)_{EI}$ is the sub algebra of EM generated by the fuzzy concepts in Λ.

$$(\Lambda)_{EI} = \left\{ \bigvee_{i\in I} (\bigwedge_{j\in J_i} a_{ij}) \;\middle|\; a_{ij} \in \Lambda, i \in I, j \in J_i, I \text{ and } J_i \text{ are any indexing sets} \right\}.$$

Exercise 9.2. Let X be a universe of discourse and M be a finite set of simple concepts. Let $\{\mu_\xi(x) \mid \xi \in EM\}$ be a set of coherence membership functions of the AFS fuzzy logic system $(EM, \vee, \wedge, ')$ and the AFS structure (M, τ, X). (seeing Definition 4.7). Let $\Lambda \subseteq EM$. Prove that for any $\beta \in (\Lambda)_{EI}$, for any $x \in X$, $\mu_\beta(x) \leq \mu_{\bigvee_{b\in\Lambda} b}(x)$.

Exercise 9.3. Let M be a set and EM be the EI algebra over M. Let A, B, C be any EI matrices with appropriate orders. Show the following assertions hold.

(1) $A(B+C) = AB + AC$;
(2) $A(BC) = (AB)C$.

Exercise 9.4. Replace the coherent membership functions and the AFS fuzzy logical operations in the algorithms used in this chapter with the "conventional" membership functions and the fuzzy logic equipped with t-norms and analyze the results.

Open problems

Problem 9.1. How to apply both $\overline{\zeta}_x$ and ζ_x defined by (9.9) and (9.10) as the fuzzy descriptions of $x \in X$ to design fuzzy clustering algorithm.

Problem 9.2. What is the relationship between Description Method A: description based on principal concepts and a category of concepts description and Method B: description based on the $1/k - A$-nearest neighbors?

Problem 9.3. The comparison and analysis of the three fuzzy cluster validity indices defined by (9.4), (9.15) and (9.40). What are relationships between them. How to improve these validity indices?

Problem 9.4. The convergence, consistency and stability of the AFS clustering fuzzy clustering algorithms presented in this chapter.

References

1. Bezdek, J.C.: Fuzzy mathematics in pattern classification. Ph.D. Dissertation, Appl. Math. Cornell Univ., Ithaca, NY (1973)
2. Bezdek, J.C.: A convergence theorem for the fuzzy ISODATA clustering algorithms. IEEE Trans. Pattern Anal. Machine Intell. 2(1), 1–8 (1980)
3. Blake, C.L., Men, C.J.: UCI Repository of machine learning databases, University of California, Imine, Dept. of Information and Computer Science (1998–2003), http://www.ics.uci.edu/~mlearn/MLRepository.html
4. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
5. Baraldi, A., Alpaydin, E.: Constructive feedforward ART clustering networks—Part I and II. IEEE Trans. Neural Netw. 13(3), 645–677 (2002)
6. Beliakov, G., King, M.: Density based fuzzy c-means clustering of non-convex patterns. European Journal of Operational Research 173, 717–728 (2006)
7. Ben-Hur, A., Horn, D., Siegelmann, H., Vapnik, V.: Support vector clustering. J. Mach. Learn. Res. 2, 125–137 (2001)
8. Blum, A., Langley, P.: Selection of Relevant Features and Examples in Machine Learning. Artificial Intelligence 97(1-2), 245–271 (1997)
9. Camastra, F., Verri, A.: A novel kernel method for clustering. IEEE Trans. Pattern Anal. Mach. Intell. 27(5), 801–805 (2005)
10. Chen, Y.L., Hu, H.L.: An overlapping cluster algorithm to provide non-exhaustive clustering. European Journal of Operational Research 173, 762–780 (2006)

11. Cord, A., Ambroise, C., Cocquerez, J.: Feature selection in robust clustering based on Laplace mixture. Pattern Recognition Letters 27, 627–635 (2006)
12. Dunn, J.C.: A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J. Cybernet 3(3), 32–57 (1974)
13. Dubois, D., Prade, H.: The three semantics of fuzzy sets. Fuzzy Sets and Systems 90, 141–150 (1997)
14. Ding, R., Liu, X.D., Chen, Y.: The Fuzzy Clustering Algorithm Based on AFS Topology. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS (LNAI), vol. 4223, pp. 89–98. Springer, Heidelberg (2006)
15. El-Sonbaty, Y., Ismail, M.A.: Fuzzy Clustering for Symbolic Data. IEEE Transactions on Fuzzy Systems 6(2), 195–204 (1998)
16. Eltoft, T., de Figueiredo, R.: A new neural network for cluster-detection-and-labeling. IEEE Trans. Neural Netw. 9(5), 1021–1035 (1998)
17. Fortuna, J., Capson, D.: Improved support vector classification using PCA and ICA feature space modification. Pattern Recognition 37, 1117–1129 (2004)
18. Graver, J.E., Watkins, M.E.: Combinatorics with Emphasis on the Theory of Graphs. Springer, New York (1977)
19. Gustafson, D.E., Kessel, W.: Fuzzy clustering with a fuzzy covariance matrix. In: Proc. IEEE Conf. Decision Control, San Diego, CA, pp. 761–766 (1979)
20. Gabrys, B., Bargiela, A.: General fuzzy min-max neural network for clustering and classification. IEEE Trans. Neural Netw. 11(3), 769–783 (2000)
21. Girolami, M.: Mercer kernel based clustering in feature space. IEEE Trans. Neural Netw. 13(3), 780–784 (2002)
22. Gath, I., Geva, A.B.: Unsupervised optimal fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell. 11(7), 773–781 (1989)
23. Graver, J.E., Watkins, M.E.: Combinatorics with Emphasis on the Theory of Graphs. Springer, New York (1977)
24. Gowda, K.C., Diday, E.: Symbolic clustering using a new dissimilarity measure. Pattern Recogn 24(6), 567–578 (1991)
25. Hathaway, R., Bezdek, J.: Fuzzy c-means clustering of incomplete data. IEEE Trans. Syst., Man, Cybern. 31(5), 735–744 (2001)
26. Halmos, P.R.: Measure theory. Springer, New York (1974)
27. Jain, A., Zongker, D.: Feature Selection: Evaluation, Application, and Small Sample Performance. IEEE Trans. Pattern Analysis and Machine Intelligence 19(2), 153–157 (1997)
28. Jenssen, R., Erdogmus, D.: Information cut for clustering using a gradient descent approach. Pattern Recognition 40, 796–806 (2007)
29. Keller, J.M., Gray, M.R., Givens Jr., J.A.: A fuzzy K-nearest neighbors algorithm. IEEE Trans. Systems Man, Cybernet. SMC 15(4), 580–585 (1985)
30. Kiang, M.Y.: Extending the Kohonen self-organizing map networks for clustering analysis. Comput. Stat. Data Anal. 38, 161–180 (2001)
31. Kohonen, T.: Self-Organized Formation of Topologically Correct Feature Maps. Biological Cybernetics 43(1), 59–69 (1982)
32. Kim, K.H.: Boolean matrix Theory and Applications. Marcel Dekker, New York (1982)
33. Kirby, M.: Data Analysis: An empirical approach to dimensionality reduction and the study of Pattern. Wiley, New York (2000)
34. Kohavi, R., John, G.: Wrappers for Feature Subset Selection. Artificial Intelligence 97(1-2), 273–324 (1997)
35. Koller, D., Sahami, M.: Toward Optimal Feature Selection. In: Proc. 13th Int'l Conf. Machine Learning, pp. 284–292 (1996)

36. Liu, X.D.: The Fuzzy Theory Based on AFS Algebras and AFS Structure. Journal of Mathematical Analysis and Applications 217, 459–478 (1998)
37. Liu, X.D.: The Topology on AFS Algebra and AFS Structure. Journal of Mathematical Analysis and Applications 217, 479–489 (1998)
38. Liu, X.D.: A new fuzzy model of pattern recognition and hitch diagnoses of complex systems. Fuzzy Sets and Systems 104, 289–297 (1999)
39. Liu, X.D.: A New Mathematical Axiomatic System of Fuzzy Sets and Systems. Journal of Fuzzy Mathematics 3, 559–560 (1995)
40. Liu, X.D.: The Fuzzy Sets and Systems Based on AFS Structure, EI Algebra and EII algebra. Fuzzy Sets and Systems 95, 179–188 (1998)
41. Liu, X.D.: Two Algebra Structures of AFS structure. Journal of Fuzzy Mathematics 3, 561–562 (1995)
42. Liu, X.D., Pedrycz, W., Zhang, Q.L.: Axiomatics Fuzzy sets logic. In: IEEE International Conference on Fuzzy Systems, vol. 1, pp. 55–60 (2003)
43. Liu, X.D., Pedrycz, W.: The Development of Fuzzy Decision Trees in the Framework of Axiomatic Fuzzy Set Logic. Applied Soft Computing 7, 325–342 (2007)
44. Liu, X.D., Wang, W., Chai, T.Y.: The Fuzzy Clustering Analysis Based on AFS Theory. IEEE Transactions on Systems, Man and Cybernetics Part B 35(5), 1013–1027 (2005)
45. Liu, X.D., Liu, W.Q.: Credit rating analysis with AFS fuzzy logic. In: Wang, L., Chen, K., S. Ong, Y. (eds.) ICNC 2005. LNCS, vol. 3612, pp. 1198–1204. Springer, Heidelberg (2005)
46. Liu, X.D., Chai, T.Y., Wang, W.: AFS Fuzzy Logic Systems and Its Applications to Model and Control. International Journal of Information and Systems Sciences 2(3), 285–305 (2006)
47. Liu, X.D.: The Development of AFS Theory Under Probability Theory. International Journal of Information and Systems Sciences 3(2), 326–348 (2007)
48. Liu, X.D., Pedrycz, W., Chai, T.Y., Song, M.L.: The Development of Fuzzy Rough Sets with the Use of Structures and Algebras of Axiomatic Fuzzy Sets. IEEE Transactions on Knowledge and Data Engineering 21(3), 443–462 (2009)
49. Liu, X.D., Zhu, K.J., Huang, H.Z.: The Representations of Fuzzy Concepts Based on the Fuzzy Matrix Theory and the AFS Theory. In: IEEE International Symposium on Intelligent Control, Texas, USA, October 5-8 (2003)
50. Liu, X.D.: The Structure of Fuzzy Matrices. Journal of Fuzzy Mathematics 2, 311–325 (1994)
51. Liu, X.D., Zhang, Q.L.: The Fuzzy Cognitive Maps Based on AFS Fuzzy Logic. Dynamics of Continuous, Discrete and Impulsive Systems 11(5-6), 787–796 (2004)
52. Liu, X.D., Chai, T.Y., Wang, W., Liu, W.Q.: Approaches to the Representations and Logic Operations for Fuzzy Concepts in the Framework of Axiomatic Fuzzy Set Theory I. Information Sciences 177, 1007–1026 (2007)
53. Liu, X.D., Chai, T.Y., Wang, W., Liu, W.Q.: Approaches to the Representations and Logic Operations for Fuzzy Concepts in the Framework of Axiomatic Fuzzy Set Theory II. Information Sciences 177, 1027–1045 (2007)
54. Liu, X.D., Ren, Y.: Novel Artificial Intelligent Techniques via AFS Theory: Feature Selection, Concept Categorization and Characteristic Description. Applied soft Computting (submitted)
55. Liu, X.D., Wang, X.C., Pedrycz, W., Chai, T.Y.: The Fuzzy Clustering via Axiomatic Fuzzy Set Theory. Pattern Recognition (submitted)
56. Liang, G.S., Chou, T.Y., Han, T.C.: Cluster Analysis Based on Fuzzy Equivalence Relation. European Journal of Operational Research 166, 160–171 (2005)

57. Lee, H., Park, K., Bien, Z.: Iterative fuzzy clustering algorithm with supervision to construct probabilistic fuzzy rule base from numerical data. IEEE Transactions on Fuzzy Systems 16, 263–277 (2008)
58. Li, Y., Dong, M., Hua, J.: Localized feature selection for clustering. Pattern Recognition Letters 29, 10–18 (2008)
59. Law, M., Figueiredo, M., Jain, A.K.: Simultaneous Feature Selection and Clustering Using Mixture Models. IEEE Trans. Pattern Anal. Mach. Intell. 26(9), 124–138 (2004)
60. Law, M.H.C., Figueiredo, M.A.T., Jain, A.K.: Simultaneous Feature Selection and Clustering Using Mixture Models. IEEE transactions on pattern analysis and machine intelligence 26(9), 1154–1166 (2004)
61. Merz, C.J., Murphy, P.M.: UCI Repository for Machine Learning Data-Bases. Dept. of Information and Computer Science, University of California, Irvine, CA (1996), http://www.ics.uci.edu/~mlearn/MLRepository.html
62. Mollineda, R., Vidal, E.: A relative approach to hierarchical clustering, in Pattern Recognition and Applications. In: Torres, M., Sanfeliu, A. (eds.) Frontiers in Artificial Intelligence and Applications, vol. 56, pp. 19–28. IOS Press, Amsterdam (2000)
63. Martinetz, T.E., Schulten, K.J.: Neural-Gas network for vector quantization and its application to time-series prediction. IEEE Trans. Neural Networks 4(4), 558–569 (1993)
64. Mitchell, T.M.: Machine Learning. China Machine Press (2003)
65. Marcelloni, F.: Feature selection based on a modified fuzzy C-means algorithm with supervision. Information Sciences 151, 201–226 (2003)
66. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: Analysis and an algorithm. Advances in Neural Information Processing Systems 14, 849–856 (2001)
67. Pedrycz, W.: Statistically grounded logic operators in fuzzy sets. European Journal of Operational Research (2008) doi:10.1016/j.ejor.2007.12.009
68. Pal, N., Bezdek, J., Tsao, E.: Generalized clustering networks and Kohonen's self-organizing scheme. IEEE Trans. Neural Netw. 4(4), 549–557 (1993)
69. Roberts, S.J., Everson, R., Rezek, I.: Maximum certainty data partitioning. Pattern Recognition 33, 833–839 (2000)
70. Ren, Y., Song, M.L., Liu, X.D.: New approches to the fuzzy clustering via AFS theory. International Journal of Information and Systems Science 3(2), 307–325 (2007)
71. Ren, Y., Wang, X.C., Liu, X.D.: Fuzzy clustering approaches based on AFS fuzzy logic I. In: Proceedings of the 6th World Congress on Control and Automation, Dalian, China, June 21 - 23 (2006)
72. Ren, Y.: Feature Selection, Concept Categorization and Characteristic Description Based on AFS Theory, Master Degree Thesis, Dalian Maritime university (2006)
73. Spath, H.: Cluster Dissection and Analysis: Theory FORTRAN Programs, Examples, translated by Goldschmidt, J., p. 226. Halsted Press, New York (1985)
74. Wu, K.L., Yang, M.S.: Alternative c-means clustering algorithms. Pattern Recognition 35, 2267–2278 (2002)
75. Wang, G.J.: Theory of topological molecular lattices. Fuzzy Sets and Systems 47, 351–376 (1992)
76. Wang, X.C.: The Fuzzy Clustering via Axiomatic Fuzzy Set Theory, Master Degree Thesis, Dalian Maritime university (2006)
77. Wang, L. D., Liu, X. D.: Concept analysis via rough set and AFS algebra. Information Sciences 178, 4125–4137 (2008)
78. Wang, G.j.: Theory of topological molecular lattices. Fuzzy Sets Systems 47, 351–376 (1992)
79. Wu, S.T., Chow, T.W.S.: Clustering of the Self-organizing Map Using a Clustering Validity Index Based on Inter-cluster and Intra-cluster Density. Pattern Recognition 37, 175–188 (2004)

80. Yang, X., Song, Q., Wu, Y.: A robust deterministic annealing algorithm for data clustering. Data & Knowledge Engineering 62, 84–100 (2007)
81. Zadeh, L.A.: Fuzzy Sets. Information and Control 8, 338–353 (1965)
82. Zhang, L.S., Liu, X.D.: Concept lattice and AFS algebra. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS (LNAI), vol. 4223, pp. 290–299. Springer, Heidelberg (2006)
83. Zhang, Y.J., Liang, D.Q., Tong, S.C.: On AFS algebra part I. Information Sciences 167, 263–286 (2004)
84. Zhang, Y.J., Liang, D.Q., Tong, S.C.: On AFS algebra part II. Information Sciences 167, 287–303 (2004)

Chapter 10
AFS Fuzzy Classifiers

In this chapter, we introduce three design strategies of classifiers which exploit the unified usage of the AFS fuzzy logic, entropy measures and decision trees. The advantage of these classifiers is two-fold. First, they can mimic the human reasoning and in this manner offer a far more transparent and comprehensible way supporting the design of the classifiers. An important aspect is concerned with the simplicity of the design methodology and the clarity of the underlying semantics. We use three well known data to illustrate the effectiveness of the classifiers and present the relationship between the parameters of the classifiers and their performance.

We are faced with a genuine abundance of data. However, knowledge representation and information acquisition from huge data constitute a serious bottleneck present in numerous engineering applications. Extracting useful information via designing effective classifiers becomes of paramount relevance. There are many approaches available aimed at alleviating this bottleneck including fuzzy sets [116]. The incorporation of fuzzy sets into the representation of fuzzy concepts enables us to combine the uncertainty handling and approximate reasoning capabilities with the comprehensibility associated with fuzzy sets [37].

Classification becomes important in a variety of fields, such as pattern recognition, artificial intelligence, and computer vision.

In this chapter, we propose a new framework for the design of fuzzy classifiers in which the AFS logic plays an essential role. The illustrative numeric examples involve the Wine, Iris and Breast Cancer data coming from the machine-learning database at the University of California, Irvine [67]. They help us reveal and emphasize the main advantages of the proposed classification environment. Actually, compared to other fuzzy classifiers, the proposed classifier comes with several advantages:

1. The proposed classifiers can be applied to high-dimensional problems without suffering from the curse of dimensionality.
2. The design of the proposed classifiers, in which each class is represented by a fuzzy set in the *EM* and the degree of the new sample belonging to the class is determined by it, is linguistically interpretable, comprehensible and similar to the classification schemes exercised by humans. Because each fuzzy set in the *EM*

X. Liu and W. Pedrycz: Axiomatic Fuzzy Set Theory and Its Applications, STUDFUZZ 244, pp. 423–489.
springerlink.com

representing a class is semantically sound in the AFS fuzzy logic, the linguistic interpretation of the proposed classifiers is viable.

3. The design of the proposed classifier is valid without making the assumptions about data set $X \subset R^{p\times n}$. This is possible owing to the fact that the AFS structure (M, τ, X) can be established for database with any types of attributes, even linguistic descriptions based on human intuition.
4. Since the new classifiers are simply built-up by the AFS logic on the fuzzy concepts in EM and amended separately by adding the information of each training sample, hence the classifiers can be designed and implemented by parallel processing and tuned online.

10.1 Classifier Design Based on AFS Fuzzy Logic

First, let us elaborate on how a human being can classify data with resorting to some training samples. Let X be a set of training samples and M be a set of simple concepts on X. Let $\Gamma \subseteq EM$ be the set of all relative fuzzy concepts. Suppose that X is classified into l classes $X_1, X_2, ..., X_l$. In order to characterize each class X_i based on the training samples, for each $x \in X_i$, we intend to find a fuzzy set $\zeta_x \in \Gamma$, such that at the largest degree x belongs to ζ_x, while for any $y \in X - X_i$, at smallest degree y belongs to ζ_x and for $z \in X_i$, $z \neq x$, at comparatively larger degree z belongs to ζ_x. In other words, x can be distinguished from any $y \in X - X_i$ by the fuzzy concept ζ_x at the maximal extent. Finally, the fuzzy set $\zeta_{X_i} = \bigvee_{x\in X_i} \zeta_x$ forms the fuzzy description of class i. For each new sample s, the degree of s belongingness (membership) to class i is $\mu_{\zeta_{X_i}}(s), i = 1, 2, ..., l$.

The above process is convincing. However, it still be facing the computational bottleneck as the number of fuzzy sets in EM is very large. In order to solve this problem in a tractable yet approximate manner, we introduce the following concepts.

Let X be the universe of discourse and M be a set of simple concepts and (M, τ, X) be an AFS structure. $\Lambda \subseteq EM$, Λ is the set of some fuzzy sets which are relatively important (essential) fuzzy features required to consider in the context of the classification problem. The fuzzy description of each element $x \in X$ in the setting of the fuzzy feature set Λ is a fuzzy set $\zeta_x \in (\Lambda)_{EI}$ such that the degree of x membership to ζ_x is the largest among other fuzzy sets in $(\Lambda)_{EI}$ while for $y \in X$, $y \neq x$, the degree of y belonging to ζ_x is quite small. In other words, by using the fuzzy concept $\zeta_x \in (\Lambda)_{EI}$, x can be distinguished among other elements in X to the maximal extent. In what follows, we will study the fuzzy description for each $x \in X$.

Now we discuss how to take advantage of Theorem 9.1 to derive the fuzzy descriptions of the samples for the classifiers. For each $\alpha \in \Lambda_x^{\varepsilon}$, since ϑ_α^x is a sub-EI algebra of $(\Lambda)_{EI}$, hence $\bigwedge_{b\in\vartheta_\alpha^x} b \in (\Lambda)_{EI}$. $\bigwedge_{b\in\vartheta_\alpha^x} b$ is a fuzzy description of x based on molecular element $\alpha \in \Lambda_x^{\varepsilon}$. It is obvious that $\overline{\zeta}_x = \bigvee_{\alpha\in\Lambda_x^{\varepsilon}} (\bigwedge_{b\in\vartheta_\alpha^x} b) \in (\Lambda)_{EI}$. This implies $\overline{\zeta}_x$ is a fuzzy set in $(\Lambda)_{EI}$ and $\zeta_x = \bigvee_{b\in\Lambda_x^{\varepsilon}} b$ is the fuzzy description of x by the molecular elements of the fuzzy sets in $(\Lambda)_{EI}$. By $\overline{\zeta}_x \geq \zeta_x$, we know that $\overline{\zeta}_x$ forms a rougher description of x than ζ_x. But we should notice that in some cases, $\zeta_x \notin (\Lambda)_{EI}$. By (4) of Theorem 9.1, one knows that for $\alpha \in \Lambda_x^{\varepsilon}$, the fuzzy set $\bigwedge_{b\in\vartheta_\alpha^x} b$ can guarantee that for any $y \neq x, y$ belongs to $\bigwedge_{b\in\vartheta_\alpha^x} b$ not at maximum degree while x belongs to $\bigwedge_{b\in\vartheta_\alpha^x} b$ to maximal degree. This implies that both $\overline{\zeta}_x$ and ζ_x are fuzzy descriptions of x. We should notice that $\overline{\zeta}_x = \zeta_x$, if for each $\alpha \in \Lambda$, it is a molecular element of $(EM, \vee, \wedge)$, i.e., $\alpha = A, A \subseteq M$.

With these theoretical analysis, we can now describe the fuzzy classification algorithm. Let X be a set of training samples with n features to describe the samples and M be a set of simple concepts formed on these features. The training samples are labeled and belong to c classes, which are X_1, X_2, ..., X_c, i.e., $X = \bigcup_{1\le i\le c} X_i,\ X_i \cap X_j = \varnothing,\ i \neq j$. Let (M, τ, X) be an AFS structure of the data. Let ρ_ν be the weight function of the simple concept $\nu \in M$ and $\{\mu_\xi(x) \mid \xi \in EM\}$ be a set of coherence membership functions of the AFS fuzzy logic system $(EM, \vee, \wedge, ')$ and the AFS structure (M, τ, X), defined by (4.29) of Theorem 4.5 in which the measure M_ν derived via (4.35) for the weight function of the simple concept ρ_ν. $\Lambda \subseteq EM$, Λ is a set of fuzzy concepts which are selected to design the classifier.

10.1.1 Classifier Design Based on AFS Fuzzy Logic

In this section, we describe the procedure of the classifier design based on AFS fuzzy logic.

Step 1: Select $\Lambda \subseteq EM$, a set of fuzzy concepts used to design the classifier. $(\Lambda)_{EI}$ is the sub EI algebra generated by Λ.

Step 2: Given small positive numbers $\varepsilon > 0, \delta > 0$, for each $i = 1, 2, ..., c$, find the fuzzy set $\zeta_{X_i} \in (\Lambda)_{EI}$, such that

$$\zeta_{X_i} = \arg\max_{\xi \in F_\varepsilon^\delta} \{\sum_{y\in X_i} \mu_\xi(y)\}, \tag{10.1}$$

where

$$E_\Lambda^\delta = \{\gamma | \gamma \in (\Lambda)_{EI}, \forall y \in X - X_i, \mu_\gamma(y) < \delta\}, \tag{10.2}$$

$$F_\varepsilon^\delta = \{\xi | \xi \in E_\Lambda^\delta, \forall y \in X_i, \mu_\xi(y) \ge \mu_{\bigvee_{b\in\Lambda} b}(y) - \varepsilon\}. \tag{10.3}$$

ζ_{X_i} is the fuzzy description for class X_i, $i = 1, 2, ..., c$. δ is a parameter to control the extent of fuzzy set ζ_{X_i} distinguishing $x \in X_i$ and $y \notin X_i$. ε is a parameter to control the degree of each training sample x ($x \in X_i$) belonging to ζ_{X_i}.

Step 3: For each testing sample s, we estimate the degree of s belonging to the fuzzy set ζ_{X_i}, $i = 1, 2, ..., c$. Although $s \notin X$, we can estimate the degree of s belonging to the fuzzy set ζ_{X_i} by the AFS structure (M, τ, X). For each simple concept $m \in M$, with the weighting function ρ_m, one can derive $\rho_m(s) \in R^+$. More specifically, for each $A \subseteq M$, let us define

$$\underline{A}(\{s\}) = \{x | x \in X, \forall m \in A, \rho_m(x) \le \rho_m(s)\},$$

$$\overline{A}(\{s\}) = X - \{x | x \in X, \forall m \in A, \rho_m(s) < \rho_m(x)\}.$$

For any fuzzy concept $\sum_{i\in I}(\prod_{m\in A_i} m) \in EM$, the *EII* algebra degree of s belonging to the fuzzy concept $\sum_{i\in I}(\prod_{m\in A_i} m)$ is $(\sum_{i\in I}\prod_{m\in A_i} m)(s)$ defined by (5.10), then the lower bound and upper bound of $(\sum_{i\in I}\prod_{m\in A_i} m)(s)$ in lattice EXM are $\sum_{i\in I}\underline{A}_i(\{s\})A_i$ and $\sum_{i\in I}\overline{A}_i(\{s\})A_i$, respectively. According to Theorem 4.5, for $\mu_{\sum_{i\in I} A_i}(s)$, the membership degree of s belonging to the fuzzy set $\sum_{i\in I}\prod_{m\in A_i} m$, we have

$$\sup_{i\in I}\prod_{\nu\in A_i}\frac{\mathscr{M}_\nu(\underline{A}_i(\{s\}))}{\mathscr{M}_\nu(X)} \leq \mu_{\sum_{i\in I} A_i}(s) \leq \sup_{i\in I}\prod_{\nu\in A_i}\frac{\mathscr{M}_\nu(\overline{A}_i(\{s\}))}{\mathscr{M}_\nu(X)}.$$

Here, we simply let

$$\mu_{\zeta_{X_i}}(s) = \sup_{i\in I}\prod_{\nu\in A_i}\frac{\mathscr{M}_\nu(\underline{A}_i(\{s\}))}{\mathscr{M}_\nu(X)} \tag{10.4}$$

to be the membership degree of s belonging to class $i, i = 1,2,...,c,$ where $\zeta_{X_i} \in EM$. For each testing sample s, it belongs to the class i if $i = \arg\max_{1\leq k\leq c}\{\mu_{\zeta_{X_k}}(s)\}$. Therefore, the designed classifier is actually of fuzzy representations of each class which can be denoted as $\mathscr{C} = (\zeta_{X_1}, \zeta_{X_2}, ..., \zeta_{X_c})$, where $\zeta_{X_i} \in EM$.

In theory, we can apply the probability distribution of the observed data X and the membership functions of fuzzy concepts in EM determined by (4.41) in Theorem 4.6 to classify any sample in the whole space. However, we have to face the complex computation of the high dimension integral in formula (4.41). Thus how to apply it to the classification remains an open problem.

It should be noted that the optimization problem in **Step 2** is very computationally demanding since it is based on the set $(\Lambda)_{EI}$, which may have a large number of elements. In order to overcome this difficulty, we can solve the problem based on the sets ϑ_α^x and Λ_x^ε according to Theorem 9.1, which will be detailed in the next section along with an illustrative example. In this case, we will obtain an approximate solution for the optimization problem present in **Step 2**. The primordial issue to obtain the approximate solution is to compute Λ_x^ε, which can be described below.

Suppose that for each $\alpha \in \Lambda$, it is a molecular element. Then for $\varepsilon > 0, x \in X$, one can follow the two steps to compute Λ_x^ε.

STEP 1: For each $\alpha \in \Lambda$ check if $\mu_\alpha(x) \geq \mu_{\bigvee_{b\in\Lambda} b}(x) - \varepsilon$ and obtain B_x^ε.
STEP 2: For each $\alpha \in B_x^\varepsilon$, with the decision tree algorithm C4.5 [65], one can find a tree with α being its root and its nodes are some other elements in B_x^ε obtained as following. $\beta \in B_x^\varepsilon$ is a child of root α if $\mu_{\alpha\wedge\beta}(x) \geq \mu_{\bigvee_{b\in\Lambda} b}(x) - \varepsilon$. For $\gamma \in B_x^\varepsilon$ a node of the tree, $\nu \in B_x^\varepsilon$ is a child of node γ if $\mu_{\zeta_\gamma\wedge\nu}(x) \geq \mu_{\bigvee_{b\in\Lambda} b}(x) - \varepsilon$, where

$$\zeta_\gamma = \bigwedge\{\varsigma \in B_x^\varepsilon \mid \varsigma \text{ is a node in the path from the tree root } \alpha \text{ to } \gamma\}$$

Thus, for each $\alpha \in B_x^\varepsilon$, it corresponds to a tree with the fuzzy operation "*and*". Therefore the "*and*" operations of all nodes in a path from the tree root to a leaf (including the root and the leaf) is an element in Λ_x^ε.

In practice, once we have obtained Λ_x^ε, the number of elements in ϑ_α^x is usually small and we can easily compute the fuzzy sets $\overline{\zeta}_x$ or ζ_x

$$\overline{\zeta}_x = \bigvee_{\alpha \in \Lambda_x^\varepsilon} \left(\bigwedge_{b \in \vartheta_\alpha^x} b \right) \tag{10.5}$$

or

$$\zeta_x = \bigvee_{b \in \Lambda_x^\varepsilon} b. \tag{10.6}$$

Then we have the approximate solution of (10.1) as follows:

$$\zeta_{X_i} = \bigvee_{x \in X_i} \overline{\zeta}_x \tag{10.7}$$

or

$$\zeta_{X_i} = \bigvee_{x \in X_i} \zeta_x. \tag{10.8}$$

10.1.2 Experimental Results

In this section we show how the AFS fuzzy logic can be used in data classification making use of the data. The wine data contains thirteen continuous attributes are measured for each wine: alcohol content (AL), malic acid content (MAC), ash content, alkalinity of ash (AA), magnesium content (MA), total phenols (TP), flavonoids (FL), nonflavanoids phenols (NFP), proanthocyaninsm (PR), color intensity (CI), hue, OD280/OD315 (O) of diluted wines, and praline (P). The number of patterns in the three classes is 59, 71, and 48, respectively. The wine data has been widely used to test the performance of various classifiers [2, 86]. Here we randomly select 106 samples to form a training set (training samples) to design the classifier and the 72 patterns are left to form the testing samples. We randomly select 60% samples as training examples from each class, i.e., class 1: 36, class 2: 43, class 3: 27. We present the data in a compact matrix form of $W = (w_{ij})_{178\times 14}$, where the first column is the class label of each sample and the i-th column is the value of each sample on feature i. Then the matrix $N = (n_{ij})_{178\times 14}$ is obtained by normalizing W, where

$$n_{ij} = \frac{w_{ij}}{\max_{1\le k\le 178}\{|w_{kj}|\}}, i = 1,2,...,178, j = 2,3,...,14.$$

Let X be the set of the 106 randomly selected training samples, i.e., we randomly select 36 rows, 43 rows, 27 rows form 1 to 59, 60 to 130, 131 to 178 row of matrix N, respectively while the indexes of the selected rows of N are:

Class 1: 2, 3, 4, 7, 8, 10, 11, 15, 16, 19, 21, 22, 24, 25, 26, 27, 28, 29, 30, 31, 33, 35, 36, 38, 39, 42, 44, 46, 48, 49, 50, 51, 53, 54, 55, 57;

Class 2: 60, 61, 67, 68, 69, 70, 72, 74, 75, 77, 78, 80, 83, 85, 87, 88, 91, 92, 94, 95, 97, 98, 99, 101, 102, 103, 104, 105, 106, 107, 109, 111, 112, 115, 116, 118, 119, 120, 121, 122, 124, 129, 130;

Class 3: 131, 133, 136, 138, 139, 142, 145, 146, 149, 151, 152, 153, 154, 155, 156, 158, 159, 162, 163, 164, 166, 167, 171, 172, 173, 174, 175;

For each $s = (s_1, s_2, ..., s_{14}) \in X$, $s_1 = 1, 2$, or 3 is the class label of sample s, i.e., X is divided into 3 classes X_1, X_2, X_3 and if $s \in X_i$, then $s_1 = i$. s_i is the value of sample s on feature $i, i = 2, 3, ..., 14$. For each feature i, we define 6 simple concepts.

"*about the center of class 1 for feature i*";
"*about the center of class 2 for feature i*";
"*about the center of class 3 for feature i*";
"*not about the center of class 1 for feature i*";
"*not about the center of class 2 for feature i*";
"*not about the center of class3 for feature i*";

Now it is ready to design the classifier. We show the design process step by step.

Step 1: Let $M = \{m_1, m_2, ..., m_{84}\}$ be the set of the $14 \times 6 = 84$ simple concepts on the 14 features where $m_1, m_2, m_3, m_{84}, m_{83}, m_{82}$ are Boolean concepts for feature 1: {class1, class2, class3, not-class1, not-class2, not-class3}. For each $m_k \in M$, $3 < k \leq 42$, $k = 3 \times q + r$ $(0 < r \leq 3)$, we define the weighting function $\rho_{m_k} : X \to R^+ = [0, \infty)$ as follows: for any $s = (s_1, s_2, ..., s_{14}) \in X$,

$$\rho_{m_k}(s) = \exp(-\frac{(s_i - \mu_{ir})^2}{2\sigma_{ir}^2}), \tag{10.9}$$

where $i = q + 1$, μ_{ir} is the mean of s_i for total $s = (s_1, s_2, ..., s_{14}) \in X_r$, i.e., the center of class r on feature i and σ_{ir} is the standard deviation of s_i for the total $s = (s_1, s_2, ..., s_{14}) \in X_r, r = 1, 2, 3$. For $k = 1, 2, 3$, $\rho_{m_k}(s) = 1 \Leftrightarrow s$ belongs to class k. If $k > 42$, then $\rho_{m_k} = 1 - \rho_{m_{(85-k)}}$. The AFS structure (M, τ, X) is established in the following manner: for any $x, y \in X$, $x \neq y$

$$\begin{aligned} \tau(x,x) &= \{m \mid m \in M, \rho_m(x) > 0\}, \\ \tau(x,y) &= \{m \mid m \in M, \rho_m(x) \geq \rho_m(y)\}. \end{aligned}$$

We can verify that τ satisfies AX1, AX2 of Definition 4.5 and (M, τ, X) is an AFS structure. Let $\{\mu_\xi(x) \mid \xi \in EM\}$ be a set of coherence membership functions of the AFS fuzzy logic system $(EM, \vee, \wedge, ')$ and the AFS structure (M, τ, X), defined by (4.29) of Theorem 4.5 in which the measure M_v derived via (4.35) for the weight function ρ_v defined by (10.9).

Step 2: For each simple concept $m \in M - \{m_1, m_2, m_3, m_{84}, m_{83}, m_{82}\}$, it is a fuzzy set describing the center of class r on some features. Hence any fuzzy set in $(\Lambda)_{EI}$ will have impact on the classification, where

$$\Lambda = \{m \mid m \in M - \{m_1, m_2, m_3, m_{84}, m_{83}, m_{82}\}\} \subset EM.$$

Because Boolean concepts $m_1, m_2, m_3, m_{84}, m_{83}, m_{82}$ are included only the label information, they are not included in Λ. Select $\Lambda \subseteq EM$ and Λ is a set of fuzzy concepts selected to design the classifier. Denote $(\Lambda)_{EI}$ as the sub EI algebra generated by Λ.

Step 3: Let $\varepsilon = 0.3, \delta = 0.1$. For each $i = 1,2,3$, find the fuzzy set $\zeta_{X_i} \in (\Lambda)_{EI}$ such that

$$\zeta_{X_i} : \sum_{y \in X_i} \mu_{\zeta_{X_i}}(y) = \max_{\xi \in F_\varepsilon^\delta} \{ \sum_{y \in X_i} \mu_\xi(y) \},$$

where

$$E_\Lambda^\delta = \{\gamma \mid \gamma \in (\Lambda)_{EI}, \forall y \in X - X_i, \mu_\gamma(y) < \delta\},$$
$$F_\varepsilon^\delta = \{\xi \mid \xi \in E_\Lambda^\delta, \forall y \in X_i, \mu_\xi(y) \geq \mu_{\bigvee_{b \in \Lambda} b}(y) - \varepsilon\}.$$

ζ_{X_i} is the fuzzy description for class X_i.

Step 4: For each testing sample s, $\mu_{\eta_{X_i}}(s)$ is the membership degrees of s belonging to class i, $i = 1,\ 2,\ 3$ and calculated by (10.4).

In the above Step 3, since there are a great number of different fuzzy concepts in $(\Lambda)_{EI}$, hence it is very difficult to find ζ_{X_i} which satisfies (10.1), (10.2), (10.3) by checking each elements in $(\Lambda)_{EI}$. In what follows, we propose an algorithm to find a fuzzy set, which can approximate ζ_{X_i}. For each $x \in X_i$, by (10.3), we have

$$\mu_{\zeta_{X_i}}(x) \geq \mu_{\bigvee_{b \in \Lambda} b}(x) - \varepsilon.$$

Therefore by (4) of Theorem 9.1, $\exists \alpha \in \Lambda_x^\varepsilon$, for any $y \in X$ we get,

$$\mu_{\zeta_{X_i}}(y) \geq \mu_{\bigwedge_{b \in \vartheta_\alpha^x} b}(y) \geq \mu_\alpha(y).$$

Actually, in this example the number of elements in ϑ_α^x is very small. Thus we can determine a fuzzy set $\zeta_x \in \bigcup_{\alpha \in \Lambda_x^\varepsilon} \vartheta_\alpha^x$ such that

$$\zeta_x \ : \ \sum_{y \in X_i} \mu_{\zeta_x}(y) = \max_{\xi \in E_\Lambda^\delta} \{ \sum_{y \in X_i} \mu_\xi(y) \}, \tag{10.10}$$
$$E_\Lambda^\delta = \{\gamma \mid \gamma \in \bigcup_{\alpha \in \Lambda_x^\varepsilon} \vartheta_\alpha^x,\ \ \forall y \in X - X_i, \mu_\gamma(y) < \delta\} \tag{10.11}$$

By (3) of Theorem 9.1, we know that for any $\eta \in \bigcup_{\alpha \in \Lambda_x^\varepsilon} \vartheta_\alpha^x$, $\mu_\eta(x) \geq \mu_{\bigvee_{b \in \Lambda} b}(x) - \varepsilon$. For each training sample x and $\alpha \in \Lambda_x^\varepsilon$, ζ_x can be obtained by parallel processing. This implies that we can efficiently obtain $\eta_{X_i} = \bigvee_{x \in X_i} \zeta_x$, here ζ_x is obtained by (10.6). By (10.11), $\forall y \in X - X_i, \mu_{\eta_{X_i}}(y) < \delta$, i.e., $\eta_{X_i} \in E_\Lambda^\delta$ and since for any $x \in X_i$, $\zeta_x \in \bigcup_{\alpha \in \Lambda_x^\varepsilon} \vartheta_\alpha^x$, hence $\forall y \in X_i, \mu_{\eta_{X_i}}(y) \geq \mu_{\bigvee_{b \in \Lambda} b}(y) - \varepsilon$, i.e., $\eta_{X_i} \in F_\varepsilon^\delta$. Although this η_{X_i} may not be the best solution of (10.1), the classifier based on the fuzzy sets $\eta_{X_i}, i = 1,2,3$ could have a satisfactory performance as illustrated by means of the experimental results.

With the above procedure, we can find all the fuzzy description ζ_x for each $x \in X$ in virtue of (10.6). For space limitation, only a few of them are listed here

$$\zeta_{x_1} = m_{22}m_{34}m_{40} + m_{35}m_{40}m_{70} + m_{19}m_{34}m_{40} + m_{11}m_{35}m_{71} + m_{18}m_{22}m_{70};$$
$$\zeta_{x_2} = m_{19}m_{31}m_{40} + m_{31}m_{37}m_{40} + m_{19}m_{25}m_{40} + m_{18}m_{25}m_{40} + m_6m_{18}m_{19}m_{40}.$$

For $x_{12} \in X_1$, there does not exist $\zeta_{x_{12}}$ satisfying (10.10) and (10.11). There are not concepts in $(\Lambda)_{EI}$ satisfying the required conditions for the samples in $D = \{x_6, x_{12}, x_{14}, x_{17}, x_{18}, x_{21}, x_{25}, x_{26}, x_{27}, x_{29}, x_{31}, x_{34}, x_{36}, x_{43}, x_{47}, x_{49}, x_{53}, x_{59}, x_{60}, x_{61}, x_{73}, x_{79}, x_{80}, x_{87}, x_{92}, x_{103}\}$. Let us note again that every fuzzy description can be easily interpreted. Furthermore Figure 10.1 shows the membership functions for each class. Figure 10.2 depicts the membership degrees of all 178 samples including 106 training samples and 72 testing samples. Table 10.1 shows the misclassified patterns (data). We notice that there does not exist fuzzy set in $(\Lambda)_{EI}$ satisfying (10.10), (10.11) for each training sample in D, and this implies that each training sample in D is not typical enough to represent the class it belongs to. Therefore we employ fuzzy set $\eta_{X_i} = \bigvee_{x \in X_i - D} \zeta_x$ in EM to describe the class i, $i = 1,2,3$, where

$$\begin{aligned}
\eta_{X_1} = {} & m_{22}m_{34}m_{40} + m_{35}m_{40}m_{70} + m_{19}m_{34}m_{40} + m_{11}m_{35}m_{71} + m_{18}m_{22}m_{70} \\
& + m_{19}m_{31}m_{40} + m_{31}m_{37}m_{40} + m_{19}m_{25}m_{40} + m_{18}m_{25}m_{40} + m_6m_{18}m_{19}m_{40} \\
& + m_{33}m_{64}m_{80} + m_8m_{13}m_{80} + m_{64}m_{65}m_{80} + m_{31}m_{46}m_{80} + m_{12}m_{25}m_{80} \\
& + m_{10}m_{28}m_{80} + m_7m_{13}m_{46} + m_7m_{44} + m_{19}m_{44} + m_{37}m_{44} + m_{31}m_{44} + m_{13}m_{44} \\
& + m_{16}m_{44} + m_{43}m_{44} + m_{44}m_{80} + m_{44}m_{70} + m_{44}m_{55} + m_{25}m_{44} + m_{16}m_{44} \\
& + m_{43}m_{56} + m_{33}m_{43} + m_{18}m_{44} + m_8m_{22}m_{33} + m_{13}m_{19}m_{33} + m_4m_{13}m_{22} \\
& + m_4m_{13}m_{25} + m_8m_{33}m_{38} + m_{43}m_{44} + m_{44}m_{61} + m_{13}m_{44} + m_{16}m_{44} + m_{28}m_{44} \\
& + m_{43}m_{62} + m_{38}m_{44} + m_{33}m_{43} + m_{26}m_{44} + m_{11}m_{31}m_{46} + m_{17}m_{40}m_{46} \\
& + m_{13}m_{17}m_{46} + m_8m_{22}m_{27} + m_6m_7m_{37} + m_{22}m_{26}m_{40} + m_{19}m_{26}m_{40} \\
& + m_{11}m_{40}m_{46} + m_{11}m_{19}m_{40} + m_{17}m_{34}m_{46} + m_4m_{18}m_{25} + m_4m_{18}m_{25}m_{31} \\
& + m_{25}m_{34}m_{40} + m_4m_{38}m_{40} + m_4m_{20}m_{25}m_{40} + m_{25}m_{29}m_{40} + m_4m_{12}m_{22} \\
& + m_6m_{20}m_{25}m_{40} + m_{19}m_{31}m_{40} + m_{19}m_{25}m_{40} + m_{19}m_{30}m_{40} + m_{22}m_{26}m_{40} \\
& + m_{19}m_{26}m_{40} + m_{16}m_{22}m_{40} + m_{18}m_{22}m_{35} + m_{19}m_{33}m_{40} + m_{19}m_{38}m_{40}m_{58} \\
& + m_{12}m_{40}m_{61} + m_{26}m_{40}m_{64} + m_4m_{12}m_{61} + m_{10}m_{37}m_{40} + m_4m_{28}m_{40}m_{64} \\
& + m_{10}m_{61}m_{62} + m_4m_{61}m_{65} + m_{10}m_{33}m_{61} + m_{22}m_{31}m_{40} + m_4m_{22}m_{29}m_{40} \\
& + m_4m_{37}m_{40} + m_{11}m_{22}m_{40}.
\end{aligned}$$

$$\begin{aligned}
\eta_{X_2} = {} & m_5m_{52} + m_{41}m_{52}m_{75} + m_5m_{52}m_{73} + m_{41}m_{52}m_{76} + m_5m_{52}m_{71} + m_{41}m_{52}m_{70} \\
& + m_{35}m_{52}m_{57} + m_{24}m_{35}m_{52} + m_{25}m_{41}m_{52} + m_{35}m_{39}m_{52} + m_{18}m_{32}m_{59} \\
& + m_6m_{37}m_{41} + m_{31}m_{37}m_{41} + m_5m_{23}m_{41} + m_5m_{14}m_{23} + m_{23}m_{25}m_{41} \\
& + m_{32}m_{57}m_{76} + m_5m_{32}m_{69} + m_5m_{67}m_{69} + m_{32}m_{58}m_{69} + m_{32}m_{58}m_{68}m_{69} \\
& + m_5m_{37}m_{69} + m_5m_{13}m_{69} + m_{32}m_{69}m_{71} + m_{32}m_{67}m_{71} + m_{69}m_{71}m_{72}
\end{aligned}$$

$$
\begin{aligned}
&+m_{32}m_{68}m_{69}m_{71}+m_{15}m_{32}m_{37}+m_{15}m_{23}m_{37}+m_{15}m_{29}m_{37}+m_{15}m_{32}m_{58}\\
&+m_{14}m_{32}m_{58}+m_{15}m_{18}m_{23}+m_{6}m_{23}m_{76}+m_{12}m_{18}m_{27}m_{37}+m_{32}m_{37}m_{41}\\
&+m_{32}m_{37}m_{76}+m_{23}m_{41}m_{76}+m_{20}m_{32}m_{41}+m_{17}m_{37}m_{41}+m_{13}m_{32}m_{41}\\
&+m_{41}m_{49}+m_{5}m_{27}m_{49}+m_{5}m_{29}m_{49}+m_{21}m_{27}m_{49}+m_{49}m_{81}+m_{49}m_{79}m_{81}\\
&+m_{12}m_{15}m_{29}m_{35}+m_{37}m_{45}+m_{5}m_{37}m_{45}+m_{11}m_{23}m_{45}+m_{11}m_{20}m_{45}\\
&+m_{13}m_{20}m_{45}+m_{9}m_{11}m_{45}+m_{32}m_{45}+m_{8}m_{23}m_{45}+m_{38}m_{45}m_{58}+m_{8}m_{13}m_{45}\\
&+m_{8}m_{18}m_{45}+m_{7}m_{29}m_{42}+m_{29}m_{42}m_{58}+m_{15}m_{29}m_{58}+m_{7}m_{24}m_{58}\\
&+m_{8}m_{15}m_{21}+m_{5}m_{28}m_{32}+m_{5}m_{25}m_{28}+m_{5}m_{15}m_{32}+m_{5}m_{23}m_{26}+m_{26}m_{52}\\
&+m_{32}m_{35}m_{42}+m_{37}m_{45}+m_{32}m_{45}+m_{5}m_{21}m_{45}+m_{23}m_{45}m_{60}+m_{23}m_{32}m_{59}\\
&+m_{11}m_{15}m_{45}+m_{5}m_{23}m_{37}+m_{23}m_{37}m_{41}+m_{5}m_{29}m_{37}+m_{26}m_{37}m_{41}\\
&+m_{5}m_{21}m_{37}+m_{32}m_{45}+m_{5}m_{23}m_{45}+m_{20}m_{32}m_{45}+m_{23}m_{26}m_{45}+m_{17}m_{23}m_{45}\\
&+m_{38}m_{81}+m_{32}m_{79}m_{81}+m_{38}m_{41}m_{55}+m_{14}m_{38}m_{81}+m_{32}m_{57}m_{79}+m_{32}m_{56}m_{81}\\
&+m_{32}m_{36}m_{81}+m_{9}m_{79}m_{81}+m_{16}m_{32}m_{81}+m_{45}m_{52}+m_{52}m_{81}+m_{52}m_{79}m_{81}\\
&+m_{49}m_{81}+m_{23}m_{81}+m_{38}m_{81}+m_{20}m_{81}+m_{11}m_{52}m_{79}+m_{49}m_{50}m_{81}\\
&+m_{49}m_{51}m_{81}+m_{15}m_{38}m_{81}+m_{45}m_{52}+m_{5}m_{52}+m_{23}m_{38}m_{45}+m_{29}m_{45}m_{52}\\
&+m_{26}m_{52}+m_{11}m_{45}m_{52}+m_{16}m_{35}m_{52}+m_{26}m_{52}+m_{26}m_{28}m_{52}+m_{9}m_{26}m_{52}\\
&+m_{32}m_{79}m_{81}+m_{19}m_{26}m_{81}+m_{14}m_{22}m_{81}+m_{17}m_{22}m_{81}+m_{22}m_{42}m_{81}\\
&+m_{28}m_{81}+m_{73}m_{79}m_{81}+m_{75}m_{81}+m_{46}m_{81}+m_{46}m_{72}m_{81}+m_{74}m_{81}\\
&+m_{61}m_{81}+m_{62}m_{73}m_{79}+m_{28}m_{81}+m_{27}m_{28}m_{81}+m_{15}m_{37}m_{76}+m_{37}m_{76}m_{78}\\
&+m_{37}m_{76}m_{77}+m_{25}m_{76}m_{77}+m_{36}m_{37}m_{76}+m_{6}m_{37}m_{77}m_{78}+m_{45}m_{52}\\
&+m_{5}m_{52}+m_{20}m_{38}m_{45}.
\end{aligned}
$$

$$
\begin{aligned}
\eta_{X_3} = {} & m_{48}m_{66}+m_{47}m_{66}+m_{48}m_{65}+m_{36}m_{47}m_{48}+m_{24}m_{33}m_{60}+m_{36}m_{39}m_{59}\\
&+m_{21}m_{33}m_{36}+m_{24}m_{41}m_{60}+m_{24}m_{41}m_{59}m_{76}+m_{30}m_{39}m_{60}+m_{4}m_{14}m_{63}\\
&+m_{14}m_{31}m_{63}+m_{31}m_{36}m_{63}+m_{26}m_{31}m_{63}+m_{30}m_{36}m_{66}+m_{24}m_{33}m_{36}\\
&+m_{11}m_{33}m_{66}+m_{5}m_{33}m_{36}+m_{25}m_{36}m_{66}+m_{33}m_{51}+m_{33}m_{50}+m_{9}m_{24}m_{33}\\
&+m_{39}m_{50}m_{51}+m_{42}m_{50}m_{51}+m_{30}m_{39}m_{51}+m_{15}m_{39}m_{50}+m_{27}m_{30}m_{39}m_{51}\\
&+m_{9}m_{17}m_{51}+m_{9}m_{47}m_{48}+m_{48}m_{53}+m_{47}m_{51}+m_{50}m_{53}+m_{47}m_{53}m_{54}\\
&+m_{33}m_{48}+m_{33}m_{47}m_{48}+m_{24}m_{33}m_{42}+m_{24}m_{33}m_{39}+m_{14}m_{33}m_{39}+m_{16}m_{24}m_{33}\\
&+m_{6}m_{24}m_{33}+m_{30}m_{33}m_{60}+m_{36}m_{39}m_{59}+m_{24}m_{33}m_{36}+m_{9}m_{33}m_{39}\\
&+m_{17}m_{33}m_{60}+m_{42}m_{53}+m_{42}m_{53}m_{54}+m_{42}m_{54}m_{80}+m_{52}m_{53}+m_{18}m_{42}m_{53}\\
&+m_{19}m_{52}m_{53}+m_{4}m_{9}m_{39}+m_{4}m_{9}m_{42}+m_{4}m_{14}m_{39}+m_{9}m_{24}m_{31}+m_{16}m_{21}m_{31}\\
&+m_{15}m_{24}m_{31}+m_{16}m_{36}m_{66}+m_{24}m_{31}m_{36}+m_{16}m_{39}m_{66}+m_{16}m_{42}m_{66}\\
&+m_{9}m_{16}m_{66}+m_{4}m_{30}m_{63}+m_{4}m_{63}m_{66}+m_{4}m_{39}m_{63}+m_{30}m_{33}m_{63}+m_{4}m_{41}m_{63}\\
&+m_{4}m_{11}m_{63}+m_{24}m_{53}+m_{21}m_{53}m_{54}+m_{31}m_{63}m_{66}+m_{17}m_{36}m_{63}\\
&+m_{17}m_{31}m_{41}m_{63}+m_{9}m_{17}m_{63}+m_{5}m_{31}m_{63}+m_{14}m_{24}m_{36}m_{39}+m_{4}m_{24}m_{77}
\end{aligned}
$$

$$+m_4m_{24}m_{78} + m_{33}m_{76}m_{77} + m_{33}m_{39}m_{77} + m_{33}m_{36}m_{39} + m_{21}m_{33}m_{36}$$
$$+m_{30}m_{76}m_{77} + m_{15}m_{17}m_{39} + m_{14}m_{33}m_{39} + m_{24}m_{33}m_{36} + m_{33}m_{36}m_{39}$$
$$+m_{21}m_{24}m_{33} + m_{24}m_{27}m_{33} + m_{16}m_{24}m_{33} + m_{18}m_{24}m_{33}.$$

Actually, in this case there are only two errors on the testing set. In what follows, we show that the proposed classifier can determine the class of a testing sample even when the testing sample is coming with some missing features. Given the list of the selected training samples, we know the sample represented by the 1-th row of the data matrix N is not a training sample, it is a test sample and the values of the features are

$$t_1 = (1, .9595, .2948, .7523, .5200, .7840, .7216, .6024, .4242, .6397, .4338, .6082, .9800, .6339)$$

In practice, one can substitute the missing feature by zero or by the mean value of the attribute. In the first case, we have

$$\bar{t}_1 = (1, .9595, .2948, .7523, .5200, .7840, .7216, .6024, .4242, .6397, .4338, .6082, 0, .6339)$$

With the classifier and $\bar{t}_1$, we obtain the degrees of this sample belonging to class 1, 2, 3 to be 0.8147, 0.0091, 0.3821, respectively. This still gives us a clear indication that this pattern belongs to class 1. If the data for 8-th, 13-th features are missing, we have,

$$\bar{t}_1 = (1, .9595, .2948, .7523, .5200, .7840, .7216, 0, .4242, .6397, .4338, .6082, 0, .6339)$$

With the classifier and $\bar{t}_1$, we can also obtain the degrees of this sample belonging to class 1, 2, 3 and they are equal to 0.8147, 0.0091, 0.3821, respectively. This

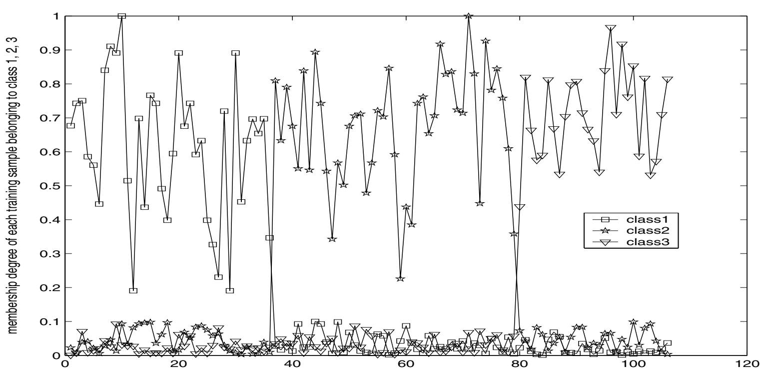


Fig. 10.1 The results of 1-th experiment: The membership degrees of training samples belonging to η_{X_i}, the fuzzy description of class i, $i = 1, 2, 3$

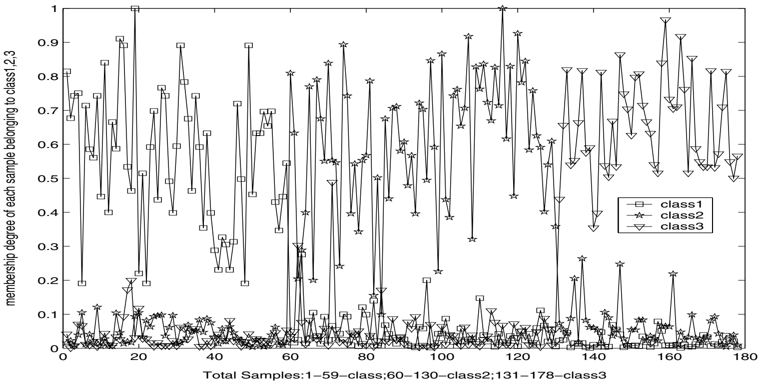


Fig. 10.2 The results of 1-th experiment: The membership degrees of total samples belonging to η_{X_i}, the fuzzy description of class i, $i = 1, 2, 3$

Table 10.1 The misclassified samples

Labels of the sample *s*	62	84
$\mu_{\eta_{X_1}}(\cdot)$	.0250	.0067
$\mu_{\eta_{X_2}}(\cdot)$	.2041	.0988
$\mu_{\eta_{X_3}}(\cdot)$	.3032	.1711

result produces the same answer as before. If the data come with the three missing features, that is 4-th, 8-th, 13-th features, one has,

$$\bar{t}_1 = (1, .9595, .2948, 0, .5200, .7840, .7216, 0, .4242, .6397, .4338, .6082, 0, .6339),$$

and the membership degrees of this sample belonging to classes 1, 2, 3 are 0.8147, 0.3461, 0.3821, respectively. Moving on with the missing features such as 4-th, 6-th, 8-th, 13-th, 14-th features one has

$$\bar{t}_4 = (1, .9595, .2948, 0, .5200, 0, .7216, 0, .4242, .6397, .4338, .6082, 0, 0),$$

and the degrees of this sample belonging to class 1, 2, 3 are 0.4613 0.3461 and 0.3821, respectively. These experiments show that even so many features are missing in the testing sample, we can still classify this sample correctly.

In addition, we carried out some other 9 experiments and the obtained results are shown in Table 10.2 which includes the number of misclassified patterns.

Table 10.2 Number of misclassified samples in other 9 experiments

i-th experiment	2	3	4	5	6	7	8	9	10
number of misclassification	4	2	4	3	5	2	5	1	1

The average number of misclassified patterns reported over 10 rounds is 3. This performance is still better than the results reported in [20, 88], where in both cases there were 4 misclassified patterns.

Based on the previous experimental results, we conclude that the new fuzzy classification based on the AFS logic comes with the following advantages.

1. In step 3, we notice that the 106 fuzzy descriptions ζ_x for the training samples can be computed independently. This implies that this design method can be implemented through parallel processing.
2. Since each simple concepts in M exhibits some semantics, hence for class 1, 2, 3, the fuzzy descriptions $\eta_{X_i} \in (\Lambda)_{EI}$, $i = 1,2,3$, also express their semantic contents separately.
3. This classifier not only can determine the class of a testing sample when it loses information about some features, but also can determine the important features of each training sample, i.e., if we eliminate some feature and this does not influence the classification results, then this feature is not important for classification in this training sample.

The main findings of this section can be summarized as follows.

- A new framework for the design of fuzzy classifier has been established.
- An optimization problem is proposed in order to design the fuzzy classifier and an approximate solution is presented.

Furthermore, the design of the new classifiers can mimic the recognition process carried out by humans via the use of some predefined fuzzy concepts, attributes, and features.

10.2 AFS Classifier Design Based on Entropy Analysis

In this section, we propose a new classifier design approach based on the AFS fuzzy logic and entropy measures. The main difference of the approach here from the above AFS classifier design based on AFS fuzzy logic is that the entropy technique is applied to select better descriptions of the samples and in this way the descriptions of the classes become simpler.

Throughout this section, we always make the following assumptions: X is the set of training samples and there are s features to describe the samples. The training samples are labeled by l classes, which are X_1, X_2, ..., X_l, *i.e.*, $X = \bigcup_{1\leq i\leq l} X_i, X_i \cap X_j = \varnothing, i \neq j$. Let $F = \{f_1, f_2, \cdots, f_s\}$ be the set of the features to describe the samples. For any $x = (v_1, v_2, \cdots, v_s), 1 \leq i \leq s, v_j = f_j(x)$ is the value of x on the features f_j. Let $M = \{m_{ij} \mid 1 \leq i \leq s, 1 \leq j \leq k_i\}$ be the set of simple concepts associating to the features. Where $m_{i1}, m_{i2}, \cdots, m_{ik_i}$ are the simple concepts, such as "*small*", "*medium*", "*not medium*", "*large*" etc associating to the feature f_i and there are k_i simple concepts on the feature f_i. Let (M, τ, X) be an AFS structure of the data. Let ρ_ν be the weight function of simple concept $\nu \in M$ and $\{\mu_\xi(x) \mid \xi \in EM\}$ be a set of coherence membership functions of the AFS fuzzy logic system $(EM, \vee, \wedge, ')$ and the AFS structure (M, τ, X), defined by (4.29) of Theorem 4.5 in which the measure M_ν derived via (4.35) for the weight function ρ_ν. $\Lambda \subseteq EM$, Λ is a set of fuzzy concepts which are selected to design the classifier.

Let Λ be all relative concepts (fuzzy or Boolean) on X. $\Lambda \subseteq EM$, where M is a set of simple concepts on X. For each $x \in X_i$, find a fuzzy set $\zeta_x \in (\Lambda)_{EI}$, such that at the largest degree x belongs to ζ_x, while for any $y \in X - X_i$, at smallest degree y belongs to ζ_x and for $z \in X_i$, $z \neq x$, at comparatively larger degree z belongs to ζ_x. In other words, x can be distinguished from any $y \in X - X_i$ by fuzzy concept ζ_x at the maximal extent. Finally, a fuzzy set $\zeta_{X_i} \in (\Lambda)_{EI}$ based on entropy idea will be selected from $\{\zeta_x \mid x \in X_i\}$ for the fuzzy characterization of class i. For each new sample y, the degree of y belonging to class i is $\mu_{\zeta_{X_i}}(y), i = 1, 2, ..., l$.

In order to present the classifier design procedure, we first introduce the concept of the fuzzy entropy.

10.2.1 Fuzzy Entropy

Entropy is a measure of the amount of uncertainty in the outcome of a random experiment, or equivalently, a measure of the information obtained when the outcome is observed. This concept has been defined in various ways such as [40, 91] and generalized in different applied fields, such as communication theory, mathematics, statistical thermodynamics etc.The entropy of a system as defined by Shannon [96] gives a measure of our ignorance about its actual structure. In what follows, we will first introduce Shannon's entropy and then describe the four de Luca-Termini axioms [40] that a well-defined fuzzy entropy measure must satisfy. Finally, the fuzzy entropy measure, which is an extension of Shannon's definition, is proposed in [60].

10.2.1.1 Shannon's Entropy

Entropy can be considered as a measure of the uncertainty of a random variable X. Let X be a discrete random variable with a finite alphabet set containing N symbols given by $x_0, x_1, \ldots, x_{N-1}$. If an output x_j occurs with probability $p(x_j)$, then the amount of information associated with the known occurrence of the output x_j is defined as

$$I(x_j) = -\log_2 p(x_j) \tag{10.12}$$

That is, for a discrete source, the information generated in selecting symbol x_j is $[-\log_2 p(x_j)]$ bits. On the average, the symbol x_j, will be selected $n \cdot p(x_j)$ times in a total of N selections, so the average amount of information obtained from n source outputs is

$$-n \cdot p(x_0)\log_2 p(x_0) - n \cdot p(x_1)\log_2 p(x_1) - \cdots - n \cdot p(x_{N-1})\log_2 p(x_{N-1}) \tag{10.13}$$

Dividing (10.13) by n, we obtain the average amount of information per source output symbol. This is known as the average information, the uncertainty, or the entropy, and is defined as follows:

Definition 10.1. ([96]) The *entropy* $H(X)$ *of a discrete random variable* X is defined as follows.

$$H(X) = -\sum_{j=0}^{N-1} p(x_j)\log_2 p(x_j) \ \ or \ \ H(X) = \sum_{j=0}^{N-1} p_j log_2 p_j,$$

where p_j denotes $p(x_j)$. Note that entropy is a function of the distribution of X. It does not depend on the actual values taken by the random variable X, but only on the probabilities. Hence, entropy is also written as $H(p)$.

De Luca-Termini Axioms for Fuzzy Entropy: Kosko [40] pointed out that a well-defined fuzzy entropy measure must satisfy the four *de Luca-Termini axioms*.

a. $E(A) = 0$ iff $A \in 2^X$, where A is a nonfuzzy set and 2^X indicates the power set of set X.
b. $E(A) = 1$ iff $\mu_{\tilde{A}}(x) = 0.5$ for all $x \in X$, where $\mu_{\tilde{A}}(x)$ indicates the membership degree of the element x to fuzzy set $\tilde{A}$.
c. $E(\tilde{A}) \le E(\tilde{B})$ if $\tilde{A}$ is less fuzzy than $\tilde{B}$, i.e., if $\mu_{\tilde{A}}(x) \le \mu_{\tilde{B}}(x)$ when $\mu_{\tilde{B}}(x) \le 0.5$ and $\mu_{\tilde{A}}(x) \ge \mu_{\tilde{B}}(x)$ when $\mu_{\tilde{B}}(x) \ge 0.5$, where $\tilde{A}$ and $\tilde{B}$ are fuzzy sets.
d. $E(A) = E(A^{'})$, $A^{'}$ is the negation of A.

10.2.1.2 Fuzzy Entropy

The fuzzy entropy based on the Shannon's entropy was defined in [60] as follows:

Definition 10.2. ([60]) Let $X = \{x_1, x_2, \cdots, x_n\}$ be a universal set with elements x_i distributed in a pattern space, where $i = 1, 2, \ldots, n$. Let $\tilde{A}$ be a fuzzy set defined on an interval of pattern space which contains k elements $(k < n)$. The membership degree of the element x_i belonging to the fuzzy set $\tilde{A}$ is denoted by $\mu_{\tilde{A}}(x_i)$. The match degree D_j with the fuzzy set $\tilde{A}$ for the elements of class j, where $j = 1, 2, \ldots, m$, is defined as

$$D_j = \frac{\sum_{x \in X_j} \mu_{\tilde{A}}(x)}{\sum_{x \in X} \mu_{\tilde{A}}(x)}. \tag{10.14}$$

The *fuzzy entropy* $FE_{C_j}(\tilde{A})$ *of the elements of class* j is defined as

$$FE_{C_j}(\tilde{A}) = -D_j log_2 D_j. \tag{10.15}$$

The *fuzzy entropy* $FE(\tilde{A})$ *on the universal set* X is defined as

$$FE(\tilde{A}) = \sum_{j=1}^{l} FE_{C_j}(\tilde{A}). \tag{10.16}$$

In (10.15), the fuzzy entropy $FE_{C_j}(\tilde{A})$ is a nonprobabilistic entropy. Therefore, the new term is coined such as "*matching degree*" that applies to D_j. The basic property of the fuzzy entropy is similar to that of the Shannon's entropy and it satisfies the four de Luca-Termini axioms; however, their ways of measuring information are not the same. The probability p_j of the Shannon's entropy is measured via the number of occurrence of elements. In contrast, the matching degree D_j in fuzzy entropy is measured via the membership values of the elements.

10.2.2 Classifier Design

With these concepts of the AFS fuzzy logic and the fuzzy entropy, we can proceed with an introduction of the algorithm. In general, we choose the number of the simple concepts associated with each feature according to the number of the classes in the training data.

AFS Classifier Design Based on Entropy Analysis

Step1: For each training sample $x_0 \in X$, compute the membership degree $\mu_\theta(x_0)$, where $\theta = \bigvee_{m\in\Lambda} m$. Actually, $\mu_\theta(x_0)$ is the maximum membership degree of x_0 belonging to the fuzzy concept in $(\Lambda)_{EI}$.

Step2: For each training sample $x_0 \in X$ and given small positive number ε, compute the following simple concepts set

$$B^{\varepsilon}_{x_0} = \{m \in \Lambda \mid \mu_m(x_0) \geq \mu_\theta(x_0) - \varepsilon\} \tag{10.17}$$

This is the set which includes all possible simple concepts in M which can characterize x_0 well.

Step 3: Given appropriate positive numbers $0 \leq \delta \leq \beta \leq 1$, and assume that x_0 belong to class X_i, we compute the following set

$$\Xi^{\varepsilon}_{x_0} = \left\{\gamma = \prod_{m\in H} m \;\middle|\; H \subseteq B^{\varepsilon}_{x_0};\;\; \mu_\gamma(x_0) \geq \beta;\;\; \forall y \in X - X_i,\;\; \mu_\gamma(y) < \delta\right\} \tag{10.18}$$

Then we also compute the following fuzzy set for the best characterization of training sample x_0.

$$\zeta_{x_0} = \begin{cases} \arg\max\limits_{\gamma\in\Xi^{\varepsilon}_{x_0}} \{\mu_\gamma(x_0)\}, & \Xi_{x_0} \neq \varnothing, \\ 0, & \Xi_{x_0} = \varnothing. \end{cases} \tag{10.19}$$

Remark 10.1. For $\alpha, \beta \in \Lambda^{\varepsilon}_{x_0}$ if $\alpha \leq \beta$ in the sub EI algebra $(\Lambda)_{EI}$, then for any $x \in X$, $\mu_\alpha(x) \leq \mu_\beta(x)$. By (10.19), the characterization ζ_{x_0} can be simply found just by checking the membership degrees of x_0 belonging to the maximal elements in $\Xi^{\varepsilon}_{x_0}$, instead of checking every one in $\Xi^{\varepsilon}_{x_0}$. In general, just very few elements in $\Xi^{\varepsilon}_{x_0}$ are its maximal elements, even $\Xi^{\varepsilon}_{x_0}$ may have a great amount of elements. For instance, if $\beta \in \Xi^{\varepsilon}_{x_0}$, then any $\alpha \in \Xi^{\varepsilon}_{x_0}$ which $\alpha \leq \beta$ need not be checked. The perfect algebraic properties of the completely distributivity lattices $(EM,\vee,\wedge,')$ can be applied to determine the maximal elements in $\Xi_{x_0}^{\varepsilon}$.

Step 4: From the last step, we can obtain the best characterization of each training sample describing by the fuzzy concept sets in EM, let

$$\mathscr{C}_{X_i}=\{\zeta_x \mid \zeta_x \neq 0, x\in X_i\} \tag{10.20}$$

Usually, some training sample $x \in X_i$ may not be typical enough or too specific to represent the characteristic of the class X_i and they may lead to the low classification rate of testing data. Thus we need to filter out these fuzzy sets in $\mathscr{C}_{X_i}$ via applying the concept of entropy. Assume further that n_i is the number of $x \in X_i$ such that $\zeta_x \neq 0$, i.e., $\mathscr{C}_{X_i}=\{\zeta_l \mid 1\leq l\leq n_i\}$. Then we compute the following set

$$\Gamma_{X_i}=\left\{\zeta_k\in\mathscr{C}_{X_i} \mid \lambda_{ik}E(\zeta_k)\leq \frac{1}{n_i}\sum_{l=1}^{n_i}\lambda_{il}E(\zeta_l)\right\}, \tag{10.21}$$

where $k=1,2,...,n_i$,

$$\lambda_{ik}=\frac{\delta+\sum\limits_{x\in X-X_i}\mu_{\zeta_k}(x)}{\sum\limits_{x\in X_i}\mu_{\zeta_k}(x)},$$

and the condition $\delta>0$ prevents us from the case where $\sum\limits_{x\in X-X_i}\mu_{\zeta_k}(x)=0$,

$$E(\zeta_k)=-\sum_{i=1}^{m}\left(\frac{\sum\limits_{x\in X_i}\mu_{\zeta_k}(x)}{\sum\limits_{x\in X}\mu_{\zeta_k}(x)}log_2\frac{\sum\limits_{x\in X_i}\mu_{\zeta_k}(x)}{\sum\limits_{x\in X}\mu_{\zeta_k}(x)}\right).$$

For the training sample x, $\zeta_x=\zeta_k\in\Gamma_{X_i}$, the lower the value of λ_{ik}, the more typical x in the class X_i. Recall from de Luca-Termini axioms that a fuzzy entropy is a function on fuzzy set that becomes smaller when the level of fuzziness of the fuzzy set is reduced. This implies that the less $E(\zeta_k)$ is, the clearer ζ_x is to represent the class X_i. Thus the fuzzy sets in Γ_{X_i} is selected to describe the characterization of the class X_i:

$$\zeta_{X_i}=\bigvee_{\zeta\in\Gamma_{X_i}}\zeta\in EM. \tag{10.22}$$

Step 5: Determine the class label of each testing sample or new sample by its membership degree belonging to ζ_{X_i}, $i=1,2,...,l$. Suppose $y\notin X$ is a testing sample or new sample. Let $\rho_\gamma(y)$ be the weight function for the simple concept $\gamma\in M$. Let the characterization of class i, $\zeta_{X_i}=\bigvee\limits_{\zeta\in\Gamma_{X_i}}\zeta=\sum_{i\in I}(\prod_{m\in A_i}m)\in EM$, $i=1,2,...,l$.

In virtue of (4.26), one has

$$A_k^{\tau}(y)=\{x\in X \mid \tau(y,x)\supseteq A_k\}\subseteq X,\ k\in I.$$

By Theorem 4.5, in which the measure M_ν has been derived via (4.35), we obtain the membership degree of y belonging to ζ_{X_i} as follows.

$$\mu_{\zeta_{X_i}}(y) = \sup_{i\in I}\left(\prod_{\gamma\in A_i} m_\gamma(A_i^\tau(y))\right) = \sup_{i\in I}\left(\prod_{\gamma\in A_i}\frac{\sum_{x\in A_i^\tau(y)\cup\{y\}}\rho_\gamma(x)}{\sum_{x\in X\cup\{y\}}\rho_\gamma(x)}\right) \tag{10.23}$$

The class label of y is expressed as $\arg\max_{1\le k\le l}\{\mu_{\zeta_{X_k}}(y)\}$.

In virtue of Theorem 4.6, the membership degree of any sample in the whole space belongingness to ζ_{X_i}, the fuzzy description of class X_i, can be determined by (4.41) or (4.43) and the class label of any new sample can be determined by its membership degree. In practice, the high dimension integral may be too complex to be obtained.

Remark 10.2. Based on this idea, we can apply AFS fuzzy logic to design classifiers for the database with various data types such as integer, real number, Boolean value, the sub-preference relations, even concepts of human intuition. With the classification algorithms presented in [1, 26], it is difficult or impossible to solve these problems.

10.2.3 Experimental Studies

The main aim of this section is to illustrate the effectiveness of the AFS fuzzy classifier design based on entropy technique. We design classifiers using the proposed procedure for the Wine data, Iris data, and the Breast data [67]. We proceed with a *tenfold cross-validation* that generates an $90-10$ split by taking 90% samples of each class as a training set and testing the remaining 10% of the data set. Furthermore, the experiments are repeated ten times by taking random splits of data into the training and testing part, respectively. In the following experimental studies, the weight functions of the fuzzy concepts in M can be defined by the following method. Let $m_{i1}, m_{i2}, m_{i3}, m_{i4}$ are the simple concepts, "*small*", "*medium*", "*not medium*", "*large*" associated with the feature f_i, respectively, according to the observed data X and their semantics, the weight functions for them can be defined as follows:

$$\rho_{m_{j1}}(x_i) = \frac{h_{j1} - f_j(x_i)}{h_{j1} - h_{j2}} \tag{10.24}$$

$$\rho_{m_{j2}}(x_i) = \frac{h_{j4} - |f_j(x_i) - h_{j3}|}{h_{j4} - h_{j5}} \tag{10.25}$$

$$\rho_{m_{j3}}(x_i) = \frac{|f_j(x_i) - h_{j3}| - h_{j5}}{h_{j4} - h_{j5}} \tag{10.26}$$

$$\rho_{m_{j4}}(x_i) = \frac{f_j(x_i) - h_{j2}}{h_{j1} - h_{j2}} \tag{10.27}$$

where $j = 1,2,\ldots,s$,

$$\begin{array}{ll} h_{j1} = \max_{1\le k\le n}\{f_j(x_k)\}, & h_{j2} = \min_{1\le k\le n}\{f_j(x_k)\}, \\ h_{j3} = \frac{\sum_{1\le k\le n} f_j(x_k)}{n}, & h_{j4} = \max_{1\le k\le n}\{|f_j(x_k) - h_{j3}|\}, \\ h_{j5} = \min\limits_{1\le k\le n}\{|f_j(x_k) - h_{j3}|\}. & \end{array}$$

By Definition 4.8 and considering the semantics of $m_{i1}, m_{i2}, m_{i3}, m_{i4}$, one can verify that $\rho_{m_{j,k}}$, $j = 1,2,\ldots,s$ are weight functions of the simple concepts $m_{j,k}$. Let ρ_ν be the weight function of simple concept $\nu \in M$ and $\{\mu_\xi(x) \mid \xi \in EM\}$ be a set of coherence membership functions of the AFS fuzzy logic system $(EM,\vee,\wedge,')$ and the AFS structure (M,τ,X), defined by (4.29) of Theorem 4.5 in which the measure M_ν derived via (4.35) for the weight function ρ_ν.

Wine Data

In each of the ten experiments of the tenfold cross-validation, there are 160 samples in the training set, and 18 samples in the testing set. Let $X = \{x_1,x_2,\cdots,x_{160}\}$. Let $F = \{f_1, f_2,\cdots, f_{13}\}$ be a set of features on X. Let

$$M = \{m_{ij} \mid 1 \le i \le 13, 1 \le j \le 3\},$$

where m_{i1}, m_{i2}, m_{i3} are the simple concepts meaning "*small*", "*medium*", "*large*", respectively on the feature f_i, and their weight functions are defined as (10.24), (10.25), (10.27) mentioned above. For example, the semantics of fuzzy concepts $m_{1,1}, m_{1,2}, m_{1,3}$ in M are: $m_{1,1}$: "*small alcohol content (AL)*", $m_{1,2}$: "*medium alcohol content (AL)*", $m_{1,3}$: "*large alcohol content (AL)*". Figures 10.3 to 10.15 display the membership functions of the simple concepts in M on each feature. Let $\Lambda = M$. The results of each step are listed below.

Step 1: Let $\theta = \bigvee_{m\in M} m$. Actually, $\mu_\theta(x)$ is the maximum membership degree of $x \in X$ belonging to the fuzzy concept in EM.

Step 2: Given $\varepsilon = 0.1$, it is easy to get set $B^{0.1}_{x_0} \subseteq M$ for each $x_0 \in X$ by (10.17). As examples, some of them are listed as follows:

$$\begin{array}{l} B^{0.1}_{x_1} = \{m_{1,3}, m_{5,3}, m_{12,3}\},\ \ B^{0.1}_{x_7} = \{m_{2,2}, m_{10,2}, m_{12,3}\}, \\ B^{0.1}_{x_{21}} = \{m_{1,2}, m_{5,2}, m_{6,2}, m_{12,3}, m_{13,2}\},\ \ B^{0.1}_{x_{40}} = \{m_{5,2}, m_{9,2}, m_{13,2}\}, \\ B^{0.1}_{x_{47}} = \{m_{4,1}, m_{6,3}, m_{7,3}\},\ \ B^{0.1}_{x_{50}} = \{m_{8,1}, m_{9,3}, m_{11,2}\}, \\ B^{0.1}_{x_{54}} = \{m_{2,1}, m_{3,1}, m_{4,1}, m_{7,1}, m_{9,1}, m_{10,1}\},\ B^{0.1}_{x_{62}} = \{m_{3,1}, m_{5,3}, m_{8,1}, m_{11,3}, m_{13,2}\}, \\ B^{0.1}_{x_{63}} = \{m_{6,1}, m_{8,2}\},\ \ B^{0.1}_{x_{79}} = \{m_{1,1}, m_{2,2}, m_{5,1}, m_{11,2}, m_{12,2}, m13,2\}, \\ B^{0.1}_{x_{85}} = \{m_{1,1}, m_{5,2}, m_{8,1}, m_{13,1}\},\ \ B^{0.1}_{x_{94}} = \{m_{3,1}, m_{5,1}, m_{6,2}, m_{7,2}, m_{12,3}\}, \\ B^{0.1}_{x_{99}} = \{m_{1,1}, m_{11,2}\},\ \ B^{0.1}_{x_{124}} = \{m_{6,1}, m_{7,1}, m_{9,1}, m_{12,1}, m_{13,2}\}, \\ B^{0.1}_{x_{148}} = \{m_{6,1}, m_{10,3}\},\ \ B^{0.1}_{x_{154}} = \{m_{2,2}, m_{4,2}, m_{7,1}, m_{9,1}, m_{10,3}, m_{11,1}\}, \\ B^{0.1}_{x_{159}} = \{m_{2,2}, m_{3,2}, m_{10,3}, m_{11,1}\}. \end{array}$$

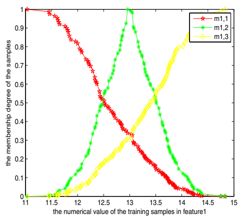


Fig. 10.3 Membership functions of the simple concepts for the feature f_1

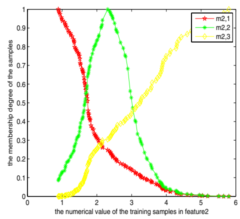


Fig. 10.4 Membership functions of the simple concepts for the feature f_2

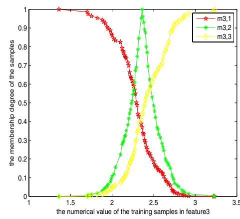


Fig. 10.5 Membership functions of the simple concepts for n the feature f_3

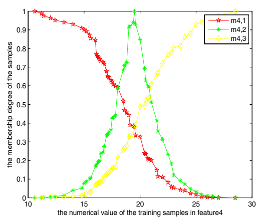


Fig. 10.6 Membership functions of the simple concepts for the feature f_4

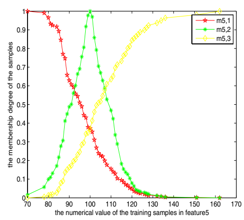


Fig. 10.7 Membership functions of the simple concepts for the feature f_5

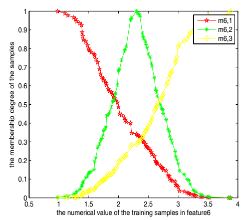


Fig. 10.8 Membership functions of the simple concepts for the feature f_6

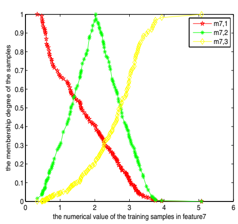


Fig. 10.9 Membership functions of the simple concepts for the feature f_7

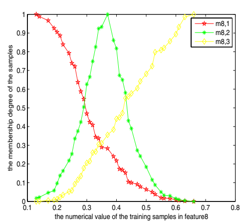


Fig. 10.10 Membership functions of the simple concepts for the feature f_8

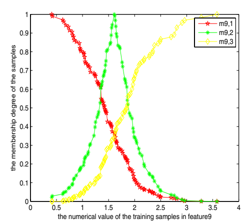


Fig. 10.11 Membership functions of the simple concepts for the feature f_9

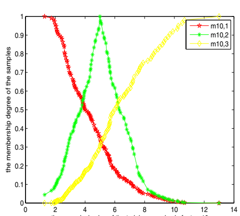


Fig. 10.12 Membership functions of the simple concepts for the feature f_{10}

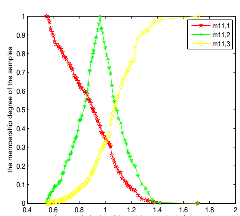


Fig. 10.13 Membership functions of the simple concepts for the feature f_{11}

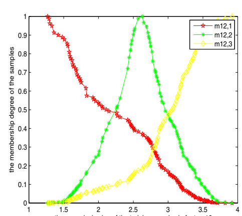


Fig. 10.14 Membership functions of the simple concepts for n the feature f_{12}

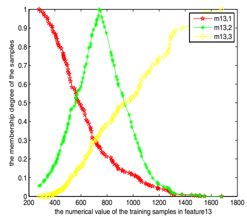


Fig. 10.15 Membership functions of the simple concepts for the feature f_{13}

Step 3: Let $\delta = 0.2, \beta = 0.5$. By (10.18), we can obtain the set of fuzzy concepts $\Xi_{x_0}^{0.1} \subseteq EM$ for every $x_0 \in X$. We just list some of them.

$$\Xi_{x_1}^{0.1} = \{m_{1,3}m_{5,3}m_{12,3}\},\ \Xi_{x_7}^{0.1} = \{m_{2,2}m_{10,2}m_{12,3}\},\ \Xi_{x_{21}}^{0.1} = \{m_{1,2}m_{12,3}m_{13,2}\},$$
$$\Xi_{x_{47}}^{0.1} = \varnothing,\ \Xi_{x_{50}}^{0.1} = \varnothing,\ \Xi_{x_{54}}^{0.1} = \{m_{2,1}m_{4,1}m_{7,1},\ m_{2,1}m_{9,1}m_{10,1},\ m_{3,1}m_{4,1}m_{7,1},$$
$$m_{3,1}m_{9,1}m_{10,1}, m_{4,1}m_{7,1}m_{9,1}, m_{4,1}m_{9,1}m_{10,1}, m_{7,1}m_{9,1}m_{10,1}\},$$
$$\Xi_{x_{62}}^{0.1} = \{m_{3,1}m_{5,3}m_{8,1}, m_{3,1}m_{11,3}m_{13,2}, m_{5,3}m_{8,1}m_{11,3}, m_{8,1}m_{11,3}m_{13,2}\},$$
$$\Xi_{x_{63}}^{0.1} = \varnothing,\ \Xi_{x_{79}}^{0.1} = \{m_{1,1}m_{11,2}, m_{1,1}m_{12,2}\},$$
$$\Xi_{x_{85}}^{0.1} = \{m_{1,1}m_{5,2}m_{8,1},\ m_{1,1}m_{8,1}m_{13,1},\ m_{5,2}m_{8,1}m_{13,1}\},$$
$$\Xi_{x_{94}}^{0.1} = \{m_{3,1}m_{5,1}m_{6,2}, m_{3,1}m_{7,2}m_{12,3}, m_{5,1}m_{6,2}m_{7,2}\},\ \Xi_{x_{99}}^{0.1} = \{m_{1,1}m_{11,2}\},$$
$$\Xi_{124}^{0.1} = \{m_{6,1}m_{7,1}m_{9,1}, m_{6,1}m_{9,1}m_{12,1}, m_{6,1}m_{12,1}m_{13,2}, m_{7,1}m_{12,1}m_{13,2}\},$$
$$\Xi_{x_{148}}^{0.1} = \{m_{6,1}m_{10,3}\},\ \Xi_{x_{154}}^{0.1} = \{m_{2,2}m_{4,2}m_{7,1},\ m_{2,2}m_{7,1}m_{9,1},\ m_{2,2}m_{9,1}m_{10,3},$$
$$m_{2,2}m_{10,3}m_{11,1},\ m_{4,2}m_{7,1}m_{9,1},\ m_{4,2}m_{9,1}m_{10,3},\ m_{4,2}m_{10,3}m_{11,1},\ m_{7,1}m_{9,1}m_{10,3},$$
$$m_{7,1}m_{10,3}, m_{9,1}m_{10,3}m_{11,1}\},\ \Xi_{x_{159}}^{0.1} = \{m_{3,2}m_{10,3}m_{11,1}\}$$

From (10.19), we can obtain the fuzzy description of each sample. Here are some illustrative examples:

$$\zeta_{x_1} = m_{1,3}m_{5,3}m_{12,3},\ \zeta_{x_7} = m_{2,2}m_{10,2}m_{12,3},\ \zeta_{x_{21}} = m_{1,2}m_{12,3}m_{13,2},$$
$$\zeta_{x_{47}} = 0,\ \zeta_{x_{50}} = 0,\ \zeta_{x_{54}} = m_{4,1}m_{9,1}m_{10,1},\ \zeta_{x_{62}} = m_{3,1}m_{5,3}m_{8,1},$$
$$\zeta_{x_{63}} = 0,\ \zeta_{x_{79}} = m_{1,1}m_{12,2},\ \zeta_{x_{85}} = m_{1,1}m_{8,1}m_{13,1},$$
$$\zeta_{x_{94}} = m_{5,1}m_{6,2}m_{7,2},\ \zeta_{x_{99}} = m_{1,1}m_{11,2},\ \zeta_{x_{124}} = m_{7,1}m_{12,1}m_{13,2},$$
$$\zeta_{x_{148}} = m_{6,1}m_{10,3},\ \zeta_{x_{154}} = m_{2,2}m_{4,2}m_{7,1},\ \zeta_{x_{159}} = m_{3,2}m_{10,3}m_{11,1}.$$

Step 4: By (10.20), we get $\mathscr{C}_{X_i}, i = 1,2,3$ as follows:

$$\mathscr{C}_{X_1} = \{m_{1,3}m_{13,3},\ m_{7,3}m_{13,3},\ m_{4,1}m_{13,3},\ m_{1,3}m_{5,3}m_{12,3},\ m_{2,2}m_{5,2}m_{9,3},$$
$$m_{2,2}m_{10,2}m_{12,3},\ m_{1,3}m_{4,1}m_{10,2},\ m_{6,2}m_{9,2}m_{10,2},\ m_{1,3}m_{4,1}m_{7,3},\ m_{1,3}m_{3,3}m_{12,2},$$
$$m_{8,1}m_{10,2}m_{11,2},\ m_{5,3}m_{12,3}m_{13,2},\ m_{1,2}m_{12,3}m_{13,2},\ m_{3,3}m_{4,2}m_{8,2},\ m_{3,2}m_{4,2}m_{13,3},$$
$$m_{3,3}m_{4,1}m_{8,2},\ m_{1,2}m_{5,2}m_{12,2},\ m_{5,2}m_{6,2}m_{12,2},\ m_{10,2}m_{12,3}m_{13,2},\ m_{1,3}m_{6,3}m_{10,2},$$
$$m_{5,2}m_{7,3}m_{8,1},\ m_{1,3}m_{6,3}m_{11,2},\ m_{1,2}m_{4,1}m_{8,1}m_{9,3}\}$$

$\mathscr{C}_{X_2} = \{m_{6,3}m_{7,2},\ m_{2,1}m_{13,1},\ m_{1,1}m_{11,3},\ m_{1,1}m_{12,2},\ m_{10,1}m_{13,1},\ m_{11,3}m_{13,1}, m_{9,3}m_{10,1},$
$m_{1,1}m_{11,2},\ m_{1,1}m_{9,3},\ m_{1,1}m_{7,2},\ m_{1,1}m_{7,3},\ m_{9,3}m_{13,1},\ m_{4,1}m_{9,1}m_{10,1},\ m_{2,1}m_{3,1}m_{9,1},$
$m_{3,1}m_{4,1}m_{5,1},\ m_{3,1}m_{5,1}m_{7,2},\ m_{2,1}m_{9,1}m_{13,2},\ m_{3,1}m_{5,3}m_{8,1},\ m_{4,3}m_{8,1}m_{11,3},$
$m_{1,2}m_{3,1}m_{7,2},\ m_{2,1}m_{8,2}m_{13,2},\ m_{5,1}m_{10,1}m_{11,3},\ m_{1,2}m_{9,2}m_{10,2},\ m_{2,1}m_{6,2}m_{7,2},$
$m_{9,2}m_{10,1}m_{11,3},\ m_{5,1}m_{6,2}m_{10,1},\ m_{5,1}m_{9,2}m_{10,1},\ m_{5,1}m_{9,2}m_{11,2},\ m_{1,1}m_{8,1}m_{13,1},$
$m_{4,2}m_{5,3}m_{9,3},\ m_{2,2}m_{9,2}m_{11,2},\ m_{2,2}m_{7,2}m_{8,2},\ m_{3,1}m_{4,2}m_{8,2},\ m_{5,1}m_{6,2}m_{7,2},\ m_{7,2}m_{8,2}m_{9,2},$
$m_{5,1}m_{6,1}m_{9,2},\ m_{6,2}m_{7,2}m_{13,1},\ m_{4,2}m_{8,2}m_{10,1},\ m_{1,2}m_{2,3}m_{13,1},\ m_{2,2}m_{8,2}m_{13,1},$
$m_{2,2}m_{4,3}m_{11,2},\ m_{2,1}m_{5,2}m_{8,3}m_{9,1}\}$

$\mathscr{C}_{X_3} = \{m_{6,1}m_{10,3},\ m_{6,1}m_{10,2}m_{12,1},\ m_{2,2}m_{6,1}m_{12,1},\ m_{3,2}m_{10,2}m_{12,1},\ m_{7,1}m_{12,1}m_{13,2},$
$m_{1,2}m_{7,1}m_{9,1},\ m_{3,2}m_{8,2}m_{9,1},\ m_{5,1}m_{6,1}m_{7,1},\ m_{3,2}m_{11,1}m_{12,1},\ m_{5,3}m_{11,1}m_{12,1},$
$m_{6,2}m_{11,1}m_{12,1},\ m_{8,3}m_{10,3}m_{11,1},\ m_{2,3}m_{12,1}m_{13,2},\ m_{4,2}m_{9,2}m_{11,1},\ m_{4,3}m_{7,1}m_{8,3},$
$m_{1,3}m_{10,3}m_{11,1},\ m_{3,2}m_{6,2}m_{11,1},\ m_{6,1}m_{7,1}m_{9,1},\ m_{2,2}m_{4,2}m_{7,1},\ m_{5,3}m_{10,3}m_{11,1},$
$m_{3,2}m_{10,3}m_{11,1},\ m_{2,3}m_{4,3}m_{8,3}m_{10,2}\}$

By (10.21) and (10.22), the following fuzzy description of characterization of each class, $\zeta_{X_i}, i = 1,2,3$ can be computed.

$$\zeta_{X_1} = m_{1,3}m_{5,3}m_{12,3} + m_{2,2}m_{10,2}m_{12,3} + m_{1,3}m_{4,1}m_{10,2} + m_{4,1}m_{13,3} + m_{1,3}m_{4,1}m_{7,3}$$
$$+ m_{1,3}m_{13,3} + m_{1,3}m_{3,3}m_{12,2} + m_{7,3}m_{13,3} + m_{5,3}m_{12,3}m_{13,2} + m_{3,3}m_{4,1}m_{8,2}$$
$$+ m_{10,2}m_{12,3}m_{13,2} + m_{1,3}m_{6,3}m_{10,2} + m_{5,2}m_{7,3}m_{8,1} + m_{1,3}m_{6,3}m_{11,2},$$

$$\zeta_{X_2} = m_{2,1}m_{13,1} + m_{1,1}m_{11,3} + m_{1,1}m_{12,2} + m_{10,1}m_{13,1} + m_{11,3}m_{13,1} + m_{9,3}m_{10,1}$$
$$+ m_{1,1}m_{11,2} + m_{1,1}m_{9,3} + m_{1,1}m_{7,2} + m_{1,1}m_{7,3} + m_{9,3}m_{13,1} + m_{4,1}m_{9,1}m_{10,1}$$
$$+ m_{2,1}m_{3,1}m_{9,1} + m_{3,1}m_{5,1}m_{7,2} + m_{2,1}m_{8,2}m_{13,2} + m_{5,1}m_{10,1}m_{11,3} + m_{2,1}m_{6,2}m_{7,2}$$
$$+ m_{9,2}m_{10,1}m_{11,3} + m_{5,1}m_{6,2}m_{10,1} + m_{5,1}m_{9,2}m_{10,1} + m_{5,1}m_{9,2}m_{11,2} + m_{1,1}m_{8,1}m_{13,1}$$
$$+ m_{3,1}m_{4,2}m_{8,2} + m_{5,1}m_{6,2}m_{7,2} + m_{5,1}m_{6,1}m_{9,2} + m_{6,2}m_{7,2}m_{13,1} + m_{4,2}m_{8,2}m_{10,1}$$
$$+ m_{2,2}m_{8,2}m_{13,1} + m_{2,2}m_{4,3}m_{11,2} + m_{2,1}m_{5,2}m_{8,3}m_{9,1},$$

$$\zeta_{X_3} = m_{6,1}m_{10,3} + m_{6,1}m_{10,2}m_{12,1} + m_{3,2}m_{10,2}m_{12,1} + m_{7,1}m_{12,1}m_{13,2} + m_{1,2}m_{7,1}m_{9,1}$$
$$+ m_{3,2}m_{11,1}m_{12,1} + m_{5,3}m_{11,1}m_{12,1} + m_{6,2}m_{11,1}m_{12,1} + m_{8,3}m_{10,3}m_{11,1} + m_{2,3}m_{12,1}m_{13,2}$$
$$+ m_{1,3}m_{10,3}m_{11,1} + m_{6,1}m_{7,1}m_{9,1} + m_{5,3}m_{10,3}m_{11,1} + m_{3,2}m_{10,3}m_{11,1}.$$

$\zeta_{X_1}, \zeta_{X_2}, \zeta_{X_3}$ come with a well-defined semantics. For instance, ζ_{X_1} states that *"small total phenols and large color intensity" or "small total phenols and medium color intensity and small OD280 /OD315 of diluted wines" or "medium ash content and medium color intensity and small OD280/OD315 of diluted wines" or "small flacanoids and small OD280/OD315 of diluted wines and medium praline" or "medium alcohol content and small flavanoids and small proanthocyaninsm" or "medium ash content and small hue and small OD280 /OD315 of diluted wines" or "large magnesium content and small hue and small*

OD280 /OD315 of diluted wines" or "medium total phenols and small hue and small OD280 /OD315 of diluted wines" or "large nonflacanoids phenols and large color intensity and small hue" or "large malic acid content and small OD280 /OD315 of diluted wines and medium praline" or "large alcohol content and large color intensity and small hue" or "small total phenols and small flacanoids and small proanthocyaninsm" or "large magnesium content and large color intensity and small hue" or "medium ash content and large color intensity and small hue"

Figure 10.16 shows the membership functions of the fuzzy sets $\zeta_{X_1}, \zeta_{X_2}, \zeta_{X_3}$ for the training data.

Step 5: By (10.23), we can assign the class label of each testing sample according to the membership degrees of the sample belonging to $\zeta_{X_1}, \zeta_{X_2}, \zeta_{X_3} \in EM$.

Figure 10.17 shows the membership functions of the fuzzy sets $\zeta_{X_1}, \zeta_{X_2}, \zeta_{X_3}$ for the test samples. Table 10.3 shows the results of the tenth experiment.

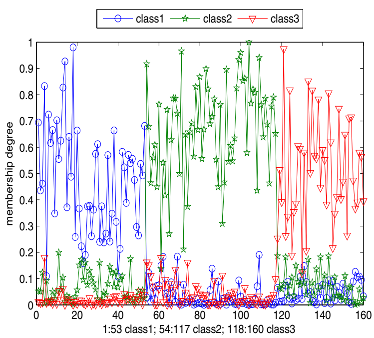


Fig. 10.16 The membership functions of the fuzzy sets $\zeta_{X_1}, \zeta_{X_2}, \zeta_{X_3}$ on training samples

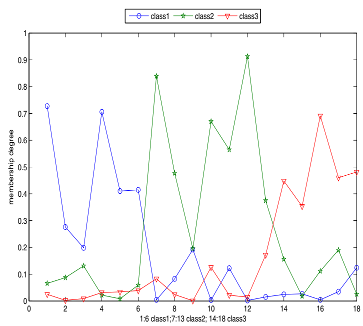


Fig. 10.17 The membership functions of the fuzzy sets $\zeta_{X_1}, \zeta_{X_2}, \zeta_{X_3}$ on testing samples

Table 10.3 The numbers of misclassifying samples and the number of rules for the Wine Data

No. of the experiment	1	2	3	4	5	6	7	8	9	10
No. of the misclassifying training samples	2	2	1	1	1	3	2	1	3	3
No. of the misclassifying testing samples	0	0	1	0	1	0	0	1	0	0
No. of the rules	58	59	51	59	60	63	57	57	54	65

The number of the rules is the total length of the fuzzy concepts $\zeta_{X_1}, \zeta_{X_2}, \zeta_{X_3}$, i.e., $||\zeta_{X_1}|| + ||\zeta_{X_2}|| + ||\zeta_{X_3}||$ (refer to Definition 5.9). In ten experiments of the tenfold cross-validation, the average accuracy and the standard deviation for the training set are 98.81% and $2.9600e-005$, and its corresponding average accuracy for testing set and the standard deviation for the testing set are 98.315% and $7.2016e-004$.

Iris Data

The Iris data set is one of the most popular data sets to examine the performance of algorithms of pattern recognition. The Iris data can be arranged into a 150×4 matrix $W = (\omega_{i,j})_{150\times 4}$ evenly distributed in three classes: iris-setosa, iris-versicolor, and iris-virginica. The vector of the sample i, $(\omega_{i1}, \omega_{i2}, \omega_{i3}, \omega_{i4})$ has four features: sepal length, sepal width, petal length, and petal width.

Let $X = \{x_1, x_2, \cdots, x_{135}\}$ be the set of 135 training samples and $M = \{m_{j,k} | 1 \leq j \leq 4, 1 \leq k \leq 3\}$, where $m_{j,1}, m_{j,2}, m_{j,3}$ are the fuzzy concepts, "*small*", "*medium*", "*large*" associating to the feature f_j, respectively. We use the numeric values of the parameters: $\varepsilon = 0.15$, $\delta = 0.25$, $\beta = 0.6$. Those are selected considering the same procedure as the one used for the Wine data. The experimental results are listed in Table 10.4.

Table 10.4 Numbers of misclassified samples and the number of rules for the Iris Data

	1	2	3	4	5	6	7	8	9	10
No. of misclassifying. in training samples	5	2	4	4	5	5	5	6	5	4
No. of misclassifying. in testing samples.	0	2	1	0	0	0	0	0	0	3
No. of rules	10	7	8	10	6	9	8	5	6	8

The average accuracy and the standard deviation for the training set are 96.67% and $6.40e-005$, and its corresponding average accuracy for testing set and the the standard deviation for the testing set are 96% and 0.0051.

10.2.3.1 Breast Data

The Wisconsin Breast Cancer Diagnostic data set contains 699 patterns distributed into two output classes, "*benign*" and "*malignant*". Each pattern consists of nine input features: clump thickness (CT), uniformity of cell size (UC), uniformity of cell shape (UCS), marginal adhesion (MA), single epithe- lial cell size (SECS), bare nuclei (BN), bland chromatin (BC), normal nuclei (NN), and mitoses (MI). The nine features have integer values in the range of $1-10$ that describe visually assessed characteristics of fine needle aspiration (FNA) samples. There are 458 patterns for benign (labeling as "2" in the dataset) and 241 patterns for malignant (labeling as "4") category.

Let $X = \{x_1, x_2, \cdots, x_{639}\}$. Since the samples are classified into two classes, hence we select two simple concepts for each feature, i.e., $M = \{m_{ij} | 1 \leq i \leq 9, 1 \leq j \leq 2\}$. Where $m_{i,1}, m_{i,2}$ are the simple concepts, "*small*", "*large*" associated with the feature f_i, respectively. We use the parameters $\varepsilon = 0.05, \delta = 0.5, \beta = 0.6$ in the tenfold cross-validation and follow the same procedure as the above wine data. The ten experimental results are listed in Table 10.5.

The average accuracy and the the standard deviation for the training set are 97.19% and $8.46e-006$, and its corresponding average accuracy for testing set and the standard deviation for the testing set are 97% and $6.12e-004$.

Table 10.5 The numbers of misclassifying samples and the number of rules for Breast Data

No. of experiment	1	2	3	4	5	6	7	8	9	10
No. of misclassifying training samples	14	19	17	16	20	20	18	18	18	17
No. of misclassifying testing samples	5	0	2	2	5	1	1	1	1	3
No. of rules	35	36	37	33	24	34	33	37	38	42

10.2.4 Experiment Studies of Parameters

In this section, we study how to select suitable values of the parameters ε, δ, and β according to the training sets to achieve the best performance of the classifiers. The influence of these parameters on the performance of the classifier are shown as the following Table 10.6 to 10.8 for the Wine, Iris and Breast data and the Figures 10.18 to 10.29. We can observe that a high accuracy can be achieved by the proposed algorithm for most parameters positioned in the range $0.05 \leq \varepsilon \leq 0.2$, $0.1 \leq \delta \leq \beta \leq 0.9$.

By analyzing Figures 10.51 to 10.93 presented in Appendix A, we arrive at the following conclusions: The accuracy of both training set and testing set is not sensitive to the choice of the parameter ε. Large differences between β and δ will result in low classification accuracy both on the training as well as the testing set.

The suitable numeric values of the parameters ε, δ, and β are selected by experimenting with the training data. We developed some guidelines regarding this choice.

Table 10.6 Experimental results for different values of the parameters of the Wine Data

ε	$\delta,\ \beta$	training variation	training average	testing variation	training average
0.05	$0.1 \leq \delta \leq \beta \leq 0.9$	0.0031321	92.545	0.0032816	90.861
0.1	$0.1 \leq \delta \leq \beta \leq 0.9$	0.0026383	94.653	0.0027112	92.605
0.15	$0.1 \leq \delta \leq \beta \leq 0.9$	0.003179	94.717	0.0028821	92.399
0.2	$0.1 \leq \delta \leq \beta \leq 0.9$	0.0038353	94.299	0.0034942	92.201

Table 10.7 Experimental results for different values of the parameters of the Iris Data

ε	$\delta,\ \beta$	training variation	training average	testing variation	testing average
0.05	$0.1 \leq \delta \leq \beta \leq 0.9$	0.0181	81.69	0.0184	80.10
0.1	$0.1 \leq \delta \leq \beta \leq 0.9$	0.0060	89.46	0.0064	87.95
0.15	$0.1 \leq \delta \leq \beta \leq 0.9$	0.0058	89.84	0.0062	88.34
0.2	$0.1 \leq \delta \leq \beta \leq 0.9$	0.0059	89.61	0.0063	88.22

Table 10.8 Experimental results for different values of the parameters of the Breast Data

ε	$\delta,\ \beta$	training variation	training average	testing variation	testing average
0.05	$0.1 \leq \delta \leq \beta \leq 0.9$	$2.1843e-006$	96.91	$4.9141e-006$	96.72
0.1	$0.1 \leq \delta \leq \beta \leq 0.9$	$4.6520e-006$	96.88	$1.2853e-005$	96.66
0.15	$0.1 \leq \delta \leq \beta \leq 0.9$	$5.2756e-006$	96.885	$1.4568e-006$	96.681

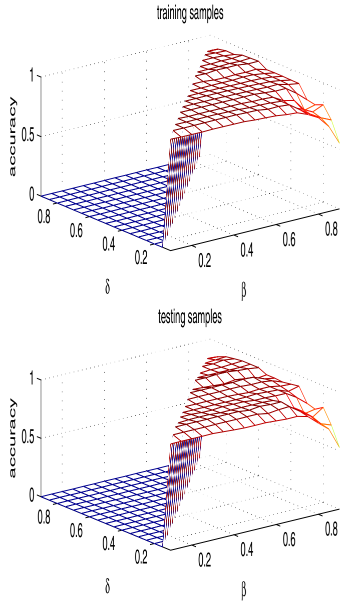


Fig. 10.18 Wine Data: The relationship between the accuracy and parameters β, δ with $\varepsilon = 0.05$

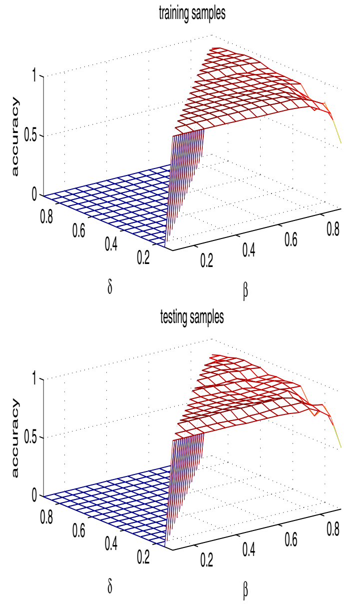


Fig. 10.19 Wine Data: The relationship between the accuracy and parameters β, δ with $\varepsilon = 0.1$

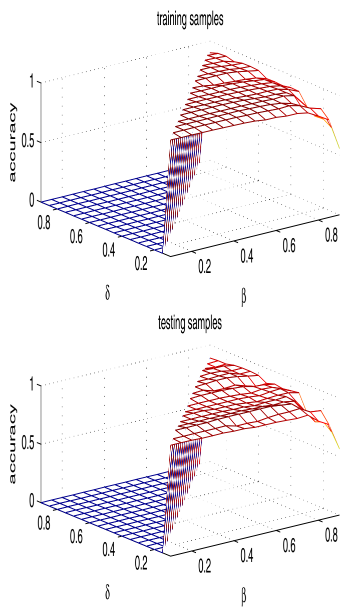


Fig. 10.20 Wine Data: The relationship between the accuracy and parameters β, δ with $\varepsilon = 0.15$

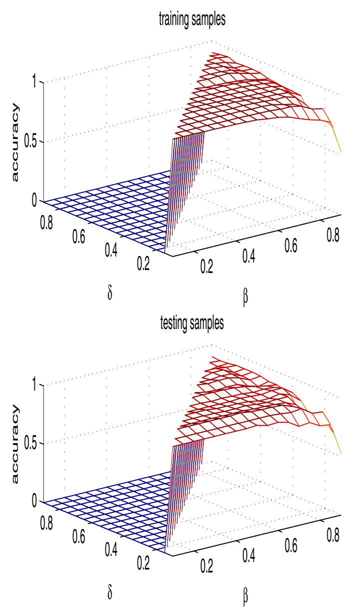


Fig. 10.21 Wine Data: The relationship between the accuracy and parameters β, δ with $\varepsilon = 0.2$

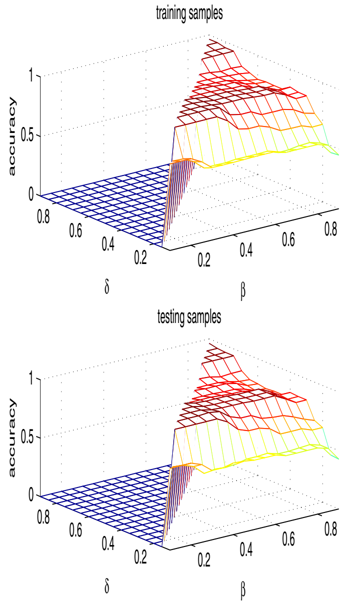


Fig. 10.22 Iris Data: The relationship between the accuracy and parameters β, δ with $\varepsilon = 0.05$

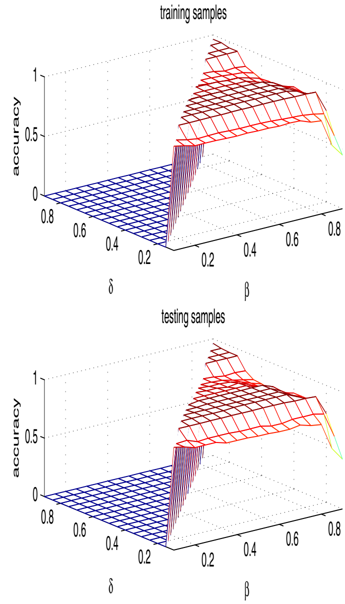


Fig. 10.23 Iris Data: The relationship between the accuracy and parameters β, δ with $\varepsilon = 0.1$

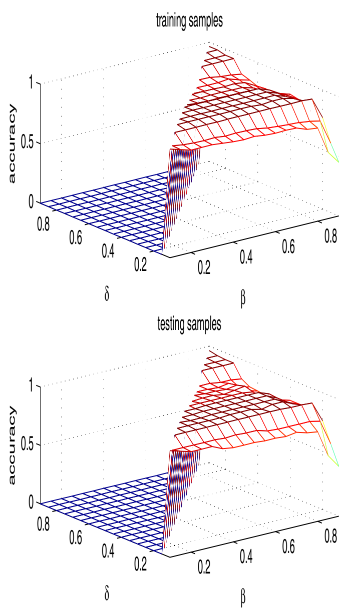


Fig. 10.24 Iris Data: The relationship between the accuracy and parameters β, δ with $\varepsilon = 0.15$

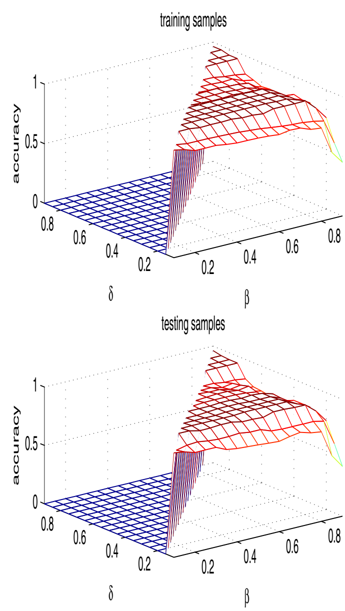


Fig. 10.25 Iris Data: The relationship between the accuracy and parameters β, δ with $\varepsilon = 0.2$

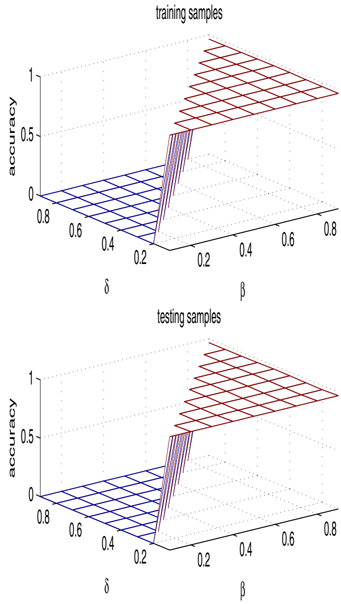


Fig. 10.26 Breast Data: The relationship between the accuracy and parameters β, δ with $\varepsilon = 0.05$

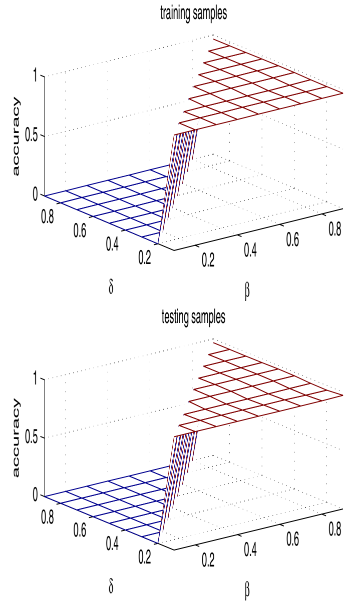


Fig. 10.27 Breast Data: The relationship between the accuracy and parameters β, δ with $\varepsilon = 0.1$

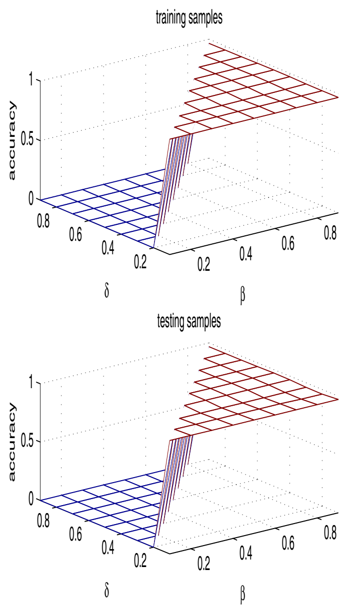


Fig. 10.28 Breast Data: The relationship between the accuracy and parameters β, δ with $\varepsilon = 0.15$

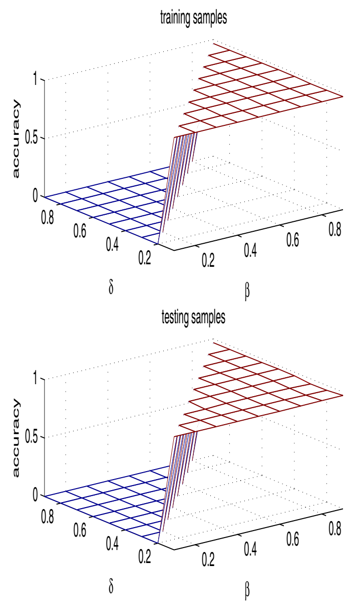


Fig. 10.29 Breast Data: The relationship between the accuracy and parameters β, δ with $\varepsilon = 0.2$

First, ε which corresponds to the smallest variation and high average accuracy for training data is preferred. For instance, $\varepsilon = 0.1$ in Table 10.6 is opted, and $\varepsilon = 0.15$ in Table 10.7 is opted. Then δ and β are selected by looking for the highest accuracy on the training data and the smallest number of the rules. Where the number of rules is $\sum_{1\leq k\leq l} ||\zeta_{X_k}||$ (refer to Definition 5.9).

10.2.5 Stability and Universality of the Coherence Membership Functions

In this section, each of the data sets: Iris data, Wine data and Breast data can be regarded as the observed data from a probability space $(\Omega, \mathscr{F}, \mathscr{P})$. The training data of the ten experiments completed for the one of Iris, Wine, and Breast can be regarded as the different data drawn from the same probability space. In what follows, we study the stability and universality of the membership functions determined by (4.40) in Theorem 4.6 via the membership functions of the fuzzy descriptions of each class obtained by the classifier design algorithm applying to the data sets in the above. Thus there are ten membership functions of each class in the ten experiments which obtained by different data. Following Figures 10.30-10.37, one can observe that the ten membership functions of ζ_{X_i} the fuzzy descriptions of the class i for a data set are quite similar, although different fuzzy descriptions $\zeta_{X_i} \in EM$ may be obtained for the different training samples. In virtue of Theorem 4.6, we know that when the number of the training samples approaches infinity, the ten membership functions will converge to one. These figures verify the stability of the coherence membership functions. In addition, each fuzzy description is obtained by the application of the classifier design to the training data (i.e.,90% of the total samples) and the membership functions shown in the figures are on the total samples. Thus these figures also show that universality of the coherence membership functions.

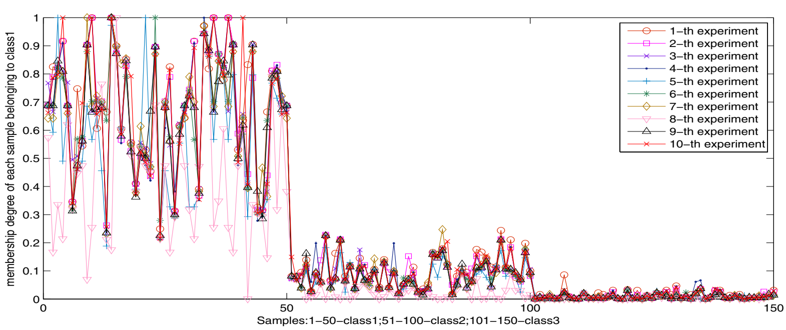


Fig. 10.30 Membership functions of the fuzzy descriptions of class 1 obtained by the algorithm applied to every training data of the ten experiments completed for the Iris Data

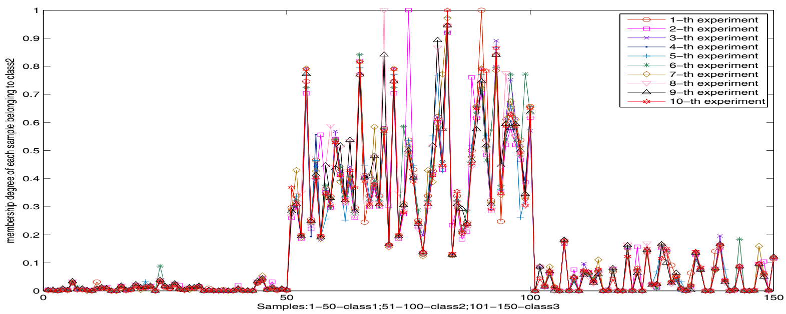


Fig. 10.31 Membership functions of the fuzzy descriptions of class 2 obtained by the algorithm applied to every training data of the ten experiments completed for the Iris Data

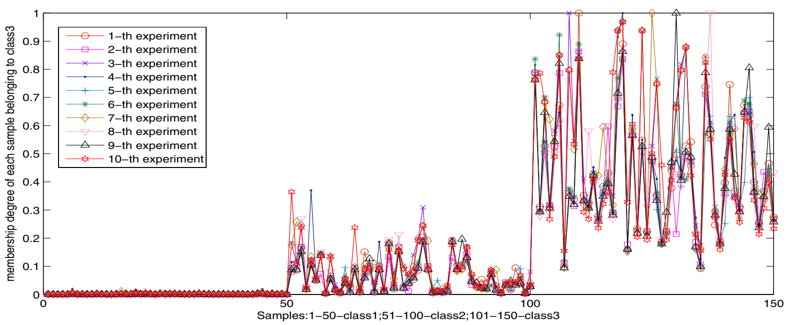


Fig. 10.32 Membership functions of the fuzzy descriptions of class 3 obtained by the algorithm applied to every training data of the ten experiments completed for the Iris Data

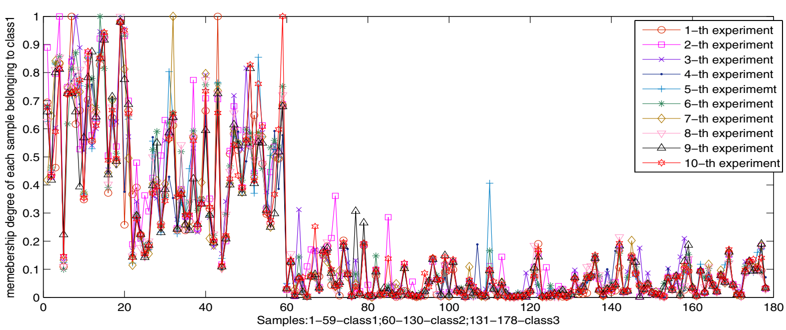


Fig. 10.33 Membership functions of the fuzzy descriptions of class 1 obtained by the algorithm applied to every training data of the ten experiments completed for the Wine Data

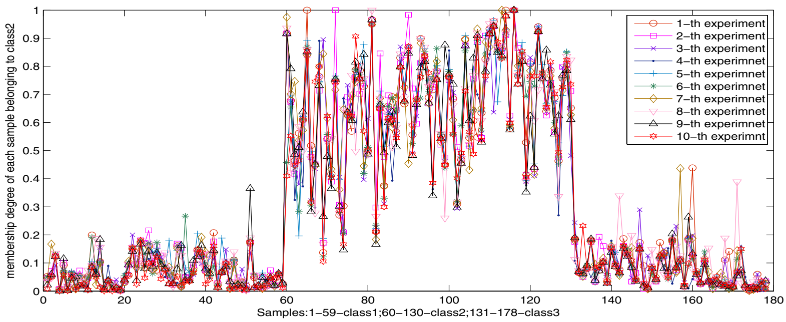


Fig. 10.34 Membership functions of the fuzzy descriptions of class 2 obtained by the algorithm applied to every training data of the ten experiments completed for the Wine Data

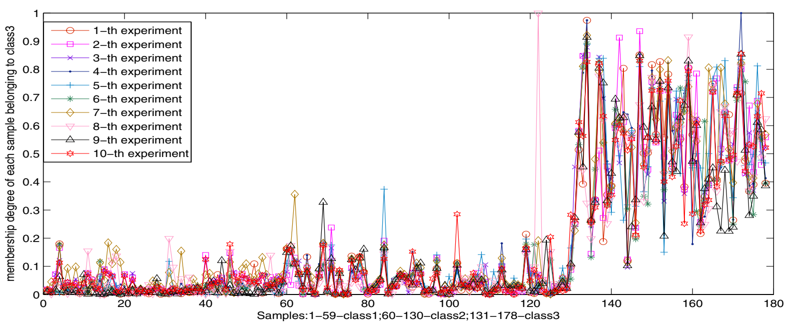


Fig. 10.35 Membership functions of the fuzzy descriptions of class 3 obtained by the algorithm applied to every training data of the ten experiments completed for the Wine Data

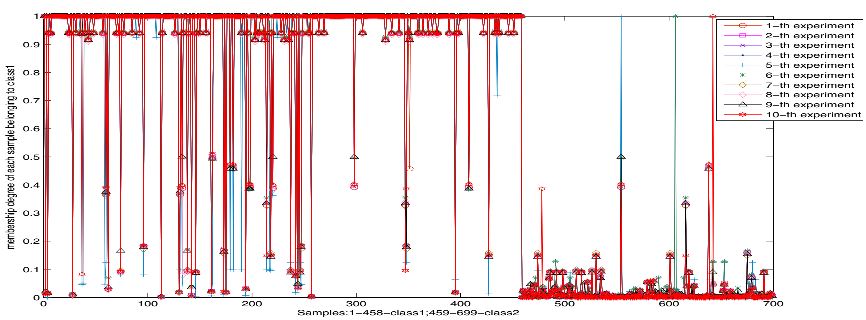


Fig. 10.36 Membership functions of the fuzzy descriptions of class 1 obtained by the algorithm applied to every training data of the ten experiments completed for the Breast Data

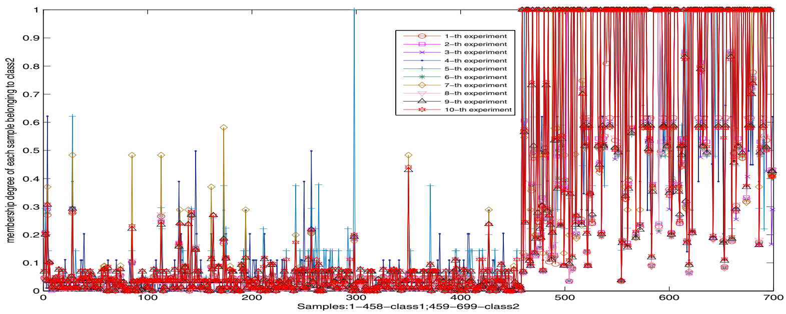


Fig. 10.37 Membership functions of the fuzzy descriptions of class 2 obtained by the algorithm applied to every training data of the ten experiments completed for the Breast Data

10.2.6 Comparative Analysis

In this section, the results of proposed algorithm are compared with two rule-based classification methods C4.5 rules [79] and RIPPER(Repeated Incremental Pruning to Produce Error Reduction) [11], BooR (a Boosting algorithm) [94] and a hybrid of rule-based methods and associative classifiers (CPAR)(Classification based on Predictive Association Rules) [112], CMAR(Classification based on Multiple Association Rules),2SARC1 and 2SARC2 (Two-Stage Approach to Classification) [3].

Rule-based classification approaches have been developed for decades in the machine learning community due to the readability of rules when compared to other classifiers. These algorithms use greedy techniques in the rule generation process. C4.5 searches the space for the best attribute according to the heuristic used, divides the search space according to the values of the attribute and then continues the process recursively in those subspaces. Rules are generated following the paths that cover the feature space. RIPPER [11] is built upon IREP (Incremental Reduced Error Pruning) algorithm [16]. Following IREP's strategy, RIPPER splits the training set in two sets. One of them is used to grow the rules and the other to prune the rules. The algorithm starts with an empty rule and it repeatedly adds conditions that maximize the information gain criterion. Once the rule is grown, conditions are deleted to maximize a function during pruning phase. When a rule has been discovered, all the examples that are covered by this rule are removed from the training set. The above process continues and we learn rules for the remaining training set.

CPAR [112] is a hybrid between associative classifiers and rule-based classifiers that use greedy techniques. It uses a greedy algorithm to search the space of attributes.The main difference is that it keeps all close-to-the-best attributes in rule generation, unlike rule-based methods which use only the best attribute.

CMAR [59], i.e., Classification based on Multiple Association Rules. The method extends an efficient frequent pattern mining method, FP-growth, constructs a class distribution-associated FP-tree, and mines large database efficiently. Moreover, it applies a CR-tree structure to store and retrieve mined association rules efficiently, and prunes rules effectively based on several measure such as confidence, correlation and database coverage. The classification is performed based on a weighted χ^2 analysis using multiple strong association rules.

The following Table 10.9 lists the accuracies over a tenfold cross-validation of various classifier designs which were reported in [3] and the proposed algorithm.

Table 10.9 Results of Several Classifiers

dataset	ARC	2SARC1	2SARC2	C4.5	Ripper	BooR	CBA	CMAR	CPAR	proposed
breast	96.42	96.13	95.85	94.71	95.28	95.85	96.28	96.40	96	97
iris	95.33	94.67	96	94	94	94.67	94.67	94	94.7	96
wine	88.79	95.50	97.18	92.12	92.12	96.67	94.96	95	95.5	98.31

In this section, a novel fuzzy classifier is proposed via AFS fuzzy logic and fuzzy entropy. The relationship between the parameters and the classifier quality is investigated through experimental studies and the optimum parameter selection principles are outlined. Considering the successive steps of the design of the classifier, one can observe that the computation and the procedure in each step are directly AFS fuzzy logic operations which are very interpretable and understandable. The final classifier results in some fuzzy concepts with well-defined semantics where each of them describing the character of a class. The experimental studies show that the proposed classifier can achieve very high accuracy both on the training set and the testing set and it is very robust to the changes in the value of the parameters. The proposed classifier can also be regarded as the knowledge representation of the data since it is simply constructed in terms of fuzzy sets with a well-defined semantics using AFS logic operations.

It is clear that the proposed classifier is just an initial stage of this approach and there is much room for possible improvement. With the strong mathematical background of AFS structure and the AFS algebra, the proposed classifier can be analyzed by many mathematical tools and implemented in a setting of parallel computing. This provides great opportunity to establish mathematical theorems for enhancing its practicality and efficiency.

10.3 AFS Fuzzy Classifier Based on Fuzzy Decision Trees

This section introduces a method to construct a fuzzy rule-based classifier via fuzzy decision trees. Given such theoretical underpinnings, these systems are referred to as AFS classifiers based on fuzzy decision trees. Compared with other fuzzy classifier systems, the classifier exhibits several essential advantages being of practical relevance. The reliability (relevance) of classification results is quantified by associated confidence levels (degrees). This quantification can be applied to the data sets with mixed data type attributes. The proposed algorithm consists of the following major steps: (a) estimation of the value of the optimal threshold to be used in the generation of the fuzzy decision trees; (b) generation and tuning of the fuzzy decision trees by information gain and the estimated threshold; and (c) pruning the rule-base by rule post-pruning. We have experimented with various data sets commonly discussed in the literature. We have also compared obtained results with those reported for C4.5 and C-decision trees. It has been found that the accuracy on test data is higher than the one produced by the other decision trees.

There have been numerous approaches to the development (extraction) of classification fuzzy rules from numerical data [24, 30, 97, 102, 110]. One of the quite often used alternatives is to construct a decision tree from the training data and afterwards extract rules from it. There are different types of decision trees. The well developed underlying methodology comes with efficient design techniques supporting their construction, cf. [8, 29, 74]. The decision trees generated by these methods have been dound useful in building knowledge-based expert systems. Due to the character of continuous attributes as well as various facets of uncertainty one has to take into consideration, there has been a visible trend to cope with the factor of fuzziness when carrying out learning from examples in the case of tree induction. In a nutshell, this trend gave rise to the generalizations known as fuzzy decision trees, cf. [25, 33]. The incorporation of fuzzy sets into decision trees enables us to combine the uncertainty handling and approximate reasoning capabilities of the former with the comprehensibility and ease of application of the latter. This combination augments the representation capabilities of decision trees with the knowledge component inherent to fuzzy logic subsequently leading to their robustness, noise immunity, and substantial applicability level in particular when dealing with situations we encounter a factor of uncertainty.

The existing fuzzy decision trees [33] assume that all domain attributes or linguistic variables come with some predefined fuzzy linguistic terms. This means that every node of the tree except for its root comes equipped with some fuzzy set that can be represented as a conjunction of several fuzzy linguistic terms. Here the conjunction operation is implemented by some standard logic operators encountered in fuzzy sets. Zhao and Hong [105] discretized continuous attributes using fuzzy numbers and the mechanisms of possibility theory. Pedrycz and Sosnowski [72], on the other hand, employed context-based fuzzy clustering for this purpose. Yuan and Shaw [110] induced a fuzzy decision tree by reducing classification ambiguity with fuzzy evidence. The input data is transformed using triangular membership functions formed around cluster centers obtained with the use of a Kohonen's feature map [38]. Wang et al. [101] presented optimization principles of fuzzy decision trees by minimizing the total number and an average depth of leaves, proving that the algorithmic complexity of the construction of a minimum tree is NP-hard. X. -Z. Wang et al. [102] studied heuristic algorithms for generating fuzzy decision trees. Mitra et al. [64] introduced a novel concept useful in measuring the goodness of a decision tree, which is expressed in terms of its compactness (size) and efficient performance. In general, the approaches shown above dwell on the generic design algorithm [33] or one of its variants. The improvement or enhancement of [33] comes with the combination of the tree with some other development machinery such as genetic algorithms, neural networks, mechanisms of fuzzy granulation, etc. An interesting commonality occurring across all of them is worth emphasizing:

such development algorithms require some knowledge about membership functions of the linguistic values of the attributes as well as specific aggregation operations (such as *t*-norms) prior to any optimization technique being considered and utilized. It becomes apparent that to significant extent the obtained fuzzy decision trees are pre-determined by the membership functions of the fuzzy terms and the fuzzy logic operators. Besides, like the classical decision trees, the class label of the node is determined by the label of the majority of the training samples falling into it. It is so simple that the difference of the membership degrees and the disproportion between the classes are ignored.

In this chapter, we use the fuzzy sets (membership functions) and underlying logic operations generated by AFS to eliminate potential subjective bias of the conventional fuzzy decision tree resulting from the use of different membership functions of the fuzzy terms. We focus on a new scheme of rule extraction. The scheme helps to deal with cases where we encountered imbalanced classes. The confidence degrees of a test samples lead to a generation of a certain confidence level associated with the resulting decision. The tuning and pruning methods help generate a sound and efficient fuzzy rule-based classifier. To offer a thorough comparative tested, we experimented with the algorithm using seven well-known real data sets coming from the UCI Repository for Machine Learning data [67], and compared the proposed decision tree with C-decision tree [75] and C4.5 [80].

10.3.1 Generation of Fuzzy Rules from AFS Decision Trees

In this section, the following assumptions are made. Let X be a set and M be a set of simple concepts on X. Let (M, τ, X) be an AFS structure of the data set and the weight function $\rho : X \to [0,1]$, $\rho(x) = 1$ for any $x \in X$. Then the set of coherence membership functions $\{\mu_\xi(x) \mid \xi \in EM\}$ can be obtained from Proposition 5.7.

$$\mu_\eta(x) = \sup_{i \in I} \frac{|A_i^\tau(x)|}{|X|}, \tag{10.28}$$

where $A_i^\tau(x)$ is calculated by (4.27). If $\sigma = 2^X$, for $W \in 2^X$, $\mathscr{M}(W) = |W|$ ($|W|$ is the cardinality of the set W, i.e., the number of elements in W) in Proposition 5.7.

The membership function defined by (10.28) depends only on the AFS structure of the data which is determined by the distribution of the data and the semantic interpretations of the simple concepts in M. Since EI algebra $(EM, \vee, \wedge)$ is closed for the logic operation $\vee$, $\wedge$ defined by (4.3) and (4.2), hence for any fuzzy concepts in EM, their membership functions and logic operations $\vee$, $\wedge$, (or, and) can be determined by (10.28) and $(EM, \vee, \wedge)$ is a logic system.

We use fuzzy logic operations defined by (4.3) and (4.2) as well as the membership functions expressed by (10.28) in the construction of fuzzy decision trees for such as the data shown as Table 10.10. The resulting classifiers are referred to as AFS decision trees.

Table 10.10 A Collection of Training Examples

Training Examples	$V_1 = inc$	$V_2 = emp$	$credit$
	U_1	U_2	Y
	u_j^1	u_j^2	y_j
x_1	0.20	0.15	0.00
x_2	0.35	0.25	0.00
x_3	0.90	0.20	0.00
x_4	0.60	0.50	0.00
x_5	0.90	0.50	1.00
x_6	0.10	0.85	1.00
x_7	0.40	0.90	1.00
x_8	0.85	0.85	1.00

10.3.1.1 Basic Notions

Let us recall some notions and definitions pertaining to the study presented in [33]. Each internal node of the tree comes with a branch for each linguistic value of the split variable, except when no training examples satisfy the fuzzy restriction.

In Table 10.10, x_i is a training example, U_i is the universe of discourse of the fuzzy variable or attribute V_i, the two fuzzy variables are $V_1 = income$ (abbreviated as inc) and $V_2 = employment$ (emp), Y is the universe of discourse of decision variable D_c. Before we define the fuzzy decision trees and rule-generate procedures, let us introduce some additional notations.

1) The set of fuzzy variables or attributes is denoted by

$$V = \{V_1, V_2, \cdots, V_n\}.$$

where V_i is a fuzzy variable over the universe of discourse U_i, i=1, 2, $\cdots$, n.

2) For each variable $V_i \in V$

- value of training example j is $u_j^i \in U_i$.
- D_i denotes the set of fuzzy terms (i.e., simple concepts) associating with V_i.
- $v_p^i \in D_i$ denotes the fuzzy term for the variable V_i. (e.g., v_{low}^{inc}, as necessary to stress the variable or with anonymous values—otherwise p alone may be used).

3) The set of fuzzy terms (simple concepts) for the decision variable is denoted by D_c. Each fuzzy term $v_k^c \in D_c$ is a fuzzy concept expressed over universe of discourse Y.

4) The set of training examples is

$$X = \{x_j \mid x_j = (u_j^1, u_j^2, \cdots, u_j^n, y_j)\}.$$

5) M is the set of all simple concepts

$$M = D_c \cup (\bigcup_{i=1}^{n} D_i).$$

(M, τ, E) is the AFS structure and EM is the EI algebra over M. In general, for each pair fuzzy terms $v_p^c, v_q^c \in D_c$ (on decision attribute Y), $v_p^c \neq v_q^c$, $\mu_{v_p^c \wedge v_q^c}(x) < \varepsilon$ (ε is a very small positive number), for any $x \in X$. This implies that the fuzzy terms in D_c implement a fuzzy classification on X, e.g., $X_{v_k^c}$, $v_k^c \in D_c$.

6) For each node N of the fuzzy decision trees

- F^N denotes the set of fuzzy restrictions on the path from the root to the node N, e.g.,
$$F^5 = \{[emp\ is\ high], [inc\ is\ high]\}$$
in Fig.10.38.
- V^N is the set of attributes appearing on the path leading to the node N
$$V^N = \{V_i \mid \exists p\ ([V_i\ is\ v_p^i] \in F^N)\}.$$
- β^N is a fuzzy concept in EM, and $\mu_{\beta^N}(x_j)$ is the membership degree of sample x_j in the node N, where
$$\beta^N = \prod_{m \in \{v_p^i \mid \exists p\ ([V_i\ is\ v_p^i] \in F^N)\}} m$$

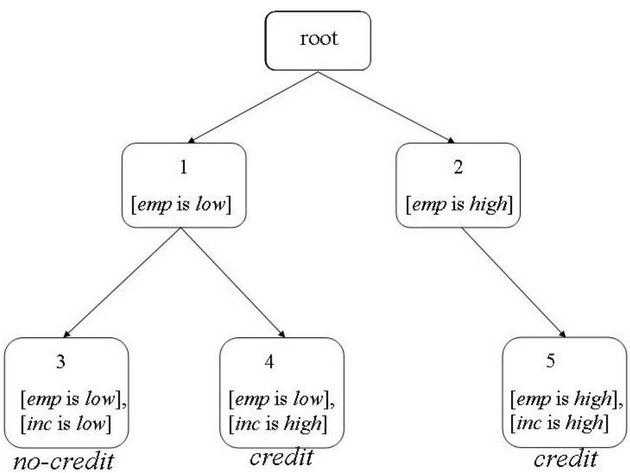


Fig. 10.38 An example of a fuzzy decision tree

- β_δ^N is the δ ($\delta \in (0, 1)$) cut set of fuzzy set β^N, i.e.,

$$\beta_\delta^N = \{x \in X \mid \mu_{\beta^N}(x) > \delta\}. \tag{10.29}$$

- $N|v_p^i$ denotes the particular child node of node N created by the use of the fuzzy attribute V_i to split N, $v_p^i \in D_i$.
- $S_{V_i}^N$ denotes the set of N's children when $V_i \in (V - V^N)$ is used for the split. Note that

$$S_{V_i}^N = \{(N|v_p^i) \mid v_p^i \in D_i^N\}, \quad D_i^N = \{v_p^i \in D_i \mid \exists x \in X,\ \mu_{\beta^N \wedge v_p^i}(x_j) > \delta\}; \tag{10.30}$$

  in other words, some fuzzy terms $v_p^i \in D_i$ which satisfy $\mu_{\beta^N \wedge v_p^i}(x) \leq \delta$ for any $x \in X$ may not be used to create sub-trees.
- $P_{v^c}^N$ denotes the example count for decision $v^c \in D_c$ in node N, where

$$P_{v^c}^N = \sum_{j=1}^{|X|} \mu_{\beta^N \wedge v^c}(x_j), \qquad P^N = \sum_{v^c \in D_c} P_{v^c}^N,$$
$$P_{v^c}^{N|v_p^i} = \sum_{j=1}^{|X|} \mu_{\beta^N \wedge v_p^i \wedge v^c}(x_j), \qquad P^{N|v_p^i} = \sum_{v^c \in D_c} P_{v^c}^{N|v_p^i}. \tag{10.31}$$

  It is important to note that unless the sets are such that the sum of all memberships for any is 1, $P_{v^c}^N \neq \sum_{v_p^i \in D_i} P_{v^c}^{N|v_p^i}$; that is, the membership sum from all children of N can differ from that of N. This is due to the existence of the fuzzy concepts, the total membership can either increase or decrease while building the tree.

- P^N and I^N denote the total example count and *information measure for the node* N, where I^N is the standard information content

$$I^N = -\sum_{v^c \in D_c} \frac{P_{v^c}^N}{P^N} \log \frac{P_{v^c}^N}{P^N} \tag{10.32}$$

- $G_{V_i}^N = I^N - I^{S_{V_i}^N}$ denotes the *information gain* when using the fuzzy attribute V_i to split N, where

$$I^{S_{V_i}^N} = \sum_{v_p^i \in D_i^N} \frac{P^{N|v_p^i}}{\sum_{v_p^i \in D_i^N} P^{N|v_p^i}} I^{N|v_p^i} \tag{10.33}$$

  is the weighted information content.

10.3.2 Construction of AFS Decision Trees

The AFS decision trees are constructed by the algorithm shown in Table 10.11. It consists of the following design steps: Discretization with Fuzzy Terms, Node Splitting Criterion, and Stopping Condition.

Table 10.11 The Algorithm of Building AFS Decision Trees

```
AFSDT=BuildTree(X, V, N, δ)
{ % X denotes the training data set with l classes, X = ⋃_1^l X_{v_k^c}; V is the set of fuzzy
  % variables or attributes; N denotes the current node; δ is the given threshold; an AFS
  % decision tree starts with N = ∅.
  1. calculate the information content I^N of node N;
  2. calculate the information gain G^N_{V_i} for each V_i from V;
  3. V_max = arg max_{V_i∈V} {G^N_{Vi}};
  4. V = V − V_max; % delete V_i from V.
  5. % check child and update the current node.
  for k = 1 : l
    if (∃x ∈ X, μ_{β^N ∧ v_k^max}(x) > δ)
      N = N ∧ v_k^max
      AFSDT=BuildTree(X, V, N, δ)
    end
  end
}
```

10.3.2.1 Discretization with Fuzzy Terms

The attributes need to be transformed into fuzzy terms. Suppose that the training samples belong to l different classes, then we assign l fuzzy terms (i.e., simple concepts) for each attribute V_i. The set of fuzzy terms for V_i is $D_i = \{v_1^i, v_2^i, \cdots, v_l^i\}$, the simple concept v_k^i with the semantics " *the value of V_i is closer to the kth cut point*". In the experimental studies, we will illustrate how to determine the values of the cut points.

10.3.2.2 Node Splitting Criterion

The growth process of the tree is guided by the maximum information gain. The information gain of the use of fuzzy attribute V_i to split the current node N is $G^N_{V_i} = I^N - I^{S^N_{V_i}}$, where I^N and $I^{S^N_{V_i}}$ are defined by (10.32) and (10.33). The fuzzy attribute V_i which exhibits the maximum information gain at the current node N is applied to split N. The children of the node N is the set $S^N_{V_i}$ (defined by (10.30)).

10.3.2.3 Stopping Condition

First, a given node N can be expanded if the samples in the set β^N_δ defined by (10.29) are not in the same class, otherwise, the node is not expanded any longer. The second stopping condition is self-evident: the current node N can be expanded if $V^N \neq V$, the set V^N is the attributes applied from root to the node N, V is the set of all the attributes. The final termination criterion is the information gain we monitor for each mode of the tree when the tree is begin built. In case when the maximum information gain at the current node N is negative or the set of N's children is empty, i.e., $S^N_{V_i} = \varnothing$, we stop expanding this node. The first two stopping criteria are a sort of precondition: if they are not satisfied, we stop expanding the node. The third one comes in a form of some postcondition: to make sure if it is satisfied, we have to expand the node first and then determine its value, if not satisfied, we should backtrack and refused to expand the particular node.

10.3.3 Rule Extraction and Pruning

After the AFS decision tree has been built for the given threshold δ, the extraction of the rule base is extracted via the algorithm shown in Table 10.12, which consists of the following steps: Rule Extraction, Pruning the Rule-Base.

Table 10.12 The Algorithm of Rule Extraction

```
Rule Extraction(X, AFSDT, δ)
{ % X denotes the training data set with l classes, X = ⋃_{1}^{l} X_{v_k^c}; AFSDT is an AFS
  % decision tree with t classifier node (terminal node).
  for i = 1 : t
    A_{r_i} = {x | x ∈ X, μ_{β^{N_i}}(x) ≥ μ_{β^{N_j}}(x), ∀ j = 1, 2, ···, t, j ≠ i}
  end
  for i = 1 : t
    if (A_{r_i} ≠ ∅)
      Class label of r_i = argmax_{1≤k≤l}{β_δ^{N_i} ∩ X_{v_k^c}}      % β_δ^{N_i} is defined by (10.29).
    else
      Class label of r_i = argmax_{1≤k≤l}{A_{r_i} ∩ X_{v_k^c}}
    end
  end
}
```

Rule Extraction

Each path starting from the root down to a classifier node (terminal node) is converted to a rule. Suppose the rules $r_1, r_2, \cdots, r_t$ are extracted form the fuzzy decision tree, the antecedent part of the rule r_i is a fuzzy set—-the conditions leading to the terminal node, i.e., $\beta^{N_i} \in EM$, where N_i is the terminal node of the corresponding path. The class labels of the rules are essential to the classifier. The class label methods for fuzzy decision tree in the existing papers are not suitable for the decision tree discussed here. Instead we consider the following scheme for the fuzzy set β^{N_i} representing the antecedent part of rule r_i. Let us introduce the notation

$$A_{r_i} = \{x \mid x \in X,\ \mu_{\beta^{N_i}}(x) \geq \mu_{\beta^{N_j}}(x),\ \forall\ j = 1,2,\cdots,t, j \neq i\}$$

$|A_{r_i}|$ measures the amount of training samples covered by the rule r_i. However, $A_{r_i} = \varnothing$ does not imply that rule r_i is incorrect. The condition $A_{r_i} = \varnothing$ implies that the contribution of rule r_i to the classification process is not so significant in the current rule set. Its importance to the classification process may increase in the process of pruning. Thus, the class label of the rule r_i is computed as $\arg\max_{1\leq k\leq l}\{|X_{r_i}\cap X_{v_k^c}|\}$, where $X_{v_k^c}$ is the set of the samples of the kth class,

$$X_{r_i} = \begin{cases} A_{r_i}, & \text{if } A_{r_i} \neq \varnothing, \\ \beta_\delta^{N_i}, & \text{otherwise,} \end{cases} \tag{10.34}$$

where $\beta_\delta^{N_i}$ is defined by (10.29) for the above given threshold δ.

10.3.3.1 Pruning the Rule-Base

The rules directly extracted from the AFS decision tree, may include redundant structures as well as involve poorly performing rules, which should be removed from the rule-base to enhance an overall performance of the classifier and improve its efficiency. In what follows, we prune the rule-base making use of the available training data:

1 Remove each rule from the rule-base, and classify the training data using the remaining rules.
2 Delete the rule, whose corresponding remaining rules have the maximal increase of accuracy on training data.
3 Repeat steps 1—2 and stop the pruning if the resulting pruned rule-base performs worse than the original one when applied to the training data.
4 Using the fuzzy logic operation "$\vee$" defined by (4.2), sum all fuzzy concepts representing the antecedents of the rules with the same consequent.

Thus, for a training data set with l classes, we can represent the rule-base with l fuzzy concepts $\xi_1, \xi_2, \cdots, \xi_l \in EM$. For each class $v_k^c \in D_c$, $k = 1, 2, \cdots, l$, a rule can be obtained and read as:

Rule k: If x is ξ_k, then x belongs to the class v_k^c, $k = 1, 2, \cdots, l$.

Example 10.1. We consider the pruning process of the tree, which is built in Experiment 1 of Wine data set. The un-pruned tree is shown in Figure 10.39.

The rules directly extracted from the AFS decision tree by the algorithm Rule extraction are listed as follows:

r_1: If x is $m_{19}m_{28}$, then x belongs to class 2;
r_2: If x is $m_{19}m_{29}$, then x belongs to class 3;
r_3: If x is $m_{19}m_{30}$, then x belongs to class 3;
r_4: If x is $m_{20}m_{28}$, then x belongs to class 2;
r_5: If x is $m_{20}m_{29}m_{39}$, then x belongs to class 1;
r_6: If x is $m_{21}m_{37}$, then x belongs to class 2;
r_7: If x is $m_{21}m_{38}m_1$, then x belongs to class 2;
r_8: If x is $m_{21}m_{38}m_3$, then x belongs to class 1;
r_9: If x is $m_{21}m_{39}m_2m_{29}$, then x belongs to class 1;
r_{10}: If x is $m_{21}m_{39}m_2m_{30}$, then x belongs to class 1;
r_{11}: If x is $m_{21}m_{39}m_3$, then x belongs to class 1.

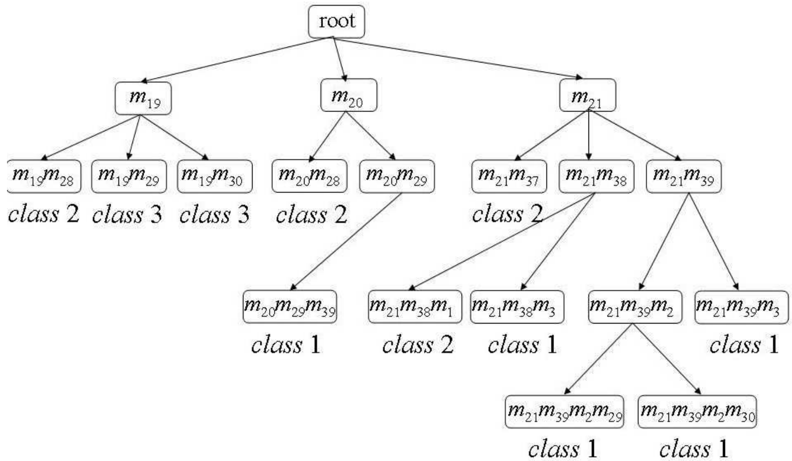


Fig. 10.39 An AFS decision tree with $\delta = 0.63$ in the experiment of Wine data set

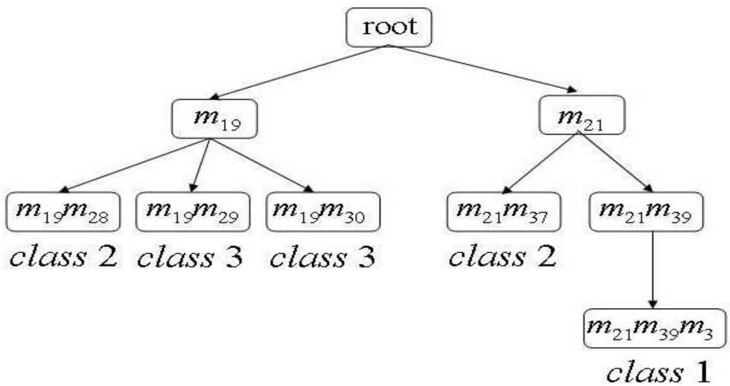


Fig. 10.40 The AFS decision tree (pruned) with $\delta = 0.63$ for the Wine data set

The classification rate achieved on training set with the rule-base shown above is 92.25%. The standard deviation of the classification rate reported for the training set while the pruning process is shown in Table 10.13. After pruning, the classification rate on the training set increased to 97.89%. Then we sum all fuzzy concepts representing the antecedents of the rules with the same consequent in the pruned rule-base, using the logic operation "$\vee$" defined by (4.2). The pruned AFS decision tree is shown in Figure 10.40 and the pruned rule-base is shown as follows:

Rule 1: If x is ξ_1, then x belongs to class 1;
Rule 2: If x is ξ_2, then x belongs to class 2;
Rule 3: If x is ξ_3, then x belongs to class 3;
where $\xi_1 = m_{21}m_{39}m_3$, $\xi_2 = m_{19}m_{28} + m_{21}m_{37}$, $\xi_3 = m_{19}m_{29} + m_{19}m_{30}$.

Table 10.13 The standard deviation of the classification rate on the training Set and the corresponding number of rules obtained in the Pruning Process for the Wine data

	1	2	3	4	5	6
Training accuracy	96.48%	97.18%	97.89%	97.89%	97.89%	97.89%
Deleted rule	r_4	r_5	r_9	r_7	r_8	r_{10}

10.3.4 Determining the Optimal Threshold δ

The amount of the "*trivial detailed information*" included in the decision trees can be controlled by value of the threshold $\delta \in (0, 1)$. The larger the value of δ, the less overlap occurs between the membership functions. In other words, we can conclude that the larger the value of δ, the more "*trivial detailed information*" becomes filtered out (ignored). In the procedure of building AFS decision trees, different values of δ can produce different trees. In fact, following (10.30), we know that when we use a smaller δ in the growth process of a tree, we build a bigger tree (consisting of more nodes). Clearly the performance of the tree depends on the threshold δ. Here, *optimal threshold* δ means that the pruned rule-base induced by the AFS decision tree with it has the highest accuracy on testing data. However, it is very difficult to determine the optimal threshold δ on a basis of the existing training data. Thus, in this study, we compute the *sub-optimal value of the threshold* δ by following Fitness Index $F(\delta)$. More specifically the sub-optimal threshold δ maximizes the following *Fitness Index* computed for the training data.

$$F(\delta) = |X| \cdot Classification\ rate - \delta \cdot Number\ of\ nodes \tag{10.35}$$

$|X|$ is the number of training samples, "$Classification\ rate$" is the classification accuracy reported on the training samples for the rule-base obtained by threshold δ via the algorithm of *Build an AFS Decision Tree* shown in Table 10.11 and the algorithm of *Rule Extraction* shown in Table 10.12, and "$Number\ of\ nodes$" is the total nodes of the pruned tree.

Example 10.2. Let $X = \{x_1, x_2, \cdots, x_{14}\}$ be a set of 14 people characterized by some attributes which are described by real numbers, Boolean values and the order relations are shown in Table 10.14. The number i in the "*black*", "*white*", "*yellow*" columns which corresponds to some $x \in X$ implies that the hair color of x has ordered ith following our perception of the color. For example, the numbers in the column *"white"* imply the order $(>)$

$$x_7 > x_{10} > x_4 = x_8 = x_{11} = x_{13} > x_2 = x_9 = x_{12} = x_{14} > x_5 > x_1 = x_3 = x_6$$

i.e., when moving from right to left, the relationship states how strongly the hair color under consideration resembles white color. In this order, $x_i > x_j$ (e.g., $x_7 > x_{10}$) states that the hair of x_i is closer to the white color than the color of hair the individual x_j. We illustrate the proposed scheme using the data shown in Table 10.14. The data set consists of 14 samples with 8 *credit* samples and 6 *not credit* samples. Each

Table 10.14 Description of attributes

	age	*height*	*self-appraisement*	*salary*	*estate*	*male*	*black*	*white*	*yellow*	*credit*
x_1	20	1.9	90	1	0	1	6	1	4	0
x_2	13	1.2	32	0	0	0	4	3	1	0
x_3	50	1.7	67	140	34	0	6	1	4	1
x_4	80	1.8	73	20	80	1	3	4	2	1
x_5	34	1.4	54	15	2	1	5	2	2	0
x_6	37	1.6	80	80	28	0	6	1	4	1
x_7	45	1.7	78	268	90	1	1	6	4	1
x_8	70	1.65	70	30	45	1	3	4	2	1
x_9	60	1.82	83	25	98	0	4	3	1	1
x_{10}	3	1.1	21	0	0	0	2	5	3	0
x_{11}	8	1.4	45	0	0	0	3	4	3	0
x_{12}	19	1.73	56	1	0	1	4	3	4	0
x_{13}	40	1.6	50	30	20	1	3	4	2	0
x_{14}	23	2	80	19	5	0	4	3	2	0

sample is described by nine condition attributes (see Table 10.14) and one decision attribute (*credit*). We take all 14 samples as training samples. On each attribute V_i, two fuzzy terms are specialized, since the training samples form two classes. The set of fuzzy terms for attribute V_i is $D_i = \{v^i_{small}, v^i_{large}\}$, and the set of fuzzy terms for the decision variable (decision attribute) is $D_c = \{v^c_{credit}, v^c_{notcredit}\}$. Let $M = \{m_1, m_2, \cdots, m_{22}\}$ be the set of simple concepts on U, where $m_{2i-1} = v^i_{small}$ with the semantics "*the value on* V_i *is small*", $m_{2i} = v^i_{large}$ with the semantic meaning "*the value on* V_i *is large*" (i=1, 2, $\cdots$, 9) and $m_{21} = v^c_{credit}$, $m_{22} = v^c_{notcredit}$. Now, we can establish the AFS structure (M, τ, X), where τ is defined by (4.26) while the membership functions of the fuzzy concepts in EM are defined by (10.28).

To start with, we choose a threshold level δ=0.8 (as it will be shown later on, the value of this threshold will be optimized). The root node starts with all the training samples without any restrictions, that is $\beta^{root} = \varnothing$. By using the node splitting criterion, the 5th attribute "*estate*" is selected to split the root node and the children of the root node form the set $\{(root|v^{estate}_{samll}), (root|v^{estate}_{large})\}$. According to the stopping condition, we obtain the decision tree shown as Figure 10.41. Node 1 contains 5 samples and all of them belong to the "*not credit*" category. There are 3 samples at node 2 which belong to the "*credit*" category. 6 samples were left out and they were neither assigned to node 1 nor node 2. This implies that these samples are not typical and significant enough for the predefined value of the threshold ($\delta = 0.8$). They may be included in the AFS fuzzy decision tree when considering smaller values of the threshold δ.

In what follows, we show how to extract rules from this tree. In order to determine the class labels of the rules with the antecedent represented by fuzzy concepts corresponding to node 1 and node 2, using (10.34), we calculate $X_{r_1} = \{x_1, x_2, x_5, x_{10}, x_{11}, x_{12}, x_{14}\}$, $X_{r_2} = \{x_3, x_4, x_6, x_7, x_8, x_9, x_{13}\}$. The class label of r_1 is $\arg\max_{k\in\{notcredit,\ credit\}}\{X_{r_1} \cap X_{v^c_k}\} = notcredit$ and the class label of r_2 is

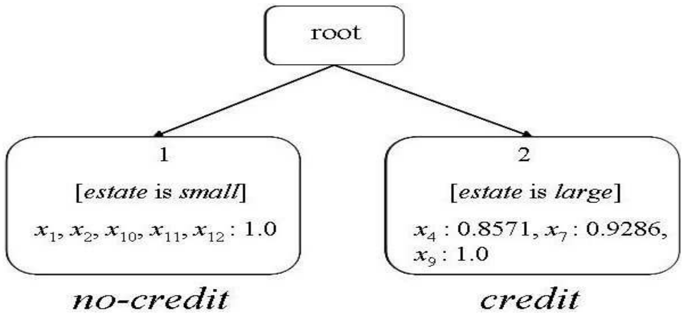


Fig. 10.41 The AFS fuzzy decision tree with $\delta = 0.8$

$\arg\max_{k\in\{notcredit,\ credit\}}\{X_{r_2}\cap X_{v_k^c}\} = credit$, where $X_{v_{credit}^c}$, $X_{v_{notcredit}^c}$ are the sets of *credit* samples and *not credit* samples. After pruning, we have *Rule 1*, *Rule 2* as follows:

Rule 1: If x is ξ_1, then x belongs to the class of *not credit*;
Rule 2: If x is ξ_2, then x belongs to the class of *credit*.
where $\xi_1 = v_{small}^{estate}$, $\xi_2 = v_{lagre}^{estate}$ and the membership functions of the fuzzy concepts ξ_1, ξ_2 are shown in Table 10.15.

Table 10.15 Membership degree of samples belonging to ξ_1, ξ_2

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
μ_{ξ_1}	1	1	0.36	0.21	0.64	0.43	0.14	0.29	0	1	1	1	0.5	0.57
μ_{ξ_1}	0	0	0.71	0.86	0.43	0.64	0.93	0.79	1	0	0	0	0.57	0.5

Next, let the threshold assume lower value of $\delta = 0.43$. In figure 10.42, the root node is split by the 5th attribute "*estate*", the children of the node is the set $\{(root|v_{samll}^{estate}),\ (root|v_{large}^{estate})\}$. The node 1 contains 8 samples have membership

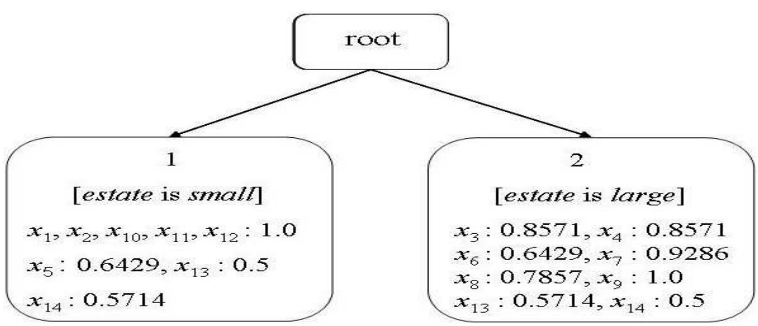


Fig. 10.42 The first layer of the AFS fuzzy decision tree with $\delta = 0.43$

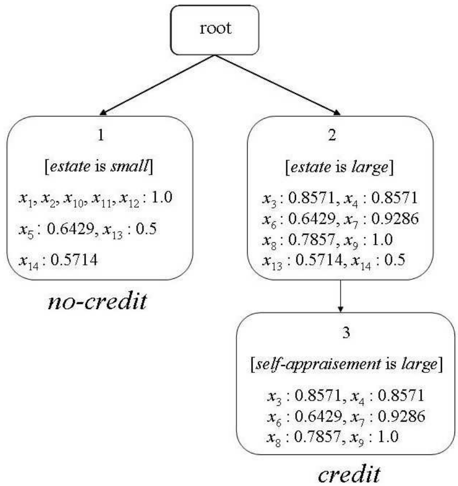


Fig. 10.43 The AFS fuzzy decision tree with $\delta = 0.43$

degrees larger than $\delta = 0.43$ and the 8 samples all belong to *not credit*. According to the stopping condition, the decision tree stops growing at this node. Unlike at node 1, the tree at node 2 will continue to grow. There are 8 samples which belong to different classes while falling into node 2, and the node 2 should be split further. The 3rd attribute "*self-appraisement*" is selected by the node splitting criterion via the formula (10.33). The child of node N_2 is $(N_2|v_{large}^{self-appraisement})$.

In figure 10.43, there are 6 samples falling into node 3 and these samples belong to the "*credit*" category. According to the stopping condition, the growth of the decision tree is terminated. After rule extraction and pruning, we have two rules, *Rule 1* and *Rule 2* as follows:

Rule 1: If x is ξ_1, then x belongs to the class of *not credit*;
Rule 2: If x is ξ_2, then x belongs to the class of *credit*.
where $\xi_1 = v_{small}^{estate}$, $\xi_2 = v_{large}^{estate} v_{large}^{self-appraisement}$. The accuracy of the tree reported on the training data is 100%.

Comparing the threshold $\delta = 0.8$ with the case $\delta = 0.43$, we find that the detailed information included in the AFS decision trees can be effectively controlled by the value of this threshold.

Now we use the Fitness Index (10.35) to determine the optimal value of the threshold δ. The plot of this index is displayed in Figure 10.44; clearly $\delta = 0.43$ leads to the maximization of the index.

For comparison, the C4.5 decision tree produced the following rules

Rule 1: If x is *estate* ≤ 20, then x is *not credit*;
Rule 2: If x is *estate* > 20, then x is *credit*.

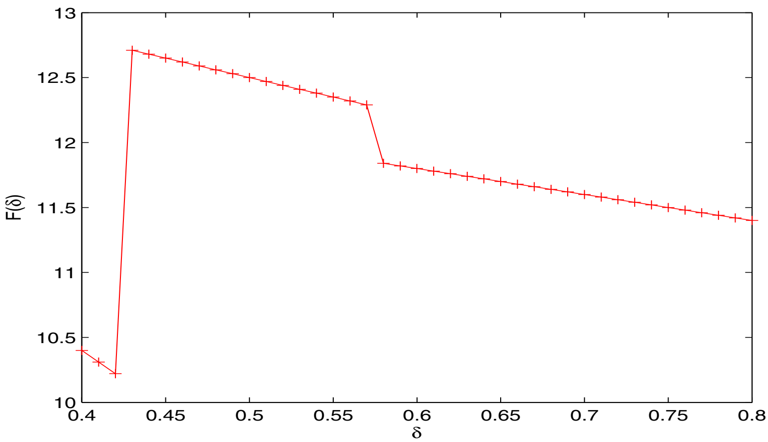


Fig. 10.44 The Fitness Index of the AFS fuzzy decision trees versus the threshold δ; the values of the threshold are shown here in the [0.4, 0.8] interval

The accuracy reported for the training data is also 100%. One can observe that the rules extracted by C4.5 are similar to the rules extracted from AFS decision tree with $\delta = 0.8$.

10.3.5 Inference of Decision Assignment and Associated Confidence Degrees

We should note that for each concept $\xi_i \in EM$ representing the antecedent of the *Rules i*, the universe of discourse of its membership function defined by (10.28) is the training sample set X. In order to predict the class label of the new samples which are not included in X, we need the fuzzy concept ξ expressed over the entire input space $U_1 \times U_2 \times \cdots \times U_n \subseteq R^n (X \subseteq U_1 \times U_2 \times \cdots \times U_n)$. In what follows, for each fuzzy concept $\xi \in EM$, we expand its universe of discourse X to $U_1 \times U_2 \times \cdots \times U_n$. For each $x = (u^1, u^2, \cdots, u^n) \in U_1 \times U_2 \times \cdots \times U_n$, $\xi = \sum_{i \in I} (\prod_{m \in A_i} m) \in EM$, the lower bound and upper bound of membership function of fuzzy concept ξ are defined over $U_1 \times U_2 \times \cdots \times U_n$ as follows:

$$\mu_\xi^U(x) = \sup_{i \in I} \left(\inf_{g \in U_{A_i}^x} (\mu_{A_i}(g)) \right), \qquad \mu_\xi^L(x) = \sup_{i \in I} \frac{|L_{A_i}^x|}{|X|}, \tag{10.36}$$

where $U_{A_i}^x \subseteq X$, $L_{A_i}^x \subseteq X$, $i \in I$ are defined as

$$U_{A_i}^x = \{x_j \in X | (x_j, x) \in R_m, \forall\, m \in A_i\},$$

$$L_{A_i}^x = \{x_j \in X | (x, x_j) \in R_m, \forall\, m \in A_i\},$$

here R_m is the binary relation of simple concept m by Definition 4.2. We call $\mu_\xi^L(x)$ the *lower bound of membership function of* ξ and $\mu_\xi^U(x)$ serves as the *upper bound of membership function of* ξ. It is clear that the following assertions hold:

- $\mu_\xi^U(x) \geq \mu_\xi^L(x)$ for each $x \in U_1 \times U_2 \times \cdots \times U_n$.
- $\mu_\xi^U(x) = \mu_\xi(x) = \mu_\xi^L(x)$ for each $x \in X$.

In virtue of (10.36), we can expand the fuzzy concept $\xi = \sum_{i\in I}(\prod_{m\in A_i} m) \in EM$, from universe of discourse X to the universe of discourse $U_1 \times U_2 \times \cdots \times U_n$. Therefore, by the fuzzy rule-base, we can establish fuzzy-inference systems whose input space is $U_1 \times U_2 \times \cdots \times U_n$. The membership functions $\mu_\xi^U(x)$, $\mu_\xi^L(x)$ are dependent on the distribution of training examples and the AFS fuzzy logic.

When we are provided with a new pattern $x \in U_1 \times U_2 \times \cdots \times U_n$ with unknown class label, we calculate the membership degree $\mu_\xi^L(x)$ by (10.36) and x belongs to the class $\arg\max_{v_k^c \in D_c}\{\mu_{\xi_k}^L(x)\}$, $k = 1, 2, \cdots, l$. Furthermore, considering the training samples, the confidence degree of the membership degree $\mu_\xi(x)$ estimated by $\mu_\xi^L(x)$ is defined as follows:

$$C_\xi(x) = 1 - (\mu_\xi^U(x) - \mu_\xi^L(x)). \tag{10.37}$$

The confident degree $C_\xi(x)$ quantifies confidence we associate with $\mu_\xi^L(x)$, the estimate of the membership degree of x belonging to ξ, $x \in U_1 \times U_2 \times \cdots \times U_n$. For sample x, the closer the upper bound of membership function of ξ to the lower bound, the larger value of $C_\xi(x)$ we have. In fact, the value of $C_\xi(x)$ depends on how many training samples are similar to x considering the fuzzy concept ξ. Larger value of $C_\xi(x)$ advises us to trust $\mu_\xi^L(x)$ as the membership degree of x belonging to ξ. Especially, if $x \in X$, $C_\xi(x) = 1$, then there exist a training sample x_0 such that the values of both x_0 and x on the attributes associating to ξ are equal. For each testing sample x, we know that it belongs to the class $\arg\max_{v_k^c \in D_c}\{\mu_{\xi_k}^L(x)\}$, $k = 1, 2, \cdots, l$, which is determined by $\mu_{\xi_k}^L(x)$, the estimate of all membership degree of x belonging to ξ_k. In order to achieve a high reliability prediction of the class label of x, every confidence degree $C_\xi(x)$ of $\mu_{\xi_k}^L(x)$, $k = 1, 2, \cdots, l$, has to be high. Thus the reliability of the classification result of each testing sample can be measured by the confidence degrees. In practice we can refuse to classify the testing samples whose confidence degree belonging to some $\mu_{\xi_k}^L(x)$ is low or require more information to produce high confidence degrees.

10.3.6 Experimental Studies

In this section, seven well-known data sets are used in our experiments. They come from the Machine Learning repository [67], which makes the experiments fully reproducible and facilitates further comparative analysis. The description of the pertinent data sets is covered in Table 10.16. Experiments 1 concerns the well-known Wisconsin breast cancer data sets, Iris data sets and Wine data sets. In experiment 2 the data sets are as follows: (a) Pima-diabetes, (b) Ionosphere, (c) Hepatitis, and (d) Auto data. In the experiments, for each experiment set, five complete five-fold cross validations are carried out, the cases are partitioned into five equal-sized subsets with similar class distributions. In turn, each subset was then used as test data for the AFS decision tree inference systems generated from the remaining four subsets. All missing values were replaced by the averages of the corresponding attributes.

Table 10.16 Descriptions of data sets from UCI repository

No.	Data set	Classes	Sizes	Missing values	Num. of attributes
1	Breast-W	2	699	Yes	9
2	Iris	3	150	No	4
3	Wine	3	178	No	13
4	Pima	2	768	No	8
5	Ionosphere	2	351	No	34
6	Hepatitis	2	155	Yes	19
7	Auto	3	398	Yes	8

10.3.6.1 Experiment 1

A Wisconsin Breast Cancer: The data set consists of 699 samples which are classified two classes, 458 *benign* samples and 241 *malignant* samples. Each sample is described by nine attributes: a) *clump thickness*, b) *uniformity of cell size*, c) *uniformity of cell shape*, d) *marginal adhesion*, e) *single epithelial cell size*, f) *bare nuclei*, g) *bland chromatin*, h) *normal nucleoli*, and i) *mitoses*. In the data set, the values of the sixth attributes of 16 samples are missing. X is the training set, on each attribute V_i, two fuzzy terms are specialized, since the training samples are two classes. The set of fuzzy terms for attribute V_i is $D_i = \{v^i_{small}, v^i_{large}\}$, and the set of fuzzy terms for the decision variable (class attribute) is $D_c = \{v^c_{benign}, v^c_{malignant}\}$. Let $M = \{m_1, m_2, \cdots, m_{20}\}$ be the set of simple concepts on U, Where $m_{2i-1} = v^i_{small}$ with the semantics "*the value on* V_i *is small*", $m_{2i} = v^i_{large}$ with the interpretation "*the value on* V_i *is large*"$(i = 1, 2, \cdots, 9)$ and $m_{19} = v^c_{benign}, m_{20} = v^c_{malignant}$. Now, we can establish the AFS structure (M, τ, X), where τ is defined by (4.26).

The tree with sub-optimal thresholds δ obtained by the Fitness Index (10.35) starts with all the training samples without any restrictions, $\beta^{root} = \varnothing$. The tree grows following the node splitting criterion, and the stopping condition shown as the algorithm of *Build an AFS Decision Tree* shown as Table 10.11. The trees in the five-fold cross validations induced by the sub-optimal thresholds have 56.8 (average of the five experiments) terminal nodes, so we extract a rule-base with 56.8 rules (on average). The average classification accurate rate on training set is 95.28% and the average accuracy on testing set is 93.51%. Then, we prune the rule-base. The number of rules has been reduced to 5.9 (on average), and for this case, the percentage of correct classification on training set by the rule-base is increase to 96.78% (on average), the average accuracy on testing data set is increase to 95.65%. The membership functions of the fuzzy concepts ξ_1, ξ_2 which are the antecedents of

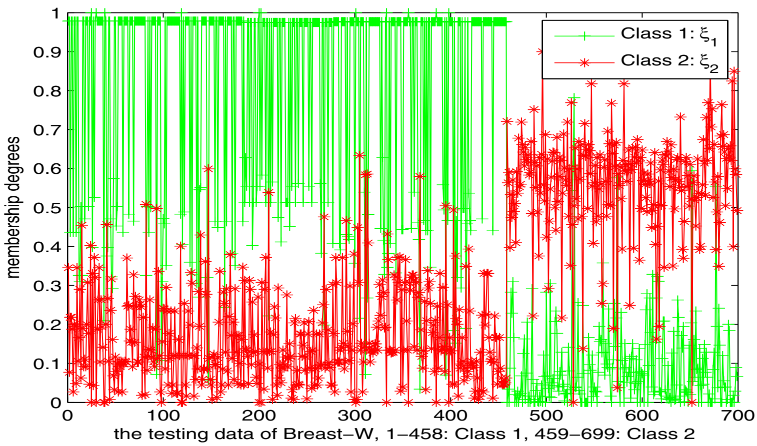


Fig. 10.45 The membership functions of ξ_1, ξ_2 in the 5th experiment of Breast data

Rule 1 and *Rule 2* are shown in Figure 10.45. The rule-base with optimal δ consists of 59.7 rules. The average classification rate on the training samples is 95.19% and on testing data is 93.96%. After pruning, the number of rules are reduced to 8.4 (on average), and the percentage of correct classification on training samples is increase to 96.52%, while 96.54% on testing samples (on average). In general, the high accuracy on training data does not imply that high accuracy on testing data, since the tree may be overfitting. Thus we estimated the optimal threshold δ by (10.35) to avoid overfitting. The percentage of misclassification on testing data of each experiment in the five fold experiment is shown in Table 10.17, where AFS_1 and AFS_2 are the results of the sub optimal threshold and optimal threshold, respectively.

Table 10.17 The percentage of misclassification of five experiments and the number of rules and the number of nodes of Breast-W

	Error: Training data (%)			Error: Testing data (%)			Number of rules			Number of nodes		
	AFS_1	AFS_2	C4.5	AFS_1	AFS_2	C4.5	AFS_1	AFS_2	C4.5	AFS_1	AFS_2	C4.5
1	3.04	3.68	2.44	4.29	3.72	5.28	7.0	7.4	8.2	31.6	30.8	15.4
2	3.15	2.72	1.44	4.15	3.58	6.44	6.0	8.6	12.6	27.4	37.6	24.2
3	3.54	4.08	1.98	4.86	3.43	5.16	5.2	8.2	10.0	20.8	31.4	19.0
4	3.00	3.61	2.02	4.43	3.43	6.00	6.8	8.6	10.2	31.6	34.8	19.4
5	3.36	3.29	1.96	4.01	3.15	5.88	4.6	9.4	10.4	22.6	41.8	19.8
mean	3.22	3.48	1.97	4.35	3.46	5.75	5.9	8.4	10.3	26.8	35.3	19.6

Figure 10.46-a shows the distribution of the number of testing samples in the 2nd experiment of the five-fold cross validations fall into different square regions of the confidence degree of the estimate of membership degree of the samples belonging to ξ_1, ξ_2 which are the fuzzy concepts to describe class 1 and class 2, respectively. Figure 10.46-b shows the distribution of the misclassification rate of testing samples

in the 2nd experiment of the five-fold cross validations fall into different square regions of confidence degree of the estimate of the membership degree of the samples belonging to ξ_1, ξ_2, the fuzzy descriptions of the classes. By Figure 10.46-a, one can observe that the confidence degrees (defined by (10.37)) of the membership degrees (defined by (10.36)) of almost of the testing samples (448 of 699) belonging to ξ_1, ξ_2 are larger than 0.9 and just 6 misclassified testing samples fall into this region. The misclassification rate in this region is 1.34%. This implies that the classification result of a testing sample x with $C_{\xi_1}(x) \geq 0.9$, $C_{\xi_2}(x) \geq 0.9$ is of high reliability.

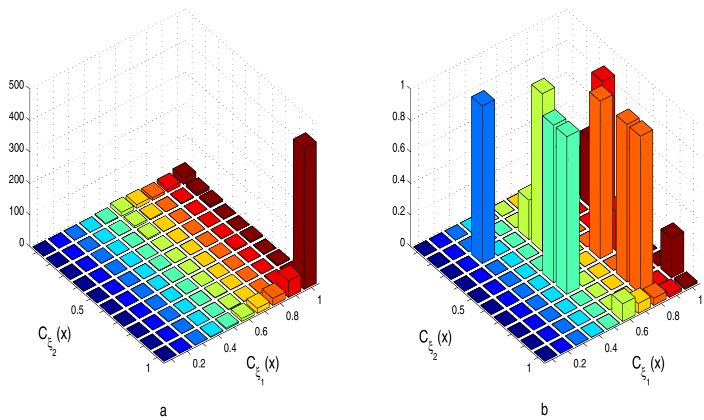


Fig. 10.46 a) The distribution of the number of testing samples falling into different square regions of the confidence degree of the estimate of membership degree of the samples belonging to ξ_1, ξ_2; b) The distribution of the misclassification rate of testing samples falling into different square regions of the estimate of confidence degree of the membership degree of the samples belonging to ξ_1, ξ_2

B Iris Plants Database: The data set consists of 150 samples involving three classes, that is 50 *setosa* samples, 50 *versicolour* samples and 50 *virginica* samples. Each sample is described by four attributes: a) *sepal length*, b) *sepal width*, c) *petal length*, d) *petal width*. X is the training set, on each attribute V_i, three fuzzy terms are specialized, since the training samples are three classes. The set of fuzzy terms for attribute V_i is $D_i = \{v^i_{small}, v^i_{mid}, v^i_{large}\}$, and the set of fuzzy terms for the decision variable (class attribute) is $D_c = \{v^c_{setosa}, v^c_{versicolour}, v^c_{virginica}\}$. Let $m_{3i-2} = v^i_{small}$ with the semantics "*the value on V_i is small*", $m_{3i} = v^i_{large}$ with the semantics "*the value on V_i is large*", $m_{3i-1} = v^i_{mid}$ with the semantics "*the value is closer to the mediacy on V_i*" $(i = 1, 2, 3, 4)$ and $m_{13} = v^c_{Setosa}, m_{14} = v^c_{Versicolour}, m_{15} = v^c_{Virginica}$.

The percentage of misclassification in five experiments and the number of rules and the number of nodes are summarized in Table 10.18, where AFS_1 and AFS_2 are the results of the sub optimal threshold and optimal threshold, respectively. The trees in the five-fold cross validations induced by the sub-optimal thresholds have 3 (average of the five experiments) terminal nodes, so we extract a rule-base with 3

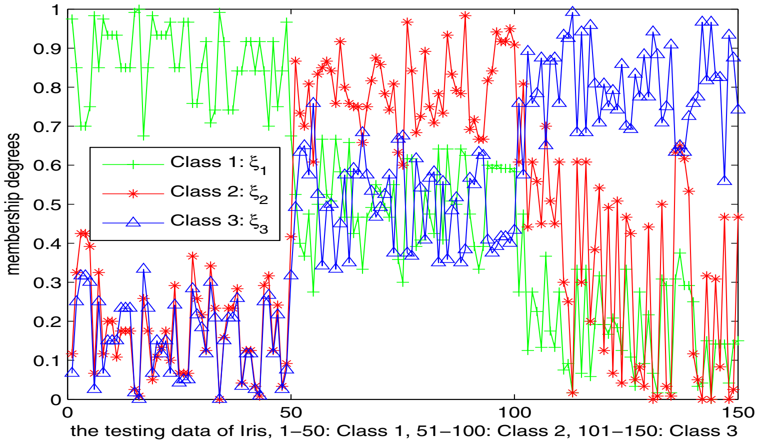


Fig. 10.47 The membership functions of ξ_1, ξ_2, ξ_3 in the 1st experiment of Iris data

rules (on average). The average classification rate on the training set is 96.83% and the average accuracy on the testing set is 96.00%. After pruning, the number of rules is 3 (on average), and the percentage of correct classification on training set is 96.83% (on average), the accuracy on the testing is 96.00%. The membership functions of the fuzzy concepts ξ_1, ξ_2, ξ_3 which are the antecedents of *Rule 1*, *Rule 2* and *Rule 3* are shown in Figure 10.47. The rule-base with optimal δ consists of 6.4 rules. The average accuracy is 94.97% and the accuracy is 96.67%for the training and testing sets, respectively. After pruning, the number of rules are reduced to 4.0 (on average), and the accuracy on training set is increase to 95.73% (average), the accuracy on testing set is 98.00%.

Table 10.18 Percentage of misclassification of five experiments and the number of rules and the number of nodes of Iris

	Error: Training data (%)			Error: Testing data (%)			Number of rules			Number of nodes		
	AFS_1	AFS_2	C4.5	AFS_1	AFS_2	C4.5	AFS_1	AFS_2	C4.5	AFS_1	AFS_2	C4.5
1	3.17	3.50	1.84	5.33	2.67	4.68	3.0	4.0	4.8	5.4	8.6	8.6
2	3.00	5.00	1.84	3.33	0.67	7.34	3.0	4.0	4.4	5.6	7.6	7.8
3	3.83	4.50	1.84	3.33	2.00	5.98	3.0	4.2	4.4	5.2	8.8	7.8
4	3.00	4.00	1.66	3.33	2.00	4.66	3.0	3.8	4.4	5.8	7.6	7.8
5	2.83	4.33	2.00	4.67	2.67	4.66	3.0	3.8	4.4	5.6	7.0	7.8
mean	3.17	4.27	1.84	4.00	2.00	5.47	3.0	4.0	4.5	5.5	7.9	8.0

C Wine Recognition Data: The data set consists of 178 samples which are classified three classes, class 1: 59 samples, class 2: 71 samples and class 3: 48 samples. Each sample is described by thirteen attributes. X is the training set, on each attribute V_i, three fuzzy terms are specialized, since the training samples are three classes. The set of fuzzy terms for attribute V_i is $D_i = \{v^i_{small}, v^i_{mid}, v^i_{large}\}$, and the set of fuzzy terms for the decision variable (class attribute) is $D_c = \{v^c_1, v^c_2, v^c_3\}$. Let $m_{3i-2} = v^i_{small}$ with the semantics "*the value on V_i is small*", $m_{3i} = v^i_{large}$ with the semantics "*the value on V_i is large*", $m_{3i-1} = v^i_{mid}$ with the semantics "*the value is closer to the mediacy on V_i*" ($i = 1, 2, \cdots, 13$) and $m_{40} = v^c_1, m_{41} = v^c_2, m_{42} = v^c_3$. Thus, the set of all fuzzy terms is $M = D_c \cup (\bigcup_{i=1}^{13} D_i) = \{m_1, m_2, \cdots, m_{39}, m_{40}, m_{41}, m_{42}\}$. Now, we can establish the AFS structure (M, τ, X), where τ is defined by (4.26).

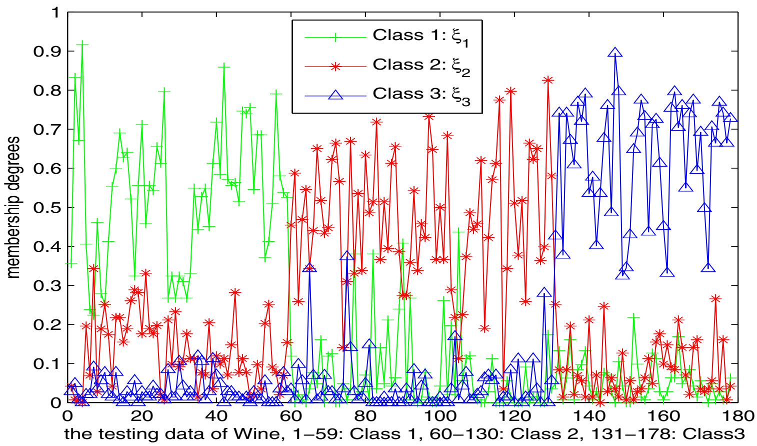


Fig. 10.48 The membership functions of ξ_1, ξ_2, ξ_3 in the 3rd experiment of Wine data

The percentage of misclassification of five experiments and the number of rules and number of nodes are summarized in Table 10.19, where AFS_1 and AFS_2 are the results of the sub optimal threshold and optimal threshold respectively. The trees in the five-fold cross validations induced by the sub-optimal thresholds have 32.2 (average of the ten experiments) terminal nodes, so we extract a rule-base with 32,2 rules (on average). The average classification rate on training samples is 98.23% and the accuracy on testing set is 94.38%. After pruning, the number of rules are reduced to 6.5 (on average), and the percentage of correct classification on training samples is increase to 99.27% (on average) and the accuracy on testing set is 95.17%. The membership functions of the fuzzy concepts ξ_1, ξ_2, ξ_3 which are the antecedents of *Rule 1*, *Rule 2* and *Rule 3* are shown in Figure 10.48. The rule-base with optimal δ consist of 15.0 rules. The average accuracy on training set is 94.94% and the average accuracy on testing set is 94.38%. After pruning, the number of rules are reduced to 6.1 (on average), and the average accuracy on training set is increase to 96.86%, the average accuracy on testing set is 98.09%.

Table 10.19 The Percentage of Misclassification of Five Experiments and the Number of Rules and the Number of Nodes of Wine

	Error: Training data (%)			Error: Testing data (%)			Number of rules			Number of nodes		
	AFS_1	AFS_2	C4.5	AFS_1	AFS_2	C4.5	AFS_1	AFS_2	C4.5	AFS_1	AFS_2	C4.5
1	0.84	5.34	1.26	6.18	1.69	9.02	6.4	6.0	5.8	15.8	11.8	10.6
2	0.70	2.38	1.40	2.81	2.25	8.44	6.0	5.8	5.8	15.8	14.2	10.6
3	0.14	2.67	0.84	4.49	1.69	6.24	8.2	6.6	5.6	19.8	14.8	10.2
4	0.56	2.67	1.40	5.62	2.25	8.42	6.4	5.4	5.8	16.2	11.6	10.6
5	1.41	2.67	1.68	5.06	1.69	10.58	5.4	6.6	5.6	15.0	14.4	10.2
mean	0.73	3.14	1.32	4.83	1.91	8.54	6.5	6.1	5.7	16.5	13.4	10.4

The performance of the rule-base extracted by AFS decision tree is summarized in Table 10.20, where AFS_1 and AFS_2 are the results of the sub optimal threshold and optimal threshold, respectively. Comparing the case Not Pruned with the case Pruned, we find that pruning of the rule-base improves the performance of the rule-base both on the training set and the testing set. It implies that there are some improper rules in the original rule-base, which are incorrectly classifying some samples both on training set and testing set. And there are many redundant rules in the original rule-base, so through pruning the number of rules in the rule-base could be decreased.

Table 10.20 Performance of the Rule-base on Different Data Sets

Data sets	Error: Training data (%)		Error: Testing data (%)		Number of rules		Number of nodes	
	Not Pruned	Pruned	Not Pruned	Pruned	Not Pruned	Pruned	Not Pruned	Pruned
Breast-W AFS_1	4.72	3.22	6.49	4.35	56.8	5.9	114.8	26.8
Breast-W AFS_2	4.81	3.48	6.04	3.46	59.7	8.4	121.3	35.3
Iris AFS_1	3.17	3.17	4.00	4.00	3.0	3.0	5.5	5.5
Iris AFS_2	5.03	4.27	3.33	2.00	6.4	4.0	11.3	7.9
Wine AFS_1	1.73	0.73	5.62	4.83	32.2	6.5	61.7	16.5
Wine AFS_2	5.06	3.14	5.62	1.91	15.0	6.1	27.4	13.4

10.3.6.2 Experiment 2

It is of interest to compare the results produced by the AFS decision tree with those obtained by C-decision tree and C4.5. In this analysis, the results obtained by C-decision tree and C4.5 are reported by [75]. The results are summarized in Table 10.21, where AFS_1 and AFS_2 are the results of the suboptimal threshold and optimal threshold respectively.

Overall, we note that the accuracy on testing set of AFS decision tree with suboptimal threshold δ are better than the C-decision trees at the level of 1%-6% and better than C4.5 at the level of 1%-16%(except for the auto data set), although the accuracy on training set is less than C-decision tree and C4.5. For Hepatitis C4.5 achieves a training error of 6.46%, the test error percentage is 43.86% which is 16% larger than that achieved by the AFS decision tree, and C-decision tree have

Table 10.21 AFS-decision Tree, C-decision Tree and C4.5: a comparative analysis for several machine learning data sets: (a) Pima-diabetes, (b) Ionosphere, (c) Hepatitis, (d) Auto data

(a)

Type of tree and their structural parameters	Error : Training data	Error : Testing data	Number of nodes
C4.5 rev.8	16.01%(average) 2.36%(st. deviation)	27.95%(average) 2.75%(st. deviation)	43(average) 9.52(st. deviation)
C-decision tree c=5 clusters, 6 iterations	10.26%(average) 1.32%(st. deviation)	27.4%(average) 3.21%(st. deviation)	30
C-decision tree c=3 clusters, 5 iterations	13.02%(average) 1.33%(st. deviation)	28.18%(average) 3.78%(st. deviation)	15
AFS_1	21.34%(average) 0.52%(st. deviation)	24.74%(average) 0.82%(st. deviation)	22(average) 7.33(st. deviation)
AFS_2	22.51%(average) 0.31%(st. deviation)	21.59%(average) 0.53%(st. deviation)	18(average) 0.67(st. deviation)

(b)

Type of tree and their structural parameters	Error : Training data	Error : Testing data	Number of nodes
C4.5 rev.8	1.78%(average) 1.52%(st. deviation)	13.54%(average) 4.60%(st. deviation)	23.8(average) 7.69(st. deviation)
C-decision tree c=4 clusters, 6 iterations	10.79%(average) 1.39%(st. deviation)	15.49%(average) 6.75%(st. deviation)	24
C-decision tree c=2 clusters, 2 iterations	13.14%(average) 2.67%(st. deviation)	16.62%(average) 4.61%(st. deviation)	4
AFS_1	7.86%(average) 0.67%(st. deviation)	15.16%(average) 0.91%(st. deviation)	27.25(average) 1.89(st. deviation)
AFS_2	10.81%(average) 0.53%(st. deviation)	10.60%(average) 0.37%(st. deviation)	21(average) 2.12(st. deviation)

(c)

Type of tree and their structural parameters	Error : Training data	Error : Testing data	Number of nodes
C4.5 rev.8	6.46%(average) 0.85%(st. deviation)	43.86%(average) 7.05%(st. deviation)	45(average) 7.87(st. deviation)
C-decision tree c=2 clusters, 6 iterations	17.58%(average) 3.34%(st. deviation)	36.13%(average) 0.08%(st. deviation)	12
C-decision tree c=9 clusters, 3 iterations	24.84%(average) 5.21%(st. deviation)	34.19%(average) 3.68%(st. deviation)	27
AFS_1	1.68%(average) 1.15%(st. deviation)	27.23%(average) 7.43%(st. deviation)	52.8(average) 5.30(st. deviation)
AFS_2	5.90%(average) 2.00%(st. deviation)	17.03%(average) 2.22%(st. deviation)	39(average) 8.58(st. deviation)

Table 10.21 (*continued*)

(d)

Type of tree and their structural parameters	Error : Training data	Error : Testing data	Number of nodes
C4.5 rev.8	8.58%(average) 3.89%(st. deviation)	25.9%(average) 4.08%(st. deviation)	139.8(average) 21.5(st. deviation)
C-decision tree c=19 clusters, 7 iterations	31.4%(average) 1.86%(st. deviation)	34.3%(average) 4.56%(st. deviation)	133
C-decision tree c=13 clusters, 7 iterations	13.14%(average) 2.67%(st. deviation)	16.62%(average) 4.61%(st. deviation)	91
AFS_1	20.83%(average) 0.35%(st. deviation)	26.48%(average) 1.60%(st. deviation)	22.5(average) 4.81(st. deviation)
AFS_2	23.25%(average) 0.49%(st. deviation)	22.36%(average) 0.59%(st. deviation)	17.1(average) 1.29(st. deviation)

a classification error on testing data 34.19% which is 6% larger than that achieved by the AFS decision tree. The standard deviation of the error is closer to C-decision tree and C4.5 for AFS decision tree. The nodes of AFS decision tree are more than C-decision tree and C4.5 for the following reasons: the depth of AFS decision tree is controlled by one threshold δ (we just use the same threshold in each node to control the size of the tree). Thus, for a well fitting of training data, more 'trivial detailed information' have to be considered, and a low level of the threshold should be used in the procedure of tree building, then we get a large tree with more nodes.

10.3.7 Structure Complexity

AFS decision trees are compact structures. The AFS decision tree has six leaf nodes shown in Figure 10.40. Figure 10.49 shows two AFS decision trees using Breast-W dataset (with classification accuracy of 95.71% and 94.29%) in the five-fold cross validation. The AFS decision trees may have complicated tree structure as each AFS decision tree may have up to $|V|^l$ leaf nodes, i.e., for Breast-W dataset, the number may be up to $2^9 = 512$. After pruning they usually, however, have not so many nodes as they are fully spanned. Figure 10.50 shows an AFS decision tree using Wine Recognition dataset in one of the five-fold cross validations. Such a tree has only 5 leaf nodes, much less than the maximum of $3^{13} = 1,594,323$. For a rough comparison, the average number of leaf nodes of FDTs constructed using Wine Recognition dataset is 59 [21]. This reveals that AFS decision trees, especially after pruning, can have more compact tree structures than FDTs. However, it can be argued that FDTs may be simplified by tree pruning as well.

In general, AFS decision trees with large threshold δ not only produce high classification accuracy, but also preserve compact tree structures, while AFS decision

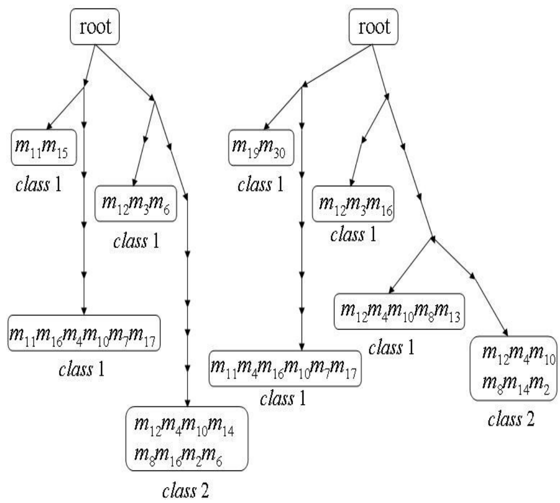


Fig. 10.49 Two AFS decision tress of Breast-W with the threshold $\delta = 0.63$ and $\delta = 0.65$, respectively

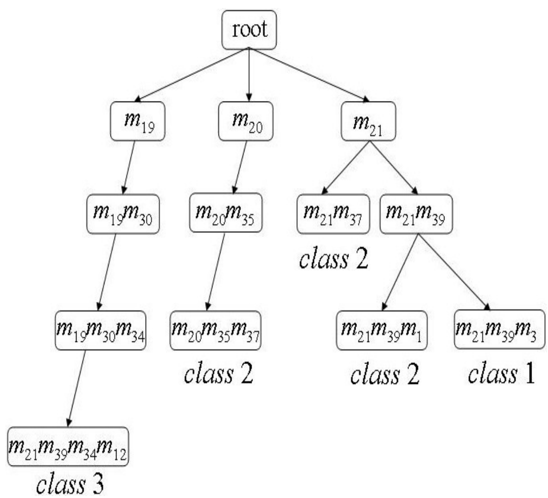


Fig. 10.50 An AFS decision tree of Wine with threshold $\delta = 0.45$

trees with smaller threshold δ can produce even better accuracy, but as a compromise produce more complex tree structures.

In this section, we have introduced and studied the AFS fuzzy rule-based classifier. We presented a way of building the AFS-decision trees, and elaborated on a way in which the rules can be extracted from the tree and pruned. We introduced the Fitness Index to estimate the threshold δ which is used to control the design of the AFS decision tree. We considered the fuzzy sets (membership functions) and the underlying logic operators generated by AFS to eliminate potential subjective bias in the construction of tree. The experiments demonstrated that the obtained results outperformed those produced by the C4.5 and the C-decision trees. We also showed the effectiveness of the rule extraction scheme. Interestingly even if the tree cannot result in an initial rule-base of good quality, the pruned rule-base can lead to a much higher performance which is consistently better both on the training and testing data.

Exercises

Exercise 10.1. Apply the design of fuzzy classifiers to other data and try to prune the terms in the fuzzy description of each class.

Exercise 10.2. Study how the parameters and the number of the simple concepts on each feature influence the prediction results.

Open problems

Problem 10.1. How to apply the new techniques of feature selection, concept categorization and characteristic description developed in chapter 9 to the design of fuzzy classifiers in this chapter?

Problem 10.2. For each classifier presented in this chapter, the fuzzy concept $\xi_{C_i} \in EM$ which is the fuzzy description of class C_i determines the membership degree of a sample belongingness to class C_i. How to apply the probability distribution of the observed data X and the membership function of ξ_{C_i} determined by (4.41) or (4.43) in Theorem 4.6 to classify any sample in the whole space?

Problem 10.3. How to design the efficient algorithms to find the fuzzy descriptions satisfying (10.1).

Problem 10.4. If the distribution of the data is given in advance, how to estimate the error bounds of the fuzzy classifiers in this chapter?

Problem 10.5. How to prove the relationship between the prediction accuracy and the confident degree $C_\xi(x)$ defined by (10.37) if the probability distribution of the data is given in advance?

Appendix A

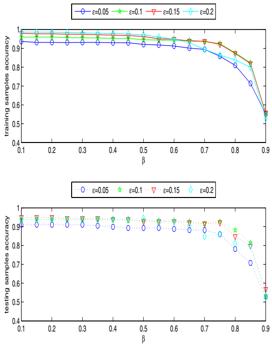


Fig. 10.51 Wine Data: Classifying accuracy with $\delta = 0.1$

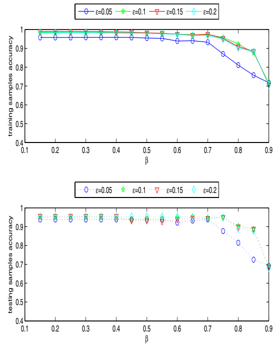


Fig. 10.52 Wine Data: Classifying accuracy with $\delta = 0.15$

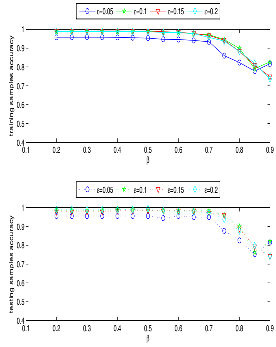


Fig. 10.53 Wine Data: Classifying accuracy with $\delta = 0.2$

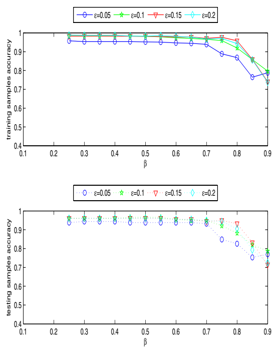


Fig. 10.54 Wine Data: Classifying accuracy with $\delta = 0.25$

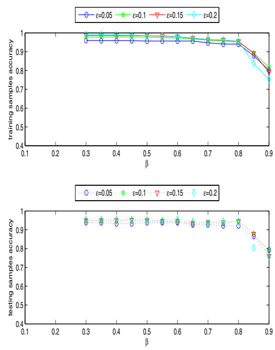


Fig. 10.55 Wine Data: Classifying accuracy with $\delta = 0.3$

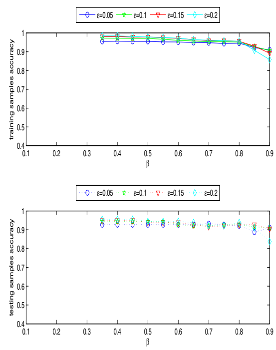


Fig. 10.56 Wine Data: Classifying accuracy with $\delta = 0.35$

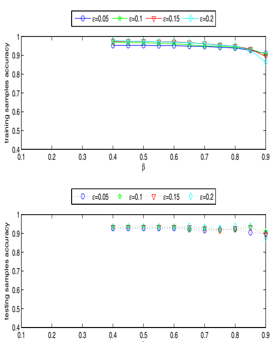


Fig. 10.57 Wine Data: Classifying accuracy with $\delta = 0.4$

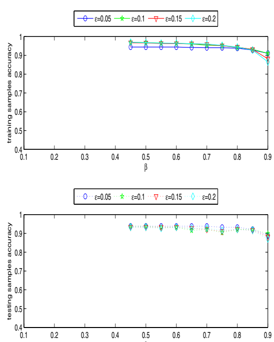


Fig. 10.58 Wine Data: Classifying accuracy with $\delta = 0.45$

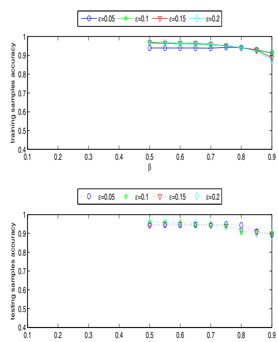


Fig. 10.59 Wine Data: Classifying accuracy with $\delta = 0.5$

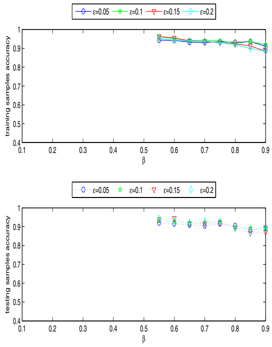


Fig. 10.60 Wine Data: Classifying accuracy with $\delta = 0.55$

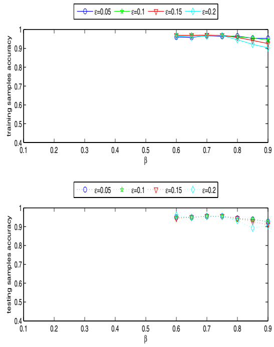


Fig. 10.61 Wine Data: Classifying accuracy with $\delta = 0.6$

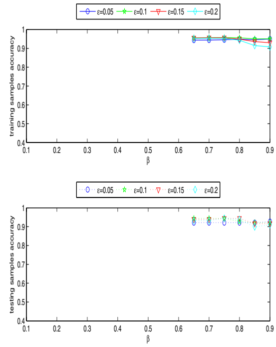


Fig. 10.62 Wine Data: Classifying accuracy with $\delta = 0.65$

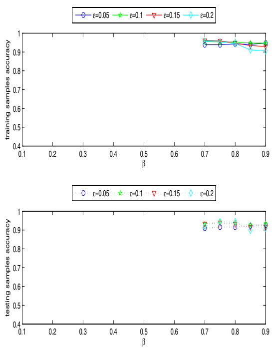


Fig. 10.63 Wine Data: Classifying accuracy with $\delta = 0.7$

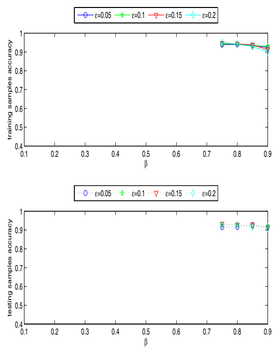


Fig. 10.64 Wine Data: Classifying accuracy with $\delta = 0.75$

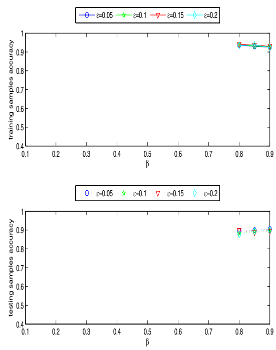


Fig. 10.65 Wine Data: Classifying accuracy with $\delta = 0.8$

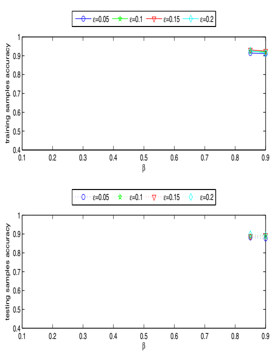


Fig. 10.66 Wine Data: Classifying accuracy with $\delta = 0.85$

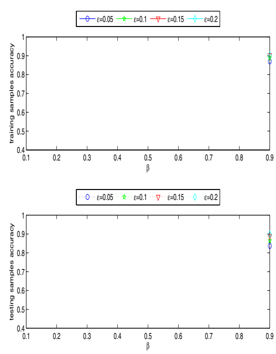


Fig. 10.67 Wine Data: Classifying accuracy with $\delta = 0.9$

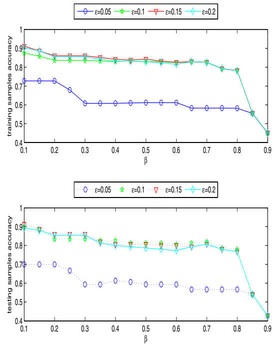


Fig. 10.68 Iris Data: Classifying accuracy with $\delta = 0.1$

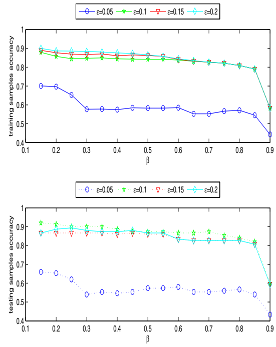


Fig. 10.69 Iris Data: Classifying accuracy with $\delta = 0.15$

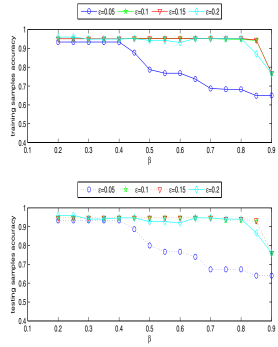


Fig. 10.70 Iris Data: Classifying accuracy with $\delta = 0.2$

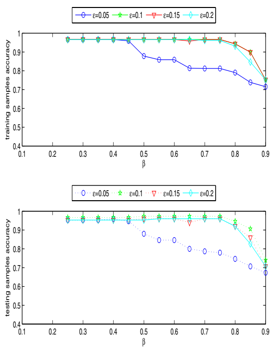


Fig. 10.71 Iris Data: Classifying accuracy with $\delta = 0.25$

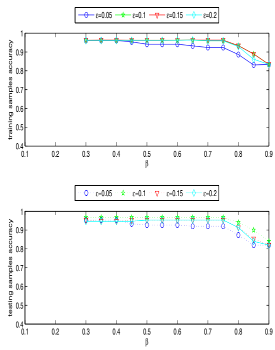


Fig. 10.72 Iris Data: Classifying accuracy with $\delta = 0.3$

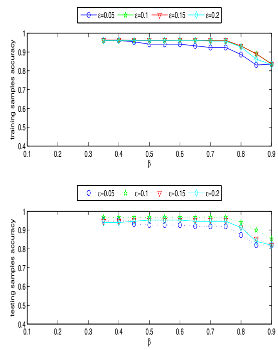


Fig. 10.73 Iris Data: Classifying accuracy with $\delta = 0.35$

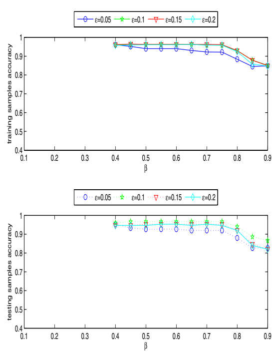


Fig. 10.74 Iris Data: Classifying accuracy with $\delta = 0.4$

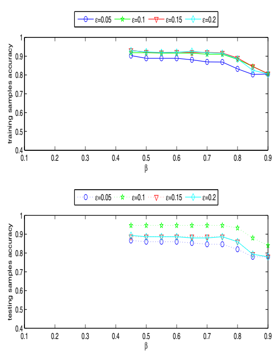


Fig. 10.75 Iris Data: Classifying accuracy with $\delta = 0.45$

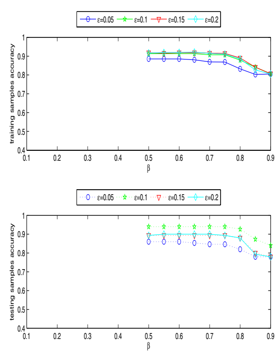


Fig. 10.76 Iris Data: Classifying accuracy with $\delta = 0.5$

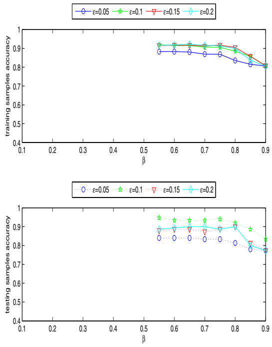


Fig. 10.77 Iris Data: Classifying accuracy with $\delta = 0.55$

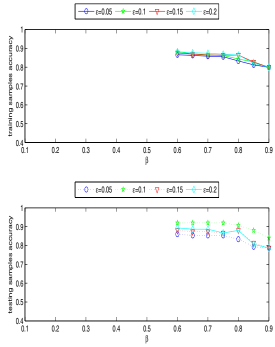


Fig. 10.78 Iris Data: Classifying accuracy with $\delta = 0.6$

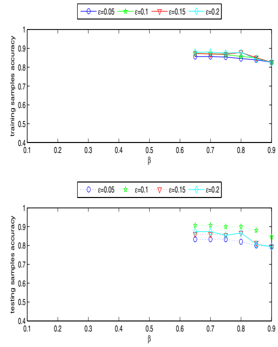


Fig. 10.79 Iris Data: Classifying accuracy with $\delta = 0.65$

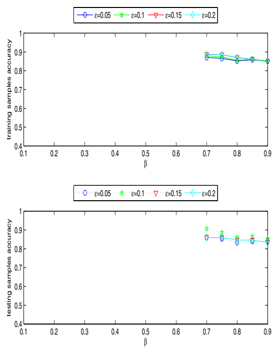


Fig. 10.80 Iris Data: Classifying accuracy with $\delta = 0.7$

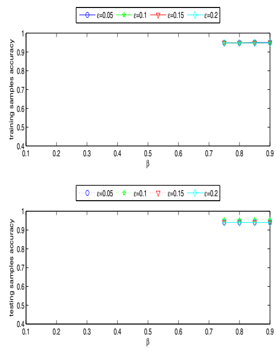


Fig. 10.81 Iris Data: Classifying accuracy with $\delta = 0.75$

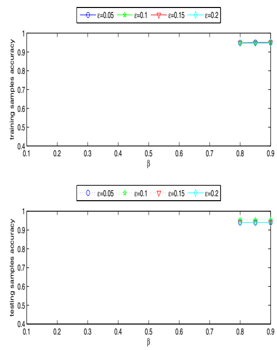


Fig. 10.82 Iris Data: Classifying accuracy with $\delta = 0.8$

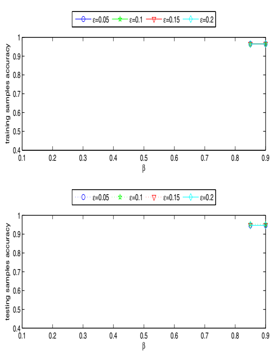


Fig. 10.83 Iris Data: Classifying accuracy with $\delta = 0.85$

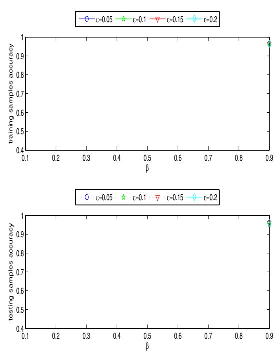


Fig. 10.84 Iris Data: Classifying accuracy with $\delta = 0.9$

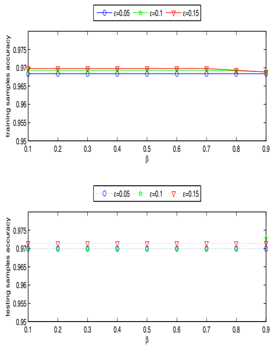


Fig. 10.85 Breast Data: Classifying accuracy with $\delta = 0.1$

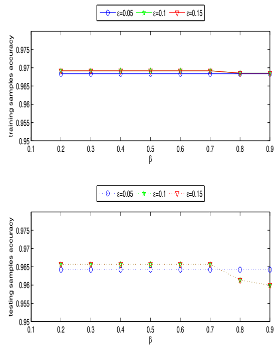


Fig. 10.86 Breast Data: Classifying accuracy with $\delta = 0.2$

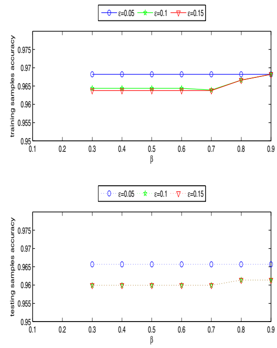


Fig. 10.87 Breast Data: Classifying accuracy with $\delta = 0.3$

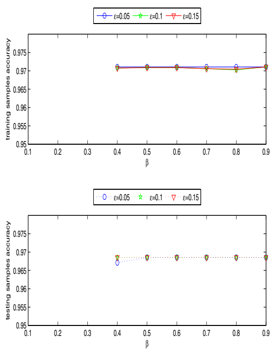


Fig. 10.88 Breast Data: Classifying accuracy with $\delta = 0.4$

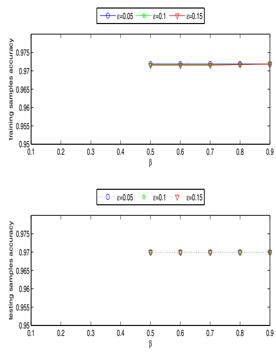


Fig. 10.89 Breast Data: Classifying accuracy with $\delta = 0.5$

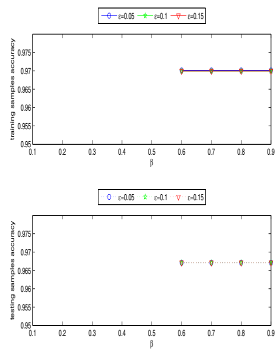


Fig. 10.90 Breast Data: Classifying accuracy with $\delta = 0.6$

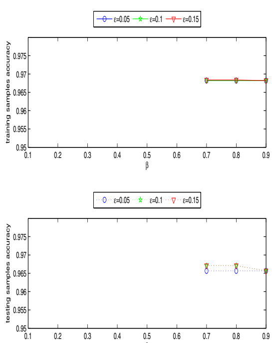


Fig. 10.91 Breast Data: Classifying accuracy with $\delta = 0.7$

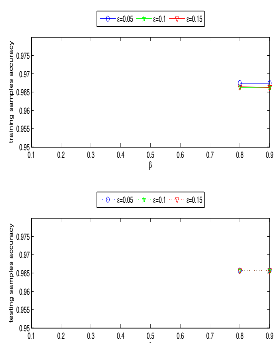


Fig. 10.92 Breast Data: Classifying accuracy with $\delta = 0.8$

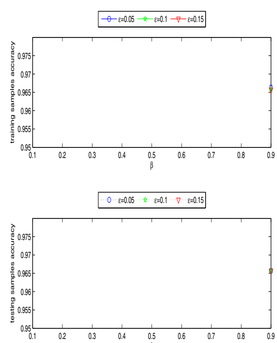


Fig. 10.93 Breast Data: Classifying accuracy with $\delta = 0.9$

References

1. Ait Kbir, M., Benkirane, H., Maalmi, K., Benslimane, R.: Hierarchical fuzzy partition for pattern classification with fuzzy if-then rules. Pattern Recognition Letters 21, 503–509 (2000)
2. Abonyi, J., Roubos, J.A., Szeifert, F.: Data-driven generation of compact accurate and linguistically sound fuzzy classifiers based on a decision-tree initialization. International Journal of Approximate Reasoning 32, 1–21 (2003)
3. Antonie, M.L., Zaiane, O.R., Holte, R.C.: Learning to Use a Learned Model: A Two-Stage Approach to Classification. In: Perner, P. (ed.) ICDM 2006. LNCS, vol. 4065. Springer, Heidelberg (2006)
4. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum, New York (1981)
5. Bellman, R., Kalaba, R., Zadeh, L.A.: Abstraction and pattern classification. J. Math. Anal. Appl. 2, 581–585 (1966)
6. Bajcsy, R., Kovacic, S.: Multi-resolution elastic matching. Computer Vision Graphics Image Processing 46, 1–21 (1989)
7. Chiang, I.J., Hsu, J.Y.J.: Integration of fuzzy classifiers with decision trees. In: Proc. Asian Fuzzy Syst. Symp., pp. 266–271 (1996)
8. Chang, R.L.P., Pavlidis, T.: Fuzzy decision tree algorithms. IEEE Trans. Syst., Man, Cybern. 7, 28–35 (1977)
9. Carse, B., Fogarty, T.C., Munro, A.: Evolving fuzzy rule based controllers using genetic algorithms. Fuzzy Sets System 80, 273–293 (1996)
10. Ciesielski, K.: Set Theory for the Working Mathematician. Cambridge University Press, Cambridge (1997)
11. Cohen, W.: Fast effective rule induction. In: Proc. of ICML, pp. 115–123 (1995)
12. Chang, X.G., Lilly, J.H.: Evolutionary design of a fuzzy classifier from data. IEEE Transactions on Systems, Man, and Cybernatics-part B: Cybernetics 34(4), 1894–1906 (2004)
13. Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets. Internat. J. General Systems 17(2-3), 191–209 (1990)
14. Dubois, D., Prade, H.: Putting fuzzy sets and rough sets together. In: Slowinski, R. (ed.) Intelligent Decision Support, pp. 203–232. Kluwer Academic, Dordrecht (1992)
15. Fernandez Salido, J.M., Murakami, S.: Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations. Fuzzy Sets and Systems 139, 635–660 (2003)
16. Furnkranz, J., Widmer, G.: Incremental reduced error pruning. In: Proc. of ICML, pp. 70–77 (1994)
17. Feng, X.H.: Extraction of Fuzzy Rules from Fuzzy Decision Trees: An Axiomatic Fuzzy Sets, Master Degree Thesis, Dalian Maritime university (2007)
18. Graver, J.E., Watkins, M.E.: Combinatorics with Emphasis on the Theory of Graphs. Springer, New York (1977)
19. Grenander, U.: General Pattern Theory. Oxford University Press, Oxford (1993)
20. Girolami, M.: Mercer kernel-based clustering in feature space. IEEE Transactions on Neural Networks 13, 780–784 (2002)
21. Huang, Z.H., Gedeon, T.D., Nikravesh, M.: Pattern Trees Induction: A New Machine Learning Method. Springer, New York (1974)
22. Hu, B.Q., Wang, S.: A novel approach to fuzzy programming part I: New arithmetic and order relations for interval numbers. Journal of Industrial and Management Optimization 2, 351–372 (2006)

23. Hu, B.Q., Wang, S.: A novel approach to fuzzy programming part II: A class of constrained nonlinear programming problems with interval objective functions. Journal of Industrial and Management Optimization 2, 373–386 (2006)
24. Hayashi, I.: Acquisition of fuzzy rules using fuzzy ID3 with ability of learning. In: Proc. ANZIIS 1996, pp. 187–190 (1996)
25. Hayashi, I., Maeda, T., Bastian, A., Jain, L.C.: Generation of fuzzy decision trees by fuzzy ID3 with adjusting mechanism of and/or operators. In: Proc. Int. Conf. Fuzzy Syst., pp. 681–685 (1998)
26. Halgamuge, S.K., Poechmueller, W., Glesner, M.: An alternative approach for generation of membership functions and fuzzy rules based on radial and cubic basis function networks. Int. J. Approx. Reas. 12, 279–298 (1995)
27. Halmos, P.R.: Measure theory. Springer, New York (1974)
28. Hartley, R.V.: Transmission of information. Bell Syst. Tech. J. 7, 535–563 (1928)
29. Ishibuchi, H., Nozaki, K., Yamamoto, N., Tanaka, H.: Selecting fuzzy if-then rules for classification problems using genetic algorithms. IEEE Trans. Fuzzy Syst. 3, 260–270 (1995)
30. Ichihashi, H., Shirai, T., Nagasaka, K., Miyoshi, T.: Neuro fuzzy ID3: A method of inducing fuzzy decision trees with linear programming for maximizing entropy and algebraic methods. Fuzzy Sets System 81(1), 157–167 (1996)
31. Ishibuchi, H., Nakashima, T., Murata, T.: Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems. IEEE Transactions on Systems, Man, and Cybernatics-part B: Cybernetics 29(5), 601–618 (1999)
32. Jain, A.K., Duin, R.P.W., Mao, J.C.: Statistical pattern recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 4–37 (2000)
33. Janikow, C.Z.: Fuzzy decision trees: Issues and Methods. IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics 28(1), 1–14 (1998)
34. Jain, A.K., Mao, J., Mohiuddin, K.M.: Artificial neural networks. Computer, 628–633 (1996)
35. Keller, J.M., Gray, M., Givens, J.: A fuzzy k-nearest neighbor algorithm. IEEE Trans.Syst., Man, Cybern. 15, 580–585 (1985)
36. Kim, K.H.: Boolean matrix Theory and Applications. Marcel Dekker, New York (1982)
37. Kosko, B.: Fuzzy Engineering. Prentice Hall, New Jersey (1998)
38. Kohonen, T.: Self-Organization and Associative Memory. Springer, Berlin (1989)
39. Kohonen, I.: Self-Organizing Maps. Springer Series in Information Sciences, vol. 30, Berlin (1995)
40. Kosko, B.: Fuzzy entropy and conditioning. Inform. Sci. 40, 165–174 (1986)
41. Liu, X.D.: The fuzzy theory based on AFS algebras and AFS structure. Journal of Mathematical Analysis and Applications 217, 459–478 (1998)
42. Liu, X.D.: The topology on AFS algebra and AFS structure. Journal of Mathematical Analysis and Applications 217, 479–489 (1998)
43. Liu, X.D.: The structure of fuzzy matrices. Journal of Fuzzy Mathematics 2, 311–325 (1994)
44. Liu, X.D.: The fuzzy sets and systems based on AFS structure, EI algebra and EII algebra. Fuzzy Sets and Systems 95, 179–188 (1998)
45. Liu, X.D.: A new fuzzy model of pattern recognition and hitch diagnoses of complex systems. Fuzzy Sets and Systems 104, 289–297 (1999)
46. Liu, X.D., Zhang, Q.L.: AFS Fuzzy logic and its applications to fuzzy information processing. Dongbei Daxue Xuebao 23, 321–323 (2002) (in Chinese)
47. Liu, X.D., Pedrycz, W., Zhang, Q.L.: Axiomatics fuzzy sets logic. In: IEEE International Conference on Fuzzy Systems, vol. 1, pp. 55–60 (2003)

48. Liu, X.D., Zhang, Q.L.: The EI algebra representations of fuzzy concept. Fuzzy Systems and Mathematics 16, 27–35 (2002)
49. Liu, X.D., Zhu, K.J., Huang, H.Z.: The representations of fuzzy concepts based on the fuzzy matrix theory and the AFS theory. In: IEEE International Symposium on Intelligent Control, Texas, USA, October 5–8 (2003)
50. Liu, X.D., Liu, W.Q.: The Framework of Axiomatics Fuzzy Sets Based Fuzzy Classifiers. Journal of Industrial and Mangenment Optimization 4(3), 581–609 (2008)
51. Liu, X.D., Wang, W., Chai, T.Y.: The Fuzzy Clustering Analysis Based on AFS Theory. IEEE Transactions on Systems, Man and Cybernetics Part B 35(5), 1013–1027 (2005)
52. Liu, X.D., Liu, W.Q.: Credit Rating Analysis with AFS Fuzzy Logic. In: Proceedings o f the Second International Conference on Fuzzy Systems and Knowledge Discovery, Changsha, China, August 27-29, 2005, pp. 1198–1204 (2005)
53. Liu, X.D., Chai, T.Y., Wang, W.: Approaches to the Representations and Logic Operations for Fuzzy Concepts in the Framework of Axiomatic Fuzzy Set Theory I, II. Information Sciences 177, 1007–1026, 1027–1045 (2007)
54. Liu, X.D., Pedrycz, W.: The Development of Fuzzy Decision Trees in the Framework of Axiomatic Fuzzy Set Logic. Applied Soft Computing 7, 325–342 (2007)
55. Liu, X.D.: The Development of AFS Theory Under Probability Theory. International Journal of Information and Systems Sciences 3(2), 326–348 (2007)
56. Liu, X.D., Liu, W.Q., Yang, B., Yan, H.: A New Framework of Axiomatic Fuzzy Set Classifiers via Entropy Analysis. Pattern Recognition (submitted)
57. Liu, X.D., Feng, X.H., Pedrycz, W.: Extraction of Fuzzy Rules from Fuzzy Decision Trees: An Axiomatic Fuzzy Sets. Knowledge and Data Engineering (submitted)
58. Liu, B., Hsu, W., Ma, Y.: Integrating classification and association rule mining. In: Proc. of SIGKDD, pp. 80–86 (1998)
59. Li, W.M., Han, J.W., Pei, J.: Accurate and Efficient Classification Based on Multiple Class-Association Rules. In: Proc. of ICDM, pp. 369–376 (2001)
60. Lee, H.M., Chen, C.M., Chen, J.M., Jou, Y.L.: An Efficient Fuzzy Classifier with feature Selection Based on Fuzzy Entropy. IEEE Transactions on Systems, Man, And Cybernetics-Part B: Cybernetics 31(3), 25–40 (2001)
61. Mizumoto, M.: Fuzzy controls under various fuzzy reasoning methods. Information Sciences 45, 129–151 (1988)
62. Maher, P.E., Clair, D.S.: Uncertain reasoning in an ID3 machine learning framework. In: Proc. IEEE ICFS 1993, pp. 7–12 (1993)
63. Michalski, R.S.: Understanding the nature of learning. In: Michalski, R., Carbonell, J., Mitchell, T. (eds.) Machine Learning: An Artificial Intelligence Approach, vol. 2, pp. 3–26. Morgan Kaufmann, San Mateo (1986)
64. Mitra, S., Konwar, K.M., Sankar, K.P.: Fuzzy decision tree, linguistic rules and fuzzy knowledge-based network: generation and evaluation. IEEE Transactions an Systems, Man, and Cybernetics–Part C: Applic. and Reviews 32(4), 328–339 (2002)
65. Mitchell, T.M.: Machine Learning. McGraw-Hill, New York (1997)
66. Mitra, S., Pal, S.K.: Self-organizing neural network as a fuzzy classifier. IEEE Trans. Syst., Man, Cybern. 24, 385–399 (1994)
67. Merz, C.J., Murphy, P.M.: UCI Repository for Machine Learning Data-Bases. Dept. of Information and Computer Science, University of California, Irvine, CA (1996), http://www.ics.uci.edu/~mlearn/MLRepository.html
68. Nozaki, K., Ishibuchi, H., Tanaka, H.: Adaptive fuzzy rule-based classification systems. IEEE Trans. Fuzzy Syst. 4, 238–250 (1996)
69. Pedrycz, W., Sosnowski, Z.A.: The design of decision trees in the framework of granular data and their application to software quality models. Fuzzy Sets and Systems 123, 271–290 (2001)

70. Pedrycz, W., Vukovich, G.: Logic-oriented fuzzy clustering. Pattern Recognition Letters 23, 1515–1527 (2002)
71. Pal, S.K., Mitra, S.: Multilayer perceptron, fuzzy sets, and classification. IEEE Trans. Neural Networks 3, 683–697 (1992)
72. Pedrycz, W., Sosnowski, Z.A.: Designing decision trees with the use of fuzzy granulation. IEEE Trans. Syst., Man, Cybern. part A 30, 151–159 (2000)
73. Perlovsky, L.I.: Conundrum of combinatorial complexity. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 666–670 (1998)
74. Pao, Y.H.: Adaptive Patter-Recognition and Neural Network. Addison-Wesley, Reading (1989)
75. Pedrycz, W., Sosnowski, Z.A.: C-fuzzy decision trees. IEEE Transactions on Systems, Man, cybernetics, Part C: Applications and Reviews 35(4), 498–511 (2005)
76. Pawlak, Z.: Rough set. International Journal of Computer and Information Sciences 11, 341–356 (1982)
77. Pedrycz, W.: Fuzzy neural networks with reference neurons as pattern classifiers. IEEE Trans. Neural Networks 3, 770–775 (1992)
78. Pal, S.K., Chakraborty, B.: Fuzzy set theoretic measure for automatic feature evaluation. IEEE Trans. Syst., Man, Cybern. 16, 754–760 (1986)
79. Quinlan, J.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Francisco (1993)
80. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo (1993)
81. Quinlan, J.R.: Induction on decision trees. Mach. Learn. 1, 81–106 (1986)
82. Quinlan, J.R.: Decision trees as probabilistic classifiers. In: Proc. 4th Int. Workshop Machine Learning, pp. 31–37 (1987)
83. Quinlan, J.R.: Unknown attribute-values in induction. In: Proc. 6th Int. Workshop Machine Learning, pp. 164–168 (1989)
84. Radzikowska, A., Kerre, E.E.: A comparative study of fuzzy rough sets. Fuzzy Sets and Systems 126, 137–155 (2004)
85. Ruspini, E.H.: A new approach to clustering. Information and Control 15, 22–32 (1969)
86. Roubos, J.A., Setnes, M., Abonyi, J.: Learning fuzzy classification rules from labeled data. Information Sciences 150, 77–93 (2003)
87. Roubos, H., Setnes, M.: Compact and transparent fuzzy models and classifiers through iterative complexity reduction. IEEE Transations on Fuzzy Systems 9, 516–524 (2001)
88. Roberts, S.J., Everson, R., Rezek, I.: Maximum certainty data partitioning. Pattern Recognition 33, 833–839 (2000)
89. Ren, Y., Song, M.L., Liu, X.D.: New approaches to the fuzzy clustering via AFS theory. International Journal of Information and Systems Sciences 3(2), 307–325 (2007)
90. Radzikowska, A.M., Kerre, E.E.: A comparative study of fuzzy rough sets. Fuzzy Sets and Systems 126, 137–155 (2002)
91. Renyi, A.: On the measure of entropy and information. In: Proc. Fourth Berkeley Symp. Math. Statistics Probability, vol. 1, pp. 541–561 (1961)
92. Schaffer, C.: Overfitting avoidance as bias. Mach. Learn. 10, 153–178 (1993)
93. Sestino, S., Dillon, T.: Using single-layered neural networks for the extraction of conjunctive rules and hierarchical classifications. J. Appl. Intell. 1, 157–173 (1991)
94. Schapire, R.E.: Theoretical views of boosting. In: Proc. of European Conference on Computational Learning Theory (EuroCOLT), pp. 1–10 (1999)
95. Setnes, M., Roubos, H.: GA-fuzzy modeling and classification: Complexity and Performance. IEEE Transactions on Fuzzy Systems 8(5), 509–522 (2000)

96. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423 (1948)
97. Tsang, E.C.C., Wang, X.Z., Yeung, D.S.: Improving learning accuracy of fuzzy decision trees by hybrid neural networks. IEEE Trans. Fuzzy Syst. 8, 601–614 (2000)
98. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its application to modeling and control. IEEE Transactions on Systems, Man, and Cybernetics-part B: Cybernetics 15, 116–132 (1985)
99. Uebele, V., Abe, S., Lan, M.S.: A neural-network-based fuzzy classifier. IEEE Trans. Syst., Man, Cybern. 25, 353–361 (1995)
100. Wang, G.J.: Theory of topological molecular lattices. Fuzzy Sets and Systems 47, 351–376 (1992)
101. Wang, X., Chen, B., Qian, G., Ye, F.: On the optimization of fuzzy decision trees. Fuzzy Sets System 112, 117–125 (2000)
102. Wang, X.Z., Yeung, D.S., Tsang, E.C.C.: A comparative study on heuristic algorithms for generating fuzzy decision trees. IEEE Transactions an Systems, Man, and Cybernetics–Part B: Cybernetics 31(2), 215–226 (2001)
103. Wang, L.D., Yen, J.: Extracting fuzzy rules for system modeling using a hybride of genetic algorithms and Kalman filter. Fuzzy Sets and Systems 101, 353–362 (1999)
104. Wiener, N.: Cybernetics. Wiley, New York (1961)
105. Xizhao, W., Hong, J.: On the handling of fuzziness for continuous valued attributes in decision tree generation. Fuzzy Sets System 99, 283–290 (1998)
106. Yang, B.: A New Framework of Axiomatic Fuzzy Set Classifiers via Entropy Analysis, Master Degree Thesis, Dalian Maritime university (2008)
107. Yao, Y.Y.: Semantics of fuzzy sets in rough set theory. In: Peters, J.F., Skowron, A., Dubois, D., Grzymała-Busse, J.W., Inuiguchi, M., Polkowski, L. (eds.) Transactions on Rough Sets II. LNCS, vol. 3135, pp. 297–318. Springer, Heidelberg (2004)
108. Yang, M.S.: A survey of fuzzy clustering. Math. Comput. Modelling 18, 1–16 (1993)
109. Yang, M.S., Ko, C.H.: On cluster-wise fuzzy regression analysis. IEEE Transactions Systems Man Cyberenet, Part B 27, 1–13 (1997)
110. Yuan, Y., Shaw, M.J.: Induction of fuzzy decision trees. Fuzzy Sets Syst. 69, 125–139 (1995)
111. Yeung, D.S., Wang, X.Z., Tsang, E.C.C.: Learning weighted fuzzy rules from examples with mixed attributes by fuzzy decision trees. In: Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics, Tokyo, Japan, October 12-15, 1999, pp. 349–354 (1999)
112. Yin, X., Han, J.: CPAR: Classification based on predictive association rules. In: Proc. of SDM (2003)
113. Zadeh, L.A.: Fuzzy logic and approximate reasoning. Synthese 30, 407–428 (1975)
114. Zadeh, L.A.: A theory of approximate reasoning. In: Hayes, Michie, Mikulich (eds.) Machine Intelligence, vol. 9, pp. 149–194. Wiley, Chichester (1979)
115. Zadeh, L.A.: The role of fuzzy logic in the management of uncertainty in expert systems. Fuzzy Sets Syst. 11, 199–227 (1983)
116. Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)
117. Zadeh, L.A.: Fuzzy logic=computing with word. IEEE Transactions on Fuzzy Systems 4, 103–111 (1996)
118. Zhang, Y.J., Liang, D.Q., Tong, S.C.: On AFS Algebra Part I, II. Information Sciences 167, 263–286, 287-303 (2004)

Bibliography

Mathematical Fundamentals

1. Lesniewski, S.: Foundations of the general theory of sets. Polish Scientific Circle, Moscow (1916)
2. Hartley, R.V.: Transmission of information. Bell Syst. Tech. J. 7, 535–563 (1928)
3. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423 (1948)
4. Kelley, J.L.: General Topology. Springer, New York (1955)
5. Renyi, A.: On the measure of entropy and information. In: Proc. Fourth Berkeley Symp. Math. Statistics Probability, Berkeley, CA, vol. 1, pp. 541–561 (1961)
6. Baum, J.D.: Elements of Point Set Topology. Prentice-Hall, Englewood Cliffs (1964)
7. Bellman, R., Kalaba, R., Zadeh, L.A.: Abstraction and pattern classification. J. Math. Anal. Appl. 2, 581–585 (1966)
8. Birkhoff, G.: Lattice Theory, 3rd edn. American Mathematics Society, Providence (1967)
9. John, L.K.: General Topology. Springer, New York (1955)
10. John, D.B.: Elements of Point Set Topology. Prentice-Hall, Inc., Englewood Cliffs (1964)
11. Halmos, P.R.: Measure Theory. Springer, New York (1974)
12. Huang, Z.H., Gedeon, T.D., Nikravesh, M.: Pattern Trees Induction: A New Machine Learning Method. Springer, New York (1974)
13. Halmos, P.R.: Measure theory. Springer, New York (1974)
14. Graver, J.E., Watkins, M.E.: Combinatorics with Emphasis on the Theory of Graphs. Springer, New York (1977)
15. Ciesielski, K.: Set Theory for the Working Mathematician. Cambridge University Press, Cambridge (1997)
16. Pu, B.M., Liu, Y.M.: Fuzzy Topology I, Neighborhood structure of Fuzzy Point and Moore-Smith Convergence. Journal of Mathematical Analysis and Applications 76, 571–599 (1980)
17. Kim, K.H.: Boolean Matrix Theory and Applications. Marcel Dekker, New York (1982)

18. Kohonen, T.: Self-Organized Formation of Topologically Correct Feature Maps. Biological Cybernetics 43(1), 59–69 (1982)
19. Seber, G.A.F.: Multivariate Observations. Wiley, New York (1984)
20. Jacobson, N.: Basic Algebra I. W H Freeman and Company, New York (1985)
21. Quinlan, J.R.: Induction on decision trees. Mach. Learn. 1, 81–106 (1986)
22. Michalski, R.S.: Understanding the nature of learning. In: Michalski, R., Carbonell, J., Mitchell, T. (eds.) Machine Learning: An Artificial Intelligence Approach, vol. 2, pp. 3–26. Morgan Kaufmann, San Francisco (1986)
23. Lindley, D.V.: The Probability Approach to the Treatment of Uncertainty in Artifical Intelligence and Expert Systems. Statistical Science 2, 17–24 (1987)
24. Bajcsy, R., Kovacic, S.: Multi-resolution elastic matching. Computer Vision Graphics Image Processing 46, 1–21 (1989)
25. Kohonen, T.: Self-Organization and Associative Memory. Springer, Berlin (1989)
26. Pao, Y.H.: Adaptive Patter-Recognition and Neural Network. Addison-Wesley, Reading (1989)
27. Martinetz, T.E., Schulten, K.J.: Neural-Gas network for vector quantization and its application to time-series prediction. IEEE Trans. Neural Networks 4(4), 558–569 (1993)
28. Blum, A., Langley, P.: Selection of Relevant Features and Examples in Machine Learning. Artificial Intelligence 97(1-2), 245–271 (1997)
29. Liu, B., Hsu, W., Ma, Y.: Integrating classification and association rule mining. In: Proc. of SIGKDD, pp. 80–86 (1998)
30. Gath, I., Geva, A.B.: Unsupervised optimal fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell. 11(7), 773–781 (1989)
31. Booker, J.M., Mayer, M.A.: Common problem in the elicitation and analysis of expert opinion affecting probabilistic safety assessments. In: CSNI Workshop on PSA Application and Limitation, Santa Fe, NM, September 4-6, pp. 353–368 (1990)
32. Quinlan, J.R.: Decision trees and decision making. IEEE Transactions on Systems, Man and Cybernetics Part C 20, 339–346 (1990)
33. Gowda, K.C., Diday, E.: Symbolic clustering using a new dissimilarity measure. Pattern Recogn. 24(6), 567–578 (1991)
34. Sestino, S., Dillon, T.: Using single-layered neural networks for the extraction of conjunctive rules and hierarchical classifications. J. Appl. Intell. 1, 157–173 (1991)
35. Wang, G.J.: Theory of Topological Molecular Lattices. Fuzzy Sets and Systems 47, 351–376 (1992)
36. Pal, N., Bezdek, J., Tsao, E.: Generalized clustering networks and Kohonen's self-organizing scheme. IEEE Trans. Neural Netw. 4(4), 549–557 (1993)
37. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo (1993)
38. Schaffer, C.: Overfitting avoidance as bias. Mach. Learn. 10, 153–178 (1993)
39. Kohonen, I.: Self-Organizing Maps, Berlin. Springer Series in Information Sciences, vol. 30, Berlin (1995)
40. Uebele, V., Abe, S., Lan, M.S.: A neural-network-based fuzzy classifier. IEEE Trans. Syst., Man, Cybern. 25, 353–361 (1995)

41. Simonoff, J.S.: Smoothing Methods in Statistics. Springer, New York (1996)
42. Koller, D., Sahami, M.: Toward Optimal Feature Selection. In: Proc. 13th Int'l Conf. Machine Learning, pp. 284–292 (1996)
43. Jain, A.K., Mao, J., Mohiuddin, K.M.: Artificial neural networks. Computer, 628–633 (1996)
44. Blum, A., Langley, P.: Selection of Relevant Features and Examples in Machine Learning. Artificial Intelligence 97(1-2), 245–271 (1997)
45. Jain, A., Zongker, D.: Feature Selection: Evaluation, Application, and Small Sample Performance. IEEE Trans. Pattern Analysis and Machine Intelligence 19(2), 153–157 (1997)
46. Kohavi, R., John, G.: Wrappers for Feature Subset Selection. Artificial Intelligence 97(1-2), 273–324 (1997)
47. Eltoft, T., de Figueiredo, R.: A new neural network for cluster-detection-and-labeling. IEEE Trans. Neural Netw. 9(5), 1021–1035 (1998)
48. Kosko, B.: Fuzzy Engineering. Prentice Hall, New Jersey (1998)
49. Perlovsky, L.I.: Conundrum of combinatorial complexity. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 666–670 (1998)
50. Gabrys, B., Bargiela, A.: General fuzzy min-max neural network for clustering and classification. IEEE Trans. Neural Netw. 11(3), 769–783 (2000)
51. Kirby, M.: Data Analysis: An empirical approach to dimensionality reduction and the study of Pattern. Wiley, New York (2000)
52. Mollineda, R., Vidal, E.: A relative approach to hierarchical clustering. In: Torres, M., Sanfeliu, A. (eds.) Pattern Recognition and Applications. Frontiers in Artificial Intelligence and Applications, vol. 56, pp. 19–28. IOS Press, Amsterdam (2000)
53. Roberts, S.J., Everson, R., Rezek, I.: Maximum certainty data partitioning. Pattern Recognition 33, 833–839 (2000)
54. Jain, A.K., Duin, R.P.W., Mao, J.C.: Statistical pattern recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 4–37 (2000)
55. Ben-Hur, A., Horn, D., Siegelmann, H., Vapnik, V.: Support vector clustering. J. Mach. Learn. Res. 2, 125–137 (2001)
56. Hathaway, R., Bezdek, J.: Fuzzy c-means clustering of incomplete data. IEEE Trans. Syst., Man, Cybern. 31(5), 735–744 (2001)
57. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: Analysis and an algorithm. Advances in Neural Information Processing Systems 14, 849–856 (2001)
58. Girolami, M.: Mercer kernel based clustering in feature space. IEEE Trans. Neural Netw. 13(3), 780–784 (2002)
59. Wu, K.L., Yang, M.S.: Alternative c-means clustering algorithms. Pattern Recognition 35, 2267–2278 (2002)
60. Girolami, M.: Mercer kernel-based clustering in feature space. IEEE Transactions on Neural Networks 13, 780–784 (2002)
61. George, C., Roger, L.B.: Statistical Inference. Wadsworth Group (2002)
62. Mitchell, T.M.: Machine Learning. China Machine Press (2003)
63. Singpurwalla, N.D., Booker, J.M.: Membership Functions and Probability Measures of Fuzzy Sets. Journal of the American Statistical Association 99(467), 867–877 (2004)

64. Fortuna, J., Capson, D.: Improved support vector classification using PCA and ICA feature space modification. Pattern Recognition 37, 1117–1129 (2004)
65. Law, M., Figueiredo, M., Jain, A.K.: Simultaneous Feature Selection and Clustering Using Mixture Models. IEEE Trans. Pattern Anal. Mach. Intell. 26(9), 124–138 (2004)
66. Wu, S.T., Chow, T.W.S.: Clustering of the Self-organizing Map Using a Clustering Validity Index Based on Inter-cluster and Intra-cluster Density. Pattern Recognition 37, 175–188 (2004)
67. Camastra, F., Verri, A.: A novel kernel method for clustering. IEEE Trans. Pattern Anal. Mach. Intell. 27(5), 801–805 (2005)
68. Grzegorzewski, P.: The coefficient of concordance for vague data. Computational Statistics and Data Analysis 51, 314–322 (2006)
69. Chen, Y.L., Hu, H.L.: An overlapping cluster algorithm to provide non-exhaustive clustering. European Journal of Operational Research 173, 762–780 (2006)
70. Cord, A., Ambroise, C., Cocquerez, J.: Feature selection in robust clustering based on Laplace mixture. Pattern Recognition Letters 27, 627–635 (2006)
71. Jenssen, R., Erdogmus, D.: Information cut for clustering using a gradient descent approach. Pattern Recognition 40, 796–806 (2007)
72. Yang, X., Song, Q., Wu, Y.: A robust deterministic annealing algorithm for data clustering. Data & Knowledge Engineering 62, 84–100 (2007)
73. Li, Y., Dong, M., Hua, J.: Localized feature selection for clustering. Pattern Recognition Letters 29, 10–18 (2008)

AFS Theory

1. Liu, X.D.: The Structure of Fuzzy Matrices. International Journal of Fuzzy Mathematics 2, 311–325 (1994)
2. Liu, X.D.: The Solution of Fuzzy Equations. International Journal of Fuzzy Mathematics 2, 326–333 (1994)
3. Liu, X.D.: A New Mathematical Axiomatic System of Fuzzy Sets and Systems. International Journal of Fuzzy Mathematics 3, 559–560 (1995)
4. Liu, X.D.: Two Algebra Structures of AFS structure. International Journal of Fuzzy Mathematics 3, 561–562 (1995)
5. Liu, X.D.: The Fuzzy Theory Based on AFS Algebras and AFS Structure. Journal of Mathematical Analysis and Applications 217, 459–478 (1998)
6. Liu, X.D.: The Topology on AFS Algebra and AFS Structure. Journal of Mathematical Analysis and Applications 217, 479–489 (1998)
7. Liu, X.D.: The Fuzzy Sets and Systems Based on AFS Structure, EI Algebra and EII algebra. Fuzzy Sets and Systems 95, 179–188 (1998)
8. Liu, X.D.: A new fuzzy model of pattern recognition and hitch diagnoses of complex systems. Fuzzy Sets and Systems 104, 289–297 (1999)
9. Liu, X.D., Zhang, Q.L.: The EI algebra Representations of Fuzzy Concept. Fuzzy Systems and Mathematics 16(2), 27–35 (2002) (in Chinese)
10. Liu, X.D., Zhang, Q.L., Wang, Y.: AFS Fuzzy Logic System and Its Applications to Fuzzy Information Processing. J. of Northeastern University 23(4), 321–324 (2002) (in Chinese)
11. Liu, X.D., Pedrycz, W., Zhang, Q.L.: Axiomatics Fuzzy sets logic. In: The Proceedings of IEEE International Conference on Fuzzy Systems, vol. 1, pp. 55–60 (2003)

12. Liu, X.D., Zhu, K.J., Huang, H.Z.: The Representations of Fuzzy Concepts Based on the Fuzzy Matrix Theory and the AFS Theory. In: IEEE International Symposium on Intelligent Control, Houston, Texas, USA, October 5-8, 2003, pp. 1006–1011 (2003)
13. Zhang, Y.J., Liang, D.Q., Tong, S.C.: On AFS Algebra Part I. Information Sciences 167, 263–286 (2004)
14. Zhang, Y.J., Liang, D.Q., Tong, S.C.: On AFS Algebra Part II. Information Sciences 167, 287–303 (2004)
15. Wang, X., Liu, X.D.: The Base of finite EI Algebra. In: 3rd IEEE International Conference on Machine Learning and Cybernetics, Shanghai, China, pp. 2005–2009 (2004)
16. Ding, R., Liu, X.D., Chen, Y.: The fuzzy clustering algorithm based on AFS topology. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS, vol. 4223, pp. 89–98. Springer, Heidelberg (2006)
17. Zhang, L.S., Liu, X.D.: Concept lattice and AFS algebra. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS, vol. 4223, pp. 290–299. Springer, Heidelberg (2006)
18. Zhang, L.S., Su, Y., Liu, X.D.: AFS Structures and Concept Lattices. In: Sixth World Congress on Intelligent Control and Automation, Dalian, China, June 21-23, vol. I3, pp. 2541–2545 (2006)
19. Zhang, L.S., Liu, X.D.: Concept Lattice and its Topological Structure. In: Sixth World Congress on Intelligent Control and Automation, Dalian, China, June 21-23, vol. 4, pp. 2633–2636 (2006)
20. Liu, X.D., Chai, T.Y., Wang, W., Liu, W.Q.: Approaches to the Representations and Logic Operations for Fuzzy Concepts in the Framework of Axiomatic Fuzzy Set Theory I. Information Sciences 177, 1007–1026 (2007)
21. Liu, X.D., Wang, W., Chai, T.Y., Liu, W.Q.: Approaches to the Representations and Logic Operations for Fuzzy Concepts in the Framework of Axiomatic Fuzzy Set Theory II. Information Sciences 177, 1027–1045 (2007)
22. Liu, X.D.: The Development of AFS Theory Under Probability Theory. International Journal of Information and Systems Sciences 3(2), 326–348 (2007)
23. Liu, X.D., Zhang, L.S., Zhou, J., Zhou, K.J., Zhang, Q.L.: The Structures of EI Algebras Generated by Information Attributes. International Journal of Intelligent Systems Technologies and Applications 3(3/4), 341–355 (2007)
24. Wang, L.D., Liu, X.D.: Concept Analysis via Rough Set and AFS Algebra. Information Sciences 178, 4125–4137 (2008)

Application of AFS Theory

1. Liu, X.D.: A New Fuzzy Model of Pattern Recognition and Hitch Diagnoses of Complex Systems. Fuzzy Sets and Systems 104, 289–297 (1999)
2. Liu, X.D., Zhang, Q.L.: The Fuzzy Cognitive Maps Based on AFS Fuzzy Logic. Dynamics of Continuous, Discrete and Impulsive Systems, Series A: Mathematical Analysis 11(5-6), 787–796 (2004)
3. Liu, X.D., Wang, Y.B., Liu, W.: The Decomposition Algorithm of AFS Structure and Its Applications to Failure Diagnosis. In: Proc. IEEE Conference on Robotics, Automation and Mechitronics, Singapore, December, 1-3, 2004, pp. 729–733 (2004)

4. Liu, X.D., Zhou, X.Y., Wang, X.: The Fuzzy Inference Rule Extraction and Attribute Reduction Based on AFS Theory and Closeness degrees. In: IEEE International Conference on Machine Learning and Cybernetics, Shanghai, China, pp. 2005–2009 (2004)
5. Liu, X.D., Wang, W., Chai, T.Y.: The Fuzzy Clustering Analysis Based on AFS Theory. IEEE Transactions on Systems, Man and Cybernetics Part B 35(5), 1013–1027 (2005)
6. Liu, X.D., Liu, W.Q.: Credit rating analysis with AFS fuzzy logic. In: Wang, L., Chen, K., S. Ong, Y. (eds.) ICNC 2005. LNCS, vol. 3612, pp. 1198–1204. Springer, Heidelberg (2005)
7. Liu, X.D.: AFS Theory and Its Applications. IEEE Systems, Man and Cybernetics, Society eNewsletter 13 (December 2005), http://www.ieeesmc.org/Newsletter/Dec2005/R8xdliu.php
8. Liu, X.D., Zhang, Q.L., Tong, S.C.: Fuzzy Identification of Systems Based on AFS Fuzzy Logic. Dynamics of Continuous, Discrete And Impulsive Systems, Series B 12b, 90–102 (2005)
9. Liu, X.D., Chai, T.Y., Wang, W.: AFS Fuzzy Logic Systems and Its Applications to Model and Control. International Journal of Information and Systems Sciences 2(3), 285–305 (2006)
10. Ding, R., Liu, X.D., Chen, Y.: The fuzzy clustering algorithm based on AFS topology. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS, vol. 4223, pp. 89–98. Springer, Heidelberg (2006)
11. Zhang, L., Liu, X.: Concept lattice and AFS algebra. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS, vol. 4223, pp. 290–299. Springer, Heidelberg (2006)
12. Wang, X.C., Ren, Y., Liu, X.D.: Fuzzy Clustering Approaches Based On AFS Fuzzy Logic II. In: Sixth World Congress on Intelligent Control and Automation, Dalian, China, June 21-23, vol. 5, pp. 4199–4203 (2006)
13. Ren, Y., Wang, X.C., Liu, X.D.: Fuzzy Clustering Approaches Based On AFS Fuzzy Logic I. In: Sixth World Congress on Intelligent Control and Automation, Dalian, China, June 21-23, vol. 5, pp. 4244–4248 (2006)
14. Ren, Y.: Feature Selection, Concept Categorization and Characteristic Description Based on AFS Theory, Master Degree Thesis, Dalian Maritime university (2006)
15. Wang, X.C.: The Fuzzy Clustering via Axiomatic Fuzzy Set Theory, Master Degree Thesis, Dalian Maritime university (2006)
16. Liu, X.D., Chai, T.Y., Wang, W.: Approaches to the Representations and Logic Operations for Fuzzy Concepts in the Framework of Axiomatic Fuzzy Set Theory II. Information Sciences 177, 1027–1045 (2007)
17. Liu, X.D.: The Development of AFS Theory Under Probability Theory. International Journal of Information and Systems Sciences 3(2), 326–348 (2007)
18. Liu, X.D., Pedrycz, W.: The Development of Fuzzy Decision Trees in the Framework of Axiomatic Fuzzy Set Logic. Applied Soft Computing 7, 325–342 (2007)
19. Ren, Y., Song, M.L., Liu, X.D.: New Approaches to the Fuzzy Clustering via AFS Theory. International Journal of Information and Systems Sciences 3(2), 307–325 (2007)
20. Feng, X.H., Xu, X.L., Liu, X.D.: Fuzzy Rule Extraction From Fuzzy Decision Tree. Advances in Information and Systems Sciences 2, 168–178 (2007)
21. Feng, X.H.: Extraction of Fuzzy Rules from Fuzzy Decision Trees: An Axiomatic Fuzzy Sets, Master Degree Thesis, Dalian Maritime university (2007)

22. Wang, L.D., Liu, X.D.: Concept analysis via rough set and AFS algebra. Information Sciences 178, 4125–4137 (2008)
23. Liu, X.D., Liu, W.Q.: The Framework of Axiomatics Fuzzy Sets Based Fuzzy Classifiers. Journal of Industrial and Mangenment Optimization 4(3), 581–609 (2008)
24. Yang, B.: A New Framework of Axiomatic Fuzzy Set Classifiers via Entropy Analysis, Master Degree Thesis, Dalian Maritime university (2008)
25. Liu, X.D., Zhang, Y.L., Ren, Y.: Dynamic System Description and Control Based on Data. International Journal of Information and Systems Sciences 4(3), 335–347 (2008)
26. Liu, X.D., Pedrycz, W., Chai, T.Y., Song, M.L.: The Development of Fuzzy Rough Sets with the Use of Structures and Algebras of Axiomatic Fuzzy Sets. IEEE Transactions on Knowledge and Data Engineering 21(3), 443–462 (2009)
27. Xu, X.L., Liu, X.D., Cheng, Y.: Applications of Axiomatic Fuzzy Set Clustering Method on Management Strategic. European Journal of Operational Research (accepted) (2008) doi: 10.1016 / j.ejor. 2008.08. 010

Fuzzy Sets: Fundamentals and Application

1. Zadeh, L.A.: Fuzzy Sets. Information and Control 8, 338–353 (1965)
2. Zadeh, L.A.: Probability Measures of Fuzzy Events. Journal of Mathematical Analysis and Applications 23, 421–427 (1968)
3. Ruspini, E.H.: A new approach to clustering. Information and Control 15, 22–32 (1969)
4. Zadeh, L.A.: Similarity relations and fuzzy orderings. Inform. Sci. 3, 177–200 (1971)
5. Bezdek, J.C.: Fuzzy mathematics in pattern classification, Ph.D. Dissertation, Appl. Math. Cornell Univ., Ithaca, NY (1973)
6. Dunn, J.C.: A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J. Cybernet 3(3), 32–57 (1974)
7. Chang, R.L.P., Pavlidis, T.: Fuzzy decision tree algorithms. IEEE Trans. Syst., Man, Cybern. 7, 28–35 (1977)
8. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1, 3–28 (1978)
9. Zadeh, L.A.: Fuzzy logic and approximate reasoning. Synthese 30, 407–428 (1975)
10. Zadeh, L.A.: Fuzzy Sets as a Basis for a Theory of Possibility. Fuzzy Sets and Systems 1, 3–28 (1978)
11. Zadeh, L.A.: A Theory of Approximate Reasoning. In: Hayes, J.E., Michie, D., Mikulich, L.I. (eds.) Machine Intelligence, vol. 9, pp. 149–194. Elsevier, New York (1979)
12. Dubois, D., Prade, H.: Fuzzy Sets and Systems: Theory and Applications. Academic Press, New York (1980)
13. Bezdek, J.C.: A convergence theorem for the fuzzy ISODATA clustering algorithms. IEEE Trans. Pattern Anal. Machine Intell. PAMI 2(1), 1–8 (1980)
14. Bezdek, J.C.: A convergence theorem for the fuzzy ISODATA clustering algorithms. IEEE Trans. Pattern Anal. Machine Intell. 2(1), 1–8 (1980)
15. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum, New York (1981)
16. Zadeh, L.A.: The role of fuzzy logic in the management of uncertainty in expert systems. Fuzzy Sets Syst. 11, 199–227 (1983)

17. Keller, J.M., Gray, M.R., Givens Jr., J.A.: A fuzzy K-nearest neighbors algorithm. IEEE Trans. Systems Man, Cybernet. SMC 15(4), 580–585 (1985)
18. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its application to modeling and control. IEEE Transactions on Systems, Man, and Cybernetics-part B: Cybernetics 15, 116–132 (1985)
19. Kosko, B.: Fuzzy entropy and conditioning. Inform. Sci. 40, 165–174 (1986)
20. Pal, S.K., Chakraborty, B.: Fuzzy set theoretic measure for automatic feature evaluation. IEEE Trans. Syst., Man, Cybern. 16, 754–760 (1986)
21. Zadeh, L.A.: Fuzzy Logic. IEEE Trans. Comput. 35, 83–93 (1988)
22. Mizumoto, M.: Fuzzy controls under various fuzzy reasoning methods. Information Sciences 45, 129–151 (1988)
23. Dubois, D., Prade, H.: Possibility Theory. Plenum Press, New York (1988)
24. Turksen, I.B.: Measurement of Membership Functions and Their Acquisition. Fuzzy Sets and Systems 40, 5–38 (1991)
25. Liang, G.S., Wang, M.J.: A fuzzy multi-criteria decision-making method for facility site selection. International Journal of Product Research 29(11), 2313–2330 (1991)
26. Wang, G.J.: Theory of Topological Molecular Lattices. Fuzzy Sets and Systems 47, 351–376 (1992)
27. Dubois, D., Prade, H.: Putting rough sets and fuzzy sets together. In: Slowinski, R. (ed.) Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory, pp. 203–222. Kluwer, Dordrecht (1992)
28. Nakamura, A., Gao, J.M.: A logic for fuzzy data analysis. Fuzzy Sets and Systems 39, 127–132 (1992)
29. Pal, S.K., Mitra, S.: Multi-layer perception, fuzzy sets, and classification. IEEE Trans. Neural Networks 3, 683–697 (1992)
30. Pedrycz, W.: Fuzzy neural networks with reference neurons as pattern classifiers. IEEE Trans. Neural Networks 3, 770–775 (1992)
31. Liou, T.S., Wang, M.J.J.: Subjective assessment of mental workload-a fuzzy linguistic multi-criteria approach. Fuzzy Sets and Systems 62, 155–165 (1994)
32. Mitra, S., Pal, S.K.: Self-organizing neural network as a fuzzy classifier. IEEE Trans. Syst., Man, Cybern. 24, 385–399 (1994)
33. Mineau, G., Godin, R.: Automatic structuring of knowledge bases by conceptual clustering. IEEE Transactions on Knowledge and Data Engineering 7, 824–829 (1995)
34. Zadeh, L.A.: Discussion: Probability Theory and Fuzzy Logic Are Complementary Rather Than Competitive. Technometrics 37, 271–276 (1995)
35. Halgamuge, S.K., Poechmueller, W., Glesner, M.: An alternative approach for generation of membership functions and fuzzy rules based on radial and cubic basis function networks. Int. J. Approx. Reas. 12, 279–298 (1995)
36. Ishibuchi, H., Nozaki, K., Yamamoto, N., Tanaka, H.: Selecting fuzzy if-then rules for classification problems using genetic algorithms. IEEE Trans. Fuzzy Syst. 3, 260–270 (1995)
37. Yuan, Y., Shaw, M.J.: Induction of fuzzy decision trees. Fuzzy Sets Syst. 69, 125–139 (1995)
38. Hayashi, I.: Acquisition of fuzzy rules using fuzzy ID3 with ability of learning. In: Proc. ANZIIS 1996, pp. 187–190 (1996)
39. Zadeh, L.A.: Fuzzy logic=computing with word. IEEE Transactions on Fuzzy Systems 4, 103–111 (1996)

40. Nozaki, K., Ishibuchi, H., Tanaka, H.: Adaptive fuzzy rule-based classification systems. IEEE Trans. Fuzzy Syst. 4, 238–250 (1996)
41. Ichihashi, H., Shirai, T., Nagasaka, K., Miyoshi, T.: Neuro fuzzy ID3: A method of inducing fuzzy decision trees with linear programming for maximizing entropy and algebraic methods. Fuzzy Sets System 81(1), 157–167 (1996)
42. Carse, B., Fogarty, T.C., Munro, A.: Evolving fuzzy rule based controllers using genetic algorithms. Fuzzy Sets System 80, 273–293 (1996)
43. Zadeh, L.A.: Toward a Theory of Fuzzy Information Granulation and its Centrality in Human Reasoning and Fuzzy Logic. Fuzzy Sets and Systems 90, 111–127 (1997)
44. Bodjanova, S.: Approximation of fuzzy concepts in decision making. Fuzzy Sets and Systems 85, 23–29 (1997)
45. Dubois, D., Prade, H.: The three semantics of fuzzy sets. Fuzzy Sets and Systems 90, 141–150 (1997)
46. Yang, M.S., Ko, C.H.: On cluster-wise fuzzy regression analysis. IEEE Transactions Systems Man Cyberenet, Part B 27, 1–13 (1997)
47. El-Sonbaty, Y., Ismail, M.A.: Fuzzy Clustering for Symbolic Data. IEEE Transactions on Fuzzy Systems 6(2), 195–204 (1998)
48. Hayashi, I., Maeda, T., Bastian, A., Jain, L.C.: Generation of fuzzy decision trees by fuzzy ID3 with adjusting mechanism of and/or operators. In: Proc. Int. Conf. Fuzzy Syst., pp. 681–685 (1998)
49. Janikow, C.Z.: Fuzzy decision trees: Issues and Methods. IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics 28(1), 1–14 (1998)
50. Xizhao, W., Hong, J.: On the handling of fuzziness for continuous valued attributes in decision tree generation. Fuzzy Sets System 99, 283–290 (1998)
51. Gbyym, S.H.: A semi-linguistic fuzzy approach to multifactor decision-making: Application to aggregation of experts judgments. Annals of Nuclear Energy 26, 1097–1112 (1999)
52. Ishibuchi, H., Nakashima, T., Murata, T.: Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems. IEEE Transactions on Systems, Man, and Cybernatics-part B: Cybernetics 29(5), 601–618 (1999)
53. Belohlavek, R.: Lattices generated by binary fuzzy relations. Tatra Mountains Mathematical Publications 16, 11–19 (1999); special issue on fuzzy sets
54. Wang, L.D., Yen, J.: Extracting fuzzy rules for system modeling using a hybride of genetic algorithms and Kalman filter. Fuzzy Sets and Systems 101, 353–362 (1999)
55. Roberts, S.J., Everson, R., Rezek, I.: Maximum certainty data partitioning. Pattern Recognition 33, 833–839 (2000)
56. Wang, X., Chen, B., Qian, G., Ye, F.: On the optimization of fuzzy decision trees. Fuzzy Sets System 112, 117–125 (2000)
57. Ait Kbir, M., Benkirane, H., Maalmi, K., Benslimane, R.: Hierarchical fuzzy partition for pattern classification with fuzzy if-then rules. Pattern Recognition Letters 21, 503–509 (2000)
58. Pedrycz, W., Sosnowski, Z.A.: Designing decision trees with the use of fuzzy granulation. IEEE Trans. Syst., Man, Cybern. part A 30, 151–159 (2000)
59. Setnes, M., Roubos, H.: GA-fuzzy modeling and classification: Complexity and Performance. IEEE Transactions on Fuzzy Systems 8(5), 509–522 (2000)
60. Tsang, E.C.C., Wang, X.Z., Yeung, D.S.: Improving learning accuracy of fuzzy decision trees by hybrid neural networks. IEEE Trans. Fuzzy Syst. 8, 601–614 (2000)

61. Lee, H.M., Chen, C.M., Chen, J.M., Jou, Y.L.: An Efficient Fuzzy Classifier with feature Selection Based on Fuzzy Entropy. IEEE Transactions on Systems,Man,And Cybernetics-Part B: Cybernetics 31(3), 25–40 (2001)
62. Wang, X.Z., Yeung, D.S., Tsang, E.C.C.: A comparative study on heuristic algorithms for generating fuzzy decision trees. IEEE Transactions an Systems, Man, and Cybernetics–Part B: Cybernetics 31(2), 215–226 (2001)
63. Pedrycz, W., Sosnowski, Z.A.: The design of decision trees in the framework of granular data and their application to software quality models. Fuzzy Sets and Systems 123, 271–290 (2001)
64. Roubos, H., Setnes, M.: Compact and transparent fuzzy models and classifiers through iterative complexity reduction. IEEE Transations on Fuzzy Systems 9, 516–524 (2001)
65. Mitra, S., Konwar, K.M., Sankar, K.P.: Fuzzy decision tree, linguistic rules and fuzzy knowledge-based network: generation and evaluation. IEEE Transactions an Systems, Man, and Cybernetics–Part C: Applic. and Reviews 32(4), 328–339 (2002)
66. Wu, K.L., Yang, M.S.: Alternative c-means clustering algorithms. Pattern Recognition 35, 2267–2278 (2002)
67. Baraldi, A., Alpaydin, E.: Constructive feedforward ART clustering networks—Part I and II. IEEE Trans. Neural Netw. 13(3), 645–677 (2002)
68. Pedrycz, W., Vukovich, G.: logic-oriented fuzzy clustering. Pattern Recognition Letters 23, 1515–1527 (2002)
69. Fern'andez Salido, J.M., Murakami, S.: On β-precision aggregation. Fuzzy Sets and Systems 139, 15–25 (2003)
70. Abonyi, J., Roubos, J.A., Szeifert, F.: Data-driven generation of compact accurate and linguistically sound fuzzy classifiers based on a decision-tree initialization. International Journal of Approximate Reasoning 32, 1–21 (2003)
71. Nguyen, H.S.: On the Decision Table with Maximal Number of Reducts. Electronic Notes in Theoretical Computer Science 82(4), 8 pages (2003), http://www.elsevier.nl/locate/entcs/volume82.html
72. Zhong, N., Dong, J.Z., Ohsuga, S.: Meningitis data mining by cooperatively using GDT-RS and RSBR. Pattern Recognition Letters 24, 887–894 (2003)
73. Marcelloni, F.: Feature selection based on a modified fuzzy C-means algorithm with supervision. Information Sciences 151, 201–226 (2003)
74. Fernandez Salido, J.M., Murakami, S.: Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations. Fuzzy Sets and Systems 139, 635–660 (2003)
75. Roubos, J.A., Setnes, M., Abonyi, J.: Learning fuzzy classification rules from labeled data. Information Sciences 150, 77–93 (2003)
76. Yao, Y.Y.: Semantics of Fuzzy Sets in Rough Set Theory. In: Peters, J.F., Skowron, A., Dubois, D., Grzymała-Busse, J.W., Inuiguchi, M., Polkowski, L. (eds.) Transactions on Rough Sets II. LNCS, vol. 3135, pp. 297–318. Springer, Heidelberg (2004)
77. Geng, L., Chan, C.W.: An Algorithm for Case Generation from a Database. Applied Mathematics Letters 17, 269–274 (2004)
78. Chang, X.G., Lilly, J.H.: Evolutionary design of a fuzzy classifier from data. IEEE Transactions on Systems, Man, and Cybernatics-part B: Cybernetics 34(4), 1894–1906 (2004)

79. Greco, S., Matarazzo, B., Slowinski, R.: Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough-set decision rules. European Journal of Operational Research 158, 271–292 (2004)
80. Liang, G.S., Chou, T.Y., Han, T.C.: Cluster analysis based on fuzzy equivalence relation. European Journal of Operational Research 166, 160–171 (2005)
81. Pedrycz, W., Sosnowski, Z.A.: C-fuzzy decision trees. IEEE Transactions on Systems, Man, cybernetics-Part C: Applications and Reviews 35(4), 498–511 (2005)
82. Beliakov, G., King, M.: Density based fuzzy c-means clustering of non-convex patterns. European Journal of Operational Research 173, 717–728 (2006)
83. Hu, B.Q., Wang, S.: A novel approach to fuzzy programming part I: New arithmetic and order relations for interval numbers. Journal of Industrial and Management Optimization 2, 351–372 (2006)
84. Lee, H., Park, K., Bien, Z.: Iterative fuzzy clustering algorithm with supervision to construct probabilistic fuzzy rule base from numerical data. IEEE Transactions on Fuzzy Systems 16, 263–277 (2008)
85. Pedrycz, W.: Statistically grounded logic operators in fuzzy sets. European Journal of Operational Research (2008) doi:10.1016/j.ejor.2007.12.009

Formal Concept Analysis (FCA) and Its Generalization

1. Wille, R.: Restructuring lattice theory: an approach based on hierarchies of concepts. In: Ivan Rival, R. (ed.) Ordered Sets, pp. 445–470. Reidel, Dordecht (1982)
2. Missikoff, M., Scholl, M.: An algorithm for insertion into a lattice: application to type classification. In: Litwin, W., Schek, H.-J. (eds.) FODO 1989. LNCS, vol. 367, pp. 64–82. Springer, Heidelberg (1989)
3. Wille, R.: Knowledge acquisition by methods of formal concept analysis. In: Diday, E. (ed.) Data Analysis, Learning Symbolic and Numeric Knowledge, pp. 365–380. Nova Science, NewYork (1989)
4. Burusco, A., Gonzalez, R.F.: The Study of the L-Fuzzy Concept Lattice. Mathware and Soft Computing 1(3), 209–218 (1994)
5. Mineau, G., Godin, R.: Automatic structuring of knowledge bases by conceptual clustering. IEEE Transactions on Knowledge and Data Engineering 7(5), 824–829 (1995)
6. Vogt, F., Wille, R.: TOSCANA—-A graphical tool for analyzing and exploring data. In: Tamassia, R., Tollis, I(Y.) G. (eds.) GD 1994. LNCS, vol. 894, pp. 226–233. Springer, Heidelberg (1995)
7. Wille, R.: The basic theorem of triadic concept analysis. Order 12, 149–158 (1995)
8. Carpineto, C., Romano, G.: A lattice conceptual clustering system and its application to browsing retrieval. Mach. Learning 10, 95–122 (1996)
9. Pollandt, S.: Fuzzy-Begriffe: Formale Begriffsanalyse Unscharfer Daten. Springer, Heidelberg (1996)
10. Bodjanova, S.: Approximation of fuzzy concepts in decision making. Fuzzy Sets and Systems 85, 23–29 (1997)

11. Godin, R., Mili, H., Mineau, G., Missaoui, R., Arfi, A., Chau, T.: Design of class hierarchies based on concept Galois lattices. TAPOS 4(2), 117–134 (1998)
12. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer, Berlin (1999)
13. Belohlavek, R.: Fuzzy Galois connections. Mathematical Logic Quarterly 45, 497–504 (1999)
14. Stumme, G., Wille, R. (eds.): Begriffliche Wissensverarbeitung—-Methoden und Anwendungen. Springer, Heidelberg (2000)
15. Diaz-Agudo, B., Gonzalez-Calero, P.A.: Formal concept analysis as a support technique for CBR. Knowledge-Based Systems 14, 163–171 (2001)
16. Huynh, V.N., Nakamori, Y.: Fuzzy Concept Formation Based on Context Model. Knowledge-Based Intelligent Information Eng. Systems and Allied Technologies, 687–691 (2001)
17. Saquer, J., Deogun, J.: Concept approximations based on rough sets and similarity measure. International Journal of Appllied Mathematics and Compututer Science 11, 655–674 (2001)
18. Stumme, G., Taouil, R., Bastide, Y., Pasquier, N., Lakhal, L.: Computing iceberg concept lattices with TITANIC. Data & Knowledge Engineering 42, 189–222 (2002)
19. Fern'andez Salido, J.M., Murakami, S.: On β-precision aggregation. Fuzzy Sets and Systems 139, 15–25 (2003)
20. Saquer, J., Deogun, J.: Approximating monotone concepts. In: Design and Application of Hybrid Intelligent System, pp. 605–613 (2003)
21. Nguyen, H.S.: On the Decision Table with Maximal Number of Reducts. Electronic Notes in Theoretical Computer Science 82(4), 8 pages (2003), http://www.elsevier.nl/locate/entcs/volume82.html
22. Zhong, N., Dong, J.Z., Ohsuga, S.: Meningitis data mining by cooperatively using GDT-RS and RSBR. Pattern Recognition Letters 24, 887–894 (2003)
23. Deogun, J.S., Saquer, J.: Monotone concepts for formal concept analysis. Discrete Applied Mathematics 144, 70–78 (2004)
24. Geng, L., Chan, C.W.: An Algorithm for Case Generation from a Database. Applied Mathematics Letters 17, 269–274 (2004)
25. du Boucher-Ryana, P., Bridge, D.: Collaborative recommending using formal concept analysis. Knowledge-Based Systems 19, 309–315 (2006)
26. Tho, Q.T., Hui, S.C., Fong, A.C.M., Cao, T.H.: Automatic Fuzzy Ontology Generation for Semantic Web. IEEE Transactions on Knowledge and Data Engineering 18(6), 842–856 (2006)
27. Zhang, L.S., Liu, X.D.: Concept lattice and AFS algebra. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS (LNAI), vol. 4223, pp. 290–299. Springer, Heidelberg (2006)
28. Shao, M., Liu, M., Zhang, W.: Set approximations in fuzzy formal concept analysis. Fuzzy Sets and Systems 23, 2627–2640 (2007)
29. Zhang, W., Ma, J., Fan, S.: Variable threshold concept lattices. Information Sciences 177, 4883–4892 (2007)
30. Zhao, Y., Halang, W.A., Wang, X.: Rough ontology mapping in E-Business integration. Studies in Computational Intelligence 37, 75–93 (2007)
31. Hu, Q., Liu, J., Yu, D.: Mixed feature selection based on granulation and approximation. Knowledge-Based Systems 21, 294–304 (2008)

32. Wang, L.D., Liu, X.D.: Concept analysis via rough set and AFS algebra. Information Sciences 178, 4125–4137 (2008)
33. Wang, L.D., Liu, X.D.: A new model of evaluating concept similarity. Knowledge-Based Systems 21, 842–846 (2008)

Rough Sets and Fuzzy Rough Sets

1. Zadeh, L.A.: A Theory of Approximate Reasoning. In: Hayes, J.E., Michie, D., Mikulich, L.I. (eds.) Machine Intelligence, vol. 9, pp. 149–194. Elsevier, New York (1979)
2. Pawlak, Z.: Rough sets. Internat. J. Comput. Inform. Sci. 11(5), 341–356 (1982)
3. Pawlak, Z.: Rough sets. International Journal of Computer and Information Science 11, 341–356 (1982)
4. Pawlak, Z.: Rough Probability. Bull. Polish Acad. Sci., Math. 32, 607–612 (1984)
5. Farinas del Cerro, L., Prade, H.: Rough sets, twofold fuzzy sets and modal logic-Fuzziness in indiscernibility and partial information. In: di Nola, A., Ventre, A.G.S. (eds.) The Mathematics of Fuzzy Systems, pp. 103–120. Verlag TU V, Rheinland (1986)
6. Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets. Internat. J. General Systems 17(2-3), 191–209 (1990)
7. Dubois, D., Prade, H.: Putting rough sets and fuzzy sets together. In: Slowinski, R. (ed.) Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory, pp. 203–222. Kluwer, Dordrecht (1992)
8. Kuncheva, L.I.: Fuzzy rough sets: application to feature extraction. Fuzzy Sets and Systems 51, 147–153 (1992)
9. Nakamura, A.: Application of fuzzy-rough classifications to logics. In: Slowinski, R. (ed.) Intelligent Decision Support: Handbook ofApplications and Advances ofthe Rough Sets Theory, pp. 233–250. Kluwer, Dordrecht (1992)
10. Skowron, A., Rauszer, C.: The Discernibility Matrices and Functions in Information Systems, Intelligent Decision Support. In: Slowinski, R. (ed.) Handbook of Applications and Advances of the Rough Sets Theory, pp. 331–362. Kluwer Academic, Dordrecht (1992)
11. Ziarko, W.: Variable precision rough set model. J. Comput. System Sci. 46, 39–59 (1993)
12. Pawlak, Z.: Hard and soft sets. In: Ziarko, W.P. (ed.) Rough Sets, Fuzzy Sets and Knowledge Discovery, pp. 130–135. Springer, London (1994)
13. Slowinski, R., Vanderpooten, D.: Similarity relations as a basis for rough approximations. In: Wang, P.P. (ed.) Advances in Machine Intelligence and Soft Computing, Bookwrights, Raleigh, NC, pp. 17–33 (1997)
14. Yao, Y.: Combination of rough and fuzzy sets based on α-level sets. In: Lin, T.Y., Cercone, N. (eds.) Rough sets and Data Mining, pp. 301–321. Kluwer, Dordrecht (1997)
15. Greco, S., Matarazzo, B., Slowinski, R.: Fuzzy similarity relations as a basis for rough approximations. In: Polkowski, L., Skowron, A. (eds.) Proc. 1st Int. Conf. on Rough Sets and Current Trends in Computing, RSTC 1998, Warsaw, Poland, pp. 283–289 (1998)
16. Yao, Y.: A comparative study of fuzzy sets and rough sets. Information Sciences 109, 227–242 (1998)

17. Greco, S., Matarazzo, B., Slowinski, R.: Rough approximation of a preference relation by dominance relations. European Journal of Operational Research 117, 63–83 (1999)
18. Slowinski, R., Vanderpooten, D.: A Generalized Definition of Rough Approximations Based on Similarity. IEEE Transactions on Knowledge and Data Engineering 12(2), 331–336 (2000)
19. Greco, S., Matarazzo, B., Slowinski, R.: Rough sets theory for multicriteria decision analysis. European Journal of Operational Research 129, 1–47 (2001)
20. Kim, D.: Data classification based on tolerant rough set. Pattern Recognition 34, 1613–1624 (2001)
21. Zaras, K.: Rough approximation of a preference relation by a multi-attribute stochastic dominance for determinist and stochastic evaluation problems. European Journal of Operational Research 130, 305–314 (2001)
22. Greco, S., Matarazzo, B., Slowinski, R.: Rough sets methodology for sorting problems in presence of multiple attributes and criteria. European Journal of Operational Research 138, 247–259 (2002)
23. Radzikowska, A.M., Kerre, E.E.: A comparative study of fuzzy rough sets. Fuzzy Sets and Systems 126, 137–155 (2002)
24. Bonikowski, Z.: Algebraic Structures of Rough Sets in Representative Approximation Spaces. Electronic Notes in Theoretical Computer Science 82(4), 1–12 (2003), http://www.elsevier.nl/locate/entcs/volume82.html
25. Fern'andez Salido, J.M., Murakami, S.: On β-precision aggregation. Fuzzy Sets and Systems 139, 15–25 (2003)
26. Dembczynski, K., Pindur, R., Susmag, R.: Generation of Exhaustive Set of Rules within Dominance-based Rough Set Approach. Electronic Notes in Theoretical Computer Science 82(4), 1–12 (2003), http://www.elsevier.nl/locate/entcs/volume82.html
27. Fernandez Salido, J.M., Murakami, S.: Rough Set Analysis of a General Type of Fuzzy Data Using Transitive Aggregations of Fuzzy Similarity Relations. Fuzzy Sets and Systems 139, 635–660 (2003)
28. Wang, Q.H., Li, J.R.: A rough set-based fault ranking prototype system for fault diagnosis. Engineering Applications of Artificial Intelligence 17, 909–917 (2004)
29. Greco, S., Matarazzo, B., Slowinski, R.: Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough-set decision rules. European Journal of Operational Research 158, 271–292 (2004)
30. Pal, S.K., Mitra, P.: Case Generation Using Rough Sets with Fuzzy Representation. IEEE Transactions on Knowledge and Data Engineering 16(3), 292–300 (2004)
31. Yao, Y.: Concept lattices in rough set theory. In: Proceedings of 2004 Annual Meeting Meeting of the North American Fuzzy Information Processing Society, pp. 796–801 (2004)
32. Radzikowska, A., Kerre, E.E.: A comparative study of fuzzy rough sets. Fuzzy Sets and Systems 126, 137–155 (2004)
33. Pawlak, Z., Skowron, A.: Rudiments of Rough Sets. Information Sciences 177, 3–27 (2007)
34. Pawlak, Z., Skowron, A.: Rough sets: Some Extensions. Information Sciences 177, 28–40 (2007)

35. Pawlak, Z., Skowron, A.: Rough Sets and Boolean Reasoning. Information Sciences 177, 41–70 (2007)
36. Wang, C., Wu, C., Chen, D.: A systematic study on attribute reduction with rough sets based on general binary relations. Information Sciences 178, 2237–2261 (2008)
37. Wang, L.D., Liu, X.D.: Concept analysis via rough set and AFS algebra. Information Sciences 178, 4125–4137 (2008)
38. Zhang, J., Wang, Y.: A rough margin based support vector machine. Information Sciences 178, 2204–2214 (2008)

Index