The Fuzzy Clustering Algorithm Based on AFS Topology

Rui Ding[1], Xiaodong Liu[1,2], and Yan Chen[1]

[1] Economics and Management College, Dalian Maritime University, Dalian,116026, P.R. China
[2] Research Center of Information and Control, Dalian University of Technology, Dalian,116024, P.R. China
joking810908@163.com

Abstract. This paper establishes a new metric space for the clustering problems. The neighbors on the object set induced by the topology molecular lattice on $^*EI$ algebra are given and a new distance based on the neighbors is proposed. In the proposed clustering algorithm, the Euclidean metric is replaced by the new distance based on the order relationship of the samples on the attributes. As a result, using the method to Iris data we show it has a better result and clearer classification than the other clustering algorithm based on the Euclidean metric. This study shows that the AFS topology fuzzy clustering algorithm can obtain an high clustering accuracy according to order relationship.

1 Introduction

Fuzzy sets and systems has been developed rapidly and applied in many fields since it was proposed by Prof. Zadeh [1]. However, a fuzzy set is a rather abstract notion. Fuzzy sets are useful for many purposes, and membership functions do not mean the same thing at the operational level in each and every context. We are often perplexed by the problem how to properly determine the membership function according to the concrete situation. In order to deal with the above discussed problems, AFS (Axiomatic Fuzzy Set) theory was firstly proposed by Liu in 1995 [4]. In essence, the AFS framework provides an effective tool to convert the information in the training examples and databases into the membership functions and their fuzzy logic operations. AFS fuzzy logic can be applied to the data sets with various data types such as real numbers, Boolean value, partial order, even human intuition descriptions, which are very difficult or unsolved for other clustering algorithm such as the fuzzy c-means algorithm (FCMA) conceived by Dunn [8] and generalized by Bezdek [9] and the fuzzy k nearest neighbor algorithm, named as k-NN algorithm [10].

We know that human can classify, cluster and recognize the objects in the ordinary data 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

L. Wang et al. (Eds.): FSKD 2006, LNAI 4223, pp. 89–98, 2006.

process is very important though it may not be based on the Euclidean metric. In [6, 11], using topological molecular theory, the topological structures on X induced by the topological molecular lattices generated by some fuzzy sets in EM have been obtained. This kind topology on X is determined by the original data and the chosen fuzzy sets in EM. It is an abstract geometry relations among the objects in X, the interpretations of the special topological structures on the AFS structures directly obtained by a given data set through the differential degrees between objects in X. With the topological space on X induced by the fuzzy concepts, the pattern recognition problems of ordinary data sets can be studied.

In this paper, we applied the topological structures induced by some concepts in EM to establish the metric for clustering problems. We study the topology molecular lattice on $^*EI$ algebra over some concepts, which based on AFS structure and AFS algebra, then give the neighbors on the object set reduced by the topology molecular lattice. We apply the neighbors of the topology to define a distance to study the fuzzy clustering analysis and the example shows that the new clustering algorithm is effective.

2 Preliminaries

In this section, we will recall the notations and definitions of AFS theory. AFS theory is made of AFS structures which is a special kind of combinatorics object [20] and AFS algebra which is a family of completely distributive lattices [7]. About the detail mathematical properties of AFS algebras please see [2-6, 11-19].

Definition 1 ([13]). *Let ζ be any concept on the universe of discourse X. R_ζ is called a binary relation (i.e., $R_\zeta \subset X \times X$) of ζ if R_ζ satisfies: $x, y \in X$, $(x, y) \in R_\zeta \Leftrightarrow x$ belongs to ζ at some degree and the degree of x belonging to ζ is larger than or equals to that of y, or x belongs to ζ at some degree and y does not at all.*

Definition 2 ([2,3]). *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:*

D5-1. If $(x, y) \in R$, then $(x, x) \in R$;
D5-2. If $(x, x) \in R$ and $(y, y) \notin R$, then $(x, y) \in R$;
D5-3. If (x, y), $(y, z) \in R$, then $(x, z) \in R$;
D5-4. If $(x, x) \in R$ and $(y, y) \in R$, then either $(x, y) \in R$ or $(y, x) \in R$.

In addition, ζ is called a simple concept or simple attribute on X if R_ζ is a sub-preference relation on X. Otherwise ζ is called a complex concept or a complex attribute on X.

Definition 3 ([2,3]). *Let X, M be two sets and 2^M be the power set of M, $\tau : X \times X \to 2^M$. (M, τ, X) is called an AFS structure if τ satisfies the following conditions:*

Table 1. Date Set

sample	age	weight	height	male	female	salary	fortune
x_1	21	50	1.69	yes	no	0	0.000
x_2	30	52	1.62	no	yes	120	200.000
x_3	27	65	1.80	yes	no	100	40.000
x_4	60	63	1.50	no	yes	80	324.000
x_5	45	54	1.71	yes	no	140	486.940.000

AX1: $\forall(x_1, x_2) \in X \times X$, $\tau(x_1, x_2) \subseteq \tau(x_1, x_1)$;
AX2: $\forall(x_1, x_2)$, $(x_2, x_3) \in X \times X$, $\tau(x_1, x_2) \cap \tau(x_2, x_3) \subseteq \tau(x_1, x_3)$.
In addition, X is called universe of discourse, M is called an attribute set and τ is called a structure.

In practice, we always suppose that every concept in M is a simple concept on X. We can verify that (M, τ, X) is an AFS structure if τ is defined by

$$\tau(x_i, x_j) = \{m|m \in M, (x_i, x_j) \in R_m\}, x_i, x_j \in X.$$

Example 1. Let $X = \{x_1, x_2, \ldots, x_5\}$ be a set of five persons. $M = \{m_1, m_2, \ldots, m_7\}$, where m_1=age, m_2=weight, m_3=height, m_4=male, m_5=female, m_6=salary, m_7=fortune, suppose there exists Table1:

According to Table1 and the preference relations, $\tau(x_1, x_1)$={age, height, weight, salary, male}. This indicates that the person x_1 has the properties m_1, m_2, m_3, m_4. Similarly for $\tau(x_i, x_i)$, $i = 2, \ldots, 10$. $\tau(x_4, x_5) = \{m_1, m_2, m_5\}$. This implies that the degree of x_4 possessing properties m_1, m_2, m_5 is larger than that of x_5 or equal. Similarly for $\tau(x_i, x_j)$, $i, j = 1, 2, \ldots, 10$. It easily verifies that τ satisfies AX1, AX2 and (M, τ, X) is an AFS structure.

In order to study fuzzy concepts and their topological structure, we introduce $^*EI$ algebra($^*EI$ algebra is the opposite of EI algebra).

Definition 4 ([11]). *Let M be sets. In general, M is a set of fuzzy or crisp concepts,*

$$EM^* = \{\sum_{i \in I} A_i | A_i \subseteq M, i \in I, I \textit{ is any no-empty indexing set}\}.$$

Each $\sum_{i \in I} A_i$ is an element of EM^*, where $\sum_{i \in I}$ is just a symbol meaning that element $\sum_{i \in I} A_i$ is composed of $A_i \subseteq M$, $i \in I$ separated by symbol "+". When I is a finite indexing set, $\sum_{i=1}^{n} A_i$ is also denoted as $A_1 + A_2 + \cdots + A_n$. $\sum_{i \in I} A_i$ represents the same element of EM^* when these $A_i (i \in I)$ are summed by different orders, for example, $\sum_{i \in \{1,2\}} A_i = A_1 + A_2 = A_2 + A_1$.

Definition 5 ([11]). *Let M be a non-vacuous set. We define a binary relation R on EM^* as follows: $\forall \sum_{i \in I} A_i$, $\sum_{j \in J} B_j \in EM^*$,*

$$(\sum_{i \in I} A_i) R (\sum_{j \in J} B_j) \Leftrightarrow \forall A_i (i \in I), \exists B_h (h \in J)$$

such that $A_i \supseteq B_h$ *and* $\forall\ B_j(j \in J), \exists A_u(u \in I)$ *such that* $B_j \supseteq A_u$. *It is obvious that* R *is an equivalence relation. We denote* EM^*/R *as* EM. *By* $\sum_{i\in I} A_i = \sum_{j\in J} B_j$, *we mean that* $\sum_{i\in I} A_i$ *and* $\sum_{j\in J} B_j$ *are equivalent under the equivalence relation* R.

Theorem 1 ([11]). *Let* $X_1, ..., X_n, M$ *be* $n + 1$ *non-empty sets. Then* $(EM, \vee, \wedge)$ *forms a completely distributive lattice under the binary operations* $\vee, \wedge$ *defined as follows:* $\forall \sum_{i\in I} A_i,\ \sum_{j\in J} B_j \in EM$,

$$\sum_{i\in I} A_i \wedge \sum_{j\in J} B_j = \sum_{k\in I\sqcup J} C_k,$$

$$\sum_{i\in I} A_i \vee \sum_{j\in J} B_j = \sum_{i\in I, j\in J} (A_i \cup B_j),$$

where $\forall k \in I \sqcup J$ *(disjoin union of indexing sets* I *and* J*),* $C_k = A_k$ *if* $k \in I$ *and* $C_k = B_k$ *if* $k \in J$. $(EM, \vee, \wedge)$ *is called the* $^*EI$ *(expanding one set* M*) algebra over* M. $\emptyset$ *are the maximum and* M *is minimum element of* EM.

Theorem 2 ([15,16]). *Let* M *be a set.* $\forall \sum_{i\in I} A_i \in EM$, *if the operator "* $'$ *" is defined as follows*

$$(\sum_{i\in I} A_i)' = \wedge_{i\in I}(\vee_{a\in A_i}\{a'\}),$$

then " $'$ *" is an order-reversing involution on* $^*EI$ *algebra* EM.

3 Fuzzy Clustering Algorithmic Based on Topological Structure and $^*EI$ Algebra

In this section, we will discuss the topological molecular lattice structures on $^*EI$ algebras; and give the relations of these topological structures. As applications, we study the topology produced by a family of fuzzy concepts on $^*EI$ algebras and apply these to analyze relations among fuzzy concepts. Using these, we believe that we can study the law of human thinking. The most important fact is that all these can be operated by computers.

Definition 6 ([6,11]). *Let* X *and* M *be sets, and* (M, τ, X) *be an AFS structure.* $\eta \subseteq EM$ *, the* $^*EI$ *algebra over* M, η *is called a closed topology . if* $\sum_{m\in M}\{m\}, M \in \eta$, *and* η *is closed under finite unions (*$\vee$*or* $*$*) and arbitrary intersections (*$\wedge$*or*$+$*).* η *is called a topological molecular lattice on* $^*EI$ *algebra over* M *of AFS structure*(M, τ, X), *denoted as*(EM, η) *(* $\sum_{m\in M}\{m\}$ *is the minimal element, and* M *is the maximal element).*

Definition 7 ([6,11]). *Let* X *and* M *be sets, and* (M, τ, X) *be an AFS structure.* η *is a topological molecular lattice on* $^*EI$ *algebra over* M *of AFS structure* (M, τ, X). *For any* $x \in X$, $\sum_{i\in I} A_i \in EM$ *, and* $\sum_{i\in I} A_i \in \eta$ *we define*

$$N_{\sum_{i\in I} A_i}(x) = \{y|\tau(x,y) \geq \sum_{i\in I} A_i\}$$

this is called the neighborhood of x inducing by $\sum_{i\in I} A_i \in \eta$.

$$N_\eta(x) = \{N_{\sum_{i\in I} A_i}(x)| \sum_{i\in I} A_i \in \eta\}$$

is called the neighborhood of x inducing by η.

Theorem 3 ([6,11]). *Let X and M be sets, and (M,τ,X) be an AFS structure. η is a topological molecular lattice on $^*EI$ algebra over M of AFS structure (M,τ,X). if*

$$B = \{N_{\sum_{i\in I} A_i}(x)|x \in X, \sum_{i\in I} A_i \in \eta\}$$

then B is a base for some topology.

The topological space (X,T_η), in which B is a base for, T_η is called the topology induced by η.

As in example1, we consider the relations among age, height, and weight. Let η be the topological molecular lattice generated by $\{m_1\},\{m_2\},\{m_3\}$, which are elements in $^*EI$ algebra over M. $\eta(m_1,m_2,m_3)$ consists of the following:
$\alpha_1 = \{m_1\}+\{m_2\}+\{m_3\}$; $\alpha_2 = \{m_1\}+\{m_2\}$; $\alpha_3 = \{m_1\}+\{m_3\}$;
$\alpha_4 = \{m_2\}+\{m_3\}$; $\alpha_5 = \{m_1\}$; $\alpha_6 = \{m_2\}$; $\alpha_7 = \{m_3\}$;
$\alpha_8 = \{m_1,m_2\}+\{m_1,m_3\}+\{m_2,m_3\}$; $\alpha_9 = \{m_1,m_2\}+\{m_1,m_3\}$;
$\alpha_{10} = \{m_1,m_2\}+\{m_2,m_3\}$; $\alpha_{11} = \{m_1,m_3\}+\{m_2,m_3\}$; $\alpha_{12} = \{m_1,m_2\}$;
$\alpha_{13} = \{m_1,m_3\}$; $\alpha_{14} = \{m_2,m_3\}$; $\alpha_{15} = \{m_1\}+\{m_2,m_3\}$;
$\alpha_{16} = \{m_2\}+\{m_1,m_3\}$; $\alpha_{17} = \{m_3\}+\{m_1,m_2\}$; $\alpha_{18} = \{m_1,m_2,m_3\}$;
$\alpha_{19} = \emptyset$;

Now we consider the base of the topology for$\{x_1,x_2,x_3,x_4,x_5\}$:
$N_{\alpha 1}(x_1) = \{x_1,x_2,x_4\}$; $N_{\alpha 2}(x_1) = \{x_1,x_2,x_4\}$; $N_{\alpha 3}(x_1) = \{x_1\}$;
$N_{\alpha 4}(x_1) = \{x_1,x_2,x_4\}$; $N_{\alpha 5}(x_1) = \{x_1\}$; $N_{\alpha 6}(x_1) = \{x_1,x_2,x_4\}$;
$N_{\alpha 7}(x_1) = \{x_1\}$; $N_{\alpha 8}(x_1) = \{x_1\}$; $N_{\alpha 9}(x_1) = \{x_1\}$;
$N_{\alpha 10}(x_1) = \{x_1\}$; $N_{\alpha 11}(x_1) = \{x_1\}$; $N_{\alpha 12}(x_1) = \{x_1\}$;
$N_{\alpha 13}(x_1) = \{x_1\}$; $N_{\alpha 14}(x_1) = \{x_1\}$; $N_{\alpha 15}(x_1) = \{x_1\}$;
$N_{\alpha 16}(x_1) = \{x_1,x_2,x_4\}$; $N_{\alpha 17}(x_1) = \{x_1\}$; $N_{\alpha 18}(x_1) = \{x_1\}$;

Therefore the neighborhoods of x_1 induced by η is
$N\eta(x_1) = \{\{x_1,x_2,x_4\},\{x_1\}\}$;
Similarly, we get the neighborhoods of x_i $(i = 2....5)$ induced by η is
$N\eta(x_2) = \{\{x_1,x_2,x_3,x_4\},\{x_1,x_2,x_4\},\{x_1,x_2,x_3\},\{x_1,x_2\},\{x_2,x_4\},\{x_2\}\}$;
$N\eta(x_3) = \{\{x_1,x_2,x_3,x_4,x_5\},\{x_1,x_3\}\}$;
$N\eta(x_4) = \{\{x_1,x_2,x_3,x_4,x_5\},\{x_1,x_2,x_4,x_5\},\{x_4\}\}$;
$N\eta(x_5) = \{\{x_1,x_2,x_3,x_4,x_5\},\{x_1,x_2,x_3,x_5\},\{x_1,x_2,x_4,x_5\},\{x_1,x_2,x_5\}\}$;

Definition 8. *Let X and M be sets, and (M, τ, X) be an AFS structure. η is a topological molecular lattice on $^*EI$ algebra over M of AFS structure (M, τ, X). if only choose $\wedge$ operation i.e. η^* is closed under arbitrary intersections. η^* is called a intersectant topological molecular lattice on $^*EI$ algebra over M of AFS structure (M, τ, X), denoted as (EM, η^*)*

As in the example, let η^* be the intersectant topological molecular lattice generated by $\{m_1\}, \{m_2\}, \{m_3\}$, which are elements in $^*EI$ algebra over M. $\eta^*(m_1, m_2, m_3)$ consists of the following:

$\alpha_5 = \{m_1\}$; $\alpha_6 = \{m_2\}$; $\alpha_7 = \{m_3\}$; $\alpha_{12} = \{m_1, m_2\}$;
$\alpha_{13} = \{m_1, m_3\}$; $\alpha_{14} = \{m_2, m_3\}$; $\alpha_{18} = \{m_1, m_2, m_3\}$;

Definition 9. *Let $N_\eta(x)$ is the neighborhood of x induced by η .*

$$N_{x_i}^{x_j} = \{\delta \in N_\eta | x_i \in \delta, x_j \notin \delta\}.$$

then the distance from x_i to x_j is $d_{i-j} = \sum_{\delta \in N_{x_i}^{x_j}} |\delta|$. ($|\delta|$ is the length of the neighbor, i.e. the number of the topology base which produced the neighbor). Similarly $d_{j-i} = \sum_{\delta \in N_{x_j}^{x_i}} |\delta|$, so define the distance between x_i and x_j is $d(i, j) = (d_{i-j} + d_{j-i})/2$.

In example1 the neighbors including x_1 but not including x_2 are $\{x_1\}, \{x_1, x_3\}$, since the neighbor $\{x_1\}$ is produced by the following 13 base in $\eta(m_1, m_2, m_3)$:

$N_{\alpha 3}(x_1) = \{x_1\}$; $N_{\alpha 5}(x_1) = \{x_1\}$; $N_{\alpha 7}(x_1) = \{x_1\}$; $N_{\alpha 8}(x_1) = \{x_1\}$;
$N_{\alpha 9}(x_1) = \{x_1\}$; $N_{\alpha 10}(x_1) = \{x_1\}$; $N_{\alpha 11}(x_1) = \{x_1\}$; $N_{\alpha 12}(x_1) = \{x_1\}$;
$N_{\alpha 13}(x_1) = \{x_1\}$; $N_{\alpha 14}(x_1) = \{x_1\}$; $N_{\alpha 15}(x_1) = \{x_1\}$;$N_{\alpha 17}(x_1) = \{x_1\}$;
$N_{\alpha 18}(x_1) = \{x_1\}$;

So $|\{x_1\}| = 13$, similarly $|\{x_1, x_3\}| = 5$, the distance from x_1 to x_2 is $d_{1-2} = 18$. the neighbors including x_2 but not including x_1 are $\{x_2, x_4\}, \{x_2\}$, the sum of the length is 5, so the distance from x_2 to x_1 is $d_{2-1} = 5$, then $d(1, 2) = (d_{1-2} + d_{2-1})/2 = 11.5$.

In the following, we describe the design method:

Let X be the universe of discourse, M be a set of simple features on X.

Step1: Consider the intersectant topological molecular lattice (EM, η^*) generated by all the correlative concepts $\Lambda \subseteq EM$, where Λ is a set of fuzzy sets which are selected to cluster the objects in X.

Step2: Establish AFS structure (M, τ, X) based on the original data and facts (refer to Example 1), and then get the neighborhoods $N_\eta(x)$ induced by the correlative intersectant topological molecular lattice (refer Definition 7).

Step3: For each $x \in X$, apply Definition 9 to calculate the distance between the objects, according to the neighborhoods.

Step4: Apply the distance between each $x \in X$ to establish the fuzzy relation matrix $M = (m_{ij})$ on $X = \{x_1, x_2, \ldots, x_n\}$. Since for any $i, j = 1, 2, ..., n$, $m_{ij} \leq m_{ii}$, hence $M^k \leq M^{k+1}$ for any $k = 1, 2, ...$ Therefore exists an integer r

such that $(M^r)^2 = M^r$, i.e., fuzzy relation matrix $R = M^r$ can yield a partition tree with equivalence classes.

Step5: Clustering analysis based on the fuzzy relation matrix R.

The distance between objects is defined acorrding to the neighbors induced by the concepts in Λ, which reflect the relation between the objects consider the concepts in Λ. We intend to cluster the objects based on the abstract geometrical relations determined by the selected concepts in Λ. The basic idea of the appraoch is based on the following observation:

(1) If none or few of the neighbors can separate the two objects x, y, then the distance of x, y is small;

(2) If any or a large numbers of the neighbors can separate the two objects x, y, then the distance of x, y is large;

(3) If the big neighbor can separate the two objects x, y, then the distance is large.

4 Example

In this section, we apply the AFS topology clustering algorithm to Fisher Iris data, which is well known to the pattern recognition community. The data set contains 150 patterns for 3 classes, each class has 50 instances, each class refers to a type of Iris plant. One class is linearly separable from other two but the latter are not linearly separable from each other. Patten classes are Iris-Setosa, Iris-Versicolor and Iris-Virginica. The four features of ordinal variables involved are the sepal length, sepal width, petal length, and petal width, respectively.

Let $X = \{x_1, x_2, \cdots, x_{150}\}$. From x_1 to x_{50} are "Iris-Setosa", from x_{51} to x_{100} are "Iris-Versicolor", and from x_{101} to x_{150} are "Iris-Virginica". In order to acquire more information from the original data, we expand the original four features to eight. Let M be a set of simple attributes on X, $M = \{m_1, m_1^{'}, m_2, m_2^{'}, m_3, m_3^{'}, m_4, m_4^{'}\}$, where m_1 =sepal length, $m_1^{'}$ =sepal short, m_2 =sepal width, $m_2^{'}$ =sepal narrow, m_3 =petal length, $m_3^{'}$ =petal short, m_4 =petal width, $m_4^{'}$ =petal narrow.

Step1:

Let $\Lambda = \{m_1, m_1^{'}, m_2, \cdots, m_4^{'}\} \subseteq EM$. Then calculate the intersectant topological molecular lattice (EM, η^*) generated by the concepts in Λ.

Since the intersect between the correlative concepts and their negation is near empty, hence we can ignore the neighbors induced by the concepts such as $m_i \wedge m_i'$, $i = 1, 2, 3, 4$. Thus the neighbor system of the topology induced by the concepts in Λ can be reduced greatly.

Step2:

Establish AFS structure (M, τ, X) based on $X = \{x_1, x_2, \cdots, x_{150}\}$ and $M = \{m_1, m_1^{'}, m_2, \cdots, m_4^{'}\}$, and get the neighborhoods of x induced by $\eta^*(m)$, which denoted as $N_{\eta^*(m)}(x_i)$ (refer to Definition 7).

Step3:

Calculate the distance between $X = \{x_1, x_2, \cdots, x_{150}\}$, establish distance matrix:

$D = (d_{ij}) =$

$$\begin{bmatrix} 0 & 749 & \cdots & 1727 & 1577 & \cdots & 1967.5 & 2023.5 & \cdots & 1849 \\ & 0 & \cdots & 1843 & 1707 & \cdots & 2152.5 & 2161 & \cdots & 1625 \\ & & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & 0 & 413 & \cdots & 971.5 & 1236.5 & \cdots & 992 \\ & & & & 0 & \cdots & 919.5 & 1156.5 & \cdots & 851 \\ & & & & & \ddots & \vdots & \vdots & \vdots & \vdots \\ & & & & & & 0 & 1146 & \cdots & 943.5 \\ & & & & & & & 0 & \cdots & 517.5 \\ & & & & & & & & \ddots & \vdots \\ & & & & & & & & & 0 \end{bmatrix}$$

Step4:
Standardize these distance:

$$d^*(i, j) = d(i, j)/max(d(:, j)),$$

then get the similar relation:

$$N(x_i, x_j) = 1 - d^*(x_i, x_j).$$

Transforming the similarity matrix into its transitive closure, the fuzzy equivalent matrix as follows, which can yield a partition tree with equivalence classes:
$R =$

$$\begin{bmatrix} 1 & 0.8758 & \cdots & 0.7242 & 0.7242 & \cdots & 0.7242 & 0.7242 & \cdots & 0.7242 \\ & 1 & \cdots & 0.7242 & 0.7242 & \cdots & 0.7242 & 0.7242 & \cdots & 0.7242 \\ & & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & 1 & 0.8568 & \cdots & 0.8408 & 0.8408 & \cdots & 0.8408 \\ & & & & 1 & \cdots & 0.8408 & 0.8408 & \cdots & 0.8408 \\ & & & & & \ddots & \vdots & \vdots & \vdots & \vdots \\ & & & & & & 1 & 0.8454 & \cdots & 0.8425 \\ & & & & & & & 1 & \cdots & 0.8425 \\ & & & & & & & & \ddots & \vdots \\ & & & & & & & & & 1 \end{bmatrix}$$

We have validated that $R^2 = R$.

According to different thresholds, we get dynamic cluster results, and finally the most accurate result is when threshold $\lambda = 0.8409$, the result we got is accord with the nature the Iris have.

When threshold $\lambda = 0.8409$, cluster one is:
$x_1, \ldots, x_{22}, x_{24}, \ldots, x_{41}, x_{43}, \ldots, x_{50}$.
cluster two is:
$x_{51}, \ldots, x_{68}, x_{70}, x_{72}, x_{74}, \ldots, x_{77}, x_{79}, \ldots, x_{83}, x_{85}, \ldots, x_{87}, x_{89}, \ldots, x_{100}$.

cluster three is:
$x_{69}, x_{71}, x_{73}, x_{78}, x_{84}, x_{88}, x_{101}, \ldots, x_{106}, x_{108}, x_{111}, \ldots, x_{117}, x_{119}, \ldots, x_{131}, x_{133}, x_{134}, x_{136}, \ldots, x_{150}$.

There are two classifying errors in the class "Iris-Setosa"; there are six patterns in class "Iris-Versicolor" distributed to class "Iris-Virginica", and there are six patterns away from class "Iris-Virginica", i.e. total 14 classification errors. The clustering accurate rate is 90.67%.

We apply Euclidean metric for traditional algorithm to establish distance matrix[21] and transform it into its transitive closure, the most accurate result is when threshold $\lambda = 0.94182$, total 29 patterns were error, clustering accurate rate is 80.67%. Using the function kmeans in MATLAB toolbox for the iris-data, which is based on the well known k-mean clustering algorithm, the clustering accuracy rate is 89.33%. And Using the function fcm in MATLAB toolbox for the iris-data, which is based on the well known fuzzy c-mean clustering algorithm, the clustering accuracy rate is also 89.33%.

5 Conclusion

In this paper, we established metric space based on the topological structures induced by the involved fuzzy concepts in the AFS framework, proposed measure for membership functions and got the fuzzy similarity relations on X, then applied the measure to study the clustering analytic problems. The AFS topology clustering algorithm is applied to the well known iris-data, and an high clustering accurate rate is achieved. By the comparison of the accuracy with the current fuzzy clustering algorithm, such as c-means fuzzy algorithm, k-nearest-neighbor fuzzy algorithm, and Euclidean metric transitive closure algorithm, one can observe that:

(1) The performance of our algorithm is quite well;

(2) The clustering algorithms based on the topological distance are more simple and understandable, they needn't repeat to convergence;

(3) The attributes of objects in it can be various data types or sub-preference relations, even human intuition descriptions. But both k-mean, fuzzy c-mean algorithms and other current fuzzy clustering algorithm can only be applied to the data sets with numerical attributes;

(4) The cluster number or the class label need not be given beforehand.

From these results, we can conclude that the performance of our proposed algorithm is comparable with many other pattern clustering algorithms and can be treated as one of the most suitable clustering algorithm.

Acknowledgement

This work is supported in part by the National Natural Science Foundation of China under Grant 60575039 60534010 and in part by the National Key Basic Research and Development Program of China under Grant 2002CB312201-06.

References

1. Zadeh, L.A.: Fuzzy sets. Inf. Control **18** (1965) 338–353
2. Liu, X.D.: The fuzzy theory based on AFS algebras and AFS structure. J. Math. Anal. Appl. **217** (1998) 459–478
3. Liu, X.D.: The fuzzy sets and systems based on AFS structure, EI algebra and EII algebra. Fuzzy Sets Syst. **95** (1998) 179–188
4. Liu, X.D.: A new mathematical axiomatic system of fuzzy sets and systems. J. Fuzzy Math. **3** (1995) 559–560
5. Liu, X.D.: Two algebra structures of AFS structure. J. Fuzzy Math. **3** (1995) 561–562
6. Liu, X.D.: The topology on AFS algebra and AFS structure. J. Math. Anal. Appl. **217** (1998) 479–489
7. Wang, G.j.: Theory of topological molecular lattices. Fuzzy Sets Syst. **47** (1992) 351–376
8. J. C. Dunn.: A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J. Cybern. vol. 3, no. 3. (1974) 32–57
9. J. C. Bezdek.: Fuzzy Mathematics in pattern classification. Ph.D. dissertation. Dept. Appl. Math. Cornell Univ. Ithaca. NY. (1973)
10. J. M. Keller., M. R. Gray. and J. A. Givens Jr.: A fuzzy K-nearest neighbors algorithm. IEEE Trans. Syst. Man, Cybern., vol. SMC-15, no. 4, Apr. (1985) 580–585
11. Liu, X.D., Zhang, Y.J.: The fuzzy theory based on AFS structure and AFS algebra. Dlian.: Dlian Maritime University Press (1998)
12. Liu, X.D., Chai, T.Y. and Wang, W.: AFS Fuzzy Logic Systems and Its Applications to Model and Control. International Journal of Information and Systems Sciences. vol. 2, no. 3. (2006) 1-21
13. Liu, X.D., Chai, T.Y. and Wang, W.: Approaches to the representations and logic operations of fuzzy concepts in the framework of axiomatic fuzzy set theory I, II. Information Sciences. Revised (2005)
14. Liu, X.D., Wang, W., and Chai, T.Y.: The Fuzzy Clustering Analysis Based on AFS Theory. IEEE Trans. Syst. Man, Cybern.-partB: Cybernetics **35** (3) (2005) 1013-1027
15. 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, Oct. (2003) 1006–1011
16. Liu, X.D., Witold, P., Zhang, Q.L.: Axiomatics fuzzy sets logic. in Proc. IEEE Int. Conf. Fuzzy Systems **1** St. Louis, MO (2003) 55–60
17. Liu, X.D., Witold, P.: The Development of Fuzzy Decision Trees in the Framework of Axiomatic Fuzzy Set Logic. Applied Soft Computing, accepted. (2005). available online.
18. Liu, X.D., Zhang, L.S., Zhu, K.J. and Zhang, Q.L.: The Structures of EI Algebras Generated by Information Attributes. Int. J. Intelligent Systems Technologies and Applications, in press.
19. Liu, X.D., Liu, W.Q.: Credit Rating Analysis with AFS Fuzzy Logic. Lecture Notes in Computer Science, LNCS 3612. (2005) 1198-1204
20. Graver J E., Watkins M E.: Combinatorics with Emphasis on the Theory of Graphs. New York. Inc.: Springer-Verlag.(1977)
21. Gao, X.B.: Fuzzy Clustering Analysis and its Applications. Xian.: Xidian University Press. (in chinese). (2004)