Rough Set Based Uncertain Knowledge Expressing and Processing

Guoyin Wang

Institute of Computer Science and Technology
Chongqing University of Posts and Telecommunications,
Chongqing, 400065, China
wanggy@ieee.org

Abstract. Uncertainty exists almost everywhere. In the past decades, many studies about randomness and fuzziness were developed. Many theories and models for expressing and processing uncertain knowledge, such as probability & statistics, fuzzy set, rough set, interval analysis, cloud model, grey system, set pair analysis, extenics, etc., have been proposed. In this paper, these theories are discussed. Their key idea and basic notions are introduced and their difference and relationship are analyzed. Rough set theory, which expresses and processes uncertain knowledge with certain methods, is discussed in detail.

Keywords: uncertain knowledge expressing, uncertain knowledge processing, fuzzy set, rough set, cloud model.

1 Introduction of Uncertainty

The methods for uncertain knowledge expressing and processing have become one of the key problems of artificial intelligence. There are many kinds of uncertainties in knowledge, such as randomness, fuzziness, vagueness, incompleteness, inconsistency, etc. Randomness and fuzziness are the two most important and fundamental ones. Randomness implies a lack of predictability (causality). It is a concept of non-order or non-coherence in a sequence of symbols or steps, such that there is no intelligible pattern or combination. Fuzziness is the uncertainty caused by the boundary region, reflecting the loss of excluded middle law. There are many theories about randomness and fuzziness developed in the past decades. Many theories and models have been proposed, such as probability & statistics, fuzzy set [20], rough set [15], interval analysis [14], cloud model [13], grey system [6], set pair analysis [22], extenics [4], etc.

In this paper, we specifically discuss fuzzy set, rough set, type-2 fuzzy set, interval-valued fuzzy set, intuitionistic fuzzy set, cloud model, grey set, set pair analysis, interval analysis, and extenics. The key ideas and basic notions of these approaches are introduced and their differences and relationships are analyzed. Some further topics and problems related to expressing and processing uncertain knowledge based on rough set are discussed too.

S.O. Kuznetsov et al. (Eds.): RSFDGrC 2011, LNAI 6743, pp. 11–18, 2011.
© Springer-Verlag Berlin Heidelberg 2011

2 Set Theory

A set is a collection of distinct objects. Set is one of the most fundamental concepts in mathematics. The basic operators of set theory are: intersection ($A \cap B$), union ($A \cup B$), subtraction ($A - B$), and complement (A^c).

3 Fuzzy Set Theory

Fuzzy set, which was proposed by Zadeh as an extension of the classical notion of set [20], whose elements have degrees of membership. In classical set theory, the membership of elements in a set is assessed in binary terms according to a bivalent condition — an element either belongs or does not belong to the set, i.e., the membership function of elements in the set is one or zero. By contrast, fuzzy set theory permits the gradual assessment of the membership of elements in a set. The membership function is valued in the real unit interval $[0, 1]$. The membership of an element x belonging to a fuzzy set A is defined as $\mu_A(x)$. Quite typically, fuzzy set operators of intersection, union, and complement are defined as $\mu_{A\cap B}(x) = \min\{\mu_A(x), \mu_B(x)\}$, $\mu_{A\cup B}(x) = \max\{\mu_A(x), \mu_B(x)\}$, and $\mu_{A^c}(x) = 1 - \mu_A(x)$, respectively.

3.1 Type-2 Fuzzy Set

In 1975, Zadeh proposed a type-2 fuzzy set [21]. In 1999, Mendel argued that "words mean different things to different people", and claimed that we need type-2 fuzzy set to handle "ambiguity" in natural language [11]. Type-2 fuzzy set is a fuzzy set whose membership grades themselves is fuzzy set.

Definition 1 [11]. A **type-2 fuzzy set**, denoted $\tilde{A}$, is characterized by a type-2 membership $\mu_{\tilde{A}}(x, u)$, where for each $x \in U$ and $u \in J_x \subseteq [0, 1]$ there is $0 \leq \mu_{\tilde{A}}(x, u) \leq 1$. $\tilde{A}$ takes a form of $\{((x, u), \mu_{\tilde{A}}(x, u))\}$ or $\int_{x\in X} \int_{u\in J_x} \mu_{\tilde{A}}(x, u)/(x, u)$, where $\int\int$ denotes the union over all admissible x and u.

Let $\tilde{A} = \int_{x\in X} \int_{u\in J_x} \mu_{\tilde{A}}(x, u)/(x, u), \tilde{B} = \int_{x\in X} \int_{w\in J_x} \mu_{\tilde{B}}(x, w)/(x, w)$ be two type-2 fuzzy sets on U, where $u, w \in J_x$ and $\mu_{\tilde{A}}(x, u), \mu_{\tilde{B}}(x, w) \in [0, 1]$. The operations of union, intersection, and complement are defined as $\mu_{\tilde{A}\cup\tilde{B}}(x) = \int_u \int_w \frac{\mu_{\tilde{A}}(x,u)*\mu_{\tilde{B}}(x,u)}{u\vee w}$, $\mu_{\tilde{A}\cap\tilde{B}}(x) = \int_u \int_w \frac{\mu_{\tilde{A}}(x,u)*\mu_{\tilde{B}}(x,u)}{u\wedge w}$, and $\mu_{\tilde{A}^c}(x) = \int_u \frac{\mu_{\tilde{A}}(x,u)}{1-u}$, respectively, where "$*$" denotes a $t-$norm.

3.2 Interval-Valued Fuzzy Set

The interval-valued fuzzy set, which was proposed by Zadeh, is defined by an interval-valued membership function.

Definition 2 [21]. Let U be a universe. Define a map $A : U \to Int([0, 1])$, where $Int([0, 1])$ is the set of closed intervals in $[0, 1]$. Then, A is called an **interval-valued fuzzy set** on U and the membership function of A can be denoted by $A(x) = [A^-(x), A^+(x)]$.

Operations take form of $A \cup B(x) = [\sup(A^-(x), B^-(x)), \sup(A^+(x), B^+(x))]$, $A \cap B(x) = [\inf(A^-(x), B^-(x)), \inf(A^+(x), B^+(x))]$, and $A^c = [1 - A^+(x), 1 - A^-(x)]$, where $A^- = \inf(A), A^+ = \sup(A)$ for any $A \subset [0,1]$. Interval-valued fuzzy set is sometimes called grey set proposed by Deng [6].

Definition 3 [18]. Let G be a **grey set** of U defined by two mappings of the upper membership function $\bar{\mu}_G(x)$ and the lower membership function $\underline{\mu}_G(x)$ as follows: $\bar{\mu}_G(x) : U \to [0,1]$; $\underline{\mu}_G(x) : U \to [0,1]$, where $\underline{\mu}_G(x) \leq \bar{\mu}_G(x)$, $x \in U$. When $\underline{\mu}_G(x) = \bar{\mu}_G(x)$, the grey set G becomes a fuzzy set.

3.3 Intuitionistic Fuzzy Set

In fuzzy set theory, the membership of an element to a fuzzy set is a single value between zero and one. But in real life, it may not always be certain that the degree of non-membership of an element to a fuzzy set is just equal to 1 minus the degree of membership, i.e., there may be some hesitation degree. So, as a generalization of fuzzy set, the concept of intuitionistic fuzzy set was introduced by Atanassov [1]. Bustince and Burillo [3] showed that vague set defined by Gau and Buehrer [8] is equivalent to intuitionistic fuzzy set.

Definition 4 [1]. $A = \{\langle x, \mu_A(x), \nu_A(x)\rangle | x \in U\}$ is called an **intuitionistic fuzzy set**, where $\mu_A : U \to [0,1]$ and $\nu_A : U \to [0,1]$ are such that $0 \leq \mu_A + \nu_A \leq 1$, and $\mu_A, \nu_A \in [0,1]$ denote degrees of membership and non-membership of $x \in A$, respectively. For each intuitionistic fuzzy set A in U, "hesitation margin"(or "intuitionistic fuzzy index") of $x \in A$ is given by $\pi_A(x) = 1 - (\mu_A(x) + \nu_A(x))$ which expresses a hesitation degree of whether x belongs to A or not.

Operations take form of $A \cup B = \{\langle x, \max(\mu_A(x), \mu_B(x)), \min(\nu_A(x), \nu_B(x))\rangle |x \in U\}$, $A \cap B = \{\langle x, \min(\mu_A(x), \mu_B(x)), \max(\nu_A(x), \nu_B(x))\rangle |x \in U\}$ and $A^c = \{\langle x, \nu_A(x), \mu_A(x)\rangle | x \in U\}$.

There are plenty of theories treating imprecision and uncertainty. Some of them are extensions of fuzzy set theory, such as type-2 fuzzy set, interval-valued fuzzy set, intuitionistic fuzzy set, etc., while the others try to handle imprecision and uncertainty in a different way. Kerre [12] gave a summary of the links between fuzzy sets and other mathematical models such as flou set, two-fold fuzzy set and $L-$fuzzy set [9]. Deschrijver [7] proved:

1. There exists an isomorphism between $L-$intuitionistic fuzzy set [2] and $L-$fuzzy set. If L is the interval $[0,1]$ provided with the usual ordering, an $L-$intuitionistic fuzzy set is an intuitionistic fuzzy set;
2. There exists an isomorphism between interval-valued intuitionistic fuzzy set and $L-$fuzzy set for some specific lattice;
3. Intuitionistic fuzzy set can be embedded in interval-valued intuitionistic fuzzy set, so interval-valued intuitionistic fuzzy set theory extends intuitionistic fuzzy set theory;
4. There exists an isomorphism between interval-valued fuzzy set and intuitionistic fuzzy set, so interval-valued fuzzy set theory is equivalent to intuitionistic fuzzy set theory.

4 Rough Set Theory

Although fuzzy set can express the phenomenon that the elements in the boundary region belong to the set partially, it can not solve the “vague” problems that there are some elements which can not be classified into either a subset or its complement. For example: no mathematical formula to calculate the number of vague elements; no formal method to calculate the membership of vague elements. Rough set, which was proposed by Pawlak in 1982 [15], uses two certain sets, that is the lower approximation set and the upper approximation set, to define the boundary region of an uncertain set based on an equivalence relation (indiscernibility relation). The “vagueness degree” and the number of the vague elements can be calculated by the boundary region of a rough set.

The information of most natural phenomenon has the following characteristics: incomplete, inaccurate, vague or fuzzy. Classical set theory and mathematical logic can not express and deal with uncertainty problems successfully. The rough set theory is designed for expressing and processing vague information. The main advantage of rough set theory in data analysis is that it does not need any preliminary or additional information about data.

Rough set theory deals with uncertain problems using precise boundary lines to express the uncertainty. For an indiscernibility relation R and a set X, it operates with $R-$lower approximation of X, $R-$upper approximation of X, and $R-$boundary region of X, which are defined as $\underline{R}X = \{x \in U | [x]_R \subseteq X\}$, $\overline{R}X = \{x \in U | [x]_R \cap X \neq \emptyset\}$, and $RN_R(X) = \overline{R}X - \underline{R}X$, respectively.

If the boundary region of a set is empty, it means that the set is crisp, otherwise the set is **rough** (inexact). Nonempty boundary region means that our knowledge about the set is not sufficient to define it precisely.

The lower approximation of X contains all objects of U that can be classified into the class of X according to knowledge R. The upper approximation of X is the set of objects that can be and may be classified into the class of X. The boundary region of X is the set of objects that can possibly, but not certainly, be classified into class of X. Basic properties of rough set are as follows [15]:

1. $\overline{R}(X \cup Y) = \overline{R}(X) \cup \overline{R}(Y)$, $\underline{R}(X \cup Y) \supseteq \underline{R}(X) \cup \underline{R}(Y)$;
2. $\overline{R}(X \cap Y) \subseteq \overline{R}(X) \cap \overline{R}(Y)$, $\underline{R}(X \cap Y) = \underline{R}(X) \cap \underline{R}(Y)$;
3. $\overline{R}(\underline{X} - Y) \subseteq \overline{R}(X) - \underline{R}(X)$, $\underline{R}(\underline{X} - Y) = \underline{R}(X) - \overline{R}(Y)$;
4. $\sim \overline{R}(X) = \underline{R}(\sim X)$, $\sim \underline{R}(X) = \overline{R}(\sim X)$.

5 The Relationships of Fuzzy Set and Rough Set

Both fuzzy set and rough set are generalizations of the classical set theory for modeling vagueness and uncertainty. A fundamental question concerning both theories is their connections and differences [16]. It is generally accepted that they are related but distinct and complementary theories [5]. The two theories model different types of uncertainty:

1. Rough set theory takes into consideration the indiscernibility between objects. The indiscernibility is typically characterized by an equivalence relation. Rough set is the result of approximating crisp sets using equivalence

classes. The fuzzy set theory deals with the ill-definition of the boundary of a class through a continuous generalization of set characteristic functions. The indiscernibility between objects is not used in fuzzy set theory.

2. Rough set deals with uncertain problems using a certain method, while fuzzy set uses an uncertain method.
3. Fuzzy membership function relies on experts' prior knowledge. Rough set theory doesn't. For uncertainty of boundary regions, fuzzy set theory uses membership to express it, while rough set theory uses precise boundary lines to express it. Hence, fuzzy set theory and rough set theory could complement each other's advantages in dealing with uncertainties.

6 Cloud Model

Languages and words are powerful tools for human thinking, and the use of them is the fundamental difference between human intelligence and the other creatures' intelligence. We have to establish the relationship between the human brains and machines, which is performed by formalization. To describe uncertain knowledge by concepts is more natural and more generalized than to do it by mathematics. Li proposed a cloud model based on the traditional fuzzy set theory and probability statistics, which can realize the uncertain transformation between qualitative concepts and quantitative values.

Definition 5 [13]. Let U be the universe of discourse, C be a qualitative concept related to U. The membership μ of x to C is a random number with a stable tendency: $\mu : U \to [0,1]$, $\forall x \in U, x \to \mu(x)$, then the distribution of x on U is defined as a **cloud**, and every x is defined as a **cloud drop**. Qualitative concept is identified by three digital characteristics: Ex (Expected value), En (Entropy) and He (Hyper entropy).

Ex is the expectation of cloud drops' distribution in the universe of discourse, which means the most typical sample in the quantitative space of the concept. En is the uncertainty measurement of qualitative concept, decided by the randomness and the fuzziness of the concept. En reflects the numerical range which can be accepted by this concept in the universe of discourse, and embodies the uncertain margin of the qualitative concept. He is a measurement of entropy's uncertainty. It reflects the stability of the drops. The special numerical characteristic of cloud lies in using three values to sketch the whole cloud constituted by thousands of cloud drops, and it integrates the fuzziness and randomness of language value represented by quality method.

In practice, the normal cloud model is the most important kind of cloud models. It is based on normal distribution, and was proved universally to represent linguistic terms in various branches of natural and social science.

7 Set Pair Analysis

The set pair analysis theory, proposed by Zhao [22], is a novel uncertainty theory that is different from traditional probability theory and fuzzy set theory. Set pair is a pair of two related sets and set pair analysis is a method to process many kinds of uncertainties. The two sets have three relations: identical, different and contrary, and the connecting degree is an integrated description of them.

Definition 6 [22]. Assuming $H = (A, B)$ is a set pair of two sets A and B. For some application, H has total N attributes and S of them are mutual attributes of A and B, and P of them are contrary attributes, residual $F = N - S - P$ attributes are neither mutual nor opposite, then the connection degree of H is defined as: $\mu = \frac{S}{N} + \frac{F}{N}i + \frac{P}{N}j$, where S/N is identical degree, F/N is different degree, and P/N is contrary degree. Usually, we use a, b and c denote them, respectively, and $a + b + c = 1$.

8 Interval Analysis

Moore proposed an interval analysis theory, the purpose of which is to process error analysis automatically [14]. Interval analysis implements the storing and computing of data using interval, and the computing results ensure including all the possible true values.

Definition 7 [14]. A continuous subset $X = [\underline{x}, \bar{x}]$ on a real number domain R is called a **real interval**, and the upper and lower endpoints of an interval are represented by $\sup(X)$ and $\inf(X)$, respectively.

Let $X = [\underline{x}, \bar{x}]$, $Y = [\underline{y}, \bar{y}]$ be real intervals. The set of operations $\{+, -, *, \div\}$ is provided as follows [14]: $X + Y = [\underline{x} + \underline{y}, \bar{x} + \bar{y}]$, $X - Y = [\underline{x} - \bar{y}, \bar{x} - \underline{y}]$, $X * Y = [\min\{\underline{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\underline{y}, \bar{x}\bar{y}\}, \max\{\underline{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\underline{y}, \bar{x}\bar{y}\}]$, and $X \div Y = X * \frac{1}{Y}$, where $\frac{1}{Y} = \{\frac{1}{y} | y \in Y\}$ if $0 \notin Y$.

Interval methods can effectively define the function scope and provide strict operation results in mathematical meaning, which enable it appropriate to solve the problems of certain nonlinear equations and global optimization [10]. In addition, uncertainty of data can be expressed by interval. It is suitable, e.g., for solving nonlinear problems of parameter uncertainty in auto control [23].

9 Extension Set

The classical set and the fuzzy set mainly describe the "static" things. For the description of transformation of object A with character A_1 to object B with character B_1, Wen proposed the extension set to solve the qualitative description for "yes (true)" and "no (false)" to quantitative description and also to the variation procedures of "from yes to no" and "from no to yes" in 1983 [4]. It provides a suitable mathematical tool for solving the contradiction problems.

Definition 8 [19]. Let U be a domain and k be a reflection from U to the real domain R. Denote by T_u, T_k, and T_U the transformation of element, transformation of correlation function, and transformation of domain, respectively. For $T \in \{T_U, T_k, T_u\}$, $\tilde{A}(T) = \{(u, y, y') | u \in U, y = k(u) \in R, y' = T_k k(T_u u)\}$ is called an **extension set** on U about T. $y = k(u)$ and $y' = T_k k(T_u u)$ are called the **correlation function** and **extension function** of $\tilde{A}(T)$, respectively.

Let $\tilde{A}_1(T_1), \tilde{A}_2(T_2)$ be extension sets for $T_i \in \{T_U^i, T_k^i, T_u^i\}$ $(i = 1, 2)$. Denote "or" and "and" as $T_1 \vee T_2$ and $T_1 \wedge T_2$, respectively. We can consider the following operations [17]:

1. $\tilde{A}_1(T_1) \cup \tilde{A}_2(T_2) = \{(u, y, y')|u \in U, y = k(u), y' = T_k k(T_u u)\}$, where $T = T_1 \vee T_2$ and $k(u) = k_1(u) \vee k_2(u)$;
2. $\tilde{A}_1(T_1) \cap \tilde{A}_2(T_2) = \{(u, y, y')|u \in U, y = k(u), y' = T_k k(T_u u)\}$, where $T = T_1 \wedge T_2$, $k(u) = k_1(u) \wedge k_2(u)$;
3. $\tilde{A}_1^c(T_1) = \{(u, y, y')|u \in U, y = -y_1, y' = -y_1'\}$.

10 Future Directions and Topics of Rough Set Based Uncertain Knowledge Expressing and Processing

Rough set itself and the integration of rough set and other methods, including vague set, neural network, SVM, swarm intelligence, GA, expert system, etc., can deal with difficult problems like fault diagnosis, intelligent decision-making, image processing, huge data processing, intelligent control, and so on. At the same time, there are also new research directions to be studied in the future:

1. The extension of equivalence relation: order relation, tolerance relation, similarity relation, etc.;
2. Granular computing based on rough set theory (Dynamic Granular Computing);
3. The interactions among attributes (features): interactions among redundant attributes might be meaningful for problem expressing and solving;
4. The generalization of rough set reduction: reduction leads to over fitting (over training) in the training samples space;
5. Domain explanation of knowledge generated from reduction: The knowledge generated from data does not correspond to the human's formal knowledge;
6. Rough set characterize the ambiguity of decision information systems, but the randomness is not studied. Extended rough set model through combing rough set and cloud model?
7. 3DM (Domain-oriented Data-driven Data Mining): Knowledge generated should be kept the same as existed in the data sets; Reduce the dependence of prior domain knowledge in data mining processes;
8. Granular computing based on cloud model: granules (concepts) could be extracted from data using the backward cloud generator automatically.

Acknowledgments. This paper is supported by National Natural Science Foundation of P. R. China under grant 61073146, Natural Science Foundation Project of CQ CSTC under grant 2008BA2041.

References

1. Atanassov, K.T.: Intuitionistic fuzzy sets. Fuzzy Sets and Systems 20, 87–96 (1986)
2. Atanassov, K.T.: Intuitionistic fuzzy sets. Physica-Verlag, Heidelberg (1999)
3. Bustince, H., Burillo, P.: Vague Sets are intuitionistic fuzzy sets. Fuzzy Sets and Systems 79, 403–405 (1996)
4. Cai, W.: The extension set and non-compatible problems. Journal of Science Explore (1), 83–97 (1983)
5. Chanas, S., Kuchta, D.: Further remarks on the relation between rough and fuzzy sets. Fuzzy Sets and Systems 47, 391–394 (1992)
6. Deng, J.L.: Grey systems. China Ocean Press, Beijing (1988)
7. Deschrijver, G., Kerre, E.E.: On the relationship between some extensions of fuzzy set theory. Fuzzy Sets and Systems 133, 227–235 (2003)
8. Gau, W.L., Buehrer, D.J.: Vague sets. IEEE Transaction on Systems Man Cybernetics 23(2), 610–614 (1993)
9. Goguen, J.A.: $L-$fuzzy sets. Journal of mathematical analysis and applications 18, 145–174 (1967)
10. Hansen, E.R.: Global optimization using interval analysis. Marcel Dekker, New York (1992)
11. Karnik, N.N., Mendel, J.M., Liang, Q.L.: Type-2 fuzzy logic systems. IEEE Transaction on Fuzzy Systems 7(6), 643–658 (1999)
12. Kerre, E.E.: A first view on the alternatives of fuzzy set theory. Computational Intelligence in Theory and Practice, 55–72 (2001)
13. Li, D.Y., Meng, H.J., Shi, X.M.: Membership clouds and cloud generators. Journal of Computer Research and Development 32, 32–41 (1995)
14. Moore, R.E.: Interval analysis, pp. 25–39. Prentice-Hall, Englewood Cliffs (1966)
15. Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences 5(11), 341–356 (1982)
16. Pawlak, Z.: Rough sets and fuzzy sets. Fuzzy Sets and Systems 17, 99–102 (1985)
17. Sun, H.: On operations of the extension set. Mathematics in Practice and Theory 37(11), 180–184 (2007)
18. Wu, Q., Liu, Z.T.: Real formal concept analysis based on grey-rough set theory. Knowledge-based Systems 22, 38–45 (2009)
19. Yang, C.Y., Cai, W.: New definition of extension set. ournal of Guangdong University of Technology 18(1), 59–60 (2001)
20. Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)
21. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning-I. Information Sciences 8, 199–249 (1975)
22. Zhao, K.Q.: Set pair analysis and its primary application. Zhejiang Science and Technology Press, Hangzhou (2000)
23. Zettler, M., Garloff, J.: Robustness analysis of polynomials with polynomials parameter dependency using Bernstein expansion. IEEE Trans. on Automatic Control 43(3), 425–431 (1998)