KALYAN MONDAL[1], SURAPATI PRAMANIK[2], BIBHAS C. GIRI[3]

[1]Department of Mathematics, Jadavpur University, West Bengal, India. E-mail: kalyanmathematic@gmail.com
[2]Department of Mathematics, Nandalal Ghosh B.T. College, Panpur, PO-Narayanpur, and District: North 24 Parganas, Pin Code: 743126, West Bengal, India. E-mail: sura_pati@yahoo.co.in
[3]Department of Mathematics, Jadavpur University, West Bengal, India. E-mail: bcgiri.jumath@gmail.com

Role of Neutrosophic Logic in Data Mining

Abstract

This paper presents a data mining process of single valued neutrosophic information. This approach gives a presentation of data analysis common to all applications. Data mining depends on two main elements, namely the concept of similarity and the machine learning framework. It describes a lot of real world applications for the domains namely mathematical, medical, educational, chemical, multimedia etc. There are two main types of indeterminacy in supervised learning: cognitive and statistical. Statistical indeterminacy deals with the random behavior of nature. All existing data mining techniques can handle the uncertainty that arises (or is assumed to arise) in the natural world from statistical variations or randomness. Cognitive uncertainty deals with human cognition. In real world problems for data mining, indeterminacy components may arise. Neutrosophic logic can handle this situation. In this paper, we have shown the role of single valued neutrosophic set logic in data mining. We also propose a data mining approach in single valued neutrosophic environment.

Keywords

Data mining, single valued neutrosophic set, single valued neutrosophic score value.

1. Introduction

Data mining [1] is actually assumed as “knowledge mining” from data. Data mining is an essential process where intelligent methods are applied to extract data patterns [2]. Data mining is a process that analyzes large amounts of data to find new and hidden information. In other words; it is the process of analyzing data from different perspectives and summarizing it into some useful information. The following are the different data mining techniques [3]: association, classification, clustering, and sequential patterns. E.Hullermeier [4] proposed fuzzy methods in data mining.

This paper focuses on real-world applications of single valued neutrosophic set [5] for data mining. Data mining decomposes into two main elements: the notion of similarity and the single valued neutrosophic machine learning techniques that are applied in the described applications. Indeed, similarity, or more generally comparison measures are used at all levels of the data mining and information retrieval tasks. At the lowest level, they are used for the matching between a query to a database and the elements it contains, for the extraction of relevant data. Then similarity and dissimilarity measures can be used in the process of cleaning and management of missing data to create a reasonable set of data. To generalize particular information contained in this reasonable set, dissimilarity measures are used in the case of inductive learning and similarity measures for case-based reasoning or clustering tasks. Eventually, similarities are used to interpret results of the learning process into an expressible form of knowledge through the definition of prototypes. Most of collective data for an investigation involves indeterminacy. Single valued neutrosophic set can handle his situation. So, there is an important role of single valued neutrosophic set in data mining.

This paper is arranged as follows. Section 2 presents some basic knowledge of single valued neutrosophic set. Section 3 considers the component of similarity, and machine learning techniques. Section 4 describes a methodical approach of data mining under single valued neutrosophic environment. Section 5 presents a numerical example for data mining. Section 6 presents concluding remarks.

2. Neutrosophic Preliminaries

2.1 Definition on neutrosophic sets [6]

The concept of neutrosophic set is originated from neutrosophy [6], a new branch of philosophy.

Definition 1:[6] Let ξ be a space of points (objects) with generic element in ξ denoted by x. Then a neutrosophic set α in ξ is characterized by a truth membership function T_α an indeterminacy membership function I_α and a falsity membership function F_α. The functions T_α and F_α are real standard or non-standard subsets of $\left]^-0, 1^+\right[$ that is T_α: $\xi \to \left]^-0, 1^+\right[$; I_α: $\xi \to \left]^-0, 1^+\right[$; F_α: $\xi \to \left]^-0, 1^+\right[$.

It should be noted that there is no restriction on the sum of $T_\alpha(x)$, $I_\alpha(x)$, $F_\alpha(x)$ i.e. $^-0 \le T_\alpha(x) + I_\alpha(x) + F_\alpha(x) \le 3^+$

Definition 2: [6] The complement of a single valued neutrosophic set α is denoted by α^c and is defined by

$$T_{\alpha^c}(x) = \{1^+\} - T_\alpha(x);\ I_{\alpha^c}(x) = \{1^+\} - I_\alpha(x)$$

$$F_{\alpha^c}(x) = \{1^+\} - F_\alpha(x)$$

Definition 3: (Containment) [6] A single valued neutrosophic set α is contained in the other single valued neutrosophic set β, $\alpha \subseteq \beta$ if and only if the following result holds.

$$\inf T_\alpha(x) \le \inf T_\beta(x),\ \sup T_\alpha(x) \le \sup T_\beta(x)$$

$$\inf I_\alpha(x) \ge \inf I_\beta(x),\ \sup I_\alpha(x) \ge \sup I_\beta(x)$$

$$\inf F_\alpha(x) \ge \inf F_\beta(x),\ \sup F_\alpha(x) \ge \sup F_\beta(x)$$

for all x in ξ.

Definition 4: (Single-valued single valued neutrosophic set)[5] .

Let ξ be a universal space of points (objects) with a generic element of ξ denoted by x.

A single-valued single valued neutrosophic set S is characterized by a true membership function $T_s(x)$, an indeterminacy membership function $I_s(x)$, a falsity membership function $F_s(x)$ with $T_s(x)$, $I_s(x)$, $F_s(x) \in [0, 1]$ for all x in ξ. When ξ is continuous a SNVS can be written as

$$S = \int_x \langle T_s(x), F_s(x), I_s(x) \rangle / x, \forall x \in \xi$$

and when ξ is discrete a SVNSs S can be written as:

$$S = \sum \langle T_S(x), F_S(x), I_S(x) \rangle / x, \forall x \in \xi$$

It should be noted that for a SVNS S,

$$0 \le \sup T_S(x) + \sup F_S(x) + \sup I_S(x) \le 3, \forall x \in \xi$$

and for a single valued neutrosophic set, the following relation holds:

$$0 \le \sup T_S(x) + \sup F_S(x) + \sup I_S(x) \le 3, \forall x \in \xi$$

Definition 5: The complement of a single valued neutrosophic set S is denoted by S^c and is defined by

$$T_S{}^c(x) = F_S(x);\ I_S{}^c(x) = 1 - I_S(x);\ F_S{}^c(x) = T_S(x)$$

Definition 6: A SVNS S_α is contained in the other SVNS S_β, denoted as $S_\alpha \subseteq S_\beta$ iff, $T_{S_\alpha}(x) \le T_{S_\beta}(x)$; $I_{S_\alpha}(x) \ge I_{S_\beta}(x)$; $F_{S_\alpha}(x) \ge F_{S_\beta}(x)$, $\forall x \in \xi$.

Definition 7: Two single valued single valued neutrosophic sets S_α and S_β are equal, i.e. $S_\alpha = S_\beta$, if and only if $S_\alpha \subseteq S_\beta$ and $S_\alpha \supseteq S_\beta$

Definition 8: (Union) The union of two SVNSs S_α and S_β is a SVNS S_γ, written as $S_\gamma = S_\alpha \cup S_\beta$.

Its truth membership, indeterminacy-membership and falsity membership functions are related to those of S_α and S_β by

$$T_{S_\gamma}(x) = \max\left(T_{S_\alpha}(x), T_{S_\beta}(x)\right);$$

$$I_{S_\gamma}(x) = \max\left(I_{S_\alpha}(x), I_{S_\beta}(x)\right);$$

$F_{S_\gamma}(x) = \min\left(F_{S_\alpha}(x), F_{S_\beta}(x)\right)$ for all x in ξ

Definition 9: (intersection) The intersection of two SVNSs, S_α and S_β is a SVNS S_δ, written as $S_\delta = S_\alpha \cap S_\beta$. Its truth membership, indeterminacy-membership and falsity membership functions are related to those of S_α an S_β as follows:

$$T_{S_\delta}(x) = \min\left(T_{S_\alpha}(x), T_{S_\beta}(x)\right);$$

$$I_{S_\delta}(x) = \max\left(I_{S_\alpha}(x), I_{S_\beta}(x)\right);$$

$$F_{S_\delta}(x) = \max\left(F_{S_\alpha}(x), F_{S_\beta}(x)\right), \forall x \in \xi$$

3. Data Mining [2]

In this section, we discuss the theoretical background common to the applications, considering successively the notion of similarity and machine learning techniques under single valued neutrosophic environment.

3.1 Similarity [2]

The notion of similarity or more generally of comparison measures, is central for all real-world applications: it aims at quantifying the extent to which two objects are similar, or dissimilar, one to another, providing a numerical value for this comparison. Similarities and dissimilarities between objects are generally evaluated from values of their attributes or variables characterizing these objects. Dissimilarities are classically defined from distances. Similarities and dissimilarities are often expressed from each other: the more similar two objects are, the less dissimilar they are, the smaller their distance. Weights can be associated with variables, according to the semantics of the application or the importance of the variables. It appears that some quantities are used in various environments, with different forms, based on the same principles. Most of the classic dissimilarity measures between two objects with continuous numerical attributes are the Euclidian distance, the Manhattan distance, and more generally Minkowski distances.

3.2 Neutrosophy Machine Learning [2]

The second part of the theoretical background common to all applications concerns the neutrosophy machine learning techniques that use the previous similarity measures. Machine learning is an important way to extract knowledge from sets of cases, especially in large scale databases. In this section, we consider only the neutrosophy machine learning methods (involving indeterminacy) that are used in the applications, leaving aside other techniques as for neutrosophy case-based reasoning or neutrosophy association rules. Three methods are successively considered: neutrosophy decision trees, neutrosophy prototypes and neutrosophy clustering. The first two belong to the supervised learning framework, i.e. they consider that each data point is associated with a category. Single valued neutrosophic set clustering belongs to the unsupervised learning framework, i.e. no decomposition of the data set with indeterminacy into categories is available.

3.2.1. Single valued neutrosophic set Decision Trees [2]

Neutrosophy decision trees (NDT) particularly can be interesting for data mining and information retrieval because they enable the user to take into account indeterminacy descriptions of the cases, or heterogeneous values (symbolic, numerical, or neutrosophical) [5]. Moreover, they are appreciated for their interpretability, because they provide a linguistic description of the relations between descriptions of the cases and decision to make or class to assign. The rules obtained through NDT make it easier for the user to interact with the system or the expert to understand, confirm or amend his own knowledge. Another quality of NDT is their robustness, since a small variation of descriptions does not drastically change the decision or the class associated with a case, which guarantees a resistance to measurement errors and avoids sharp differences for close values of the descriptions.

3.2.2. Single valued neutrosophic set Prototype Construction [2]

Neutrosophy prototypes are another approaches to the characterization of data categories: they provide descriptions or interpretable summarizations of data sets, so as to help a user to better apprehend their contents: a prototype is an element chosen to represent a group of data, to summarize it and underline its most characteristic features. It can be defined from a statistical point of view, for instance as the data mean or the median; more complex representatives can also be used as the most typical value [7] for instance. The prototype notion was also studied from a cognitive science point of view, and specific properties were pointed out in [8]: it was shown that a prototype underlines the common features of the category members, but also their distinctive features as opposed to other categories, underlining the specificity of the group. Furthermore, prototypes were related to the typicality notion, i.e. the fact that all data do not have the same status as regards the group: some members of the group are better examples, more representative or more characteristic than others. It was also shown that the typicality of a point depends both on its resemblance to other members of the group (internal resemblance), and on its dissimilarity to members of other groups (external dissimilarity). More precisely, the method consists of computing internal resemblance and external dissimilarity for each data point. Internal resemblance and external dissimilarity are respectively defined as the aggregation (mean or median) of the resemblance to the other members of the group, and as the aggregation of the dissimilarity to members of other groups, for a given choice of the resemblance and dissimilarity measures.

4. Single Valued Neutrosophic Logic in Data Mining

The tools that have been proposed in single valued neutrosophic set (SVNS) have the potential to support all of the steps that neutralized a process of knowledge discovery. SVNS can be used in the data selection and preparation phase for data modeling. For any data analysis associated with an experiment or investigation, it is observed that much information involve indeterminacy. Single valued neutrosophic set logic is capable of dealing with this situation. So, for the case of data mining single valued neutrosophic set logic has an important role.

Standard methods of data analysis can be extended in a rather generic way by means of an extension principle. For example, the functional relation between the data points and the decision making function can be extended to the case of single valued neutrosophic data, where the observations are described in terms of single valued neutrosophic sets. If single valued neutrosophic data is not used in the data preparation phase, they can still be employed in a later stage in order to analyze the original data.

Various techniques are widely used for data mining from gathering data within a domain of expertise. Delphi method [9] and BIRCH method [10] are very popular for data mining. Rekha and Swapna [2] studied the role of fuzzy logic in data mining. Literature review reflects that there is no single valued neutrosophic approach for data mining till now.

4.1. Neutrosophic data mining method

Generally, there are many attributes in decision making problems, where some of them are important and others may not be so important. So it is crucial to select the proper attributes for decision-making situation. Now, we shall propose a methodical approach for data mining with single valued neutrosophic information to prepare a panel of attributes which are technically sound. All steps of this proposed approach are given as follows.

Step 1: Problem field selection

Consider a multi-attribute decision making problem with m alternatives and n attributes (large numbers of data). Let A_1, A_2, ..., A_m and C_1, C_2, ..., C_n denote the alternatives and attributes respectively. In decision making process, we have to select a finite but more important attributes from given n attributes. All attributes are expressed in single valued neutrosophic number.

Table 1: *Single valued neutrosophic set decision matrix*

$$D = \langle d_{ij} \rangle_{m\times n} =$$

$$\begin{array}{c|cccc} & C_1 & C_2 & \cdots & C_n \\ \hline A_1 & \langle d_{11}\rangle & \langle d_{12}\rangle & \cdots & \langle d_{1n}\rangle \\ A_2 & \langle d_{21}\rangle & \langle d_{22}\rangle & \cdots & \langle d_{2n}\rangle \\ . & \cdots & \cdots & \cdots & \cdots \\ . & \cdots & \cdots & \cdots & \cdots \\ A_m & \langle d_{m1}\rangle & \langle d_{m2}\rangle & \cdots & \langle d_{mn}\rangle \end{array} \quad (1)$$

Here, d_{ij} (i = 1, 2, …, m and j = 1, 2, …, n) are all single valued neutrosophic numbers.

Step 2: Single valued neutrosophic set score matrix

Definition 10: Single valued neutrosophic score function (SVNSF)

Single valued neutrosophic score function (*SVNSF*) corresponding to each attribute is defined as follows.

$$SVNSF(C_j) = \frac{1}{m}\sum_{r=1}^{m} \frac{2+T_{rj}-I_{rj}-F_{rj}}{3} \quad (2)$$

Where, j = 1, 2, …, n

Using equation (2) we calculate single valued neutrosophic score matrix as follows.

Table: *Single valued neutrosophic score matrix*

$$SVNSF(C_j) = \begin{array}{c|c} attributes & Single\ valued\ neutrosophic\ score\ value \\ \hline C_1 & SVNSF(C_1) \\ C_2 & SVNSF(C_2) \\ \vdots & \vdots \\ C_n & SVNSF(C_n) \end{array} \quad (3)$$

Step 3: Selection zone

Single valued neutrosophic score values are classified into three zones. These are described as follows.

Definition 11: *SVNSF* of all the attributes are classified in three categories and it is defined as follows

Highly acceptable zone: $0.50 \le SVNSF(C_j) \le 1$

Tolerable acceptable zone: $0.25 \leq SVNSF(C_j) \leq 0.50$

Unacceptable acceptable zone: $0.00 \leq SVNSF(C_j) \leq 0.25$

Step 4: Ranking of attributes

According to the single valued neutrosophic score values, we can set up a panel of all attributes in descending order and we can choose important attributes from large number of attributes into decision making process considering highly acceptable zone and tolerable acceptable zone

Step 5: End

5. Numerical Example

In this section we demonstrate a numerical problem for applicability and effectiveness of this proposed approach. The methodical steps are as follows.

Step 1: Problem field selection

Suppose a person who wants to purchase a SIM card for mobile connection. So, it is necessary to select suitable SIM card for his/her mobile connection. There is a panel with four possible alternatives (SIM cards) for mobile connection. The alternatives (SIM cards) are presented as follows:

A_1: Airtel

A_2: Vodafone

A_3: BSNL

A_4: IDEA.

For this purpose, the following attributes about SIM cards may be arise in decision making process. These are stated as follows.

1. Service quality of the corresponding company (C_1)
2. Cost (C_2)
3. Call rate per second (C_3)
4. Internet facilities (C_4)
5. Tower facility (C_5)
6. Call drops (C_6)
7. Risk factor (C_7)

Table3: Single valued neutrosophic decision matrix

$$D = \langle d_{ij} \rangle_{4\times 7} =$$

$$\begin{array}{c|ccccccc} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 \\ \hline A_1 & \langle 0.8,0.3,0.2\rangle & \langle 0.7,0.3,0.2\rangle & \langle 0.7,0.2,0.4\rangle & \langle 0.8,0.1,0.2\rangle & \langle 0.3,0.5,0.5\rangle & \langle 0.2,0.3,0.6\rangle & \langle 0.1,0.6,0.5\rangle \\ A_2 & \langle 0.8,0.3,0.3\rangle & \langle 0.7,0.1,0.2\rangle & \langle 0.7,0.3,0.4\rangle & \langle 0.8,0.1,0.1\rangle & \langle 0.3,0.4,0.5\rangle & \langle 0.2,0.5,0.6\rangle & \langle 0.1,0.4,0.5\rangle \\ A_3 & \langle 0.8,0.2,0.2\rangle & \langle 0.7,0.3,0.1\rangle & \langle 0.7,0.4,0.4\rangle & \langle 0.8,0.2,0.2\rangle & \langle 0.3,0.5,0.6\rangle & \langle 0.2,0.1,0.5\rangle & \langle 0.1,0.6,0.3\rangle \\ A_4 & \langle 0.8,0.1,0.2\rangle & \langle 0.7,0.0,0.2\rangle & \langle 0.7,0.1,0.4\rangle & \langle 0.8,0.3,0.2\rangle & \langle 0.3,0.3,0.3\rangle & \langle 0.2,0.3,0.3\rangle & \langle 0.1,0.5,0.4\rangle \end{array} \quad (4)$$

Step 2: Single valued neutrosophic score matrix

Using equation (2) we calculate single valued neutrosophic score matrix as follows.

Table 4: Single valued neutrosophic set score matrix

$$SVNSF(C_j) = \begin{array}{c|c} attributes & Single\ valued\ neutrosophic\ score\ value \\ \hline C_1 & 0.7833 \\ C_2 & 0.7833 \\ C_3 & 0.6833 \\ C_4 & 0.8167 \\ C_5 & 0.4667 \\ C_6 & 0.4667 \\ C_7 & 0.3833 \end{array} \quad (5)$$

Step 3: Selection zone

Single valued neutrosophic score values are classified into three zones. These are described as follows.

Definition 11: *SVNSF* of all the attributes are classified in three categories and it is defined as follows

Highly acceptable zone: $0.50 \le SVNSF(C_j) \le 1$

Tolerable acceptable zone: $0.25 \le SVNSF(C_j) \le 0.50$

Unacceptable acceptable zone: $0.00 \le SVNSF(C_j) \le 0.25$

Step 4: Ranking of attributes

From equation (5) we can write single valued neutrosophic score values of all attributes in descending order as follows.

$$SVNSF(C_4) \succ SVNSF(C_1) \succ SVNSF(C_2) \succ SVNSF(C_3) \succ SVNSF(C_5) \succ SVNSF(C_6) \succ SVNSF(C_7)$$

So, attributes corresponding to single values neutrosophic score values (highly acceptable and tolerance zone) can be chosen as important attributes for decision making process.

Step 5: End

6. Conclusion

In this paper we briefly present first two of the essential pillars of data mining: similarity measures and machine learning in single valued neutrosophic environment. We showed that neutrosophic logic can perform an important role in data mining method. We define single valued neutrosophic score function (*SVNSF*) to aggregate attribute values of each alternative. We also propose an approach for data mining with single valued neutrosophic information from large amounts of data and furnish a numerical example for the proposed approach. In future this method can be extended in interval neutrosophic environment for data mining.

References

1. Jiawei Han, Micheline Kamber, Jian Pei, Data mining concepts and techniques, 2012
2. M. Rekha and M. Swapna. Role of fuzzy logic in data Mining, International Journal of Advance Research in Computer Science and Management Studies, 2(12), 2014.
3. E.Hullermeier Fuzzy methods in data mining: status and prospects. Fuzzy Sets and Systems, 156(3), 387-406, 2005.
4. C.Z Janikow. Fuzzy decision trees: issues and methods IEEE transactions on systems Man and Cybernetics, 28(1), 1-14, 1998.
5. H. Wang, F. Smarandache, Y. Q. Zhang and R. Sunderraman. Single valued neutrosophic sets, Multispace and Multistructure, 4(2010), 410-413.
6. F. Smarandache. A unifying field in logics. neutrosophy: neutrosophic probability, set and logic. Rehoboth: American Research Press, 1998.
7. M.Friedman and M.Ming and A.Kandel. On the theory of typicality, International journal of Uncertainity, Fuzzyness and Knowledge-Based Systems, 3(2): 127-142, 1995.
8. E.Rosch and C.Mervis .Family Resemblance: Studies of internal structure of categories, Cognitive psychology, 7, 573-605, 1975.
9. C. C Hsu and B A. Sandford. The Delphi technique: making sense of consensus, practical assessment, Research and Evaluation, 12(10), 2007.
10. Y. Rani, Manju and H. Rohil. Comparative Analysis of BIRCH and CURE Hierarchical Clustering Algorithm using WEKA 3.6.9, The SIJ Transactions on Computer Science Engineering & its Applications (CSEA), 2 (1), 2014.