Neutrosophic Sets and Systems
{Special Issue: Impact of Neutrosophy in solving the Latin American's social problems}, Vol. 37, 2020
University of New Mexico

Neutrosophic K-means Based Method for Handling Unlabeled Data

Ned Vito Quevedo Arnaiz[1], Nemis Garcia Arias[2], and Leny Cecilia Campaña Muñoz[3]

[1] Docente de la carrera de Derecho, Universidad Regional Autónoma de los Andes (UNIANDES), Avenida La Lorena, CP 230150, Santo Domingo de los Tsáchilas, Ecuador, Ecuador. E-mail: us.nedquevedo@uniandes.edu.ec
[2] Docente de la carrera de Derecho, Universidad Regional Autónoma de los Andes (UNIANDES), Avenida La Lorena, CP 230150, Santo Domingo de los Tsáchilas, Ecuador, Ecuador. E-mail: us.nemisgarcia@uniandes.edu.ec
[3] Docente de la carrera de Derecho, Universidad Regional Autónoma de los Andes (UNIANDES), Avenida La Lorena, CP 230150, Santo Domingo de los Tsáchilas, Ecuador, Ecuador. E-mail: us.lenycampana@uniandes.edu.ec

Abstract. Nowadays, incalculable volumes of data are generated due to the technological development achieved by the current society of information. The exponential growth of information significantly supports people's decision making in their daily activities. In Ecuador, there are many institutions that store the data of their processes. The tourism sector represents an example of this assertion. However, the data generated exceeds the power of analysis and processing of human beings, sometimes relevant information is presented it is not visible for persons. The present investigation proposes a solution to the described problem starting from the development of a method for the treatment of unlabeled data. The proposed method is based on the unsupervised k-means algorithm. We used rough neutrosophic sets to reduce the number of attributes. The proposal has been implemented from the stored dataset of the tourism sector in the City of Riobamba.

Keywords: Machine learning; data mining; rough neutrosophic sets; entropy

1. Introduction

Tourism represents an important source of income in Ecuador's internal economy. Each region of the country has attractions that make it unique as a tourist destination. The city of Riobamba in Ecuador is characterized by representing a very attractive tourist area, it is a city with great cultural heritage that attracts even the most exquisite vacationers [1].

Tourism management itself generates high demands for products and services that include a wide range of different activities such as: transportation, accommodation, supply, shopping, travel agency services, inbound and outbound tourism operators, among others [2, 3]. Without doubt, tourism represents a fundamental source of income and it generates a large amount of data [4].

From the different operations that are carried out in the City of Riobamba, there are stored historical data of the different operations in tourism management. However, the existing data is not properly labeled, which makes it impossible to obtain objective information that contributes to decision-making for the tourism sector [5].

Problems of this nature have been addressed in the scientific literature from data mining techniques for the cleaning, transformation and treatment of unlabeled data [6-8]. This investigation has the objective to develop a method based on k-means for the treatment of unlabeled data.

We use three main tools for solving this problem. One is the combination of rough sets with single-valued neutrosophic sets, to deal with the so-called rough neutrosophic sets, which extends the existing rough fuzzy sets, but including the indeterminacy. The other two techniques are k-means and entropy. The hybridization of these methods creates a more accurate result.

This paper is divided into the following sections: Section 2 contains the preliminary concepts; section 3 exposes the materials and methods that will be used in this paper. Section 4 summarizes the results applied to an actual case study related with a database of touristic industry in the city of Riobamba in Ecuador. The paper ends with the conclusions.

2. Preliminaries

This section introduces an approximation of the main theoretical references that support the research proposal. It begins with a characterization of machine learning. The fundamental elements on the rough sets and some criteria

for comparing k-means algorithms are presented. The section continues with the significant elements associated with entropy and information. Finally, the used k-means algorithm is described.

2.1 Machine learning

Machine learning introduces a new paradigm that refers to the study of computational algorithms that automatically incorporate experiences to improve its operation [9, 10]. Machine learning systems simulate the processes that humans perform when they are executing a task.

A machine learning process needs to train a model by applying learning techniques. For the training process, the data that the machine will use to learn this procedure are provided [11-13]. This type of learning has been used in data mining applications with the aim of discovering rules and patterns in large data sets and filtering information [14, 15].

The classification of machine learning techniques can be divided into:

- Supervised or predictive learning: where the objective is to learn to map from X inputs to Y outputs, given a labeled set of N input-output pairs; this set is called Training set.
- Unsupervised or descriptive learning: aims to find interesting patterns in the N entries.
- Reinforcement of learning: it is used to know how it acts or behaves when certain occasional signs of reward or punishment are given.

2.2 Rough sets

Rough sets (RS) are based on the assumption that each object x in the universe of discourse U has associated certain information that represents data and knowledge [16]. It is expressed through attributes that describe the object. Among the advantages of RS for data analysis are [17-19]:

- It is based on the original data and does not require external information, so there is no need to make any assumptions about the data.
- It allows the analysis of qualitative and quantitative traits.

Then a rough set is formalized as follows:

Definition 1. Let U be a finite universe. Let R be an equivalence relation defined in U, which partitions U. (U, R) is a collection of all equivalence classes, called the *approximation space*. Let $w_1, w_2 w_3, \cdots, w_n$ elements of the approximation space (U, R). This connection is known as the *knowledge base* [20]. Then, for any subset B of U, the *upper approximation* $\overline{B}$ and the *lower approximation* $\underline{B}$ are defined as [21, 22]:

The ordered pair $(\overline{B}, \underline{B})$ is called an *approximate set* and its elements are defined as follows:

The *lower approximation* of the set X with respect to R is defined in Equation 1:

$$\underline{B}(X) = \cup_{x \in U}\{R(X): R(X) \subseteq X\} \quad (1)$$

The *upper approximation* of the set X with respect to R is defined in the following form:

$$\overline{B}(X) = \cup_{x \in U}\{R(X): R(X) \cap X \neq \emptyset\} \quad (2)$$

And it also has:

$POS(B) = \underline{B}$: It is certainly a member of X.

$NEG(B) = U \setminus \overline{B}$: It is certainly not a member of X.

$BR(B) = \overline{B} \setminus \underline{B}$: It is possibly a member of X.

Where:

$POS(B)$ refers to the positive region of B,

$NEG(B)$ refers to the negative region of B,

$BR(B)$ refers to the border region of B.

An *approximate set* can be defined by neutrosophic numbers. Neutrosophic logic is a general framework for unifying many existing logics. Generalize fuzzy logic (especially fuzzy intuitionist logic). In 1995, Florentin Smarandache extended Paradoxism (based on opposites) to a new branch of philosophy called Neutrosophy (based on opposites and their neutral), that gave birth to many areas of application [23, 24]. The important idea of NL is to characterize each logical statement in a neutrosophic 3 dimensional $[0,1]^3$-space, where each dimension of the space represents the truth (T) respectively, the falsehood (F), and indeterminacy (I) of the proposition, where T, I , F are standard or not standard real subset of]-0, 1+[[25].

The classic interval unit [0,1] can be used. T, I, F are independent components leaving room for incomplete information (when they sum up <1), for consequent and contradictory information (when they sum up > 1) or complete information (sum of components = 1)[25-27].

Definition 2. *Neutrosophic sets* are a fuzzy set generalization (spatially fuzzy intuitive set). Let U, be a universe of discourse, and M a set included in U. An element x of U is denoted with respect to the set M as x(T, I, F) and it belongs to M in the following way: It is t% true, it is i% indeterminate, and it is f% false, where t varies in T, i varies in I, and f varies in F[28, 29].

Statistically T, I, F are subsets of [0, 1], but dynamically T, I, F are functions or operations dependent on many unknown or known parameters [26].

In order to facilitate practical application to decision-making and engineering problems, the proposal was made for single-value valued neutrosophic sets [16, 30, 31], which allow us to use linguistic variables [32, 33] increasing the interpretability in the recommendation models and the use of indeterminacy.

Definition 3. Let X be a universe of discourse. A Single-valued neutrosophic set (SVNS) A over X is an object of the form.

$$A = \{\langle x, u_A(x), r_A(x), v_A(x)\rangle : x \in X\} \quad (3)$$

Where $u_A(x): X \rightarrow [0,1]$, $r_A(x), : X \rightarrow [0,1]$ and $v_A(x): X \rightarrow [0,1]$ with $0 \leq u_A(x) + r_A(x) + v_A(x) :\leq 3$ for all $x \in X$. The intervals $u_A(x), r_A(x)$ and $v_A(x)$ denote the memberships to true, indeterminate, and false of x in A, respectively. For reasons of convenience, a single-valued neutrosophic number will be expressed as $A = (a, b, c)$, where $a, b, c \in [0,1]$, and $0 \leq a + b + c \leq 3$.

Definition 4. [34] Let U be a non-null set and R be an equivalence relation on U. Let F be neutrosophic set in U with the membership function μ_F, indeterminacy function ν_F and non-membership function ω_F. The *lower* and the *upper approximations* of F in the approximation (U, R) denoted by $\underline{N}(F)$ and $\overline{N}(F)$ are respectively defined as follows: $\underline{N}(F) = \{< x, \mu_{\underline{N}(F)}(x), \nu_{\underline{N}(F)}(x), \omega_{\underline{N}(F)}(x) > | y \in [x]_R, x \in U\}$,

$\overline{N}(F) = \{< x, \mu_{\overline{N}(F)}(x), \nu_{\overline{N}(F)}(x), \omega_{\overline{N}(F)}(x) > | y \in [x]_R, x \in U\}$.

Where:

$\mu_{\underline{N}(F)}(x) = \wedge_{y \in [x]_R} \mu_F(y)$, $\nu_{\underline{N}(F)}(x) = \wedge_{y \in [x]_R} \nu_F(y)$, $\omega_{\underline{N}(F)}(x) = \wedge_{y \in [x]_R} \omega_F(y)$,

$\mu_{\overline{N}(F)}(x) = \vee_{y \in [[x]_R} \mu_F(y)$, $\nu_{\overline{N}(F)}(x) = \vee_{y \in [x]_R} \nu_F(y)$, $\omega_{\overline{N}(F)}(x) = \vee_{y \in [x]_R} \omega_F(y)$.

Then, $0 \leq \mu_{\underline{N}(F)}(x) + \nu_{\underline{N}(F)}(x) + \omega_{\underline{N}(F)}(x) \leq 3$ and $\mu_{\overline{N}(F)}(x) + \nu_{\overline{N}(F)}(x) + \omega_{\overline{N}(F)}(F)(x) \leq 3$, where "$\vee$" and "$\wedge$" mean "max" and "min" operators respectively, $\mu_F(y)$, $\nu_F(y)$, and $\omega_F(y)$ are the membership, indeterminacy and non-membership of y with respect to F.

Especially for decision-making, we define R as the similarity relation, such that there is a similar or equal labels between two elements, see [35].

The membership, indeterminate-membership, and non-membership are defined from the rough set as follows:

$T(x) = \frac{card(S_B(x) \cap \underline{B}(x))}{card(S_B(x))}$, $I(x) = \frac{card(S_B(x) \cap (\overline{B}(x) \setminus \underline{B}(x)))}{card(S_B(x))}$, and $F(x) = \frac{card(S_B(x) \cap \overline{B}(x))}{card(S_B(x))}$, respectively.

2.3 Entropy and information gain

The entropy in a data source represents the magnitude that measures the information provided about the data source. Entropy provides information about a specific data source or fact [36-38].

Definition 5. Given two classes P and N in a sample space S, where:

$$S = P \cup N \quad (4)$$

Where the cardinality is given by:

$$|P| = p \text{ and } |N| = n \quad (5)$$

Entropy refers to the amount of information necessary to decide whether a sample of S belongs to P or to N and it is defined as Equation 6, see [39, 40]:

$$E(S) = \frac{p}{p+n}\log_2\left(\frac{p}{p+n}\right) - \frac{n}{p+n}\log_2\left(\frac{n}{p+n}\right) \quad (6)$$

When selecting an attribute b the sample space is divided into child subsets of b, the way to determine how much information an attribute b contributes in a total set of attributes A, is given by [41, 42]:

$$Input(b) = E(A) - \sum(\forall \text{ the child sets of } B) \quad (7)$$

Finally, if we have k classes, N instances in the data set, the entropy of the entire set is E, the entropy of each of the subsets is E1 and E2, the number of instances in one class is k1 and in the other k2, then the minimum contribution of information is defined as in Equation 8, see [43, 44]:

$$\frac{\log_2 N - 1}{N} + \frac{\log_2 3^k - 2 - K^E + k * E1 * E2}{N} \quad (8)$$

2.4 K-means algorithm

K-means is one of the most widespread algorithms for grouping. Clustering represents a technique implemented in Data Mining. The idea of k-means is to place all objects in a certain space and given their characteristics to form groups of objects with similar but different features to those that make up other groups. K-means is an unsupervised learning algorithm that has the following characteristics [45, 46]:

- The data set is randomly partitioned into K groups (clusters).
- K center points are selected at random, one from each group (centroids).
- For each data, the distance from the point to each central point of the groups is calculated and the data becomes part of the group whose distance is less than its center.
- If the data is closer to its own group, it stays in its group, otherwise it becomes part of the group of the closest center.
- The previous process is repeated until no point passes the group.

However, the algorithm has some drawbacks:

- The final grouping depends on the initial centroids.
- Convergence in the global optimum is not guaranteed, and for problems with many specimens, it requires a large number of iterations to converge.

3. Materials and methods

This section describes the implementation scenario based on the stored dataset of the tourism sector in the City of Riobamba.

Data set	Type of data	Instances	Attributes
Destination evaluation	Multivariate	1382	4
Historical Tours	Multivariate	720	6
Hotel accommodation	Univariate	1080	5
Transportation	Multivariate	2801	4

Table 1. Data stored from the tourist sector in the City of Riobamba.

A data set with decision attributes was used to execute the algorithms excluding this attribute and then compare the results obtained with those originally indicated by these attributes.

Having data sets for which the decision attribute is known, allows determining the number of groups of the k-means algorithm.

Figure 1 shows a diagram of the flow of the k-means method for the case under study.

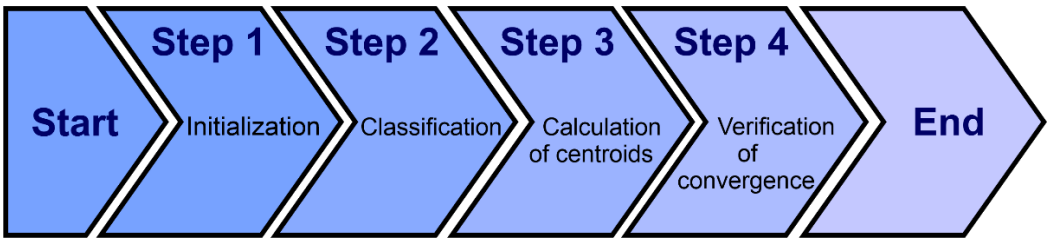

Figure 1. Diagram of the operation of the k-means method.

As in most data mining processes, each repository had to go through the stages of: clean, integrate, select, transform, mine, interpret and present [47]. Figure 1 shows the flow of the k-means method for the case under study. The following is a description of the different steps that describe the method:

Step 1. Initialization: Definition of a set of objects to which the clustering process is applied, which consists of separating the data into groups and a centroid (geometric center of the clusters) for each one. Initial centroids can be determined randomly, while in other cases they process the data and centroids are determined by calculations.

Step 2. Classification: For each data, the square Euclidean distance from the centroids is calculated, the closest centroid to each of the data is determined, and the object is appended to the cluster of the centroid that was selected.

Step 3. Calculation of centroids: The centroids are recalculated for each of the clusters.

Step 4. Verification of convergence: If one of the algorithm's conditions has been met and it must stop, this is called the convergence or stop condition. A set of conditions are defined for processing:

a) The number of iterations.

b) When the centroids obtained in two successive iterations do not change their value.

c) When the difference between the centroids of two successive iterations does not exceed a certain threshold.

d) When there is no transfer of objects between groups in two successive iterations.

If any of the convergence conditions is not met, steps two, three and four of the algorithm are repeated.

For computational processing the algorithms were coded in Python 3.8.12 and it was run on the following platform:

- Intel(r) core(tm) i3-2100 CPU @ 3.10ghzprocessor.

- Operating system: Ubuntu/Linux.

From the k-means algorithm comparison criteria, the one was chosen to maximize the number of success cases, since the last interest is to determine how well the grouping did.

In order to compare the results, three processes were run with the same data sets under the following conditions:

- 90 iterations were made with random processes in order to determine the average effect of the algorithm.
- When ranges were used, there was no point in repeating it more than once as the algorithm is deterministic for a given data set.

3.1 Classic k-means with random centroids

K-means was used as a grouping algorithm so that the resulting groups were then used to label the objects in their decision attribute (D); using the group number in which the object was grouped as the value of the decision attribute (D).

K-means using only the attributes with a contribution of information superior to a border. The entropy of each of the attributes and its information gain were calculated. The method used was as follows:

- Let E (C) be the entropy of the entire set of attributes.
- How much information is provided by the entropy of each of the c condition attributes (C) is calculated.
- Let E(ci) be the entropy of the condition attribute ci.
- Since the selection of the criterion in which value, of the Vc values, to divide the attribute c to calculate the entropy can be very different for each attribute, it is decided to order the Vc values from least to greatest and take the mean as the division criterion.
- The information input of attribute c is equal to: $C - \sum(\forall \text{ the child sets of } C)$
- The condition attributes that provide the greatest amount of information such as those selected are used to choose the initial centers for the k-means algorithm from them.

Once the attributes to be considered have been chosen, if it is desired that the decision attribute (D) take different values from it, then k-means is run to form groups, using as distances only the attributes selected for their greatest contribution of information. The centers can either be initialized randomly or divide the total range of the values of attribute c into k uniform pieces and take these values as initial centers of the k-means algorithm.

Given that they are the attributes that provide the most information, it was decided to initialize the centers with uniform ranges.

3.2 k-means using only attributes selected by rough neutrosophic sets

We used the theory of rough neutrosophic sets to determine which condition attributes are indispensable and which are dispensable and therefore, proceed to the reduction of attributes, calculating the relation of indispensableness of each one of them.

Being P the set of attributes, $a \in P$, the attribute *a* is dispensable in P if:

$$\text{IND}(P) = \text{IND}(p\{a\}) \quad (9)$$

Similarly, once the attributes to be considered have been chosen, if it is desired that the decision attribute (D) take Vd different values, then k-means is run to form Vd groups using only the indispensable attributes for the calculation of distances. The centers can be initialized randomly or the total range can be divided into k uniform pieces. In order to compare the results, the centers with uniform ranges were initialized.

4. Results

From the data recovered from the tourist sector in the City of Riobamba are processed. After applying the previously proposed experiments, we obtained the results that are pr in Table 2.

Characteristic/Data sets	Valuation of fate	Historical Tours	Hotel accommodation	Transport job
Total records	1382	720	1080	2001
Total attributes including decision	4	6	5	4
		***k-means* classic**		
Classic *k-means* success rate mean	63.40	35.20	53.20	54.50

Standard deviation of the classic *k-means* success rate	3.80	7.38	3.01	8.45
Variation coefficient of the classic *k-means* success rate	0.08	0.43	0.07	0.15
***k-means* using information gain**				
Number of attributes removed due to information gain	1	1	1	1
Average success rate using the remaining attributes	56.80	30.40	56.20	48.30
Standard deviation of success rate using only the remaining attributes	0.00	0.00	0.00	0.00
Variation coefficient of success rate using only the remaining attributes	0.00	0.00	0.00	0.00
Rough neutrosophic sets				
Number of attributes removed by rough sets	2	1	0.00	0.00
Average success rate using the remaining attributes	57.6	32.32	54.20	42.40
Standard deviation of success rate using only the remaining attributes	0.00	0.00	0.00	0.00
Variation coefficient of success rate using only the remaining attributes	0.00	0.00	0.00	0.00

Table 2. Results obtained for the different data sets.

From the analysis of the results presented in Table 2, the following discussions are presented:

1. The classical k-means algorithm is highly dependent on the selection of the initial centers. Random center initialization tends to have high standard deviations, therefore high coefficients of variation.
2. Using entropy and information gain, only the attributes that provide more information are used, uniform ranges are used for centroids instead of random centers. The process becomes deterministic for the same data set; therefore, the standard deviation and coefficient of variation are displayed at zero.
3. Once the data has been labeled, or if already labeled data sets are available, and although the determination of the indispensable and dispensable attributes using rough sets is an expensive process in computational time, once determined, the reduction of attributes benefits likewise the future classification process.
4. The classic k-means with random centers showed that in some cases it obtained a higher success rate than the others. The problem being that its standard deviation is high and, therefore, as the average case will not always occur, it can perfectly yield the worst case, or cases close to it, and in these scenarios its performance is lower than when using information gain or rough sets.

Conclusions

This investigation proposed a machine learning method for dealing with unlabeled data sets which bases its operation on:

To use entropy and information gain to select from which attributes calculate the k-means centers.

Use k-means only with the attributes selected from the previous step to label the data in your decision attribute.

Once the objects have been labeled with the previous steps, approximate sets can be used to determine which attributes are dispensable and which are indispensable and, therefore, proceed to the reduction of attributes.

The calculation of entropy, the information gain and the neutrosophic approximate sets requires a computational effort before calculating the k-means.

By implementing the k-means algorithm on the stored data set of the tourism sector in the City of Riobamba, a classification of the information is obtained from relevant data. The proposal provides a tool for decision-making based on achieving better opportunities for the sector.

References

1. M. V. Brachtl, J. L. Durant, C. P. Perez, J. Oviedo, F. Sempertegui, E. N. Naumova, and J. K. Griffiths, "Spatial and temporal variations and mobile source emissions of polycyclic aromatic hydrocarbons in Quito, Ecuador," *Environmental Pollution,* vol. 157, no. 2, pp. 528-536, 2009.
2. K. R. Lomas, and C. A. Trujillo, "Environmental educational model for community tourism of fakcha llakta community-Ecuador," *International Journal of Professional Business Review: Int. J. Prof. Bus. Rev.,* vol. 3, no. 1, pp. 95-110, 2018.
3. J. C. Sierra, "Estimating road transport fuel consumption in Ecuador," *Energy Policy,* vol. 92, pp. 359-368, 2016.
4. F. Pizzitutti, S. J. Walsh, R. R. Rindfuss, R. Gunter, D. Quiroga, R. Tippett, and C. F. Mena, "Scenario planning for tourism management: a participatory and system dynamics model applied to the Galapagos Islands of Ecuador," *Journal of Sustainable Tourism,* vol. 25, no. 8, pp. 1117-1137, 2017.
5. J. E. Ricardo, M. L. Poma, A. A. Pazmiño, A. A. Navarro, L. M. Estévez, and N. B. Hernandez, "Neutrosophic model to determine the degree of comprehension of higher education students in Ecuador," *Neutrosophic Sets & Systems,* vol. 26, 2019.
6. J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of massive data sets*: Cambridge university press, 2020.
7. P. Shi, and J. M. Goodson, "A Data Mining Approach Identified Salivary Biomarkers That Discriminate between Two Obesity Measures," *Journal of Obesity,* vol. 2019, 2019.
8. K. Baalaji, and V. Khanaa, "A Review on Process of Data Mining Approaches in Healthcare Sectors," *Indian Journal of Public Health Research & Development,* vol. 11, no. 1, 2020.
9. E. Alpaydin, *Introduction to machine learning*: MIT press, 2020.
10. M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*: MIT press, 2018.
11. J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *Nature,* vol. 549, no. 7671, pp. 195-202, 2017.
12. J. González, "Genetic classification algorithm proposal," *RCI,* vol. Vol. 4 no. No.2, pp. 37-42, 2013.
13. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, and M. Isard, "Tensorflow: A system for large-scale machine learning." pp. 265-283.
14. S. Mullainathan, and J. Spiess, "Machine learning: an applied econometric approach," *Journal of Economic Perspectives,* vol. 31, no. 2, pp. 87-106, 2017.
15. R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models." pp. 3-18.
16. M. Leyva-Vázquez, F. Smarandache, and J. E. Ricardo, "Artificial intelligence: challenges, perspectives and neutrosophy role.(Master Conference)," *Dilemas Contemporáneos: Educación, Política y Valore,* vol. 6, no. Special, 2018.
17. H.-Y. Zhang, and S.-Y. Yang, "Feature selection and approximate reasoning of large-scale set-valued decision tables based on α-dominance-based quantitative rough sets," *Information sciences,* vol. 378, pp. 328-347, 2017.
18. E. Eiben, D. Hermelin, and M. Ramanujan, "On approximate preprocessing for domination and hitting subgraphs with connected deletion sets," *Journal of Computer and System Sciences,* vol. 105, pp. 158-170, 2019.
19. J. M. Fraser, and H. Yu, "Approximate arithmetic structure in large sets of integers," *arXiv preprint arXiv:1905.05034*, 2019.
20. L. K. Á. Gómez, D. A. V. Intriago, A. M. I. Morán, L. R. M. Gómez, J. A. A. Armas, M. A. M. Alcívar, and L. K. B. Villanueva, "Use of neutrosophy for the detection of operational risk in corporate financial management for administrative," *Neutrosophic Sets and Systems, Book Series, Vol. 26, 2019: An International Book Series in Information Science and Engineering,* vol. 26, pp. 75, 2019.
21. S. Machado, "Approximate lattices and Meyer sets in nilpotent Lie groups," *arXiv preprint arXiv:1810.10870*, 2018.
22. N. Van Hung, D. H. Hoang, and V. M. Tam, "Painlevé-Kuratowski convergences of the approximate solution sets for vector quasiequilibrium problems," *Carpathian Journal of Mathematics,* vol. 34, no. 1, pp. 115-122, 2018.
23. T. Popescu, P. Georgelin, F. Smarandache, and L. Popescu, "The aesthetics of paradoxism," 2002.
24. F. Smarandache, *Paradoxism's Manifestos and International Folklore*: Infinite Study, 2010.
25. F. Smarandache, and S. Pramanik, *New trends in neutrosophic theory and applications*: Infinite Study, 2016.

26. M. L. Vázquez, and F. Smarandache, *Neutrosofía: Nuevos avances en el tratamiento de la incertidumbre*: Infinite Study, 2018.
27. F. Smarandache, and M. Leyva-Vázquez, *Fundamentos de la lógica y los conjuntos neutrosóficos y su papel en la inteligencia artificial*: Infinite Study, 2018.
28. O. Mar, I. Santana, and J. Gulín, "Algoritmo para determinar y eliminar nodos neutrales en Mapa Cognitivo Neutrosófico," *Neutrosophic Computing and Machine Learning,* vol. 8, pp. 4-11, 2019.
29. F. Smarandache, and T. Paroiu, *Neutrosofia ca reflectarea a realităţii neconvenţionale*: Infinite Study, 2012.
30. F. Smarandache, J. E. Ricardo, E. G. Caballero, M. Y. L. Vasquez, and N. B. Hernández, "Delphi method for evaluating scientific research proposals in a neutrosophic environment," *Neutrosophic Sets and Systems*, pp. 204, 2020.
31. N. B. Hernandez, M. B. Ruilova Cueva, and B. N. Mazacón, "Prospective analysis of public management scenarios modeled by the Fuzzy Delphi method," *Neutrosophic Sets and Systems,* vol. 26, no. 1, pp. 17, 2019.
32. H. Wang, F. Smarandache, Y. Zhang, and R. Sunderraman, "Single valued neutrosophic sets," *Review of the Air Force Academy*, no. 1, pp. 10, 2010.
33. M. Y. L. Vázquez, K. Y. P. Teurel, A. F. Estrada, and J. G. González, "Modelo para el análisis de escenarios basados en mapas cognitivos difusos: estudio de caso en software biomédico," *Ingenieria y Universidad: Engineering for Development,* vol. 17, no. 2, pp. 375-390, 2013.
34. S. Broumi, F. Smarandache, and M. Dhar, "Rough Neutrosophic Sets," *Neutrosophic Sets and Systems,,* vol. *3*, pp. 60-65, 2014.
35. M. Abdel-Basset, and M. Mohamed, "The role of single valued neutrosophic sets and rough sets in smart city: Imperfect and incomplete information systems," *Measurement,,* vol. *124*, pp. 47-55, 2018.
36. [36] G. U. Crevecoeur, "Entropy growth and information gain in operating organized systems," *AIP Advances,* vol. 9, no. 12, pp. 125041, 2019.
37. O. Mar, I. Ching, and J. Gulín, "Competency assessment model for a virtual laboratory system at distance using fuzzy cognitive map," *Investigación Operacional,* vol. 38, no. 2, pp. 169-177, 2018.
38. M. Naghiloo, J. Alonso, A. Romito, E. Lutz, and K. Murch, "Information gain and loss for a quantum maxwell's demon," *Physical review letters,* vol. 121, no. 3, pp. 030604, 2018.
39. S. Isler, R. Sabzevari, J. Delmerico, and D. Scaramuzza, "An information gain formulation for active volumetric 3D reconstruction." pp. 3477-3484.
40. H. Carrillo, P. Dames, V. Kumar, and J. A. Castellanos, "Autonomous robotic exploration using a utility function based on Rényi's general theory of entropy," *Autonomous Robots,* vol. 42, no. 2, pp. 235-256, 2018.
41. A. Sadri, Y. Ren, and F. D. Salim, "Information gain-based metric for recognizing transitions in human activities," *Pervasive and Mobile Computing,* vol. 38, pp. 92-109, 2017.
42. S.-B. LIU, "Information-theoretic approach in density functional reactivity theory," *Acta Physico-Chimica Sinica,* vol. 32, no. 1, pp. 98-118, 2016.
43. A. Ben Jaafar, and Z. Bargaoui, "Generalized Split-Sample Test Interpretation Using Rainfall Runoff Information Gain," *Journal of Hydrologic Engineering,* vol. 25, no. 1, pp. 04019057, 2020.
44. J. B. Kalmbach, J. T. VanderPlas, and A. J. Connolly, "Applying Information Theory to Design Optimal Filters for Photometric Redshifts," *arXiv preprint arXiv:2001.01372*, 2020.
45. L. Bai, X. Cheng, J. Liang, H. Shen, and Y. Guo, "Fast density clustering strategies based on the k-means algorithm," *Pattern Recognition,* vol. 71, pp. 375-386, 2017.
46. H. Yang, H. Peng, J. Zhu, and F. Nie, "Co-Clustering Ensemble Based on Bilateral K-Means Algorithm," *IEEE Access*, vol. 8, pp. 51285-51294, 2020.
47. R. G. Ortega, M. Rodríguez, M. L. Vázquez, and J. E. Ricardo, "Pestel analysis based on neutrosophic cognitive maps and neutrosophic numbers for the sinos river basin management," *Neutrosophic Sets and Systems,* vol. 26, no. 1, pp. 16, 2019.

Received: April 13, 2020. Accepted: August 15, 2020