Filomat 34:1 (2020), 89–97
https://doi.org/10.2298/FIL2001089B

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

Document Classification in Information Retrieval System based on Neutrosophic sets

O.G El Barbary[a]

[a]*Mathematics Department, Faculty of Science, Tanta University, Egypt*

Abstract. The aim of the present paper is to suggest a neutrosophic set model for information retrieval and to develop methods and algorithms for information retrieval based on the neutrosophic set model. A neutrosophic set has the possibility of being a common structure for the vagueness analysis of data sets also including big data sets. Here we defined useful techniques like distance and similarity between two neutrosophic sets that have been applied to document classification in information retrieval. We define an innovative algorithm for classifying documents based on Euclidian distance between two neutrosophic sets. About 2500 documents in seven different categories are used for evaluation this new algorithm. Our experiments show that Neutrosophic Classification Algorithm achieved 95% performance.

1. Introduction

In recent years, there has been tremendous growth in the volume of electronic text documents available on the Internet. In these vast volumes of unstructured texts lie nuggets of useful information and knowledge if only we can mine them successfully. Hence there is a growing need for automated techniques to efficiently organize, classify, label, and extracts relevant information.

Automatic text classification is significant for organizing vast amounts of data. Many document retrieval systems have been developed along with advances in areas such as keyword retrieval, similar file retrieval, and automatic document classification. Document classification independent of existing classified information is certainly important. But, humans can find one document, classify in the field health or in the field beauty. The problem is that what the most suitable field for the document is. In the other words, what is the degree of truth that the document belongs to the field.

In the present paper, we consider some aspects of neutrosophic information retrieval, that is, efficient frame for neutrosophic retrieval and representation of the neutrosophic retrieval process. Moreover, classify documents using the neutrosophic set idea of defining a degree of truth, indeterminacy and falsity for a document to a field. These two aspects have not been studied before by other researchers.

This paper is prepared as follows, In Section 2 we give some basic concepts of information retrieval system. Section 3 discusses the basics of neutrosophic sets. Section 4 is a wide previous work in neutrosophic set application and information retrieval application specially for document categorization. In Section 5 discusses the basics of information retrieval using neutrosophic sets. Section 6 is devoted to introduce

2010 *Mathematics Subject Classification*. 68P20, 68T50
Keywords. Information retrieval; neutrosophic set; document classification.
Received: 08 January 2019; Revised: 28 February 2019; Accepted: 08 March 2019
Communicated by Biljana Popović
Email address: Omnia.elbarbari@science.tanta.edu.eg, omniaelbarbary@yahoo.com (O.G El Barbary)

our approach to document classification using distance between neutrosophic sets. Also, an application example is introduced to classify documents in information retrieval systems. Our experiment and its results in section 7. The paper conclusion is given in Section 8.

2. Information Retrieval System

The main aim of an Information Retrieval (IR) structure is to retrieve documents most relevant to the user's query, and the best IR system ranks best the more relevant documents than less relevant ones. The documents are ranked based on terminology in the query and terms in the retrieved documents. In many cases, queries do not contain enough terms to disambiguate the user's knowledge need, so the IR system may return irrelevant documents.

The design and operations of large-scale information system have become of concern to an ever-increasing segment of the scientific and professional world. Furthermore, as the amount and the difficulty of the accessible information have continued to grow, the use of automated or partly mechanized procedures for various information storage and retrieval tasks has also be converted into more widespread. Also, there has been a great growth of online textual information related to digital libraries. So, there is a immense need to search and arrange vast amounts of information in documents.

As a result, a number of large information systems are now in operation in which at least the search operations that is, the comparison of incoming search request with stored information is carried automatically. While these operational information systems are thus able to rapidly to search vast storage files, often containing many hundreds of items, most of other operations other even the search itself are performed manually with the help of human experts.

In particular, all the content analysis and indexing operations of index terms to the stored documents and to the incoming search requests are normally performed by specialists who know the given topic area, as well as the performance characteristics of the retrieval environment within which they operate.

Many of the information systems which based their operations on manual indexing, but largely automatic search methods are quite successful in isolating, from the large mass of largely irrelevant stored material, many of the items which prove pertinent to the users' information need. Many different criteria may suggest themselves for measuring the performance of an information system. In the appraisal work approved with the collection system, the success of an information retrieval system is assumed to depend on its ability to suit the users' information needs by retrieving required material, while rejecting not needed items. Two scales have been generally used for this purpose, known as recall and precision, and instead of respectively the proportion of appropriate material really retrieved, and the percentage of retrieved material actually relevant. Ideally, all relevant items should be retrieved, even as at the same time, all non-relevant items should be rejected, as reflected by perfect recall and precision value to equal 1. It should be noted that both recall and precision measures achieved by a given system are adjustable, in the sense that recreation of the search conditions often lead to high recall, while a tightening of the search conditions leads to high precision. Furthermore, many document retrieval systems developed in areas: Keyword retrieval, similar file retrieval, automatic document classification and document summarization.

3. Neutrosophic Sets

In recent times a new theory has been introduced which is identified as neutrosophic logic and sets. The expression neutro-sophy means knowledge of neutral thought and this neutral represents the main difference between fuzzy and intuitionistic fuzzy logic and set. The first presenter for neutrosophic logic is Florentin Smarandache [27]. It is a judgment in which each proposition is estimated to have a scale of truth (T), a scale of indeterminacy (I) and a scale of falsity (F). A set where every constituent of the universe has a three degree, that is truth, indeterminacy and falsity respectively, and which lies between [0, 1] *, the non-standard unit interval, this is label a Neutrosophic set [20].

The single valued neutrosophic set is a generalization of classical set, fuzzy set, intuitionistic fuzzy set and paraconsistant sets and its defined as follows:

Definition 3.1. *Let D be a universal set, a single valued neutrosophic set is an object* $N = \{\langle d,\ T_N(d), I_N(d), F_N(d)\rangle :$ $d \in D$,*that characterized by three membership functions.*$T_N(d) : D \to [0,1]$ *is a truth membership function,* $I_N(d) : D \to [0,1]$ *is an indeterminacy membership function and* $F_N(d) : D \to [0,1]$ *is a falsity membership function. The sum* $T_N(d) + I_N(d) + F_N(d)$ *of any element* $d \in D$ *deceits in the closed interval [0, 3].*

4. Previous Works

Neutrosophy is the foundation of neutrosophic logic, neutrosophic probability, neutrosophic set, and statistics used in the computer science field and engineering applications. The following is a summarized for some application on neutrosophic. In [15] Faruk Karaaslan introduced the notion of the possibility neutrosophic soft set (NSS) and defined a number of related notions such as a possibility NSS separation, possibility NSS null set, and possibility NSS universal set. Then, based on qualifier on n-norm and n-conorm, he definite set theoretical operations on possibility neutrosophic soft sets such as union, intersection and complement, and suggested some property of these operations.

Deepika Koundal and others in [16] introduced an accurate approach using spatial neutrosophic clustering that helped to reduce the number of false positives, improved the accurate detection and resourceful diagnosis of thyroid nodules. This paper illustrated an automated delineation method that included spatial information with neutrosophic clustering and level-sets for exact and effective segmentation of thyroid nodules in ultrasound images. Xianying Qi and others in [22] proposed a pixel-wise adaptive neutrosophic purify depend on NI feature to remove high-level salt-and-pepper noise. They investigated an algorithm using NS to measure the match of pixels of images.

A.Q. Ansari and others in [3] introduced a neutrosophic classifier (NC), which would utilize neutrosophic logic for its working. The projected NC employs neutrosophic logic for its effective and is an extension of the commonly used fuzzy classifier. B. Kavitha, Dr. S. Karthikeyan and P. Sheeba Maybell in [17] has introduced a budding approach for an invasion detection system using (NC). It is capable of usage fuzzy, vague, partial and inconsistent information in one framework. They exhibit the efficiency of handling vagueness in intrusion recognition precisely using (NC) based intrusion detection system.
The single valued neutrosophic set was introduced for the first time by Smarandache, in 1998 in his book [ ]. Jun Ye in [31] also defined the single valued neutrosophic set (SVNS), which offer a supplementary option to represent vague, imprecise, incomplete, and inconsistent information. It is more appropriate to apply indefinite information and inconsistent information measures. He also proposed the cross entropy of SVNSs that he calls it the single valued neutrosophic cross entropy, which is an extension of the cross entropy of fuzzy sets. He then established a multi-criteria decision-making scheme depends on the proposed single valued neutrosophic cross entropy. Him with others in [11, 32, 33] suggest improved cosine similarity measures of SNSs based on cosine function, including single valued neutrosophic cosine similarity measures and interval neutrosophic cosine similarity measures. Then he introduced weighted cosine similarity measures of SNSs by attractive into account the significance of each element. Furthermore, he improved a medical diagnosis method using the cosine similarity measures to solve medical diagnostic problems with simplified neutrosophic information. Finally, he concluded that the improved cosine measures of SNSs based on cosine function can overcome some drawbacks of existing cosine similarity measures of SNSs in vector space, and their diagnosis technique is very appropriate for management the medical diagnostic troubles with simplified neutrosophic information and demonstrates the efficiency and sagacity of medical diagnoses.

Yanhui Guo, Abdulkadir Sengur in [12] introduced an innovative cluster algorithm and neutrosophic c-means (NCM) meant for doubtful data clustering. They have been introduced two new sample of negative response: the vagueness rejection which concerns the patterns lying near the cluster boundaries, and the distance negative response dealing with patterns that are far away from all the cluster. These procedures are capable to handle uncertainty due to the incomplete definition of the clusters. Finally, the method they proposed was applied into the image separation algorithm. Their experimental outcome demonstrates that the proposed algorithm be able to consider as a hopeful tool for data clustering and image processing.

Umberto Rivieccio in [23] reviewed Smarandache book [28] on NS and presented a serious introduction to NS, focus on the problem of defining appropriate neutrosophic propositional connectives and talk about the relationship among neutrosophic logics and other familiar frameworks for reasoning with vagueness and uncertainty. Jun Ye and Jing Fu in [33] have proposed a similarity measure connecting single valued neutrosophic sets based on tangent function and a multi-period medical diagnostic method based on the similarity evaluation and the biased aggregation of multi-period information. Adrian Rubio Solis and George Panoutsos in [24] have presented a new structure for creating Granular Computing, Neural-Fuzzy modelling structures through the use of Neutrosophic Logic to speak to the issue of ambiguity during the data granulation process. They are offered and discuss the theoretical and computational aspects of this approach, as well as a case of study using real industrial data. Yanhui Guo and H. D. Cheng in [13] has proposed method that employs to perform image segmentation using a γ-means clustering. They have conducted experiments on a variety of images. The experimental outcome demonstrates that the planned approach can segment the images mechanically and effectively. Moreover, there are multiple application for neutrosophic set found in [2, 4, 10,19, 25, 30].

There are many suggested methods modified information retrieval using different tools, one of this is a vector space model with FA words. El-Monsef, M E Abd and others in [1] investigated three different methods of vector space models using FA words. They are developed K- nearest neighbor classification algorithm, Rocchio document classification algorithm and centroid based algorithm. Paper [14] presents an original content-based image retrieval method stand on the Gaussian mixture probability form. The proposed technique provides the explanation toward similar arbitrary images based on color, shape and texture. The glyph structure of the image, which inclines on the form and texture attribute, is modeled and used for content matching.

Paper [7] developed a classifier method for documents using k-means clustering and field association words. In addition, in [8] she developed another method that makes use of FA words to classify the Arabic news. Another development can be found in [9].

5. Information Retrieval based on Neutrosophic Sets

This section clearly discusses the basics of information retrieval using neutrosophic sets.

Definition 5.1. *Let D be a finite set of documents for retrieval,* $D = \{d_1, d_2, \ldots, d_n\}$ *.W is a set of keywords,* $W = \left\{w_1, w_1, \ldots, w_j\right\}$, $w_j \in d_i$ *A Neutrosophic set N in D is described by a truth-membership function* t_N*, an indeterminacy-membership function* i_N *and a falsity-membership function* f_N*, where* $t_N, i_N, f_N : D \to [0,1]$ *are functions and*
$\forall d \in D,\ d = d\,(t_N d\,(w)\,, i_N d\,(w)\,,\ f_N d\,(w)) \in N$ *, is a single valued neutrosophic factor of N.*

A single valued neutrosophic set N over a finite universe $D = \{d_1, d_2, \ldots, d_n\}$
is represented as follows:

$$N = (d_1, <\ t_N d_1\,(w_i)\,, i_N d_1\,(w_i)\,,\ f_N d_1\,(w_i) >) + (d_2, <\ t_N d_2\,(w_i)\,, i_N d_2\,(w_i)\,,\ f_N d_2\,(w_i) >)$$

$+\ldots+(d_n, <\ t_N d_n\,(w_i)\,, i_N d_n\,(w_i)\,,\ f_N d_n\,(w_i) >)$ *where,* $t_N d_i\left(w_j\right) = \frac{S - r_{d_i}\left(w_j\right)}{S}$, $i_N d_i\left(w_j\right) = \frac{r_{d_i}\left(w_j\right)}{M}$, $f_N d_i\left(w_j\right) = \frac{r_{d_i}\left(w_j\right)}{S}$,
$S = \sum_{j=1}^{n} r d_i\left(w_j\right)$, $M = \sum_{k=1}^{m} r d_k\left(w_j\right)$.
Where r is the number of appearance the word w_j *in the document* d_i *, S is the number of appearance the word* w_j *in the set D and M is the number of appearance the word* w_j *in the subset N.*

Example 5.1. *Assume that* $\mathrm{D} = \{\mathrm{d}_1, \mathrm{d}_2, \mathrm{d}_3, \mathrm{d}_4, \mathrm{d}_5, \mathrm{d}_6, \mathrm{d}_7\}$ *is a set of documents taken from our list of [5, 18, 21, 23, 28, 26, 34] and the set of keywords are* W ={ *Neutrosophic, Fuzzy, Intuitionistic, Indeterminacy, Membership*}*, N is a subset of documents from* $N = \{\mathrm{d}_1, \mathrm{d}_3, \mathrm{d}_5, \mathrm{d}_7\}$ *. They selected according to the occurrence of the set of keywords W, where* $t_N d_i\left(w_j\right)$, $i_N d_i\left(w_j\right)$ *and* $f_N d_i\left(w_j\right)$ *a degree of 'strong occurrence of keywords', a degree of 'indeterminacy of key words' and a degree of 'poor occurrence of keywords ' respectively. The number of occurrence keywords in the documents given in Table 1. A is a single valued for neutrosophic set N given in Table 2.*

D	Neutrosophic	Fuzzy	Intuitionistic	Indeterminacy	Membership
d_1	40	16	5	5	7
d_2	29	7	3	21	67
d_3	32	58	25	7	17
d_4	42	31	18	11	23
d_5	64	48	13	18	0
d_6	23	57	29	7	17
d_7	117	13	0	39	5
S	253	135	43	69	29
M	347	230	93	108	136

Table 1: The number of occurrences of keywords in documents

D	Neutrosophic			Fuzzy			Intuitionistic			Indeterminacy			Membership		
	T_N	I_N	F_N	T_N	I_N	F_N	T_N	I_N	F_N	T_N	I_N	F_N	T_N	I_N	F_N
d_1	0.84	0.12	0.16	0.88	0.07	0.12	0.88	0.05	0.12	0.93	0.05	0.07	0.76	0.05	0.24
d_3	0.87	0.09	0.13	0.57	0.25	0.43	0.42	0.27	0.58	0.90	0.06	0.10	0.41	0.13	0.59
d_5	0.75	0.18	0.25	0.64	0.21	0.36	0.70	0.14	0.30	0.74	0.17	0.26	1.00	0.00	0.00
d_7	0.54	0.34	0.46	0.90	0.06	0.10	1.00	0.00	0.00	0.43	0.36	0.57	0.83	0.04	0.17

Table 2: A single valued for a neutrosophic set of keywords

6. Document Classification using Distance Between Neutrosophic Sets

Single valued neutrosophic sets [17, 20] are particular instance of neutrosophic sets which motivated from a realistic point of view that can be used in real technical and engineering applications. According to [20] "Again, distance and similarity are key concepts in an amount of field such as linguistics, psychology, computational intelligence etc". So we can use the distance between two neutrosophic sets to document classification.

6.1. Euclidean distance between two neutrosophic sets

In Section 3 we define the documents as a neutrosophic set. This section introduces the notion of distance between two documents as a single valued neutrosophic.

Let the sets d_1 and d_2 defined over the finite universe $D = \{d_1, d_2, \ldots, d_n\}$, and Let d_1 *and* d_2 are two single valued neutrosophic sets in $D = \{d_1, d_2, \ldots, d_n\}$.Then the distance between d_1 *and* d_2 as follows:

$$d_n(d_i,\ d_j) =$$

$$\sum_{k=1}^{m}\left[\sum_{\substack{i,\, j=1\\ i?j}}^{n}\left\{\left|t_{d_i}(w_k)-t_{d_j}(w_k)\right|+\left|i_{d_i}(w_k)-i_{d_j}(w_k)\right|+\left|f_{d_i}(w_k)-f_{d_j}(w_k)\right|\right\}\right]$$

The Euclidian distance between $d_i$and d_j is defined as follows:

$$e\left(d_i,\ d_j\right)=\sqrt{\sum_{k=1}^{m}\sum_{\substack{i,\, j=1\\ i?j}}^{n}\left\{(t_{d_i}(w_k)-t_{d_j}(w_k))^2+(I_{d_i}(w_k)-I_{d_j}(w_k))^2+(f_{d_i}(w_k)-f_{d_j}(w_k))^2\right\}}$$

The normalized Euclidian distance between $d_1$and d_2 is defined as follows:

$$q_N\left(d_i,\ d_j\right) = \sqrt{\frac{1}{3n}\sum_{k=1}^{m}\sum_{\substack{i,j=1\\ i?j}}^{n}\left\{(t_{d_i}(w_k) - t_{d_j}(w_k))^2 + (I_{d_i}(w_k) - I_{d_j}(w_k))^2 + (f_{d_i}(w_k) - f_{d_j}(w_k))^2\right\}}$$

Example 6.1 From the data of Example 4.1 and Tables 1& 2, the normalized Euclidian distance between $d_1$and d_3 is given as follows:

$$q_N(d_1,\ d_3) = \sqrt{\frac{1}{3n}\sum_{k=1}^{m}\sum_{\substack{i,j=1\\ i?j}}^{n}\left\{(t_{d_i}(w_k) - t_{d_j}(w_k))^2 + (I_{d_i}(w_k) - I_{d_j}(w_k))^2 + (f_{d_i}(w_k) - f_{d_j}(w_k))^2\right\}}$$

$$= \sqrt{\frac{1}{3n}\sum_{k=1}^{m}\left\{(t_{d_1}(w_k) - t_{d_3}(w_k))^2 + (I_{d_1}(w_k) - I_{d_3}(w_k))^2 + (f_{d_1}(w_k) - f_{d_3}(w_k))^2\right\}}$$

$$= \sqrt{\frac{1}{3n}\sum_{k=1}^{m}\left\{(t_{d_1}(w_1) - t_{d_3}(w_k))^2 + (I_{d_1}(w_k) - I_{d_3}(w_k))^2 + (f_{d_1}(w_k) - f_{d_3}(w_k))^2\right\}}$$

$$= \sqrt{\frac{1}{3n}\left\{\begin{array}{l}\left[(0.84-0.87)^2 + (0.12-0.09)^2 + (0.16-0.13)^2\right] +\\ \left[(0.88-0.57)^2 + (0.07-0.25)^2 + (0.12-0.43)^2\right] +\\ \left[(0.88-0.42)^2 + (0.05-0.27)^2 + (0.12-0.58)^2\right] +\\ \left[(0.93-0.90)^2 + (0.05-0.06)^2 + (0.07-0.10)^2\right] +\\ \left[(0.76-0.41)^2 + (0.05-0.13)^2 + (0.24-0.59)^2\right]\end{array}\right\}}$$

$$= \sqrt{\frac{1}{6}(0.0027 + 0.2178 + 0.4716 + 0.0019 + 0.2514)}$$

$= 2.45$

6.2. *Neutrosophic Classification Algorithm (NCA)*

Classification Neutrosophic (C, D)

1. KW← Extract Keywords(D)
2. N← neutrosophic set on (D)
3. n← Count Docs(D)
4. $r \leftarrow count\ w_j\ (d_i)$
5. $S \leftarrow count\ w_j\ (D)$
6. $M \leftarrow count\ w_j\ (N)$
7. for each $w_i \in KW$
8. for each $d_j \in N$
9. do
10. $t_N d_i\left(w_j\right) = \frac{S - r_{d_i}(w_j)}{S}$,
11. $i_N d_i\left(w_j\right) = \frac{r_{d_i}(w_j)}{M}$,
12. $f_N d_i\left(w_j\right) = \frac{r_{d_i}(w_j)}{S}$

13. *$d_i \wedge d_j$are in the same class C iff* $q_N(d_i, d_j) \geq \infty$

The classification neutrosophic algorithm is amongst the simplest of all machine learning algorithms: an entity is classified by a majority vote of its neighbors. By using the distance between two (NS) with the object being assigned to the class most common amongst its nearest to another document.

This scheme is a generality of linear interpolation. The neighbors are in use from a set of objects for which the correct classification (or, in the case of regression, the value of the property) is known. The neutrosophic classification algorithm is sensitive to the local structure of the data. Nearest neighbor rules in construct calculate the decision border in an implicit manner. It is also probable to calculate the decision boundary itself explicitly, and to do so in a professional behavior.

7. Experimental

The objective of this evaluation is to decide whether the classification is useful to infer the better classification accuracy and performance.

Our experiments, trained the system using different documents collected from the Internet. It is collected from Wikipedia. A dataset with 2500 documents classified in seven different categories is used for evaluation. The selected dataset contains seven categories of document: multimedia, networking, information system, electronics, modeling and meteorology. All the seven categories are easily differentiated The model is built based on the "neutrosophic" classifier. Table 3 summarizes the result of using neutrosophic classifier to classify the documents. However, it surprisingly finds that the results reach to (90%).

We measure our experiment using (Recall, Precession, F-Measure)

Precision, Recall and F-measure are used to estimate relevancies of the presented methods and defined as follows [6]:

$$Recall(R) = \frac{Correct \cdots Classified \cdots Documnts}{Totall \cdots Corrected \cdots Classified}$$

$$\Pr ecison(P) = \frac{Correct \cdots Classified \cdots Documnts}{Totall \cdots Retrieved \cdots Classified}$$

$$F - measure = \frac{2 \times P \times R}{P + R}$$

Name of field	**Precision**	**Recall**	**F- measure**
Multimedia	0.75	0.86	0.80
Networking	0.82	0.91	0.86
Information system	0.79	0.87	0.83
Modeling	0.89	0.97	0.93
Electronics	0.74	0.82	0.77
Meteorology	0.91	0.98	0.94

Table 3: the classification accuracy achieved by the neutrosophic classification algorithms .

In conclusion, the results of F-measure records a high performance for classifying documents.

8. Conclusions and Future Works

The objectives of this work are to study new alternative methods of document classifications in information retrieval systems depend on neutrosophic sets. The advantage of using neutrosophic sets is that neutrosophic sets are a good mathematical tool for data classification by using distance between two neutrosophic set. It seems to be advantageous for mining of document retrieval as well as for other information retrieval techniques.

The major purpose of the future work of this paper is to suggest a modification of our method by using field association words instead of keywords. Also, we can extend their method of maximizing the deviation by using the refined neutrosophic set, i.e. the truth value T is refined into the types of truths such as T1, T2, etc., Similarly indeterminacy, I is split/refined into the types of indeterminacies I1, I2, etc., and the falsehood F is split into F1, F2, etc. Another one of the modification is to gain the advantages of rough sets and topological spaces with our methods. An advantageous way may be the use of some generalizations of fuzzy sets and rough sets with topology, which enable accurate work with document retrieval.

References

[1] M. E. Abd El-Monsef,; Atlam, El-Sayed, O. G. El-Barbary, Combining FA words with vector space models for Arabic text categorization ,International Information Institute (Tokyo) Information 16 (6)(2013) 3517-3528.

[2] C. Antony Crispin Sweety& I. Arockiarani, Topological structures of fuzzy neutrosophic rough sets, Neutrosophic Sets and Systems, Vol. 9, 2015, 50-57

[3] A.Q. Ansari, Ranjit Biswas, Swati Aggarwal, Neutrosophic classifier: An extension of fuzzy classifier, Applied Soft Computing, Volume 13, Issue 1, January 2013, Pages 563-573.

[4] Jiqian Chen, Jun Ye, A Projection Model of Neutrosophic Numbers for Multiple Attribute Decision Making of Clay-Brick Selection, Neutrosophic Sets and Systems, Vol. 12, 2016, 139-142.

[5] H. D. Cheng, and Yanhui Guo, a new neutrosophic approach to image thresholding, New Mathematics and Natural Computation, Vol. 4, No. 3 (2008) 291–308.

[6] Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze : Introduction to Information Retrieval, ISBN 978-0-521-86571-5, Cambridge University Press (2008).

[7] O. G. El-Barbary, Using Field Association Words and K-means Cluster for Arabic Document Classification, Ciniea ea tecnia journal, Vol. 30 (3)(2015) 287-299.

[8] O. G. El-Barbary, Arabic News Classification Using Field Association Words, Advances in research, 6(1)(2016) 1-9.

[9] O. G. El-Barbary, Using Arabic skeleton morphology and Maximum Entropy for Arabic Document classification, British Journal of Mathematics & Computer Science 14(3)(2016) 1-9.

[10] Y. Guo, R. Xia, Şengür, A. et al., A novel image segmentation approach based on neutrosophic c-means clustering and indeterminacy filtering, Neural Comput & Applic Vol. 27 (2016) 1-11.

[11] Yanhui Guo, Abdulkadir Şengür, Jun Ye, A novel image thresholding algorithm based on neutrosophic similarity score, Measurement, Volume 58, December 2014, Pages 175-186.

[12] Yanhui Guo, Abdulkadir Sengur, NCM: Neutrosophic c-means clustering algorithm, Pattern Recognition, Volume 48, Issue 8, August 2015, Pages 2710-2724.

[13] Yanhui Guo, H.D. Cheng, New neutrosophic approach to image segmentation, Pattern Recognition, Volume 42, Issue 5, May 2009, Pages 587-595.

[14] Y.D. Khan, F. Ahmad & S.A. Khan, Content-based image retrieval using extroverted semantics: a probabilistic approach, Neural Comput & Applic Vol. 24(7)(2014) 1735-1748.

[15] Faruk Karaaslan, Possibility neutrosophic soft sets and PNS-decision making method, Applied Soft Computing, In Press, Corrected Proof, and Available online 15 July 2016.

[16] Deepika Koundal, Savita Gupta, Sukhwinder Singh, Automated delineation of thyroid nodules in ultrasound images using spatial neutrosophic clustering and level set, Applied Soft Computing, Volume 40, March 2016, Pages 86-97.

[17] B. Kavitha, Dr. S. Karthikeyan, P. Sheeba May bell, An ensemble design of intrusion detection system for handling uncertainty using Neutrosophic Logic Classifier, Knowledge-Based Systems, Volume 28, April 2012, Pages 88-96.

[18] Pawalai Kraipeerapun, Chun Che Fung, Binary classification using ensemble neural networks and interval neutrosophic sets, Neurocomputing 72 (2009) 2845–2856.

[19] Chunfang Liu, Yuesheng Luo, A new method to construct entropy of interval-valued Neutrosophic Set, Neutrosophic Sets and Systems, Vol. 11, 2016, 8-11.

[20] P. Liu, & Y. Wang, Multiple attributes decision-making method based on single-valued neutrosophic normalized weighted Bonferroni mean, Neural Comput & Applic, Vol. 25(7)(2014) 2001–2010.

[21] Juan-juan Peng, Jian-qiang Wang, Jing Wang, Hong-yu Zhang and Xiao-hong Chen, Simplified neutrosophic sets and their applications in multi-criteria group decision-making problems, International Journal of Systems Science, Vol. 47, 2015, 2342-2358.

[22] Xianying Qi, Boqiang Liu, Jianwei Xu, A neutrosophic filter for high-density Salt and Pepper noise based on pixel-wise adaptive smoothing parameter, Journal of Visual Communication and Image Representation, Volume 36, April 2016, Pages 1-10.

[23] Umberto Rivieccio, Neutrosophic logics: Prospects and problems, Fuzzy Sets and Systems, Volume 159, Issue 14, 16 July 2008, Pages 1860-1868.

[24] Adrian Rubio Solis, George Panoutsos, Granular computing, neural-fuzzy modelling: A neutrosophic approach, Applied Soft Computing, Volume 13, Issue 9, September 2013, Pages 4010-4021.

[25] A. A. Salama , Basic Structure of Some Classes of Neutrosophic Crisp Nearly Open Sets & Possible Application to GIS Topology, Neutrosophic Sets and Systems, Vol. 7, 2015, 18-22.

[26] Florentin Smarandache, Neutrosophic Sets and Systems, (t, i, f)-Neutrosophic Structures & I-Neutrosophic Structures Vol. 8, (2015) 3-9.

[27] F. Smarandache , A Unifying Field in Logics, Neutrosophy: NeutrosophicProbability, Set and Logic. American Research Press (1999).

[28] Florentin Smarandache, *Neutrosophy. Neutrosophic Probability, Set, and Logic*, ProQuest Information & Learning, Ann Arbor, Michigan, USA, 105 p., 1998; http://fs.unm.edu/eBook-Neutrosophics6.pdf (edition online);

[29] Zhang-peng Tian , Hong-yu Zhang , Jing Wang , Jian-qiang Wang and Xiao-hong Chen, Multi-criteria decision-making method based on a cross-entropy with interval neutrosophic sets, International Journal of Systems Science, Vol. 47,2015, 3598-3608.

[30] W. Yang, J. Shi, Y. Pang, et al., Linear assignment method for interval neutrosophic sets Neural Comput & Applic Vol. 27 (2016) 1-12.

[31] Jun Ye, Single valued neutrosophic cross-entropy for multicriteria decision making problems, Applied Mathematical Modelling, Volume 38, Issue 3, 1 February 2014, Pages 1170-1175.

[32] Jun Ye, Improved cosine similarity measures of simplified neutrosophic sets for medical diagnoses, Artificial Intelligence in Medicine, Volume 63, Issue 3, March 2015, Pages 171-179.

[33] Jun Ye, Jing Fu, Multi-period medical diagnosis method using a single valued neutrosophic similarity measure based on tangent function, Computer Methods and Programs in Biomedicine, Volume 123, January 2016, Pages 142-149.

[34] Jun Ye, Multi criteria decision-making method using the correlation coefficient under single-valued neutrosophic environment, International Journal of General Systems, 2013, Vol. 42, No. 4, 386–394.