IOSR Journal of Engineering (IOSR JEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 29-34

Computational Model for Document Classification Based on New Fuzzy Soft Distance

Dr. Neeraj Sahu[1]
[1]*Department of Computer Science, GHRIIT, Nagpur, Maharashtra, India*

Abstract: *In this paper, we introduced new fuzzy soft distance based document classification. The Fuzzy Soft Distance (FSD) is a powerful classification for Document classification. The different distance are based on Hamming distance, normalized Hamming distance, Euclidean distance, normalized Euclidean distance, Hausdorff metric, normalized hausdorff distance, generalized Hamming distance, generalized Euclidean distance and generalized hausdorff distance. An application of Fuzzy Soft distance is illustrated for the document classification. We use here MATLAB 7.14 software tools for calculating result analysis of efficiency and accuracy of document classification.*

Keywords: *Fuzzy sets, Soft sets, Fuzzy Soft sets, Document Classification, Fuzzy Soft Distance*

I. Introduction

Document classification is an important issue in text mining. Classification has been widely applicable in different areas of science, technology, social science, biology, economics, medicine and stock market. Classification problem appears in other different field like pattern recognition, statistical data analysis, bio-informatics, etc. There exist many classification methods in the literature. In this paper a fuzzy distance measure between two generalized fuzzy numbers is developed.

In last recent years, lot of research work has been done on Document classification. Some contributions are as follows:

D. Molodtsov gave "Soft set theory—first results,"[2], In this paper the soft set theory offers a general mathematical tool for dealing with uncertain, fuzzy, not clearly defined objects.

D. Guha and D. Chakrabortygave "A new approach to fuzzy distance measure and similarity measure between two generalized fuzzy numbers[6],D. Molodtsov gave "The Theory of Soft Sets[5] In this paper also definethe soft set theory offers a general mathematical tool for dealing with uncertain, fuzzy, not clearly defined objects.

P.K. Maji, R. Biswas and A.R. Roy gave "Fuzzy soft sets [3] In this paper introduce the concept of possibility fuzzy soft expert set.

P.K. Maji, R. Bismas and A.R. Roy gave "Soft set theory [4] In this paper define introduce the concept of possibility fuzzy soft expert set.

L.A. Zadehgave "Fuzzy Sets[1], In this paper the theory of possibility described in this paper is related to the theory of fuzzy sets by defining the concept of a possibility distribution as a fuzzy restriction which acts as an elastic constraint on the values that may be assigned to a variable.

G.A. Papakostas gave "Distance and similarity measures between intuitionistic fuzzy sets:A comparative analysis from a pattern recognition point of view," [7], A detailed analysis of the distance and similarity measures for intuitionistic fuzzy sets proposed in the past is presented in this paper.

Chang Wang and AnjingQu gave "Entropy similarity measure and distance measure of vague soft sets and their relations[8], In this paper define Soft set theory offers a general mathematical tool for dealing with uncertain, fuzzy, not clearly defined objects.

Z. Xu and M. Xia, gave "Distance and similarity measures for hesitant fuzzy sets," [10], In this paper, propose a variety of distance measures for hesitant fuzzy sets, based on which the corresponding similarity measures can be obtained. We investigate the connections of the aforementioned distance measures and further develop a number of hesitant ordered weighted distance measures and hesitant ordered weighted similarity measures.

H. Zhang and L. Yu, gave "New distance measures between intuitionist fuzzy sets and interval-valued fuzzy sets,"[9], In this paper, propose A New Similarity Measure between Intuitionistic Fuzzy Sets and Its Application to Pattern.

H. Liao, Z. Xu and X. Zeng, gave "Distance and similarity measures for hesitant fuzzy linguistic term sets and their application in multi-criteria decision making," [11], In this paper, the hesitant fuzzy linguistic term sets (HFLTSs), which can be used to represent an expert's hesitant preferences when assessing a linguistic variable, increase the flexibility of eliciting and representing linguistic information

In section1, we describe introduction to document classification. section 2 we are describe preliminaries and basic definitions and in section 3 are we describe proposed fuzzy soft distance formula for classification and in section 4 are we describe proposed fuzzy soft distance experimental result analysis and in section 5 are we describe proposed fuzzy soft distance conclusion.

II. Preliminaries And Basic Definitions

The soft set theory [2,5,3], proposed by Russian researcher Molodtsov [3], in 1999 as a new generic mathematical tool for dealing with the uncertain data.

Definition 1([3]). Let U be an initial universe set and E be a set of parameters. Let P (U) denotes the power set of U and $A\subset E$. A pair (F, A) is called a soft set over U, where F is a mapping given by $F: A\rightarrow P(U)$. In other words, a soft set over U is a parameterized family of subsets of the universe U. For $\varepsilon\in A$, F (ε) may be considered as the set of ε-approximate elements of the soft set (F, A).

Example 1: Suppose that Mr. X wants to buy a car. Let $U=\{c_1,c_2,c_3,c_4,c_5\}$be a set of car under consideration. Let A= $\{e_1, e_2, e_3\}$ be a set of parameters where e_1=expensive,e_2=more comfort and e_3=in the good average. Suppose that

$F(e_1) = \{c_1, c_5\}$
$F(e_2) = \{c_1, c_3, c_5\}$
$F(e_3) = \{c_1, c_4\}$

The soft set (F, A) is describe the "attractiveness of the car". $F(e_1)$ means "car (expensive)" whose function-value is the set $\{c_1,c_5\}$, $F(e_2)$ means "car (more comfort)" whose function-value is the set $\{c_1,c_3,c_5\}$and $F(e_3)$ means "car(in the good average)" whose function-value is the set $\{c_1, c_4\}$.

Fuzzy soft sets

The real world is inherently uncertain, imprecise and vague. Traditional mathematical tools cannot deal with such problems. The fuzzy sets theory is widely employed in such kinds of problems. Maji and Biswas et al. [3, 5] proposed the notion of fuzzy soft sets and an example of decision-making discussed.

Definition 2. [3,5]: Let U be an initial universe set and E be a set of parameters. Let P (U) denotes the set of all fuzzy sets of U and $A \subset E$. A pair (F,A) is called a fuzzy soft set over U, where F is a mapping given by $F: A\rightarrow P(U)$.

Example 2: Consider example 1.In real life, infect much information is fuzzy, we can't describe fuzzy information with only two numbers 0 and 1, we often use a membership function instead of the crisp number 0 and 1 to characterize it. Then the fuzzy soft set (F,A) can describe the "attractiveness of the car" under the Fuzzy circumstances.

$F(e_1) = \{ h_1 /0.3, h_2/0.4, h_3 /0.6,h_4/0.8, h_5/0.7\}$
$F(e_2) = \{ h_1 /0.6, h_2/0.4, h_3 /0.9, h_4/0.3, h_5/0.2\}$
$F(e_3) = \{ h_1 /0.3, h_2/0.4,h_3 /0.8, h_4/0.5, h_5/0.6\}$

III. Proposed Fuzzy Soft Distance Formule For Classification

In this section of paper, we describe fuzzy soft distance based document classification. The Fuzzy Soft Distance (FSD) is a powerful classification for Document classification. The different distance are based on Hamming distance, normalized Hamming distance, Euclidean distance, normalized Euclidean distance,Hausdorff metric, normalized Hausdorff distance, generalized Hamming distance, generalized Euclidean distance and generalized Hausdorff distance. fuzzy soft distance for calculating result analysis of efficiency and accuracy of document classification.

Table 1: FUZZY Soft Distance

S.No	Name	Formulae
1	Fuzzy Soft Hamming Distance(f_1)	$d_2((F,M),(G,N)) = \sum_{i=1}^{m} \sum_{j=1}^{n} \lvert F(ei)(xj) - G(ei)(xj)\rvert$;
2	Fuzzy Soft Euclidean Distance (f_2)	$d_3((F,M),(G,N)) = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} \lvert F(ei)(xj) - G(ei)(xj)\rvert^{2} \right)^{1/2}$;
3	Fuzzy Soft Normalized Euclidean Distance(f_3)	$d_4((F,M),(G,N)) = \frac{1}{n} \left(\sum_{i=1}^{m} \sum_{j=1}^{n} \lvert F(ei)(xj) - G(ei)(xj)\rvert^{2} \right)^{1/2}$;
4	Fuzzy soft Haudorff metric distance(f_4)	$d_5((F,M),(G,N)) = \max \lvert F(ei)(xj) - G(ei)(xj) \rvert$;
5	Fuzzy soft normalized Hamming distance(f_5)	$d_6((F,M),(G,N)) = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{lxi} \sum_{j=1}^{lxi} \lvert F(ei)_{M}^{\sigma(j)}(xj) - G(ei)_{N}^{\sigma(j)}(xj)\rvert \right]$
6	Generalized fuzzy soft normalized Hausdorff distance(f_6)	$d_7((F,M),(G,N)) = \left[\frac{1}{n} \sum_{i=1}^{n} \max_j \lvert F(ei)_{M}^{\sigma(j)}(xj) - G(ei)_{N}^{\sigma(j)}(xj)\rvert^{\lambda} \right]^{1/\lambda}$
7	Fuzzy soft normalized Hamming Hausdorff distance(f_7)	$d_8((F,M),(G,N)) = \frac{1}{n} \sum_{i=1}^{n} \max_j \lvert F(ei)_{M}^{\sigma(j)}(xj) - G(ei)_{N}^{\sigma(j)}(xj)\rvert$
8	Fuzzy soft normalized Euclidean Hausdorff distance (f_8)	$d_9((F,M),(G,N)) = \left[\frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{lxi} \sum_{j=1}^{lxi} \lvert F(ei)_{M}^{\sigma(j)}(xj) - G(ei)_{N}^{\sigma(j)}(xj)\rvert^{2} \right] \right]^{1/}$
9	Hybrid fuzzy soft normalized Hamming distance (f_9)	$d_{10}((F,M),(G,N)) = \left[\frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{lxi} \sum_{j=1}^{lxi} \lvert F(ei)_{M}^{\sigma(j)}(xj) - G(ei)_{N}^{\sigma(j)}(xj)\rvert^{\lambda} \right] \right]^{1/\lambda}$
10	Hybrid fuzzy soft normalized Euclidean distance(f_{10})	$d_{11}((F,M),(G,N)) = \left[\left[\frac{1}{n} \sum_{i=1}^{n} \max_j \lvert F(ei)_{M}^{\sigma(j)}(xj) - G(ei)_{N}^{\sigma(j)}(xj)\rvert^{2} \right] \right]^{1/2}$
11	Generalized hybrid fuzzy soft normalized distance(f_{11})	$d_{12}((F,M),(G,N)) = \frac{1}{2n} \sum_{i=1}^{n} \left(\frac{1}{lxi} \sum_{j=1}^{lxi} \lvert F(ei)_{M}^{\sigma(j)}(xj) - G(ei)_{N}^{\sigma(j)}(xj)\rvert + \max_j \lvert F(ei)_{M}^{\sigma(j)}(xj) - G(ei)_{N}^{\sigma(j)}(xj)\rvert \right.$
12	Generalized fuzzy soft weighted distance(f_{12})	$d_{15}((F,M),(G,N)) = \left[\sum_{i=1}^{n} w_i \left(\frac{1}{lxi} \sum_{j=1}^{lxi} \lvert F(ei)_{M}^{\sigma(j)}(xj) - G(ei)_{N}^{\sigma(j)}(xj)\rvert^{\lambda} \right) \right]^{1/\lambda}$
13	Generalized fuzzy soft weighted Hausdorff distance(f_{13})	$d_{16}((F,M),(G,N)) = \left[\sum_{i=1}^{n} w_i \max_j \left(\lvert F(ei)_{M}^{\sigma(j)}(xj) - G(ei)_{N}^{\sigma(j)}(xj)\rvert^{\lambda} \right) \right]^{1/\lambda}$
14	Fuzzy soft weighted Hamming distance(f_{14})	$d_{17}((F,M),(G,N)) = \sum_{i=1}^{n} w_i \left[\frac{1}{lxi} \sum_{j=1}^{lxi} \lvert F(ei)_{M}^{\sigma(j)}(xj) - G(ei)_{N}^{\sigma(j)}(xj)\rvert \right]$
15	Fuzzy soft weighted Hamming–Hausdorff distance(f_{15})	$d_{18}((F,M),(G,N)) = \sum_{i=1}^{n} w_i \max_j \lvert F(ei)_{M}^{\sigma(j)}(xj) - G(ei)_{N}^{\sigma(j)}(xj) \rvert$
16	Fuzzy soft weighted Euclidean distance(f_{16})	$d_{19}((F,M),(G,N)) = \left[\sum_{i=1}^{n} w_i \left(\frac{1}{lxi} \sum_{j=1}^{lxi} \lvert F(ei)_{M}^{\sigma(j)}(xj) - G(ei)_{N}^{\sigma(j)}(xj)\rvert^{2} \right) \right]^{1/2}$

IV. Results And Discussion

In this section fuzzy soft distance are useful to calculating here distance between two fuzzy soft distance in the domains such as in social, economic system, pattern recognition, medical diagnosis and game theory coding theory. Then we find fuzzy soft distance of these two fuzzy soft (FS). If they are significantly similar then we conclude that the accuracy and efficiency is percentages calculated for function f_1 to f_{15}.

The algorithm 1 of this method is as follows:

Algorithm 1: Calculating accuracy and efficiency of (Fuzzy Soft Distance) FSD

Step1. Compute fuzzy soft distance between (F,M) and (G, N).
Step2. Compute fuzzy soft distance accuracy.
Step3. Compute fuzzy soft distance efficiency.
Step4. Display result analysis of accuracy and efficiency.
Step5. Use similarity to estimate the results.

Table 2: Accuracy AND EFFICIENCY Percentages for FUNCTION F1 to f15

Functions	Accuracy%	Efficiency%
f_1	88	79
f_2	87	88
f_3	74	81
f_4	89	89
f_5	98	89
f_6	89	90
f_7	79	79
f_8	78	78
f_9	91	89
f_{10}	98	89
f_{11}	98	88
f_{12}	87	86
f_{13}	84	85
f_{14}	87	78
f_{15}	87	81
f_{16}	91	91

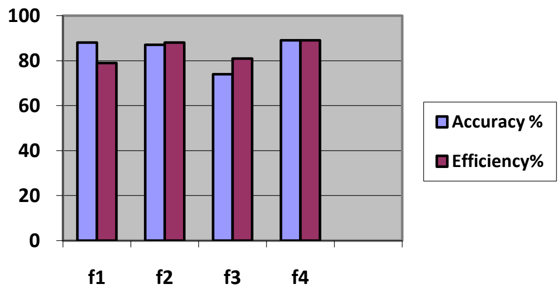

Figure1. Function f_1 to f_4 accuracy and efficiency percentages

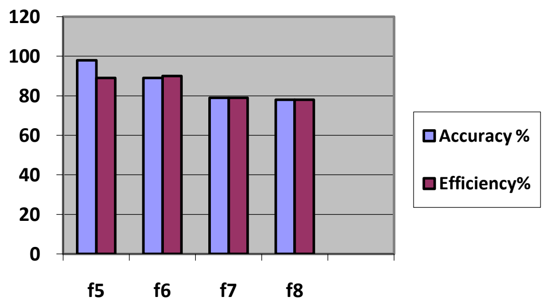

Figure2. Function f_5 to f_8 accuracy and efficiency percentages

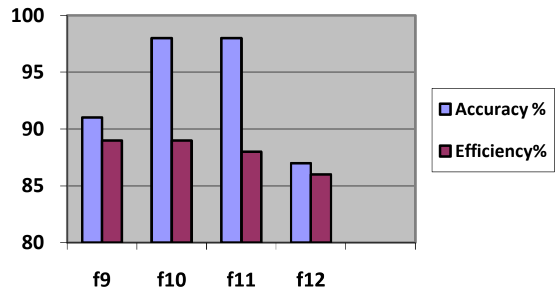

Figure.3 function f_9 to f_{12} accuracy and efficiency percentages

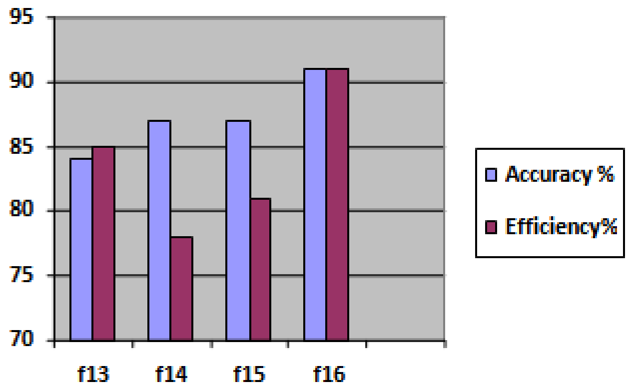

Figure.4function f_{13} to f_{16} accuracy and efficiency percentages

In Table I we describe fuzzy soft distance and Table II we describe accuracy and efficiency percentages for function f1 to f_{16}

From figure 1 to figure 4 in experimental results we describe classification accuracy and efficiency percentages shows for functions f_1 to f_{16}.

V. Conclusion

The types of new distance are proposed for fuzzy soft sets. The Fuzzy Soft distance can be applied in the variety of scientific fields such as decision making, pattern recognition, machine learning and market prediction, The different distance and similarity measures are based on Hamming distance, Euclidean distance, Hausdorff metric, normalized Hamming distance, normalized Euclidean distance, normalized Hausdorff distance, generalized Hamming distance, generalized Euclidean distance and generalized Hausdorff distance. The proposed algorithm is useful for document classification. We are calculating here accuracy and efficiency of document classification and result analysis. The accuracy and efficiency percentages calculated by MATLAB 7.14 Software tools. Outcomes are very efficient and accurate and give good outcomes to existing algorithms.

Acknowledgment

This research work is done at GHRIIT Research and Development Lab, Nagpur, M.H. India. Authors would like to thank to principal and staff for their helpful nature.

References

[1]. L.A. Zadeh, “Fuzzy Sets,” Inform. Control, vol. 8, pp.338–353, 1965.
[2]. D. Molodtsov, “Soft set theory—first results,” Comput. Math. Appl, vol.37, pp. 19–31, 1999.
[3]. P.K. Maji, R. Biswas and A.R. Roy, “Fuzzy soft sets,” J. Fuzzy Math. vol. 9, pp. 589–602, 2001.
[4]. P.K. Maji, R. Bismas and A.R. Roy, “Soft set theory,” Comput. Math.Appl. vol.45. pp. 555–562, 2003.
[5]. D. Molodtsov, “The Theory of Soft Sets,” URSS Publishers, Moscow, 2004.
[6]. D. Guha and D. Chakraborty, “A new approach to fuzzy distance measure and similarity measure between two generalized fuzzy numbers,” Applied Soft Computing vol. 10, pp. 90–99, 2010.
[7]. G.A. Papakostas “Distance and similarity measures between intuitionistic fuzzy sets:A comparative analysis rom a pattern recognition point of view,” Pattern Recognition Letters vol. 34, pp.1609–1622, 2013.
[8]. Chang Wang and Anjing Qu , “Entropy similarity measure and distance measure of vague soft sets and their relations,” Information Sciences vol. 244 pp. 92–106, 2013

[9]. H. Zhang and L. Yu, “New distance measures between intuitionist fuzzy sets and interval-valued fuzzy sets,” Information Sciences vol. 245 pp.181–196, 2013.
[10]. Z. Xu and M. Xia, “Distance and similarity measures for hesitant fuzzy sets,” Information Sciences vol. 181 pp. 2128–2138, 2011.
[11]. H. Liao, Z. Xu and X. Zeng, “Distance and similarity measures for hesitant fuzzy linguistic term sets and their application in multi-criteria decision making,” Information Sciences vol. 271 pp. 125–142, 2014.