A New Feature Selection Method based on Intuitionistic Fuzzy Entropy to Categorize Text Documents

M. B. Revanasiddappa, B. S. Harish *

Department of Information Science and Engineering, Sri Jayachamarajendra College of Engineering, Mysuru, Karnataka (India)

Abstract

Selection of highly discriminative feature in text document plays a major challenging role in categorization. Feature selection is an important task that involves dimensionality reduction of feature matrix, which in turn enhances the performance of categorization. This article presents a new feature selection method based on Intuitionistic Fuzzy Entropy (IFE) for Text Categorization. Firstly, Intuitionistic Fuzzy C-Means (IFCM) clustering method is employed to compute the intuitionistic membership values. The computed intuitionistic membership values are used to estimate intuitionistic fuzzy entropy via Match degree. Further, features with lower entropy values are selected to categorize the text documents. To find the efficacy of the proposed method, experiments are conducted on three standard benchmark datasets using three classifiers. F-measure is used to assess the performance of the classifiers. The proposed method shows impressive results as compared to other well known feature selection methods. Moreover, Intuitionistic Fuzzy Set (IFS) property addresses the uncertainty limitations of traditional fuzzy set.

Keywords

I. Introduction

In recent years, rapid development in internet technology generated massive amount of text documents by private and public sectors. To handle such a huge amount of text documents, automatic text categorization has become a popular technology to manipulate and manage [1][2]. Text Categorization (TC) is a process of automatically categorizing unknown text documents into one or more pre-defined classes by their contents. TC has been successfully proposed for many applications viz., retrieving useful information in search engines, spam filtering, document organization, automatic document indexing etc. Due to these applications automatic text categorization is an important research area in text mining and information retrieval [3][4]. Text categorization process includes feature extraction, pre-processing, feature selection and categorization. In feature extraction, the features are extracted from the text documents [5]. Each term (word) of the text document is considered as a feature and most of the features are unwanted and irrelevant. Further, during pre-processing, tokenization, stop-word elimination and stemming are employed to eliminate irrelevant and unwanted features [6]. The pre-processed text documents are represented in machine understandable form by employing a representation model. Then, feature selection method selects the most informative features from the representation model [7]. Feature selection plays high influence on the performance of classifiers and it is mainly used for dimensionality reduction [8][9]. Finally,

selected feature subset is fed into a classifier to categorize the text documents. The text categorization suffers from high dimensionality of feature space. Due to this, the performance of the classifiers degrades and also takes more time for categorization [10][11]. The reduction of high dimensional feature matrix is a significant challenge in text categorization. To tackle this challenge, many researchers have proposed various feature selection methods [8-9][12-15].

Feature selection methods are generally partitioned into 3 groups: filter, wrapper and hybrid [11][15]. Presently there are various feature selection methods reported in the literature viz., Document Frequency [14], Term Frequency-Inverse Document Frequency [16], Term Strength [14], Mutual Information [17], Information Gain [18], Chi-Square [11][13], Ambiguity Measure [19], Term Frequency-Relevance Frequency [20], Symbolic Feature Selection [21], Distinguish Feature Selection [12], Entropy based Feature Selection [22] and many more. Among these feature selection methods, entropy based feature selection method computes the amount of uncertainty and the quality of information content present in the text. The concept of entropy was introduced by Shannon [23] and is called as Shannon entropy (information entropy). It is fundamentally based on information theory, which estimates the entropy value of each feature using probability. A lower value of entropy indicates higher contribution of information in decision making process and larger value of entropy indicate lesser contribution of information. The concept of entropy is described in several manners and applied in different areas [24].

The extended version of Shannon entropy is fuzzy entropy, which is non-probabilistic entropy [25]. It adopts a new term named match-degree to estimate entropy value [26][27]. Match degree satisfies the four properties of de Luca-Termini axioms [25][27]. The probability of the entropy is computed using the number of occurring terms. In

* Corresponding author.

E-mail addresses: revan.cr.is@gmail.com (M. B. Revanasiddappa), bsharish@sjce.ac.in (B. S. Harish).

contrast, the match degree in fuzzy entropy is computed using the membership values of the occurring terms [26]. The fuzzy entropy value does not only consider the number of features, but also considers the actual distribution of features by membership function. Hence, fuzzy entropy reflects more information in the actual distribution of the features than Shannon entropy in the feature space [27].

Fuzzy Entropy is based on the fuzzy set, which measures the fuzziness. The Fuzzy set is used to solve various real world problems, which mainly deals with impreciseness and vagueness [29]. In fuzzy set theory, the membership value of a feature varies between '0' to '1'. Fuzzy set assumes that the non-membership is complementary of membership, but in real world this assumption fails due to hesitation. This hesitation originates, while defining the membership function, due to lack of precise knowledge and it is another type of uncertainty. To address this hesitation, Atanassov [30] developed an Intuitionistic Fuzzy Set (IFS), which is an enhanced version of Fuzzy Set. Unlike Fuzzy set which considers only the membership degree, the IFS considers the degree of non-membership and the hesitation along with degree of membership. In IFS, the degree of non-membership is less than or equal to the complement of the degree of membership due to the hesitation degree. The IFS gives mathematical model to deal with the vagueness arising from the inherent uncertainty and imprecision or insufficiency of imperfect information. Considering these advantages, many researchers have recently developed IFS based clustering techniques. Chaira [31] proposed a novel Intuitionistic fuzzy c-means (IFCM) clustering method. IFCM gives more accurate clusters, and shows better performance in the presence of uncertainty and it clusters the data with less iteration than the traditional FCM [32].

Motivated by the significant advantages of IFCM and Fuzzy Entropy, in this article a new feature selection method called Intuitionistic Fuzzy Entropy-Feature Selection (IFE-FS) is proposed for text categorization. This method selects feature subsets based on intuitionistic fuzzy entropy for text document categorization. Basically, it contains two phases: In the first phase, Intuitionistic membership degree is computed with the help of the IFCM clustering method. In second phase, the Intuitionistic fuzzy entropies on the basis of the Intuitionistic membership degree via match degree are estimated. Further, entropy values are arranged in ascending order and then the feature subset selection is pruned by the threshold value. Finally, selected feature subset is considered as input to the classifiers to categorize the text documents. More popular classifiers such as K-Nearest Neighbor (KNN) [33], Support Vector Machine (SVM) [34] and Radial Basis Function-Neural Network (RBF-NN) [35] are used. These classifiers are used due to their widespread use in the area of text mining and give competitive results on standard benchmark datasets [33]. The proposed method is experimented on 3 standard benchmark datasets viz., 20 Newsgroups, Reuters-21578 and TDT2. The performance of IFE-FS method is compared with CHI-Square ($\chi2$) [11], Mutual Information (MI) [7], Information Gain (IG) [18] and Entropy based Feature Selection (EFS) [22]. All these feature selection methods have their variant characteristics and also are the well-known feature selection methods for text categorization.

The main contribution of this article is as follows:

1. Proposes a new feature selection method (IFE-FS) based on Intuitionistic Fuzzy Entropy, which reduces high dimensionality of feature matrix and enhances the performance of classifiers.

2. Resolves uncertainty (hesitation) limitation of Fuzzy Entropy by making use of IFS.

3. Conducts extensive experiments on three standard benchmark datasets. The experimental results acknowledge that the IFE-FS method shows outstanding performance in terms of F-measure.

The rest of the article is organized as follows: existing and entropy based feature selection methods are reviewed in section II. The preliminary outlines on the Intuitionistic fuzzy set are presented in section III. The proposed method (IFE-FS) along with illustration is described in detail in section IV. The experimental results and discussion are presented in section V. Finally, the conclusion along with future work is presented in section VI.

II. Literature Review

High dimensionality of the feature matrix is a major challenge during text categorization. Feature selection plays an effective role to identify the relevant features to minimize dimensionality. Feature selection method gives reduced text feature collection, reduces storage size, lesser model building and computation time, and better model interpretation [9][15][36]. However, many studies [37][38] [39] indicate that there are no feature selection methods to provide accurate discriminative feature for text categorization. Given the importance of feature selection for text categorization, many methods were proposed and some of them are presented below.

Recent researchers adopted the concept of entropy in feature selection methods to select the discriminate features. The concept of Entropy is based on the information theory proposed by Shannon which is also called as Shannon Entropy [23]. It measures the expected uncertainty of probability distribution with the results predicted by random experiments. Tang et al., [40] proposed a feature selection method based on the information theory, which uses Jeffreys-Multi-Hypothesis (JMH) divergence information measure. This method ranks the original features and maximizes the discriminative capacity for text categorization. Largeron et al., [22] proposed a feature selection method called Entropy based Category Coverage Difference (ECCD), which is based on the entropy. This method computes the occurrence of terms inside various classes with the assistance of entropy. Cai and Song [41] used the maximum entropy modeling with different feature selection methods to categorize text documents and also proposed novel feature selection method named as "count difference". This method considers features of both relevant and irrelevant classes to compute the frequency differences between relative documents. Vaghela et al., [42] presented an entropy based feature selection method, which uses InfoDist and Pearson's Correlation parameters. This method selects the features using InfoDist, which adopts conditional entropy to compute the relevancy of feature and category. Further, it eliminates the irrelevant features using Pearson's correlation. Liu and Song [43] presented the portion set of key words based on proximity degree. Key words are selected based on entropy, semantic field and association degree. Later, fuzzy classification is used to categorize documents.

Many researchers have identified the significance of entropy and have developed fuzzy entropy measures from various perspectives [44]. Fuzzy entropy has been widely used in pattern recognition, image processing and clustering analysis. De and Termini [25] proposed the first non-probabilistic entropy based on the fuzzy set theory. Khushaba et al., [45] described the fuzzy entropy in terms of match degree. The fuzzy entropy uses match degree instead of probability to estimate value of the entropy. Parkash et al., [46] used the principle of maximum weighted fuzzy entropy to develop two new weighted fuzzy entropy measures to remove the redundancy. Luukka [47] proposed a feature selection method based on fuzzy entropy, in which fuzzy entropy is estimated by using the membership degree in fuzzy set. This method minimizes the dimensionality of feature space and also enhances the efficiency of classifiers. Ahmadizar et al., [48] proposed a hybrid feature selection method, which has two stages. In the first stage, dimensionality of the feature matrix is reduced using the fuzzy entropy feature selection method. Fuzzy entropy values are ranked in ascending order and then the feature subset with lowest entropy value is selected. In the second stage, ant colony optimization is used to

select the features from feature subset for text categorization.

All the above mentioned methods fail to handle uncertainty that arises while defining the membership function. To handle this uncertainty, Atanassov [30] developed Intuitionistic Fuzzy Set (IFS) in 1986, which is an advanced version of fuzzy set. IFS considered degree of non-membership and hesitation along with degree of membership. In IFS, non-membership degree is computed by employing intuitionistic fuzzy complement generator [49][50]. Intarapaiboon [51] proposed a framework for text categorization based on similarity measure of intuitionistic fuzzy sets. In this framework, each document is represented in terms of the IFS, where the IFS uses sugeno's integral method to represent the document. Different types of similarity measures using IFS are described in [52][53]. Szmidt and Kacprzyk [54] proposed a feature selection method based on the IFS for text categorization. This method uses the degree of membership, non-membership and hesitation to resolve the imbalance and overlap class problem of categorization.

Chaira [31] proposed a novel Intuitionistic fuzzy c-means (IFCM) clustering method. The IFCM was evaluated on various CT scan brain images and achieved better performance than Type 2 fuzzy and traditional Fuzzy c-means algorithms. In literature, many researchers applied IFCM clustering method to solve image processing problems [55-58]. However, very less amount of work is reported on text categorization [51][54]. In addition to clustering methods, researchers developed entropy methods based on IFS. Szmidt and Kacprzyk [59] depicted entropy as far as non-probabilistic type of entropy. Burillo and Bustince [60] depicted entropy as far as degree of intuitionism of an intuitionistic fuzzy set. Hung and Yang [61] proposed two IFS based entropy measures and provided the axiomatic definition of entropy for IFS. Vlachos and Sergiadis [62] proposed new Intuionistic Fuzzy Entropy and also presented the connection between intuitive and mathematics of entropy for fuzzy set and IFS.

From literature survey, it is observed that a lot of work is reported on entropy and fuzzy entropy based feature selection methods for text categorization. In addition, it is also observed from the literatures that IFS addresses the limitation of fuzzy set by considering the hesitation degree. In this article, we developed a new Feature Selection method based on Intuitionistic Fuzzy Entropy (IFE). To the best of our knowledge this work is first of its kind, where a feature selection method based on IFE to categorize the text documents is proposed. The next section describes the construction of Intuitionistic Fuzzy Set for IFCM.

III. Construction of Intuitionistic Fuzzy Set (IFS)

In this section, the mathematical background on IFS is explained. The Intuitionistic Fuzzy Set, in general sense is portrayed by the membership degree, the non-membership degree and the hesitation degree [30].

Generally the fuzzy set A is defined on text document D_i as

$$A = \left\{ \langle T_l, \mu_A(T_l) \rangle \mid T_l \in D_i \right\} \tag{1}$$

Where, $\mu_A(T_l) \to [0,1]$ denotes the membership degree of A. T_l is the lth feature in the ith document D_i. The membership value $\mu_A(T_l)$ defines the degree of belongingness of $T_l \in D_i$ in A. An Intuitionistic fuzzy set $\tilde{A}$ is described on document D_i as:

$$\tilde{A} = \left\{ \langle T_l, \mu_{\tilde{A}}(T_l), \vartheta_{\tilde{A}}(T_l) \rangle \mid T_l \in D_i \right\} \tag{2}$$

Where, $\mu_{\tilde{A}}(T_l)$ and $\vartheta_{\tilde{A}}(T_l)$ denotes membership and non-membership degree of lth feature T_l respectively. In the fuzzy set, another uncertainty emerges while defining the membership function due to imprecise knowledge. IFS handles this uncertainty by considering the hesitation degree. Unlike in fuzzy set, where the non-membership degree is calculated by taking the complement of membership degree, IFS computes the non-membership degree with the help of intuitionistic fuzzy complement generator. In this work, the Sugeno's intuitionistic fuzzy complement generator is used to calculate the non-membership degree [49]. The Sugeno's intuitionistic fuzzy complement is computed as

$$\vartheta_{\tilde{A}}(T_l) = \frac{1 - \mu_{\tilde{A}}(T_l)}{1 + \lambda \mu_{\tilde{A}}(T_l)} \tag{3}$$

Where, λ is the constant and $\lambda > 0$. When we set $\lambda = 1$, IFS becomes the traditional fuzzy set. Hesitation degree of a term $T_l \in D_i$ in $\tilde{A}$ is computed as

$$\pi_{\tilde{A}}(T_l) = 1 - \mu_{\tilde{A}}(T_l) - \vartheta_{\tilde{A}}(T_l) \tag{4}$$

Where $\pi_{\tilde{A}}(T_l)$ is the hesitation degree of lth feature T_l. It is evident from the equation (4)

$$0 \leq \pi_{\tilde{A}}(T_l) \leq 1, \, \forall T_l \in D_i$$

Non-membership degree is computed from Sugeno's intuitionistic fuzzy complement. Thus, using Sugeno's intuitionistic fuzzy complement, the IFS becomes

$$\tilde{A}_\lambda = \left\{ \left\langle T_l, \mu_{\tilde{A}}(T_l), \frac{1 - \mu_{\tilde{A}}(T_l)}{1 + \lambda \mu_{\tilde{A}}(T_l)} \right\rangle \mid T_l \in D_i \right\} \tag{5}$$

IV. Proposed Method

A new feature selection method based on Intuitionistic Fuzzy Entropy (IFE) named as Intuitionistic Fuzzy Entropy-Feature Selection (IFE-FS) is proposed in this section. The Intuitionistic Fuzzy Entropy estimates the value of entropy on the basis of intuitionistic membership degree via match degree. The intuitionistic membership degrees are computed from the Intuitionistic Fuzzy C-Means (IFCM) Clustering method. The IFCM is based on the Intuitionistic Fuzzy Set (IFS), which considers the degree of membership, non-membership and hesitation. The proposed method consists of two phases: in the first phase, an intuitionistic membership value is computed using IFCM. In the second phase, entropy is estimated using the intuitionistic membership values and a feature subset is selected on the basis of the entropy value.

A. Intuitionistic Fuzzy C-Means (IFCM) Clustering Method

Intuitionistic Fuzzy C-Means (IFCM) is an advanced version of the traditional Fuzzy C-means (FCM) and it is based on the IFS. Unlike, FCM which clusters input document based on the membership value, IFCM considers the non-membership and hesitation degree along with the membership degree.

Let us assume that there are k number of pre-defined classes $C_a, a = 1, 2, 3, ..., k$. Each class contains n number of documents $D_i, i = 1, 2, 3, ..., n$, and m number of features (terms) $T_l, l = 1, 2, 3, ..., m$. Then the total number of documents is denoted by $z = [k \times n]$. The text document contains sequence of terms (features), where each term is treated as a feature. In text document, most of the features are irrelevant and redundant. These features results in high dimension in feature space and also degrade the performance of classifier. It is necessary to use pre-processing techniques to eliminate irrelevant and redundant features. The most common pre-processing task is

tokenization, elimination of stop-word and stemming. Tokenization is the process of partitioning the text into terms (features), called tokens. In the process of stop-word elimination, the features that do not have important information are eliminated from the feature space. For example, features like: "a", "is", "the", etc., occur very frequently in all text documents and do not convey any meaning for class prediction. The process of stemming is to reduce the inflected terms to their root form. The stemming process helps to group the frequencies of different inflection to single term. For example, the features: "loves", "loved" and "loving" have the similar meaning as of its root "love". Further, the preprocessed text documents are represented using the Term Document Matrix (TDM) form. The TDM is considered as the input to the IFCM clustering method. IFCM assigns membership values to a document with respect to each class. But in this work, the aim is to compute the intuitionistic fuzzy entropy of each feature with respect to each class. So in order to compute the intuitionistic fuzzy entropy of each feature, we applied IFCM to each feature rather than the document. The objective function (J) of IFCM is as follows

$$J(U,V) = \sum_{j=1}^{c}\sum_{l=1}^{m}\mu_{jl}^{p}d_{jl}^{2}$$

(6)

Where, U is the membership matrix, V indicates the cluster centers matrix, μ_{jl} is the intuitionistic membership degree of lth term in j^{th} cluster, p is the fuzzy coefficient. $d_{jl} = d(T_l, v_j)$ describes the distance measure between cluster center v_j and term T_l, v_j means j^{th} cluster center, T_l means l^{th} term and c is number of cluster. The intuitionistic fuzzy membership (μ_{jl}) is the combination of membership degree (μ_{jl}^{*}) and hesitation degree (π_{jl}).

$$\mu_{jl} = \mu_{jl}^{*} + \pi_{jl}$$

(7)

The membership degree μ_{jl}^{*} is computed as

$$\mu_{jl}^{*} = \frac{1}{\sum_{g=1}^{c}\left[\frac{d_{lj}^{2}}{d_{lg}^{2}}\right]^{1/(p-1)}}.$$

(8)

In order to compute the hesitation degree, firstly the non-membership degree (ϑ_{jl}) has to be computed by using the intuitionistic fuzzy complement generator. In this article, we used the Sugeno's intuitionistic fuzzy complement generator [49] to compute the non-membership degree. The non-membership degree (ϑ_{jl}) is computed using the equation

$$\vartheta_{jl} = \frac{1-\mu_{jl}^{*}}{1+\lambda\mu_{jl}^{*}}.$$

(9)

Where, λ is the constant and $\lambda > 0$. The value of non-membership degree changes by varying the value of λ. The hesitation degree computed using the membership degree (equation (8)) and the non-membership degree (equation (9)) is given by

$$\pi_{jl} = 1 - \mu_{jl}^{*} - \vartheta_{jl}$$

(10)

The Intuitionistic membership degree is computed using the equation (7) and further cluster center is calculated using the following equation

$$v_j = \frac{\sum_{l=1}^{m}\mu_{jl}^{*}T_l}{\sum_{l=1}^{m}\mu_{jl}^{*}}$$

(11)

The IFCM optimizes the objective function by continuously updating the membership degree and the cluster center until it meets the convergence criteria value ε, i.e., $\left|J^{(t)} - J^{(t-1)}\right| < \varepsilon$. Here, t is the iteration and ε is the user specified convergence criteria. Further, using this membership value, match degree is computed to estimate the entropy.

B. Intuitionistic Fuzzy Entropy (IFE)

The Intuitionistic Fuzzy Entropy is the non-probabilistic entropy in which we use match degree to estimate the entropy value. The match degree in IFE is computed using the intuitionistic membership value. IFE maximizes the capacity of discriminative features and generates more affluent information. Match degree is described as

$$X_{al} = \frac{\mu_a(T_l)}{\sum_{a=1}^{k}\mu_a(T_l)}$$

(12)

Where, $\mu_a(T_l)$ indicates the membership of lth feature T_l in a^{th} class C_a. The match degree X_{al} is the ratio of intuitionistic membership of each feature T_l in class and by the summation of intuitionistic membership of feature T_l in all classes. The match degree X_{al} in intuitionistic fuzzy entropy uses the membership values to calculate the matching degree of feature in class. Later the Intuitionistic fuzzy entropy (IFE) of feature T_l in class C_a is computed using the following equation

$$IFE_{al} = -X_{al}\log X_{al}$$

(13)

Further, each features' IFE value of class is summed to calculate the overall Intuitionistic fuzzy entropy of each feature

$$IFE_l = \sum_{a=1}^{k}IFE_{al}$$

(14)

The lower entropy value of a feature indicates major contribution about the classes. Therefore, we select r number of low ranked features. Here, the threshold value (r) indicates the number of features selected and also r varies depending on the datasets. But during experimentation, the major issue is in choosing an accurate threshold value. It will be varied multiple times before selecting the best value. The selected feature subset is then given to the classifiers to categorize the text documents. We used three different classifiers: K-Nearest Neighbor (KNN), Support Vector Machine (SVM) and Radial Basis Function Neural Network (RBF-NN). The performance of IFE-FS method is measured based on the F-measure. Algorithm 1 explains the individual steps involved in the proposed method.

In order to provide more meaningful theory of our proposed method, the next section illustrates the proposed method in detail by considering a simple term document matrix.

Algorithm 1: Intuitionistic Fuzzy Entropy Feature Selection (IFE-FS)

Data: k Number of class with n number of documents and m number of terms (features), Fuzzy Coefficient (p), Convergence Criteria (ε), λ is constant and Threshold (r)

Result: Class label

Step 1: Initialize cluster centers v_j randomly

Initialize number of iteration t=0

Repeat

Step 2: Compute membership degree using equation (8)

Step 3: Compute non-membership degree using equation (9)

Step 4: Compute hesitation degree using equation (10)

Step 5: Compute Intuitionistic Fuzzy membership degree using equation (7)

Step 6: Compute Intuitionistic Fuzzy cluster center using equation (11)

Step 7: Compute objective function using equation (6)

Until $\left| J^{(t)} - J^{(t-1)} \right| < \varepsilon$ is satisfied

Step 8: Compute match degree using equation (12)

Step 9: Compute class-wise Intuitionistic Fuzzy entropy for each feature using equation (13)

Step 10: Compute overall Intuitionistic Fuzzy entropy for each feature using equation (14)

Step 11: Select feature subset by using Threshold value (r)

Step 12: Selected feature subset will be the input to classifiers

---

C. Illustration

This section illustrates the individual steps involved in the proposed feature selection method. For illustration we considered an example of Term Document Matrix (TDM) of size 10 x 8, where 10 documents (D) are distributed among 3 classes (C) with 8 unique features (T). The same is presented in Table I.

The proposed feature selection method is based on the Intuitionistic Fuzzy Entropy, which uses the Intuitionistic Fuzzy C-Means (IFCM) clustering method to compute the intuitionistic membership degree for entropy estimation. So we computed each features' membership degree rather than that of the documents in all classes. The objective function of IFCM is shown in equation (7).

TABLE I. Term Document Matrix (TDM)

Documents	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	Classes
D_1	2	2	3	0	0	8	2	1	
D_2	3	1	3	0	0	0	1	6	C_1
D_3	1	1	1	1	1	1	0	1	
D_4	1	1	1	1	1	1	1	0	
D_5	1	1	2	0	0	0	0	1	
D_6	1	1	1	0	3	0	1	0	C_2
D_7	1	1	1	1	5	1	1	2	
D_8	1	1	1	1	1	0	0	1	
D_9	1	1	1	0	3	0	1	0	C_3
D_{10}	1	1	1	1	2	1	1	1	

In step 1, we initialized the cluster centers randomly. Since there are 3 classes and 10 documents, the matrix size will be 3 x 10, which is presented in Table II. Using these cluster centers, the membership degree is calculated.

In step 2, each feature membership degree is computed using equation (9) with respect to the class. Here, we set Fuzzy Coefficient p=2. The dimension of membership matrix is 3×8 and it is shown in Table III.

In step 3, the non-membership degree is computed using the Sugeno's Fuzzy Complement generator, which is mentioned in equation (10). Here, for illustration we set $\lambda = 0.5$. The obtained matrix will be in the form 3×8 as shown in Table IV.

In step 4, the hesitation degree is computed, using the membership and non-membership degree, using equation (11). The obtained resultant matrix is presented in Table V. Further, the intuitionistic fuzzy membership degree is computed using equation (8) in step 5 and the membership values are presented in Table VI.

After the 1st iteration, the cluster centers are updated using equation (12) in step 6. Now the updated cluster centers are shown in Table VII.

The objective function value after 1st iteration is J=77.2455, which is computed using equation (7) in step 7. Similarly, same steps are repeated for every iteration. After the 15th iteration, the computed intuitionstic membership degrees and cluster centers are presented in Table VIII and Table IX respectively, and the objective function value is 76.6000. The steps 2 to 7 are repeated until it satisfies the user defined convergence criteria.

TABLE II. Cluster Centers

D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	Classes
2.3709	2.3669	0.9335	0.8242	0.8542	0.6182	1.3264	0.8154	0.6182	1.0377	C_1
1.6367	1.0015	0.6376	0.9946	0.3689	0.7297	1.0364	0.6127	0.7297	1.0077	C_2
2.1788	1.5690	0.9856	0.8224	0.4923	1.4360	2.7697	0.8068	1.4360	1.3980	C_3

TABLE III. Membership Degree Matrix

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	Classes
0.7263	0.2395	0.6315	0.2294	0.2302	0.3569	0.1748	0.4503	C_1
0.1508	0.6232	0.1821	0.5726	0.2797	0.3270	0.6864	0.2641	C_2
0.1229	0.1373	0.1864	0.1980	0.4901	0.3161	0.1388	0.2856	C_3

TABLE IV. Non-Membership Degree Matrix

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	Classes
0.1556	0.6189	0.2292	0.6312	0.6302	0.4860	0.7006	0.3910	C_1
0.7328	0.2360	0.6910	0.2787	0.5715	0.5183	0.1856	0.5896	C_2
0.7721	0.7516	0.6854	0.6704	0.3530	0.5302	0.7495	0.5647	C_3

TABLE V. Hesitation Degree Matrix

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	Classes
0.1181	0.1416	0.1393	0.1394	0.1396	0.1571	0.1246	0.1587	C_1
0.1164	0.1408	0.1269	0.1487	0.1488	0.1547	0.1280	0.1463	C_2
0.1050	0.1111	0.1282	0.1316	0.1569	0.1536	0.1117	0.1497	C_3

TABLE VI. Intuitionistic Membership Degree Matrix

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	Classes
0.8444	0.3811	0.7708	0.3688	0.3698	0.5140	0.2994	0.6090	C_1
0.2672	0.7640	0.3090	0.7213	0.4285	0.4817	0.8144	0.4104	C_2
0.2279	0.2484	0.3146	0.3296	0.6470	0.4698	0.2505	0.4353	C_3

TABLE VII. Updated Cluster Centers

D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	Classes
2.5155	2.6050	0.9634	0.8486	0.9868	0.7969	1.3746	0.8556	0.7969	1.0558	C_1
1.9649	1.0953	0.7366	0.9331	0.4029	0.7802	1.3586	0.6445	0.7802	1.0729	C_2
2.1477	1.4133	0.9483	0.8438	0.4132	1.2625	2.5368	0.7663	1.2625	1.3451	C_3

TABLE VIII. Intuitionistic Membership Degree Matrix After 15th Iteration

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	Classes
0.8444	0.3811	0.7708	0.3688	0.3698	0.5140	0.2994	0.6090	C_1
0.2672	0.7640	0.3090	0.7213	0.4285	0.4817	0.8144	0.4104	C_2
0.2279	0.2484	0.3146	0.3296	0.6470	0.4698	0.2505	0.4353	C_3

TABLE IX. Cluster Centers of 16th Iteration

D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	Classes
2.5155	2.6050	0.9634	0.8486	0.9868	0.7969	1.3746	0.8556	0.7969	1.0558	C_1
1.9649	1.0953	0.7366	0.9331	0.4029	0.7802	1.3586	0.6445	0.7802	1.0729	C_2
2.1477	1.4133	0.9483	0.8438	0.4132	1.2625	2.5368	0.7663	1.2625	1.3451	C_3

TABLE X. Intuitionistic Membership Degree Matrix at 32nd Iteration

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	Classes
0.8444	0.3811	0.7708	0.3688	0.3698	0.5140	0.2994	0.6090	C_1
0.2672	0.7640	0.3090	0.7213	0.4285	0.4817	0.8144	0.4104	C_2
0.2279	0.2484	0.3146	0.3296	0.6470	0.4698	0.2505	0.4353	C_3

TABLE XI. Cluster Centers at 32nd Iteration

D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	Classes
2.3891	2.8796	0.9821	0.8299	1.0720	0.8305	1.4171	0.9003	0.8305	1.0617	C_1
2.1315	1.0463	0.7994	0.9161	0.3644	0.9130	1.7058	0.6594	0.9130	1.1555	C_2
2.1289	1.0456	0.7998	0.9160	0.3634	0.9137	1.7088	0.6598	0.9137	1.1562	C_3

In this illustration, the algorithm converges in the 32nd iteration. The final Intuitionistic membership values and cluster centers are shown in Table X and Table XI respectively and the objective function value is 76.6067.

In step 8, the final Intuitionistic membership values are given to the match degree to compute the actual distribution of features with respect to the class by using equation (13). The obtained match degree of the feature is shown in Table XII.

In step 9, class-wise Intuitionistic Fuzzy Entropy is estimated for each feature with respect to the class by using equation (14) and the estimated entropy value are presented in Table XIII. Later, all rows are summed up to compute the overall Intuitionistic Fuzzy Entropy for each feature by using equation (15), which is shown in Table XIV with a dimension of 1×8.

The obtained entropy value of each feature is sorted in ascending order to select the features with lower entropy values, which is shown in Table XV. The features are ranked $T_1 \succ T_3 \succ T_7 \succ T_2 \succ T_8 \succ T_4 \succ T_5 \succ T_6$ based on their entropy value. The lower entropy

TABLE XII. MATCH DEGREE MATRIX

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	Classes
0.7467	0.1616	0.6273	0.2249	0.2733	0.3135	0.1517	0.4597	C_1
0.1266	0.4197	0.1864	0.3874	0.3631	0.3433	0.4244	0.2701	C_2
0.1265	0.4186	0.1861	0.3876	0.3635	0.3431	0.4237	0.2701	C_3

Table XIII. INTUITIONISTIC FUZZY ENTROPY

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	Classes
0.0946	0.1279	0.1270	0.1457	0.1539	0.1579	0.1242	0.1551	C_1
0.1136	0.1582	0.1359	0.1595	0.1597	0.1594	0.1579	0.1535	C_2
0.1136	0.1583	0.1359	0.1595	0.1597	0.1593	0.1580	0.1535	C_3

Table XIV. OVERALL INTUITIONISTIC FUZZY ENTROPY

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8
0.3219	0.4444	0.3989	0.4648	0.4734	0.4767	0.4402	0.4622

TABLE XV. FEATURES RANKED IN ASCENDING ORDER

0.3219	0.3989	0.4402	0.4444	0.4622	0.4648	0.4734	0.4767
T_1	T_3	T_7	T_2	T_8	T_4	T_5	T_6

TABLE XVI. SELECTED DISCRIMINATIVE FEATURES

T_1	T_3	T_7
0.3219	0.3989	0.4402

value indicates higher relevant feature with respect to the class and contributes with more information. The higher entropy value indicates less contribution to the class. Thus, we select lower entropy values for categorization.

In next step, we select the number of features based on Threshold value (r). The Threshold value (r) is used to prune the feature subset. The r value is selected after multiple iterations by considering distinct values for r times and we found that r =3 is the most suitable value for the given illustration. Here r value is less than total number of features. The selected discriminative features are presented in Table XVI.

Finally, the selected feature subset is considered as the input to the classifiers to categorize the text documents

V. EXPERIMENTS

This section presents detailed experimentations carried out to demonstrate the efficiency of the proposed method.

A. Datasets Description

In order to assess the effectiveness of the Intuitionstic Fuzzy Entropy-Feature Selection (IFE-FS) method, experiments are conducted on 20-Newsgroups [63], Reuters-21578 [64] and TDT2 [65] standard benchmark datasets. The 20-Newsgroups dataset is a set of 18846 online newsgroup documents, split evenly into 20 different classes with 26214 features. The Reuters-21578 dataset is from newswire, which contains 8293 documents and is non-uniformly divided into 65 categories with 18933 features. The TDT2 (Topic Detection and Tracking) dataset consists of 9394 documents and 36771 features,

spread across 30 categories. In each dataset, distribution of features is varied according to their corresponding classes.

B. Experimental Setup

During the experimentation, it is necessary to split the dataset into training and testing set (to validate the proposed method). The large training data results in overfitting of the model. On the other hand, small training data results in underfitting the model. The whole reason for split comes from the fact that, we often have limited and finite data. So we want to make the best use of it and train on as much data as we can. In this paper, to validate the proposed IFE-FS method, we split the dataset into training and testing set in 60:40 ratios respectively. The training set is 60% documents of each class of dataset, used to build our proposed method. On the other hand, 40% testing set is applied on proposed model to assess the performance.

The performance of the proposed method mainly depends on the IFCM parameter values. Authors in [31] investigated the parameter values of IFCM. Based on the backdrop of [31], in this paper we initialized the parameters as follows: fuzzy coefficient p =2, convergence criteria $\varepsilon = 0.0001$ and λ =0.5. The performance of the IFE-FS method is compared against the four widely used feature selection methods viz., Chi-Square, Mutual Information (MI), Information Gain (IG) and Entropy based Feature Selection (EFS). We conducted experiments on three standard benchmark datasets using KNN, SVM, and RBF-NN classifiers. We used the F-Measure metric to assess the performance of the classifiers. The F-Measure is widely used in text categorization, which indicates the overall categorization performance and also combined effectiveness measure determined by precision and recall.

A number of features are selected based on the threshold value (r), where r indicates the number of features. Initially, we conducted experiments by fixing r value as 500 empirically. Further, we varied the value of r from 500 to 7000, with an increment of 500. However, decrease in the value of r below 500 and increase in value of r above 7000, does not yield good results. Hence, we restricted the value of r between 500 to 7000.

C. Experimental Results

Initially, we conducted the experiment without employing feature selection method. The number of features obtained for 20-Newsgroups is 26214, for Reuters-21578 there is 18933 and TDT2 has 36771 features. The F-measure using KNN, SVM and RBF-NN classifiers on three standard benchmark datasets are presented in Table XVII.

Fig. 1 depicts the categorization performance of the proposed method (IFE-FS) using KNN classifier in terms of F- measure on 20-NewsGroups dataset. It can be observed that each feature selection method obtained their best results with variant number of features, in terms of F-measure. The IFE-FS method achieved better result of 0.662 for 2500 features compared to other feature selection methods using KNN classifier. Our proposed feature selection method identifies discriminative features when the value of r is 2500, which leads to achieve the maximum result. Similarly, we conducted the same set of experiments using SVM and RBF-NN classifiers on 20-NewsGroups.

Fig.1. Performance comparisons using KNN classifier (20-NewsGroups dataset).

Fig. 2 and Fig. 3 show the performance comparison of the IFE-FS method with existing feature selection methods on 20-NewsGroups for SVM & RBF-NN classifier. The proposed IFE-FS method performed better compared to existing feature selection methods, when the number of features are ranging from 4500 to 7000. In Fig. 3, the performance of RBF-NN based on MI, IG and Entropy, completely coincides with each other. However, the F-measure curve of RBF-NN based on the proposed IFE-FS method is significantly higher than that of the existing feature selection methods. It is evident from Fig, 1, 2 and 3 that the performance of IFE-FS method in terms of F-measure is superior to that of the existing feature selection methods when value of r is 2500 for KNN, 5000 for SVM and 5500 for RBF-NN classifiers.

Further, the same set of experimentations were carried out on Reuters-21578 and TDT2 dataset. In Reuters-21578, the performance of the proposed IFE-FS method with that of the existing feature selection methods using KNN, SVM and RBF-NN classifiers are shown in Fig. 4, 5 and 6, respectively. It can be observed from Fig. 4-6 that the performance of the proposed IFE-FS method is lower than other feature selection methods, when the number of features ranging from 500 to 2500.
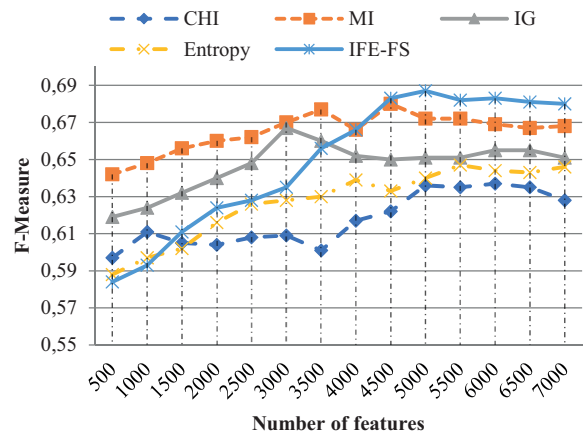

Fig. 2. Performance comparisons using SVM classifier (20-NewsGroups dataset).
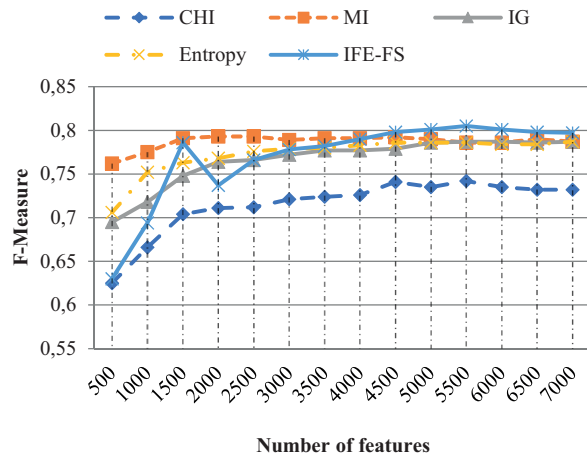

Fig. 3. Performance comparisons using RBF-NN classifier (20-NewsGroups dataset).

From Fig. 4-6, we can note that the performance of IFE-FS method in terms of F-measure is significantly higher than other existing feature selection methods, when the value of r is ranging from 2500 to 7000. From Fig. 4, 5 and 6, the KNN classifier obtains a F-measure of 0.711 when r =2500, the SVM classifier obtains a F-measure of 0.751 when r =2500 and RBF-NN classifier obtains a F-measure of 0.874 when r =6000.
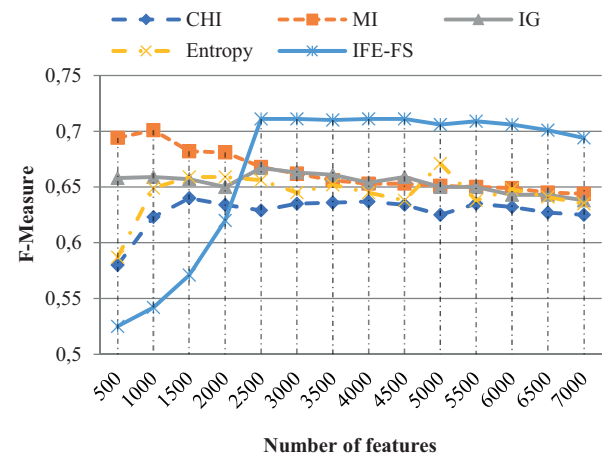

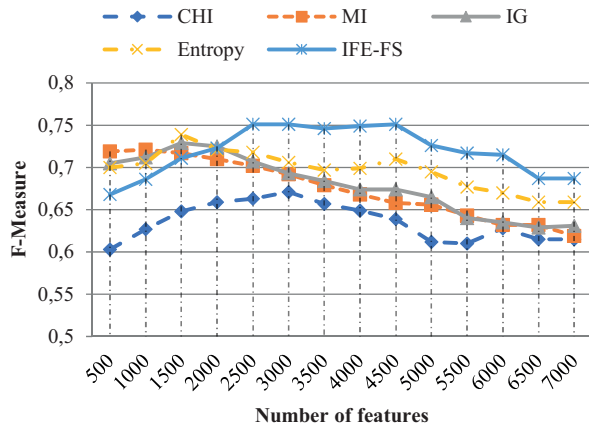Fig. 4. Performance comparisons using KNN classifier (Reuters-21578 dataset).

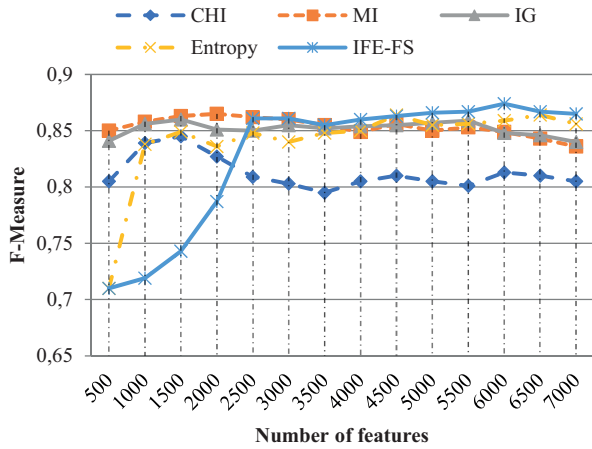Fig. 5. Performance comparisons using SVM classifier (Reuters-21578 dataset).

Fig. 7. Performance comparisons using KNN classifier (TDT2 dataset).

Fig. 6. Performance comparisons using RBF-NN classifier (Reuters-21578 dataset).

Fig. 8. Performance comparisons using SVM classifier (TDT2 dataset).

Fig. 7, 8 and 9 show the comparison of the proposed Intuitionstic Fuzzy Entropy-Feature Selection (IFE-FS) method with the existing feature selection method using KNN, SVM and RBF-NN classifiers on TDT2 dataset. It can be observed from Fig. 7-9, that the proposed IFE-FS method outperforms other existing feature selection methods. Fig. 7 shows that F-measure curve of KNN classifier based on the proposed IFE-FS method is higher than that of the other existing feature selection methods, when the value of r is ranging from 2500 to 6000. Similarly, the proposed IFE-FS method achieved best result when value of r is ranging from 3000 to 6000 using SVM. From Fig. 9, we can observe that the proposed IFE-FS method achieved best result when the value of r is in the range of 5500 and 6000 using RBF-NN classifier. The proposed IFE-FS method obtained a result of 0.837 using KNN classifier when the value of r is 3500, SVM achieved 0.915 result when the value of r is 3500 and RBF-NN achieved 0.967 result when the value of r is 5500, which are presented in Fig. 7, 8 and 9.

From Fig. 1-9, we can infer that, in terms of F-measure, the performance of IFE-FS method improved when the value of r is ranging from 2500 to 6000. In addition, the proposed IFE-FS method identifies discriminative features when the value of r is ranging from 2500 to 6000, which lead to achieve maximum result for all three classifiers. Table XVII presents the F-measure results using KNN, SVM and RBF-NN classifiers without feature selection method on the three datasets. From Table XVII and Fig. 1-9, we can observe that the result of the proposed IFE-FS method with all three classifiers is superior than the results of the same classifier without feature selection.
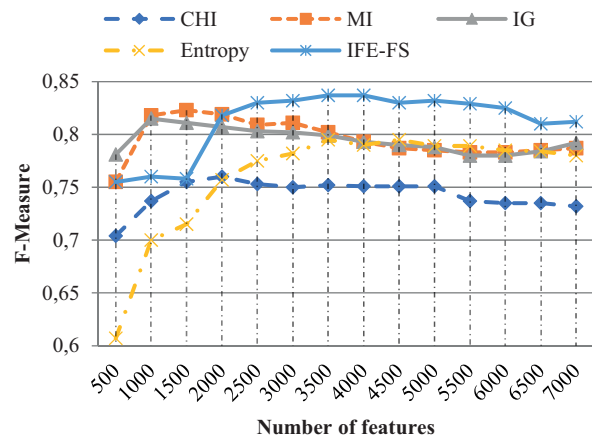

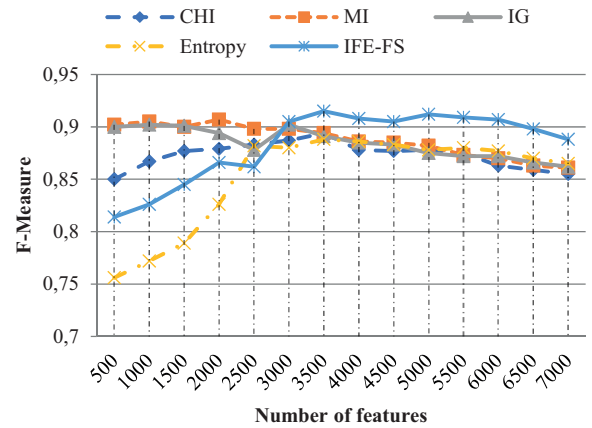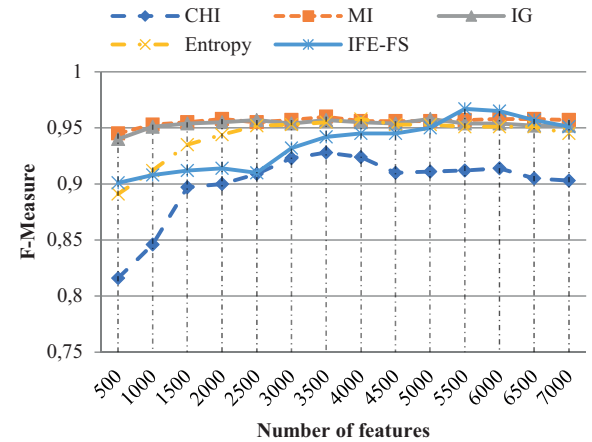Fig. 9. Performance comparisons using RBF-NN classifier (TDT2 dataset).

Besides, the following observations were made during the experimentation:

- From experimental results, it is noted that the IFE-FS method obtained better results by considering less number of features from the original feature set.
- From Fig. 1-9, the proposed IFE-FS method achieved significantly better results on all standard benchmark datasets using KNN, SVM and RBF-NN classifiers.

TABLE XVII. Performance of Classifiers with and Without Feature Selection Method on Three Datasets

Dataset	Classifiers	Without Feature Selection		With Feature Selection	
		Number of Features Selected	F-Measure	Number of Features Selected	F-Measure
20-Newsgroups	KNN	26214	0.469	2500	0.662
	SVM	26214	0.618	5000	0.687
	RBF-NN	26214	0.716	5500	0.805
Reuters-21578	KNN	18933	0.484	2500	0.711
	SVM	18933	0.570	2500	0.751
	RBF-NN	18933	0.810	6000	0.874
TDT2	KNN	36771	0.745	3500	0.837
	SVM	36771	0.745	3500	0.915
	RBF-NN	36771	0.891	5500	0.967

D. Discussions

The proposed IFE-FS method performance is measured in terms of F-Measure. It is evident from Fig, 1-9 that the proposed IFE-FS method performs superior compared to Chi-Square, MI, IG and Entropy based Feature Selection (EFS) methods using KNN, SVM and RBF-NN classifiers on the three standard datasets. The Chi-Square, estimates the lack of independence between terms and class. The proposed IFE-FS method computes the distribution of terms in each class. Thus, the proposed IFE-FS method obtained higher result than Chi-Square on all the three datasets. The MI estimates the mutual dependency of two terms using joint probability distribution and marginal probability distribution. The proposed IFE-FS method is non-probabilistic, which estimates the intuitionistic fuzzy information by intuitionistic fuzzy set. IG evaluates the quantity of bits of information acquired by knowing the presence or absence of a term in the document for categorization prediction. However, the proposed IFE-FS method does not consider absence or presence of term in a document.

On the other hand, the EFS method considers only randomness uncertainty and estimates the entropy of the terms in a class, based on probability. However, the proposed IFE-FS method considers uncertainty like randomness, ambiguity and vagueness to provide the terms with importance in the class. Moreover, it maximizes the discriminative capacity of the features and produces rich information. The proposed IFE-FS method shows higher performance using RBF-NN classifier compared to other KNN and SVM classifier. It can be concluded from the experiments that the proposed IFE-FS method shows significant improved performance using classifiers by selecting a number of discriminative features based on the distribution of terms in the classes. The proposed IFE-FS method performs better compared to other well known feature selection methods. However, the performance of the proposed method mainly depends on the number of features, initial cluster centers and parameter values. These parameter values differ across the datasets and it makes very difficult task to select the parameter values.

VI. Conclusion

In this article, a new feature selection method called Intuitionistic Fuzzy Entropy-Feature Selection (IFE-FS) is proposed for text categorization. The IFE is based on the intuitionistic fuzzy set. Unlike traditional Fuzzy Entropy, which considers only membership degree, Intuitionistic Fuzzy Entropy considers non-membership and hesitation degree along with membership degree. Thus, it handles uncertainty to better extent. To evaluate the performance of proposed IFE-FS method, extensive experimentations were conducted on three standard benchmark datasets: 20-NewsGroups, Reuters-21578 and TDT2. From the experiments, we can conclude that the IFE-FS method reduces high dimensionality of feature matrix and enhances the performance of classifier. It can also be concluded that IFE-FS selects good subsets of features, which contains most discriminative features. The experimental results reveal that the IFE-FS method outperformed other feature selection methods in terms of F-measure.

In future, we intend to use optimization technique in the proposed IFE-FS method to automatically select the threshold value r. Additionally, we intend to add the kernel method in IFCM, which could improve the categorization performance.

References

[1] J. H. Du, "Automatic text classification algorithm based on Gauss improved convolutional neural network", *Journal of Computational Science*, *21*, pp.195-200, 2017.

[2] M. Kang, J. Ahn and K. Lee, "Opinion mining using ensemble text hidden Markov models for text classification", *Expert Systems with Applications*, *94*, pp.218-227, 2018.

[3] A. M. Jalil, I. Hafidi, L. Alami and E. Khouribga, "Comparative study of clustering algorithms in text mining context", *International Journal of Interactive Multimedia and Artificial Intelligence*, *3*(7), pp.42-45, 2016.

[4] A. Moreno and T. Redondo, "Text analytics: The convergence of big data and artificial intelligence", *International Journal of Interactive Multimedia and Artificial Intelligence*, *3*(6), pp.57-64, 2016.

[5] A. Onan, S. Korukoğlu and H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification", *Expert Systems with Applications*, *57*, pp.232-247, 2016.

[6] A. K. Uysal and S. Gunal, "The impact of preprocessing on text classification", *Information Processing & Management*, *50*(1), pp.104-112, 2014.

[7] J. Huang, G. Li, Q. Huang and X. Wu, "Joint feature selection and classification for multilabel learning", *IEEE Transactions on Cybernetics*. 2017.

[8] D. Agnihotri, K. Verma and P. Tripathi, "Variable Global Feature Selection Scheme for automatic classification of text documents", *Expert Systems with Applications*, *81*, pp.268-281, 2017.

[9] M. Labani, P. Moradi, F. Ahmadizar and M. Jalili, "A novel multivariate filter method for feature selection in text classification problems", *Engineering Applications of Artificial Intelligence*, *70*, pp.25-37, 2018.

[10] Y. Lu and Y. Chen, "A Text Feature Selection Method Based on the Small World Algorithm", *Procedia Computer Science*, *107*, pp.276-284, 2017.

[11] B. Tang, S. Kay and H. He, "Toward optimal feature selection in naive Bayes for text categorization", *IEEE Transactions on Knowledge and Data Engineering*, *28*(9), pp.2508-2521, 2016.

[12] A. K. Uysal, "An improved global feature selection scheme for text classification", Expert Systems with Applications, 43:82–92, 2016.

[13] D. Zhu and K. W. Wong, "An evaluation study on text categorization using automatically generated labeled dataset", *Neurocomputing*, *249*, pp.321-336, 2017.

[14] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization", In *ICML,* Vol. 97, pp. 412-420, 1997.

[15] A. Rehman, K. Javed, and H. A. Babri, "Feature selection based on a normalized difference measure for text classification", *Information Processing & Management*, *53*(2), pp.473-489, 2017.

[16] L. P. Jing, H. K. Huang and H. B. Shi, "Improved feature selection approach TFIDF in text mining", In *Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on,* Vol. 2, pp. 944-946, 2002. IEEE

[17] Y. Xu, G. J. Jones, J. Li, B. Wang and C. Sun, "A study on mutual information-based feature selection for text categorization". *Journal of Computational Information Systems*, *3*(3), pp.1007-1012, 2007.

[18] C. Lee and G. G. Lee, "Information gain and divergence-based feature selection for machine learning-based text categorization", *Information processing & management*, *42*(1), pp.155-165, 2006.

[19] S. S. Mengle and N. Goharian, "Ambiguity measure feature-selection algorithm", *Journal of the American Society for Information Science and*

Technology, *60*(5), pp.1037-1050, 2009.

[20] M. Lan, S. Y. Sung, H. B. Low and C. L. Tan, "A comparative study on term weighting schemes for text categorization", In *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on,* Vol. 1, pp. 546-551, 2005. IEEE

[21] M. B. Revanasiddappa, B. S. Harish and S. Manjunath, Document classification using Symbolic classifiers", In *Contemporary Computing and Informatics (IC3I), 2014 International Conference on,* pp. 299-303, November 2014. IEEE

[22] C. Largeron, C. Moulin and M. Géry, "Entropy based feature selection for text categorization", In *Proceedings of the 2011 ACM Symposium on Applied Computing,* pp. 924-928, 2011. ACM

[23] C. E. Shannon, "A mathematical theory of communication", *ACM SIGMOBILE Mobile Computing and Communications Review*, *5*(1), pp.3-55, 2001.

[24] S. V. Vaseghi, "*Advanced digital signal processing and noise reduction*", 2008. John Wiley & Sons

[25] A. De Luca and S. Termini, "A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory", *Information and control*, *20*(4), pp.301-312, 1972.

[26] H. M. Lee, C. M. Chen, J. M. Chen and Y. L. Jou, "An efficient fuzzy classifier with feature selection based on fuzzy entropy", *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *31*(3), pp.426-432, 2001.

[27] C. Palanisamy and S. Selvan, "Efficient subspace clustering for higher dimensional data using fuzzy entropy", *Journal of Systems Science and Systems Engineering*, *18*(1), pp.95-110, 2009.

[28] X. Liu and W. Pedrycz, "*Axiomatic fuzzy set theory and its applications*", Vol. 244, 2009. Heidelberg: Springer

[29] L. A. Zadeh, "Probability measures of fuzzy events". *Journal of mathematical analysis and applications*, *23*(2), pp.421-427, 1968.

[30] K. T. Atanassov, "Intuitionistic fuzzy sets", *Fuzzy Sets and Systems*, *20*(1), pp.87-96, 1986.

[31] T. Chaira, "A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images", *Applied Soft Computing*, *11*(2), pp.1711-1717, 2011.

[32] K. Atanassov, "Intuitionistic Fuzzy Sets: Theory and Applications", 1999. Springer, Heidelberg

[33] B. S. Harish, D. S. Guru and S. Manjunath, "Representation and classification of text documents: A brief review", *IJCA, Special Issue on RTIPPR (2)*, pp.110-119, 2010.

[34] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features", Machine Learning: ECML-98. pp.137-142, 1998.

[35] E. P. Jiang, "Semi-supervised text classification using RBF networks", In *International Symposium on Intelligent Data Analysis,* pp. 95-106, August 2009. Springer Berlin

[36] S. Goswami, A. K. Das, A. Chakrabarti and B. Chakraborty, "A feature cluster taxonomy based feature selection technique", *Expert Systems with Applications*, *79*, pp.76-89, 2017.

[37] H. T. Ng, W. B. Goh and K. L. Low, "Feature selection, perceptron learning, and a usability case study for text categorization", In *ACM SIGIR Forum* (Vol. 31, No. SI, pp. 67-73, July 1997. ACM

[38] F. Sebastiani, "Machine learning in automated text categorization", *ACM computing surveys (CSUR)*, *34*(1), pp.1-47, 2002.

[39] S. Vora and H. Yang, "A comprehensive study of eleven feature selection algorithms and their impact on text classification", In *Computing Conference,* pp. 440-449, 2017.

[40] B. Tang, S. Kay and H. He, "Toward optimal feature selection in naive Bayes for text categorization", *IEEE Transactions on Knowledge and Data Engineering*, 28(9), pp.2508-2521, 2016.

[41] J. Cai and F. Song, "Maximum entropy modeling with feature selection for text categorization", In *Asia Information Retrieval Symposium*, pp. 549-554, January 2008. Springer Berlin Heidelberg

[42] V. B. Vaghela, K. H. Vandra and N. K. Modi, "Entropy based feature selection for multi-relational Naive Bayesian Classifier", *Journal of International Technology and Information Management*, *23*(1), p.2, 2014.

[43] W. Liu and N. Song, "A fuzzy approach to classification of text documents", *Journal of Computer Science and Technology*, *18*(5), pp.640-647, 2003.

[44] J. L. Fan and Y. L. Ma, "Some new fuzzy entropy formulas", *Fuzzy sets and Systems*, *128*(2), pp.277-284, 2002.

[45] R. N. Khushaba, A. Al-Jumaily and A. Al-Ani, "Novel feature extraction method based on fuzzy entropy and wavelet packet transform for myoelectric control", In *Communications and Information Technologies, 2007. ISCIT'07. International Symposium on,* pp. 352-357, October 2007. IEEE

[46] O. Parkash, P. K. Sharma and R. Mahajan, "New measures of weighted fuzzy entropy and their applications for the study of maximum weighted fuzzy entropy principle", *Information Sciences*, *178*(11), pp.2389-2395, 2008.

[47] P. Luukka, "Feature selection using fuzzy entropy measures with similarity classifier", *Expert Systems with Applications*, *38*(4), pp.4600-4607, 2011.

[48] F. Ahmadizar, M. Hemmati and A. Rabanimotlagh, "Two-stage text feature selection method using fuzzy entropy measure and ant colony optimization", In *Electrical Engineering (ICEE), 2012 20th Iranian Conference on,* pp. 695-700, May 2012. IEEE

[49] M. Sugeno and T. Terano, "A model of learning based on fuzzy information", *Kybernetes*, *6*(3), pp.157-166, 1977.

[50] R. R. Yager, "On the measure of fuzziness and negation part I: membership in the unit interval", *Int. J. Gen. Syst.* **5**, pp.221–229, 1979.

[51] P. Intarapaiboon, "An application of intuitionistic fuzzy sets in text classification", In *Information Science, Electronics and Electrical Engineering (ISEEE), 2014 International Conference on,* Vol. 1, pp. 604-608, April 2014. IEEE

[52] Y. Li, D. L. Olson and Z. Qin, "Similarity measures between intuitionistic fuzzy (vague) sets: A comparative analysis", *Pattern Recognition Letters*, *28*(2), pp.278-285, 2007.

[53] G. A. Papakostas, A. G. Hatzimichailidis and V. G. Kaburlasos, "Distance and similarity measures between intuitionistic fuzzy sets: A comparative analysis from a pattern recognition point of view", *Pattern Recognition Letters*, *34*(14), pp.1609-1622, 2013.

[54] E. Szmidt and J. Kacprzyk, "Using intuitionistic fuzzy sets in text categorization", In *International Conference on Artificial Intelligence and Soft Computing,* pp. 351-362, 2008. Springer, Berlin, Heidelberg

[55] H. M. Zhang, Z. S. Xu and Q. Chen, "On clustering approach to intuitionistic fuzzy sets", *Control and Decision*, *22*(8), p.882, 2007.

[56] D. Iakovidis, N. Pelekis, E. Kotsifakos and I. Kopanakis, "Intuitionistic fuzzy clustering with applications in computer vision", In *Advanced Concepts for Intelligent Vision Systems* (pp. 764-774), 2008. Springer Berlin/Heidelberg

[57] Z. Xu, J. Chen and J. Wu, "Clustering algorithm for intuitionistic fuzzy sets", *Information Sciences*, *178*(19), pp.3775-3790, 2008.

[58] Z. Xu and J. Wu, "Intuitionistic fuzzy C-means clustering algorithms", *Journal of Systems Engineering and Electronics*, *21*(4), pp.580-590, 2010.

[59] E. Szmidt and J. Kacprzyk, "Entropy for intuitionistic fuzzy sets", *Fuzzy Sets and Systems*, *118*(3), pp.467-477, 2001.

[60] P. Burillo and H. Bustince, "Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets". *Fuzzy Sets and Systems*, *78*(3), pp.305-316, 1996.

[61] W. L. Hung and M. S. Yang, "Fuzzy entropy on intuitionistic fuzzy sets", *International Journal of Intelligent Systems*, *21*(4), pp.443-451, 2006.

[62] I. K. Vlachos and G. D. Sergiadis, "Intuitionistic fuzzy information–applications to pattern recognition", *Pattern Recognition Letters*, *28*(2), pp.197-206, 2007.

[63] 20newsgroups: http://people.csail.mit.edu/jrennie/20Newsgroups/

[64] Reuters-21578: http://www.daviddlewis.com/resources/testcollections/reuters21578/

[65] X. Hu, X. Zhang, C. Lu, E. K. Park and X. Zhou, "Exploiting Wikipedia as external knowledge for document clustering", In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining,* pp. 389-396, 2009. ACM

M. B. Revanasiddappa

He received his B.E degree in Information Science and Engineering and M.Tech degree in Software Engineering from Visvesvaraya Technological University, Belagavi, Karnataka, India. He is currently pursuing Ph.D degree in Computer Science from Visvesvaraya Technological University, Belagavi, Karnataka, India. His area of research includes Machine Learning, Text Mining and Soft Computing.

B. S. Harish

He obtained his B.E in Electronics and Communication (2002), M.Tech in Networking and Internet Engineering (2004) from Visvesvaraya Technological University, Belagavi, Karnataka, India. He completed his Ph.D. in Computer Science (2011); thesis entitled "Classification of Large Text Data" from University of Mysore. He is presently working as an Associate Professor in the Department of Information Science & Engineering, JSS Science & Technology University, Mysuru. He was invited as a Visiting Researcher to DIBRIS - Department of Informatics, Bio Engineering, Robotics and System Engineering, University of Genova, Italy from May-July 2016. He delivered various technical talks in National and International Conferences. He has invited as a resource person to deliver various technical talks on Data Mining, Image Processing, Pattern Recognition, Soft Computing. He is also serving and served as a reviewer for National, International Conferences and Journals. He has published more than 50 International reputed peer reviewed journals and conferences proceedings. He successfully executed AICTE-RPS project which was sanctioned by AICTE, Government of India. He also served as a secretary, CSI Mysore chapter. He is a Member of IEEE (93068688), Life Member of CSI (09872), Life Member of Institute of Engineers and Life Member of ISTE. His area of interest includes Machine Learning, Text Mining and Computational Intelligence.