

Automatic Irony Detection using Feature Fusion and Ensemble Classifier

H. M. Keerthi Kumar¹, B. S. Harish² *

¹ JSSRF, JSS TI Campus, Mysuru, Karnataka (India)

² Department of Information Science and Engineering, JSS Science and Technology University, Mysuru, Karnataka (India)

Received 15 March 2019 | Accepted 12 June 2019 | Published 2 July 2019



ABSTRACT

With the advent of micro-blogging sites, users are pioneer in expressing their sentiments and emotions on global issues through text. Automatic detection and classification of sentiments like sarcastic or ironic content in micro-blogging reviews is a challenging task. It requires a system that manages some kind of knowledge to interpret the sentiment expressed in text. The available approaches are quite limited in their capabilities and scope to detect ironic utterances present in the text. In this regards, the paper propose feature fusion to provide knowledge to the system by alternative sets of features obtained using linguistic and content based text features. The proposed work extracts five sets of linguistic features and fuses with features selected using two stages of a feature selection method. In order to demonstrate the effectiveness of the proposed method, we conduct extensive experimentation by selecting different feature subsets. The performances of the proposed method are evaluated using Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Decision Tree (DT) and ensemble classifiers. The experimental result shows the proposed approach significantly out-performs the conventional methods.

KEYWORDS

Classification, Feature Selection, Feature Fusion, Ensemble Method, Irony, K-means Clustering, Sentiment Analysis.

DOI: 10.9781/ijimai.2019.07.002

I. INTRODUCTION

IN the era of internet, micro-blogging sites like twitter, facebook, and review forums allow users to express their sentiment or opinion [56]. The sentiment or opinion in micro-blogs may relate to product, event or political debate in the form of text, image and video clips, where text plays an essential role in expressing opinions. Micro-blogging textual information is rich, progressively expanding in large volume of data, with a variety of information ranging from product to political events [1]. This textual information plays a vital role in determining sentiment of the population. An enormous amount of textual information provides valuable insight to governments, business organizations and individual decision makers [2]. The manual summarization of micro-blog textual information is time consuming. Hence, the automatic summarization of subjective information is very essential to determine polarity of the population [3]. The automatic text polarity identification process is known as Sentiment Analysis (SA) or Opinion Mining.

Sentiment Analysis (SA) aims to classify a given text into positive, negative or neutral polarity [4] [53]. There are many challenges related to SA which need to be addressed and resolve. Some of the challenges are: (a) language utterances, (b) punctuation marks to express sentiments, (c) shorten form of words (mainly in micro-blogs), (d) sarcasm/irony present in text snippet and many more. Sarcasm/irony detection in text is one of the major challenges in sentiment analysis.

The sarcasm has been studied by multidisciplinary endeavors such as sociology [5], psychologists [6], linguists [7] and computer scientists [8] for different types of text: twitter tweets, product reviews, internet dialogs, etc. [19]. Over the time, human have developed the ability to recognize sarcastic/ironic intent in utterances from childhood through social interaction [10]. Sarcasm is portrayed as ironic, intended to insult, mock or amuse. However, irony or sarcasm is a complicated mode of communication, which is informally connected with the expression of feelings, attitudes and emotions [11]. Sarcasm is closely related to irony [16]. Irony shifts the polarity of an apparently positive/negative utterances into its opposite [12]. Human intervention to recognize irony is extremely studious and time consuming. Due to this, researchers aim to develop an automatic system to recognize the ironic utterances present in the text.

Understanding ironic utterances from stance of both semantic and grammatical is another practice of Nature Language Processing [13]. Irony detection techniques are roughly categorized into machine learning and lexicon based approaches [14]. A lexicon based approach uses dictionary/corpus using statistical and semantic features to detect ironic utterances in a given text. On the other hand, a machine learning approach uses text features to classify ironic utterances using supervised/unsupervised techniques based on label or unlabeled text. Both approaches perform well in detecting ironic utterances present in sentences [47].

In irony detection, feature extraction and selection plays a vital role in determining the ironic utterances present in sentences. The features are extracted based on linguistic and content based approaches. A linguistic approach is an extremely broad phrase to extract textual features such as lexical, hyperbole and pragmatic features. The lexical

* Corresponding author.

E-mail addresses: hmkeerthikumar@gmail.com (H. M. Keerthi Kumar), bsharish@jssstuniv.in (B. S. Harish).

approach uses text properties such as unigram, bigram, n-grams, etc. for detecting irony in text [42]. In a lexical approach, the dictionary or corpus related to vocabularies of words are used to identify irony present in text snippet [15]. Similarly, hyperbole is another key feature often used in irony detection from textual data. A hyperbolic text contains interjection (wow, aha, etc.), punctuation marks (question marks and exclamation mark), quotes (‘ ’, “ ”) and intensifiers (noun, adverbs, adjectives) to detect irony in tweets [43]. The pragmatic feature includes symbolic or figurative texts such as emoticons of happy, sad, laughing, and crying etc., expressed in the sentences [15]. Researchers [17] [47] [51] [52] used various linguistic features to detect ironic utterances in short texts. However, identifying appropriate patterns to detect ironic utterances remains an open challenge.

In addition to the wide range of linguistic features, many researchers [9] [22] [46] [55] studied the content based approach i.e., presence or absence of term/features in reviews. In the content based approach, the number of features plays a vital role in accurate classification. The high dimensionality and sparsity is one of the major challenges faced during classification task. To curtail the dimensionality, many researchers reported in [48] [49] [50] used feature selection methods to select discriminative features from a high dimensional feature space. The conventional feature selection methods such as Chi-square (χ^2), Information Gain (IG), Mutual Information (MI) are used to select discriminative features from high dimensional feature spaces. However, the selected feature subset may have features which convey similar information [26]. On this line, we propose a two stage feature subset selection using conventional feature selection methods and a clustering method to select the most discriminative features from a high dimensional feature space. On the other hand, linguistic features are extracted to detect sarcastic utterances in short text. The five groups of linguistic features that are extracted viz: Rating Feature, Word Feature, Acronym Feature, Symbol Feature and Emoticon Feature. Further, features are fused to capture various dimensions of characteristics of review. The fused features are classified using various classifiers such as Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF) and Decision Tree (DT). To enhance the performance of the classifiers, we ensemble Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF) and Decision Tree (DT) using weighted majority voting schema.

The main contribution of the paper is as follows:

- Five types of features are extracted using linguistic approach viz: Rating, Word, Acronym, Symbol and Emoticon Features.
- Two stage feature subset selection to select the most discriminative features.
- Features are fused (linguistic and content based feature subsets) to capture various dimensions of characteristics of review.
- Use of weighted majority voting schema to ensemble decision of each classifier.

The remainder of the paper is structured as follows: Section II depicts related works on irony detection. Section III presents the methodology of the proposed work. Experimentation and related results along with discussion are presented in Section IV. Finally, the work is concluded with future research directions in Section V.

II. RELATED WORK

In recent decades, prominent research works are carried out by various researches [17]-[21] [27] for automatic detection of irony in various micro-blogs such as twitter, product reviews and movie reviews. A brief survey on automatic sarcasm detection by Joshi et al. [16] described various datasets, approaches, trends and issues in sarcasm detection. Some of the related works in literature are reviewed

based on supervised, semi-supervised and rule based approaches. Similarly, in [13] Wicana et al., described sarcasm detection from the machine-learning perspective. The research tried to explore supervised, unsupervised, rule based approaches and hybrid approaches to process data. Dave and Desai in [14], examine various lexicons based and machine learning techniques for sarcasm detection on textual data. The comprehensive survey highlights the use of hybrid techniques, i.e. usage of both lexicons based and machine learning techniques together for sarcasm classification.

Ravi and Ravi in [17], proposed a framework to automatically detect satire, sarcasm and irony found in news and customer reviews. The framework extracts features based on linguistic, semantic, psychological and unigram features. The various feature selection techniques are used to select the relevant feature subset from unigram features. The extracted and selected features subsets are fused and classified using Support Vector Machine (SVM) with various kernels, Logistic Regression (LR), Random Forest (RF), Naive Bayes (NB), Multilayer perceptron (MLP), etc. Similarly, Buschmeier et al. [18], described impact of features in a classification approach to detect irony in product review [20]. The method uses 29 special features such as positive/negative imbalance of reviews, hyperbole, positive/negative word with punctuation, quotes etc., along with bag-of-word features (21,773 features). The various features set comparison are drawn on different classifiers such as Linear SVM, LR, Decision Tree (DT), RF, and NB classifiers. Filatova in [19], identifies the sentiment shifts in sarcasm product review dataset [20]. The method demonstrated sentiment flow shifts (from negative to positive and likewise) using bi-gram feature along with 8 classification features (very negative-positive, very negative-very positive, negative-positive, negative-very positive and likewise). Justo et al., [21] proposed to detect sarcasm and nastiness in the social web. The various features such as mechanical turk, statistical cues, linguistic information, semantic information using Linguistic Inquiry Word Count (LIWC) and n-gram distribution of Part-of-Speech (POS) taggers are used to extract features. The feature subset is selected using the chi-square (χ^2) feature selection method. The binary classification was performed using rule-based and NB classifiers.

In literature, many researchers proposed various feature extraction techniques to classify ironic review. The existing methods uses NLP and machine learning approaches to extract various patterns to classify ironic content in reviews. However, some of the observations made from literature are as follows: (a) the number of features was too large, (b) feature extraction using POS tagger, and (c) searching each word in sentiment dictionary is clumsy. Hence, the proposed research developed a new approach to address these issues such as (a) Feature extraction (use of linguistic and content based features), (b) Applying feature selection methods to select discriminative features in content based approach and (c) Fusion of both features to provide useful insights of ironic contents. The content based feature subset is selected using a two stage feature selection method. In the first stage, the conventional feature selection methods such as Chi-square (χ^2), Information Gain (IG) and Mutual Information (MI) are used to select relevant feature subsets from a high dimensionality feature space. The selected feature subset may have features which convey similar information. Due to this reason, the features exhibiting similar information are grouped and features belonging to each group are selected. The second stage of feature selection is used to select the representative feature from each first stage feature subset. In this second stage, the features are grouped based on features exhibiting similar information. The clustering algorithm is used to cluster or group the similar information features subset. In this work, k-means clustering algorithm is used to cluster such that the members in each group are as similar (close) as possible to one another. The feature nearer to each cluster center is consider as

the representative feature among other features within the cluster. On the other hand, the linguistic features are extracted and categorized into five groups such as Rating, Word, Acronym, Symbol and Emoticon Features. Overall twenty special features are extracted using linguistic based feature extraction and categorized into these groups. The special features symbolize frequency of occurrences of the each feature in a review. Further, the special feature and content based feature subset are fused and evaluated using various classifiers such as Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF) and Decision Tree (DT) on product review dataset [20]. Furthermore, we construct an ensemble of four classifiers: SVM, LR, RF and DT based on combination rule to enhance the classification of the classifiers using weighted majority voting scheme.

III. METHODOLOGY

The proposed approach is a hybrid feature fusion method, which integrates linguistic features and content based text features. The proposed approach is used to classify product reviews into ironic or non-ironic content based on a feature fusion method. The general architecture of the proposed feature fusion approach is outlined in Fig. 1.

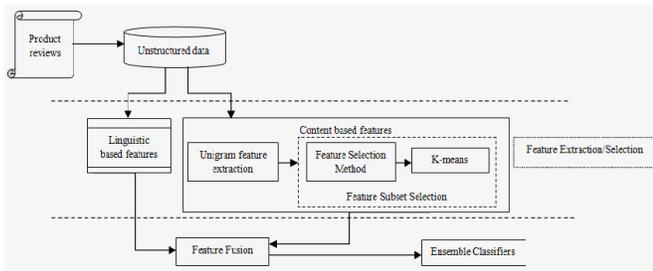


Fig. 1. Block diagram of the proposed feature fusion method.

As depicted in Fig. 1, the proposed approach comprises three phases: (i) feature extraction and selection, (ii) feature fusion and (iii) ensemble classification. The details of product review dataset are presented in section IV (A) and the rest of the above are described in this section. In order to identify ironic customer reviews, we developed feature fusion of linguistic and content based features. The fused features are classified using individual and ensemble classifiers such as Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT) and Random Forest (RF) classifiers.

A. Feature Extraction and Selection for Irony Detection

In the classification of irony on product reviews, feature extraction and selection plays a vital role. Usually, ironic utterances expresses opposite meaning of the intended content. To extract ironic utterances present in reviews, linguistic based feature extraction is used. On the other hand, text features present in reviews yield promising results in the area of text classification [28] and sentiment analysis [25]. Hence, content based feature extraction and selection of features are used to detect ironic utterances in text.

B. Linguistic Feature Extraction

In this work, the features are extracted using a linguistic approach (hereafter, special features) to identify ironic utterances in text. To extract special features from sentences, we use syntactic information such as interjections, pragmatic, intensifier and many more. The special features are grouped into five sets of features. Table I provides an overview of group of features extracted using the linguistic approach.

In order to extract ironic feature, in [22] briefed that irony always express opposite of its actual content. The rating feature set groups the star rating (i.e 1* to 5* rating) for the reviews and the imbalance feature.

The imbalance between star rating and over-all polarity of words in the review are considered. The work in [18], assumes there is imbalance when the star rating (i.e., 4* and 5*) is considered as positive review but polarity of reviews are negative. Similarly, imbalance exists when star-rating (1* and 2*) is considered as negative review but polarities of reviews are positive. This imbalance between the start rating and polarity of the text is considered as imbalance feature. The polarity of the review text is determined based on dictionary of [23], which consists of 6,800 words with positive and negative polarity words.

TABLE I. FEATURE GROUPS USED FOR IRONY DETECTION IN PRODUCT REVIEW DATASETS

No.	Groups	Features
1	Rating Features	Star rating, Imbalance feature
2	Word Feature	Hyperbole, Quotes, Ellipsis, Punctuation, Interjection
3	Acronym Feature	Acronym for laughter, Onomatopoeia for laughter, Acronym for Grin
4	Symbol Feature	Exclamation mark, Question mark, Ellipsis only, Ellipsis and Punctuation mark, Question and Exclamation mark.
5	Emoticon Feature	Sad, Happy, Laughing, Surprise, Winking

In general, user tends to exaggerate his/her sentiment through quoting certain words in sequences or between symbols. The word feature set consists of the features related to polarity of words present in the reviews. The feature hyperbole [24] implies the exaggeration present in sentences, which are extracted, based on appearance of three consecutive positive or negative words in a row. The feature quotes considers two consecutive intensifiers such as noun, adverbs and adjectives, which have positive or negative polarity in quotation marks [18]. In linguistic, ellipsis refers to the omission of words rather than repeating them unnecessarily and it is represented as three consecutive dots (“...”). In this work, feature ellipsis is considered as any positive or negative word end with an ellipsis. The feature punctuation considers positive or negative word with punctuation mark such as question marks and exclamation mark. The feature interjection indicates the occurrence of terms such as “wow”, “ah”, “aha” and many more in the sentences.

The use of language is constantly changing across space and across social group. The usage of acronymic word in micro blog text has grown enormously over the time. The feature such as acronym for laughter (lol, lawl, luls and many more) is used as a short form of laugh. The Onomatopoeia feature mimic the verbal conversation for laugh such as “haha”, “mu-ba”, “hehe”, “hihi” and many more in the sentence. The feature acronym for Grin (*g*, *gg* and many more) depicts expression smiling broadly. These features are grouped into acronym feature set to describe emotions jargon present in reviews.

Usually, user tends to highlight his/her emotion by making more intense through exclamation mark, question mark, ellipsis and combination of these symbols. The symbol feature set groups these set of features to intensify the sentiment present in review text. The feature exclamatory mark (“!”) symbolizes to express strong emotions in a review. Similarly, Question marks (“?”) represent uncertainty about something in the sentence. Hence, exclamatory mark and question mark symbols are considered as set of features. The ellipsis only feature indicates the situation in which words are left out of a sentence but the sentence can still be understood by its context. The consecutive three dots (“...”) are scrutinized as ellipsis feature. Further, the combination of ellipsis and punctuation (ellipsis followed by multiple exclamation or combination of exclamation and question mark), question mark and exclamation mark are considered as another set of features.

In social media text, emoticon plays a vital role in expressing a

person's feelings or mood by using punctuation marks, numbers, and letters. The emoticon feature set consists of emoticons such as Sad (“:(”, “- :”, “:”) and “o:”), Happy (“:-)”, “:0)”, “:)”, “(:” and many more), Laughing (“:-D” and “xD”), Surprise (“:-0”, “0-:” and “0.0”) and Winking (“;-)”, “; 0)” and “;:”) as features. Overall, twenty linguistic features are extracted from five sets of features and hereafter linguistic features are referred as Special Feature (SF). However, in [18] the features are concentrated more on twitter related data. In this work, we generalize the linguistic features more on review text.

C. Content based Feature Extraction

In addition to special features, the content based features are extracted. To extract content based features, raw data are preprocessed by removing non-informative and trivial information. The review text consists of digits, punctuation, HTML tags and stop words, which occur more often and do not contribute to the analysis [57]. The preprocessed text are represented using unigram features with term frequency (tf) schema. The work in [32] suggests that unigram with term frequency (tf) performs well on sentiment analysis for micro-blogging data. Hence, we considered, unigram with term frequency (tf) schema to represent review text into machine understandable form. The extracted features are high in dimension feature space, which need to be reduced to low dimensionality feature space by applying various feature selection methods. The aim of feature selection methods is to select relevant and non-redundant features from high dimensionality feature space. In this work, the two stages of Feature Subset Selection (FSS) are used to select the most discriminative feature from high dimensionality feature space.

D. Feature Subset Selection (FSS)

The feature subset selection consists of two stages: In the first stage, the conventional feature selections method is used to select relevant feature subsets from a high dimensionality feature space. The selected feature subset may have features which convey similar information. Due to this reason, the features exhibiting similar information are grouped and feature belonging to each group are selected. The second stage feature selection is used to select the representative feature from each first stage FSS group.

In the first stage, the conventional feature selection methods such as CHI-square (χ^2), Information Gain (IG) and Mutual Information (MI) feature selection methods are used. These features selection methods are widely used to select relevant feature subset in text processing domain. The Chi-square (χ^2) is a statistical method used to test specific feature correlated with the class. The higher value of feature (χ^2) score indicates the likelihood of feature occurrence is highly dependent on the occurrence of the class. The IG is frequently used in the field of machine learning to determine the term of goodness criterion. IG measures the information that is gained by knowing the value of the attribute, which is the difference between the entropy of the distribution before the split and the entropy of the distribution after the split. The higher IG value indicates features contribute with more information for category prediction of the review, whereas lesser IG value indicates they do not add much information. Similarly, MI measures how much information a feature contains about a class. If the feature distribution is the same in intra-class and inter-class, the MI value reaches the minimum. Otherwise, MI value reaches the maximum when feature distribution is in intra-class only.

Let there be m the number of reviews and n the number of total features in the feature space. The preprocessed texts are represented in the form of a Document Term Matrix $DTM (m \times n)$. The feature selection methods are applied to select relevant features among the features space. The conventional feature selection methods generate scores (S) corresponding to each feature. The feature scores are arranged in descending order to select top rated feature scores. The

feature subsets are selected by fixing threshold value (l) empirically. The first stage feature subset selection using conventional feature selection methods has been shown in Algorithm 1.

Algorithm 1: First stage feature subset selection

Data: Document Term Matrix $DTM (m \times n)$; m = number of reviews; n = total number of features; S = Feature score;
 Result: l number of features
 Initialize threshold value to l
 //compute score for each feature using Feature Selection Method (FSM)
 Step1: $S = FSM[DTM]$
 Step 2: $S = [S_1, S_2, \dots, S_n]$ // n = total number of features
 Step 3: Sort S in descending order // to select top ranked feature scores
 Step 4: Select first l number of features from S , the selected feature subset l is represented using Document Term Matrix $DTM (l \times m)$, where m represents the number of reviews and l indicates the number of feature subset selected from feature selection methods
 ($l \ll n$).

The work presented in [26] suggested that features may convey similar information in the feature space. In conjecture to that, features evinige similar information are grouped to select the most representative features from each group. Due to this reason, the second stage feature selection is applied to select the most representative features from the feature subset obtained from the first stage.

In the second stage, the features are grouped based on features exhibiting similar information. The clustering algorithm is used to cluster or group the similar information feature subset. In this work, k-means clustering algorithm is used to group or cluster, such that the members in each group are as similar (close) as possible to one another. The k-means clustering algorithm works iteratively to assign features to one of the k clusters based on the similar information features. To determine the optimal number of clusters (k) as mentioned in [26] [29], number of cluster (k) is varied from \sqrt{l} to $\frac{l}{2}$. The feature nearer to each cluster centers is considered as the representative feature among other features within the cluster. The cosine similarity measure is used to determine the similarity between the features and cluster center.

The $DTM (m \times l)$ is obtained from the first stage of FSS and is transposed and represented as Term Document Matrix $TDM (m \times l)$, where l indicates the number of features in the subset and m represents the number of reviews. The algorithm 2 presents the second stage of FSS to select the most discriminative features among the features exhibiting similar information.

Algorithm 2: Second stage of feature subset selection

Data: Term Document Matrix $TDM (m \times l)$, l = number of features, m = number of reviews, k = clusters
 Result: $TDM' (k \times m)$, k = number of discriminative features ($k < l$)
 Initialize k points at random as cluster centers, $t=0$, $F = \{\}$,
 $U^0 = \{u_1^0, \dots, u_k^0\}$
 Repeat
 Step 1: for $i \leftarrow 1$ to k do
 Step 2: for $j \leftarrow 1$ to l do
 //compute distance between cluster center U_i and feature F_j
 $D_{ij} = dist(U_i, F_j)$
 end
 end
 Step 3: Assign F_j to nearest cluster center u_i
 // Update cluster center

Step 4: for $i \leftarrow 1$ to k do

$$u_i^{(t+1)} = \frac{\sum_{j \in I} F_j}{|u_i|}$$

end

Until $U^{(t-1)} - U^{(t)} < \zeta$, $t = t + 1$

//find representative feature for each cluster center

Step 5: for $i \leftarrow 1$ to k do

$$F_j = \min_j \text{dist}_{j \in I} (u_i, F_j)$$

$$F = F \cup F_j$$

end

Further, the feature subset F , which consists of k number of discriminative features is represented in TDM' ($k \times m$), where m is the number of reviews and k is the reduced number of features $k < l < n$.

E. Feature Fusion

In addition to linguistic features, content based features play a significant role in sentiment analysis [17]. In order to capture various dimensions of characteristics of review, linguistic features and subset of content based features are fused. Overall twenty special features are extracted using linguistic based feature extraction and categorized into five groups. The special features symbolize frequency of occurrences of each feature in a review. On the other hand, content based feature subsets selections are applied to select the most discriminative features from a high dimensionality feature space. The special feature and content based Feature Subset Selection (FSS) are fused and evaluated using various classifiers.

F. Ensemble Classifiers

In this work, the predictive classifiers such as Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT) and Random Forest (RF) are used to evaluate the efficiency of the proposed method. The SVM classifier works by finding the hyper-plane, which maximizes the margin between the two classes. The vectors that define the hyper-plan are known as support vectors. The SVM is used for both classification and regression problems [34]. The LR is a statistical method used for binary classification problems (problems with two class values) [33]. The DT is a non-parametric approach used to construct a tree in top-down and recursive divide-and-conquer manner [17]. DT is mainly used for classification and regression problem. Similarly, RF is an automatic learning technique which combines the concepts of random subspaces and bagging [36]. Random Forest operates by constructing a multitude of decision trees at training time and outputting the class based on decision of individual trees [35]. These four classifiers are widely used in classification of ironic content in reviews [18] [30].

On other hand, work of [37] [38] emphasizes that individual prediction of various classifiers can be ensemble so that a more robust and accurate classification model can be built. The ensemble learning plays a vital role in recent research activity of pattern recognition and machine learning [38]. The main aim of the ensemble learning is to weigh several individual classifiers and combine the prediction of the multiple classifiers, which outperforms prediction of individual classifiers [31]. The majority voting and weighted majority voting are the most popular combination schemas, which are widely used in ensemble classification [39]. The simple majority voting schema selects one of many alternatives of the predicted classes with the most votes [40]. The weighted majority voting schema assigns weight for each prediction of the classifiers based on the performance of the classifier [41]. In this work, weighted majority voting schema is used to assign the weight for each decision of the classifiers. For

weighted majority voting schema, first let there be (H_1, H_2, \dots, H_l) classifiers with accuracies (A_1, A_2, \dots, A_l) , respectively. Then, h_{ij} be defined as the decision of j^{th} classifier which chooses j^{th} label from class C . In weighted majority voting schema, weights w_i are assigned to the individual decision of the classifiers h_i . The ensemble classifier decision using weighted majority voting schema $H(X)$ is as follows:

$$H(X) = \max_{j=1, \dots, C} \sum_{i=1}^l w_i h_{i,j} \quad (1)$$

In weighted majority voting schema, the optimal weight w_i for the classifiers are assigned based on the accuracy of the classifiers h_i , i.e., $w_i \propto \frac{a_i}{(1-a_i)}$ where a_i is the accuracy of the independent classifiers h_i . In [54], the weight w_i equip each classifier output with different weights between 0 and 1 value ($0 \leq w_i \leq 1$) and sum of the weight w_i is equal to 1. Hence, weights w_i are assigned based on the performance of the individual classifiers as indicated in [54].

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the effectiveness of feature fusion is examined on a publically available product review dataset [20].

A. Dataset Description

The Amazon product review dataset consists of 1,254 reviews, consists of 437 ironic and 817 non-ironic or regular reviews created by [20]. The structure of the dataset contains * rating ranging from 1* to 5* star, along with labeled ironic and regular review content. Table II depicts the distribution of reviews by star-rating.

TABLE II. DISTRIBUTION OF REVIEWS BY STAR-RATING

	Number of reviews					
	1*	2*	3*	4*	5*	
Ironic	437	262	27	20	14	114
Regular	817	64	17	35	96	605

B. Experimental Setup

In this work, we conducted an experiment based on 80 training and 20 testing splits on product review dataset [20]. The experiments are conducted similarly to [18], [30] to give comparison with existing methods. The experiments are conducted based on two baselines: Feature Subset Selection (FSS) that accomplishes two stages feature selection and Feature Fusion baseline that exploits feature fusion of linguistic features and FSS. Initially, twenty Special Feature (SF) are extracted using the linguistic approach from each review. On the other hand, various preprocessing techniques are applied to extract content based features in review text. The 20,985 distinct features are extracted and represented using unigram term frequency schema. Further, extracted features are processed to select discriminative feature using FSS. The Feature Subset Selection (FSS) consists of two stages: In the first stage, conventional feature selection methods such as Chi-square, IG and MI are used to select feature subsets. The features are arranged in descending order based on the scores obtained from individual feature selection methods. The threshold value is used to select top scored features from the feature space. The threshold value is varied between 1,000 to 20,000 numbers of features by empirically. In the second stage, k-means clustering is applied to select the most discriminative features from first stage feature subsets. The optimal number of cluster is determined based on [26] and it is explained in previous section. In Feature Fusion baseline, the extracted linguistic and selected content based features are fused. The extracted linguistic and selected content based features are represented using frequency of occurrence of each feature. Hence, both the features are fused and evaluated using SVM, LR, DT and RF classifiers. Further, individual predictions of classifiers

are ensemble based on weighted majority voting schema as explained in section III (F). The efficiency of the proposed method is evaluated based on Precision, Recall and F-measure obtained from individual classifiers and ensemble classifier.

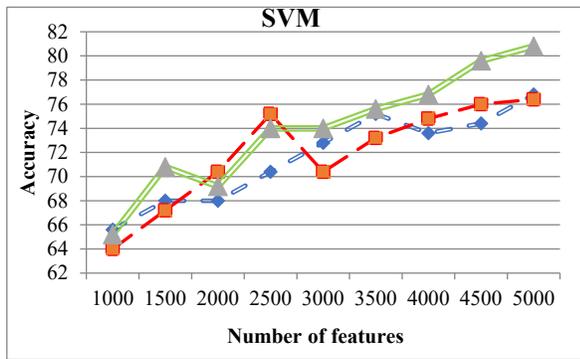
C. Experimental Results

To report the performance of the proposed method, the experiments are conducted in two baselines. In the first baseline, FSS consists of the two stage feature selection method: in the first stage threshold values are varied from 1,000 to 20,000 features and empirically is found that 10,000 features depict competitive results for various classifiers. Table III elucidates the comparisons of overall features with conventional feature selection method of the first stage of FSS using SVM, DT and RF classifiers. The Mutual Information (MI) with 10,000 features outperforms the other selection methods in accuracy, precision, recall and f-measure using SVM, DT and RF classifiers. The Information Gain (IG) with 10,000 feature subset achieves 0.658 recall compared to other feature selection methods using DT classifier. Overall from Table III, we can observe that feature selection method plays a vital role in reducing high dimensional feature space and maximizing the performance of the classifiers. Further, the second stage of FSS is applied on 10,000 features obtained from various feature selection methods.

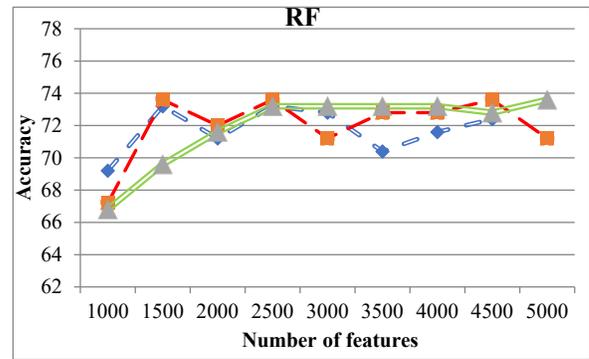
In the second stage of FSS method, k-means clustering algorithm is used by varying cluster numbers with difference of 500 from 100 to 5000 features. The k-means clustering with cosine distance yields promising results compared to Euclidean, City block, Correlation and Hamming Distance metrics [44] [45]. Hence, k-means clustering with cosine distance is used to select most discriminative features from the first stage FSS method. As explained in Algorithm 2, similar features are clustered and the features nearer to the cluster center are considered as the most discriminative feature subset. Fig. 2 depicts the classification accuracy of the proposed method with varying feature subsets using SVM, RF, LR and DT classifiers. Fig. 1 shows the feature subset ranging from 1000 to 5000 features with differences of 500 features. However, feature subset less than 1000 does not yield promising results. Hence, the feature subsets less than 1000 are excluded in Fig. 2. From Fig. 2 (a), we observe that the FSS of MI with k-means gradually increases and reaches maximum accuracy for 5000 features using SVM classifier. Similar variation is observed in Fig. 2 (b) and Fig. 2 (c), where MI with k-means clustering reaches maximum accuracy for 5000 features using RF and LR classifiers, respectively. In Fig. 2 (d), CHI-square with k-means clustering achieves maximum accuracy for 2500 features using DT classifiers. Overall, performance of MI with k-means feature subset increases with increasing number of feature subset using SVM, RF, LR and DT classifiers.

TABLE III. PERFORMANCE COMPARISON OF OVERALL FEATURES WITH REDUCED SET OF FEATURES;
A-ACCURACY, P-PRECISION, R-RECALL, F-F-MEASURE

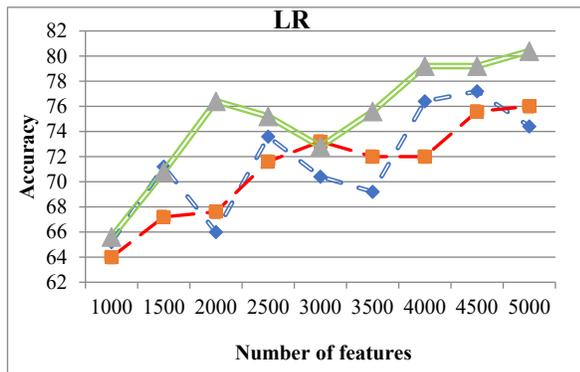
	No. of Features	SVM				DT				RF			
		A	P	R	F	A	P	R	F	A	P	R	F
Total features	20985	68.40	0.646	0.597	0.595	64.80	0.606	0.601	0.603	69.20	0.748	0.566	0.529
CHI	10000	73.60	0.771	0.637	0.638	68.40	0.645	0.626	0.631	69.60	0.755	0.571	0.539
IG	10000	73.60	0.756	0.642	0.646	70.80	0.675	0.658	0.663	70.00	0.761	0.577	0.548
MI	10000	75.60	0.791	0.666	0.675	72.00	0.693	0.657	0.664	71.60	0.821	0.595	0.572



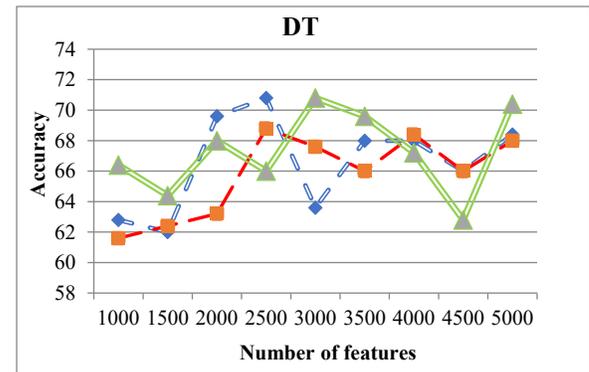
2 (a)



2 (b)



2 (c)



2 (d)

- ◆ - CHI-square + k-means - ■ - Information Gain + k-means - ▲ - Mutual Information + k-means

Fig. 2. Classification accuracy of feature selection methods with various feature subsets.

The feature fusion baseline consists of Special Features (SF) and FSS features. Both features are in same weighting scheme i.e., frequency of occurrence of each feature. From Fig. 2, we observe that feature subset ranging from 4000 to 5000 yield promising result using various classifiers. Hence, feature fusion experiments are conducted by varying FSS from 4000 to 5000 features with 20 Special Features using 10 fold cross validation method. The number of features considered for feature fusion (FSS+SF) experiments are 4020, 4520 and 5020. In Fig. 3, the box plot representation is given to describe the variations observed in F-measure of feature fusion with various FSS using different classifiers. The Fig. 3 represents minimum, maximum and mean of 10 fold cross validation using SVM, LR, RF and DT classifiers. Fig. 3 (a), (b) and (c) present the result of feature fusion of FSS (Chi) + SF feature subsets using SVM, LR, RF and DT classifiers. In Fig. 3 (b), the feature fusion of FSS (Chi-square) achieves maximum F-measure for 4520 features using LR classifier. Fig. 3 (d), (e) and (f) present F-measure results of feature fusion of FSS (IG) + SF feature subsets. Fig. 3 (e) shows feature fusion of FSS (IG) + SF achieving maximum F-measure for 4520 feature subset using RF classifier. Fig. 3 (g), (h) and (i) present the results of feature fusion of FSS (MI) + SF using SVM, LR, RF and DT classifiers. From Fig. 3 (h), we observe that feature fusion of FSS (MI) + SF features achieve maximum F-measure for 4520 feature subset using the LR classifier. From Fig. 3, we can

conclude that feature fusion (FSS (MI) + SF) for 4520 feature subset achieves maximum F-measure using SVM, LR, RF and DT classifiers. Hence, we consider feature fusion method (FSS (MI) + SF) with 4520 features for comparing with other existing methods.

Further, the classifiers are ensemble to seek collective opinions of the classifiers to enhance the performance of the models. Here, weighted majority voting schema is used to assign the weight for each decision of the classifiers. As explained in section III (F), weights are varied between 0 and 1, where the sum of the weights are equal to 1. The best combinations of weights are assigned empirically for each model. The Table IV presents the maximum accuracy obtained in 10 folds cross validation for the proposed feature fusion method (FSS (MI) + SF) with 4520 features using various classifiers. From Table IV, it can be inferred that SVM, DT and RF outperform LR classifier by a good margin in terms of accuracy, precision, recall and F-measure. The LR model performs least among other models due to the possibility of overfitting. Hence, the best possible weights assigned are 0.3 for SVM, 0.1 for LR, 0.3 for RF and 0.3 for DT classifiers, respectively. Table IV elucidates the performance comparison of the ensemble model with individual classifiers. The ensemble model outperforms other classifiers with a maximum margin in Accuracy, Precision, Recall and F-measure. Hence, the performance of the ensemble model

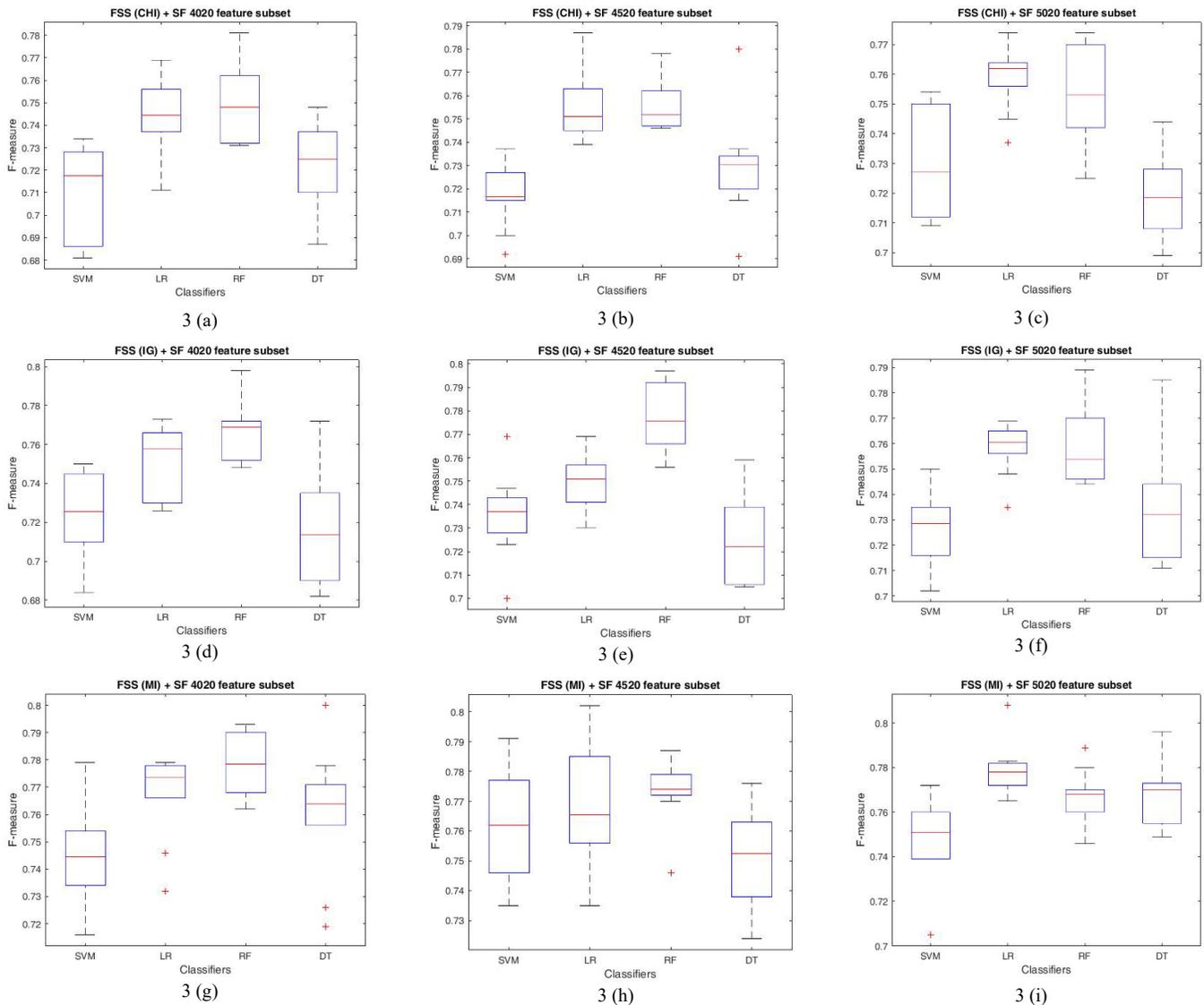


Fig. 3. F-measure of Feature Fusion with various Feature Subset Selection (FSS) using SVM, LR, RF and DT classifiers.

is considered for comparison with existing methods on product review dataset [20].

TABLE IV. PERFORMANCE COMPARISON OF ENSEMBLE MODEL

Classifiers	Accuracy	Precision	Recall	F-Measure
SVM	82.00	0.814	0.779	0.791
LR	78.80	0.769	0.754	0.760
DT	80.00	0.781	0.772	0.776
RF	82.40	0.838	0.769	0.787
Ensemble	86.00	0.871	0.818	0.835

D. Comparison with Existing Methods

From literature, Buschmeier et al. [18] and Reganti et al., [30] describe irony/satire detection on product review dataset [20] using SVM, LR, DT and RF classifiers. The similar set of experiment is conducted and compared the results. The Table V elucidates the comparisons of the proposed method with [18] and [30] method. In [18], models are evaluated based on precision, recall and F-measure using linear SVM, LR, DT and RF classifiers. The features are extracted and fused using lexicon based (29 special features) and content based approach (21744 Bag of Words features). The total number of features considered for the model are 21773 features. On the other hand, Reganti et al. [30] performed feature fusion on 42 special features and baseline features. The baseline features such as character n-gram, word n-gram and word skip gram are used but the total number of features are not stated clearly. The fused features are classified using linear SVM, LR, DT and RF classifiers. Further, the decisions of individual classifiers are ensemble using weighted majority voting schema by varying weighted from 0 to 1 value. The performance of the model was evaluated using F-Measure. Hence, we conducted our experiment on same lines with 20 Special Features (SF) and 4500 text features using individual classifiers and ensemble classifiers using weighted majority voting schema. However, product review dataset [20] is used to describe the impact of sarcasm detection in sentiment shift [19]. Therefore, the proposed method is compared with the methods proposed in [18] and [30]. From Table V, it can be observed that the proposed feature fusion method outperforms the existing methods in terms of Precision, Recall and F-measure using individual classifiers and ensemble classifiers.

TABLE V. PERFORMANCE COMPARISON WITH EXISTING METHODS

Classifiers	Performance Measure	Methods		
		Buschmeier et al., [18]	Reganti et al., [30]	Proposed Method
SVM	Precision	0.680	--	0.814
	Recall	0.613	--	0.740
	F-Measure	0.643	0.740	0.791
LR	Precision	0.752	--	0.769
	Recall	0.622	--	0.754
	F-Measure	0.678	0.768	0.760
DT	Precision	0.598	--	0.781
	Recall	0.550	--	0.772
	F-Measure	0.569	0.681	0.776
RF	Precision	0.732	--	0.838
	Recall	0.241	--	0.769
	F-Measure	0.353	0.710	0.787
Ensemble	Precision	--	--	0.871
	Recall	--	--	0.818
	F-Measure	--	0.781	0.835

E. Discussion

Feature fusion is a process of integrating various characteristics of features to produce more consistent, accurate, and useful information. In irony detection, feature extraction and selection plays a vital role in determining the ironic utterances present in sentences. In this work, linguistic based and content based Feature Subset Selection (FSS) are extracted and features are fused to provide various characteristics of features of product review dataset [20]. The Table II elucidates distribution of product reviews by its star-rating. The 20 linguistic features are extracted and represented based on frequency of occurrence of features in each review. Table I presents 20 Special Feature (SF) extracted using linguistic approach. These features are used to detect ironic utterance present in reviews. On the other hand, content based features play a vital role in accurate classification of ironic reviews. The high dimensionality and sparsity is one of the major challenges faced during classification task. To curbe the dimensionality, two stages of content based Feature Subset Selection (FSS) is proposed. In the first stage FSS, conventional Feature Selection Method (FSM) is applied to select relevant features set. In this experiment, conventional FSM such as Chi-square, IG and MI are used to select the relevant feature set from original features based on scores of individual FSM. It is noticeable from Table III, that first stage FSS using conventional FSM reduces high dimensionality feature space by selecting a subset of relevant features. The MI feature selection method yields maximum performance for 10,000 features compared to other FSM using SVM, LR, DT and RF classifiers, respectively. However, selected feature subset may have features which convey similar information. Due to this reason, second stage FSS is applied to select more discriminative features among feature subset. In the second stage, k-means clustering algorithm is used to group the features which convey similar information. Fig. 2 (a), (b), (c) and (d) present the classification accuracy of the proposed method with varying number of features (1000 to 5000) using SVM, RF, LR and DT classifiers. The MI with k-means consistently outrages other feature selection combination using various classifiers because MI compares the probability of observing features and class together (joint probability) instead of observing independently. MI reaches its maximum value, when the feature is a perfect indicator for class membership. On the other hand, k-means groups the similar information. The feature closer to cluster center is considered as the most representative feature within each cluster. The representative features are considered as discriminative features among other clusters. Hence, the combination of these methods exhibits maximum accuracy on all the classifiers. Further, features are fused to detect ironic utterances present in review using linguistic and content based FSS. A crucial observation is noted from Fig. 3 (a) – 3 (i), which depicts the F-measure of 10 folds cross validation of feature fusion using various classifiers. Fig. 2 (h) elucidates maximum F-measure on feature fusion (FSS (MI + k-means) +SF) of 4520 feature set using various classifiers. Further, the prediction of each classifier is ensemble to enhance the classification performance. The weighted majority voting scheme is used to ensemble different classifier based on the prediction of each classifier. Table IV elucidates the performance comparison of ensemble model with individual classifiers. The ensemble model outperforms other classifiers with a maximum margin of Accuracy, Precision, Recall and F-measure. To evaluate the effectiveness of the proposed method, the proposed method is compared with existing methods present in the literature. From Table V, it can be observed that the proposed method outperforms other existing methods in terms of minimal number of feature set and the maximum classification performance of individual and ensemble classifiers.

V. CONCLUSION AND FUTURE WORK

In this paper, we employed feature fusion to capture various dimensions of characteristics of the reviews. The proposed approach extracts features based on linguistic and content based text features. The five types of features are extracted using linguistic approach, viz: Rating, Word, Acronym, Symbol and Emoticon Features. On the other hand, content based text feature consists of two stages of Feature Subset Selection (FSS) to select the most discriminative features. Both the features are fused and classified using SVM, LR, RF and DT classifiers. With the series of experimentation, we demonstrated the proposed approach has an ability to capture ironic utterances present in the reviews. To enhance the performance of the classifiers, we make use of weighted majority voting schema to create an ensemble from the decision of each classifier. The results show that the proposed feature fusion out-performs the existing methods on benchmark dataset.

In future, the proposed approach can be extended to (i) extract more number of linguistic features to identify ironic utterances present in text, (ii) use of more sophisticated feature selection methods and (iii) employing various clustering algorithms along with ensemble methods which enhances the classification accuracies. Further, the proposed approach can be also extended to various fields such as sentiment classification, spam detection and many more.

ACKNOWLEDGMENT

H M Keerthi Kumar has been financially supported by UGC under Rajiv Gandhi National Fellowship (RGNF) Letter no: F1-17.1/2016-17/RGNF-2015-17-SC-KAR-6370/(SA-III Website), JSSRF (University of Mysore), Karnataka, India.

REFERENCES

- [1] M. P. S. Bhatia, and A. K. Khalid, Information retrieval and machine learning: Supporting technologies for web mining research and practice. Webology, 2008, vol. 5, pp. 2–19.
- [2] G. Wang, J. Sun, J. Ma, K. Xu, and J. Gu, “Sentiment classification: The contribution of ensemble learning”. Decision Support Systems, 2014, vol. 57, pp. 77–93.
- [3] E. Fersini, E. Messina, and F. A. Pozzi, “Sentiment analysis: Bayesian ensemble learning”. Decision Support Systems, 2014, vol. 68, pp. 26–38.
- [4] A. Onan, S. Korukoğlu, and H. Bulut, “A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification”. Information Processing & Management, 2017, vol. 53(4), pp. 814-833.
- [5] V. David. “Towards a sociology of attitudes”, Sociological Research Online, 2011, vo. 19.1, pp. 12.
- [6] R. J. Davidson, K. R. Sherer, and H. H. Goldsmith, “Handbook of affective sciences”, Oxford University Press, 2009.
- [7] J. R. Martin, and R. Peter White, “The language of evaluation”. 2003, vol. 2. Basingstoke: Palgrave Macmillan.
- [8] B. Pang, and L. Lee, “Opinion mining and sentiment analysis”. Foundations and Trends in Information Retrieval, 2008 vol. 2(1–2), pp. 1-135.
- [9] C. C Florian, F. A. Kunneman, and A. P. J. Van den Bosch, “The perfect solution for detecting sarcasm in tweets #not”, In Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA 2013.
- [10] I. Hernández-Farías, J. M. Benedí, and P. Rosso, “Applying basic features from sentiment analysis for automatic irony detection”, In Iberian Conference on Pattern Recognition and Image Analysis, 2015, pp. 337-344. Springer, Cham.
- [11] S. Attardo, “Irony markers and functions: Towards a goal-oriented theory of irony and its processing”, Rask, 2000, vol. 12(1), pp. 3-20.
- [12] R. González-Ibáñez, S. Muresan, and N. Wacholder, “Identifying sarcasm in Twitter: a closer look”, In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2, pp. 581-586, Association for Computational Linguistics. 2011
- [13] S. G. Wicana, T. Y. İbisoglu, and U. Yavanoglu, “A Review on Sarcasm Detection from Machine-Learning Perspective”, In Semantic Computing (ICSC), 2017 IEEE 11th International Conference on, pp. 469-476. IEEE.
- [14] A. D. Dave, and N. P. Desai, “A comprehensive study of classification techniques for sarcasm detection on textual data”, In Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on, pp. 1985-1991. IEEE, 2016.
- [15] S. K. Bharti, B. Vachha, R. K. Pradhan, K. S. Babu, and S. K. Jena, “Sarcastic sentiment detection in tweets streamed in real time: A big data approach”, Digital Communications and Networks, 2016, vol. 2(3), pp. 108-121.
- [16] A. Joshi, P. Bhattacharyya, and M. J. Carman, “Automatic Sarcasm Detection: A Survey”, ACM Computing Surveys (CSUR), 2017, vol. 50(5), pp. 73.
- [17] K. Ravi, and V. Ravi. “A novel automatic satire and irony detection using ensembled feature selection and data mining”, Knowledge-Based Systems, 2017, vol. 120, pp. 15-33.
- [18] K. Buschmeier, P. Cimiano, and R. Klinger. “An Impact Analysis of Features in a Classification Approach to Irony Detection in Product Reviews”, In WASSA@ACL, pp. 42-49. 2014.
- [19] E. Filatova, “Sarcasm Detection Using Sentiment Flow Shifts”, Florida Artificial Intelligence Research Society Conference, 2017, pp 264-269.
- [20] E. Filatova, “Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing”, In LREC, 2012, pp. 392-398.
- [21] R. Justo, T. Corcoran, S. M. Lukin, M. Walker, and M. I. Torres, “Extracting relevant knowledge for the detection of sarcasm and nastiness in the social web”. Knowledge-Based Systems, 2014, 69, pp. 124-133.
- [22] D. Davidov, O. Tsur, and A. Rappoport, “Semi-supervised recognition of sarcastic sentences in twitter and amazon”, In Proceedings of the fourteenth conference on computational natural language learning, 2013, pp. 107-116. Association for Computational Linguistics.
- [23] M. Hu, and B. Liu, “Mining and summarizing customer reviews”, Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2004.
- [24] R. W. Gibbs, “Irony in talk among friends”, Metaphor and symbol, 2000, vol. 15(1-2), pp. 5-27.
- [25] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques”, In Proceedings of the ACL-02 conference on Empirical methods in natural language processing, 2002, vol. 10, pp. 79-86. Association for Computational Linguistics.
- [26] S. Dey Sarkar, S. Goswami, A. Agarwal, and J. Aktar, “A Novel Feature Selection Technique for Text Classification Using Naïve Bayes”, International Scholarly Research Notices, 2014.
- [27] A. Reyes, and P. Rosso, “Mining subjective knowledge from customer reviews: A specific case of irony detection”, In Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis, 2011, pp. 118-124. Association for Computational Linguistics.
- [28] J. Fürnkranz, “A study using n-gram features for text categorization”, Austrian Research Institute for Artificial Intelligence, 1998, vol. 3, pp. 1-10.
- [29] G. Li, X. Hu, X. Shen, X. Chen, and Z. Li. “A novel unsupervised feature selection method for bioinformatics data sets through feature clustering”, In Granular Computing, 2008, IEEE International Conference on, pp. 41-47. IEEE,
- [30] A. N. Reganti, T. Maheshwari, U. Kumar, A. Das, and R. Bajpai, “Modeling Satire in English Text for Automatic Detection”, In Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference on (pp. 970-977). IEEE.
- [31] L. Rokach, “Ensemble-based classifiers” Artificial Intelligence Review 33, no. 1-2 (2010), pp. 1-39.
- [32] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, “Sentiment analysis of twitter data”. In Proceedings of the workshop on languages in social media, 2011, pp. 30-38. Association for Computational Linguistics.
- [33] A. B. Musa, “Comparative study on classification performance between support vector machine and logistic regression”, International Journal of Machine Learning and Cybernetics, 2013, Vol. 4(1), pp. 13-24.
- [34] I. H. Witten, E. Frank, M. A. Hall, C. J. Pal, DataMining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.

- [35] A. Bouaziz, C. Dartigues-Pallez, C. da Costa Pereira, F. Precioso and P. Lloret, "Short text classification using semantic random forest", in: International Conference on Data Warehousing and Knowledge Discovery, Springer, 2014, pp. 288–299.
- [36] Y. Al Amrani, M. Lazaar, and K. E. El Kadiri, "Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis", *Procedia Computer Science*, 2018, vol. 127, pp. 511-520.
- [37] T. G. Dietterich, "Ensemble methods in machine learning", In International workshop on multiple classifier systems, 2000, pp. 1-15, Springer, Berlin, Heidelberg.
- [38] A. Onan, S. Korukoğlu, and H. Bulut, "A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification", *Information Processing & Management*, 2017, vol.53(4), pp. 814-833.
- [39] Y. Zhang, H. Zhang, J. Cai, and B. Yang, "A weighted voting classifier based on differential evolution". In *Abstract and Applied Analysis*, 2014 vol. 2014. Hindawi.
- [40] L. I. Kuncheva, and J. J. Rodríguez, "A weighted voting framework for classifiers ensembles". *Knowledge and Information Systems*, 2014, vol. 38(2), pp. 259-275.
- [41] A. Ekbal, and S. Saha, "Weighted vote-based classifier ensemble for named entity recognition: A genetic algorithm-based approach". *ACM Transactions on Asian Language Information Processing (TALIP)*, 2011, vol. 10(2), pp. 9.
- [42] R. J. Kreuz, and R. M. Roberts, "Two cues for verbal irony: Hyperbole and the ironic tone of voice". *Metaphor and symbol*, 1985, vol. 10(1), pp. 21-31.
- [43] A. Utsumi, "Verbal irony as implicit display of ironic environment: Distinguishing ironic utterances from nonirony", *Journal of Pragmatics*, 2000, vol. 32(12), pp. 1777-1806.
- [44] A. Strehl, J. Ghosh, and R. Mooney, "Impact of similarity measures on web-page clustering", In *Workshop on Artificial Intelligence for Web Search (AAAI 2000)*, pp. 58-64.
- [45] L. Sahu, and B. R. Mohan, "An improved K-means algorithm using modified cosine distance measure for document clustering using Mahout with Hadoop", In *Industrial and Information Systems (ICIIS)*, 2014 9th International Conference on, pp. 1-5. IEEE.
- [46] K. Chakraborty, S. Bhattacharyya, R. Bag, and A. E. Hassanien, "Comparative Sentiment Analysis on a Set of Movie Reviews Using Deep Learning Approach", In *International Conference on Advanced Machine Learning Technologies and Applications*, 2018, pp. 311-318.
- [47] O. Tsur, D. Davidov, and A. Rappoport. "ICWSM-A Great Catchy Name: Semi-Supervised Recognition of Sarcastic Sentences in Online Product Reviews", In *ICWSM*, 2010, pp. 162-169.
- [48] B. S. Harish, M. B. Revanasiddappa, "A comprehensive survey on various feature selection methods to categorize text documents". *International Journal of Computer Applications*, 2017, vol. 164, pp. 1–7.
- [49] A. Yousefpour, R. Ibrahim, and H. N. A. Hamed, "Ordinal-based and frequency-based integration of feature selection methods for sentiment analysis". *Expert Systems with Applications*, 2017, vol. 75, pp. 80-93.
- [50] A. Tripathy, A. Agrawal, S. K. Rath, "Classification of sentiment reviews using n-gram machine learning approach". *Expert Systems with Applications*, 2016, vol. 57, pp. 117–126.
- [51] A. Joshi, V. Sharma, P. Bhattacharyya, "Harnessing context incongruity for sarcasm detection", *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics Int. Joint Conf. Natural Lang. Process. (ACL-IJCNLP)*, 2015, vol. 2, pp. 757-762.
- [52] M. Bouazizi, and T. O. Ohtsuki, "A pattern-based approach for sarcasm detection on Twitter", *IEEE Access*, 2016, vol. 4, pp. 5477-5488.
- [53] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey". *Ain Shams Engineering Journal*, 2014, vol. 5(4), pp. 1093-1113.
- [54] H. Toman, L. Kovacs, A. Jonas, L. Hajdu, and A. Hajdu, "Generalized weighted majority voting with an application to algorithms having spatial output", In *International Conference on Hybrid Artificial Intelligence Systems*, 2011, pp. 56-67.
- [55] U. Yavanoglu, T. Y. Ibisoglu, and S. G. Wicana, "Technical Review: Sarcasm Detection Algorithms". *International Journal of Semantic Computing*, 2018, vol. 12(03), pp. 457-478.
- [56] H. M. Kumar, B. S. Harish, and H. K. Darshan. "Sentiment

Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method". *International Journal of Interactive Multimedia and Artificial Intelligence*, 2019, vol. 5(5), pp. 109-114.

- [57] H. M. Kumar and B. S. Harish. "Classification of Short Text Using Various Preprocessing Techniques: An Empirical Evaluation." In *Recent Findings in Intelligent Computing Techniques* pp. 19-30. Springer, Singapore, 2018.



H. M. Keerthi Kumar

He received B.E in Information Science and Engineering and M.Tech in Software Engineering from Visvesvaraya Technological University, India. He is currently pursuing Ph.D degree in Computer Science from University of Mysore, India. His area of research includes Data Mining, Pattern Recognition and Machine Learning.



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, and 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.