iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** IV **Month of publication:** April 2017

DOI: http://doi.org/10.22214/ijraset.2017.4044

www.ijraset.com Volume 5 Issue IV, April 2017
IC Value: 45.98 ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

A Hybrid Approach for Emotion Extraction

Mr. Ahefaz Pothiyawala[1], Mrs. Kavita R. Singh[2]
[1]M.Tech Student , [2]Associate Professor, Department of Computer Science and Engineering
Yeshwantrao Chavan College of Engineering Nagpur, Nagpur, India

Abstract: With the growth of internet the amount of content that gets produced everyday has increased significantly. Many different types of textual/image documents are produced on varying types of platforms such as social media, news, blogs etc. This paper presents a hybrid approach for emotion extraction from the text documents. With such large amounts of textual content being produced every day, one cannot ignore the importance of mining and extracting opinions, sentiments and also emotions that are embedded in the text. The task of opinion mining, extraction and sentiment analysis has been extensively used in many fields. This is because emotions form an important part of any communication and opinion and have profound effect on the sentiments that the text represents. This makes emotion recognition or extraction from text as well as other media of communication such as speech images videos a crucial task. Emotion analysis is the task of making the computers makes sense of emotions. Emotion extraction is focused on delivery of the exact emotional state of a user at the time when the document was being prepared. In this paper, the main focus of the proposed work is to improve emotion extraction and the study of different approaches and methods that have already been developed and deployed for use. A number of approaches have been developed for emotion extraction and each of them has some strengths and weaknesses. Many machine learning techniques have been used such as Support Vector Machines, Naive Bayes classifier, Vector Space Model and others. Many others combine machine learning techniques with other rule-based techniques or keyword-based techniques. Our approach combines the SVM with k-NN algorithm while still employing the rule based technique to take the advantage of all while avoiding their weaknesses. Our algorithm will be implemented and evaluated using a standard benchmark GRIMM's data set. The new approach will be evaluated on the basis of accuracy of the emotions that are assigned to the sentences in a document

Keywords: Opinion Mining, Sentiment Analysis, Emotion Extraction, SVM, k-NN

I. INTRODUCTION

Emotions are an intrinsic part of any communication that occurs. Many day to day decisions made by a person are based on the type of emotions that a person feels in the moment. Emotions also play an important role in influencing ones behavior towards other entities or objects. Emotions go deep into a person's state of mind. Therefore, in order to understand why a person engaged in an activity we need to be able to extract and analyse the type of emotion the person felt prior to his/her engagement in the activity. Also the rapid increase in the emotion rich content available on the internet has caught the attention of researchers regarding emotion extraction, sentiment analysis and opinion mining. Due to the growing use of internet, more and more emotional information is gathered at one place from which emotions are required to be extracted and analysed. Emotion extraction is a method of finding emotions of an author about particular entity. As a result emotion analysis has received considerable attention in recent years.

Emotion analysis is also closely related to the tasks of opinion mining and sentiment analysis, however it is a fine grained task as it attempts to assign a single exact emotion value from a pool of emotion values or labels that exist in real life as opposed to the basic orientation values of positive, negative and neutral that are often used in opinion mining.

Different approaches for emotion analysis [8], text classification and opinion mining have been developed. First is the keyword-based approach used in [1] [2] [4] [6] [13] [14] [11] that depend on the presence of opinion and emotion conveying words. Next, is the learning-based approach used in [3] [5] [9] [10] [12] [13] that utilizes a trained classifier to perform the task and third is hybrid approach [8] [12] that combine keyword-based approach and learning-based approach. The use of any particular approach depends on the application. The applications for emotion analysis generally include different domains. Some of these applications include applications in government [1] [7] [15], applications in shopping and in marketing [4] [6] [11] [10] [13] [14] [16] and applications for entertainment [2] [3] [4] and other purposes [17].

The work presented in this paper is focused on the development of a hybrid method which will accept a text document in english and evaluate the emotion for every sentence in the document. The main focus is to design a approach that would efficiently and accurately extract the emotions from a given text.

A hybrid approach which utilizes both rule based engine and machine learning techniques is developed. To solve the multi class problem a rule-based engine and an SVM model is used for each category of the emotion. The sentences which are not assigned any emotion by the SVM are further classified using the k-NN classifier. A set of syntactic and semantic features are selected for each sentence to build the rules and train the classifiers. The rules are generated manually based on a set of lexical and emotional clues.

In the following section we have provided a brief introduction to the different systems that have been developed for emotion extraction.

II. RELATED WORK

A short study of the related work and different systems proposed for emotion extraction and detection is done here. Different systems have used different approaches which are broadly classified into three categories namely keyword-based methods, learning-based methods and hybrid methods by Nilesh M. Shelke et al., in [8]. Keyword-based emotion detection approaches are those in which the emotions are detected based on the related set(s) of keywords found in the input text. Learning-based methods try to detect emotions based on a previously trained classifier, which applies different theories of machine learning. The last of the approach is hybrid method and it combines keyword-based methods with learning-based methods. The approach adopted by Wingyan Chung et al., in [1] is keyword-based where the authors have used a lexicon of 13901 words. The focus of the research in [1] is limited to only US immigration policies. The advantage of the work done is that it helps to pinpoint the most influential users in a social networking environment and also helps compute the emotion entrainment score of the influential users which is helpful in analysing the effect of the emotions that the most influential users have on others. The authors Akshi Kumar et al., in [2] have performed emotion extraction from tweets for a particular hashtag or username. The advantage of this is that the approach in [2] can be applied to a wide range of tweets and is not limited to only a particular topic. The authors have also presented different methods to score a tweet. This is helpful in case if another application uses a different set of features to extract emotions for a different topic then a different method for scoring the tweets can be chosen and applied. The authors Lily Dey et al., in [3] have proposed an approach for emotion extraction from a real time chat messenger. The authors have constructed a lexicon of 5000 words and have also focused on emotion detection from emoticons which are now widely used in day to day communication and are used to express different types of emotions. The emotions are extracted at different levels by considering simple, compound and complex sentences as well. The authors ChengYu Lu et al., in [4] have proposed novel approach to emotion extraction. The advantage of the approach adopted by the authors is that it tries to model real life events. The approach is a combination of keyword-based and a semantic role labeling technique. The system developed is used to extract emotions in online chatting environment. The keyword-based approach is used in presence of emotion conveying words, however when such emotion words are not present the semantic role labeling with web scale mining extracts the events that can occur between any two entities in real life and assigns the emotion accordingly. This way an emotion, that a person actually feels in real life, in the presence of some other entity gets assigned. In [5] the authors Mansur Alp Tocoglu et al. , have proposed a system to extract emotions from turkish text. The approach used by the authors is learning-based where in the authors have manually constructed a dataset and used the Vector Space Model as the classifier for emotions. In [6] the authors Luwen Huangfu et al., have made use of the well established OCC model for emotion extraction from online reviews. A dictionary of emotion words is constructed by the authors for every dimension they have used from the OCC model and then the dictionary is subsequently refined using WordNet and publish word list for errors and noise. The dictionary is expanded using WordNet and syntactic templates. The OCC model employs rules to detect emotions in different dimensions that it contains. The authors Azadeh Nikfarjam et al., have proposed a hybrid approach for emotion extraction from suicide notes in [7]. The authors have used support vector machines in combination with rule engines to extract emotions. They have extracted 15 emotions. The authors have extracted clues and features for the system based on semantic roles and other syntactic constructs. The rules are manually generated and syntactic and semantic features that have been extracted are used for rule generation as well as for training the classifier. Many systems for opinion mining, sentiment analysis and emotion extraction use machine learning algorithms for classification these include SVM, k-NN, Naive Bayes, Vector Space models and others. In [9] the authors J. S. Raikwal et al., have provided an insight into the working of the SVM and k-NN classifiers and evaluated the performance of the two based on parameters like model build time, search time, memory required and performance under dataset of different sizes.SVM clearly outperforms k-NN. In [10] the authors Taysir Hassan A. Soliman et al., have used keyword-based approach for opinion mining. The authors have proposed a two level classification of the reviews in which the first classification is done based on the feature and the second classification on the polarity assigned to the words describing the feature. The approach is suitable and works for systems that include explicit words for describing the features of a product. The two level classification further helps in deciding about which feature of the product an opinion talks about. An approach which uses more semantic features for processing of opinions is

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

combined with keyword-based approach in [11] by authors Xiaojun li et al.,. The authors have constructed an emotional lexicon and a feature lexicon for the system. The authors have used the emotion words and assigned intensities to each and computed an orientation value by taking into account impact of the adverbs on them. The advantage of the system is that the opinion is classified using both the emotion role and the feature role contained in it. To improve text classification the authors Yun Li et al,. in [12] have proposed the combination of SVM and k-NN algorithm. The combined approach for classification clearly outperforms either SVM or k-NN for the classification task of 10000 webpages used for experimentation by the authors. The system provides improved results because SVM fails to classify the web pages that fall between the margins of the maximum margin hyperplane and these web-pages in turn are then classified using k-NN classifier. In [13] the authors Jibran Mir et al., have proposed keyword-based approach to a conceptual model for aspect based opinion mining .In [14]the authors Vijay B. Raut et al., have used three different machine learning algorithms namely SVM, Naive Bayes and decision tree for performing classification of hotel reviews and compared the outputs of all.

III. METHODOLOGY

The idea is to extract the emotions using a technique which combines different fairy tales from standard GRIMM's dataset of fairy tales to identify nine artistic emotions as described as: pleasure, mirth, anger, energy, fear, disgust, astonish, serenity, sorrow. The architecture of the proposed work is shown in the figure 1.The input to the system is a single text document in English Natural Language and the output is the emotion label for each sentence. The proposed approach undergoes three phase's viz., Pre-processing Phase, Feature Extraction Phase,

Polarity Measurement Phase and Emotion Extraction Phase. The Pre-processing Phase consists of sentence segmentation, tokenization stop word removal and stemming. The Feature Extraction Phase consists of parsing the sentences in the document, Part-of-Speech tagging of the sentences and Name Entity Recognition for the sentences in the document.The Polarity Measurement Phase consists of the calculation of the sentiment polarity as negative and positive for each sentence in the document. The Emotion extraction phase uses the Rule-Engine and the combination of the SVM and k-NN algorithms to extract and label the emotions for each sentence present in the document.

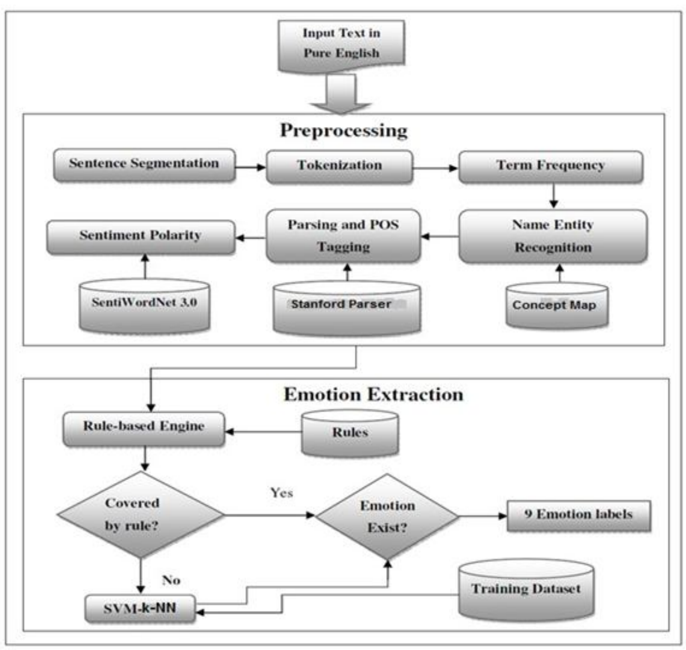


Fig.1. Architecture of Proposed approach

A. Algorithm

Input: Accept a document in pure English (.txt)

Output: Emotion label (pleasure, mirth, sorrow, anger, energy, fear, disgust, astonishment, serenity)

Accept a text document in Pure English as input

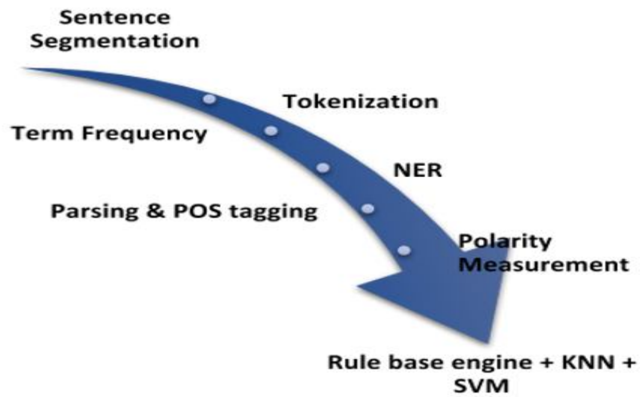
For each document perform

1) Sentence Segmentation
2) Tokenization
3) Calculate the term frequency
4) Perform Name Entity Recognition (NER)
5) Parsing and POS Tagging
6) Calculate Sentiment Polarity
7) For each sentence perform Emotion Detection by using rule base engine, SVM algorithm
8) If sentence undetected then for each sentence perform Emotion Detection by using rule base engine, KNN and SVM algorithm

B. Sample Input Document

Cinderella thanked him, went to her mother's grave, and planted the branch on it, and she wept so much that her tears fell upon it and watered it. It grew and became a beautiful tree.

C. Output: Emotion Labels



D. The Proposed Approach Consists of Four Phases

1) *Pre-Processing Phase:* A computer, as we know, is not capable of thinking or taking decision on its own. Rather it is impossible for a computer to independently analyze the data provided as an input and give the solution. So there is a need of a program to perform all these functions for the computer, where given a text in English, it needs to be pre-processed. Initially, the entire document is considered to be an input.

a) *Sentence Segmentation:* Sentence Segmentation is a process of breaking the string text into sentences. This is done by searching for punctuation mark, particularly full stop in English Language. Once it is found the string of text is split into component sentences. These sentences are then saved in a separate file in order to provide it as an input to the next module of tokenization.

Algorithm for Sentence Detection

Step 1: Accept the filtered document as an input.

Step 2: Compare each character of the filtered document with the special character “.”.

If match found split the sentence and display it.

Step 3: Repeat the step till every character from the filtered document gets processed.

Input Cinderella thanked him, went to her mother's grave, and planted the branch on it, and she wept so much that her tears fell upon it and watered it. It grew and became a beautiful tree.

Output: Cinderella thanked him, went to her mother’s grave, and planted the branch on it, and she wept so much that her tears fell upon it and waters it. It grew and became a beautiful tree.

b) *Tokenization:* Tokenization is a task of breaking sentences into tokens. This is done by searching a space after each word. All the generated tokens are saved in a separate file for further processing. The list of tokens becomes input for further processing such as part-of-speech tagging and named entity recognition. It locates word boundaries.

Algorithm for Tokenization

Step 1: Accept the input document

Step 2: From beginning scan the input document and identify the white space

Step 3: If space is found then word before the space/ word between two spaces are identified as a token

Step 4: Repeat the process until all the text is considered.

Input: It grew and became a beautiful tree.

Output (tokens)

It grew and became a beautiful tree

c) *Term Frequency:* Term frequency is defined as the number of times a word occurs in the document. Term frequency generally depends on the length of the document, i.e., a term may occur more frequently in a large document as compared to a small document.

Algorithm for calculating Term Frequency

Step 1: Accept the tokens from tokenization module

Step 2: Initialize count to 0

Step 3: For each token increment the count and store it in separate file.

Repeat step until all the tokens are completed.

Input Cinderella thanked him, went to her mother's grave, and planted the branch on it, and she wept so much that her tears fell upon it and watered it. It grew and became a beautiful tree.

Output Cinderella <0> thanked <0> him <0>, went <0> to<0> her<1> mother's<0> grave<0>, and<3> planted<0> the<0> branch<0> on<0> it<3>, and<3> she <0> wept<0> so<0> much<0> that<0> her<1> tears<0> fell<0> upon<0> it<3> and<3> watered<0> it<3>. It<3> grew<0> and<3> became<0> a<0> beautiful<0> tree<0>.

2) *Feature Extraction phase:* This module takes the input from previous phase and generates parsing and POS and NER for Feature extraction. In this step feature are getting calculated for each sentences on the basis of different factors which checks for linguistics as well as statistics for every sentences.

a) *Name-Entity Recognition (NER):* Named Entity Recognition (NER) labels sequences of words in a text which are the names of things, such as person and company names, or gene. It comes with well-engineered feature extractors for Named Entity Recognition, and many options for defining feature extractors. Good named entity recognizers for English, particularly for the 3 classes (PERSON, ORGANIZATION, and LOCATION) are available. Stanford NER tool and Open NLP tool are available for doing the tasks.

Input Cinderella thanked him, went to her mother's grave, and planted the branch on it, and she wept so much that her tears fell upon it and watered it.

Output Cinderella -> Person

b) *Parsing and Part-of-Speech Tagging (POS):* A Part-Of-Speech Tagger (POS Tagger) is a piece of software that reads text in some language and assigns parts of speech to each word, such as noun, verb, adjective, etc. It is an extremely powerful and accurate tool. It can be used in any application that deals with natural language text to analyze words/tokens and classify them into categories. Stanford POS tagging tool is available and can be used as a plug in our project. The Stanford POS Tagger uses a probability model to predict the correct pos tag out of the tag set. To limit the possible tags for a token a tag dictionary can be used which increases the tagging and runtime performance of the tagger.

Input Cinderella thanked him, went to her mother's grave, and planted the branch on it, and she wept so much that her tears fell upon it and watered it.

Output (ROOT (S (S (NP (NNP Cinderella)) (VP (VBD thanked) (SBAR (S (NP (PRP him)) (VP (VP (VBD went) (PP (TO to) (NP (NP (PRP$ her) (NN mother) (POS 's)) (NN grave)))) (CC and) (VP (VBD planted) (NP (DT the) (NN branch)) (PP (IN on) (NP (PRP it))))))))) (CC and) (S (NP (PRP she)) (VP (VBD wept) (ADVP (RB so) (RB much)) (SBAR (IN that) (S (NP (PRP$ her) (NNS tears)) (VP (VP (VBD fell) (PP (IN upon) (NP (PRP it)))) (CC and) (VP (VBD watered) (NP (PRP it)))))))) (. .)))#(ROOT (S (NP (PRP It)) (VP (VP (VBD grew)) (CC and) (VP (VBD became) (NP (DT a) (JJ beautiful) (NN tree)))) (. .)))#

3) *Sentiment Polarity Calculation:* Here we need to assign the Score (polarity) to each sentence which is based on the weight of feature terms. Sentiments are generally classified into two groups such as positive or negative. But determining a sentiment as positive or negative is not just enough. It is also necessary to analyze the intensity of that sentiment i.e. how positive or how negative the sentiment is. This is done with the help of SentiWord Net 3.0. As per SentiWordNet, Sentiment Polarity is given as follows.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Sentiment Polarity = 1 – (Pos Score + Neg Score).
Algorithm for Sentiment Polarity calculation
Step 1: Accept the input text document
Step 2: For each sentence in the document ObjScore = 1- (PosScore + NegScore)
Step 3: Repeat the step until all the sentences in the document are scanned.
Input Cinderella thanked him, went to her mother's grave, and planted the branch on it,
and she wept so much that her tears fell upon it and watered it.
Output (Polarity) -0.301025

4) *Emotion Extraction Phase:* This is the main phase of the system wherein the emotion labels are extracted using rule-based engine, SVM and KNN. In this phase set of rules are applied to the input text document.

Here we have two different flow for analysis of the sentence processing by rule_based combines with SVM and Rule_based combines with k-NN and then it will relate to SVM for final output.

a) *Rule Based +SVM Approach:* Rules are always help us to judge the proper emotions that covers by the system. In this section, we covers around nine rules to find correct emotions by specifying rules.

Rules covered in proposed approach:
Pleasure
For eg.: I am so happy.
Rule 2: IF the person describes him/herself afraid from someone and the sentence is negative
then
Fear
For eg.: Daddy will scold me.
Rule 3: IF the person describes him/herself being angry on someone and the sentence is negative
then
Anger
Rule 4: IF the person is in shock or in a state of confusion
then
Astonish
For eg.: The surprise attack was devastating
Rule 5: IF the person express dislike or hate towards someone or something and the sentence is negative.
then
Disgust
For eg.: Due to her aversion to the outdoors she complained throughout the entire camping trip.
Rule 6: IF the person feels endurance or forcefulness and the sentence may be positive or negative.
then
Energy
For eg.: 1) The outbreak of World War II in 1939 gave a new impetus to receiver development.
2) He recited the story with great animation.
Rule 7: IF the person feels glad or cheerful
then
Mirth
For eg.: 1) The decoration added greatly to the gaiety of the room.
2) His presentation pleased the executives
Rule 8: If the person feels peace of mind or calm
then
Serenity
For eg.: 1) She looked at her students with joviality and serene mentality.
2) Her soul rested in harmony.
Rule 9: IF the person feels miserable or heartbreaking
then

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Sorrow

For eg.: She was worn out from so much grief.

Some of the developed rules for emotion extraction are implemented as below

If is_first_person(subj) and verb_aux is member of[do,shall,have..] and verb polarity >0 and verb_tense== present and verb is a member of [happy,good..] and sentence polarity is positive then assign emotion label as pleasure

if is_first_person(subj) and verb_aux is member of[will,shall,..] and verb polarity >0 and verb_tense== future/present and verb is a member of [calm ,serene,rest..] and sentence polarity is positive then assign emotion label as serene

if is_first_person(subj) and verb_aux is member of[might,been, may..] and verb polarity <0 or is_negated(verb) and verb is a member of [miserable,break,..] and sentence polarity is negative then assign emotion label as sorrow

if is_first_person(subj) and verb_aux is a member of [can,could,will..] and is_negated(verb) or verb polarity >0 and verb_aux is a member of [not ,can't..] and verb_tense== past and verb is a member of[run,kill,..] and sentence polarity is negative then assign emotion label as anger

if is_first_person(subj) and verb_aux is member of[was,has,did..] and is_negated(verb) or verb polarity <0 and verb_tense== past and verb is a member of[shock,confuse,..] then assign emotion label astonishment

if is_first_person(subj) and verb_aux is member of[could,would,should..] and verb polarity <0 or is_negated(verb) and verb is a member of [hate,like,dislike..] and sentence polarity is negative then assign emotion label as disgust

if is_first_person(subj) and verb_aux is member of[do,shall,have..] and verb polarity >0 and verb_tense== present and verb is a member of [endurance,force,..] then assign emotion label as energy

if is_first_person(subj) and verb_aux is member of[could,was,will..] and verb polarity <0 and verb_tense== past/future and verb is a member of [scold,afraid,..] and sentence polarity is negative then assign emotion label as fear

if is_first_person(subj) and verb_aux is member of[do,shall,have,..] and verb polarity >0 and verb is a member of [glad,cheer,great,..] and sentence polarity is positive then assign emotion label as mirth

If the rule-based engine does not extract the emotional label then such sentences are forwarded to the SVM. SVM builds a hyper-plane based on the polarity calculated. Here the positive values are on one side of the hyper-plane while the negative values are on the other side.

While implementing this system on the Grimm's dataset it was observed that there are some polarity values that lie on the hyper-plane and close to hyper plane boundary and the emotion labels for such valued sentences are not extracted. In order to overcome this problem, k-NN algorithm is used.

b) *Rule Based + k-NN + SVM Approach:* Rule engine is not able to cover all the emotions efficiently so we need classifier to extract the emotions, to do this we have used a trained SVM classifier. The SVM classifier is trained using a corpus of 500 emotion conveying words which are manually collected by carefully studying the GRIMM's dataset of fairy tales. SVM basically covers all the data which falls in dense regions but does not cover data on the hyper-plane. The sentences whose polarity value belong to the hyper-plane thereafter are covered by k-NN algorithm which considers an optimal value of k(2 in our case) and finds all the nearest neighbors. This reduces the work of SVM and SVM finally provides the emotion label.

Different standard distance metrics were evaluated and studied namely the manhattan distance ,the euclidean distance and the minowski distance. The euclidean distance measure is chosen and used for computation of the nearest neighbors for each of the input sentence in the text. By experimenting with different values of k(1,2,3,4) for the k-NN ,it was observed that for k=1 the recall of the system increases significantly and that for k=3 and k>3 the accuracy gets reduced. Therefore the optimal value of k=2 has been chosen for the k-NN classifier. A data-set of 1000 sentences annotated with different emotion class labels has been developed by carefully examining and studying the standard GRIMMS data-set of stories.

IV. EXPERIMENTAL RESULTS

The system is evaluated to check whether the output generated by the current system is efficient than that of the existing system or not. Various performance measures such as accuracy, precision, recall and f-measure score has been used to check the same. To check the accuracy and efficiency of the system number of sentence in the document are used as unit for evaluation. In order to perform these evaluations, some notations are used.

They are as follows:

DS1, DS2, DS3,……. – Represents input documents from Grimm's test dataset used for testing.

C1, C2,…… C9 - Represents the class of emotions such as Anger, Astonish, Disgust, Energy, Fear, Mirth, Pleasure, Serenity and

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Sorrow respectively, in which the sentences are classified.

PC1, PC2, …….., PC9 – Represents the predicted class of emotions such as Anger, Astonish, Disgust, Energy, Fear, Mirth, Pleasure, Serenity and Sorrow respectively, in which the sentences are classified.

Sample Input Cinderella thanked him, went to her mother's grave, and planted the branch on it, and she wept so much that her tears fell upon it and watered it. It grew and became a beautiful tree.

Cinderella went to this tree three times every day, and beneath it she wept and prayed. A white bird came to the tree every time, and whenever she expressed a wish, the bird would throw down to her what she had wished for.

Now it happened that the king proclaimed a festival that was to last three days. All the beautiful young girls in the land were invited, so that his son could select a bride for himself.

When the two stepsisters heard that they too had been invited, they were in high spirits. They called Cinderella, saying, "Comb our hair for us. Brush our shoes and fasten our buckles. We are going to the festival at the king's castle."

Cinderella obeyed, but wept, because she too would have liked to go to the dance with them.

She begged her stepmother to allow her to go.

Table 1: Output given by existing system

Sr.No	Sentence	Polarity	No. of words	Emotion
1	Cinderella thanked him went mother grave and planted the branch on it and she wept so much that tears fell upon it and watered it	-0.301025390625	25	NATURAL
2	It grew and became a beautiful tree	0	7	PLEASURE
3	Cinderella went this tree three times every day and beneath it she wept and prayed	-0.243896484375	15	NATURAL
4	A white bird came the tree every time and she expressed a wish the bird throw her she had wished for	0.0234603881835	21	NATURAL
5	Now it happened that the king proclaimed a festival was last three days	-0.034912109375	13	NATURAL
6	the beautiful young girls in the land were invited so that son select a bride for himself	0.15625	17	PLEASURE
7	the two stepsisters heard that they too had been invited they were in high spirits	0.4396352767944	15	NATURAL
8	They called Cinderella saying Comb hair for us	-0.281164648941	8	NATURAL
9	Brush shoes and fasten buckles	-3.814697265625E-05	5	NATURAL
10	We are going the festival at the king castle	0.004150390625	9	NATURAL
11	Cinderella obeyed but wept because she too have liked go the dance with them	-0.25	14	PLEASURE
12	She begged stepmother allow her go	0.0625	6	NATURAL

Table 2: Output given by our system

Sr.No	Sentence	Polarity	No. of words	Emotion
1	Cinderella thanked him went mother grave and planted the branch on it and she wept so much that tears fell upon it and watered it	-0.301025390625	25	SORROW
2	It grew and became a beautiful tree	0	7	ASTONISH
3	Cinderella went this tree three times every day and beneath it she wept and prayed	-0.243896484375	15	SORROW
4	A white bird came the tree every time and she expressed a wish the bird throw her she had wished for	0.0234603881835	21	PLEASURE
5	Now it happened that the king proclaimed a festival was last three days	-0.034912109375	13	ENERGY
6	the beautiful young girls in the land were invited so that son select a bride for himself	0.15625	17	MIRTH
7	the two stepsisters heard that they too had been invited they were in high spirits	0.4396352767944	15	MIRTH
8	They called Cinderella saying Comb hair for us	-0.281164648941	8	ENERGY
9	Brush shoes and fasten buckles	-3.814697265625E-05	5	ENERGY
10	We are going the festival at the king castle	0.004150390625	9	PLEASURE
11	Cinderella obeyed but wept because she too have liked go the dance with them	-0.25	14	SORROW
12	She begged stepmother allow her go	0.0625	6	SORROW

A. Accuracy

Accuracy is defined as the ratio of number of sentences for emotion labels are extracted correctly to the total number of sentences in the document.

$$\text{Accuracy}=\frac{\text{number of sentences for which emotions labels are extracted correctly}}{\text{total number of sentences for which emotions labels are extracted by algorithm}} * 100$$

Table 3: Accuracy Comparison

No. of input document (DS)	No. sentences in input document	Accuracy (%)	
		Rule-base + SVM	Rule-based + KNN + SVM
DS1.txt	5	70	70
DS2.txt	10	60	70
DS3.txt	12	65	73.33
DS4.txt	18	62.22	67.77
DS5.txt	20	70	75
DS6.txt	23	63.91	72.60
DS7.txt	27	67.77	75.18
DS8.txt	30	76.66	83.33
DS9.txt	32	74.84	77.87
DS10.txt	35	78.57	84.28
DS11.txt	38	71.57	74.21
DS12.txt	40	65	70
DS13.txt	44	76.36	80.90
DS14.txt	48	65	79.58

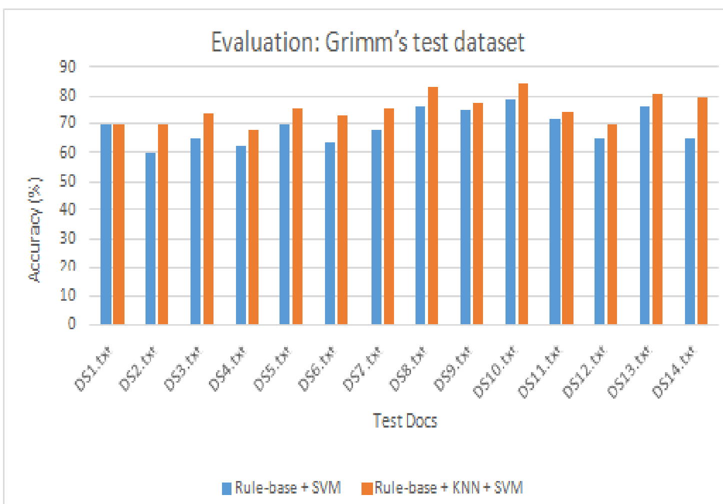


Fig 2. Accuracy comparison graph

B. Precision and Recall

As the system deals with multi class classification problem, the only way to evaluate the performance of the system is to compute the precision and recall for each class label and analyze the individual performance on class labels or average on class labels or average the values to get the overall precision and recall. For evaluating precision and recall for the system. 10 sample sentences are considered for each class of emotion

1) *Precision:* Precision is defined as the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved.

$$\text{Precision} = \frac{\text{True positive}}{\textit{True Psitive+False Positive}} * 100$$

2) *Recall:* Recall is defined as the ratio of the number of relevant records retrieved to the total number of relevant records in the database. It is usually expressed as a percentage.

$$\text{Recall} = \frac{\text{True positive}}{\textit{True Psitive+False Negative}} * 100$$

Precision recall table for existing system

	C1	C2	C3	C4	C5	C6	C7	C8	C9
PC1	**6**	0	0	0	1	0	0	0	0
PC2	**0**	3	0	0	0	1	0	0	0
PC3	**0**	0	7	0	0	0	0	0	1
PC4	**0**	0	0	6	0	0	0	0	0
PC5	0	0	0	0	5	0	0	0	0
PC6	0	1	0	0	0	5	0	0	0
PC7	0	1	0	1	0	2	7	0	0
PC8	0	0	1	0	0	0	0	6	0
PC9	2	1	0	0	2	0	2	1	8

Precision recall table for our system:

	C1	C2	C3	C4	C5	C6	C7	C8	C9
PC1	**7**	0	0	0	1	0	0	0	0
PC2	**0**	4	0	0	0	1	0	0	0
PC3	**0**	0	7	0	0	0	0	0	1
PC4	**0**	0	0	7	0	0	0	0	0
PC5	0	0	0	0	5	0	0	0	0
PC6	0	1	0	0	0	6	0	0	0
PC7	0	1	0	1	0	3	8	0	0
PC8	0	0	0	0	0	0	0	7	0
PC9	2	2	0	0	3	0	2	1	9

Class Labels	Precision by existing system	Precision by our system	Recall by existing system	Recall by our system
C1	85.71	87.5	75	77.77
C2	75	80	50	50
C3	87.5	87.5	87.5	100
C4	100	100	85.71	87.5
C5	85.71	100	62.5	55.55
C6	83.33	85.71	62.5	60
C7	61.53	62.5	77.77	80
C8	87.5	85.71	85.71	87.5
C9	50	45	80	90

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

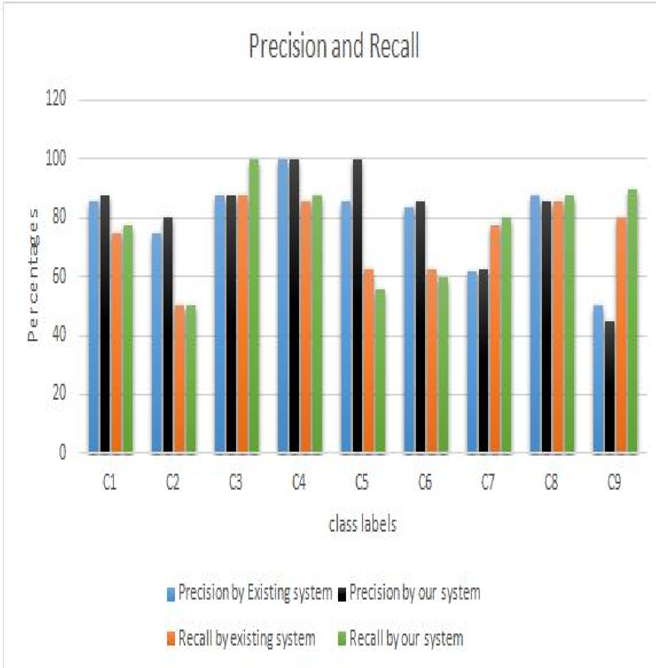


Fig. 3: Precision and Recall

V. CONCLUSION

The system has performed extremely well in terms of accuracy while showing the signs that it can be further scaled to much bigger and different dataset with better performance. In this research, the main focus was on analyzing emotions at sentence level which can further explored to paragraph level and topic level. The work can also be expanded in future by introducing methods that increase the accuracy by handling problems like change of emotions when the personal pronoun changes which still needs to be evolved properly.

REFERENCES

[1] Wingyan Chung, Saike He, Daniel Dajun Zeng, Victor Benjamin, “Emotion extraction and entrainment in social media: The case of U.S. immigration and border security”, Proceedings of 2015 IEEE International Conference on Intelligence and Security Informatics (ISI), pp. 55-60, 2015.
[2] Akshi Kumar, Prakhar Dogra, Vikrant Dabas, “Emotion analysis of Twitter using opinion mining”, Proceedings of the Eighth IEEE International Conference on Contemporary Computing (IC3), pp.285-290, Noida, 2015.
[3] Lily Dey; Mahim-Ul Asad; Nadia Afroz; Rudra Pratap Deb Nath, “Emotion extraction from real time chat messenger”, Proceedings of 2014 International Conference on Informatics, Electronics & Vision (ICIEV),pp. 1-5, 2014.
[4] Cheng-Yu Lu, William W. Y. Hsu, Hsing-Tsung Peng, Jen-Ming Chung, Jan-Ming Ho, “Emotion Sensing for Internet Chatting: A Web Mining Approach for Affective Categorization of Events”, Proceedings of 2010 13th IEEE International Conference on Computational Science and Engineering, pp. 295 – 301, 2010.
[5] Mansur Alp Tocoglu; Adil Alpkocak, “Emotion Extraction from Turkish Text”, Proceedings of 2014 European Network Intelligence Conference, pp. 130 – 133, 2014.
[6] Luwen Huangfu; Wenji Mao; Daniel Zeng; Lei Wang, “OCC model-based emotion extraction from online reviews”, Proceedings of 2013 IEEE International Conference on Intelligence and Security Informatics, pp. 116-121, 2013
[7] Azadeh Nikfarjam, Ehsan Emadzadeh, “A Hybrid System for Emotion Extraction from Suicide Notes”, Biomedical Informatics Insight, pp. 165-174, 2012.
[8] Nilesh M. Shelke and Shriniwas Deshpande and Vilas Thakre, “Approaches towards Emotion Extraction from TEXT”, IJCA Proceedings on National Conference on Innovative Paradigms in Engineering & Technology , pp. 10-14, 2013.
[9] J. S. Raikwal Kanak Saxena, “Performance Evaluation of SVM and K-Nearest Neighbor Algorithm over Medical Data set”, International Journal of Computer Applications (0975 – 8887), vol. 50 , no.14, July 2012.
[10] Taysir Hassan A. Soliman; Mostafa A. Elmasry; Abdel Rahman Hedar; M. M. Doss, “Utilizing support vector machines in mining online customer reviews”, Proceedings of 22nd International Conference on Computer Theory and Applications (ICCTA), pp. 192-197, 2012.
[11] Xiaojun li, Lin Dai, Hanxiao Shi, “Opinion mining of camera reviews basedon Semantic Role Labeling”, Proceedings of the Seventh IEEE International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), pp. 2372-2375, 2015.
[12] Yun Li, Jie Wang, “Research on Text Classification Based on SVM-KNN”, Proceedings of 5th IEEE International Conference on Software Engineering and Service Science (ICSESS), pp. 7-29, 2014.
[13] Jibran Mir, Muhammad Usman, “An effective model for aspect based opinion mining for social reviews”, Proceedings of the IEEE Tenth International

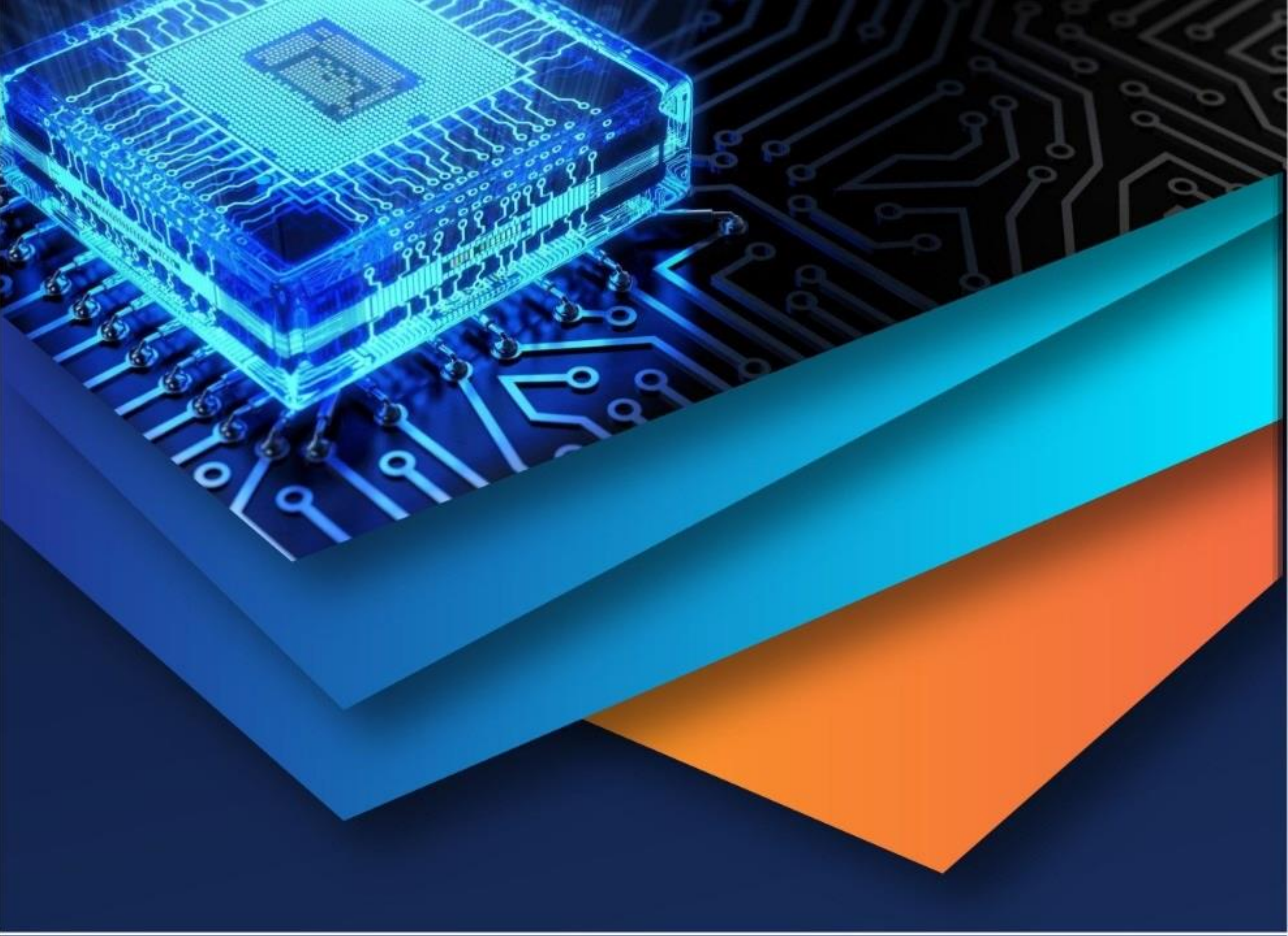
Conference on Digital Information Management (ICDIM), pp.49-56, 2015.

[14] Vijay B. Raut, D.D. Londhe, “Opinion Mining and Summarization of Hotel Reviews”, Proceedings of the IEEE International Conference on Computational Intelligence and Communication Networks (CICN), pp.556-559,Bhopal, 2014.

[15] E. Aramaki, T. Imai, K. Miyo, and K. Ohe, “Patient status classification by using rule based sentence extraction and bm25-knn based classifier,” In Workshop on Challenges in Natural Language Processing for Clinical Data, 2006.

[16] Tomomi Nomura; Yasue Mitsukura, “Extraction of unconscious emotions while watching TV commercials”, Proceedings of 41st Annual Conference of the IEEE Industrial Electronics Society, pp. 368-373, 2015.

[17] Money Goyal; Mooninder Singh; Mandeep Singh, “Classification of emotions based on ERP feature extraction”, Proceedings of 1st International Conference on Next Generation Computing Technologies (NGCT), pp- 660-662, 2015.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429