TexEmo: Conveying Emotion from Text- The Study

Mukesh C. Jain
ME (Scholar) Computer Engineering
MIT, Pune

V. Y. Kulkarni
Department of Computer Engineering
MIT, Pune

ABSTRACT

Human Computer Interface for communication is a very powerful and most current area of research because the human world is getting more and more digitize and one wants the digital systems to behave like a human being. This requires the digital systems to imitate the human behavior accurately. Emotion is one aspect of human behavior which plays an important role in human feeling and decision making thus influencing the way people interact in the society. In human computer interaction, the computer interfaces need to recognize the emotion of the end users in order to exhibit a truly intelligent behavior. Human express the emotion in the form their facial expression, through their speech, and by writing text.

The automatic identification of emotions in texts is important for applications such as: opinion mining and market analysis, affective computing, natural language interfaces, and e-learning environments, including educational games.

This paper is mainly focused on to conveying the emotion expressed by a text documents. In simple way I can say that main aim of this study is to classify the emotion expressed by the text, based on pre-defined list of emotion i.e. Anger, Joy, Sad, Fear, Disgust, and Surprise. In order to collect the emotion evoking word, this study paper is mainly enlightened on ISEAR dataset, WPARD and Word-Net Affect dataset. A lot of attention is paid to the normalization of text and expand the knowledge base of emotion word. Uses of Vector Space Model for Information retrieval and classification.

General Terms

Convey the Emotion from Text.

Keywords

Information Retrieval, Vector Space Classification.

1. INTRODUCTION

Emotions have mesmerized researchers for long. Emotions identification is currently widely being studied in neuroscience, psychology and cognitive sciences, computer sciences and behavior sciences, as they are an important element of human nature. Most of the attention of researchers in computer science is paid to in the field of human computer interaction, especially in the field of emotion recognition because it is main field in human computer interaction. Human emotion demonstrates itself in the form of facial expressions, speech utterances, writings, and in gestures and actions. Consequently, scientific research in emotion has been pursued along several dimensions and has drawn upon research from various fields. This paper addresses the task of emotion recognition by attempting to automatically learn emotions from text.

Text communicates informative contents as well attitudinal information which includes emotional states. The most well known form of communication on web is in the form of text, offering a platform for computer systems to behave more intelligently based on the user's mood. Large amounts of textual data available in the form of blogs, social networking sites, emails, etc. This textual data is useful to generate better human centered communication system which needs to be able to analyze the text and infer the sentiment/emotion of the user. Although communication systems can identify the user's emotional states from different communication modalities, the variety and complexity of language makes it difficult for researchers to recognize emotional states from pure textual data.

In order to capture emotion from text document we require the classification which aims at inferring the emotion conveyed by the documents based on predefined lists of emotion. Predefined list of emotion contains six basic emotion named as Joy, Anger, Fear, Disgust, Sad and Surprise.

This topic is mainly focused on two main tasks.
1) The test data i.e. text document collected from any news articles etc. required for understanding the emotions evoked by words and phrases is huge. This is because a different word evokes different emotions learnt from our day to day experiences. For this purpose, need is to enhanced dictionary with emotion word from WordNet Affect, WPARD, ISEAR to improve in result.
2) Since the scope of words is larger in this scenario, the usage of words and their inflected form is large too and negations are major emotion modifiers. So these problems need to be solved properly. For this purpose, need is to text normalization mainly to handle negation.

In section 2 we discuss application area. Section 3 gives overview of previous work done on this area. Section 4 deals with Existing System of Emotion Detection and Section 5 explain different classic Information Retrieval Models. In Section 6 give details about system description and this is divided into two sub sections like
- Data Preparation- Dataset Selection, Data Processing, Data collection from External Knowledge Source
- Feature Extraction- Bag of Words approach, BOW limitation
- Classification Schema- Vector Space Model- Term Weighting, Cosine Normalization and Similarity Measure

Last Section 7, 8 contain conclusion and References.

2. APPLICATION AREAS

The application areas of textual emotion detection are manifold:

2.1 Sentiment Analysis / Opinion Mining:

Sentiment Analysis /Opinion Mining are the new and emerging area of research. It mainly focuses on information retrieval & knowledge discovery from text. The goal of Sentiment Analysis is to make computer able to detect & express emotions. Thus Sentiment Analysis is also referred as Opinion Mining. Business Firms & Individuals are interested in to find consumers opinion about products, services, issues & event for finding the best choices.

Opinion can be collected from web forums, blogs, discussion groups & comment boxes. Opinion Mining is one of an important application of web data. It is used to collect user opinion & extract meaningful patterns from it. It can be

positive or negative. During decision making process we can take good decision on the basis of opinion of others. This shows the growing importance & need of Opinion Mining.

2.2 Text – to – Speech Generation:

This research addresses the reader's emotions provoked by the text. The goal of this is to classify the emotional resemblance of sentences in the storyline text, for appropriate expressive representation of Text- to-Speech synthesis.

Storyline text is especially flat even though it having emotional content. In verbal communication, spokespersons effectively express emotions by modifying manner of speech or communication, including pitch, intensity & durational signals in the speech signal. Thus, in order to make Text - to – Speech combination, so sound must be as usual & appealing as possible, it is important to express the emotional posture in the text. Use of this system implies the appropriate emotional meaning of the corresponding text passage.

An application of this system is to solve two basic problems. First, describe the emotions appropriately in certain text passage and Second, from given a text passage and specified emotional mark up how to provide the prosodic form in order to convey the emotional content.

The Classification of documents into Text - to – Speech production has novel applications. One of them is to integrate Text - to – Speech emotion classification into a web search engine to allow users to retrieve documents that contain relevant contents and at the same time produce proper emotions. Another is for websites to organize contents according to Text - to – Speech sentiment categories and offer users a convenient browse. This makes Text - to – Speech production a very productive area of research.

2.3 Better Computer Interaction System:

Various kinds of the Human centered communication systems, such as dialogue systems, automatic answering systems and human like robots etc. It can apply emotion recognition techniques so that a user can feel that the system is more like human. A better response system, based on the user's current mood / emotion, makes users and computer work in sync.

Finally, we summarize this (the application area) section as achievement in this domain can be most helpful in next generation intelligent robotics, artificial intelligence, psychology blogs, product reviews, CRM and service oriented companies like Right Now Technology and NICE system produces customer service software Smart Sence[TM] and Smart Perform[TM] respectively which recognizes customer emotion using keyword spotting techniques.

3. LITERATURE SURVEY

3.1 Literature Study on Emotion Theories:

The basic hypothesis presented in this paper is to categorize emotions into primary emotions and secondary emotions.
More recently, researchers have investigated several aspects of human emotion in order to arrive at a set of emotion categories that are universally acceptable. Several works in this direction have been reported in this literature survey. Table 1 lists the basic emotion categories identified by the different researchers.

Table 1: Basic Emotion Categories Identified by Researchers

Tomkins (1962)	Izard (1977)	Plutchik (1980)	Ortony (1988)	Ekman (1992)
Joy Anguish Fear Anger Disgust Surprise Interest Shame	Enjoyment Sadness Fear Anger Disgust Surprise Interest Shame Shyness Guilt	Joy Sorrow Fear Anger Disgust Surprise Acceptance Anticipation	Joy Sadness Fear Anger Disgust Surprise	Happiness Sadness Fear Anger Disgust Surprise

Some psychologists have investigated facial expressions of emotion to identify the basic distinguishable expressions among them, and mapped them to basic human emotions. Ekman has defined basic emotions as those that have universally accepted distinctive facial expressions [7].

3.2 Literature Survey on Emotion Recognition:

A lot of research has been done in the area of emotion classification from text. The majority of work being lexicon-based supervised learning approaches.

Shritiranjan Satapathy et al. [2] works on identifying the emotion conveyed by the sentence at different granularities like one is coarse grained classification means classify the sentence into only positive and negative emotions based on their polarity , other is fine grained classification means classify the emotion conveyed by sentence based on predefined emotion set. They are also proposes a bag of word approach with lower order dependencies captured through the use of bigrams and trigrams. They work on ISEAR dataset and WordNet Affect. They are mainly focused on normalization of text as the inflected forms of a word occur in plenty. For classification they have used Multinomial Naïve Bayes and compared with VSM classifiers with MNB giving most promising result.

Taner Danisman et al. [1] Proposes a bag of word approach to emotion classification from text. They work on ISEAR dataset i.e. International Survey on Emotion Antecedents and Reaction data set and SemEval dataset. For the classification they have used Vector Space Model with a total of 801 news headlines provided by "Affective Task" in SemEval 2007 workshop and which is focusing on classification of emotions and valence in text. They have compared their result with different classifiers including Concept Net, Naïve Bayes, and SVM and conclude that VSM classification gives better performance those other classifiers.

Carlo Strapparava et al [5] provide unsupervised learning experimenting with the automatic analysis of emotions in text, They proposes and evaluate several knowledge based and corpus based methods for the automatic identification of emotions in text. They developed dataset for emotion analysis using news headlines that are drawn from major newspaper such as New York Times, CNN, and BBC news as well as from Google News search engine for the experiment. Two data set with 250 annotated headlines and other is the test data set with 1000 annotated headlines. This task was carried out in an unsupervised setting. They have implemented five different systems for emotions analysis using the knowledge based and corpus based approaches. They compared their overall average results obtained by the five proposed system

with three SemEval System. Through comparative evaluation of several knowledge based and corpus based methods they tried to identify the methods that work best for the annotation of emotions.

Shoushan LI and Chengging Zong et al. [6] combine multiple SVM classifiers on sentiment classification with several feature sets. Their experiments did sentiment classification based at assigning a document to a predefined category according to the polarity of its subjective information. They present the classifier combination approaches to perform this task. First, different classifiers are generated through training the review data with different features: unigrams and some POS features. Then, classifier selection method is used to select a part of the classifier for the next step combination. Finally, these selected classifiers are combined using several combining rules. Their experimental results proved that all the combinational approaches with different combining rules outperform individual classifier and the sum rule achieves the best performance with improved result over the best individual classifier.

Sugimoto et al. [8] considered the emotional expressions for text-to-speech engines and emotional reading. They partitioned the text into nouns, adjectives and adverbs and used the frequency of words to determine the emotional class.

Cecila Ovesdotter Alm et al. [4] provides experiments and empirical study on text based emotion prediction problem, using supervised machine learning with SNoW (Sparse Network of Windows) learning architecture. SNoW is a multiclass classifier that is specifically tailored for large scale learning tasks. They focused on the basic task of recognizing emotional passages and on determine their valence. They predict finer emotional meaning distinctions according to emotional categories in the speech. They work on data set of 22 fairy tales that show encouraging results over a naïve baseline and BOW approach for classification of emotional versus non-emotional contents, with some dependency on parameter tuning.

Liu et al. [9] used large-scale real-world knowledge about the inherent affective nature of everyday situations to classify sentences into "basic" emotion categories. This commonsense approach has new robustness implications. Open Mind Commonsense was used as a real world corpus of 400,000 facts about the everyday world. Four linguistic models are combined for robustness as a society of commonsense-based affect recognition. These models cooperate and compete to classify the affect of text. Such a system that analyzed the affective qualities sentence by sentence is of practical value when people want to evaluate the text they are writing. As such, the system is tested in an email writing application. The results suggest that the approach is robust enough to enable plausible affective text user interfaces.

4. EXISTING EMOTION DETECTION SYSTEM

This section gives brief idea about existing feature detection system. Many studies use it for emotion detection and problems related to this system.

4.1 Keyword-based Methods:

Keyword-based methods are the most perceptive ways to detect textual features or emotion. Emotions are detected based on the related sets of keywords found in the input text;

Keyword-based emotion detection methods have three limitations described below [10].

4.1.1 Ambiguity in Keyword

Use of emotion keywords is a straightforward way to detect associated emotions. But it has one big problem that the meanings of keywords could be multiple and unclear. Most words could change their meanings according to different usages and contexts, and it is just not feasible to include all possible combinations into the dataset.

Fig 1: Ambiguous word in ISEAR dataset after finding set difference

4.1.2 Failure of Recognizing Sentences without Keywords

This approach is totally based on the set of emotion keywords. Therefore, sentences without any keywords would denote that they do not contain any emotions at all, which is obviously wrong.

For example, "I helped old lady to cross the river" and "Hooray! I helped old lady to cross the river" Both statement indicate the same emotion i.e. Joy, but the previous statement without "hooray" possibly will remain undetected if "hooray" is the only keyword to detect this emotion.

4.1.3 Deficiency of Linguistic Information

Ignoring linguistic information also poses a problem to keyword-based methods. Syntax structures and semantics also have manipulates on articulated emotions.

For example, "I laughed at him" and "He laughed at me" would suggest different emotions from the first person's perspective.

4.2 Learning-based Methods:

Emotions are detected based on previous training result with respect to specific statistic learning methods to determine which emotion category should the input text belongs.

Still this method also suffer from some problems, those are listed below.

4.2.1 Determining Emotion Indicators

The first problem is, this method still needs keywords, but just in the form of features because this method automatically determine the probabilities between features and emotions, the most sensitive features may be emoticons, which can be seen as emotion label in the texts.

4.2.2 Simplified Emotion Categories

Most of the learning-based methods can only classify sentences into two categories because of lacking in efficient

features other than emotion keywords, which are positive and negative.

4.3 Hybrid Methods:

This Features detection is based on the combination of keyword based and learning based approach, and other supplementary information;

The main advantages of this approach is that

1) It can give up higher accuracy results from training and adding knowledge-rich linguistic information from dictionaries and thesauri.

It will balance the high cost involved for information retrieval tasks and minimize difficulties encountered while adding different lexical resources [10].

5. INFORMATION RETRIEVAL MODEL

This section presents information retrieval models that can be applied on any text collection. Not all the IR models are easily scaled up to be able to deal with a very large collection, such as pages collected from the Web. The most important IR models are: the Boolean Model, the Statistical model which includes the Vector Space Model, and the Probabilistic Model, and the Linguistic and Knowledge-based models. Various extensions of these models are possible.

5.1 The Boolean Model:

The Boolean model is often referred as "exact match" model. Its main goal is capture conceptual structure and contextual information. A document is represented as a set of keywords. Queries are Boolean expressions of keywords, connected by AND, OR, and NOT, including the use of brackets to indicate the scope of these operators.

For example, the query "all the hotels in Everest India or Bhutan Nepal, but not China" is typed by the user as:

[[Everest & India] | [Bhutan & Nepal]] & hotel & !China]

The output of the system is a list of documents that are relevant, but there will be no partial matches or ranking. The Boolean model is very strict: AND means "all"; OR means "any". All matched documents will be returned, making it difficult to control the number of documents retrieved.

It has the following strengths:

1) It is simple to perform and it is computationally efficient. Hence, it is the normally used for the current large-scale, operational retrieval systems and many of the major on-line information services use it.
2) It allows users to express structural and notional constraints to describe important linguistic features. Users able to find that synonym specifications reflected by OR-clauses and phrases represented by nearness relations are useful in the formulation of queries.
3) The Boolean approach possesses a great clarity and expressive power. Boolean retrieval is very effective if a query requires a complete and explicit selection.
4) It offers a numbers of techniques to enlarge or narrow a query.
5) The Boolean approach can be especially effective in the later stages of the search process, because of the clarity and exactness with which relationships between concepts can be represented.

The standard Boolean approach has the following shortcomings:

1) Users find it difficult to construct effective Boolean queries.
2) Only documents that satisfy a query exactly are retrieved.
3) It is hard to control the number of retrieved documents.
4) The traditional Boolean approach does not provide a relevance ranking of the retrieved documents. It does not represent the degree of uncertainty or error due the vocabulary problem.

5.2 The Statistical Model:

The statistical retrieval approach contains two major models one is vector space model and another is probabilistic models. These models use the statistical information to determine the relevance of documents with respect to a query in the form of term frequencies. Even though they are different from each other because of the way they use the term frequencies, both produce as their output a list of documents ranked by their expected significance [11].

Table 3: Comparisons of Vector Space Model and Probabilistic Model

Model	Statistical Model	
	Vector Space Model	**Probabilistic Model**
Motivation	It make simpler query formation skill to control output	It deals with uncertainty in query representations
Goal	It grade the output based on Similarity	It grade the output based on Probability and Relevance
Methods	Cosine Measure	Use of different models
Source	**Vector Space Model** $$\text{Similarity(Q,D)} = \frac{\sum(w_{iq} * w_{id})}{\sqrt{\sum w_{iq}^2 * \sum w_{id}^2}}$$ Where $w_{iq} = tf_{iq}\cdot idf_i$ $w_{id} = tf_{id}\cdot idf_i$ $tf_i(q \text{ or } d)$= Term frequency $tf(t,(d \text{ or } q))$: frequency of a term t in document d or query q. idf_i = Inverse document frequency $(idf(t))$: measures the rarity of a term t in the documents. **Probabilistic Model** Term weight = $\log [\{r_t/R - r_t\}/\{\{n_r - r_t\}/\{\{N - n_t\} - \{R - r_t\}\}\}]$ $= "\{hits/misses\}/\{false\ alarms/correct\ misses\}"$ $\text{Similarity}_{jk} = \sum(C + idf(i) * tf(i,j))$ Where, tf(i,j)= K+ (i-K) (freq(i,j) / maxfreq(j))	
Issues	• How to express NOT? Proximity searches? • Limited expressive power • Computationally intensive • Assumes that terms are independent	• Estimation of needed probabilities • Prior knowledge needed • Independence assumption • Boolean relations lost • Which model is best?
	• Lack of structure to represent important linguistic features • How to better visualize the retrieved set?	

5.3 Linguistic and Knowledge-based Model:

In automatic text retrieval, users enter a string of keywords. This approach retrieves documents based solely on the presence or absence of exact single word strings as specified by the logical representation of the query. Clearly this approach will neglect many relevant documents because it does not capture the complete or deep meaning of the user's query.

Linguistic and knowledge-based approaches have also been developed to address the problem occurred in above model by performing a morphological, syntactic and semantic analysis to retrieve documents more effectively. In a morphological analysis, roots and attaches are analyzed to determine the part of speech (noun, verb, adjective etc.) of the words. Next complete phrases have to be parsed using some form of syntactic analysis. Finally, the linguistic methods have to resolve word ambiguities and/or generate relevant synonyms or quasi-synonyms based on the semantic relationships between words. The development of a sophisticated linguistic retrieval system is difficult and it requires complex knowledge bases of semantic information and retrieval heuristics. Hence these systems often require techniques that are commonly referred to as artificial intelligence or expert systems techniques.

6. SYSTEM DISCRIPTION

6.1 System Architecture of Conveying Emotion from Text using VSM

This section is divided into four main modules; listed as below:1) Data Preparation 2) Feature Extraction 3) VSM Classification 4) Negation Handling. Following figure and table gives brief overview and implementation details of these modules and its theory give brief idea about the system.
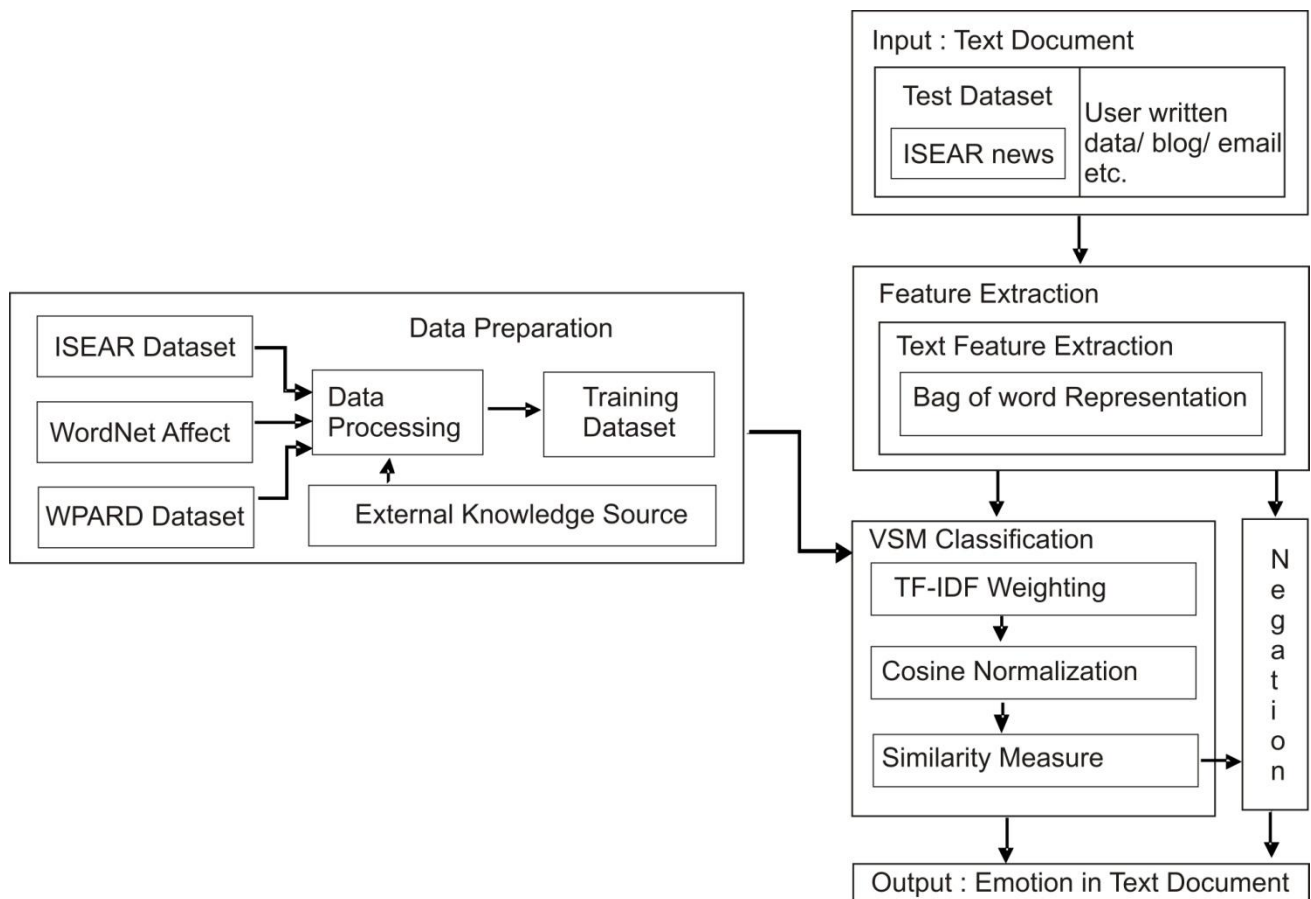

Fig 2: System Architecture of Emotion Identification from Text using VSM

Table 4: Implementation Details of Emotion Identification From Text using VSM

	User Query	Emotion Classes
Feature Extraction	Set of Stemmed Keywords i.e. Terms	
Feature and Term Matrix Generation	$\begin{matrix} & q \\ T_1 & w_{11} \\ T_2 & w_{12} \\ T_3 & w_{13} \\ \vdots & \vdots \\ \vdots & \vdots \\ T_N & w_{1N} \end{matrix}$	$\begin{matrix} & T_1 & T_2 \ldots \ldots & T_N \\ d_1 & w_{11} & w_{21}\ldots .. & w_{t1} \\ d_2 & w_{12} & w_{22}\ldots .. & w_{t2} \\ d_3 & w_{13} & w_{23}\ldots .. & w_{t3} \\ \vdots & \vdots & \vdots \; \vdots \;\; \ldots & \vdots \\ d_N & w_{1N} & w_{2N}\ldots .. & w_{tN} \end{matrix}$
TF-IDF Weighting	$wiq = tfiq \cdot idfi = tfiq \, log\,(N/dfi).$	$wij = tfij \cdot idfi = tfij \, log\,(N/dfi)$
Cosine Similarity Measures	$cosSim(\overrightarrow{d_J} \cdot \vec{q}) = \dfrac{\overrightarrow{d_J} \cdot \vec{q}}{|\overrightarrow{d_J}| \cdot |\vec{q}|} = \dfrac{\sum_{i=1}^{t}(w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^{t} w_{ij}^2 \cdot \sum_{i=1}^{t} w_{iq}^2}}$	
Handling Negation	Reverse the Emotion	
Output Generation	The classification Result is VSM(q)=$avgmax(Sim(\overrightarrow{d_J} \cdot \vec{q}))$	

6.1.1 Data Preparation

The first and most difficult task is data preparation because a problem arises from selection of dataset for this work, in order to cover most of the word. This study prepared the dataset for basic six emotions like anger, disgust, fear, joy, sad and surprise. To prepare dataset it takes help from ISEAR, WordNet-Affect and WPARD and synonyms of all collected emotional word and also to expand dataset many needs to take help from web. In dataset preparation after study we conclude that we can use sentences from ISEAR dataset, emotional words from WordNet-Affect and polarity of words from WPARD datasets. In data preparation stage we can create training dataset and select the testing dataset.

6.1.2 Feature Extraction

It is the process consists in transforming arbitrary data, such as text or images, into numerical features usable for machine learning. At the end it is a machine learning technique applied on these features. Here we are mainly focused on text feature extraction using Bag of word approach. In this approach, a document is considered to be an unordered collection of words whereas the position of words in the document bears no importance. "'Bag of words'" is the simplest representation of textual data.

6.1.3 VSM Classification-

Classification is the task of choosing the correct class label for a given input. The vector space model is a very successful statistical method of information retrieval. In this model the documents and queries are vectors in a high-dimensional space. We can model the documents by using single words (stemmed etc.) and stop words removal components. For a given query this study follows following sequence:

- First is to convert the query to a vector of terms and Weight each component.
- Second is consult the index to find all documents containing each term.
- Third is converting each document to a weighted vector.
- Fourth is match the query vector against each document vector and sort the documents by their similarity.
- Finally present the *n* most similar documents to the user.

The main advantages of the vector space model are:

1) It is simple and based on clear mathematical theory.
2) It considers both local (*tf*) and global (*idf*) word occurrence frequencies.

3) It provides partial matching and ranked results.
4) It tends to work quite well in practice and allows efficient implementation for large document collections.

The main weaknesses are:

1) It does not account for semantic information and syntactic information.
2) It lacks the control of a Boolean model.

6.1.4 Negation Handling:

Varieties of techniques have been proposed to handle negation. If a sentence is given using 'not', previous approaches focused on to append 'not' to every word following it until the next punctuation appears. This results in the creation of many useless words, which may hinder feature selection as the original word is lost.

This study uses a more careful approach, where every-word of the sentence is first simplified with stop-words removal and finding out keyword that are used to detect the emotion. If 'not' is present in the sentence it means it can be used for negation handling and 'not' is commonly used in conjunction with emotion indicating word. This study first find out emotion from the sentence without considering 'not' and after that it generate the reverse result of the emotion.

7. CONCLUSION

This paper addresses an important and less examined area of sentiment research, that is, emotion identification from text. The major contribution of this work is to show that it is practical to apply computational methods to identify and distinguish various types of emotions in text.

The goal of the emotion identification experiment is to manually add emotion information to each sentence in a dataset of blogs collected from the web or standard dataset. This manually annotated data can be used to train computer based systems to automatically identify emotion information on a large-scale.

This paper is mainly focused on automatic emotion detection and also suggested that to achieve good performance, it is important to include a wide variety of words that go away from the conventional emotion words. This study also introduced a novel approach of automatically building emotion lexicon utilizing the classification system. In the approach introduced here, a variety of emotion-related words were learned, and their usability demonstrated by their effectiveness in VSM classification methods for emotion detection.

In Future, This system should also detect not only the existence of keywords, but also their linguistic information to detect emotions more accurately

8. REFERENCES

[1] Taner Danisman and Adil Alpkocak, 2008. Feeler: Emotion Classification of Text Using Vector Space Model. AISB 2008 Convention, Scotland.

[2] Shrutiranjan Satapathy and Sumit Bhagwani, 2012. Capturing Emotion in Sentences. cse.iitk.ac.in, Kanpur

[3] Khairullah Kan, Baharum Baharudin, Aurangzeb Khan and Fazal-e-Malik, 2009. Mining Opinion from Text Documets: A Survey. 3rd IEEE International Conference on Digital Ecosystems and Technologies.

[4] Cecilia Ovesdotter Alm, Dan Roth and Richard Sproat. 2005. Emotions from text: Machine Learning for Text-based emotion prediction. HLT '05.

[5] Carlo Strapparava and Rada Mihalcea. 2008. Learning to Identify Emotions in Text. Proceedings of the 2008 ACM SAC'08

[6] Shoushan LI, Chengqing Zong and Xia WANG, 2007. Sentiment Classification through Combining Classifiers with Multiple Feature Sets. IEEE International Conference.

[7] Saima Aman and Stan Szpakowicz. 2007. Identifying Expressions of Emotion in Text. TSD'07 Proceedings of the 10th international conference on Text, Speech and Dialogue.

[8] Futoshi Sugimoto and Masahide Yoneyama. 2006. A method for classifying emotion of Text bsed on Emotion-Dictionaries for Emotional Reading. AIA'06 Proceedings of the 24th IASTED international conference on Artificial intelligence and applications.

[9] Hugo Liu, Henry Lieberman and Ted Selker. 2003. A model of textual affect sensing using real-world knowledge. Proceedings of the 8th International Conference on Intelligent User Interfaces IUI'03.

[10] Edward Chao-Chun Kao Chun-Chieh Liu, Ting-Hao Yang, Chang-Tai Hsieh, Von-Wun Soo, 2009. Towards Text-based Emotion Detection A Survey and Possible Improvements. IEEE International Conference on Information Management and Engineering ICIME.

[11] Chapter 2 – Information Retrieval Model http://comminfo.rutgers.edu/~aspoerri/InfoCrystal/Ch_2.ps

[12] Bo Pang, Lillian Lee and Shivakumar Vaithyanathan. 2002.Thumbs up? Sentiment Classi_cation using Machine Learning Techniques. EMNLP'02.

[13] Shilpa Arora, Elijah May_eld, Carolyn Penstein-Ros and Eric Nyberg. 2010. Sentiment Classi_cation using Automatically Extracted Subgraph Features. Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text.

[14] Vincent Ng, Sajib Dasgupta and S. M. Niaz Ari_n. 2006. Examining the Role of Linguistic Knowledge Sources in the Automatic Identification and Classification of Reviews. Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions.

[15] C. Strapparava and A. Valitutti, 2004. WordNet-Affect: an affective extension of WordNet,. 4th International Conference on Language Resources and Evaluation (LREC ˝04), pp. 1083–1086, Lisbon, Portugal.

[16] S.K. Shandilya, S.K. Jain, A.K. Nagar, 2011. Opinion Mining and Information Retrieval Techniques for E-Commerce. In Edited Book: Ambient Intelligence and Smart Environments: Trends and Perspectives", edited by Dr. Fulvio Mastrogiovanni & Nak-Young Chong, ISBN: 978-16-1692-857-5, to be published in 2011 by IGI Global (IGI) – Information Science Publishing, USA.

[17] Alexander Pak and Patrick Paroubek. 2011. Text Representation Using Dependency Tree Subgraphs for Sentiment Analysis. DASFAA Workshops'11.

[18] A McCallum and Kamal Nigam. 1998. A Comparison of Event Models for Naive Bayes Text Classication. AAAI-98 Workshop on Learning for Text Categorization.

[19] Harry Zhang. 2004. The Optimality of Naive Bayes. FLAIRS'04.

[20] Andrew Y. Ng and Michael I. Jordan. 2002. On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. NIPS'02

[21] "ISEAR Dataset",http://emotion-research.net/toolbox/toolboxdatabase

[22] "Porter Stemming", http://en.wikipedia.org/wiki/Stemming

[23] Mohammad O.Wedyan Aarti Singh 2013. On the Design and Implementation of an Efficient Information Retrieval System for Arabic Language. International Journal of Electronics and Electrical Engineering Vol. 1, No. 1, March 2013