

SENTIMENT ANALYSIS IN MESSAGING SYSTEMS: A REVIEW

¹Manvi Siwach, ²Jyoti Bhardwaj

¹Assitant Professor, ²Research Scholar(M.tech)

¹Department of Computer Science,

¹J.C.Bose University of Science and Technology, YMCA, Faridabad

Abstract : Sentiment analysis is a process of determining and categorizing the nature of a textual data in order to ascertain whether it expresses a neutral, positive or negative sentiment. A number of applications are based on indentifying the sentiments expressed in texts. It includes performing sentiment analysis on words, sentences, paragraphs, literary texts, tweets etc. Sentiment analysis can be done on products for improving customer relation model in e-commerce. Sentiment analysis in messaging systems has a lot of research work going on. This paper summarizes a brief review on sentiment analysis in messaging systems. A comparison has also been performed among the various sentiment analysis algorithms.

IndexTerms - Emotions; Sentiment analysis; Lexicons; Algorithms

I. INTRODUCTION

In order to perform efficient computing, the HCI(human computer interaction) enables a device to have the capability of detecting and aptly responding to emotions of its users and additional stimuli. An evaluating device having this ability can easily gather indication from a variety of resources to predict users' emotions. Using various means such as speaking, writing and reading, human exposes his/her emotions in every turn of life. Emotions play a very important role in one's life, for example, people become very productive when they are happy than when they are sad or depressed. In our daily life emotions play a very noteworthy role. In everyday's life, with the high usage of internet, the communication between human and machine emphasises on recognition of emotional reactions(or non-verbal information). Written communication when blended along with sentimental expressions leaves a prodigious effect on the brain. As an aftermath, it calls for a need to explore some new features for example emotions and sentiment that may introduce the way for differentiating the collected data in an efficient as well as effective way. This will hence improve the effectiveness of informational retrieval. Numerous methodologies have been established by many researchers, over the last decades, for extracting the emotions conveyed through texts, but surprisingly minimal efforts have been commenced in real time messengers which can analyse or predict emotional state of the user. Here we have incorporated a rule based approach that will bring out the emotional state from the chat box with some emotional categorization such as 'joy', 'interest', 'anger', 'fear', 'surprise', 'shame', 'sad', 'guilt', 'disgust'.

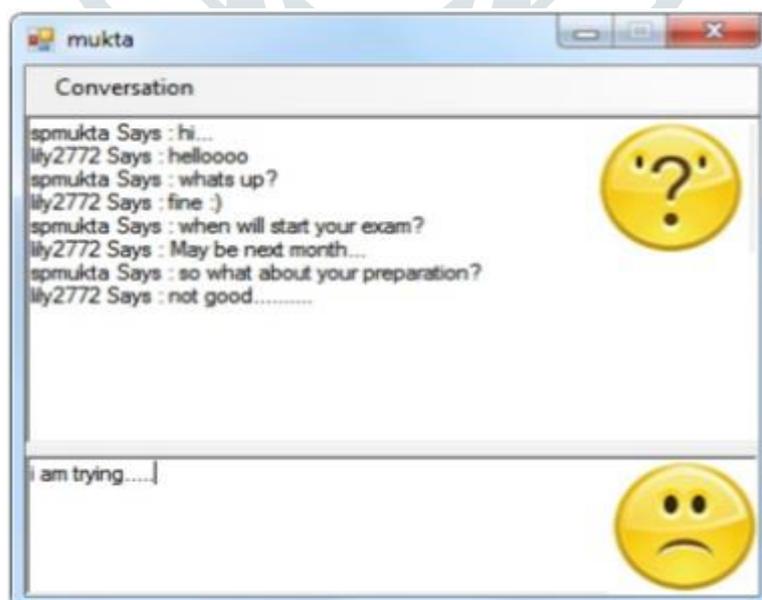


Figure.1: A sample messenger

Here they have constructed a dictionary (or lexicon) containing nearly 4000 meaningful words, which convey emotions and more than 100 abbreviations, emoticons and so forth. They have designed a sample version of chat messenger that can predict emotions

(as shown in figure 1). Usually text reflects the communicator's emotional state, which can also be spotted by the use of various methods. There have been a number of machine learning and statistical classifier techniques developed so far for effective computing, which includes Naïve Bayes, decision tree, Support Vector Machines, vector-space-model and so on. There are various applications of sentiment analysis such as:

- **Opinion Analysis**

Opinion analysis aims at receiving negative as well as positive comments about a particular product as reviews. It helps in examining the customers' opinion about their products.

- **Better HCI systems**

In HCI(Human computer Interaction) systems, various emotion recognition techniques have been employed for recognizing the user's emotions and to make the system have the feel of human.

- **Text to Speech generation**

In order to generate speech form a given context, we need to examine the emotions behind the text. This way leads to detection of emotions from a given text and this eventually becomes the research area for generating speech from text. In huge databases, data mining plays a very important role in analyzing patterns and correlations among various fields. Data mining is used for analyzing the data from various perspectives and then summarized into effective information. The essential theme in data mining is Text Mining. Now-a-days almost all the information related to industry, government, business etc are stored in electronic form in large textual databases. Text databases such as research papers, news, articles, digital libraries, web pages and e-mail messages consists of large amount of the available information. As emotions are implicit in the text, in many cases it becomes tough for a primer to take out important words and sentences with the emotions. Sentiment analysis is an information extraction task and also as natural language processing which targets to fetch out writer's feeling which are expressed as positive and negative emotions through text. Sentiment analysis is mainly used for predicting the approach of the communicator while he communicates via text messages. The internet and web is a large pool of unstructured and structured data. In extracting emotions from texting is a very challenging task.

II. LITERATURE SURVEY

In order to compute effectively a large number of machine learning classifier and statistical tools and techniques have been developed such as support vector machines, Naïve Bayes, vectors-space-model, k-nearest neighbour, decision tree, regression model, support vector machines, n-gram based and so on. These tools and techniques are applicable in various areas such as news categorization, text genre categorization, spam filtering, language identification and so forth.

A hybrid model has already been proposed which incorporates a lexicon based keyword spotting, emotion classification at sentence level with the help of machine learning, CRF-based emotion cue identification as various natural language processing techniques. The model provides a set of full text suicide notes as input and the output comes out to be a set of chosen sentences. Atleast one probable emotion expression is contained by this and is also marked a proportional emotion label. The results thus generated are merged with the help of distinct vote based merging techniques.

Language representation for spoken natural language processing system, Semantic grammars are a very usual form[4]. The text that is incoming (whether output from the speech recognizer or typed) gets mapped with these typically domain dependent grammars to unmask semantics.[5] Has put forward a framework for solving the problems such as uncertainty and ambiguity in semantic analysis for pulling out information. The deep understanding about the subject is the novel approach for this framework and the for information retrieval the most significant preceding sentence has been used for resolving the ambiguity in understanding the semantics. Fuzzy sets as well as Possibility theory have been used to extract the most probable semantics against the text which is based upon the knowledge about the subject as well as the preceding sentence.

VSM or Vector Space Model has wide use in information retrieval where we use vectors, which is depicted by each document and a separate term is depicted by the dimension. They have taken a non zero vector, if the term occurs in the document. Suppose our dictionary (or lexicon) contains n distinct terms. The lexicon or dictionary ,l, has been presented as a set of ordered terms. They have formally defined 'l' as follows:

$$l = t_1, t_2, t_3, \dots, t_n \quad (1)$$

VSM for classification has been used as the basic hypothesis as it is contiguous where the documents belonging to the same class tends to form a adjoining region, and different classes that have distinct regions need not overlap[10]. They have used five emotion classes, namely, joy ,sad, disgust, fear, anger, which will form the junction between ISEAR and SemEval dataset. Hence the number of emotion class, s=5.

A lexicon or dictionary of emotion conveying words have been built to accomplish the task of emotion extraction. We have used nine emotion labels for encoding the emotion-related words namely, 'surprise', 'interest', 'shame', 'joy', 'sadness', 'fear', 'guilt', 'disgust', 'anger'. Emotion intensities that range from 0.0 to 1.0 have been taken corresponding to these words. Our dictionary or lexicon comprises emoticons, adjective, adverb, noun and so forth. The lexicon can be manually enriched corresponding to the emotional category intensity. An example dictionary has been shown in table 1.

Table I: Sample words from the lexicon

Word	Anger	Disgust	Fear	Guilt	Interest	Joy	Sadness	Shame	Surprise
Happy	0	0	0	0	0.3	0.9	0	0	0
Ashamed	0	0	0	0	0	0	0.5	0.8	0
Argument	0.5	0.2	0.02	0	0	0	0.2	0	0.7

Table II: Sample emotions from the lexicon

Emotion Name	Emotion Symbol	Category	Intensity
Angry	X-(Anger	0.8
Happy	☺	Joy	0.6

It is high time, as social media is becoming prevalent, emoticons tend to play a very significant role in communication. Through texting, these emoticons tend to proffer another range of 'Feeling' and 'tone'. We have annexed more than 100 emoticons from numerous internet resources, where each entry has been assigned an emoticon label presumably in Table II. We have decided to use the subset of emotional states such as 'joy', 'shame', 'sadness', 'surprise', 'fear', 'guilt', 'disgust', 'anger' and 'interest', for efficient text classification[7].

III. COMPARISON

Three metrics; Precision, recall and F-score govern the effectiveness of various means used for identifying sentiments as well as emotions in text [13]. Recall measures the sensitivity or completeness of a classifier while precision measures the accuracy of a classifier. F-score is basically a harmonic mean of recall and precision measures. Table III gives an overview of the work done while detecting emotions. It basically provides a summary of earlier work done on identifying emotions in textual data.

IV. CHALLENGES

While performing sentiment analysis we face some key challenges:

1. Named entity recognition: for e.g. 300 Spartans means a movie or a group of Greeks.
2. Abbreviations, lack of capitals, poor grammar, poor spellings, poor punctuation.
3. Anaphora Resolution: for e.g. 'We went to dinner after watching the movie; it was awful'. 'It' refers to movie or dinner. This is the problem of resolving what a noun phrase or a pronoun phrase refers to.
4. Sarcasm: Not knowing about the author, we may be unable to decide if 'bad' means bad or good.
5. Parsing: We try to find out about the object and subject of the sentence and which one does the verb or adjective actually refers to.
6. Varied texting styles may often lead to confusion.
7. Difficulty in analyzing paragraphs and subjective sentences.

Author	Task Description	Emotion Model	DataSet	Accuracy
Yong-Soo Seol	Emotion detection from text using knowledge base artificial neural network	KBANN	Sentences	-
Christos Troussas, maria Virvou	Sentiment analysis of facebook statuses for language learning	Naïve Bayes Classifier	Facebook Statuses	By comparing 3 classifiers naïve bayes yield the best accuracy
Tejasvini Patil and Sachin Patil	Automatic generation of emotions for social networking websites using text mining	Naïve Bayes Classifier	Blogs, Tweets and Chatting Data	89.38%
Ricardo A. Calix	Emotion detection in text for 3-D facial expression rendering	SVM	Children stories	70.55%
Lily Dey, Nadia Afroz and Rudra Pratap Deb Nath	Emotion extraction from real time chat messenger	Vector space model	Chatting Data	44.70%
Sivaraman Sriram and Xiaobau Yuan	Classifying emotions using customized decision tree algorithm	Decision trees	Digg Dataset	84.36%

Table III: Comparison between different Emotion Models

V. CONCLUSION

From this survey paper we can easily conclude that emotion recognition and sentiment analysis in text plays a very crucial role in HCI(Human Computer Interaction systems). Texting as a communication medium is a very conventional approach. Detecting emotions from text aims to perceive a particular emotion through in-depth study of the text. This survey paper sums up the work that has already been done in this area by various researchers. After observing the various classifiers used in this survey paper it can easily be concluded that Naïve Bayes classifier has come out with the best accuracy. The future work undertakes a more advanced approach for emotion classification by using more improved featureset and recognizing emotions more accurately from sentences containing negation.

REFERENCES

- [1] D. Tsonos and G. Kouroupetroglou, "A methodology for the extraction of reader's emotional state triggered from text typography," *Tools in Artificial Intelligence*, pp. 139–154, 2008.
- [2] H. Yang, A. Willis, A. De Roeck, and B. Nuseibeh, "A hybrid model for automatic emotion recognition in suicide notes," *Biomedical informatics insights*, vol. 5, no. Suppl 1, p. 17, 2012.
- [3] J. Cichosz and K. Slot, "Emotion recognition in speech signal using emotion-extracting binary decision trees," *Proceedings of Affective Computing and Intelligent Interaction*, 2007.
- [4] C. Guinn and R. Hubal, "Extracting emotional information from the text of spoken dialog," in *Proceedings of the 9th international conference on user modeling*. Citeseer, 2003.
- [5] S. Jusoh and H. Fawareh, "Semantic extraction from texts," in *Proceedings of International Conference on Computer Engineering and Applications IPCSIT*, 2011.
- [6] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, in *Proceedings of the 23rd International Conference on Computational Linguistics*. Association for Computational Linguistics, 2010, pp. 806–814.

- [7] —, “Emoheart: conveying emotions in second life based on affect sensing from text,” *Advances in Human-Computer Interaction*, vol. 2010, p. 1, 2010.
- [8] C. E. Izard, *Human emotions*. Boom Koninklijke Uitgevers, 1977.
- [9] X. Zhe and A. Boucouvalas, “Text-to-emotion engine for real time internet communication.”
- [10] T. Danisman and A. Alpkocak, “Feeler: Emotion classification of text using vector space model,” in *AISB 2008 Convention Communication, Interaction and Social Intelligence*, vol. 1, 2008, p. 53.
- [11] Sriram, Sivaraman, and Xiaobu Yuan. "An enhanced approach for classifying emotions using customized decision tree algorithm", 2012 Proceedings of IEEE Southeastcon, 2012.
- [12] Chee Kian Leong and Yew Haur Lee, "Mining sentiments in SMS texts for teaching evaluation", *Expert systems with applications* 39(2012).
- [13] G. Vinodhini and RM. Chandrasekaran, "Sentiment analysis and opinion mining: a survey" in *proceedings of International journal of advanced research in computer science and software engineering*, June 2012.
- [14] Yong-soo seol and Dong-joo kim, "Emotion recognition from text using Knowledge-based ANN", *ITC-CSCC 2008*.

