ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

A Survey on Identification of Emotion from Text Corpus

Simpal Kawade, Prof. K. C. Waghmare

M.E Student, Dept. of Computer Engineering, Pune Institute of Computer Technology, Pune, Maharashtra, India

Assistant Professor, Dept. of Computer Engineering, Pune Institute of Computer Technology, Pune, Maharashtra, India

ABSTRACT: Emotion detection is the most important and interesting topic for human-computer interaction as huge data is available in digital form. The feelings can be expressed by facial expression, audio speech or written text. In recent years, the most common way of communication is the written text. Emotion detection from text is a research issue as it is difficult to automate the recognition of feelings through the computer. There are various methods for emotion detection based on the text corpus and different emotion models is also used to identify different class of emotion. The papers presents an overview of the different emotion detection methods and the limitation of these methods. The methods are divided based on four approaches: Keyword-Spotting approach, Lexicon-Rule based approach, Machine learning based approach, and Hybrid approach.

KEYWORDS: Emotion Word Ontology, Human Computer Interaction, Sentiment Analysis, Text Classification, Textual Emotion Detection.

I. INTRODUCTION

In recent years, social networks have become very popular because of the availability and affordability of internet enabled devices and due to which the network of interactions or relations has proliferated. Life of human is filled with emotions that are encountered in day to day life in a different manner. A person can express his emotions by audio speech, facial expressions, hand gestures and written text. One of the most common way for communicating the views and the messages in social networks is the written text. Earlier detection of emotions from the speech and facial expression has grabbed the attention of the researchers but today detection of emotions from the textual data needs more attention. There is a huge amount of data on the social network in the textual form which needs to extract the emotions from the text for the purpose of analysis and prediction of human behaviour in the business. Feelings can be easily recognized by the speech or the facial expressions but it is challenging to correctly predict the emotion or feeling from the textual information[2].

Automatic emotion detection from the text has attracted growing attention due to its potentially useful applications. For example, psychologists can better assist their patients by analyzing their session transcripts for any subtle emotions; reliable emotion detection can help develop powerful human-computer interaction devices and deep emotional analysis of public data such as tweets and blogs could reveal interesting insights into human nature and behaviour.

In luxurious products available online, the emotional aspects like brand, uniqueness and prestige for purchasing decisions are weighted more by the customer as compared to rational aspects such as technical, functional and cost[6]. If a customer is emotionally satisfied he can purchase a product and also encourage another customer to purchase that product. Emotional Marketing targets the customer's emotions to encourage him to opt for a particular brand and so results in an increase of product/service sales. Nowadays there is a wide range of products available, but the main target of the emotion detector is to create confidence in customer about a product/service he uses[3].

The short messaging language has the ability to interrupt and falsify Natural language processing tasks done on text data. Natural Language Processing tools are trained and adopted to work properly with plain text. It is very

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

sensitive to map short messaging language words to plain text words can be very sensitive in some cases. A wrong mapping can result in alternations of the meaning or it may destroy semantics under the applied context.

Emotions can be expressed in a word or set of words. Sentence level emotion detection method plays a crucial role to detect emotions or to find the clues for recognizing such emotions. Globally, the emotions are divided into six types which are joy, love, surprise, anger, sadness and fear [2]. Emotion detection is an NLP application that benefits from being able to distinguish subjective from an objective language. Subjective language is a language used to express opinions, evaluations, emotions, or speculations. Objective language is unbiased and not influenced by the writer's opinions or tastes. Both types of language are useful in text analysis. Subjective language is used for automatic subject analysis and objective language is used to extract information. Emotions are expressed in subjective language, thus subjectivity analysis can only be considered beneficial for emotion detection from the text.

The rest of the paper is organized as follows. Section 2 describes the four different approaches used for emotion detection from text and the limitations of the approaches. Section 3 presents the work of different researches on detecting emotion from text using the approaches explained in section 2. Finally, in section 4 we conclude the paper.

II. APPROACHES FOR EMOTION DETECTION FROM TEXT

There are four different approaches for emotion detection based on the text: Keyword Spotting Approach, Lexical Rule Based Approach, Machine Learning Based Approach, Hybrid Approach[11].

1.Keyword Spotting Approach

The keyword spotting approach uses the keyword pattern matching, finding the occurrences of keywords from a given set of substrings in the given text corpus. This approach identifies the emotion keywords from the given text and matches it with the predefined lexical of each emotion categories. The keyword spotting technique uses the synonyms and antonyms words from the lexical database for the specific language. Some of the lexical database in the English language are Wordnet, Dante, SentiWordnet, etc. The words in the lexical database are classified into different emotion categories like happy, sadness, anger, fear, disgust, surprise, depressed, etc.

In [7] a bootstrapping approach is proposed, which uses a small set of given seed opinion words to find their synonyms and antonyms in WordNet to predict the semantic orientation of adjectives. In WordNet, adjectives are organized into two polarity clusters and the synonyms/antonyms are oriented accordingly. To assign orientation of an adjective, the synset and the antonym set of the given adjective are searched. As the synset of an adjective always contains a sense that links it to the head synset, the search range is rather large. Keyword spotting method is shown as in the figure-1. The first step is the tokenization of the text data and then the emotion related keywords are extracted. Next, after the intensity analysis, negation check is done and then the emotion class is predicted.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

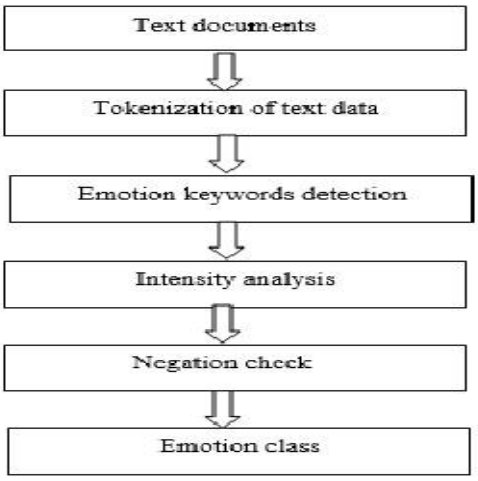

Fig-1: Keyword Spotting Approach[11]

1.1 Limitation of keyword spotting:

1. Keywords with ambiguous definition

   The detection of emotion from text by using emotion keywords is the straightforward way, but the meaning of the keywords may differ. Most of the words have different meaning according to the usage and context. Some of the words express emotion themselves. Moreover, some of the set of emotion keywords without all their synonyms could have different emotions in some extreme cases.

2. Sentences without keywords are not recognized

   The Keyword-based approach is totally based on the set of emotion keywords. Therefore, sentences without any keywords would be considered as neutral that means it does not contain any emotions at all, which is not always true. For example, “She has given me a beautiful gift” and “Hooray! She has given me a beautiful gift” should imply the same emotion (joy), but the former without “hooray” could remain undetected if “hooray” is the only keyword to detect this emotion.

3. Lack of Linguistic Information

   The syntactic and semantic structure of the sentence also have influences on emotions. For example, “I slapped him” and “He slapped me” would imply different emotions from the person’s perspective. Thus, keyword-based methods may poses a problem if linguistic information is ignored. The emotion detection method should detect not only the the existences of keywords, but also the linguistic information more accurately.

2.Lexical Rule Based Approach

Emotion recognition based on related keywords is easy to use and straight forward method. Keyword spotting technique is extended into lexical affinity method which assigns a probabilistic affinity for each emotion class depending on arbitrary words, apart from picking up emotional keyword. The lexical affinity based approach works on the word level and semantic structure of the sentence.

For example, lexical affinity might assign a 70-percent probability to the word “accident” indicating a negative effect, as in “bike accident” or “hospitalized by accident”. There are two main issues in this method. First, sentences such as “I escaped an accident” and sentences with other meanings such as “I met you by accident” creates a problem in lexical affinity, as they work only on the word level. Second, this method is often biased towards a particular type. Due to which it is not possible to reuse the same model for different systems. Thus, this approach is difficult to build and may not detect the emotion class accurately.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

3.Machine Learning Based Approach

The machine learning approach is divided into supervised learning and unsupervised learning. In machine learning based approach, the emotion is detected by using classification technique based on the training dataset. The classification of the text into a particular emotion class is based on the previous history of a document which was used in training set. The classifier automatically learns the properties of categories from the pre-classified training documents. ML-based classification is called as supervised learning because this process is guided by the labeled training set.

The unsupervised machine learning approach classifies the unlabeled data to build emotion classification model. With respect to works based on categorical emotion model, (Strapparava, C., & Valitutti, A, 2004, May) the unsupervised techniques combining LSA with WordNet Affect (Agrawal, A., & An, A, 2012, December) is applied. This proposal used the Ekman's basic emotions. (Calvo, R. A., & Mac Kim, S, 2013) proposes a novel unsupervised context-based approach based on a methodology that does not depend on any existing affect lexicon, thereby their model is flexible enough to classify sentences beyond Ekman's model of six basic emotions.

4.Hybrid Approach

There are disadvantages in each of the approaches explained above and any of the approach cannot give the accurate result. The keyword-based methods with thesaurus and naive learning-based methods could not acquire satisfactory results. Thus some of the system uses the hybrid approach by combining both keyword spotting technique and learning based method, which help to improve accuracy. This feature 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 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 [9].

III. RELATED WORK

The emotions from the text is recognized by using the different approaches explained above. The section outlines the use of the approaches for various application of emotion detection from text.

A general overview of the many diverse approaches that recognize emotion in text is presented in the paper. Subjective language is language used to express opinions, evaluations, and emotions. The results of any emotion detection algorithm needed to be compared with a human-labeled text to check the accuracy. In emotion detection, text is annotated on polarity, emotion, and intensity. In emotion detection from text, first the keywords or phrase associated with emotions is extracted. A list of emotions keywords that express each emotion class is called an emotional lexicon. The WordNet Affect Lexicon, The General Inquirer are few lexicons that can be used for emotion detection. Emotional lexicons are not successful in classifying the texts according to their appropriate emotion as all the words may not be present in the lexicon. The disadvantage of this approach is the inability to compare the majority of the algorithm and the emerging and changing message language or new form of language in chat rooms[1].

Ghazi et al.followed the hierarchical classification to classify the six Ekmans emotions. Multiple levels ofhierarchy was used while classifying emotions by first classifying whether a sentence contains an emotion or is neutral, then classifying whether the emotion is either positive or negative and finally classifying into different eight emotions classes on a fine-grained level. The main disadvantage of this approach is that it is not context sensitive[5].

Hancock et al.classify emotions as positive or negative by using content analysis Linguistic Inquiry and Word Count (LIWC). They found that positive emotions are expressed in text by using more exclamation marks and words, while negative emotions are expressed using more affective words. However, this method is limited to positive and negative emotions[7].

Yang et al.proposed a hybrid model for emotion classification that includes lexicon-keyword spotting,CRFbased (conditional random field) emotion identification, and machine-learning-based emotion classification using SVM, Naive Bayesian. This approach gives the better result but here the recognized emotions are negative only[4].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

H. Binali proposed a hybrid based architecture consisting of the keyword spotting approach and machine learning based approach. Subtle sentences expressing emotions is identified based on syntactic and semantic details. The emotion being expressed is identified based on effective lexicon resources and trained classifiers. The emotions are classified into positive and negative classes that is two dimensionalclassification technique is used. Semantic analysis was performed to enhance the detection accuracy. The limitations of the paper is that the emotions are classified as positive and negative polarity instead of different emotion categories[3].

IV. CONCLUSION AND FUTURE WORK

The detection of emotions from the textual way of expressing views help in improving the human-computer interaction. The paper, discuss different approaches for detecting emotion from text and the limitations of these approaches. The detection of emotion from the short messaging text is difficult and can be solved by using different techniques. The hybrid approach of emotion detection may give the better result than other approaches depending on the dataset. The paper also explains the use of the approaches using different dataset for emotion detection from text. The machine learning approach and the hybrid approach gives an accurate result for most of the dataset.

REFERENCES

1. S. N. Shivhare, S. Garg, A. Mishra, "EmotionFinder: Detecting Emotion From Blogs and Textual Documents", International Conference on Computing, Communication and Automation (ICCCA2015), vol. 4, pp.52-57, 2015.
2. Ameeta Agrawal, Aijun An, "Unsupervised Emotion Detection from Text using Semantic and Syntactic Relations",IEEE Computer Society, Washington, DC, USA, pp. 346-354, 2012.
3. H. Binali, Chen W.,V. Potdar, "Computational Approaches for Emotion Detection in Text",4th IEEE International Conference on Digital Ecosystems and Technologies, pp. 172-177, 2010.
4. Yang, H., Willis, A., De Roeck, A., Nuseibeh, "A hybrid model for automatic emotion recognition in suicide notes", Biomedical informatics insights, vol. 5, pp. 17-24, 2012.
5. Ghazi, D., Inkpen, D., Szpakowicz,"Hierarchical versus flat classification of emotions in text", Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, Association for Computational Linguistics,vol.3 issue 2, pp. 140-146,June 2010.
6. J. Kaur and J. R. Saini, "Emotion Detection and Sentiment Analysis in Text Corpus: A Differential Study with Informal and Formal Writing Styles", International Journal of Computer Applications, vol.101, issue 9, pp. 1-9, September 2014.
7. Hancock, J. T., Landrigan, C., Silver, C., "Expressing emotion in textbased communication", Proceedings of the SIGCHI conference on human factors in computing systems, ACM,San Jose, California, USA, pp. 929-932, April 2007.
8. S. Banerjee, Prof. U. Dutta, "An Intelligent Text Processing for Emotion Detection in Text", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 6,issue 2, pp. 467-473, February 2016.
9. D. Yasmina, M. Hajar, A. Hassan, "Using Youtube comments for textbased emotion recognition", The 7th International Conference on Ambient Systems, Networks and Technologies,vol. 83, issue 7, pp. 292-299, 2016.
10. Saleh Alsaleem, "Automated Arabic Text Categorization Using SVM and NB", International Arab Jornal of e-Technology, vol. 2, issue 2, pp. 94-102, June 2011.
11. C. Strapparava and R. Mihalcea, "Learning to identify emotions in text", in Proceeding of the ACM symposium on Applied computing, vol. 4, issue 6, pp. 1556-1560, 2008.
12. J. A. Bullinaria and J. P. Levy, "Extracting Semantic representation from word co-occurrence statistics", Behaviour Research Methods, vol. 39, issue 5, pp. 510-526, 2007.
13. C. Maaoui, A. Pruski, and F. Abdat, "Emotion recognition for human machine communication", Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 08), IEEE Computer Society, vol. 6, issue 4, pp. 1210-1215, September 2008.
14. C. M. Lee, S. S. Narayanan, and R. Pieraccini, "Combining Acoustic and Language Information for Emotion Recognition," Proc. 7th International Conference on Spoken Language Processing (ICSLP02),Philadelphia, PA, USA, pp. 873-876,2002.
15. S. Aman and S. Szpakowicz, "Using Roget's Thesaurus for Fine-grained Emotion Recognition," in Third International Joint Conf. on Natural Language Processing IJCNLP 2008, Hyderabad, India,vol. 9, issue 8, pp. 296-302, 2008.
16. F. Sebastiani, "Machine Learning in Automated Text Categorization," ACM Computing Surveys, vol. 34, pp. 1-47,2002.
17. Shadi Shaheen,Wassim El-Hajj,Hazem Hajj,Shady Elbassuoni, "Emotion Recognition from Text Based on Automatically Generated Rules", in IEEE International Conference on Data Mining Workshop,Berlin, pp. 383-393, 2014.
18. R. Mihalcea and H. Liu., "A corpus-based approach to finding happiness", In Proc. of Computational approaches for analysis of weblogs, AAAI Spring Symposium 2006, Stanford, pp. 1157-1161, March 2006.

BIOGRAPHY

Simpal Kawadeis a Student pursuing M.E. in the Computer Engineering Department, Pune Institute of Computer Technology, Pune. Her research interests are Data Mining and Machine Learning.

Prof. K. C. Waghmareis a Assistant Professor in the Computer Engineering Department, Pune Institute of Computer Technology, Pune. Her research interests are Data Mining, Data Structure, Design and Analysis of Algorithms.