



COMPREHENSIVE STUDY FOR EMOTION DETECTION USING SPEECH DIALOGUE

RAJAT YADAV and ANURAG MISHRA

Computer Science and Engineering Department

ABES Engineering College

Ghaziabad, U.P. 201009, India

E-mail: rajat.16bcs1178@abes.ac.in

anurag.mishra@abes.ac.in

Abstract

Emotion is a very important part of any interaction whether it is human-human or human-computer interaction as an emotion conveys the essence of the interaction. Decision making is also influenced by the emotional state of a person. If we can capture emotion(s) from a human-computer interaction then it will be very beneficial for the computer to learn from that interaction. It can be efficiently used by an artificial intelligence system to understand and give response accordingly. The method proposed in this paper uses a combination of different approaches to detect emotions from the text obtained from the speech dialogue.

Introduction

An emotion describes the feeling of a person which can be expressed in more than one way like when a person communicates then the tone of the speech signifies a lot about the emotional state of the person at that time. In speech synthesis a lot of factors have to be taken into consideration to detect the emotion. Detecting emotion using speech synthesis is a different domain and since our focus is to detect emotion from text which will be generated from the speech, we will limit ourselves to the emotional analysis of text. There are six basic emotions which were determined by psychologist Paul Ekman and they are known as Ekman's six basic emotions which are anger, disgust, fear, happiness, sadness and surprise. But recent studies show that

2010 Mathematics Subject Classification: 62H30, 68T10.

Keywords: Emotion detection, Emotion detection from text, Text classification, Sentiment analysis.

Received May 20, 2020; Accepted July 31, 2020

there are more than six emotions such as admiration, boredom, confusion, satisfaction and more like these. The objective of this paper is to detect emotion from a text using a combination of two approaches which are keyword spotting and context-based approach. Keyword spotting approach is based on spotting specific keywords for detecting emotions. Context-based approach detects emotion on a sentence level by understanding the context of a sentence. The model used for the context-based approach is a machine learning model. The problem here is to identify a single emotion and also multiple emotions if present. The text may contain multiple mixed emotions. So, the goal is to detect all the emotions present in a text. However, to find an exact emotion from a text is a very difficult task so the method proposed in this paper detects emotion which can be exact or very close to the actual emotion.

Related Work

Emotion classification on the basis of text is usually based on supervised machine learning techniques which uses labeled training data. Two types of labeled datasets are : manually labeled and pseudo labeled. The quality of manually labeled dataset is high since each sample is annotated by multiple humans. Pseudo labeled dataset is generated by tagging social media posts with emoticons. To retain the high quality with small size datasets an ensemble method is used that uses both a bag-of-words (BOW) based linear model and a pre-trained word vectors based non-linear model. This is the first research paper where use of pretrained word vectors was analyzed for improvement in emotion detection from text. A technique CLASS was introduced which can represent a document as a dense vector based on the significance of words in the document. By using the CLASS technique with combination of BOW and embedded representations in an ensemble, better results were obtained in domain-specific datasets in comparison to older methods. This method requires very less computing power and fits a less number of model parameters as compared to various deep-learning approaches. [1]

Keyword spotting technique has been used extensively in the field of speech and emotion recognition. This study has combined both keyword

spotting and semantic analysis for emotion detection in blog reviews. A tool known as Rapid Miner which is a powerful analytics platform was used to obtain the texts from review sites. Due to the limitations of keyword spotting approach semantic analysis was used to match the words with the appropriate emotion with the help of WordNet dictionary. Synonym sets were used to link the relations with the help of classifiers. Classification algorithms like *K*-Nearest Neighbor (KNN), Naïve Bayes and Support Vector Machine (SVM) were used in the analysis. The study involved six months of blogs reviews of restaurants and hotels in the Philippines. For validation of results annotation was done by experts in the post-processing stage. The developed system produced a low result because of the small training dataset containing only over 1000 combinations. As the blog reviews were written by professionals they were technical in nature and they contained words which did not produce any emotion. Due to this neutral emotions were collected mostly. Results were better with 1000+ combinations as compared to 200 combinations. It would be very useful to include dictionaries of proper nouns to identify words and to use adjectives and adverb to transform the semantics. Using a domain with more emotional content such as blogs on human relationships was recommended. [2]

Various approaches are being used for emotion recognition from text such as learning based method, lexical affinity and keyword spotting. This paper proposes an architecture based on spotting keywords and on ontology of emotions. The use of ontology makes this model more effective in recognizing Text emotions. An ontology describes the concepts and relationships that may exist for an agent or an agent community. Ontology have definitional aspects like high level schemas and aspects like entities and attributes interrelationship is between entities, domain vocabulary. The emotion detector algorithm given in this paper compares the calculated scores for every emotion class of the primary level with their respective secondary and tertiary level classes and emotion class with the maximum score is assigned as the emotion of the input text. This paper explores the prevailing research of emotion recognition in textual data and proposes an emotion recognition system for improving emotion detection abilities in an efficient way with the help of ontology. [3]

Many emotion detection methods use supervised learning techniques

which require a lot of annotated data for training the model. The methods which use dictionaries of emotion keywords and use an affect lexicon based approach are limited to fixed number of emotion categories in the dictionary. This paper suggests an unsupervised approach that recognizes the emotion based on context and does not rely on dictionaries and annotated training data to influence. When beginning a small set of representative terms, an effect-bearing word's emotional vector is determined by measuring the semantic relation between the word and the emotion. Three types of syntactic dependencies are used for deciding the context of the word to enhance the emotion vectors. In this paper, a context-sensitive, unsupervised technique is used to identify emotions from text. This method requires no annotated data set, or any comprehensive lexicon effect. The results indicate that the method yields more accurate results than other recent unsupervised techniques and comparable results with some of the supervised techniques. One of the weaknesses of this method is that the Semantic Relationship scores depend on the text corpus from which they are taken. The Wikipedia corpus was observed to be better than the other two corpora, and the context-free approach was outperformed by context-based approach. [4]

The approach used in this paper can evaluate the level of joy, sadness, anger and fear in a text. The aim here is to understand the meaning of the emotions observed and to identify the inferences of the obtained results. It is shown that in texts of the same literary genre as children's stories, the general emotional propensity of the writer, the variations in the texts of the same writer, variations in the emotions expressed by semi-formal writing, the proposed approach will identify different emotions. In this paper is presented the emotional analysis performed on the stories of some children from five different writers. Around 20 writings of two well-known Spanish writers from different periods were analyzed in each category of prose and poetry: Gustavo Adolfo Bécquer (1836-1870) and Mario Benedetti (1920-2009) to identify a general emotional profile for both writers. It can be inferred that the identification of emotions in texts is an efficient way of increasing user models in various contexts such as *e-learning* and help desk systems. Another possible function for the method discussed is to construct a given user's emotional profile by analyzing the emotional pattern of the

user's texts. The system could track and raise an alarm after the profile has been developed when a text has an emotional signature too far from the user's usual pattern. [5]

The first goal of this paper is to detect different predefined speech styles on the basis of only the text content of human-computer interaction. The second goal is to use the recognized speech style in the dialogs between the intelligent entities for signing aspects such as speech style or other nonlinguistic aspects. The third goal is to recognize the emotion of the participant in the communication. The system described in this paper is implemented in the Hungarian language. For training the text a small set was gathered comprising of emotions such as sadness, happiness, anger and surprise. Little corpus was also gathered for the speech style. The tests show that both the speech styles and the emotions have been recognized well. The solution for the speech style classification has been implemented in the Java programming language. A test bed known as VirCA (Virtual Collaboration Arena) which is a modular, virtual test bed with 3D functionality was used for our system. VirCA's network communication is based on Open RTM software platform. The system used in this paper is integrated into this platform as RT-Component (Automatic Text Classification component). The application includes spoken dialogue systems, audio-visual games, vehicle-human interaction. [6]

As compared to emotion recognition in prosody, physiological state and facial expressions there is a lack of work on emotional recognition from text. The obligation to understand emotions is supported by a report "W3C is to work on a web standard for Emotion Markup". The approaches such as keyword based, learning based and hybrid based presently dominate the emotion detection task. Mainly syntactic features such as n -grams, pos tags, phrase patterns and semantic features like synonym sets are used by these approaches for emotion detection. This paper exhibits that in what ways emotion detection theories have been used as a basis for building models that have been used in this paper. This paper demonstrated a literature review of the recent research in this area. A hybrid approach for emotion detection was proposed in this paper that shows that syntactic and semantic knowledge can greatly improve the prediction accuracy which is justified with experimental results. [7]

This paper focuses on recognizing emotions from short blog texts in which the author's mood is not described explicitly. This paper explores which emotional signals are accessible in a computer-mediated environment and measures the effects of top down content analysis through specific emotion and linguistic categories. Additionally data-driven techniques are applied to classify opinion and mood. Emotional linguistic features were explored in blog-text samples of 50 and 200 words. It was discovered by using automated content analysis (LIWC) that angry authors use a lot of negative effect words and happy authors use more positive effect words. The LSA method which is a semantic space co-occurrence technique was used to classify fear texts. The combination of the discussed methods may enable more accurate recognition of finer grained emotions. [8]

The field of emotion recognition and affective analysis can provide valuable contribution in different scenarios like computer assisted creativity, sentiment analysis, verbal expressivity in human computer interaction. The purpose of this paper is to recognize emotions from news headlines. Several algorithms have been discussed in this paper ranging from a simple heuristic algorithm to advanced algorithm. The simple heuristic algorithm directly checks for specific affective lexicons and the advanced algorithm checks for sameness between emotions representations in the latent semantic space and use Naïve Bayes classifiers trained using mood labeled blogposts. The approaches used are unsupervised and supervised. This paper compared results of many knowledge based and corpus based which were performed on a big data set of 1000 headlines to find out the method that works best for emotion annotation. [9]

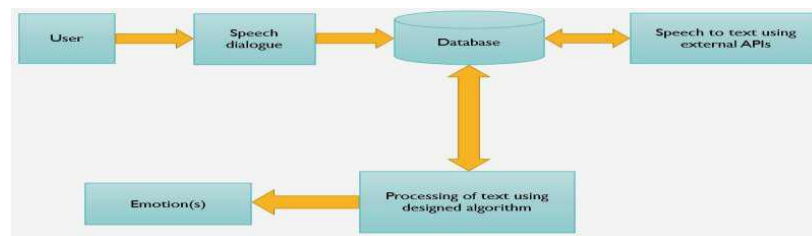
The purpose of this paper is to recognize emotions such as happiness, sadness and anger from sources containing text like e-mails and forums. Since emotion pairs like love and joy, fear and sadness have a close relationship among one another, only three emotions were selected for this study so that the emotions are clearly distinct from each other. Two different methods which are semantic network and keyword spotting method were analyzed and it was concluded that semantic emotion recognition engine was better at detecting emotions since there was no dependency on emotion keywords. This study only concentrates on texts regarding couples' break up. Similar approach can be used for detecting emotion from texts regarding

domestic violence, job dissatisfaction, and effect of teenage problems on studies. The method of emotion detection used in this paper cannot detect emotions from technical or scientific texts. [10]

Proposed Methodology

The proposed method takes a speech recording as an input from the user in the form of audio then stores it into the database. Then using the available external APIs the speech recording is converted to text. There is a limit of 100-150 words on the number of words which can be present in the speech text to increase the efficiency of the algorithm.

When the text has been generated then it is analyzed using the designed algorithm based on keyword spotting and context-based methods. Django framework is used to implement the algorithms to analyze the text. As the framework is based on python it is suitable for machine learning algorithms. After analyzing the text output is generated which is either a single emotion or multiple emotions. The emotion can be exact or near to the actual emotion which was expected.



Conclusions

Detecting emotions is always being the most curious event of interaction, so get it extracted from the delivery would definitely pull a major area of concern. Many more elaborative fields can be judged and embarked on the basis of these parameters and findings.

References

- [1] Jonathan Herzig, Michal Shmueli-Scheuer and David Konopnicki, Emotion Detection from Text via Ensemble Classification Using Word Embeddings. ICTIR '17 Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval (2017), 269-272.

- [2] Mary Jane C. Samonte, Richard Julian Paul G. Santiago, Hector Irvin B. Punzalan and Peter Joshua L. Linchangco, Emotion Detection in Blog Posts Using Keyword Spotting and Semantic Analysis, ICCIP '17 Proceedings of the 3rd International Conference on Communication and Information Processing (2017), 6-13.
- [3] Shiv Naresh Shivhare, Shakun Garg and Anitesh Mishra, EmotionFinder: Detecting Emotion From Blogs and Textual Documents, International Conference on Computing, Communication & Automation DOI: 10.1109/CCAA.2015.7148343 (2015)
- [4] Ameeta Agrawal and Aijun An, Unsupervised Emotion Detection from Text Using Semantic and Syntactic Relations, IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology DOI: 10.1109/WI-IAT.2012.170 (2012)
- [5] Linda Barros, Pilar Rodríguez and Alvaro Ortigosa, Emotion Recognition in Texts for User Model Augmenting. INTERACCION '12 Proceedings of the 13th International Conference on Interacción Persona-Ordenador Article No. 45 (2012).
- [6] Gábor Szucs and Gábor Magyar, Classification, Recognition and Feedback in Text based Metacommunication, 2nd International Conference on Cognitive Infocommunications (CogInfoCom) (2011)
- [7] Haji Binali, Chen Wu and Vidyasagar Potdar, Computational Approaches for Emotion Detection in Text. 4th IEEE International Conference on Digital Ecosystems and Technologies DOI: 10.1109/DEST.2010.5610650 (2010)
- [8] Alastair J. Gill, Robert M. French, Darren Gergle and Jon Oberlander, The Language of Emotion in Short Blog Texts, CSCW '08 Proceedings of the 2008 ACM conference on Computer supported cooperative work (2008), 299-302.
- [9] Carlo Strapparava and Rada Mihalcea, Learning to Identify Emotions in Text, SAC '08 Proceedings of the 2008 ACM symposium on Applied computing (2008), 1556-1560.
- [10] Huan Su Ling, Ranaivo Bali and Rosalina Abdul Salam, Emotion detection using Keywords Spotting and Semantic Network, International Conference on Computing & Informatics, DOI: 10.1109/ICOI.2006.5276495 (2006).