Emotion Extraction using Rule based and SVM-KNN Algorithm

Mohini Chaudhari
Department of Computer Engineering
PIIT, New Panvel, India

Sharvari Govilkar
Department of Computer Engineering
PIIT, New Panvel, India

ABSTRACT
Language is not only a powerful tool to communicate and convey information but is also a means to express emotion. Emotions are the important factors while interacting socially as emotion can easily connect people and improve health and other aspects of daily life. Emotions manipulate the way human thinks, percept and behave. We propose a hybrid system that consists of a rule-based engine and trained a Support Vector Machine (SVM) classifier. For each possible emotion, a rule-based engine is used to find whether the rule is present or not and if the rule is not present for the emotion, we require the SVM classifier in order to get the proper final decision. A set of syntactic and semantic features are extracted from sentences for building the rules and training the classifier.

General Terms
Emotion Extraction, English

Keywords
Emotion Extraction, Tokenizer, Stemmer, Rule Based Engine, Machine Learning, SVM

1. INTRODUCTION
For human beings, expressing emotions is a part of everyday communication. These emotions can be judged by a combination of clues such as the patterns of sounds and rhythms, facial expressions, gestures, and actions. Emotions are also effectively expressed by written texts. Inspired by works in sentiment analysis, the thesis explores hybrid approach for detection of emotions in text. Recognition of emotions from text is just one of the several studies of the task of making the computers understand emotions. It is a type of a content-based classification problem.

Understanding and expressing emotions between people is a complex interactional phenomenon that forms an intricate web. People can distinguish different expressed emotions very easily because they understand the meaning of the words and phrases because of their past experiences and knowledge. Humans can easily generate expressions sentences to react according to different emotions they encountered. A computer system which analyzes and interprets different emotions associated in a given text is a difficult task to accomplish.

Emotion detection from text is one of the applications of NLP. Hybrids approaches, Keyword i.e. (rule) base and Machine learning-based are the main approaches used to detect emotion. Using a Grimm's emotion-annotated dataset [12] which combines different fairy tales emotion extraction is preformed. To detect emotion different features sets like bags of words, and N-grams, were used. The paper presents a hybrid system to detect the emotions in the given text provided in English Language. In section 2, the literature survey of the past work in presented. Basic working of the system is discussed in section 3. Finally, paper is concluded in section 4.

2. LITERATURE SURVEY
This section lists the relevant past literature that use the various emotion extraction techniques. Most of the system extracts the emotions from sentences, blogs, and text documents.

Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka [1] describe an application related to Affect Analysis Model in Second Life and a lexical rule-based approach to recognize emotions from text. Based on the result of the Affect Analysis Model, they developed EmoHeart that triggers animations of facial expressions and visualizes emotion by heart-shaped textures. They propose a two-fold focus in their research: (1) recognize the affective content conveyed through text, and (2) automatic visualization of emotional expression of avatars, that allows avoiding manual control by the user and enriching remote communications effortlessly.

Daniel Dichiu, Sunita Andreea Moga, Ana Lucia Pai, Catalin Buiu, [2] said that emotions plays a vital role in the social as well as cognitive development of human beings. Robot artists are referred to as emotional robots that express themselves using various forms of art in order to evoke an aesthetic experience in the audience. Since past few years, emotion recognition from text has been an active field of research that produced several detection methods. The paper presents a cognitive system whose aim is to extract the dominant emotion from a text and change it into a graphical form using a drawing robot. This comes in the general context of a project that aims for developing robot artists who are capable of expressing artistic emotions.

The project is based on the Indian theory of Rasa, that are responsible for identifying nine artistic emotions viz. disgust, pleasure, fear, mirth, anger, sorrow, energy, astonish and serenity. The text-to-emotion module extracts various emotions from a text using an algorithm for classification which is based on a constructed graph that provides information on the words expressing emotion conveyed through the text, offers a score for each emotion, and the connections between them. The emotion-to-drawing module physically represents the dominant emotion by sketching a human face that illustrates the given emotion by using resulted graph and other input parameters. This was accomplished using a computer controlled robot as a drawing agent.

Azadeh Nikfarjam, Ehsan Emadzadeh, [3] has proposed a hybrid approach combining machine learning and rule based techniques. They designed a rule-based engine and trained a Support Vector Machine (SVM) classifier for each possible emotion. A set of syntactic and semantic features are extracted from sentences to build the rules and train the classifier. In

order to generate the sentence features we propose a new approach to identify a sentence's clauses and its constitutive grammatical elements and to use them to measure the polarity of a given sentence. By polarity in context with emotion extraction means a quantitative measure of the positive or negative feelings reflected in the text. They designed a rule-based engine along with a trained Support Vector Machine (SVM) classifier for every possible emotion. To handle the multiclass classification problem they implemented an emotion detector component for each of the 15 emotion categories. The main components of the system are: pre-processing, rule engine and SVM classifier.

Emotions are mental states accompanied by physiological changes. Ekman [4] identified six basic emotions – happiness, sadness, anger, disgust, surprise, and fear – in text form. The text considered for study comprises of data collected from blogs, representing texts rich in emotion content and therefore suitable for this study.

Emotion is expressed as joy, sadness, anger, surprise, hate, fear, so on. W. Gerrod Parrot [5] wrote a book named "Emotions in Social Psychology", which explains the emotion system and formally classifies the human emotions. This classification is done on the basis of an emotion hierarchy which is divided into six different classes at primary level which are Love, Anger, Joy, Sadness, Fear and surprise as shown in following table. Certain other words exists that can be categorized into secondary and tertiary levels.

Ruchika Sharma and Amit Arora[6] has made extension of the work proposed by Mullen and Collier. The system consists of a learning phase and a feature extraction phase; based on which the overall sentiment of the document is analyzed. Some of the machine learning approaches namely Perceptron, K-nearest, Naives Bayes, Maximum Entropy, SVM and Kernels were explored and Multiple Kernel outperforms them all. Multiple Kernel produces an accuracy of 90% and 92% for cross validation in 5 fold and 10 fold respectively.

Tony Mullen and Nigel Collier[8] introduces an approach to sentiment analysis using support vector machines (SVMs) to bring together diverse sources of potentially relevant information, including several favorable measures for adjectives and phrases and, where available, knowledge of the topic of the text.

3. PROPOSED SYSTEM

We proposed an idea to extract the person's state of mind (emotion) using a Grimm's emotion-annotated dataset which combines different fairy tales to identify disgust, pleasure, fear, mirth, anger, sorrow, energy, astonish and serenity artistic emotions and classify them using Linear Classifier and Rule based techniques.

The proposed approach consists of three phases:

1. Pre-processing phase

2. Feature Extraction phase

3. Emotion Extraction and Classification Phase

First the user provides input text in English (Grimm's Dataset) to the system. The system then pre-processes the input text and gives the pre-processed text to the feature extraction phase where for each word in the document, apply POS tagging, NER and stemming. Then for each sentence in the input document, emotions are detected and are given as an input to the classifiers. Finally the system gives the output as the emotion and then classifies it as positive, negative or neutral.

3.1 Preprocessing

This phase analyzes the input text; here sentence are segmented and then tokenized and also Stop words are removed and lastly Stemming is performed on the tokenized input.

3.1.1 Sentence Segmentation:

Here sentence boundary is detected where larger processing units which consist of one or more word are extracted.

3.1.2 Tokenization:

Here sequence of characters in the given text is split into tokens by finding the word boundaries. Along with the generation of token this process also evaluates the frequency values of the tokens present in the input document.

3.1.3 Stop Word Removal

This process removes the many of the most frequently used word in English which are worthless called as stop words. For example:- to, of, the, etc.

3.1.4 Stemming

The purpose of this process is to obtain the root or stem word of each word which emphasizes its semantics.

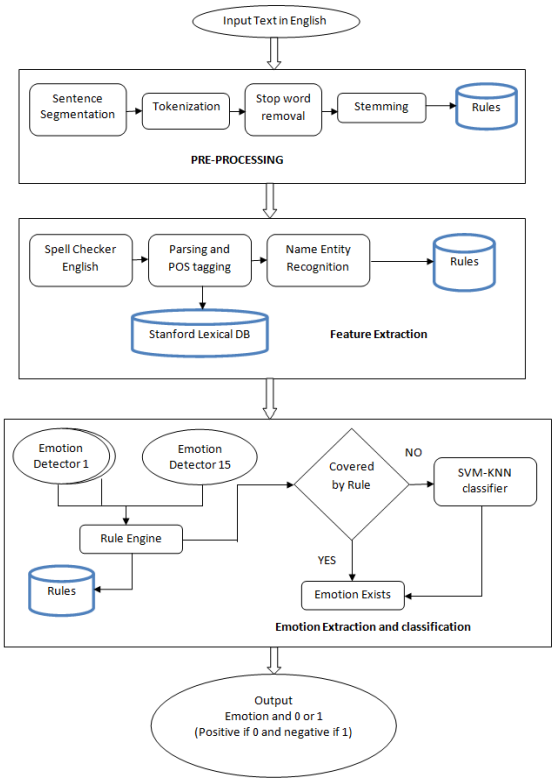

Figure 1 Proposed System

3.2 Feature Extraction Phase

This module accepts the pre-processed input sentences as input and generates POS and NER for further use. In this step vector of feature terms is used to represent every sentences which checks for statistics and linguistics of every sentence. Score is assigned to each sentence which is based on the weight of feature terms. And these scores are used to rank sentence. Here ranges between 0 to1 is used to represent feature term values.

3.3 Emotion Extraction and Classification Phase

This module accepts the Features as input and generates the nine categories of emotions such as mirth, pleasure, anger, fear, astonish, disgust, sorrow, energy, serenity and also classifies them as positive and negative as output. It consists of three parts:

3.3.1 Emotion Detector

This part detects the presence of emotion in the sentences based on three rules i.e. 0 if emotion is not present, 1 if emotion is present and 2 if emotion cannot be determined by the rules.

3.3.2 Rule Engine

Two sets of rules for each emotion i.e. positive and negative are present in rule engine. Here in case of positive rules if the rule premise is satisfied then the emotion is mostly to be present in the sentence and in negative rules if the rule premise is present then the emotion is likely to not exist in the sentence. Lexical as well as emotional clues in the sentences are used for creating these rules.

3.3.3 SVM-KNN Classifier

This phase provide the final output by passing the sentences that are not covered by any rule present in the previous sections to the SVM-KNN classifier [8]. If the SVM still fails to classify any emotion then such emotions are given to the KNN classifier. Thus we finally get the output as the emotion and its classification as positive or negative.

Following example demonstrates the working of our system:

Input: Once upon a time, there lived an unhappy young girl.

Sentence Segmentation: Once upon a time, there lived an unhappy young girl

Tokenization: Once<1> upon<1> a<1> time<1> there<1> lived<1> an<1> unhappy<1> young<1> girl<1>

Stop Word Removal: Once upon time there lived unhappy young girl

Stemming: Once upon time there live unhappy young girl

Spell Checker: Once upon time there live unhappy young girl

Parsing and POS Tagging: Once/RB upon/IN a/DT time/NN there/EX lived/VBD an/DT unhappy/JJ young/JJ girl/NN ./.

Name Entity Recognition: No output as there are no entities as such.

Emotion Detector: The output for the sentence will be 1 i.e. emotion is present

Rule Based Engine: Negative

Output: sorrow: negative.

4. CONCLUSION

Computational approaches to emotion analysis have focused on various emotion modalities, resulting in a large number of multi-modal emotion-annotated data. The method to determine the overall sentiment of the document using sentiment orientation of the words has drawbacks, as it relies only on superficial features, whereas sentiment is often communicated through the composite meaning of the text, rather than exclusively through the use of affect words. Thus, we propose to recognize artistic emotions using Grimm's emotion-annotated dataset. However, recognizing emotions using supervised machine learning approach (Hybrid) makes it more useful. In future this system can be extended to other language like Hindi and Marathi.

5. ACKNOWLEDGMENTS

Our thanks to the experts who have contributed towards development of this paper. I remain immensely obliged to Prof. Sharvari Govilkar, for providing me with the idea of this topic, and for her invaluable support in garnering resources for me either by way of information or computers also her guidance and supervision.

6. REFERENCES

[1] P. Ekman, (1992), "An Argument for Basic Emotions" International Journal of Cognition and Emotion, Vol. 6(3), published by Lawrence Associates Ltd, US, Jan 1992, p.p. 169-200.

[2] Daniel Dichiu, Ana Lucia Pai, Sunita Andreea Moga, Catalin Buiu, "A Cognitive System for Detecting Emotions in Literary Texts and Transposing them into Drawings", IEEE, 2010.

[3] Azadeh Nikfarjam, Ehsan Emadzadeh, (2012), "A Hybrid System for Emotion Extraction from Suicide Notes", Biomedical Informatics Insights 2012:5 (Suppl. 1) 165-174.

[4] Neviarouskaya A, Prendinger H, Ishizuka M. EmoHeart: Conveying Emotions in Second Life Based on Affect Sensing from Text. Advances in Human-Computer Interaction. 2010;(2):1–13.

[5] W.G. Parrot, "Emotion in Social Psychology," in Psychology Press, Philadelphia 2001, p.p. 102-105.

[6] Ruchika Sharma and Amit Arora, "Improve Sentiment Analysis Accuracy using Multiple Kernel Approach" International Journal of Computer Applications (0975 – 8887) Volume 71– No.20, June 2013

[7] Yun Li, Jie Wang, "Research on Text Classification Based on SVM-KNN", Published in Software Engineering and Service Science (ICSESS), 2014 5th IEEE International Conference on, 7-29 June 2014.

[8] T. Mullen and N. Collier, "Sentiment analysis using support vector machines with diverse information sources," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '04), pp. 412–418, Barcelona, Spain, July 2004.

[9] R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, B. Webber, "The Penn Discourse Treebank 2.0", Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008), Morocco.

[10] C. Fellbaum, "WordNet: An Electronic Lexical Database", MIT Press,1998.

[11] Alm, C. O. Affect in Text and Speech. PhD Dissertation. University of Illinois at Urbana- Champaign, 2008.

[12] StanfordParser,http://nlp.stanford.edu:8080/parser/index.jsp, June 15, 2012.