

Dynamic Mood Detection in Chat Application Using Text Pattern Analysis

Dheepthi M, M. Hemalatha

Department of Computer Science, Karpagam University, India

Article Info

Article history:

Received Sep 12, 2015

Revised Oct 10, 2015

Accepted Oct 26, 2015

Keyword:

Age group

Mood detection

Text

ABSTRACT

In modern communication and social networking the peoples of different ages are of different moods while chatting. This paper deals with detecting the modes and identifying human emotions through text mining. This paper explored to detect the mood variation of different age group that swings is maximum as compared to other age group. The moods can be classified in the basis of gender, age group and the emotion while texting. The random sample is taken from public chat in that the users are manually classified for strength of positive and negative emotions. By classifying emotions and using decision tree different variations are analyzed in this paper. Outlier study is used to recognize emotion distinction in child having any kind of disability. The pattern of the text is analysed and clustered and with the help of Besiyan classifier the text is classified in accordance with their emotions.

Copyright © 2015 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Dheepthi M,

Department of Computer Science,

Karpagam University,

Coimbatore-641021, India.

Email: dheepthimurugesan@gmail.com

1. INTRODUCTION

The computer-aided detection, analysis and application of emotion, particularly in text, has been a growth area in recent years. Almost all of this research has focused on detecting opinions in large bodies of text. For example, a program might scan a large number of customer comments or reviews of a manufacturer's products and report which aspects of which products tended to receive positive and negative feedback. Known as opinion mining or sentiment analysis, this approach typically works by identifying positive words or phrases in free text (e.g., "I like", or "rocked!") and tying them to the objects referred to (e.g., "the leather seats", "the package of extras").

From a wider social perspective, emotion is important to human communication and life and so it seems that the time is ripe to exploit advances and intuitions from opinion mining in order to detect emotion in a wider variety of contexts and for primarily social rather than commercial goals. In particular, is it now possible to detect emotion in people's textual communications and use this to gain deeper insights into issues for which emotion can play a role? For instance, how important is emotional Expression for: effective communication between friends or acquaintances, winning an online argument, automatically detecting abusive communication patterns in chat rooms, or detecting predatory behaviour online.

This article begins the process of moving from opinion mining to emotion detection by using a case study to demonstrate that it is possible to extract emotion-bearing comments on a large scale, to gain preliminary results about the social role of emotion and to identify key problems for the task of identifying emotion in informal textual communications online. Hence, although it is preliminary and exploratory it is designed to report useful information for future emotion detection research and for those interested in social network communication. Large scale data collection and analysis from social network sites has already been used for social science research goals but not yet in combination with emotion detection.

1.1. Gender and Language Usage

Gender variations in language use are relevant to the research questions in this paper, especially when they relate to emotion expression. Two key factors that are known to vary by gender are affinitive and assertive language use. Affinitive language affirms or positively engages the other person, for example by showing support or expressing agreement. In contrast, assertive language includes directive statements and criticism. Women tend to use affinitive language more and men tend to use assertive language more.

This is in broad agreement with other research that suggests that “women more often focus on the social or affective function of talk, while men tend to orientate to its referential function”. In order to partially explain findings like these, have suggested that girls learn to use language to create and maintain closeness with others through supportive and inclusive forms of speech. In contrast, boys learn to use words to assert dominance through commands and challenging statements. Of the two categories, affinitive language seems most likely to contain positive emotions and assertive language seems most likely to contain negative emotions. Some psychology theories also claim that girls have, in general, a slight advantage in language development over boys because of evolutionary pressures leading to differing brain organization and functioning and hence women may simply perform better in some communication environments.

It seems possible that increased affective style could be part of this improved performance. In terms of different forms of computer-mediated communication, it seems that offline patterns of emotion use are similar to online patterns in a variety of environments. Of particular relevance for the research in the current paper, women only discussion groups seemed to involve more emotion-related communication, with male only groups using less, with the latter groups apparently suffering as a result. Another study of internet discussion groups found women to conform to offline patterns of relatively high levels of supportive positive communication, although men were more likely to post negative comments. Research for blogs has given different results, however, with similarity rather than gender difference being the norm and with teenage males using more emoticons than females.

As a result, whilst the default assumption for a study of any new form of computer-mediated communication, such as social network site comments, should be that gender differences can occur, this should not be assumed.

2. LITERATURE REVIEW

Mining social emotion for document categorization so that it is useful for online user to select the document based on their emotional preferences, for this they have propose a joint emotion topic model with the help of latent dirichlet allocation with an intermediate layer for emotion modelling, this model provide us connection between online document and user generated social emotion. By text mining they mine the affective words and make connections with relative emotion. By this model we can uncover hidden topic that exhibits strong emotion. But problem may arise that if same word has different meaning & they may convey different emotion. These methodology can be applied in songs, emotion aware recommendation of advertisements. Further I am studying on some new techniques to detect emotion and their application area.

An enhanced approach for classifying Emotions using Customized decision tree algorithm. As there are different way to recognize emotion like from textual conversation, facial recognition, dynamic gesture recognition capturing the human body movements but as I have read emotion detection can also be done with the help of decision tree or nearest neighbour algorithm in this emotion generated rules are used, here artificial neural network is also used for emotion detection, here we find out mean and root mean square for all values in dataset, as in dataset have all seven emotion. It is also be used in real time situation such as data mining or gene prediction system but proposed paper implements this above approach in application like to classify vedio's according to their emotion.

Lyrics based emotion classification using feature selection by partial syntactic analysis:-Songs fell emotionally diff to listeners depending on their lyrical contents even melodies are similar. in this a method for lyrics based emotion classification is text-based using feature selection by partial syntactic analysis. Classification of emotion require the choice of emotion model, as such existing research on music emotion use Thayer model, tellegen-watson-Clark model. Thayer model is efficient including twp pillar representing stress and energy to classify emotion polarity. In this study examined emotion extracted through application of the syntactic analysis rule and classified them on basis of lyrics.

Emotion-Inspired Age and Gender Recognition Systems emotion-inspired age and gender recognition systems are developed. In this work, the proposed age and gender recognition systems based on arousal intensities of speaker's emotions. First, speech frames of a speaker's utterance are classified into two groups that are higher and lower than the mean of arousal intensities of speech frames. The speech frames with higher and lower arousal intensities are used for age and gender recognitions, respectively.

Here, only three ages of young, adult and senior are identified. Hence, according to the emotional arousal intensity, speech frames of a speaker's utterance are classified into two groups which are above and below the mean of arousal intensities of speech frames. Mel-scale Frequency Cepstral Coefficients (MFCC) are often used to differentiate age and gender of a human. Additionally, the classifier can be realized by k nearest neighbors, it is emotion-inspired recognition systems to identify the age and gender of a speaker. Four emotions of angry, happy, calm and sad are analyzed first. The emotions of angry and happy usually have higher arousal than those of calm and sad. From our simulations, the gender recognition prefers calm whereas the age recognition prefers angry and happy. Therefore, the recognition systems proposed herein can be widely applied to various voice interface applications.

Using Body Movement and Posture for Emotion Detection in Non-Acted Scenarios .In this paper body position and movement are used for automatically detection people's emotion in non acted scenarios. As people use to play game different emotion is observed like triumph, frustration, defeat, and concentration. our goal is to detect the subjects' emotions using their full body while they were playing games naturally a person may be able to control their facial expressions but still reveal their true emotion through their body languages. The dataset which is using having rotation angle of the people who are playing game. Different machine learning techniques are used. a range of approaches have been applied for automatically detecting emotion including recording and analyzing the person physiological responses, electroencephalography (EEG) signal, facial expression and body movement. combined information of the velocity and position of their body joints. in this different:

1. Algorithm 2 (Directed Symmetry),
2. Algorithm 1 (Pose Symmetry),
3. Algorithm 3 (Pose Difference).

Effectively detecting emotions in real time can facilitate a range of applications, e.g. help game designers dynamically adjust the content of the game and create a personalized experience for players. In this work, we explored the use of meta-features that incorporate the players body movements and posture for detecting emotions in a non-acted standing setting. we plan to improve the accuracy of our emotion detection algorithms by incorporating additional features that represent the status of the game interaction. This paper completely relying on a bottom-up approach for detecting emotions, and we will be able to consider the cognitive factors of the person as well as their experiences.

3. MOOD DETECTION BY TEXT MINING

3.1. Identifying Writing Styles

The technique SFG (Systemic Functional Grammar) explicitly identifies and represents various aspects of the general grammar, which makes it directly adoptable to the stylistic classification. SFG is the only approach that convenient to this classification. Hence the writing styles are vary from the peoples of different age groups.

Conjunctions:

Here we expose the usage of conjunctions and phrases in author's textual communication. Words and phrases that adjoin clauses such as "while", "in other words", are organized in CONJUNCTION system network. The top three levels of conjunctions are 1) Elaboration, 2) Extension and 3) Enhancement.

1. **Elaboration:** It focuses the content in its context by refocusing (for example "in other words").
2. **Extension:** It specifies adding new related information such as ("and","or","furthermore", "on the other hand").
3. **Enhancement:** It qualifies the logical connection such as ("and then ","because", "similarly").

Prepositions:

Prepositions provide the meaning expansion of phrase or clause by connecting to it a phrase. This structure of preposition system is similar to that of conjunction. Which consist of three top level option.

1. **Enhancement:** It qualifies the context spatially, temporally, or casually ("around", "thanks to", "during").
2. **Extension:** it adds the related information of an event or object ("of","without", "besides").
3. **Projection:** Using an object to interpret the meaning of another such as ("against","regarding", "according to").

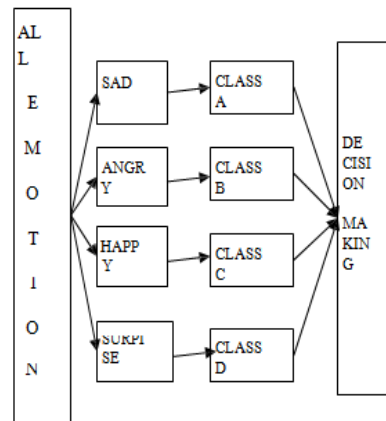


Figure 1. Classification of mood

Modality:

The MODALITY system describes the choices of typicality or necessity of facts and events, which can be realized through the modal verbs such as (“can”, “might”, “should”, “must”) and adverbial adjuncts (“probably”, “preferably”) or projective clauses. The modal expression has three attributes.

1. **Moralization:** The typical words such as (“probably”, “seldom”).
2. **Median:** The normal amount such as (“likely”, “usually”).
3. **Outer:** Extreme (either high or low) and amount (“never”, “maybe”, “must”).

Implementation:

These features have been seen both in authorship profiling and authorship attribution, here we try to determine characteristics of author such as age, sex, personality etc, even in the lack any unique candidate authors. Here we illustrate some research of authorship profiling for author age, sex, language and personality.

Even though the features used in a particular experiment, we denote the document as a numerical vector X . once we assign training documents in this way, we can apply some algorithms such as machine learning algorithms to learn classifiers that allocate new documents to categories. We learn a weight vector for each category, so that we use the effective algorithm called Plackett-Luce regression to learn the weight vector. By using this Plackett-Luce regression algorithm we can find both accuracy and efficiency and also it is well suited for large number of variables.

3.2. Training Data

As mentioned before, emotions are a matter of subjective feelings and as such they can be hardly described and objectively measured. And this is a problem. When we are to deal with training an emotion classification learning model, an absolute necessity is to have quality training data. A classification model can have only such good accuracy as the training data had For the purpose of emotion classification and identification in text data there are already several English training data sets available (such training data are often called a corpus). They are often not directly linked to emotions but in general to semantic meaning.

As training data (so called corpus in the context of text classification) approximately 1000 randomly chosen data sets were used. Each text was labelled by eight independent persons and each emotion class was marked from 0 % to 100 % according to how a person subjectively felt the strength of a particular emotion. Individual text were displayed separately in random order. Thus we hope that the impact of previous texts on the evaluation of a current result was reduced. An average value over answers of all eight persons was used as the resulting value. This was accomplished for each emotion: Anger, Disgust, Fear, Joy, Sadness, and Surprise. In case the level of resulting emotion strength did not exceed at least 20 %, it was considered not to contain any emotion and “No emotion” class was assigned.

3.3. Data Pre-processing

Before the data were provided for a learning algorithm, they needed to be pre-processed first. The pre-processing was accomplished in several steps. First, all the texts were converted to lower-case. Thus, for example, the word “pain” is considered to have the same meaning, no matter whether it is written as “Pain” or “pain”. In the next step, the sentence is tokenized. The so-called tokenization means a process when the

sentence is divided into words. To cope with various morphologies of the same word, the so-called stemmer is used, which converts the word to one common form of the word. Some words, which are insignificant to classification and potentially can have negative impact for classification accuracy (so-called stop-words), were removed from the set of tokens. These stop-words are listed at the end of this document in Appendix A. Finally, the resulting set of tokenized sentences was passed through Term Frequency–Inverse Document Frequency (TF-IDF) to determine word relevance according to each particular emotion class. It represents an example of affected words for each category of emotions.

3.4. Learning Model Evaluation

To assess the accuracy of identification and classification of emotions a 10-fold cross validation technique is used. The cross-validation is based on the principle that the training set is split into 10 disjunctive subsets. Nine subsets are used for training and the remaining subset is used for testing. This is repeated 10 times; each subset is left out of training and subsequently used for testing. Thus we get an evaluation of the model on the whole training set and the evaluation is independent of the training data. This technique is especially profitable in the case of a limited number of training data. A total of 9 different learning algorithms were used and evaluated as the learning algorithm. The list of these algorithms is provided.

4. RESULTS

In this paper an identification of the presence of an emotion (Anger, Disgust, Fear, Joy, Sadness, Surprise and No emotion) and classification according to the most dominant emotion present is analyzed. The accuracy of detection of emotional states: Anger, Disgust, Fear, Joy, Sadness and Surprise are depicted. The detection was evaluated using nine different learning algorithms as described. Each column marked with letters A, B, C, D, E, F, G, H, and I in the figure stands for its respective algorithm listed. Resulting averaged accuracy and standard error deviation among ten cross-validation evaluation is depicted for each column. This standard error is computed, where \bar{x} stands for mean value, X_i for accuracy of evaluation on a partial result of cross-validation, N stands for a number of folds of cross-validation, i.e. 10 in our case. As can be read from the figure, the most accurate results for Anger detection gave the SVM method with a linear kernel. The results were $87.3\% \pm 5.65\%$. Thus the model can identify an angry emotion in the content of news headline with this accuracy. In case of detection of Disgust emotion in the data, the algorithm with the highest accuracy was the SVM method with radial kernel. The resulting accuracy was $95.01\% \pm 2.73\%$. Compared with other emotion groups the emotion Disgust is of relatively low frequency, but a relatively high degree of accuracy with a relatively low dispersion of mean deviation is achieved. At a closer look at randomly selected training data it is clear that there are words that probably are not too numerous in the other emotional categories and, also, they can be expected to evoke exactly the Disgust emotion.

They are words like porn, cancer, campaign, be over, resign, etc. It is also interesting to note that keywords associated with politics are also relatively important in terms of Disgust emotion, regardless of whether the result is positive or negative. Detection of Fear achieved the highest accuracy with the SVM method with linear kernel. The resulting accuracy was $81.32\% \pm 2.13\%$. Detection of Joy achieved the highest accuracy with the SVM method with linear kernel. The resulting accuracy was $71.60\% \pm 2.22\%$. For Sadness the highest accuracy was achieved with the SVM method with linear kernel.

The resulting accuracy was $75.41\% \pm 4.60\%$. The last emotion Surprise also achieved the best accuracy with the SVM method with linear kernel and results are $71.09\% \pm 2.21\%$. Averaged values for these values are 80.28. The results obtained by different algorithms show that some algorithms, such as various variants of the SVM method, consistently give relatively good results. On the other hand, algorithms such as the k -NN or Bayesian-Network methods often fail. This behavior can probably be explained by the nature of the SVM methods. The SVM methods inherently tend to look for solutions with ideal hyperspace distribution. As a result, even in the case of a multi-dimensional limited training set, the SVM method gives, in comparison with other methods, interesting results (about a thousand headlines used in the training set is still a small amount compared with the number of words occurring in Czech). It is obvious that the most successful variant was obtained by SVM methods with linear or radial kernel. It is probably due to the nature of training data.

The linear kernel can easily separate classes on the basis of individual words in a linear way. On the other hand, the radial kernel should be able to easily find clusters, which lead to classification into one of the appropriate emotion classes.

5. CONCLUSION

This paper provides a description and evaluation of learning models of identification and classification of emotions in Czech news headlines. As mentioned, it is relatively difficult to accurately identify each emotion. One of the reasons can be that it is closely related to semantic understanding of a text. For example, the headline “Jestřábi se ujímají vedení” (the Hawks are taking on leadership) can have a different meaning in a sports' magazine and in a hunters' journal. However, even when someone fully understands the semantic content of the message, two different people can still have invoked totally different emotions. For example the message that “Euro kleslo na nejnižší úroveň” (the Euro has dropped to the lowest level) can be bad news for people living in Euro zone but for people having most of their deposits in the US dollars it can be good news. Based on an analysis of examples where the greatest problems in classification appear, it is clear, that future work should consist in using a significantly larger training set or combining the approach described in this paper with the ontological base of the Czech language such as the Czech WordNet database. Any method that would be able to take into account the semantic meaning of sentences and identify, whether the event being described is usual or extraordinary, would also bring significant improvements.

REFERENCES

- [1] http://www.hhs.gov/opa/familylife/tech_assistance/etraining/adolescent_brain/Overview/what_is_adolescence/index.html
- [2] http://wiki.answers.com/Q/What_are_some_of_the_emotional_changes_during_the_age_of_adolescence
- [3] G. K. Gupta, “Introduction to data mining and case studies”, Second edition.
- [4] Shenghua Bao, Shengliang Xu, “Mining Social Emotions from Affective Text”, 2012 IEEE, September 2012.
- [5] Sivaraman sriram, xiaobu yuan, “An enhanced approach for classifying emotions using customized decision tree algorithm”, 2012 IEEE, 2012.
- [6] Minh Kim, Hyuk-Chul Kwon, „Lyrics-based Emotion Classification using Feature Selection by Partial Syntactic Analysis”, 2011 IEEE, 2011.
- [7] Oscar T. C. Chen, Jhen Jhan Gu, Ping-Tsung Lu and Jia-You Ke, “Emotion-Inspired Age and Gender Recognition Systems”, 2012 IEEE, 2012.
- [8] Michael Garber-Barron, Mei Si, ”Using Body Movement and Posture for Emotion Detection in Non- Acted Scenarios”, 2012 IEEE, 2012.
- [9] <http://answers.yahoo.com/question/index?qid=20080614190226AAAtyXkm>