Machine learning for text-based emotion

Swati D. Bhutekar
Mtech , Second Year , CSE, Shri Ramdeobaba Kamla Nehru College of Engg., Nagpur, India

Manoj. B. Chandak
Head of Department, Computer Science of Engg. Shri Ramdeobaba Kamla Nehru College of Engg., Nagpur, India

ABSTRACT
This paper presents an machine learning approach as classification concept will be needed to emotion recognition from textual content. This paper also focuses on Emotion engine available, Corpus needed , textual emotion recognition module etc. In text analysis, all emotional keywords and emotion modification words are manually defined. The emotion intensity levels of emotional keywords and emotion modification words are estimated based on a collected emotion corpus. The final emotional state is determined based on the emotion outputs from the acoustic and textual analyses.

Keywords
Emotion, Machine learning, Cognitive Styles

1 INTRODUCTION
Text does not only communicate informative contents, but also attitudinal information, including emotional states. Emotion engine is also used to extract the emotions from textual content. But there are certain questions like, How can emotions be recognized in text? 2) What word features have the highest contribution to detection of emotions in text? 3) How can emotion detection be used to render facial expressions? and 4) Are automatically selected word features as good as human selected word features in emotion classification? & To address these questions, study used supervised machine learning techniques to classify sentences into one of a set of predefined emotion classes.

The supervised classification approach relies on a corpus of children's stories which has already been annotated with affect labels . To perform the classification task, each sentence is represented as a feature vector and classified using support vector machines (SVM). Performance was measured using precision and recall accuracy scores. After the classification task, a feature analysis was performed to identify feature contribution to accuracy and to highlight possible approaches to emotion expression in 3-D face rendering. The solution builds on recently developed corpora for emotion detection. It also illustrates mechanisms which can be used in a pipeline-based architecture in existing and emerging technologies for 3-D graphics generation and animation. One of the supervised machine learning is SNoW learning architecture.

2 LITERATURE REVIEW

2.1 Corpora
Emotion annotated corpora are not as common in the research literature as other types of general purpose corpora such as the Brown corpus or the Penn Treebank. There are several implementations such as MPQA , ISEAR, EARL, and Movie Reviews but none that could be considered a gold standard corpus. Most of these corpora are hybrids which started with opinion annotations and which were later extended to include some emotional tags. There is currently no standard emotion mark-up language which can be used to develop a corpus with emotion tags. For the purposes of this study, we selected the UIUC children's stories corpus . This is a new corpus designed specifically for emotion classification which was developed as part of a dissertation project by Alm at UIUC. Here in this study, we provide a word feature-based analysis of the corpus, and we use it for emotion classification from text.

2.2 Emotion Classes
There are hundreds of words in the English language that can be used to represent emotions. Most of these, however, do not represent unique emotions but instead are varying levels of the degree to which a person experiences an emotion. An example of this is being ecstatic, which is a level of the degree of happiness that a person experiences. With this in mind, almost all emotions can be classified as either positive or negative. Some previous studies have added a third category to represent emotions such as astonishment which are neither positive nor negative . An important early work on human emotions can be traced back to Descartes' "The Passions of the Soul". In this work, Descartes proposes new ideas about sensations and perceptions of man. This work was important because it allowed emotions to be analyzed from a scientific point of view. Researchers like Plutchik reason that emotions can be categorized into two main groups. The first group consists of eight primary emotions. These emotions are: anger, fear, sadness, disgust, surprise, curiosity, acceptance, and joy. Plutchik argues that these eight emotions are biological in nature and are essential for human survival. The second group of emotions consists of combinations of the emotions in the first group. Most implementations in the literature seem to follow this basic set of classes and for the purposes of comparison, this study uses a modified version of primary emotions with eight categories. Preliminary analysis used the eight classes but subsequent analysis used a reduced set of six classes . The six classes used in this study were: anger, fear, happiness, sadness, surprise, and neutral (which is the absence of emotion). It is important to note that six emotion classes are common in the field and are referred to as "The Big Six".

2.3 Emotion Detection—Previous Work
Supervised learning techniques in natural language processing and semantic analysis are commonly used to detect and classify information in a document. Supervised learning has been extensively used in topic detection , entity recognition , and extraction of temporal relations , to name a few applications. Some s authors conducted empirical studies to determine the emotional affinity of sentences in the domain of children's stories. To achieve this, Alm developed the affect corpus and explored different types of features and their contribution to emotion classification. Alm also explored affect in speech, as she was particularly interested in expressive text-to-speech synthesis. In certain cases, the authors use a knowledge-based approach to detect emotion in

text. They argue that common sense knowledge about the world is required to disambiguate and detect emotions.

2.4 Word Feature Selection

Previous research has performed a semantic analysis of English words which relate to emotions. There are basic sets of words that relate to each of the basic emotions. Other Works also suggest that emotion states are associated with specific sets of words. The use of feature selection techniques in emotion detection from text is critical to achieving good prediction results. The use of these techniques has ranged from simple bag-of-words approaches to more recent implementations using semantic representations such as the work by Mathieu. Of the studies that we reviewed, most used lists of emotion words along with other features such as POS count. These lists can be obtained manually or automatically from different sources such as Word-Net and its extension Word Net-affect . There are many feature selection methods in statistical learning for text categorization. These are generalized approaches which take a given set of features and try to either map the features into a new set of transformed features or try to reduce the number of features without transformation . Since the relation between words and emotions is our main focus for understanding the types of features that have greater influence in emotion detection, this study only uses the feature reduction approach to feature selection.

Some of the main reduction-based feature selection techniques used in the literature include: document frequency (DF), mutual information (MI), the chi-square statistic, and information gain (IG) . Document frequency (DF) is a method used to calculate the number of documents in which a word appears. IG determines the number of bits of information contributed by a word for classification by knowing the absence or presence of a word in a given document . MI is a method which calculates the two-way contingency between words and categories. It measures the mutual dependence between two variables . Chi-square is similar to MI with a major difference being that chi-square values are normalized . This study focuses on effects of MI on the classification task. MI was selected because of its tendency to favour terms that occur with less frequency . Its use seems appropriate in emotion detection since terms can occur infrequently but have high contribution to classification. Formally, MI can be defined as follows: (1) where represents the mutual dependence between and , the numerator is the joint probability between and , and are the individual probabilities for and respectively, is the term, and is the class.

2.5 Machine Learning: Supervised Learning

One of the learning mechanism used in this study was SVM . This supervised learning machine was implemented because it provides the option of evaluating the data under different spaces and functions through kernels that range from simple linear to radial basis function (RBF) . Its wide use in the field of machine learning research and ability to handle a large set of features makes it an attractive tool to explore this relatively new emotions corpus . SVM is a binary classification method based on statistical learning theory which maximizes the margin that separates samples from two classes . The maximization of the margin is based on the training samples that are closest to the optimal line (also known as support vectors). Because the method tries to maximize the margin between the samples of two classes under a set of constraints, it ultimately becomes an optimization problem to find the maximum separation band. The function that represents the margin is quadratic and can be easily solved using quadratic programming techniques with Lagrange operators. The "objective function" and constraints can be represented as follows: where represents each training sample, is the weight vector (normal vector) that defines the maximum margin, is the bias, is the class for each training sample , and is the feature vector for each training sample . The term in the model represents the slack error for each sample. The cost parameter represents the penalty for each error. In this formulation of the "objective function", the optimization is a trade-off between the maximal margin and the slack error for each sample. Although the SVM is used for binary classification, multiclass implementations can be achieved by creating different classifiers for each pair of class comparisons. The implementation used in this study Chang and Lin's relies on a one-against-one approach. The class imbalance challenge in the data set is an important issue that needs to be addressed when using SVMs and most other machine learning techniques Machine learning model is used to determine the emotion of a linguistic unit can be cast as a multi-class classification problem. For the flat case, let *T* denote the text, and *s* an embedded linguistic unit, such as a sentence, where *s 2 T*. Let *k* be the number of emotion classes *E* = *fem*1*;em*2*; ::;emkg*, where *em*1 denotes the special case of *neutrality*, or absence of emotion. The goal is to determine a mapping function *f* : *s ! emi*, such that we obtain an ordered labeled pair (*s;emi*).The mapping is based on *F* = *ff*1*; f*2*; ::; fng*, where *F* contains the features derived from the text. Furthermore, if multiple emotion classes can characterize *s*, then given *E' ½ E*, the target of the mapping function becomes the ordered pair (*s;E0*). Finally, as the hierarchical case of label assignment requires a sequential model that further defines levels of coarse versus fine-grained classifiers, for the *question classification* problem.

3 SENTENCES, EMOTIONS AND COGNITIVE STYLES :

Chomsky and Saussure demonstrated that sentences can be broken down into two groups of words: function words and content words. Nouns, adjectives, verbs and adverbs belong to content words while function words are prepositions, conjunctions and auxiliary verbs. According to what groups the words belong to and the words preceding and subsequent, the sentences can be decomposed and be analysed.

Rewrite rules or so-called tree presentation are widely used tools to analyse sentence structure. Rewrite rules apply to a string of characters and replace some of these characters by another string. Chomsky showed that grammar could be described in terms of a finite number of rewrite rules capable of generating all grammatical sentences of a given language.

The categories are happiness, sadness, anger, fear, disgust and surprise. Within each of the categories a wide range of expression intensity and variation of detail expression exists. User CS is the consistent and constant in time underlying method of an individual's thinking and perceiving that subsequently affects the way in they perceives and responds to events and ideas. The classification of CS has been found to explain the behaviour of subjects performing simple tasks using text.

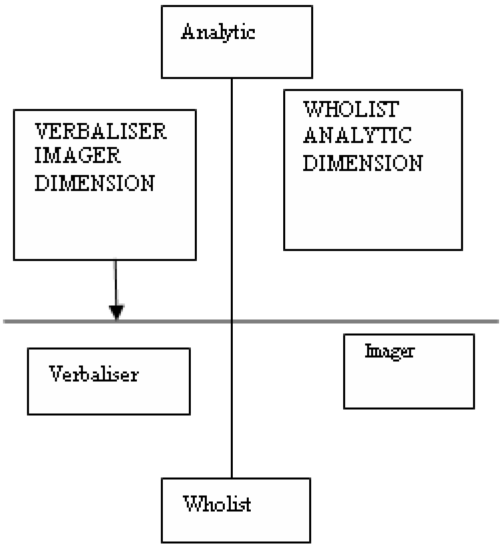

Fig. 1: Cognitive style dimensions

The CS Analysis (CSA) test package divides CS as a function of two independent dimensions; the Wholist-Analytic dimension and the Verbal-Imagery dimension. The Wholist-Analytic dimension determines whether people understand situations as a whole or see things in parts. The Verbal-Imagery dimension determines whether individuals are outgoing and verbal or more introverted and think in terms of mental pictures or images. The CS independent dimensions are shown in figure 1.

4 EMOTIONAL KEYWORD DEFINITION

For each emotional keyword, the corresponding emotion descriptor is manually defined. The emotion descriptor is a set of descriptions of the emotion reactions corresponding to the keywords. Basically, it contains an emotional state label and an intensity value, which ranges Multi-Modal Emotion Recognition from Speech and Text 53 from 0 to 1. The emotional state label can be one of the following six labels: happiness, sadness, anger, fear, surprise, and disgust. The intensity value describes how strongly the keyword belongs to this emotional state. In many cases, however, a word may contain one or more emotional reactions. Accordingly, there may be more than one emotion descriptor for each emotional keyword. For example, two emotional states, sadness and anger, are involved in the keyword "disappointed." However, the keyword "depressed" is annotated with only one emotional state: sadness. After the tagging process is completed, the emotion descriptors of the word disappointed" are {(2, 0.2), (3, 0.6)}, and the emotion descriptor of the word "depressed" is {(3, 0.6)}. The numbers 2 and 3 in the parentheses indicate the emotional states anger and sadness, respectively. The numbers 0.2 and 0.6 represent the degree of the emotional states. In the following, we describe how the emotional state is calculated. Consider the following input sentence at time *t*: ***St*** : "We felt very ***disappointed*** and ***depressed*** at the results." Here, the *ith* emotional keyword is represented by *t ki* , 1£ *i* £ *Mt* , and *Mt* is the number of keywords in sentence *St*. In this example, 1 *k t* and 2 *t k* represent the words "***disappointed***" and "***depressed,***" respectively, and the value of *Mt* is 2. For each emotional keyword *t i k* , the corresponding emotion descriptor is (*ti* , *ti*) *r r l v* , 1 *t i* £ *r* £ *R* , where *t i R* represents the number of emotion descriptors of *t i k* . The variable *ti r l* is the *rth* emotional state label, and *ti r v* is the *rth* intensity value of *t i k* . The value of the emotional state label can range from 1 to 6, corresponding to six emotional states: happiness, sadness, anger, fear, surprise, and disgust. In this case, the values of 1 *Rt* and 2 *t R* are 2 and 1, respectively. For 1 *kt* , the values of 1 1 *lt* , 1 2 *t l* , 1 1 *t v* , and 1 2 *t v* are 2, 3, 0.2, and 0.6, respectively. For 2 *t k* , the values of 2 1 *t l* and 2 1 *t v* are 3 and 0.6, respectively.

The emotion descriptors of each emotional keyword are manually defined based on a Chinese lexicon containing 65620 words. In order to eliminate errors due to subjective judgment, all the words are firstly tagged by three people individually and then cross validated by the other two people. For each word, if the results tagged by different people are close, the average of these values will be set as the emotion descriptors of the word. If the three people cannot reach a common consensus, an additional person will be asked to tag the word, and the result will be taken into consideration. Based on experience, only a few words need additional suggestions.

The final tagged results for the emotion descriptors . A total of 496 words are defined as emotional keywords, and there are some ambiguities. Only 423 of them have unique emotional label definitions, 64 words have 2 emotional label definitions, and 9words have 3 emotional label definitions. Most of the ambiguities occur in the anger and sadness categories. For example, the word "unhappy" may indicate an angry emotion or a sad emotion, according to the individual's personality and situation.

Number of tagged emotion labels of an emotional keyword			Total
1	2	3	496
423	64	9	

Table 1. The ambiguity of tagged emotion labels for an emotional keyword.

5 TEXTUAL EMOTION RECOGNITION MODULE

The most popular method for performing emotion recognition from text is to detect the appearance of emotional keywords. Generally, not only the word level but also the syntactic and semantic levels may contain emotional information. Figure 3 shows a diagram of the textual emotion recognition module. A front-end speech recognizer is first used to convert the speech signal into textual data. To extract the emotional state from the text input, we assume that every input sentence includes one or more emotional keywords and emotion modification words. The emotional keywords provide a basic emotion description of the input sentence, and the emotion modification words can enhance or suppress the emotional state. Finally, the final emotional state is determined by combining the recognition results from both textual content and speech signal.

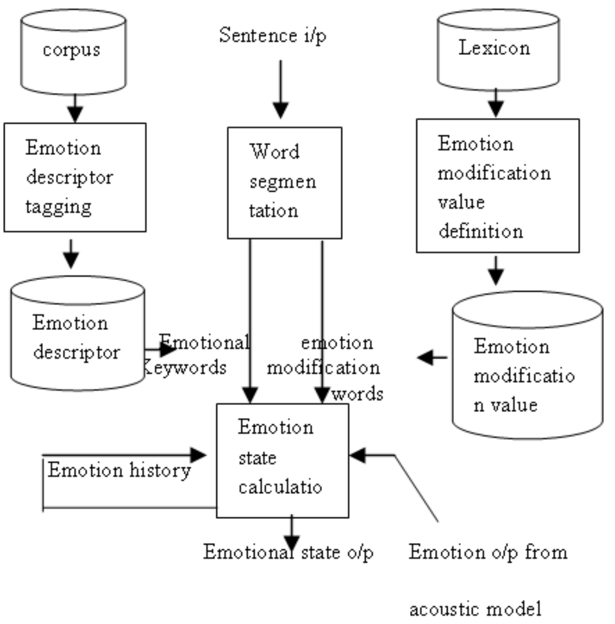

Fig. 2. Diagram of textual emotion recognition module.

6 EMOTION EXTRACTION ENGINE

We have developed an emotion extraction engine as part of a client-server text based internet communication system which instead of transmitting pictures through the network, only transmits the parameters needed for generating the emotional images. As a result the bandwidth requirement is extremely low.

The emotion extraction engine includes three parts: input analysis, tagging system and parser. Here we present a general description, more detailed information, is given in. The working flow of the engine is shown in figure 2.

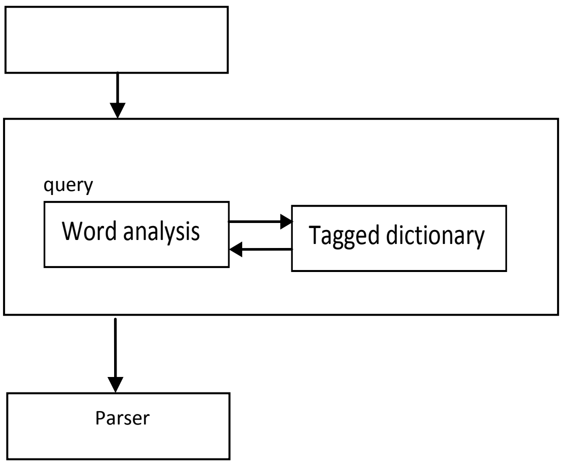

Fig. 3.: The working flow of the engine

6.1 Input analysis function

The engine analyses only one sentence at a time. The input sentences are analysed individually without context information. The punctuation is replaced with the pre-defined characters.

6.2 The Tagging system

The word analysis function splits the sentences from input analysis function into words and searches the tagged dictionary to find the corresponding tag category. The outputs include the word category and the emotional tag. Daily communications involve about two thousand words . In order to identify the words, a special designed dictionary was set up. In this project, a database containing 16400 words was used. The database includes three fields: word, category and emotional tag. The category field contains the corresponding word category (noun, verb, adjective etc) and the emotional tag field describes whether that word belongs to one of the six emotion types. Unlike tagging methods used in some existing systems, the entire word is appropriately tagged in the dictionary in order to keep the response time to a minimum.

6.3 Parser

Receiving the output from the tagging system, the parser will try to identify the emotion content. The parser's analysis is accomplished through the use of rewrite rules and tree representations . According to pre-defined rules, the parser will search for the current emotional words, the person to whom the emotional words refer to and the intensity of the emotional words. The parser’s outputs are the emotional parameters and are sent across the network.

An emotion extraction engine, which can analyse text sentences typed by the users, is presented. The sentences are analysed and the detected emotive content and the appropriate expressions are displayed automatically. The intensity and duration of the expression are also calculated and displayed in real time automatically

7 PREVIOUS WORK

For a complete general overview of the field of *affective computing* is a rare study in text based inference of sentence-level emotional affinity. The authors adopt the notion of *basic emotions*, and use six emotion categories: ANGER, DISGUST, FEAR, HAPPINESS, SADNESS, SURPRISE. They critique statistical NLP for being unsuccessful at the small sentence level, and instead use a database of common-sense knowledge and create affect models which are combined to form a representation of the emotional affinity of a sentence. At its core, the approach remains dependent on an emotion lexicon and hand-crafted rules for conceptual polarity. In order to be effective, emotion recognition must go beyond such resources; the authors note themselves that lexical affinity is fragile. The method was tested on 20 users’ preferences for an email-client, based on user-composed text emails describing short but color ful events. While the users preferred the emotional client, this evaluation does not reveal emotion classification accuracy, nor how well the model generalizes on a large data set. Whereas work on emotion classification from the point of view of natural speech and human computer dialogues is fairly extensive, this appears not to be the case for text-to-speech synthesis (TTS). A short study by addresses sentence-level emotion recognition for Japanese TTS. Their model uses a composition assumption: the emotion of a sentence is a function of the emotional affinity of the words in the sentence. They obtain emotional judgements of 73 adjectives and a set of sentences from 15 human subjects and compute words' emotional strength based on the ratio of times a word or a sentence was judged to fall into a particular emotion bucket, given the number of human subjects. Additionally, they conducted an interactive experiment concerning the acoustic rendering of emotion, using manual tuning of prosodic parameters for Japanese sentences. While the authors actually address the two fundamental problems of emotional TTS, their approach is impractical and most likely cannot scale up for a real corpus. Again, while lexical items with clear emotional meaning, such as *happy* or *sad*, matter, emotion classification probably needs to consider additional inference mechanisms. Moreover, a natıve compositional approach to emotion recognition is risky due to simple linguistic facts, such as context-dependent semantics, domination of words with multiple meanings, and emotional negation. Many NLP problems address attitudinal meaning distinctions in text, e.g. detecting *subjective* opinion documents or expressions, measuring *strength* of subjective clauses , determining word *polarity* or texts' attitudinal valence,. Here, it suffices to say that the targets, the domain, and the intended application differ; our goal is to classify emotional text passages in children's stories, and eventually use this information for rendering expressive child-directed storytelling in a text-to-speech application. This can be useful, e.g. in therapeutic education of children with communication disorders

8 APPLICATION AREA: TEXT-TO-SPEECH

Narrative text is often especially prone to having emotional contents. In the literary genre of fairy tales, emotions such as HAPPINESS and ANGER and related cognitive states, e.g. LOVE or HATE, become integral parts of the story plot, and thus are of particular importance. Moreover, the story teller reading the story interprets emotions in order to orally convey the story in a fashion which makes the story come alive and catches the listeners' attention. In speech, speakers effectively express emotions by modifying prosody, including pitch, intensity, and durational cues in the speech signal. Thus, in order to make text-to-speech synthesis sound as natural and engaging as possible, it is important to convey the emotional stance in the text. However, this implies first having identified the appropriate emotional meaning of the corresponding text passage. Thus, an application for emotional text-to-speech synthesis has to solve two basic problems. First, what emotion or emotions most appropriately describe a certain text passage, and second, given a text passage and a specified emotional mark-up, how to render the prosodic contour in order to convey the emotional content. The *text-based emotion prediction* task (TEP) addresses the first of these two problems.

9 REFERENCES

[1] Xu Zhe, David John and Anthony C. Boucouvalas, " Text-to-Emotion Engine: Tests of User preferences" Multimedia Communications Research Group, Bournemouth University,

[2] Ricardo A. Calix, Sri Abhishikth Mallepudi, Bin Chen, and Gerald M. Knapp, "Emotion Recognition in Text for 3-D Facial Expression Rendering," IEEE Transactions on Multimedia, vol. 12, no. 6, Oct. 2010, pp- 544

[3] Fazel Keshtkar, Diana Inkpen," A Corpus-based Method for Extracting Paraphrases of Emotion Terms", Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, pages 35–44,Los Angeles, California, June 2010. Association for Computational Linguistics

[4] Curry Guinn and Rob Hubal," Extracting Emotional Information from the Text of Spoken Dialog", RTI International, 3040 Cornwallis Road, Research Triangle Park, North Carolina, USA, 27709

[5] Ze-Jing Chuang and Chung-Hsien Wu," Multi-Modal Emotion Recognition from Speech and Text", Computational Linguistics and Chinese Language Processing Vol. 9, No. 2 , August 2004, pp. 45-62 45 , The Association for Computational Linguistics and Chinese Language Processing

[6] Xu Zhe, David John and Anthony C. Boucouvalas," Emotion Extraction Engine: Expressive Image generator", Multimedia Communications Research Group, School of Design, engineering and Computing, Bournemouth University,

[7] Bassili, J.N., 1979. Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. Journal of Personality and Social Psychology, 37, pp 2049-2058.

[8] Buck, R., 2001. Emotional Expression, Communication, and Competence: Applications of a Developmental-Interactionist Theory of Biological, Social, Cognitive and Moral Emotions. University of Connecticut. http://wattlab.coms.uconn.edu/ftp/users/rbuck/Germany01/

[9] Zhe, X. & Boucouvalas, A.C., 2002. Text-to-Emotion Engine for Real Time Internet Communication. International symposium on Communication Systems, Networks and DSPs, 15-17 July 2002, Staffordshire University,UK, pp 164-168