**A Survey on Emotion Detection from Text**

Abhishek Mahakal

Pennsylvania State University, Harrisburg

amm10344@psu.edu

**1 Introduction:**

The recognition of emotions has been a topic of interest for several decades, with researchers developing techniques to detect emotions through various physiological characteristics such as voice, facial expressions, hand gestures, or body movements. Emotions can also be detected using speech and image signals, with some researchers combining different emotion recognition models to achieve better results.

However, there has been little research on recognizing emotions from only textual input, even though natural language plays an important role in emotion recognition. Traditionally, research in this area has focused on the discovery and utilization of emotional keywords, which are specific words that express the speaker's emotional state. While using emotional keywords is the most direct way to recognize a user's emotions from text input, it has its limitations, such as ambiguity in defining all emotional keywords, recognizing emotions from sentences with no emotional keywords, and the lack of semantic and syntactic information for emotion recognition.

Some researchers have used other textual information clues such as pragmatic intent, text content plausibility, and paragraph structure to recognize emotions from the text. Litman and Forbes integrated all dialog information, including acoustic and linguistic features, dialog acts, and the sequence of speakers, to recognize the emotional state in a dialog system. Schuller et al. integrated both acoustic and linguistic information in emotion recognition.

With further analysis, some researchers believe that textual data is rich with emotion at the semantic level, that is, that emotional content is also contained in the semantic structure. Chan and Franklin analyzed input sentences and constructed a symbolic network to reduce language model perplexity. Woods used a transition network to analyze natural language. Chuang and Wu proposed a semantic network-based emotion recognition mechanism using emotional keywords, semantic/syntactic information, and emotional history to recognize the emotional state of a speaker.

Research in emotion recognition has focused on discerning emotions along the dimensions of valence (positive/negative) and arousal (calm/excited) and on recognizing distinct emotion categories. Liu et al. used a real-world commonsense knowledge base to classify sentences into Ekman's basic emotion categories. They used an ensemble of rule-based affect models to determine the emotional affinity of individual sentences. Neviarouskaya et al. also used rules to determine the emotions in sentences in blog posts, relying on a manually prepared database of words, abbreviations, and emoticons labeled with emotion categories.

Although communication systems can identify the users' emotional states from different communication modalities, the variety and complexity of language makes it difficult for researchers to recognize emotional states from pure textual data. Recognizing emotion is extremely important for some text-based communication tools, such as the dialog system, which is a kind of human-machine communication system that uses only text input and output. Recognizing the users' emotional states enables the dialog system to change the response and answer types. Text is still the main communication tool on the Internet.

In conclusion, emotion recognition is an important area of research, and while significant progress has been made in recognizing emotions through physiological characteristics and speech and image signals, recognizing emotions from only textual input remains a challenge. While keyword-based approaches have been used traditionally, they have their limitations, and researchers are exploring other textual information clues and semantic networks to improve emotion recognition. Recognizing emotions from text is important for various communication systems, including dialog systems, and will continue to be an area of interest for researchers in the future.

**2. Summary of Papers before 2010:**

* 1. **Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models**

The paper proposes a new approach to emotion recognition from text using semantic labels and separable mixture models. The authors argue that traditional approaches to emotion recognition often rely on pre-defined sets of emotion categories and do not capture the complexity and variability of emotional experiences. Instead, they propose a method that is able to identify the specific emotions expressed in a piece of text by modeling the underlying semantic structure of the language.

Diagram

Description automatically generated

The proposed approach consists of two main components: semantic labeling and separable mixture models. In the first component, the authors use a pre-trained semantic parser to identify the syntactic and semantic structure of the input text. The parser breaks down the text into its constituent parts (e.g. subject, verb, object) and assigns semantic labels to each part based on its function in the sentence (e.g. agent, patient, instrument).

Once the input text has been labeled semantically, the authors use a separable mixture model to identify the specific emotions expressed in the text. A separable mixture model is a probabilistic model that can be used to represent the joint distribution of multiple variables. In this case, the variables are the semantic labels assigned to the input text and the emotions expressed in the text. The model is "separable" because it allows the authors to model the relationship between the semantic labels and the emotions separately.

To train the separable mixture model, the authors first create a training dataset consisting of text samples annotated with the specific emotions expressed in the text. They then use a clustering algorithm to group similar semantic labels together, which reduces the number of variables in the model and makes it more computationally efficient. Finally, they use a maximum likelihood estimation algorithm to fit the model to the training data.

The experimental results in the paper were divided into four experiments:

Experiment 1 aimed to determine the best parameter combination for the entire system. The results showed that the optimal parameter combination for the proposed approach was a combination of semantic labels and separable mixture models.

Experiment 2 evaluated the accuracy of emotion recognition using the proposed approach. The experimental results showed that the proposed approach outperformed other approaches, such as the maximum entropy approach and the SVM approach.

Experiment 3 tested the generalization ability of the proposed approach. The system was ported into another domain, which was a dialogue system for customer service. The experimental results showed that the proposed approach maintained high accuracy in recognizing emotions in a different domain.

Experiment 4 compared the proposed approach with other approaches in the literature. The experimental results showed that the proposed approach achieved higher accuracy in recognizing emotions than other approaches.

Two dialogue corpora were collected for the experiments. Corpus A consisted of 26 volunteer college students who recorded dialogs about their daily life. The data was collected over one month, and the students were asked questions about their daily life. They were divided into two groups, and the emotions of the students in Group 1 were recorded for 40 dialogs for each emotion, while for the remaining 20 students, 10 dialogs were recorded for each emotion.

Corpus B contained the dialogs from a broadcast drama. The emotions of the performers in the drama were manually tagged with one of the three emotional states: happy, unhappy, and neutral. The dialogs of the protagonists were preserved in the corpus, and sentences with three words or less were discarded.

In both corpora, the sentences were annotated with one of the three emotional states, and sentences that did not belong to these three emotional states were deleted from the corpus. The experimental results showed that the proposed approach achieved high accuracy in recognizing emotions in both corpora.

Overall, the experimental results showed that the proposed approach, which used semantic labels and separable mixture models, achieved high accuracy in recognizing emotions in text, even when applied to a different domain. The results also showed that the proposed approach outperformed other approaches in the literature.

* 1. **Emotion Recognition From Textual Input Using An Emotional Semantic Network**

The paper focuses on the problem of recognizing emotions from textual input, which is an important task in natural language processing. Emotion recognition is challenging because emotions are often expressed implicitly in text, and it can be difficult to capture the relationships between words and emotions.

To address this problem, the authors propose an approach that involves constructing an emotional semantic network from a large corpus of text. The network maps words to emotions and captures the relationships between emotions. The authors then use the network to classify text into different emotion categories.

Diagram

Description automatically generated

The authors begin by collecting a large corpus of Chinese texts and extracting the emotional content of the texts using a set of emotion keywords. They then use this data to construct an emotional semantic network. The network consists of nodes that represent words and emotions. They are basically the emotion carrier. There are 5 components in a node: Word Name, trigger value acceptor, initial emotion vector, emotion propagation value, emotion vector. And edges that represent the relationships between them. There are 4 types of edges: subject to object, subject to indirect object, bi-directional link between subject and complement, and equal link between subject and object. The authors use a clustering algorithm to group emotions together and to capture the relationships between them.

Once the emotional semantic network is constructed, the authors use it to classify text into different emotion categories. The classification process involves mapping words in the text to nodes in the network and propagating the emotions through the network to obtain a final classification. The authors use a probabilistic approach to propagate the emotions through the network, taking into account the strength of the relationships between the emotions. The propagation of emotions is triggered by a new input sentence and ends with the emotion propagation values below a threshold.

If concept C is defined by *k* definitions (D1 to Dk) with *k* relationships (r1 to rk), shown as:

Text, letter

Description automatically generated

The decay function is defined as:

Text, letter

Description automatically generated

The authors evaluated their approach on a dataset on a broadcast drama that consisted of lead male and lead female dialogues. There were 558 sentences contained in 137 dialogues from the leading man and 453 sentences contained in 136 dialogues from the leading woman. According to the experimental results, using this proposed system, then emotion recognition of textual input achieved an accuracy of about 60%.

Table

Description automatically generated

The authors also discuss the potential applications of their approach in various domains, such as in sentiment analysis, customer feedback analysis, and opinion mining. They note that their approach could be applied to a wide range of text-based applications where emotion recognition is important.

In conclusion, the paper presents an approach for recognizing emotions from textual input using an emotional semantic network. The approach involves constructing a network that maps words to emotions and captures the relationships between emotions, and using the network to classify text into different emotion categories. The authors evaluate their approach on a dataset of Chinese texts and report promising results, and discuss the potential applications of their approach in various domains.

* 1. **Using Roget’s Thesaurus for Fine-grained Emotion Recognition**

The paper proposes a method for fine-grained emotion recognition in text using Roget's Thesaurus. Emotion recognition is an important task in natural language processing and has many applications, such as sentiment analysis, opinion mining, and dialogue systems. However, existing approaches to emotion recognition typically rely on a limited set of predefined emotion categories, which may not capture the complexity and nuance of human emotions.

To address this limitation, the authors propose a method that leverages Roget's Thesaurus, which contains a comprehensive list of words and their semantic relationships. Specifically, the authors use a set of 20 emotion categories derived from Roget's Thesaurus, which cover a wide range of emotions such as joy, anger, fear, and sadness.

Roget's Thesaurus is a comprehensive list of English words and their semantic relationships. The authors use Roget's Thesaurus to define a set of 20 emotion categories, each of which is associated with a set of words that are indicative of that emotion. For example, the emotion category "joy" is associated with words such as "happy", "pleased", and "delighted", while the emotion category "anger" is associated with words such as "mad", "upset", and "irritated". By using Roget's Thesaurus to define these emotion categories, the authors are able to capture a wide range of emotions and their semantic relationships.

Table

Description automatically generated

The authors first annotate a dataset of sentences with the 20 emotion categories using crowd-sourcing. They then extract lexical features from each sentence, such as unigrams, bigrams, and part-of-speech tags, and use these features to train a machine learning classifier. The classifier predicts the most likely emotion category for a given sentence based on its lexical features.

They used a corpus of blog sentences annotated with emotion labels for their experiments. They trained classifiers with unigram features for each emotion class using Support Vector Machine (SVM) for predicting the emotion category of the sentences in their corpus. SVM has previously given good performance in sentiment classification experiments.

In their experiments, they used ten-fold cross-validation experiments conducted using the SMO implementation of SVM in Weka. In each experiment, they represented a sentence by a vector indicating the number of times each feature occurs. They conducted three experiments: (1) using only corpus-based unigram features, (2) using features from the emotion lexicon acquired from Roget’s Thesaurus (RT), and (3) combining corpus-based unigrams with RT features and features from WordNet-Affect.

The results showed that the combination of corpus-based unigram features and features derived from emotion lexicons contributed to improved performance, better than given by any one type of feature group alone. They found that the correlation between the features and the predicted class was highest for the "happy" class, indicated by a precision of 0.813 and recall of 0.698, the highest among all classes. This suggests that it is easier to discern happiness in text than Ekman's other basic emotions.

Table

Description automatically generated

In conclusion, the authors were able to demonstrate that a combination of corpus-based unigram features and features derived from emotion lexicons can help automatically distinguish basic emotion categories in written text. The effectiveness of their emotion lexicon derived from Roget's Thesaurus based on semantic relatedness of words to a set of basic emotion words for each emotion category was also demonstrated.

Code results:

Text, letter

Description automatically generated

Text

Description automatically generated

Text, letter

Description automatically generated

**3. Summary of Papers between 2014 & 2010**

**3.1 Unsupervised Emotion Detection from Text Using Semantic and Syntactic Relations**

The paper proposes a novel unsupervised context-based emotion detection method that does not rely on any affect dictionaries or annotated training data, unlike the traditional methods that depend on supervised learning methods that has a large set of annotated data. But the availability of such large annotated data is very rare and a model trained on a particular domain is not easily ported to another domain. The authors work with a small set of representative words which are used to compute an emotion vector of an affect-bearing word by calculating the semantic relatedness score between this word and an emotion concept. After this, fine-tuning is carried out based on the context of the word. This approach shows promising results in the end.

Diagram

Description automatically generated

In this approach, emotion detection is modelled as a classification problem where one or more nominal labels are assigned to a sentence from a pool of target emotion labels. The four main components of this framework include pre-processing, semantic, syntactic, and sentence analysis. The approach starts by extracting the affect words from the sentence, also called as NAVA (Noun Adjective Verb Adverb) words. Also, it is essential to look up the surrounding words in a sentence to get some idea about the emotion expressions of a particular word, for this, a syntactic dependency is employed. Focus is being made on three types of dependencies namely, adjectival complement, adjectival modifier, and negation modifier.

Usually, the emotion of a NAVA word can be defined as a vector whose elements each represent the strength of the affinity of the word for an emotion category. But if the sentence is deprived of an emotion keyword, it’s difficult for a system to identify the emotion, so sematic relatedness is used to compute the emotion vector. It is observed that a meaning of a word can be found by observing it’s usage across a large sample of language. If 2 words co-occur more often then they tend to be semantically related. To calculate this, PMI is used. PMI between two words *x* and *y* is calculated as:

A picture containing text

Description automatically generatedA picture containing chart

Description automatically generated

For computing emotion vector without context, PMI score between a word and the word representing an emotion concept has been used. Then the emotion vector of the dependent word is fine-tuned using context information. Lastly, the emotion vector of a sentence is calculated by aggregating the emotion vectors of all the affect words, and if the emotion vector score is above a threshold, it is classified as that emotion or else neutral.

To evaluate this framework, it was tested against 3 different corpus namely Wikipedia data, Gutenberg corpus, and Wiki-Guten. The authors also found that stemming does improve the accuracy of the emotion detection process and they used stemmed text for further experiments. This approach performs better than the keyword baseline and Alm’s lextag methods and the context-based approach is slightly better than the context-free one. This method also proved significantly better than other methods on the ISEAR datasets with Four Emotions classification and seven emotions classification.

A screenshot of a computer

Description automatically generated with low confidence

Overall, the proposed method does not require annotated data or a detailed affect lexicon. The results show that the approach outperforms other recent unsupervised methods and is comparable to some supervised methods. However, the weakness of the approach is that the semantic relatedness scores depend on the text corpus from which they are derived. The study also found that the context-based approach consistently outperforms the context-free approach. In future research, the authors plan to derive semantic relatedness scores from multiple measures and test other syntactic dependencies.

**3.2 Emotion Recognition from Text Based on Automatically Generated Rules**

In this work, the authors propose a framework for emotion classification in English sentences where emotions are treated as generalized concepts extracted from the sentences. By looking at the syntactic and semantic structure of a sentence, an intermediate emotional data representation is generated which is later generalized to obtain emotion recognition rules (ERR). Then using classification algorithms, these generated ERRs are compared with the training set ERRs to get the emotion labels for the sentences. This paper aims to recognize six emotions as suggested by Ekman namely happiness, sadness, anger, fear, disgust, and surprise. Finally, to test out this approach, it was tested against the Twitter dataset and Aman dataset. In both cases, this approach proved to be better than the existing methods.

In this approach, emotions are treated as concepts extracted from the sentence and the representation of concepts depends on 2 principles; rules of composition are formed after removing the lexical parts of the sentence, and syntactic and semantic analysis must be performed to understand a sentence. The syntactic and semantic analysis is done by constructing the dependency parsing tree of the sentence and then pruning it as per some set rules. The output of this is basically the EER.

The authors describe 2 main phases of their methodology: the offline phase, and the comparison and classification phase. The main role of the offline phase is to process each annotated sentence in the dataset by the Sentence Processor module to generate the corresponding EER.

Diagram

Description automatically generated

The sentence processor determines the POS tag for each word and after tagging, the Stanford dependency parser is applied which gives us the dependency tree. Pruning of this tree is based on 2 rules: Separation Rules (2 rules), and Deletion Rules (3 rules). After pruning the tree, what we are left with is the required EER of the sentence.

Diagram

Description automatically generated

The next phase includes comparison and classification. The goal is to compare the two EERs i.e the input sentence EER and the EER generated from the offline phase. For comparison, a customized KNN algorithm is used. The KNN is customized in the sense that it is based on semantic and keyword similarity. However, if the input is rejected by the KNN classifier, an attempt is made to classify it using the PMI classifier, if it gets rejected by PMI as well, it is classified using PMI-IR. In the end, we will get a similarity score and the EER to which the input EER has the most similarity, the input is classified as per the emotion of the matched EER.

This system follows a 2-tier client-server architecture where the client is responsible for getting input from the user and the server is responsible for processing it and sending back the results to the client. The method is tested on two different datasets. Firstly, on the Aman dataset, the training sentences were of length in the range of 0 to 10. The computation was compared against the baseline approach which classifies based on keywords. It was found that the average F-score of 84% was obtained on all the 6 emotions. For the second experiment consisting of the Twitter dataset, an F-score of 84% was obtained as well for the 5 emotions excluding disgust. A comparison was also made with the Emoheart classifier in both experiments.

Chart, bar chart

Description automatically generated

Overall, the work presented a new approach to classifying emotions from textual data by performing complex syntactic and semantic analysis of the sentence and using ontologies such as Wordnet and ConceptNet. The approach was evaluated on two different datasets and outperformed the state-of-the-art method in emotion classification. The classifier was shown to be context-sensitive and able to generalize the training set for better coverage of emotion rules. The approach also demonstrated that comparing the relations between the words of the sentence could lead to better accuracy than assigning individual emotional rates for each word. The proposed classifier was shown to be flexible and easily extendable to classify any number of emotions by providing a reasonably-sized training set that covers the required emotions.

**3.3 Emotion Detection in suicide notes**

Suicide can be defined as death caused by self-directed injurious behavior with any intent to die, is a major public health concern and a leading cause of death worldwide. Suicide prevention efforts focus on early risk recognition and referral support. However, due to information overload on social media, it’s difficult to continuously monitor them. There is a need for automatic suicidal message detection. Most of the previous works on this is based on manual analysis and detection of important features that can distinguish between a fake note and a genuine one. Also, it is worth noting that suicide-related notes need not contain words related to suicide. The sentiment of such notes can also reside in a combination of words that might not have any suicide-related keywords.

This paper aims at the fine-grained detection of a large set of emotions (15) in suicide notes, where the emotions are detected at the sentence level. In order to do so, a supervised machine learning approach is used and trained an SVM on a manually labelled suicide note corpus.

The dataset used consists of suicide notes written by 1319 people, collected between 1950 and 2011. Each note was annotated manually by at least 3 annotators. The most frequent labels found were, in order, *instructions, hopelessness, love, information,* and *guilt.* The data showed that most emotions were strongly lexicalized i.e. some words are typically associated with them.

The authors adopted a supervised machine learning methodology to predict the presence of one or more emotions in unseen data. The data was preprocessed with MBSP with well defined features as follows: Lemmas, Lemmas + POS, Pruned lemmas + POS tags, Trigrams, WordNet synsets, SentiWordNet information, Subjective clues. To determine the optimal feature combinations, the 7 feature groups were combined into 17 feature sets. The dataset also contained spelling errors which could negatively affect the performance along the way. So, spelling were corrected and experiments were performed on both original and spellchecked data.

Chart, scatter chart

Description automatically generated

For the classification part, 15 binary classifiers were used whose outputs were combined. Since SVM is shown to work well in other NLP tasks, it was used. The results of the classifier are reported as F1-scores that place equal emphasis on both precision and recall. But in our case of finding suicide-related content, some precision can be traded up in favor of recall because a false negative is worse than a false positive. For this, bootstrap resampling has been used. It was also used to determine which threshold maximized F-score for each classifier.

Chart, bar chart

Description automatically generated

The bootstrap resampling experiments were conducted using SVM classifiers with 17 features set on 15 emotions. For emotions with a frequency of more than 40 annotations per 1000 sentences, F-score was above 40% but it was worse for rare emotions as there were fewer training data for them. Emotions such as *forgiveness, thankfulness,* and *love* performed better than would be expected from frequency alone as they were lexicalized more often. Also, it was found that making spelling corrections was marginally beneficial. The 15 best classifiers use only 6 feature sets: trigrams and subjectivity clues (set J, 6classifiers), lemmas and trigrams (set F, 3 classifiers), lemmas, trigrams, and subjectivity clues (set K, 3 classifiers), lemma-POS pairs (set B, 1 classifier), WordNet (set E, 1 classifier) and lemmas, trigrams, WordNet, and subjectivity clues (set O, 1 classifier). The final system output is produced by aggregating the outputs of the best performing classifier for each emotion.

Overall, the paper outlines a machine-learning approach for detecting emotions in suicide notes, using support vector machines and a variety of lexical and semantic features. Results suggest that the most successful features are trigrams and lemma bags-of-words and subjectivity clues, but data sparseness in rare emotions remains a challenge. The paper suggests that deeper semantic analysis and more advanced spelling correction could yield informative features for emotion classification. The detection of emotions in suicide notes has applications in forensic linguistics and suicide prevention, and the paper suggests that future research could use the presence of emotions as features for predicting suicidality in text.