Emotion Detection using Lexical Chains

M.Naveen Kumar
Assistant professor
Department of Computer Science and Engineering
Saveetha Engineering College

R.Suresh
Assistant professor
Department of Information Technology
SAMS College of Engineering and Technology

ABSTRACT
Emotion detection has become an indispensable task in Natural Language Processing (NLP) in recent times. Emotion detection has been done using various knowledge based and corpus based methods. In this paper, we propose a methodology for emotion detection using lexical chains. Lexical chains easily identify the coherent concepts present in the text and so the emotion detection is also done with much ease. We have handled six types of emotions namely happiness, sadness, anger, fear, disgust and surprise. We have evaluated our method with the existing methods and we have shown better performance.

General Terms
Emotion Detection, Sentiment Analysis

Keywords
Lexical chains, Emotion Detection, Sentiment Analysis

1. INTRODUCTION
The massive increasing documents on the web calls for an efficient text processing and representation technique. Text classification becomes one of the essential task in Natural Language Processing (NLP) arena. Text classification involves categorizing the texts with pre defined domains such as sports, art, politics, stories etc. Apart from classifying texts domain wise, they can also be classified based on the emotion they exhibit as whatever may be the domain, emotion is an omnipresent element in a text. Eliciting the emotion can help in various NLP applications such as Image/Video Retrieval, Summary Generation and Question Answering (QA) systems. Identifying the core emotion of the text will reveal the theme of the text. [1].

There has been many works done on emotion labeling Emotions has been identified and rated from the blogs by Gill et al [2]. They have used experts to identify the emotions using 200 sample texts. Manually identifying emotions may yield accurate results but it becomes a daunting task when the size of the data increase

. Carlo Strapparava et al have identified six emotions namely Anger, Joy, Surprise, Disgust and Fear manually using supervised and unsupervised methods[3]. The valence of the text is also identified (i.e) whether the text expresses more positive or negative emotion has also been done. The human annotators have used resources like WordNet for emotion labeling. They have used News articles as the data set. When it comes to emotion labeling for huge number of documents, this method may not be efficient to follow as manual annotation becomes a laborious task for large data set.

Yang et al have labeled four emotions using words as features [4. They have extracted features at sentence level and have applied to the documents using Support Vector Machines (SVM) and Conditional Random Field (CRF). Using machine learning approach for emotion labeling may be efficient to handle huge number of documents but it highly depends on the feature set that is supplied to the machine learning algorithm. The limitations in the feature may reflect in the system performance.

Alm et al have come up with a supervised learning called SNoW learning to annotate the various types of emotions prevalent in children fairly tales[5]. They have used fourteen features and have shown good results. This method has been tested with only twenty two stories and it also suffers from feature set limitations.

Recently, Bao et al have proposed an emotion detection methodlog that identifies emotions using co- occurrence of emotion centric words [6]. The methodology uses Latent Dirichlet Allocation (LDA) to build the topic model. Using frequently co-occuring terms to identify emotions does yield better performance but the underlying semantics might get deviated if the co-occurring words convey a different context.

We propose a simple but an efficient emotion detection algorithm that detects six emotions using lexical chains.

Lexical chains are predominantly used in identifying the coherent pieces in the text [7]. A lexical chain represents a string of concepts that are semantically bonded throughout the text. We make use of these lexical chains to identify the topic emotion of the text. Unlike the existing approaches which suffer from limitations in feature set, the method proposed is robust towards these limitations and it is scalable and can handle any number of documents.

The rest of the paper is organized as follows. Section 2 describes the proposed emotion detection approach. Section 3 discusses about the evaluation of the proposed approach and Section 4 conveys the conclusion and future work.

2. Proposed work
Figure 1 shows the flow diagram of the emotion detection algorithm proposed in this paper. It can be observed from Figure 1 that we have used WordNet as the lexical data base and an Emotion key word list as the knowledge base. Given a text document, Lexical chains are formed for each emotion key word present in the text. The lexical chains are scored and the strongest chains are identified from which the emotion which is inherent in the text is identified. This process is repeated for all the input corpora . The details of each process is described in the next section

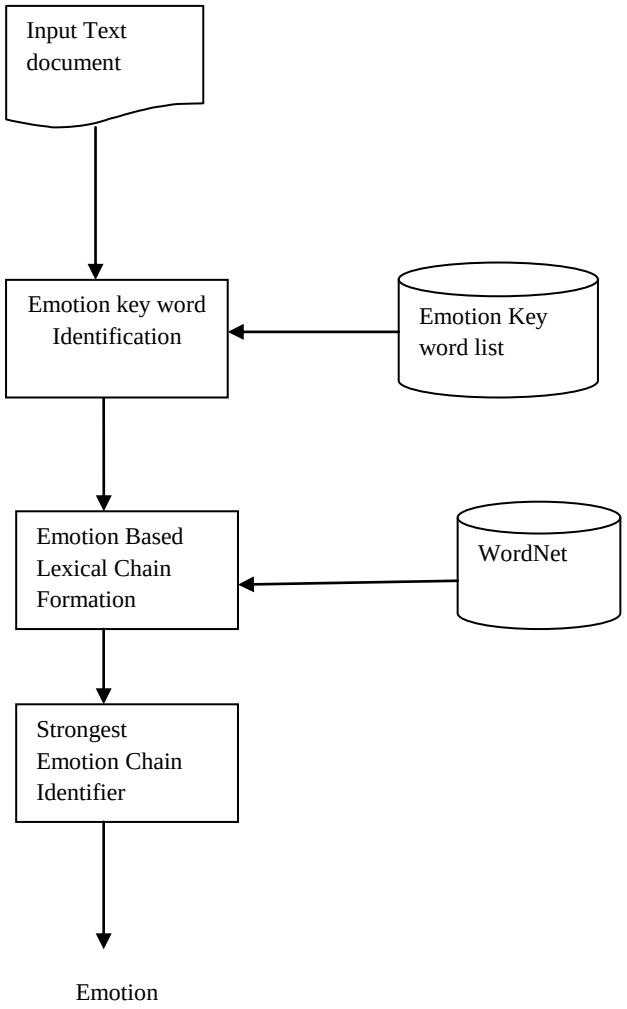

Fig 1: Flow Diagram of the Proposed Emotion Detection System

2.1 Lexical Chain Formation

As discussed previously, lexical chains are formed for each emotion keyword word in the text which becomes the candidate word. The semantic bond may be the mere repetition of the word for which the lexical chain is built, a synonym or it may be a concept related through semantic relations such as meronym, hyponym, holonym etc. For instance, the word, "smile" and "laughter" may be found in the same lexical chain as the word, "smile" is related to the word "laughter" through the semantic relation, "synonym". These semantic relations can be identified with the help of lexical resources such as ontology or Wordnet. We have used WordNet as the lexical resource. Let the lexical chain formed for the word, "smile" be denoted as $Lex_{smile.}$. The text window for choosing and adding the words in the lexical chain $Lex_{smile.}$ depends on the semantic relationship between the words to be added and the candidate word under consideration, "smile".

Mere repetition of the candidate word can be searched till the last sentence of the text and added in the lexical chain. For instance, the text window for the word, "smile" (other than the candidate word) can be the full text document. Whereas, the text window for the other words which are related through the semantic relations are usually fixed to three sentences from the candidate word. For instance, if the word, "smile" occurs in the first sentence, then the semantically related words can be searched till the fourth sentence and inserted in the lexical chain $Lex_{smile.}$

As discussed above, the lexical chains are formed for all the candidate emotion word in the text document. Each lexical chain is scored depending on the length of the chain and the semantic relationships that exist between the chaining words and the candidate words for which the chain is built. This is discussed in the next section.

2.2 Scoring Lexical Chains

The lexical chains built for a text document are scored depending upon the various factors namely the length of the chain and Homogeneity index of each member present in the chain.[7]. The length of the chain and the homogeneity index is calculated as follows.

Length of a lexical chain= Number of occurrences of each member of the chain.

Homogeneity index of a lexical chain=1- number of distinct occurrence of each member of the chain/length of the chain.

Score of a lexical chain=Length* Homogeneity index.

First strong chains that satisfy the strength criterion are chosen.

Strength Criterion= Average score of all chains+2*standard deviation of the scores of all chains.

Strong chains are those which exceed the strength criterion. The candidate emotion word of each strong chain depict the emotions that are exhibited by the text. Instead of choosing the strong chain that has the highest score, we analyse the text window covered by the members of the chain. The emotion that is dominant in the text will be prevalent right from beginning and till the end of the text. We identify this by tagging the sentence identifiers of each member of the chain while adding them to the chain. The lexical chain that encompasses a wide text window becomes the dominant chain and the candidate emotion word of the chain becomes the topic emotion of the text.

3. Evaluation

We have used 2000 expository news articles posted online from May 2006 to till date collected from http://en.wikinews.org as the data set. We have compared our approach with the existing approach described in section1. The Table 1 shows the statistics of the proposed emotion detection approach.

Table 1. Statistics of the proposed approach

Emotion	Number of Documents for which emotion was identified
Joy	264
Anger	405
Fear	106
Disgust	280
Surprise	254
Sadness	350

It can be observed from the Table 1 that, out of 2000 news articles, emotion detection has been done for only 1660 articles. This is due to various reasons such as missing emotion centric words in our lexical data base and the corpus datum which is not sufficient enough for our algorithm to form strong chains. We have used the publically available Princeton WordNet 2.1 for our emotion detection task. The emotion keyword list we have consists of 1000 emotion keywords which are manually collected from the web . Table 2 shows some of the emotions centric words found in the frequently occurred strong emotion based lexical chains for each emotion.

Table 2. Frequently observed Emotion centric keywords

Emotion	Emotion centric keywords
Joy	Happiness, bliss, beatitude, joy, joyous, happy, joyful, elated,
Anger	Enraged, annoyed, hateful, furious, irritated, fierce, choleric
Fear	Trembling, tremor, scare, qualm, terror, trepidation
Disgust	Revulsion. Hatred, abhorrence, nausea, satiation
Surprise	Astonishment, bewilderment, miracle, fortune, amazement
Sadness	Grief, misery, melancholy, distress, dejection, anguish

We have also done a comparison with our approach with the existing approach proposed by Yang et al[5]. that are described in section1. We have used our wiki news corpus and have implemented Yang et al method of emotion detection for comparison purpose. We collected another 2000 news articles apart from the datum collected to test our approach and used this corpora as the training data. The corpora of 2000 news articles used in our approach has been used as the test data. As mentioned previously, in the existing appraoch, they have used SVM and CRF classifiers to find emotions. We have used the same LIBSVM4 and MALLET CRF tool kit used by Yang et al. Features are typically key words that are extracted from each sentence. We made use of the lexical chains to extract key words instead of manually extract the emotion based features. Table 3 shows the statistics obtained using the existing approach and Table 4 shows the performance comparison. Figure 2 shows the graphical representation of the comparison. The graph shows the precision and recall values in percentages.

Table 3. Statistics of the Yang et al approach

Emotion	Number of Documents for which emotion was identified
Joy	253
Anger	350
Fear	100
Disgust	250
Surprise	212
Sadness	421

Table 4. Comparison between the proposed and the existing methods

Method	Precision	Recall	F measure (%)
Proposed	0.96	0.8	86
Yang et al	0.81	0.65	71.2

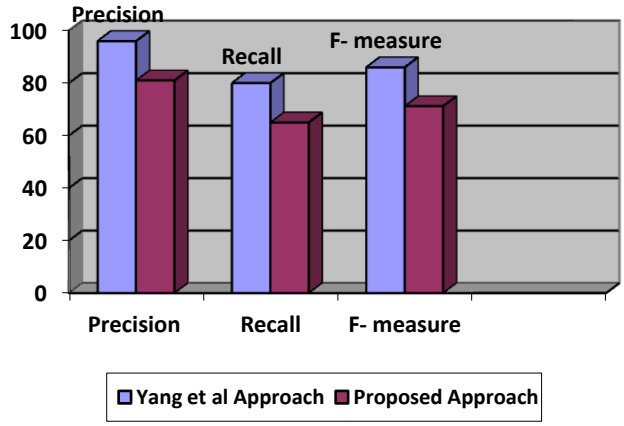

Fig 2: Graph showing Performance Comparsio

It can be observed from the statistics discussed in this section that the existing approach by Yang et al shows a lower performance compared to our method. As discussed in the section, this is mainly due to the poor scalability of the machine learning approaches in encompassing the features. Though they are capable of handling small amount of data efficiently but there occurs a fall back in performance while handling huge number of text documents in one shot. The next section discusses about the conclusions and about the future work.

4. Conclusion and Future work

We have proposed an emotion detection approach using lexical chains. Lexical chains find the coherent terms that are hidden in the text. This concept is applied to find the hidden emotion of the text. Lexical chains can robustly handle large set documents unlike machine learning approaches which suffer from insufficient feature sets. Though the proposed approach is mainly built for emotion detection it can also be

extended for various other NLP tasks such Information Extraction (IE), IR and summary generation as lexical chains go well with these applications.

More number of emotion key words need to be added to the emotion key word dictionary in order to increase the recall. Also we are currently working on exploring more emotions other than the emotions that are described in this paper. Analyzing texts of various genres such as stories, dialogues using the proposed approach is also in progress.

5. ACKNOWLEDGMENTS

Our thanks to the experts who have contributed towards development of the template.

6. REFERENCES

[1] Andrea Esuli , Fabrizio Sebastiani. 2006. SENTIWORDNET- A publically available open source for opinion mining In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06.

[2] Alastair J. Gill, Darren Gergle, Robert M. French, Jon Oberlander 2008. Emotion Rating from Short Blog Texts.In Proceedings of CHI 2008, April 5–10, 2008, Florence, Italy.

[3] Carlo Strapparava, Rada . 2008. Learning to Identify Emotions in Text. In *SAC'08* March 1620, , Fortaleza, Cear´a, Brazil.

[4] Cecilia Ovesdotter Alm, Dan Roth. Richard Sproat. 2005. Emotions from text: machine learning for text-based emotion prediction. In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing.

[5] Changhua Yang Kevin Hsin-Yih Lin Hsin-Hsi Chen. 2007. Emotion Classification Using WEB Blog Corpora. WI '07 Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence.

[6] Shenghua Bao, Shengliang Xu, Li Zhang, Rong Yan, Zhong Su, Dingyi Han, and Yong Yu. 2012. Mining Social Emotions from Affective Text. In IEEE Transactions on Knowledge and Data Engineering, VOL. 24, NO. 9, SEPTEMBER 2012

[7] Regina Barzilay and Michael Elhadad. 1997. Using Lexical Chains for Text Summarization. In In Proceedings of the ACL Workshop on Intelligent Scalable Text Summarization