Approaches towards Emotion Extraction from Text

Nilesh M. Shelke
Assistant Professor, Department of IT, Priyadarshini Indira Gandhi College of Engineering, Nagpur, M.S., India

Shriniwas Deshpande, Ph.D
Associate Professor and Head in P. G. Department of Computer Science and Technology, DCPE, HVPM, Amravati. (MS) India.

Vilas Thakre, Ph.D
Professor and Head in P.G. Department of Computer Science and Engg., SGB Amravati University, Amravati.

ABSTRACT

With the growth of internet community, many different text-based documents are produced. This paper presents an overview of the emerging field of emotion detection from text and describes the current generation of detection methods of emotions from the text. Emotion recognition in text is just one of the several dimensions of the task of making the computers make sense of emotions. In this study the main research focus will be on suggestions for designing more efficient and adaptive Natural Language Processing System for the detection of various emotions (sentiment analysis) on the basis of study of important recent techniques.

Keywords

Sentiment, sentiment score, polarity, valence, semantic, ontology.

1. INTRODUCTION

Emotions are fundamental to human lives and decision-making. Understanding and expression of emotional feeling between people forms an intricate web. This complex interactional phenomena, is a hot topic for research. “Emotion” is a large topic area with varied use of terminology and many concepts, theories and models associated with it. The term “emotion extraction /emotion computing /affective computing/ Sentiment Analysis” is now commonly used in the human-computer interaction (HCI) field to embrace systems that address human emotion in some way and this has become a major topic of interest in the community over the last fifteen years. Emotion Detection in text documents is essentially a content – based classification problem involving concepts from the domains of Natural Language Processing as well as Machine Learning. In this paper emotion recognition based on textual data and the techniques used in emotion detection are discussed.

Emotion is expressed as joy, sadness, anger, surprise, hate, fear and so on. In 2001, W. Gerrod Parrot [1] wrote a book named “Emotions in Social Psychology”, in which he explained the emotion system and formally classified the human emotions through an emotion hierarchy in six classes at primary level which are Love, Joy, Anger, Sadness, Fear and Surprise as shown in fig. 1. Certain other words also fall in secondary and tertiary levels. People are able to perfectly distinguish the expressed emotions because they understand the meaning of the words and phrases. They also are able to generate expressions and sentences for different emotions.

However, developing a computer system to analyze and interpret different emotions in a given text is a difficult task.

2. LITERATURE SURVEY

A great body of work exists in the field of emotion extraction. The work done in this area includes distinguishing subjective portions in text, finding sentiment orientation and, in few cases, determining fine-grained distinctions in sentiment, such as emotion and appraisal types. Work exclusively on emotion detection is comparatively rare and lacks empirical evaluation.

Emotions are mental states accompanied by physiological changes. Ekman identified six basic emotions: happiness, sadness, anger, fear, disgust and surprise [2].

To characterize emotional interactions in social networks, and then using these characteristics to distinguish friends from acquaintances was proposed in [3]. The goal was to extract the emotional content of texts in online social networks. The interest is in whether the text is an expression of the writer’s emotions or not. For this purpose, text mining techniques are performed on comments retrieved from a social network.

A model for statistical analysis of collective emotions for product reviews was proposed in [4]. This research work is limited to the extraction of emotional parameters like helpfulness, unhelpfulness and rating. This work was for providing the guidelines to manufacturers on how to increase customer satisfaction

The prevalent approach to sentiment classification is based on the premise that the overall sentiment of a document is the aggregate of the sentiment of the words comprising it. These techniques therefore look for the presence of appropriate affect words in text. Some words are quite unambiguously affect words, while others convey affect to some degree. This method either uses a corpus-driven approach to assign affective orientation or scores to words, or it relies on some existing affect lexicons.

Many researchers have become interested in sentiment analysis, as more people learn of the scientific challenges posed, and the scope of new applications enabled, by the processing of subjective language. The papers studied in [5] are a relatively early representative sample of research in the area.

A lexicon model for subjectivity description of Dutch verbs that offers a framework for the development of sentiment analysis and opinion mining applications based on a deep syntactic-semantic approach was presented [6,7]. The model aims to describe the detailed subjectivity relations that exist between the participants of the verbs, expressing multiple attitudes for each verb sense.

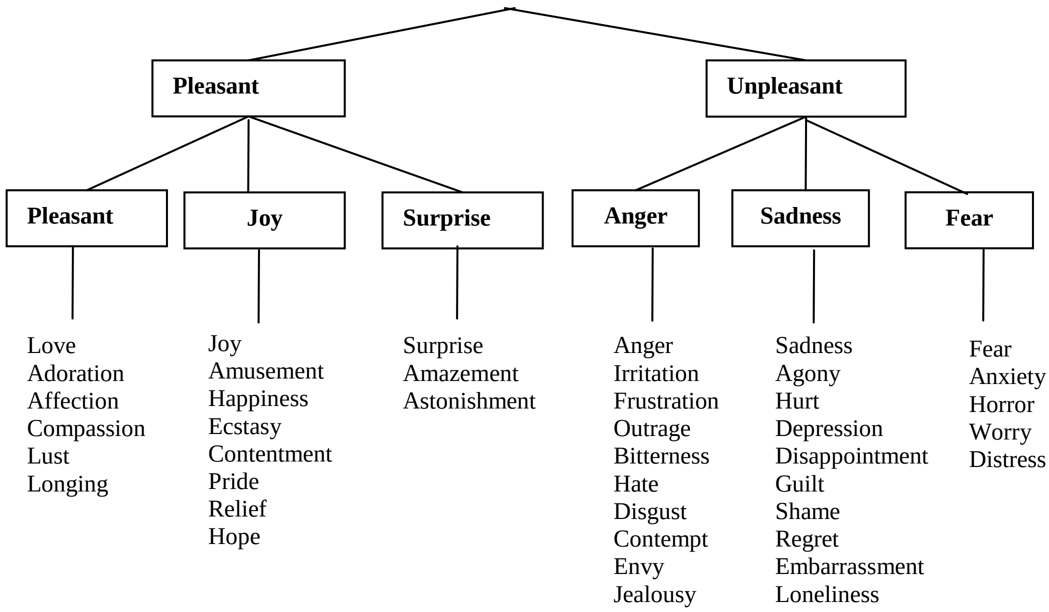

Fig. 1 Hierarchical Structure of Emotions

Table 1. Basic Emotions as classified by different researcher

Lists of Basic Emotions	
Ekman	anger, disgust, fear, joy, sadness, and surprise
Izard	anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, and surprise
Plutchik	anger, anticipation, disgust, fear, joy, sadness, surprise, and trust

Table 2. Most frequent Emotion Indicators in the data

Emotion	**Words**
Anger	greedy, hatred, hostility, envy, jealousy,hateful, angry, horror , disgust , sinful , noisome
Disgust	obscene , detestable, repulsive, wicked, sicken, dishonest, amoral, horror , disgust , sinful , noisome
Fear	fearful , awful , horrible, hysterical , cruel, diffidence, panic, horrific, fear, scare, anxious, terror, upset
Joy	good, amorous, great,fascinating, celebrate, comforting, glad , romantic, pleased, happy , satisfaction, comforting, merry
Sadness	glum, bad, grim, misery, dismay, poor, regret, dismal, sorry, guilty, mourning, sadness rueful ,glooming,dark, tearful, oppression
Surprise	wonderful , astonish, wonderment, amaze, surprising, terrific, fantastic admiration

Polarity scores to a large list of words, based on their WordNet distance from positive and negative seed words was assigned in [8].

Deep language analysis techniques for machine translation to extract sentiment scores for words in documents were used in [9].

The sentiment of customer reviews was determined by counting positive and negative terms, taking into account contextual valence shifters, such as negations and intensifiers [10, 11]. A computable metric of positive or negative polarity in financial news text was proposed in [12].

Research work on determining whether opinions expressed on different topics in a conditional sentence are positive, negative or neutral. Conditional sentences are one of the commonly used language constructs in text. In a typical document, there are around 8% of such sentences. This research work presents a linguistic analysis of such sentences, and then builds some supervised learning models to determine if sentiments expressed on different topics in a conditional sentence are positive, negative or neutral. Experimental results on conditional sentences from 5 diverse domains are given to demonstrate the effectiveness of the proposed approach [13].

As opinion mining and sentiment analysis applications tend to utilize more and more the composition of sentences use the value and properties of the verbs expressed by its dependency trees, there is a need for specialized lexicons where this information can be found. For the analysis of more complex opinionated text like news, political documents, and (online) debates the identification of the attitude holder and topic are of crucial importance. Applications that exploit the relations between the verb meaning and its arguments can better determine sentiment at sentence level and trace emotions and opinions to their holders [14].

Corpus-based methods use a corpus of documents that are labeled with polarity to train a sentiment classifier. Various classification models and linguistic features were introduced to improve the classification performance. Structured Model was explored for jointly classifying the sentiment of text at different levels of granularity. Research carried out is related to the Effect of Negation on Sentiment Analysis and Retrieval Effectiveness [15, 16, 17].

Following is an example of dynamic progression of emotion across sentences

1. Today was just kind of stressful. (*sadness*)
2. My night last night turned out amazing towards 8:00-9:00. (*happiness*)
3. I had a really nice conversation with Jack*, another amazing one with Pat*, and then another one with Sid*. (*happiness*)
4. It was definitely a good night of conversation, although I did stay up until 12:30, and I payed for that this morning in 1st period (Physics) (*mixed emotion*)
5. But all of a sudden today it's hit me that I have all this work due. (*surprise*).

3. PROPOSED CLASSIFICATION OF EMOTION DETECTION

Emotion Detection methods can be classified as follows:

1. Keyword-based detection: Emotions are detected based on the related set(s) of keywords found in the input text;

2. Learning-based detection: Emotions are detected based on previous training result with respect to specific statistic learning methods;

3. Hybrid detection: Emotions are detected based on the combination of detected keyword, learned patterns, and other supplementary information; Besides these emotion detection methods that infer emotions at sentence level, there has been work done also on detection from paragraphs or articles[18]. For example, though each sentence in a blog article may indicate different emotions, the article as a whole may tend to indicate specific ones, as the overall syntactic and semantic data could strengthen particular emotion(s).

3.1. Keyword-based Methods

Keyword based approaches use synonyms and antonyms in WordNet to determine word sentiments based on a set of seed opinion words.

In [19] a bootstrapping approach is proposed, which uses a small set of given seed opinion words to find their synonyms and antonyms in WordNet (wordnet.princeton.edu) to predict the semantic orientation of adjectives. In WordNet, adjectives are organized into bipolar clusters and share the same orientation of their synonyms and opposite orientation of their antonyms. To assign orientation of an adjective, the synset of the given adjective and the antonym set are searched. If a synonym/antonym has known orientation, then the orientation of the given adjective could be set correspondingly. As the synset of an adjective always contains a sense that links it to the head synset, the search range is rather large. Given enough seed adjectives with known orientations, the orientations of all the adjective words can be predicted [20].

As was observed in [21], keyword-based emotion detection methods have three limitations described below.

1) Ambiguity in Keyword Definitions

Though using emotion keywords is a straightforward way to detect associated emotions, the meanings of keywords could be multiple and vague. Except those words standing for emotion labels themselves, most words could change their meanings according to different usages and contexts. Moreover, even the minimum set of emotion labels (without all their synonyms) could have different emotions in some extreme cases such as ironic or cynical sentences.

2) Incapability of Recognizing Sentences without Keywords

Keyword-based approach is totally based on the set of emotion keywords. Therefore, sentences without any keywords would imply they do not contain any emotions at all, which is obviously wrong. For example, "I passed my qualify exam today" and "Hooray! I passed my qualify exam today" should imply the same emotion (joy), but the former without "hooray" could remain undetected if "hooray" is the only keyword to detect this emotion.

3) Lack of Linguistic Information

Syntax structures and semantics also have influences on expressed emotions. For example, "I laughed at him" and "He laughed at me" would suggest different emotions from the first person's perspective. As a result, ignoring linguistic information also poses a problem to keyword-based methods. In summary, keyword-based methods should also detect not only the existence of keywords, but also their linguistic information to detect emotions more accurately.

3.2 Learning-based Methods

Researchers using learning-based methods attempt to formulate the problem differently. The original problem that determining emotions from input texts has become how to classify the input texts into different emotions. Unlike keyword-based detection methods, learning-based methods try to detect emotions based on a previously trained classifier, which apply various theories of machine learning such as support vector machines [22] and conditional random fields [23], to determine which emotion category should the input text belongs.

However, comparing the satisfactory results in multimodal emotion detection, the results of detection from texts drop considerably. The reasons are addressed below:

1) Difficulties in Determining Emotion Indicators

The first problem is, though learning-based methods can automatically determine the probabilities between features and emotions, learning-based methods still need keywords, but just in the form of features. The most intuitive features may be emoticons [], which can be seen as author's emotion annotations in the texts. The cascading problems would be the same as those in keyword-based methods.

2) Over-simplified Emotion Categories

Nevertheless, lacking of efficient features other than emotion keywords, most learning-based methods can only classify sentences into two categories, which are positive and negative. Although the number of emotion labels depends on the emotion model applied, we would expect to refine more categories in practical systems.

3.3 Hybrid Methods

Since keyword-based methods with thesaurus and naïve learning-based methods could not acquire satisfactory results, some systems use a hybrid approach by combining both or adding different components, which help to improve accuracy and refine the categories. The most significant hybrid system so far is the work of Wu, Chuang and Lin [24], which utilizes a rule-based approach to extract semantics related to specific emotions, and Chinese lexicon ontology to extract attributes. These semantics and attributes are then associated with emotions in the form of emotion association rules. As a result, these emotion association rules, replacing original emotion keywords, serve as the training features of their learning module based on separable mixture models. Their method outperforms previous approaches, but categories of emotions are still limited.

4. MOTIVATIONS AND PRACTICAL APPLICATIONS

The motivation for this work has come from the recent growing interest in the sentiment analysis field. The rapid growth of the World Wide Web has facilitated increased online communication and opened up newer avenues for the general public to post their opinions online. This has led to generation of large amounts of online content rich in user opinions, sentiments, emotions and evaluations. We need computational approaches to successfully analyze this online content, recognize and aggregate relevant information and draw useful conclusions.

Among the less explored sentiment areas is the recognition of types of emotions and their intensity. Recognizing emotions conveyed by a text can provide an insight into the author's intent and sentiment, and can lead to better understanding of the text's content.

Some of the practical applications based on proposed research work are:

4.1 Applications as a Sub-Component Technology

Emotion extraction engine helps in augmentation to *recommendation systems*, since it might behoove such a system not to recommend items that receive a lot of negative feedback.

Detection of "flames" (overly-heated or antagonistic language) in email or other types of communication is another possible use of subjectivity detection and classification.

Additionally, there are potentially relations to citation analysis, where, for example, one might wish to determine whether an author is citing a piece of work as supporting evidence or as research that he or she dismisses.

4.2 Applications in Business and Government Intelligence

The field of opinion mining and sentiment analysis is well-suited to various types of intelligence applications. These opinions greatly impact on customers to make their choices regarding online shopping, choosing events, products and entities. Indeed, business intelligence seems to be one of the main factors behind corporate interest in the field.

4.3 Applications across Different Domains

As is well known, opinions matter a great deal in politics, proposed work has high utility on understanding what voters are thinking, whereas other projects have as a long term goal the clarification of politicians' positions, such as what public figures support or oppose, to enhance the quality of information that voters have access to.

4.4 Applications to Mobile Phones

Today's mobile phones represent a rich and powerful computing platform, given their sensing, processing and communication capabilities. Application of the proposed work has high utility for developing an exceptionally suitable tool for conducting social and psychological experiments in an unobtrusive way on the mobile phones.

5. CONCLUSION

Directions to improve the capabilities of current methods of text-based emotion detection are proposed in this paper.

Attitude or emotional information exists in text documents in addition to the informatory and/or explanatory content. Classification of documents based on this piece of information can serve important functions in more than one applied areas, (like in TTS applications). However, emotional classification is very different from simple topic classification, because it is not merely comparison of the document content against a set of 'keywords' or 'keyphrases'.

Following are some of the issues in emotional classification of text documents that make this problem quit a challenging one:

Emotion builds up through the sentences of textual content- There are no specific set of words which can completely define a particular emotion. Given a set of words, more than one emotional feeling can be implicit from those 'happy' and 'excited' may convey the feeling of 'happiness' as well that of 'surprise'. Many times, a certain set of words may not be having any emotional information, but when placed in the sentence and the context of the document, the underlying 'emotion' becomes clear, Example – 'I don't understand why people tell lies!' There is no emotional word in this sentence;

Presence of ambiguity – a given sentence may convey more than one emotional meaning. For example with the above sentence anger might also be associated.

we have outlined the major approaches for text based emotion detection and shown how syntactic and semantic information can be beneficial for emotion detection. Current methods are lacking in in-depth semantic analysis for detecting hidden phrase patterns and more investigations need to be done to identify, build and incorporate knowledge rich linguistic resources that have a focus on detecting emotions.

The context plays an important role for the emotion detection. Research fraternity will require to stress on contextual based emotion recognition from the text.

REFERENCES

[1] W.G, "Emotions in Social Psychology," in Psychology Press, Philadelphia 2001, p.p. 102-105.

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

[3] Yassine, Hajj (2010), "A Framework for Emotion Mining from Text in Online Social Networks", IEEE International Conference on Data Mining Workshops, Sydney, IEEE publications, Dec 2010, p.p. 1136-1143.

[4] Garcia, Schweitzer, Chair of Systems Design, ETH Zurich, Kreuzplatz, (2011), "Emotions in Product Reviews – Empirics and Models", 2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, Boston, MA, IEEE publications, Oct 2011, p.p. 483-488.

[5] Esuli Baccianella Stefano, Esuli Andrea and Sebas-tiani Fabrizio, (2010), "SentiWordNet 3.0: An Enhanced Lexical Re-source for Sentiment Analysis and Opinion Mining". In Proceedings of the 7th Conference on Language Resources and Evaluation, p.p.2200-2204.

[6].Turney Mohammad, S. and Turney, P.D. (2010), "Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon". Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Los Angeles, ACM publications, June 2010, p.p. 26-34.

[7] Isa Maks, Piek Vossen, (2011), "A Verb Lexicon Model for Deep Sentiment Analysis and Opinion Mining Applications", Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, ACL-

[8] Rup Penhofer, J. M. Ellsworth, M. Petruck, C. Johnson, and J. Scheffzcyk, (2010), Framenet II: "Theory and Practice (e-book)", http://framenet.icsi. berkeley.edu/ book/book.pdf, accessed on date: 5 June 2012.

[9] Soo-Min Kim and Eduard Hovy, (2004), "Determining the Sentiment of Opinions", In Proceedings of the International Conference on Computational Linguistics (COLING), Geneva, ACM Publications, June 2004, p.p. 50-57.

[10] Hiroshi Kanayama and Tetsuya Nasukawa (2006). "Fully Automatic Lexicon Expansion for Domain-Oriented Sentiment Analysis", In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Sydney, ACM publications, July 2006, p.p. 355–363.

[11] Alistair Kennedy and Diana Inkpen. (2006), "Sentiment Classification of Movie Reviews using Contextual Valence Shifters". International Journal of Computational Intelligence, ATLAN TIS Press, Vol. 22(2), Sept 2006, p.p. 110–125.

[12] Devitt et al, (2007), "Sentiment Polarity Identification in Financial News: A Cohesion-based Approach", International Conference on Linguistics, Prague, Czech Repbulic, ACL Publications, June 2007, p.p. 102-107.

[13] Ramanathan Narayanan, Bing Liu, Alok Choudhary (2009), "Sentiment Analysis of Conditional Sentences", Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, Singapore, ACM publications, Sept 2009, p.p. 180–189.

[14] Choi Y. and C. Cardie (2008), "Learning with Compositional Semantics as Structural Inference for Sentiment Analysis". Proceedings of Recent Advances in Natural Language Processing (RANLP), Honolulu, Hawaii, ACM publications, p.p. 793-801.

[15] Jia, L., Yu, C.T., Meng, W. (2009), "The Effect of Negation on Sentiment Analysis and Retrieval Effectiveness", CIKM-2009, Hong Kong, ACM publications, p.p. 18-27.

[16] Bo Pang, Lillian Lee. (2008), "Opinion Mining and Sentiment Analysis" International Conference on Current Trends in IT, CA, USA, Vol.2 (1-2), Now publishers, p.p. 1-135.

[17] W.H. Lin, T.Wilson, J.Wiebe, and A. Hauptmann. (2006). "Which Side Are You on?: Identifying Perspectives at the Document and Sentence Levels". In Proceedings of the Tenth Conference on Computational Natural Language Processing, New York City, USA, ACM publications, June 2006, p.p. 160-174.

[18] L. Devillers, I. Vasilescu, and L. Lamel, "Annotation and detection of emotion in a task-oriented human-human dialog corpus," Proc. ISLE Workshop on Dialogue Tagging for ulti-Modal Human-Computer Interaction, Dec. 2002.

[19] Hu and Liu, 2004, Kim and Hovy, 2004 Hu, M and Liu, B. (2004). *Mining and Summarizing Customer Reviews.* Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD"04).

[20] Lee, D., Jeong, OR and Lee, SG. (2008). *Opinion Mining of Customer Feedback Data on the Web.* Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication, Korea.

[21] C.-H. Wu, Z.-J. Chuang, and Y.-C. Lin, "Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models," ACM Transactions on Asian Language Information Processing (TALIP), Vol. 5, issue 2, Jun. 2006, pp. 165-183,doi:10.1145/1165255.1165259.

[22] Teng, F. Ren, and S. Kuroiwa, "Recognition of Emotion with SVMs," in Lecture Notes of Artificial Intelligence 4114, D.-S. Huang, K. Li, and G. W. Irwin, Eds. Springer, Berlin Heidelberg, 2006, pp. 701-710, doi: 10.1007/11816171_87.

[23] C. Yang, K. H.-Y. Lin, and H.-H. Chen, "Emotion classification using web blog corpora," Proc. IEEE/WIC/ACM International Conference on Web Intelligence. IEEE Computer Society, Nov. 2007, pp. 275-278, doi: 10.1109/WI.2007.50

[24] C.-H. Wu, Z.-J. Chuang, and Y.-C. Lin, "Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models," ACM Transactions on Asian Language Information Processing (TALIP), Vol. 5, issue 2, Jun. 2006, pp. 165-183, doi:10.1145/1165255.1165259.