Emotion And Sentiments Analysis In Social Media Communication (Text)

Sarfaraz Ahmad[1] Md. Tabrez Nafis[2]
[1]M Tech Scholar
[1,2]Department Of Computer Science and Engineering
[1]Al-Falah School Of Engineering and Technology, Dhauj, Faridabad, India, [2]Jamia Hamdard University,New Delhi, India.

Abstract— People when they speak, interact and write convey emotions. The life is an emotion. Emotions play a key role in any kind of decision in affective, social or business area. The emotions are manifested in verbal, facial expressions but also in written texts. Nowadays, with the growth of internet and social media like facebook twitter , many humans interact, in textual manner with each other in large , and exchange ideas, opinions in the web channel or web communities. Emotions in text are very important if we consider that textual interface with the computer is one of the most used. Emotions are treated largely in affective computing that focuses on improving the interaction between user and computer. People express emotions as part of everyday communication. Emotions can be judged by a combination of cues such as facial expressions, prosodies, gestures, and actions. Emotions are also articulated by written texts. Inspired by works in sentiment analysis, this paper explores approaches to automatic detection of emotions in text. I draw from emotion theories in the fields of psychology and linguistics, and use natural language processing and machine learning techniques for automatic emotion detection. In this paper, I describe studies and experiments in manual and automatic recognition of expressions of the six basic emotions (Ekman, 1992) – happiness, sadness, anger, disgust, surprise, and fear – in text form.

***Key words*:** Emotion, Facebook, entiments, facial expression

I. INTRODUCTION

Language is a powerful tool to communicate and convey information. It is also a means to express emotion. Natural Language Processing (NLP) techniques have long been applied to automatically identify the information content in text. Applications such as topic-based text categorization, summarization, question-answering systems, and information retrieval systems typically focus on the information contained in text. This work is an endeavour to apply NLP techniques to identify emotions expressed in text.

In recent years, research inspired by Artificial Intelligence (AI) has focused increasing efforts on developing systems that incorporate emotion. Emotions are crucial to several natural processes that are modeled in AI systems. These include perception, reasoning, learning, and natural language processing. Emotion research is significant for developing affective1 interfaces – ones that can make sense of emotional inputs, provide appropriate emotional responses, and facilitate online communication through animated affective agents. Such interfaces can greatly help improve user experience in Computer-Mediated Communication (CMC) and Human-Computer Interaction (HCI). Emotion research is also vital for text-to-speech (TTS) synpaper systems. Emotion-aware TTS systems can identify emotional nuances in written text and accordingly provide more natural rendering of text in spoken form. Automatic emotion detection and analysis methods are also useful in many applications with psychological basis. For example, they can be successfully applied to learn user preferences and interests from users' personal writings and speeches. These methods are often studied in the scope of the domain of personality modeling and consumer feedback analysis. Similarly, e-learning systems can benefit from affective tutoring approaches. Emotion research has recently attracted renewed attention of the scientific community, as evident from the increased number of events related to it. The *International Conference on Affective Computing and Intelligent Interaction (ACII-2007)*[1]

II. PROBLEM DEFINITION AND SCOPE

Before embarking on the task of emotion recognition, it is imperative to define precisely the goals of this work. I address the task of determining emotions expressed in text at the sentence level. More specifically, the goal is to assign automatically an emotion label to each sentence in the given dataset, indicating the predominant emotion type expressed in the sentence. The possible labels are *happiness*, *sadness*, *anger*, *disgust*, *surprise*, *fear* and *no emotion*. Those are the six basic emotion categories identified by Ekman (1992), and an additional label to account for the absence of a clearly discernible emotion. I also address the task of determining the intensity of the emotion expressed in a sentence. I consider four levels – *high, medium, low*, and *neutral*. The data for this work is drawn from blogs. Blogs mostly comprise unedited, first person narratives related to a variety of interpersonal and public issues, which makes them potentially rich in emotive content. Emotional states have cognitive bases and are shaped by several factors. Emotions manifest themselves in the form of facial expressions as well as linguistic expressions – both verbal and written. The scope of this work is limited to determining the emotional orientation of sentences, as much as is evident from the written text. The criterion for determining what constitutes emotional text and what is the type of emotion expressed is human judgment. It is common experience that just as emotion can be expressed in many ways, the same expression may be interpreted differently by different readers. The techniques introduced in this work were evaluated against data whose emotional orientation was judged the same by at least two humans.

III. Motivation and Context

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(facebook, orkut, twitter) 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. Much of the current work in this direction has typically focused on recognizing the polarity of sentiment (*positive*/*negative*). Among the less explored sentiment areas is the recognition of types of emotions and their intensity – the focus of this work. 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. The inspiration for this work has also come from studies in psychology, which focus on analyzing emotive texts to gain a deeper understanding of the way people express different kinds of emotions. These studies are typically carried out in controlled laboratory settings, or drawn from academic writings or medical domain (Pennebaker et al., 2003). Acquiring this kind of data from the Web – as done in this work – brings attention to a hitherto unexplored source of emotive text.

IV. Literature

An emotion is a mental and physiological state associated with a wide variety of feelings, thoughts, and internal (physical) or external (social) behaviors. Love, hate, courage, fear, joy, sadness, pleasure and disgust can all be described in both psychological and physiological terms. An emotion is a psychological arousal with cognitive aspects that depends on the specific context. According to some researcher, the emotions are cognitive processes. Emotion is a process in which the perception of a certain set of stimuli, follows cognitive assessment which enables people to label and identify a particular emotional state. At this point there will be an emotional physiological, behavioral and expressive response. For example, the primordial fear, that alert us as soon when we hear a sudden noise, allows to react to dangerous situations and provides instantly resources to face them as escape or close the door. The emotional stimuli may be an event, a scene, a face, a poster, an advertising campaign. These events, as a first reaction, put on alert the organism with somatic changes as heart rate, increase of sweat, acceleration of respiratory rhythm, rise of muscle tensions. Emotions give an immediate response that often don't use cognitive processes and conscious elaboration and sometimes they have an effect on cognitive aspects as concentration ability, confusion, loss, alert and so on. This is what is asserted in evaluation theory, in which cognitive appraisal is the true cause of emotions [2]. Two factors that emerge permanently are those related to signals of pleasure and pain and characterizing respectively the positive and negative emotions. It's clear that these two parameters alone are not sufficient to characterize the different emotions. Many authors debate on primary and secondary emotions, other on pure and mixed emotions, leaving the implication that emotions can somehow be composed or added. From the variations, shades, nuances of primary emotions it is possible arise others complex emotions. The basic emotions can be detected from the study of emotional expressions (facial or textual) and from their invariance respect to different individuals and different cultures. Various lists, proposed in many studies, include the following primary emotions: fear, joy, sadness, anger, disgust, surprise. The systems based on the analysis of physiological response as blood pressure, heart rate, respiration change present an initial phase where the signals are collected in configurations to be correlated with different emotional states and a subsequently recognition basing on the measure of indicators. One of the interesting early work on the emotions was that one of Ortony [3]. From this work, through componential analysis, other authors constructed an exhaustive taxonomy on affective lexicon. According to Ortony, stimuli that cause emotional processes are of three basic types: events, agents and objects corresponding to three classes of emotions: satisfied/ unsatisfied (reactions to events), approve/disapprove(reaction to agents), appreciate/unappreciate (reaction to objects).

According to Osgood [4] an emotion consists of a set of stages: stimulus (neural and chemical changes), appraisal and action readiness. Continuing the studies of Charles Darwin, the canadian psychologist Paul Ekman [5] has confirmed that an important feature of basic emotions is that they are universally expressed, by everybody in any place, time and culture, through similar methods.Some facial expressions and the corresponding emotions are not culturally specific but universal and they have a biological origin. Ekman, analyzed how facial expressions respond to each emotion involving the same type of facial muscles and regardless of latitude, culture and ethnicity. This study was supported by experiments conducted with individuals of Papua New Guinea that still live in a primitive way. Damasio [6] affirms that the decisions are choices mainly emotionals. To support this assertion, Damasio shows the cases of some patients who, with neurological damage in certain brain areas, are completely unable to make a decision, despite being perfectly able to make a correct evaluation of all factors involved. Daniel Goleman [7] is one of the major experts in the world of emotional intelligence. Goleman, in his experiments, noted the success of people without a great cognitive intelligence of logical-mathematical type but with a strong emotional sensitivity.

V. Social Media

With the advance of Internet and Web technologies, the accessibility of social data through social networks, blogs, forums and news sites has increased rapidly and has pulled much attention all over the world in recent research These social data can be used in marketing, decision making, destabilizing terrorist networks, behavior evolution, discovering social relationship from multiple entities as well as many other applications. But those social data need to be reflected for suitable usages. To process social data for suitable usages, social network analysis is one of the most important issues. Many data mining algorithms can be used to analyze social networks. For last few years, some research works have been developed on social networks. But all of those researches are mainly for social interaction and connectivity analysis. Those researches have not been focused on mining intellectual knowledge and comparison analysis. Again, none of them has been done on the analysis of individual social data in the social usages In real life, individuals become friends when they share common interests or passions. Sociologists have termed this tendency of human beings as 'homophily'. Similarly, on online social networks (OSNs), like Facebook or Orkut, users establish friendships when they discover similar profile characteristics. The growth of LinkedIn, a social networking website, demonstrates the impact of profile information very well. Its purpose is to help people build professional networks and find career development opportunities.

A. *Language in Online Social Networks*

Texts in online social networks have their specificity that must be taken into account. Indeed, it is common in these sites for users to use an informal and less structured language to communicate with their friends. Corney et. al. presents some features of this "online language". They are presented here in addition to other interesting features:

- Intentional misspelling, in particular the repetition of a letter in the same word, (e.g. "hellooooooo").
- Interjections and lexical surrogates for vocalizations (e.g. "mwah" indicating a kiss or "hmmmm").
- Grammatical markers such as the use of upper-case letters and the excessive use of punctuation (e.g. repetition).
- Social Acronyms: Acronyms of popular expressions used in online chatting systems and online social networks. For instance "BRB" denotes the expression "be right back".
- Emoticons: visual arrangements of characters in order to form facial expressions conveying emotions. For instance ":)" indicates joy and ":(" indicates sadness. In addition to this informal language, sentences may also lack proper syntax structure and words may be misspelled. Furthermore, specific to the non-English users is the use of non Latin-based languages transliterated into the English Alphabet; the use of other languages such as French and Spanish is also common.

VI. Emotion Theories

Emotions have fascinated researchers for long, as is evident in the vast body of research work related to emotion in fields of psychology, linguistics, social sciences, and communication. Human emotion manifests itself in the form of facial expressions, speech utterances, writings, and in gestures and actions. Consequently, scientific research in emotion has been pursued along several dimensions and has drawn upon research from various fields. This paper addresses the task of emotion recognition by attempting to automatically learn emotions from text. The French philosopher René Descartes' treatise, *Les passions de l'âme* (Passions of theSoul), published in 1649, is considered to be among the earliest works to theorize emotions (Anscombe and Geach, 1970; Cowie, 2000). The basic hypopaper presented in the treatise categorizes emotions into primary emotions and secondary emotions.

More recently, researchers have investigated several aspects of human emotion in order to arrive at a set of emotion categories that are universally acceptable (Picard, 1997). Several works in this direction have been papered in the literature (Tomkins, 1962; Izard, 1977; Plutchik, 1980; Ortony et. al., 1988; and Ekman, 1992). Table 1. lists the basic emotion categories identified by the different researchers.

Table 2.1 Basic Emotion Categories Identified by Researchers

Tomkins (1962)	Izard (1977)	Plutchik (1980)	Ortony et.al. (1988)	Ekman (1992)
joy	enjoyment	joy	joy	happiness
anguish	sadness	sorrow	sadness	sadness
fear	fear	fear	fear	fear
anger	anger	anger	anger	anger
disgust	disgust	disgust	disgust	disgust
surprise	surprise	surprise	surprise	surprise
interest	interest	acceptance		
shame	shame	anticipation		
	shyness			
	guilt			

Some psychologists have investigated facial expressions of emotion to identify the basic discriminable expressions among them, and mapped them to basic human emotions. Ekman (1992) has defined basic emotions as those that have universally accepted distinctive facial expressions. The six basic emotions defined on this basis are *happiness*, *sadness*, *fear*, *anger*, *disgust*, and *surprise*. This work uses Ekman's emotion categories since these emotions have been most widely accepted by the different researchers (see Table 2.1). Ekman's emotion categories have also been previously used in other computational approaches to emotion recognition (Liu et al., 2003; Alm et al., 2005; and Neviarouskaya et al., 2007a,b). Some researchers such as Schlosberg (1954) have referred to continuous dimensions of emotion instead of distinct emotion categories. There is agreement among researchers on at least two of these dimensions: *valence* (positive/negative) and *arousal* (calm/excited) (Barrett, 1998). Plutchik (1980) and Frijda et al. (1992) have highlighted the role of the intensity component in the study of emotion. The Circumplex Theory of Affect (Watson and Tellegen, 1985) identifies two main dimensions of positive and negative affect, which range from high to low. It also recognizes an alternative set of dimensions in the model, which consist of *pleasantness-unpleasantness* and *engagement-disengagement*. These two systems of axes define eight emotion octants in the model (Figure 2). Each emotion octant is marked with three to six characteristic affect words. These words were chosen on the basis of self-papered experiences of emotions by human subjects (Rubin et al., 2004).

Several works have been papered in the literature on the study of emotion expression in texts. The communicative function model of language introduced by the Russian-American linguist, Roman Jakobson (1960) identifies emotive function as one the six functions of language. The written expression of emotion lacks gestures, tones, and facial expressions, and instead relies on creative use of words for communicating emotion. Johnson-Laird and Oatley (1989) have deduced basic emotions by analyzing 590 English words, which describe emotion.

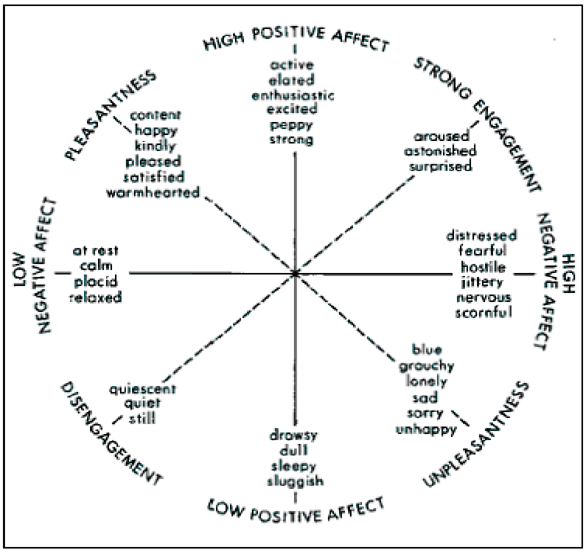

Fig. 1: Watson and Tellegen's Circumplex Theory of Affect (Watson and Tellegen, 1985)

Osgood's theory of Semantic Differentiation (Osgood et al., 1957) deals with assigning emotive meanings to words along three dimensions. Osgood et al. performed factor analysis of texts to identify three main factors on which the affective words can be rated. The three factors are: *evaluative factor* (good or bad), *potency factor* (strong or weak), and *activity factor* (active or passive). The evaluative factor carries the strongest relative weight, and several works have focused on it (Kamps and Marx, 2002; Turney and Littman, 2003; and Mullen and Collier, 2004). Some words convey emotion explicitly, while some other words can be used to convey emotion implicitly depending on the context (Clore et al., 1987). Strapparava and Valitutti (2006) have classified words into '*direct affective words*' (explicit) and '*indirect affective words*' (implicit) categories. My work has utilized both these types of words. The experiments papered in this paper show that it is important to take into account a variety of emotion-related words for automatic recognition of emotion, including direct and indirect affective words.

A. Sentiment Analysis

Sentiment Analysis is a rapidly growing area of research. It focuses on developing automatic systems that can analyze natural language texts to determine the sentiment expressed in them. The word "sentiment" is often used in a wide sense to refer to expressions of subjectivity, opinion, affect, attitude, orientation, feelings, emotions, and tone in the text. Much of the current work in sentiment analysis has focused on the task of determining the presence of sentiment in the given text, and on determining its valence, that is, the classification of sentiment according to positive or negative orientation. The sentiment classification task is quite often contrasted with that of topic-based categorization (Pang et al., 2002; Whitelaw et al., 2005). Pang et al. (2002) have empirically investigated if sentiment classification can be regarded as a special case of topic-based categorization, with positive and negative sentiment being taken as the two topics. They found that accuracy achieved in the sentiment classification problems is not as high as achieved in the conventional topic-based categorization, despite the different types of features tried for representing text. The results indicate that sentiment classification is comparably more challenging, because while topics are frequently expressed by relevant keywords, sentiment information is often embedded in the text in more subtle ways. The automatic recognition of sentiment, particularly in large volumes of documents, can have a variety of applications, notably in summarizing popular sentiment about any product or issue. Such information may particularly be of interest to policy-makers, economists and market researchers, political analysts, and social scientists. One of the earliest areas to attract the attention of researchers in sentiment analysis is that of movie reviews (Pang et al., 2002; Turney, 2002; Turney and Littman, 2003; Pang and Lee, 2004; Whitelaw et al., 2005; Kennedy and Inkpen, 2006). There are many popular movie review sites on the Web. Addition of sentiment information to movie reviews can provide additional information to the readers and increase the popularity of the reviews. It can also help in automatic analysis of movie reviews to determine the viewers' opinions. Pang et al. (2002) have attempted to classify the movie reviews drawn from the Internet Movie Database (IMDB)**[8]** into positive and negative categories using machine-learning techniques. Turney (2002) has proposed an unsupervised method of classifying movie reviews into *recommended* and *not recommended* categories based on the semantic orientation of adjectives and adverbs in the review. Kennedy and Inkpen (2006) have investigated the role of valence shifters in movie review classification. Contextual valence shifters can change the intensity of sentiment or even the sentiment itself according to the context (Polanyi and Zaenen, 2004). The authors have investigated the role of three types of valence shifters – intensifiers, diminishers, and negations – in predicting the sentiment of movie reviews and found that taking valence shifters into account helps improve the classification results. Pang and Lee (2005) have gone beyond the binary classification of movie reviews to address the task of fine-grained classification of the reviews on a multi-point scale. Their objective is to capture the reviewers' numerical ratings similar to the five-star ratings provided by the review authors. The main challenges cited for this task are the presence of certain degrees of similarities between the labels, which makes categorization into distinct classes difficult; and the misclassification of the borderline cases. This paper also deals with multiple emotion classes to which the sentences have to be assigned. Here also I found some similarity in classes, as all of them are basically emotion classes. Particularly notable in this regard are two classes – *anger* and *disgust*, which the human annotators found hard to distinguish in many cases. In another example of multi-class sentiment assignment, Nadeau et al. (2006) address the task of automatic dream sentiment classification on a 4-level scale from neutral to highly negative. In their study, the dreams gathered from a dream bank were manually annotated for use in different classification approaches. The best accuracy achieved was 50%, indicating the difficulty of the task. The baseline for the task was 33% using a majority class guessing approach. Besides movie reviews, the Web also abounds in reviews for all kinds of products and services ranging from consumer goods to restaurants and vacation spots. In an article in Forbes.com, Hoffman (2005) acknowledges the fact that "successful applications (based on sentiment analysis) could help automate market and product research". Market research needs feedback from the consumers, and the Web provides a convenient and popular public forum from where the feedback can be collected. However, the feedback information is often present in the form of unstructured free-form text on review sites, blogs, and other discussion forums, and can be useful only if automatic techniques for consumer sentiment analysis are available. Many researchers are involved in development of automatic systems that can aggregate the consumer sentiment from this vast amount of consumer-generated media, classify them into positive or negative, and also track them over time. The work papered in paper also draws its data from the Web. Hu and Liu (2004) process customer reviews collected from C|Net**[9]** and Amazon**[10]** to detect customers' opinion on different product features. They first find the product features mentioned in the reviews, and then identify the related opinion sentences and their orientation. Finally, results from all the customer reviews are summarized. This summarization differs from conventional topic-based summarization, as it only extracts those sentences in which the customers have commented on the product features. Such product feature summaries can help potential customers to know about the product and make an informed decision. It can also help the manufacturers to know the customers' opinions about various product features. Mishne and Glance (2006) analyze blog data to show correlation between popular blogger sentiment about movies and corresponding movie sales. They collect relevant blog posts discussing movies from the posts appearing in the BlogPulse**[11]** Index (Glance et al., 2004), and estimate the blogger sentiment prior to and after the movies' release. Experiments revealed that positive sentiment about movies correlates heavily with the movies' success at the box office. Recognizing sentiment from blogs is a more challenging task, as the data in blogs tends to be more informal and less focused compared to the data drawn from dedicated review sites (Mishne and Glance, 2006). This task is further complicated by use of informal language and lack of proper structure in blogs (Java et al., 2006; and Mishne, 2005). For a more detailed description of the nature of information present in blogs and of the challenges associated with mining information from the blogosphere, see (Mishne, 2006). Zhang et al. (2006) draw attention to the variety of challenges posed by the kind of language used in non-standard text. These include the presence of misspellings, slang, ungrammaticality, abbreviations, and onomatopoeic elements (such as "grrr", "hmm") as well as use of upper case, special punctuation (such as "!!!"), and repetitions of letters or words (such as "sweeeet") for affective emphasis. They apply several preprocessing steps to their corpus to address these challenges. These steps include a look-up table to deal with abbreviations and a small dictionary, containing base forms of certain special words. These resources are used to provide the appropriate replacements for non-standard usage in text. In addition, they have also used two spelling-correction algorithms. Another work, which has discussed the various kinds of noise present in the online text-based communication, is by Sokolova et al. (2005b). They discuss the challenges posed by the presence of noise in the context of the electronic negotiations data. To address the problem, they have used spellchecking methods based on the frequency counts to use the most appropriate replacement for misspelled words. One particular area of sentiment analysis, which focuses on identifying expression of opinion in the text as well as of their orientation, is called Opinion Mining or Opinion Analysis. Hurst and Nigam (2004) combine shallow NLP techniques and statistical machine learning to find the opinion sentences about the given topics from a corpus of data containing both relevant and irrelevant sentences. Opinion classification has also been studied by Liu et al., (2005), Wilson et al. (2006) and Riloff et al. (2006). Hiroshima et al. (2006) propose a web search engine that can search for the sentences expressing opinion about a given topic. Search engines can improve the usefulness of results provided by them if these results present a comparison of the different opinions on any topic. While in movie reviews, product reviews, and blogs, the opinions are most likely to be that of the review author, news articles may contain the opinions of various people. Kim and Hovy (2006) have investigated the task of finding the opinion-holder and the topic of opinion from the online news articles. Mullen and Malouf (2006) have used the posts from a political discussion forum to classify them on the basis of political sentiments. All posts are selflabeled with the writer's stated political affiliations, and the authors have mapped those affiliations into two broad political classes: right and left. Some researchers focus on identifying the instances of subjective language in the text as a precursor to finding the sentiment information. Subjective language is used to express opinions, evaluations, and other private states (Wiebe et al, 2004). The motivation behind this approach is that sentiment is often expressed using subjective language, and once the subjective content is sifted out of the text, further techniques can be applied to find the overall orientation, as well as the specific instances of opinion or emotion from it (Hatzivassiloglou and Wiebe, 2000; and Wiebe et al., 2004). Pang and Lee (2004) demonstrate the effectiveness of extracting subjective portions of text before performing polarity classification of movie reviews. Chesley et al. (2006) have also investigated the problem of subjectivity and polarity classification of blog posts.

VII. CONCLUSION

This is an investigation of expressions of emotion in social media text. The paper covers manual works performed to identify textual expressions of emotion, as well as the computational methods adopted to learn emotions in text. The presentation begins with an introduction of the problem. The high-level objective of this work was to explore the applicability of automatic approaches to recognizing emotions expressed in written text. The scope of this work was limited to learning emotions that can be interpreted from textual expressions. Emotion research is an inter-disciplinary area and draws upon earlier works in Psychology, Linguistics, and Natural Language Processing. I provided a description of the previous works in these fields. The differences in human judgment related to the assignment of *emotion category* and *emotion intensity* to individual sentences in the corpus indicate that interpretation of emotion expressed in text is subjective; in many cases, different judges interpreted differently the emotive content of the same sentences. The primary objective of this work is to develop automatic methodologies for learning emotions from text. The resources I used – General Inquirer, WordNet-Affect, and *Roget's* Thesaurus – are all publicly available24. The former two implicitly categorize words on the basis of semantic distinctions, many of which are relevant to emotion. The latter's classification system was utilized to automatically build a lexicon of emotion-related words based on the similarity of words to a small primary set of manually identified emotion words. The results of the experiments suggest that combining a variety of features – representing explicitly emotion words as well as words that may be conceptually related to emotion – helps distinguish the fine-grained emotions in text. The performance achieved in these experiments was significantly better than that using the baseline methods. I also addressed the task of automatic recognition of emotion intensity level of text.. The authors note that opinion pieces are not entirely composed of subjective sentences and that as many as 30% sentences in opinion pieces were found to be objective. In the above passages, I summarized the work presented in this paper. In next two sections, I present major contributions of this work and a roadmap for future research.

VIII. CONTRIBUTIONS

This work addresses an important and less investigated area of sentiment research, that is, emotion detection in text. The major contribution of this work is to show that it is viable to apply computational methods to identify and distinguish various types of emotions in text. Similar works in fine-grained emotion detection lack mention of conventional performance metrics such as precision and recall, which prevents proper assessment of their approaches (Liu et al., 2003; Neviarouskaya et al., 2007a). They demonstrated the efficacy of their methods through user evaluation studies only. This paper papers the first empirical results in this domain that addresses the problem of detecting basic emotion categories at the sentence level. The experiments performed in automatic emotion detection also suggested that to achieve good performance that it is important to include in the ambit of consideration a wide variety of words that go beyond the stereotypical emotion words. This finding might appear commonplace, but the fact is that it is not emphasized in earlier works, and not particularly proven empirically. Another significant contribution of this work is to produce a 5205-sentence corpus of emotion-annotated data. The annotations include fine-grained distinctions of various emotion categories, the emotion intensity levels, as well as the emotion indicators in text. No comparable work focusing exclusively on emotions exists in the public domain. This paper also introduced a novel approach of automatically building emotion lexicon utilizing the classification system of *Roget's* Thesaurus. In earlier works, emotion (or sentiment) words were typically acquired using WordNet or corpus-based approaches. In the approach introduced here, a variety of emotion-related words were learned, and their usability demonstrated by their effectiveness in ML methods for emotion detection.

IX. FUTURE WORK

The work presented in this paper can be pursued further in several directions. One of the tasks to be addressed is to explore the relation between emotion categories and intensity. Some steps could be taken to address the special needs of the kind of informal language used in online communication. This would help improve performance. Another direction for future work is to consider the emotion intensity classification problem as that of ordinal classification – that is, classification with ordered categories. The intensity levels of high, medium, low, and neutral form a natural ordering, which can be taken into account during classification. One of the approaches used for this kind of problem is to assign numerical labels to classes and then apply regression (Wilson et al., 2004). Another approach is to transform the ordinal problem into a series of binary class problems that incorporate ordering information (Frank and Hall, 2001). These methods could be applied to the emotion intensity classification task addressed in this paper. The data prepared as part of this work is rich in emotion annotations and offers several exciting possibilities for further research. Future work may attempt to automatically identify the emotion indicators in sentences. A corpus-driven approach can allow a lexicon of emotion words to be built starting from the set of emotion indicators identified during the annotation process. This set could be further extended based on existing syntactic patterns in the set and similarity measures to identify similar words. Content Analysis of emotion-labeled data is yet another possible line of research. If thesalient words in emotion sentences are analyzed on the basis of their affiliations to semantic categories as defined in resources such as *Roget's* Thesaurus, it can help pinpoint the sources that evoke particular emotions.

REFERENCES

[1] http://gaips.inesc-id.pt/acii2007/index.html
[2] W. L. Jarrold, "Towards a theory of affective mind: computationally modeling thegenerativity of goal appraisal", Ph.D. dissertation, 2004, supervisor : Schallert, Diane L.
[3] C.-G. Ortony, A. and A. Collins, « The cognitive structure of emotions », New York:Cambridge University Press, 1990.
[4] M. M. C.E. Osgood, W.H. May, Cross-cultural Universals of Affective Meaning. Urbana: University of Illinois Press, 1975.
[5] Ekman, Paul, "Emotions Revealed: Recognizing Faces and Feelings to ImproveCommunication and Emotional Life". OWL Books, NY, 2007.
[6] A. Damasio, The Feeling of What Happens: Body and Emotion in the Making ofConsciousness, was named as one of the ten best books of 2001 by New York Times BookReview.
[7] D. Goleman "Emotional Intelligence: Why It Can Matter More Than IQ", Bantam Books,1996.
[8] http://reviews.imdb.com/Reviews/
[9] http://www.cnet.com/
[10] http://www.amazon.com/
[11] http://www.blogpulse.com/