**Emotion Detection in Text: A review**

<https://arxiv.org/abs/1806.00674>

**Abstract**

Emotion Detection in text has exploded in popularity due to its potential applications in various fields such as marketing, science and psychology. In this paper, such progress on the emotion detection is being reviewed. Arguably, although many techniques, methodologies and models have been created for emotion detection, there is a need for improvement in the design and architecture of the current systems. Factors such as the complexity of human emotions, and the use for implicit and metaphorical language in expressing led us to think that just re-purposing standard methodologies will not be enough to capture these complexities, and that it is important to pay attention to the linguistic intricacies of emotion expression.

**Introduction**

Emotion detection in computational linguistics is the process of identifying discrete emotions expressed in text, whereas emotion analysis can be viewed as a natural evolution of sentiment analysis and its more fine-grained model. Sentiment analysis, on the other hand is a well-established field in natural language processing with thousands of articles written about its method and applications in various fields such as marketing and advertising (Qiu et al., 2010; Jin et al., 2007).

The amount of useful information which can be gained from moving past the negative and positive sentiments and towards identifying discrete emotions can help improve many applications mentioned above, and open ways for new use cases.

Not all negative or positive sentiments are created equal, one good example is the comparison between two emotions; Fear and Anger. Fear is a passive emotion which affects the person to have a pessimistic view of the future while anger can lead to action and affect them to have an optimistic view of the future. These precise types of information indicate potential uses of emotion detection.

**Psychological Models of Emotion**

This part provides the definitions, terminologies, models and theories.

In psychology, based on the appraisal theory, emotions are viewed as states that reflect evaluative judgements of the environment, the self and other social agents, in light of the organisms goal and beliefs, which motivate and coordinate adaptive behavior (Hudlicka, 2011). In psychology, emotions are categorized into basic emotions and complex emotions.

Most papers for emotion detection used the Ekman’s six basic emotions (Ekman, 1992) comprised of: *sadness, happiness, anger, fear, disgust and surprise* for detecting emotions as a multi-class classification problem, along with some that are based on Plutchik’s wheel of emotions (Plutchik 1984); *joy, trust, fear, surprise, sadness, disgust, anger and anticipation*.

Upon a different perspective, dimensional model of emotions define emotion based on two or three dimensions, based on the hypothesis that all emotions are the result of a common and interconnected neurophysiological system. The Circumplex Model (Russell, 1980) suggests that emotions can be shown in a two-dimensional circular space; one dimension for arousal and another for valance. The dimensional models are barely used in emotion detection literature, but shows as promising model to represent emotions in textual data. (Calvo and Mac Kin 2013).

**Complexity of Expressing Emotions in Language**

Emotion expression is very sensitive and complex. Ben-Ze’ve (2000) relates this complexity to various reasons; first is its sensitivity to multiple personal and contextual circumstances, second is the fact that these expressions often consist of a cluster of emotions rather than merely a single one, and finally the confusing linguistic use of emotional terms. Bazzanella (2004) argues that complexity of emotions can be seen in multiple levels “the nested interplay with mind/language/behavior/culture, the lexical and semantic problem, the number of correlated physiological and neurological features, their universality or relativity, etc.”.

It has been shown that context is very important, and is crucial in understanding emotions (Oatley et al., 2006). Most recent studies in textual emotion detection in LNP are based on explicit expression of emotion using emotion-bearing words. But emotion expression is mostly done by expressing emotions-provoking situations, which can be interpreted in an affective manner (Balahur and Montoyo, 2008; Pavlenko, 2008).

There are sentiment analysis literatures that attempted on the areas of detecting implicit expression of emotions. Greene and Resnik (2009) used syntactic packaging for ideas to asses the implicit sentiment in text, and to improve the state of the art sentiment detection techniques. Cambria et al (2009) proposed an approach to overcome this issue by building a knowledge base that merges Common Sense and affective knowledge, the goal is to move past the methods that rely on explicit expression of emotion; their reasoning for choosing this approach was based on the notion that most are emotions expressed through concepts with affective valence. Lakoff (2008), in a case study of *Anger* argues that emotions have a very complex conceptual structure, and that the structure could be studied by systematic investigation of expression that are understood metaphorically; he also argued that many expressions of anger are metaphorical, thus could be assessed by the literal meaning of expression. This fact makes it more difficult to create a lexical, or machine learning method to identify emotions in text without first solving the problem of understanding of metaphorical expressions.

**Resources for Detecting Emotions in Text**

In this section, some of the most prominent and publicly available sources will be introduced. These data can be separates into two groups: Labeled texts and emotion lexicons.

**Labeled text**

One of the most prominent and well-known sources for emotionally labeled text is the Swiss Center for Affective Science (SCA). The most used resources they provide is International Survey on Emotion Antecedents and Reaction (ISEAR) which consists of responses from about 3000 people around the world who were asked to report situations in which they experience each from the seven major emotions (joy, fear, anger, sadness, disgust, shame and guilt), and how they reacted to such situations, resulting to a promising dataset of 7600 records of emotion-provoking texts to be used in testing many methods for emotion extraction and classification.

EmotiNet knowledge base (Balahur et al., 2011), Balahur et al. argued that word level attempt to detect emotion would lead to a low performance system because ”expressions of emotions are most of the time not presented in text in specific words”, rather from the ”interpretation of the situation presented” in the text. They based their insight on *Appraisal Theory* in psychology (Dalgleish and Power, 2000). They managed to create a new knowledge base containing action chains and their corresponding emotional label. This approach showed promise, but it could not present itself as a viable and generally applicable in its current form due to the small size of knowledge base and the structure of information they used (limited to four-tuple of actor, action, object and emotion).

Vu et al. (2014) focused on discovery ad aggregation of emotion-provoking events. They created a dictionary of such events through a survey of 30 subjects, and used that to aggregate similar events from the web by applying Espresso pattern expansion (Pantel and Pennacchiotti, 2006) and bootstrapping algorithms.

One of the frequently used dataset is the SemEval-2007 (Strapparava and Mihalcea, 2007), which consists of 1250 news headlines extracted from news websites, and annotated with Ekman’s six emotions. The other example, Alm’s annotated fairy tale dataset (Alm et al., 2005), consists of 1580 sentences from children fairy tales, also annotated with six Ekman’s emotions. These datasets have been mostly used as benchmark in the literature.

**Emotion Lexicons**

The lack of reasonable size lexicon for emotions led Mohammed and Turney (2010) to create an emotion word lexicon. Later, Mohammed and Turney (2013) used Amazon Mechanical Turk to annotate around 14000 words in English Language (along with lexicons in other languages and are available on their website).

Another popular emotion lexicon used in literature is WordNet-Affect. Strapparava et al. (2004) tried to create a lexical representation of affective knowledge by starting from WordNet (Miller and Fellbaum, 1998), a well-known lexical database. Then, they used selection and tagging of a subset of synsets which represents the affective concepts, with the goal of introducing “affective domain labels” to the hierarchical structure of WordNet. WordNet-Affect, despite its small size (2874 synsets and 4787 words), was a great attempt to extract emotional relations of words from WordNet, and was used in many early applications of sentiment analysis, opinion mining (Balahur et al., 2013), and in emotion detection especially for extending affective word sets from the basic set of emotions.

Another emotion lexicon, developed by Staiano and Guerini (2014) called DepecheMood. They used crowd-sourcing to annotate thirty-five thousand words. They showed that lexicons could be used in several approaches in sentiment analysis, as a feature for classification in machine learning methods (Liu and Zhang, 2012) or to generate an affect score for each sentence, based on the scores of the words which are higher in the parse tree (Socher et al., 2013b).

LIWC lexicon (Pennebaker et al., 2001) consists of 6400 words annotated for emotions

ANEW (Affective Norm for English Words) developed by Brandley and Lang (1999) which consists of 2000 words which has been annotated based on dimensional model of emotions with three dimensions of valance, arousal and dominance.

**Word Embedding**

Word Embedding is a technique based on distributional sematic modeling. It is rooted in the idea that words which frequently co-occur in a relatively large corpus similar in some sematic criteria. In such methods, each word is represented as a vector in an n-dimensional space, called the vector space, and in a way that the distance between vectors corresponds to the semantic similarity of the words they represent. These vector space models have been shown to be useful in many natural language processing tasks, such as named entity recognition (Turian et al., 2010), machine translation (Zou et al., 2013), and parsing (Socher et al., 2013).

**Methodologies for Detecting Emotions in Text: Supervised Approaches**

Due to the lack of emotion-labeled datasets, many supervised classifications for emotions have been done on data gathered from microblogs using hashtags or emotions as the emotional label for the data, under the assumption that these signals show the emotional state of the writer.

Suttles and Ide (2013) where the four pairs of opposite emotions in the Plutchik’s wheel of emotion were used to create four binary classification tasks. With hashtags, emoticons and emoji as labels for their data, they reached between 75-91% accuracy on a separate manually labeled dataset

The study of Purver and Battersby (2012) on Twitter data using SVM classifier showed that the classifiers performed well on emotions like happiness, sadness and anger but not well for others. They concluded that using hashtags and emoticons as labels is a promising labeling strategy and an alternative to manual labeling.

Hasan et al. (2014) also used hashtags as their label and created their features using the unigram model, removing any word from tweets which were not in their lexicon (created using 28 basic emotion word in Circumplex model and extended with WordNet synsets). Four classifiers (Naive Bayes, SVM, Decision Tree, and KNN) achieved accuracies close to 90% in classifying four main classes of emotion categories in Circumplex model. Hasan et al. (2018) created an automatic emotion detection in system to identify emotions in stream of tweets. This approach included two tasks: training an offline emotion classification model based on their 2014 paper, and a two-step classification to identify tweets containing emotions and to classify such emotional tweets into more fine-grained labels using soft classification techniques.

**Methodologies for Detecting Emotions in text: Unsupervised Approach**

Kim et al. (2010) used an unsupervised method to automatically detect emotions in text based on both categorical and dimensional models of emotion. They used three datasets: SemEval-2007 Affective text, ISEAR and children’s fairy tales. For categorical model, they used WordNet-Affect t as the lexicon, and evaluated three dimensionality reduction methods: Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (PLSA), and Non-negative Matrix Factorization (NMF). And for the dimensional model, they used ANEW (Affective Norm for English Words) and WordNet-Affect as a means to extend ANEW. They assigned the emotion of the text based on closeness (cosine similarity) of its vector to the vectors for each category or dimension of emotions. Their study showed that NMF-based categorical classification performs best among categorical approaches, and dimensional model had the second best performance with highest F-measure of 0.73.

PMI(x, y) = cooccurrence(x, y) / occurrence(x) ∗ occurrence(y)

Agrawal and An (2012) started by extracting NAVA words (i.e. Nouns, Adjectives, Verbs, and Adverbs) from a sentence, and then extracting syntactic dependencies between extracted words in each sentence to include contextual information in their model. They then used semantic relatedness to compute emotion vectors for words, based on the assumption that the affect words (NAVA words) which co-occur together more often tend to be semantically related. They use Point-wise Mutual Information (PMI) as the measure of semantic relatedness of two words (Equation above) and computed a vector for each word using the PMI of the word with all words related to each emotion, then adjust the vectors by considering the contextual information in syntactic dependency of words. After computing vectors for each word they generated a vector for each sentence by aggregating the emotion vectors of all the affect words in it. By evaluating on multiple data sources, they showed that their method preformed more accurate, compared to other unsupervised approaches, and had comparable results to some supervised methods.

Another lexicon base approach by Mohammad (2012) showed how detecting emotions in text can be used to organize collection of text for affect-based search, and how books portray different entities through co-occuring emotion words by analyzing emails and books. He used NRC lexicon to see which of the emotion words exist in the available text, and calculated ratios such as the number of words associated to a particular emotions compared to other emotions, to determine if a document have more expressed emotions compared to other documents in the corpus. He compiled three datasets for emotional emails: love letters, hate mails, and suicide notes. He goes on to analyze presents of different emotions based on criteria like, workplace emails, emails written by women/men, or emails written by men to women vs men.

**Discussion and Open Problems**

The difficulties of detecting expressed emotions may be attributed to many problems from *complex nature of emotion expression in text*, *inefficiency of current emotion in detection models* *and lack of high quality data to be utilized by those models*.

**Complex Nature of Emotion Expression**: Expression of emotion is a complex phenomena, in such a way that a shortest phrase cam express multiple emotions with different intensity that cannot be understood at first glance even by humans. On the other hand, the intricacy of emotional language, resulting from the vast use of metaphorical language, context dependent nature of emotion expression, and implicit nature of such expression makes this task even harder. To address these issues, it is important to pay attention to the complexity of emotional language when building emotion detection systems. These systems should also be designed based on the linguistic complexities of emotion expression to be able to grasp the implicit expression of emotions, and untangle the metaphorical nature of these expression. It is also crucial to consider the contextual information in which the expression is occurring.

**Shortage of Quality data**: The quality and quantity of data has a huge effect on the performance of emotion classification algorithms. Although huge amount of textual data is currently available, for any supervised model, a large amount of annotated data is required. To overcome this problem, there has been use of self-annotated microblog data but it does not yet possess the qualities which are required for an applicable system. Additionally, the niche nature of the language used in microblog text prevents the systems trained on these texts to be used to classify other types of text (e.g. tweets vs. news comments). Furthermore, the imbalance nature of currently available emotional text, will cause the classifier to severely under-preform for emotions that are underrepresented in the dataset. Therefore, any attempt to create a large balanced dataset, with high quality labels could provide a brighter future for the field.

**Inefficiency of Current Models**: There have been many attempts to approach the need for performance in such systems, with the most frequently used being; converting the task of multi-class to multiple binary classification, either by having one classifier for each emotion (e.g. anger vs not anger), or one classifier for a pair of opposite emotions (e.g. joy vs sadness). Further improvement in classification algorithms, and trying out new ways is necessary in order to improve the performance of emotion detection methods. Some suggestions include developing methods that can go above BOW representations and consider the flow and composition of language. In addition, specific neural network designs or ensemble methods are possible approaches that has been shown to be useful in other areas of natural language processing. New ways to increase the emotional qualities of embeddings and vector models could be beneficial in unsupervised methods, or be used as features in neural networks.

**Conclusion**

While many successful methodology and resources was introduced for sentiment analysis in recent years, researchers, by understanding the importance of more fine-grained affective information in decision making, turned to emotion detection in order to distinguish between different negative or positive emotions. In addition, having large amount of textual data with the rise of social media in past couple of decades, and therefore the availability of vast self-expression text about any major or minor event, idea, or product, points to a great potential to change how entities and organizations can use these information as a basis for their future decision making processes