



Multi-Class Twitter Emotion Classification: A New Approach

R C Balabantaray
Department of CSE
IIIT Bhubaneswar
IIIT Bhubaneswar, Odisha

Mudasir Mohammad
Department of CSE
IIIT Bhubaneswar
IIIT Bhubaneswar, Odisha

Nibha Sharma
Department of CSE
IIIT Bhubaneswar
IIIT Bhubaneswar, Odisha

ABSTRACT

Micro blogging today has become a very popular communication tool among Internet users. Millions of users share opinions on different aspects of life every day. Therefore micro blogging web-sites are rich sources of data for opinion mining and sentiment analysis. Because micro blogging has appeared relatively recently, there are a few research works that are devoted to this topic. In this paper, we are focusing on using Twitter, the most popular micro blogging platform, for the task of Emotion analysis. We will show how to automatically collect a corpus for Emotion analysis and opinion mining purposes and then perform linguistic analysis of the collected corpus and explain discovered phenomena. Using the corpus, we will build a Emotion classifier that will be able to determine the emotion class of the person writing.

Keywords

Emotion Analysis; Sentiment Analysis; Opinion Mining; Text Classification

1. INTRODUCTION

Emotions play important role in human intelligence, rational decision making, social interaction, perception, memory, learning, creativity, and more. Opinion mining and sentiment analysis is a subtopic of natural language processing and text mining that deals with the automated discovery and extraction of knowledge about people's sentiments, evaluation and opinions from textual data such as personal blogs, review websites and customer feedback forms. Opinion mining and sentiment analysis is an area that has received significant interest in recent times because of its practical usage and application in today's environment.

Previous researches for emotional analysis of texts have included a variety of text contents: web-logs [2,3], stories[4,5], news, text messages[6], spoken dialogs[7], etc. For many applications, identifying emotions only on document level may not be sufficient. A text-based emotion prediction system would benefit from identifying the emotional affinity of sentences. The emotion analysis on sentence level may also be important for more detailed emotion analysis systems.

So far emotion analysis of text classifies text into three categories only i.e. Positive, negative and neutral, however we would extend this and classify it into six basic emotion classes i.e. positive, negative, fear, joy, surprise, hate, disgust. In this paper we discuss the design an emotion classifier engine capable of such classification. We discuss the methodology and results of our proposed system. Our goal is to investigate the expression of emotion in tweets through a corpus annotation study and to prepare an annotated corpus for use in automatic emotion analysis experiments we explore syntactic, semantic and contextual-based approach for automatically classification. The results of the initial experiments show an improved performance over baseline accuracy.

The data for our experiments came from Twitter [8]. We wanted to identify the Emotions of different people, so Twitter was the best available source to get data for them. People freely communicate over micro-blogging sites and thus micro-blogging is rich in providing the emotion perspective of a person.

This paper is organized as follows. In section 2, the related work is reviewed. Section 3, describes the data preparation part. In section 4, describes Automatic Emotion Classification. Finally in section 5, conclusions are given.

2. RELATED WORK

Some researchers have studied emotion in a wider framework of private states [9]. Wiebe et al. [10] worked on the manual annotation of private states including emotions, opinions, and sentiment in a 10,000-sentence corpus (the MPQA corpus) of news articles. Expressions of emotions in text have also been studied within the Appraisal Framework [11], a functional theory of the language used for conveying attitudes, judgments and emotions [12, 13]. Neither of these frameworks deals exclusively with emotion, the focus of this paper. In a work focused on learning specific emotions from text, Alm et al. [14] have explored automatic classification of sentences in children's fairy tales according to the basic emotions identified by Ekman [15]. The data used in their experiments was manually annotated with emotion information, and is targeted for use in a text-to- speech synthesis system for expressive rendering of stories. Read [16] has used a corpus of short stories, manually annotated



with sentiment tags, in automatic emotion-based classification of sentences. These projects focus on the genre of fiction, with only sentence-level emotion annotations; they do not focus on context of emotion in a sentence, as we do in our work.

In other related work, Liu et al. [17] have utilized real-world knowledge about affect drawn from a common-sense knowledge base. They aim to understand the semantics of text to identify emotions at the sentence level. They begin with extracting from the knowledge base those sentences that contain some affective information. This information is utilized in building affective models of text, which are used to label each sentence with a six-tuple that corresponds to Ekman's six basic emotions [15]. Neviarouskaya et al. [18] have also used a rule-based method for determining Ekman's basic emotions in the sentences in blog posts.

Mihalcea and Liu [19] have focused in their work on two particular emotions –happiness and sadness. They work on blog posts which are self-annotated by the blog writers with happy and sad mood labels. Our work differs in the aim and scope from those projects: we have prepared a corpus annotated with rich emotion information that can be further used in a variety of automatic emotion analysis experiments.

3. THE EMPIRICAL STUDY

3.1 Data Sets

We worked with tweets that we collected from web [20]. The resource was prepared by downloading 1000 randomly selected twitter users and their tweets. Each user's tweets reside within a file, the filename consists of user-name and the suffix .tweet. Each line is formatted as time-stamp text. For example, a tweet sent by Joe would be in joe. tweets and be formatted as follows:

2008-11-14T16:01:38+00:00 I had cat food for dinner.

The dataset had 3,05,310 multi-lingual tweets .We removed foreign language tweets from our dataset, resulting in dataset comprised of only tweets in English language .The time-stamp not being relevant was removed from the data.

We prepared a list of seed words for six basic emotions categories proposed by Ekman [15]. These categories represent the distinctly identifiable facial expressions of emotion –happiness, sadness, anger, disgust, surprise and fear. Using these seed words we retrieved tweets from the resulting dataset, which was further refined by removing URL's from it. Has tags and user-names were retained. The resulting dataset has 8150 tweets.

Table 1: Sample Examples of the Refined Tweets

I feel so incredibly sick this morning."My friends." I'm black so I know I ain't his friend. #current
@Dina1504 Hi..did u hear weird audio on 8TV lately? like tape slowing down momentarily? seems astro wont believe us! thnx for help :)

20 degrees outside :D now let's see what it's like when
So far so Zbieczuk plus or minus. Just to increase the
@numberall Thanks not sure I get it yet but Ill play
I'm under very strict time constraints due to the fact that today is only today for a certain amount of time.

3.2 Data Annotation

Emotion labeling is reliable if there is more than one judgment for each label. Five judges manually annotated the corpus; each sentence was subject to two judgments. Annotations were performed by five people. The annotators received no training, though they were given samples of annotated sentences to illustrate the kind of annotations required. The annotators were required to label each sentence with the appropriate emotion category, which describes its affective content. If they didn't found the tweet belonging to any category it was assigned neutral label .Thus to Ekman's six emotions, we added no emotion (neutral), resulting in seven categories to which a sentence could be assigned. All sentences that had no emotion content were to be assigned to the no emotion (neutral) category. eg.

@mebner I would say social media is a cocktail party - sometimes the conversations "click" and/or we move on.

In the final annotated corpus, the no emotion category was the most frequent. It is important to have no emotion sentences in the corpus, as both positive and negative examples are required to train any automatic analysis system.

3.3 Measuring Annotations Agreements

The interpretation of sentiment information in text is highly subjective, which leads to disparity in the annotations by different judges. Difference in skills and focus of the judges, and ambiguity in the annotation guidelines and in the annotation task itself also contribute to disagreement between the judges [21]. We seek to find how much the judges agree in assigning a particular annotation by using metrics that quantify these agreements.

First we measure how much the annotators agree on classifying a sentence as an emotion sentence. Cohen's kappa [25] is popularly used to compare the extent of consensus between judges in classifying items into known mutually exclusive categories. Table 2 shows the pair-wise agreement between the annotators on emotion/non-emotion labeling of the sentences in the corpus. We report agreement values for pairs of annotators who worked on the same portion of the corpus.

Table 2. Pair-wise agreement in emotion/non-emotion labeling

	A<->C	B<->D	C<->E	D<->A	B<->E	Avg
Kappa	0.68	0.79	0.81	0.64	0.63	0.71



Table 3: Pair-wise agreement in emotion Categories

Category	A<>C	B<>D	C<->E	D<>A	B<->E	Avg
Happy	0.81	0.74	0.54	0.63	0.82	0.70
Sad	0.63	0.72	0.53	0.42	0.79	0.61
Anger	0.73	0.76	0.82	0.61	0.83	0.75
Disgust	0.84	0.68	0.75	0.63	0.71	0.73
Surprise	0.78	0.63	0.54	0.59	0.62	0.63
Fear	0.81	0.24	0.61	0.45	0.73	0.57

Within the emotion sentences, there are seven possible categories of emotions to which a sentence can be assigned. Table 3 shows the value of kappa for each of these emotion categories for each annotator pair.

3.4 Learning Model

Our emotion classifier is based on multi-class SVM kernels. The seed words obtained in Data normalization process are used as Vectors for training the SVM. The stem words are then preprocessed for the use of SVM. This is done by converting the stem words into numeric data that can be used by SVM.

We use SVM multi-class formulation described in [1]. But it optimizes it with an algorithm that is very fast in the linear case. For a training set $(x_1, y_1) \dots (x_n, y_n)$ with labels y_i in $[1..k]$, it finds the solution of the following optimization problem during training.

$$\min 1/2 \sum_{i=1..k} w_i * w_i + C/n \sum_i = 1..n \xi_i$$

$$\text{s.t. for all } y \text{ in } [1..k]: [x_1 \cdot w_{y1}] \geq [x_1 \cdot w_y] + 100 * \Delta(y_1, y) - \xi_1$$

$$\text{for all } y \text{ in } [1..k]: [x_n \cdot w_{yn}] \geq [x_n \cdot w_y] + 100 * \Delta(y_n, y) - \xi_n$$

C is the usual regularization parameter that trades off margin size and training error. $\Delta(y_n, y)$ is the loss function that returns 0 if y_n equals y , and 1 otherwise. [1]

3.5 Feature Set

The features were extracted using java. Tweets were preprocessed using Stanford CoreNlp Package [22]. The Support Vector Machine Framework (SVM) has been used for training as well as classifying tweets into above mentioned six emotion tags and one neutral tag. We have used syntactic, Semantic as well as Contextual features for our classification work. Feature is of following format

Feature: weight

Where feature represents the feature no and weight represents the weight the feature is having. We have used following features in our classification task:

- 1) Unigrams
- 2) Bigrams
- 3) Personal-pronouns
- 4) adjectives
- 5) Word-net Affect emotion lexicon
- 6) Word-net Affect emotion lexicon with left/right context
- 7) Word-net Affect emotion POS
- 8) POS
- 9) POS-Bigrams
- 10) Dependency-Parsing Feature
- 11) Emoticons

Each word was stemmed using Porter Stemmer [23], and stop words were removed before feature extraction.

Feature groups 3,4,8,9 are extracted automatically from the tweets with Stanford Penn-Bank POS-tagger.

Word-net-Affect assigns a variety of affect labels to a subset of syn-sets in Word-net. We utilized the publicly available lists extracted from WNA, consisting of emotion-related words. There are six lists corresponding to the six basic emotion categories identified by Ekman .

we utilized these lists to capture our feature group 5.

Feature group 6 tries to capture the contextual information of the emotion-word. We obtain a window defined context of the emotion-word. Window is defined in the configuration files and can be defined to obtain the desired context .It Captures both left and right context of the emotion word.

Feature group 7 was captured using combination of Word-net Affect emotion lexicon and Stanford POS-tagger. It captures the POS tags corresponding to Word-net-Affect Emotion Lexicon Feature Group 10 uses Stanford Dependency Parser [24].

we have used Typed Dependency using “edu/stanford/nlp/models/lexparser/englishPCFG.ser.gz” parser. We used Dependency of Personal-pronoun with the Verb as a feature.

Feature group 11 uses The Emoticons and their consecutive occurrences generally contribute as much as real sentiment to the words or sentences that precede or follow it.

4. RESULTS AND EXPERIMENTS

For our classification experiments, we used Support Vector Machines (SVM), which have been popularly used in sentiment classification tasks conclusion and future work. All Experiments were performed using Leave-One -Out Cross-validation. Each sentence was represented by a vector composed of multiple features in the sentence along with the feature weight. Examples depicted in table IV



Table 4: Example Vectors

5 9371:1.0 11946:1.0 13770:1.0 13820:1.0 13861:2.0
7 539:1.0 567:1.0 587:1.0 863:1.0 1550:2.0 1561:1.0 1828:1.0 2082:1.0 2484:2.0 2639:1.0 3178:1.0 3659:1.0 5588:1.0 6167:1.0 6764:1.0 8485:1.0 8498:1.0 8784:1.0 8825:1.0 9481:1.0 10312:2.0 11252:1.0 11257:1.0 11303:1.0 11840:1.0 12268:1.0 12817:1.0 13008:1.0 13650:1.0
2 2269:1.0 2556:4.0 3416:1.0 4124:2.0 5106:1.0 7130:1.0 7298:1.0 7861:2.0 8365:1.0 8649:1.0 9863:3.0 12933:2.0
5 71:1.0 446:1.0 894:1.0 895:1.0 898:1.0 1398:1.0 2103:1.0 2418:1.0 3218:2.0 3692:1.0 5502:1.0 6436:1.0 6575:1.0 6789:1.0 7475:2.0 7632:1.0 8194:1.0 8649:1.0 9090:2.0 9419:1.0 9545:2.0 9742:1.0 9863:3.0 10063:1.0 10282:1.0 10353:2.0 11461:1.0 12493:1.0 13069:1.0

The Above vectors are in the format class-label feature: weight.

To explore the contribution of different feature groups to the classification performance, we conducted experiments using different combination of features and different assignment of weights. The combinations we used were (1) features from Word-net-Affect only, (2) Word-net-Affect + Personal-pronoun + Unigram, (3) Word-net-Affect + Personal-pronoun + Unigram+ POS , (4) Word-net-Affect + Personal-pronoun + Unigram + POS-Bigram , (5) Word-net-Affect + Personal-pronoun + Unigram + Dependency Parsing (6) Word-net-Affect + Personal-pronoun + Unigram + Dependency Parsing +adjectives ,(7) Word-net-Affect + Personal-pronoun + Unigram + Dependency Parsing +adjectives with weight of Word-net-Affect features set to 2.0 ,(8) Word-net-Affect + Personal-pronoun + Unigram + Dependency Parsing +adjectives + weight of dependency parsing feature set to 2.0 , (9) all features (including the nonlexical features).

We achieved best results when all features were combined. Table 5 summarizes the results and figure 1 depicts it in graphical form.

Table 5: Emotion Classification Accuracy

Legend	Feature	Accur acy
A	Word-net-Affect only	68.16
B	Word-net-Affect + Personal-pronoun + Unigram	71.3

C	Word-net-Affect + Personal-pronoun + Unigram+ POS	72.02
D	Word-net-Affect + Personal-pronoun + Unigram + POS-Bigram	72.54
E	Word-net-Affect + Personal-pronoun + Unigram + Dependency Parsing	71.16
F	Word-net-Affect + Personal-pronoun + Unigram + Dependency Parsing +adjectives	71.44
G	Word-net-Affect + Personal-pronoun + Unigram + Dependency Parsing +adjectives with weight of Word-net-Affect features set to 2.0	72.92
H	Word-net-Affect + Personal-pronoun + Unigram + Dependency Parsing +adjectives with weight of Word-net-Affect features set to 2.0 + weight of dependency parsing feature set to 2.0	72.94
I	All	73.24

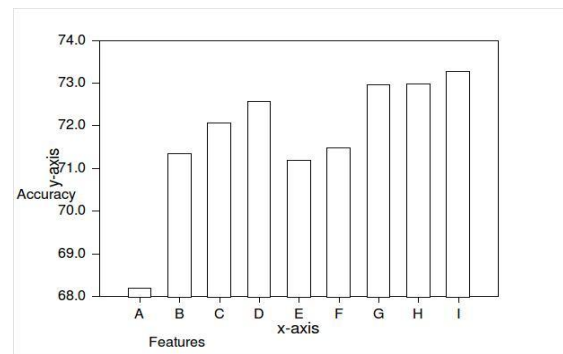


Fig. 1 Showing accuracy achieved using different features

Legends:

A: Word-net-Affect only

B: Word-net-Affect + Personal-pronoun + Unigram

C: Word-net-Affect + Personal-pronoun + Unigram+ POS

D: Word-net-Affect + Personal-pronoun + Unigram + POS-Bigram

E: Word-net-Affect + Personal-pronoun + Unigram + Dependency Parsing

F: Word-net-Affect + Personal-pronoun + Unigram + Dependency Parsing +adjectives



G: Word-net-Affect + Personal-pronoun + Unigram + Dependency parsing +adjectives with weight of Word-net-Affect features set to 2.0

H: Word-net-Affect + Personal-pronoun + Unigram + Dependency parsing +adjectives with weight of Word-net-Affect features set to 2.0 + weight of dependency parsing feature set to 2.0

I: All

We have also added the commonly used acronyms like lol , wtf, shit, rofl etc to the word-net-affect Lexicon to capture the slangs used on twitter.

5. CONCLUSION AND FUTURE WORK

We address the problem of identifying expressions of emotion in text. We describe the task of annotating sentences in a Tweet corpus. An annotation agreement study shows variation in agreement among judges for different emotion categories. We found the annotators to agree most in identifying instances of happiness and anger.

We also present the results of automatic emotion classification experiments, which utilized knowledge resources in identifying emotion-bearing words in sentences. The accuracy is 73.24%, significantly higher than previous work done on the same theme.

Table 6: Comparison of different approaches

Approach	Accuracy
Seol et al.(2008) used neural networks for an 8-way classification.	45-65 %
Chuang and Wu (2004) used SVM and manually defined keyword features over a seven-way classification consisting of the same six-class taxonomy plus a neutral category	Average accuracy of 65.5 %
Our Approach	Accuracy of 72.34

In future we would aim to use more features to the system and try to exploit more dependency relations. we will also use the annotated data for fine-grained classification of sentences on the basis of emotion categories. We also plan to incorporate methods for addressing the special needs of the kind of language used in on-line communication.

6. REFERENCES

- [1] Information gathered from http://svmlight.joachims.org/svm_multiclass.html
- [2] G. Mishne. Experiments with Mood Classification in Blog Posts, In Proceedings of the Style2005: The 1st Workshop on Stylistic Analysis of Text for Information Access, SIGIR 2005; Salvador, Brazil; Aug 15-19, 2005..
- [3] Y. Jung, H. Park and S. H. Myaeng, A Hybrid Mood Classification Approach for Blog Text, Lecture Notes in Computer Science, 4099, pp.1099-1103, 2006.
- [4] H. Liu, H. Lieberman and T. Selker, A model of textual affect sensing using real-world knowledge, In Proceedings of the 2003 international conference on intelligent user interfaces, pp.125- 132, 2003.
- [5] C. O. Alm, D. Roth and R. Sproat. Emotions from Text: Machine Learning for Text-based Emotion Prediction, In Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Vancouver, British Columbia, Canada, pp.579-586, 2005.
- [6] A. Neviarouskaya, H. Prendinger and M. Ishizuka, Textual Affect Sensing for Social and Expressive Online Communication, In Proceedings of the 2nd international conference on Affective Computing and Intelligent Interaction. pp.218-229, 2007.
- [7] C. M. Lee and S. S. Narayanan. Toward Detecting Emotions in Spoken Dialogs, Journal of the American Society for Information Science. IEEE Trans. on Speech and Audio Processing 13(2), pp.293-303, 2005.
- [8] Information from Social networking site: www.twitter.com
- [9] Quirk, R., Greenbaum, S., Leech, G., Svartvik, J.: A Comprehensive Grammar of the English Language. Longman, New York (1985)
- [10] Wiebe, J., Wilson, T., Cardie, C.: Annotating expressions of opinions and emotions in language. Language Resources and Evaluation 39(2-3), 165–210 (2005)
- [11] Martin,J.R..White,P.R.R.:The Language of Evaluation : Appraisal in English, Palgrave, London (2005), <http://grammatics.com/appraisal/>
- [12] Read, J., Hope, D., Carroll, J.: Annotating expressions of Appraisal in English. In: The Proc. of the ACL Linguistic Annotation,Workshop, Prague (2007)
- [13] Whitelaw, C., Garg, N., Argamon, S.: Using Appraisal Taxonomies for Sentiment Analysis. In: Proc. of the 2nd Midwest Comp., Linguistic Colloquium, Columbus (2005)
- [14] Alm, C.O., Roth, D., Sproat, R.: Emotions from text: machine learning for text-based emotion prediction. In: Proc. of the Joint Conf. on Human Language Technology/Empirical Methods in Natural Language Processing (HLT/EMNLP), pp. 579– 586 (2005)
- [15] Ekman, P.: An Argument for Basic Emotions. Cognition and Emotion. 6, 169–200 (1992)



- [16] Read, J.: Recognising affect in text using pointwise mutual information. Master's thesis, University of Sussex (2004)
- [17] Liu, H., Lieberman, H., Selker, T.: A Model of Textual Affect Sensing using Real-World Knowledge. In: Proc. of the Int'l Conf. on Intelligent User Interfaces (2003)
- [18] Neviarouskaya, A., Prendinger, H., Ishizuka, M.: Analysis of affect expressed through the evolving language of online communication. In: Proc. of the 12th Int'l Conf. on Intelligent
- [19] User Interfaces (IUI-07), Honolulu, Hawaii, pp. 278–281 (2007) Mihalcea, R., Liu, H.: A corpus-based approach to finding happiness. In: The AAAI Spring Symposium on Computational Approaches to Weblogs, Stanford, CA (2006)
- [20] Information gathered from <http://infolab.tamu.edu/resources>
- [21] Passonneau, R.J.: Measuring agreement on set-valued items (MASI) for semantic and pragmatic annotation. In: Proc. 5th Int'l Conf. on Language Resources and Evaluation (2006)
- [22] Information gathered from <http://nlp.stanford.edu/software/corenlp.shtml>
- [23] Information gathered from <http://tartarus.org/martin/PorterStemmer/>
- [24] Information gathered from <http://nlp.stanford.edu/software/stanford-dependencies.shtml>
- [25] Information gathered from <http://www-sers.york.ac.uk/~mb55/msc/clinimet/week4/kappash2.pdf>