Using a Heterogeneous Dataset for Emotion Analysis in Text

Conference Paper · May 2011

DOI: 10.1007/978-3-642-21043-3_8 · Source: DBLP

Soumaya Chaffar
University of Ottawa

Diana Inkpen
University of Ottawa

Using a Heterogeneous Dataset for Emotion Analysis in Text

Soumaya Chaffar and Diana Inkpen

School of Information Technology and Engineering, University of Ottawa
Ottawa, ON, Canada

{schaffar, diana}@site.uottawa.ca

Abstract. In this paper, we adopt a supervised machine learning approach to recognize six basic emotions (*anger*, *disgust*, *fear*, *happiness*, *sadness* and *surprise*) using a heterogeneous emotion-annotated dataset which combines news headlines, fairy tales and blogs. For this purpose, different features sets, such as bags of words, and N-grams, were used. The Support Vector Machines classifier (SVM) performed significantly better than other classifiers, and it generalized well on unseen examples.

Keywords: Affective Computing, Emotion Analysis in Text, Natural Language Processing, Text Mining.

1 Introduction

Nowadays the emotional aspects attract the attention of many research areas, not only in computer science, but also in psychology, healthcare, communication, etc. For instance, in healthcare some researchers are interested in how acquired diseases of the brain (e.g., Parkinson) affect the ability to communicate emotions [10]. Otherwise, with the emergence of Affective Computing in the late nineties [11], several researchers in different computer science areas, e.g., Natural Language Processing (NLP), Human Computer Interaction (HCI), etc. are interested more and more in emotions. Their aim is to develop machines that can detect users' emotions and express different kinds of emotion. The most natural way for a computer to automatic emotion recognition of the user is to detect his emotional state from the text that he entered in a blog, an online chat site, or in another form of text.

Generally, two approaches (knowledge-based approaches and machine learning approaches) were adopted for automatic analysis of emotions in text, aiming to detect the writer's emotional state. The first approach consists of using linguistic models or prior knowledge to classify emotional text. The second one uses supervised learning algorithms to build models from annotated corpora. For sentiment analysis, machine learning techniques tend to obtain better results than lexical-based techniques, because they can adapt well to different domains [7]. In this paper, we adopted a machine learning approach for automatic emotion recognition from text. For this purpose, we used a heterogeneous dataset collected from blogs, fairly tales and news headlines.

The rest of the paper is organized as follows: Section 2 identifies the several datasets that we used for our emotion detection in text. In Section 3, we describe the methodology that we adopted for this purpose. Section 4 presents and discusses the results by comparing different machine learning techniques for detecting emotion in texts. Finally, Section 5 concludes the paper and outlines the future direction of our research.

2 Datasets

Five datasets have been used in the experiments reported in this paper. We describe each one in details below.

2.1 Text Affect

This data consists of news headlines drawn from the most important newspapers, as well as from the Google News search engine [12] and it has two parts. The first one is developed for the training and it is composed of 250 annotated sentences. The second one is designed for testing and it consists of 1,000 annotated sentences. Six emotions (*anger*, *disgust*, *fear*, *joy*, *sadness* and *surprise*) were used to annotate sentences according to the degree of emotional load. For our experiments, we further use the most dominant emotion as the sentence label, instead of a vector of scores representing each emotion.

2.2 Neviarouskaya et al.'s Dataset

Two datasets produced by these authors were used in our experiments [8, 9]. In these datasets, ten labels were employed to annotate sentences by three annotators. These labels consist of the nine emotional categories defined by Izard [8] (*anger*, *disgust*, *fear*, *guilt*, *interest*, *joy*, *sadness*, *shame*, and *surprise*) and a *neutral* category. In our experiments, we considered only sentences on which two annotators or more completely agreed on the emotion category. We briefly describe in the following the two datasets.

- *Dataset 1*

This dataset includes 1000 sentences extracted from various stories in 13 diverse categories such as education, health, and wellness [8].

- *Dataset 2*

This dataset includes 700 sentences from collection of diary-like blog posts [9].

2.3 Alm's Dataset

This data include annotated sentences from fairy tales [1]. For our experiments, we used only sentences with high annotation agreement, in other words sentences with four identical emotion labels. Five emotions (*happy*, *fearful*, *sad*, *surprised* and *angry-disgusted*) from the Ekman's list of basic emotions were used for sentences annotations. Because of data sparsity and related semantics between *anger* and *disgust*, these two emotions were merged together by the author of the dataset, to represent one class.

2.4 Aman's Dataset

This dataset consists of emotion-rich sentences collected from blogs [3]. These sentences were labelled with emotions by four annotators. We considered only sentences for which the annotators agreed on the emotion category. Ekman's basic emotions (*happiness*, *sadness*, *anger*, *disgust*, *surprise*, and *fear*), and also a *neutral* category were used for sentences annotation.

3 Emotion detection in text

To find the best classification algorithm for emotion analysis in text, we compared the three classification algorithms from the *Weka* software [14] with the BOW representation: J48 for Decision Trees, Naïve Bayes for the Bayesian classifier and the SMO implementation of SVM.

To ensure proper emotional classification of text, it is essential to choose the relevant feature sets to be considered. We describe in the following the ones that we employed in our experiments:

- Bag-Of-Words (BOW)

Each sentence in the dataset was represented by a feature vector composed of Boolean attributes for each word that occurs in the sentence. If a word occurs in a given sentence, its corresponding attribute is set to 1; otherwise it is set to 0. BOW considers words as independent entities and it does not take into consideration any semantic information from the text. However, it performs generally very well in text classification.

- N-grams

They are defined as sequences of words of length *n*. N-grams can be used for catching syntactic patterns in text and may include important text features such as negations, e.g., "not happy". Negation is an important feature for the analysis of emotion in text because it can totally change the expressed emotion of a sentence. For instance, the sentence "I'm not happy" should be classified into the *sadness* category and not into *hapiness*. For these reasons, some research studies in sentiment analysis claimed that N-grams features improve performance beyond the BOW approach [4].

- Lexical emotion features

This kind of features represents the set of emotional words extracted from affective lexical repositories such as, WordNetAffect [13]. We used in our experiments all the emotional words, from the WordNetAffect (WNA), associated with the six basic emotions.

4 Results & Discussion

For an exploratory purpose, we conducted several experiments using the labelled datasets for classifying emotional sentences.

4.1 Cross-validation

First of all, it is important to prepare the data for proper emotional sentence classification. For classifying text into emotion categories, some words such as “I” and “the” are clearly useless and should be removed. Moreover, in order to reduce the number of words in the BOW representation we used the *LovinsStemmer* stemming technique from the *Weka* tool [14], which replaces a word by its stem.

Another important way for reducing the number of words in the BOW representation is to replace negative short forms by negative long forms, e.g., “don’t” is replaced by “do not”, “shouldn’t” is replaced by “should not”, and so on. Applying this method of standardizing negative forms gave us better results for BOW representation and can consider effectively negative expressions in N-grams. In this later, the features include words, bigrams and trigrams.

In the spirit of exploration, we used five datasets to train supervised machine learning algorithms: Text Affect, Alm’s dataset, Aman’s dataset and the Global dataset (see Table 2). We also used the ZeroR classifier from Weka as a baseline; it classifies data into the most frequent class in the training set.

Table 1. Results for the training datasets using the accuracy rate (%)

	Baseline	***Naive Bayes***	***J48***	***SMO***
Text Affect	31.6	**39.6**	32.8	**39.6**
Alm’s Dataset	36.86	54.92	47.47	**61.88**
Aman’s Dataset	68.47	73.02	71.43	**81.16**
Global Dataset	50.47	59.72	64.70	**71.69**

The results presented in Table 2 show that in general the SMO algorithm has the highest accuracy rate for each dataset. The use of the global dataset for the training is much better, because, on one hand it contains heterogeneous data collected from blogs, fairly tales and new headlines, and on the other hand the difference between accuracy rates for the SMO algorithm and the baseline is higher compared to Aman’s dataset. With the global dataset, SMO is statistically better than the next-best classifier (J48) with a confidence level of 95% based on the accuracy rate (according to a paired t-test).

Specifically, for Aman's dataset, we achieved an accuracy rate of 81.16%, which is better than the highest accuracy rate (73.89%) reported in [2]. Compared to their work, we used not only emotional words, but also non-emotional ones, as we believe that some sentences can express emotions through underlying meaning and depending on the context, i.e., "Thank you so much for everyone who came". From the context, we can understand that this sentence expresses *happiness*, but it does not include any emotional word.

4.2 Supplied test set

Given the performance on the training datasets, one important issue that we need to consider in emotion analysis in text is the ability to generalize on unseen examples, since it depends on sentences' context and the vocabulary used. Thus, we tested our model (trained on the global dataset) on the three testing datasets using three kinds of feature sets (BOW, N-grams, emotion words from WordNetAffect). The results are presented in Table 3 below.

Table 2. SMO results using different feature sets.

Test sets	**Feature sets**	**Accuracy rate (%)**	
		baseline	***SMO***
Text Affect	*WNA*	36.20	36.55
	BOW		38.90
	BOW +WNA		36.55
	N-grams		**40.30**
Neviarouskaya et al.'s dataset 1	*WNA*	24.73	44.76
	BOW		**57.81**
	BOW +WNA		56.28
	N-grams		49.47
Neviarouskaya et al.'s dataset 2	*WNA*	35.89	48.91
	BOW		**53.45**
	BOW +WNA		52.56
	N-grams		50.69

As shown in Table 3, using the N-grams representation for Text Affect gives better results than the BOW representation, but the difference is not statistically significant. However, the use of N-grams representation for Neviarouskaya et al.' datasets decreased the accuracy rate compared to the BOW representation. As we notice from the table, using features sets from WordNetAffect did not help in improving the accuracy rates of the SMO classifier.

5 Conclusion

In this paper, we presented a machine learning approach for automatic emotion recognition from text. For this purpose, we used a heterogeneous dataset collected from blogs, fairly tales and new headlines, and we compared it to using each homogenous dataset separately as training data. Moreover, we showed that the SMO algorithm made a statistically significant improvement over other classification algorithms, and that it generalized well on unseen examples.

Acknowledgments.

We address our thanks to the **Natural Sciences and Engineering Research Council (NSERC)** of Canada for supporting this research work.

References

1. Alm, C. O. (2008). *Affect in Text and Speech*. PhD Dissertation. University of Illinois at Urbana- Champaign.
2. Aman, S. & Szpakowicz. S. (2007). Identifying Expressions of Emotion in Text. *Lecture Notes in Computer Science*. 4629: 196--205.
3. Aman, S. (2007). Identifying Expressions of Emotion in Text, Master's thesis, University of Ottawa, Ottawa, Canada.
4. Arora, S., Mayfield, E., Penstein-Ros, C. & Nyberg, E. (2010). Sentiment Classification using Automatically Extracted Subgraph Features. *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*.
5. Ekman, P. & Friesen, W. V. (1978) *Facial action coding system: Investigator's guide*, Palo Alto, Consulting Psychologists Press.
6. Izard, C. E. (1971). The Face of Emotion. Appleton-Century-Crofts, New York.
7. Melville, P. Gryc, W. & Lawrence, R. (2009). Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification. In*Proc. of KDD*, pages 1275–1284.
8. Neviarouskaya, A., Prendinger, H. & Ishizuka, M. (2010). @AM: Textual Attitude Analysis Model. *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, Los Angeles, USA.
9. Neviarouskaya, A., Prendinger, H. & Ishizuka, M. (2009). Compositionality Principle in Recognition of Fine-Grained Emotions from Text. *Proceedings of the International Conference on Weblogs and Social Media*, AAAI, San Jose, US.
10. Paulmann, S. & Pell, M.D. (2010). Dynamic emotion processing in Parkinson's disease as a function of channel availability. *Journal of Clinical and Experimental Neuropsychology, 32* (8), 822-835.
11. Picard, R. W. (1997). *Affective Computing*, Cambridge, MA: MIT Press.
12. Strapparava, C., & Mihalcea, R. (2007). Semeval- 2007 task 14: Affective text. In Proceedings of the 4th International Workshop on the SemEval 2007, Prague.
13. Strapparava, C., Valitutti, A., & Stock, O. (2006). The affective weight of lexicon. In Proceedings of the Fifth International Conference on Language Resources and Evaluation, Genoa, Italy.
14. Witten, I. H., & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques. 2nd Edition, Morgan Kaufmann, San Francisco.