Active Learning for Segment-based Emotion Analysis System

Guanhong Zhang  
Department of Computer Science and Technology, Hefei University, Hefei, Anhui, China

Odbal*  
Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei, Anhui, China  
*Corresponding author(E-mail: xishazgh@163.com)

Abstract
Emotion detection is one of the challenging tasks in the fine-grained sentiment analysis fields. Traditional approaches focus on supervised learning with features based on surface word forms, without linguistic structure data, existing sparse problems and cannot mine implicit information including collocation and semantic prosody, which result in poor performance, especially for identifying implicit emotion expression. To improve the performance of emotion detection system, this paper deals with assigning Ekman’s six basic emotion tags to sentences in Chinese written text. Baseline supervised system is developed based on segment based unit, and an active learning strategy using unlabeled data is presented to improve the performance of baseline supervised system. Experimental evaluation shows that our active learning procedure helps to improve the classification accuracy.

Key words: Emotion Analysis, Segment-based Unit, Active Learning

1. INTRODUCTION

Emotion analysis aims to identify the emotion categories (e.g., happy, angry, sadness and surprise) of a given written text. For example, the Chinese sentence “xiao meimei yi ting ku le qi lai, bian ku bian shuo: "nan dao ni men jiu bu neng de jiu ma?"” (The little sister wept and said, "Can you not be set free?"). which emotion category is “sadness”. Labeling the written text with their emotion categories has drawn the attention of the natural language processing research community due to its many applications.

Recently the most widely used methods for emotion analysis are supervised learning algorithms, such as Naive Bayes (NB), Support Vector Machine (SVM) and Condition Random Fields (CRF) etc. Supervised-based models (Sunita Sarawagi, 2004; Karl Moritz Hermann, 2013) usually require hand-labeled data in order to achieve high-quality performance. However, the scarcity of hand-labeled data seriously compromises the development of automatic emotion analysis systems (EAS). One way of improving the performance of EAS (Richard Socher, 2012) is to use extra annotation, typically unlabeled data, which are nowadays pervasive in digital format and relatively easy and inexpensive to collect (e.g., documents off the web).

Unfortunately, the application of either post-editing or EAS to unlabeled data with massive data volumes is still too expensive, simply because manual supervision of all instances requires huge amounts of manpower. For such massive data streams the need of employing active learning (AL) is compelling. AL techniques (Jeff Mitchell, 2008) for EAS selectively ask a human annotator to supervise a small portion of the incoming sentences. Sentences are selected so that EAS models estimated from them classify emotion category as accurately as possible.

This paper describes an active learning algorithm for EAS. The model has been extended to incorporate segment-based emotion analysis as presented here. As the unlabeled data size becomes huge in a practical setting, and the active learning considers the amount of manual annotation for each sentence, the sentence sampling techniques are necessary. We tested our model in a Chinese emotion analysis system, and obtained better results than SVM. We also implement active learning for our model. The experimental results show that our active learning consistently improves the performance of the segment-based model in all test cases.

2. RELATED WORK

Recognizing emotions in texts has grown popular among researchers in recent years. The most common technology used to improve the performance of EASs supervised learning with more information from the natural language, which usually lexical, syntactic and semantic cues. Alm et al. (2005) explored the text-based emotion prediction problem empirically, using supervised machine learning. Das and Bandyopadhyay (2009)
assigned emotion tags on the Bengali blog sentences using CRF. Chaffar and Inkpen (2011) adopted a supervised 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. Purver and Battersby (2012) described a set of experiments using automatically labeled data to train supervised classifiers for multi-class emotion detection in Twitter messages with no manual intervention. Saif Mohammad (2012) used word-level affect lexicons to provide significant improvements in sentence-level emotion classification. There are also studies that analyzed the deeper level information, such as color-concept-emotion associations (Volkova et al., 2012), emotion causes detection (Chen et al., 2010) and learning hashtags to improve emotion classification performance (Qadir and Riloff, 2013).

Another group of related studies for improving the performance of EASIs to use the extra annotation data, typically unlabeled data. Active learning strategy can efficiently use large amount of unlabeled data, together with the labeled data, to build better sentiment classifiers (Smajlovic et al., 2014; Kranjc et al., 2015). To the best of our knowledge, very few works considered AL for emotion analysis. In contrast to sentiment analysis, emotion analysis belongs to multi-class problem that as we increase the number of classes, building an accurate classifier becomes more and more difficult.

In this work, the AL frameworks is extended in an effort to improve the performance of EAS. In short, we propose an AL framework for EAS that can select the most uncertain sentence which results are could be incorrect. Our baseline system allows us to have more contexts to model the changing probability distribution of the sentence and results in a more accurate sampling of the changing pool of sentences.

3. EMOTION ANALYSIS SYSTEM

Our baseline system was described in (Odbal and Wang, 2014), called segment-based emotion analysis model, which analyzes the emotional state of the Chinese sentence by applying semi-Markov conditional random fields (semi-CRFs) at each segment units. Given a sequence of observations $x = x_1^J = < x_1, \ldots, x_J >$ to be labeled with a corresponding emotion class $y$, where $y$ is one of the Ekman’s six basic emotion types such as happiness, sadness, fear, surprise, anger and disgust etc. The fundamental equation of segment-based EAS is defined as follows:

$$
p(y | x) = \sum_{x_1 \ldots y_j} p(s_1^K | x) \cdot p(y_j^K | x_1^K, x) \cdot p(y | y_1^K, s_1^K, x)$$

As shown in the above formula (1), the segment-based EAS based on the semi-CRFs model consists of three processes (as shown in Figure 1): at first, a sentence or short text is divided into non-fixed length segments according to dependency grammar. We construct segment units from the dependency parse tree of each sentence, and then build up possible segment candidates based on those units. More specifically, the dependency subtrees that contain the path from the root node (e.g., core verb “zhao gu” (take care of)) to leaf node are selected for the candidate segmentation. For instance, let us consider the subjective sentence "shan liang de gu liang xi xin de zhao gu zhe zhi ruo xiao de mao” (Good girl carefully take care of the small cat) (as shown in Figure 2). We can select four dependency subtrees: (“shan liang de gu liang” “zhao gu”) (good girl, take care of), (“xi xin de”, “zhao gu”) (carefully, take care of), (“zhao gu”, “zhe zhi”, “mao”) (take care of, the cat), and (“zhao gu”, “ruo xiao de mao”) (take care of, the small cat) as the candidate segmentations.

In the next process, the segmentation strings as an observation inputs. The semi-CRF model utilized the context-informed features, such as POS tags, emotion word lists and context-informed dependency relations etc, to assess the emotion classes of each segment. That is, instead of determining $y$ directly from $x$, we introduce hidden variables $z = (z_1, \ldots, z_m)$ as intermediate decision variables, where $z_i = (s_i, y_i)$ and $y_i \in \{happiness, sadness, fear, surprise, anger, disgust, none\}$, so that $y_i$ represents whether $s_i$ is a phrase with happiness, sadness, fear, surprise, anger, or disgust, or none of the above. In the above example, we can obtain the emotion label of each segment $y = \{\text{happy, happy, happy, sadness}\}$.

At last, once the intermediate decision variables are decided, a probabilistic model based on log linear model is utilized to combine segment-level emotion categories.

For simplicity, the probability $p(y | x)$ can be introduced by two probability distribution models: segment-level emotion detection model and emotion tag distribution model. Specifically, the discriminate function can be defined as follows:
\[ p(y \mid x) = \sum_{s, y_1^K} p(y_1^K \mid x) \cdot p(y \mid y_1^K) \]  
\[ (2) \]

There are two probability distributions:
- \( p(y, s \mid x) \) indicates segment-level emotion detection model, and this model describes the distribution of the sequence of segmentation \( s_1(1:k) \) and its emotion tag \( y_1(1:k) \). This distribution can be calculated directly by the semi-CRF model.
- \( p(y \mid y_1^K) \) indicates emotion tag distribution model, and this model describes the probability distribution of the emotion classes. Where \( y_1^K \) is segment-level emotion tag and \( y \) represents the overall emotion tag. This distribution can be calculated by similar 1-gram model.

In this study, we use the maximum a posteriori estimation with Gaussian priors for parameter estimation, and the Viterbi algorithm for the inference problem.

![Graphical presentation for semi-CRF segment based model](image1.png)

**Figure 1.** Graphical presentation for semi-CRF segment based model

![Dependency parse tree example](image2.png)

**Figure 2.** A dependency parse tree example There are four segment units in this sentence

### 4. ACTIVE LEARNING FOR EAS

The aim of the EAS is to obtain high-quality classifications while minimizing the required human effort. Despite the fact that EAS may reduce the required effort with respect to post-editing, it still requires the user to supervise all the sentences. To address this problem, we propose to use AL techniques to select only a small number of sentences whose results are worth to be supervised by the human expert.

In this section we describe two different AL algorithms used to retrain a segment-based EAS model for improving the classification performance based on exploitation of unlabeled data. For all the algorithms described, we assume the following premises: (1) A small set of labeled data \( L \) exists, where—as above—\( L = \{ (x_1, y_1), \ldots, (x_j, y_j) \} \), \( x_i \) is a \( d \)-dimensional feature vector \( x_i \in \mathbb{R}^d \), and \( y_i \) is the label for each set of data; (2) a large set of unlabeled data \( U \) is available, where \( U = \{ x_1', \ldots, x_u' \} \) and \( u \geq 1 \) and \( x_j' \) is a \( d \)-dimensional feature vector; and (3) at each iteration, a subset of \( n \) instances is selected from \( F \) for labeling (by a human annotator), where \( F \) is sentence sampling function.

Algorithm 1 illustrates the basic algorithm to implement AL for EAS. This algorithm starts by classifying all sentences of the unlabeled data pool \( U \) using the segment-based EAS model previously trained on the labeled data \( L \). Then, if this initial emotion label is classified as incorrect or “worth of supervision”, the human annotator is required to label it. If not, we directly return the initial automatic emotion label and no effort is required from the human annotator. AL selects the incorrect or “worth of supervision” sentences according to sentence sampling function. At the end of the process, we use the new sentence-label pair \((x, y)\) available to refine the segment-based models used by the EAS system. In this scenario, the human annotator only checks a small number of sentences, thus, final emotion labels are not error-free as in conventional EAS.

The sentence sampling function is a primary part of the active learning process. A good sentence sampling strategy must be able to select those sentences that along with their correct emotion labels improve most the performance of the EAS model. To do that, the sampling strategy has to correctly discriminate “informative” sentences from those that are not. We can make different approximations to measure the informativeness of a given sentence. This paper describes two different sampling strategies tested in our experimentation.
(1) Random sampling

Arguably, the simplest sampling approach is random sampling, where the sentences are randomly selected to be interactively annotated. Although simple, it turns out that random sampling perform surprisingly well in practice. In this method, a few sentences are picked at random and get their labels by manually in each cycle, and the new labeled data are added for retraining the current EAS models.

(2) The most uncertainty sampling

Another technique is to consider that the most informative sentence is the one the current EAS model generates worst. We apply segment-based EAS model as our base learner throughout this paper and employ a utility function which is based on the conditional probability of the most likely label sequence $y'$ for an observation sequence $x$:

$$Fs(x) = 1 - P(y' | x)$$ (3)

Sequences for which the current EAS model is least confident on the most likely emotion label sequence are preferably selected. These selected sentences are fully manually annotated, and then would be added to previous train data to retrain the EAS model.

Algorithm 1 Pseudo-code of the proposed algorithm to implement AL for EAS from unlabeled data

Given:
- U: Unlabeled Data Set
- L: Labeled Data Set
- C: a supervised learning classifier
- Fs: sentence sampling function

Repeat:
1. use L to learn a classifier CA
2. use CA to label the unlabeled data U
3. select most confidently predicted subset Ua according to sentence sampling function Fs
4. submit the selected subset Ua to human annotation
5. remove Ua from unlabeled data, U=U-Ua
6. add Ua to the labeled data, L=L+Ua

5.EXPERIMENTS AND RESULTS

5.1.Data Construction

This subsection simply explains the datasets and lexicons used in our experiments. These experimental datasets are created at our previous work (Odbal and Wang, 2014). Table 1 shows the details of these datasets.

<table>
<thead>
<tr>
<th></th>
<th>Entries</th>
<th>Average Length</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese emotion lexicon</td>
<td>1810</td>
<td>---</td>
<td>This emotion lexicon is created based on the English WordNet Affect lists (Strapparava et al., 2004), and it includes Ekman’s six basic emotion types (happy, fearful, sad, surprised, angry and disgusted). In the production process of the lexicon, the English-Chinese bilingual dictionary is used to translate the English entries into Chinese one.</td>
</tr>
<tr>
<td>Fairy tales dataset</td>
<td>1223</td>
<td>34.76</td>
<td>This data set is based on Alm’s dataset (Alm et al., 2005), and it includes annotated sentences from fairy tales, and annotate six emotion tags from the Ekman’s basic emotion state.</td>
</tr>
<tr>
<td>News dataset</td>
<td>1135</td>
<td>27.09</td>
<td>This corpus is created manually by two annotators. Each annotator marks the sentence with one or two of six primary emotions, and then calculates the kappa value to assess such reliability regarding emotion categories with a value of 0.7 or above it indicating complete agreement.</td>
</tr>
<tr>
<td>Unlabeled corpora</td>
<td>4000</td>
<td>---</td>
<td>We downloaded additional unlabeled corpora, which contains 2000 sentences from Chinese version of Andersen’s and Green’s fairy tales, and 1000 sentences from news domain.</td>
</tr>
</tbody>
</table>
5.2. Experimental Results of Supervised Model

As described in Section 3, we use segment-based Semi-CRFs as the modeling paradigm for evaluating the various machine learning algorithms, and we include the following features for detecting emotional state in written text. Stanford toolkits are employed for Chinese segmentation, part-of-speech tagging and dependency parsing.

(1) Bag-of-words (BOW): Surface forms of word unigrams and bigrams in the sentence are used as features.

(2) Part-of-speech (POS): The part-of-speech of the current word and the surrounding words are used a feature for emotion classification.

(3) Content bag-of-words (contentBOW): N (noun), V (verb), JJ (adjective) words by POS is used as features.

(4) Emotion word lists (Emotion): the features are based on the emotion-word itself. The emotion class of a word can be assigned as the word’s prior emotion according to the Chinese emotion lexicon, which is a translation and extension version of WordNet-Affect lexicon.

(5) Dependency relations (Dependency): the features are binary indicators of whether the leaf phrase in the dependency parse tree belongs to one of the emotion classes. The dependencies are all binary relations: a grammatical relation holds between a governor (head) and a dependent (modifier).

In terms of performance evaluation, we use accuracy as the primary performance measure. Table 2 shows the accuracy result of our segment-based model with five different kinds of features (BOW, contentBOW, POS, Emotion and Dependency) and their combination features, setting 10-fold cross validation as a testing option.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Features</th>
<th>SVM</th>
<th>Our approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>news dataset</td>
<td>BOW</td>
<td>46.84</td>
<td>53.33</td>
</tr>
<tr>
<td></td>
<td>contentBOW</td>
<td>48.59</td>
<td>55.74</td>
</tr>
<tr>
<td></td>
<td>contentBOW+POS</td>
<td>48.64</td>
<td>56.69</td>
</tr>
<tr>
<td></td>
<td>contentBOW+Emotion</td>
<td>51.46</td>
<td>58.55</td>
</tr>
<tr>
<td></td>
<td>contentBOW+Emotion+POS</td>
<td>50.2</td>
<td>61.53</td>
</tr>
<tr>
<td></td>
<td>contentBOW+Emotion+POS+Dependency</td>
<td>54.45</td>
<td>65.12</td>
</tr>
<tr>
<td>fairy tales dataset</td>
<td>BOW</td>
<td>39.59</td>
<td>40.09</td>
</tr>
<tr>
<td></td>
<td>contentBOW</td>
<td>39.98</td>
<td>42.11</td>
</tr>
<tr>
<td></td>
<td>contentBOW+POS</td>
<td>40.19</td>
<td>43.95</td>
</tr>
<tr>
<td></td>
<td>contentBOW+Emotion</td>
<td>45.86</td>
<td>46.49</td>
</tr>
<tr>
<td></td>
<td>contentBOW+Emotion+POS</td>
<td>46.15</td>
<td>48.95</td>
</tr>
<tr>
<td></td>
<td>contentBOW+Emotion+POS+Dependency</td>
<td>48.23</td>
<td>50.81</td>
</tr>
</tbody>
</table>

As shown in Table 2, we can observe that our approach has better results than SVM, and using the combination features of contentBOW+Emotion+POS+Dependency has the highest accuracy rate for each dataset and each classifier. When the baseline system use the contentBOW features, the POS, Emotion and Dependency representation improve the accuracy rates of our classifier for each dataset. For example, on news dataset, segment-based model with contentBOW has the accuracy rate of 55.74% and adding emotion words has the accuracy rate of 58.55%, showing the improvements of 2.81%. For the experimental datasets in different fields, overall performances on the news dataset are better than on the fairy tales dataset. The reasons perhaps are that the syntactic structures of the sentences from fairy tales dataset are less restricted and highly variable.

5.3. AL Performance

The following baseline methods are implemented in order to compare the effectiveness of proposed AL algorithm:

(1) Active learning with random sampling: in each cycle, t examples were randomly selected from unlabeled data for human labeling and added to the training data with corresponding labels.

(2) Active learning with uncertainty sampling: this model is based on the uncertainty sampling approach. The uncertainty value is obtained according to a utility function, as shown in formula (3).

Figure 3 shows the classification accuracy of various active learning based methods on two different evaluation datasets. As shown in this figure, by comparing active learning with uncertainty sampling with random sampling, the classification accuracy of the uncertainty sampling strategy improved very significantly in each cycle (especially in the tale dataset). This was because the sentences selected based on uncertainty were more representative than those sentences selected based on random in active learning. The results presented in this figure also showed that the decrease point at random sampling happens at random cycle. But the results of uncertainty sampling are not the case. This was most likely due to the augmentation of the most confident
automatic classified examples, along with the manually labeled examples, into the training data during the learning process.

![Figure 3](image)

**Figure 3.** Classification accuracy over the number of manually labeled examples on two different datasets

### 6. Conclusion

In this work, the AL framework is extended in an effort to improve the performance of fine-grained emotion analysis system. Baseline supervised system is developed based on segment based unit, and an active learning algorithm with different sentence sampling strategies for using unlabeled data to improve the performance of baseline supervised learning algorithm is presented. Two different datasets, which contains news content and fairly tales, are used to test our proposed model, and the experimental results show that our approach make a statistically significant improvement over the classification algorithms, reflecting its potential usage in the emotion detection task.

### Acknowledgements

This work is supported by the grant of Anhui Provincial Natural Science Foundation (No. 1308085QF115, No.1508085QF116), Training Object Project for Academic Leader of Hefei University (No. 2014dtr08), Key Constructive Discipline Project of Hefei University (No. 2014xk08).

### REFERENCES


Ashequl Qadir, Ellen Riloff. (2013) “Bootstrapped Learning of Emotion Hashtags”, Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp.2-11.


