Joint Learning on Sentiment and Emotion Classification

Wei Gao[†] Shoushan Li[†‡*] Sophia Yat Mei Lee[‡] Guodong Zhou[†] Chu-Ren Huang[‡]

[†]Natural Language Processing Lab
Soochow University, Suzhou, China
{wei.gao512, shoushan.li } @gmail.com,
gdzhou@suda.edu.cn

[‡] Department of Chinese and Bilingual Studies
The Hong Kong Polytechnic University, Hong Kong
{sophiaym, churenhuang} @gmail.com

ABSTRACT

Sentiment and emotion classification have been popularly but separately studied in natural language processing. In this paper, we address joint learning on sentiment and emotion classification where both the labeled data for sentiment and emotion classification are available. The objective of this joint-learning is to benefit the two tasks from each other for improving their performances. Specifically, an extra data set that is annotated with both sentiment and emotion labels are employed to estimate the transformation probability between the two kinds of labels. Furthermore, the transformation probability is leveraged to transfer the classification labels to benefit the two tasks from each other. Empirical studies demonstrate the effectiveness of our approach for the novel joint learning task.

Categories and Subject Descriptors

I.2.7 [**Artificial Intelligence**]: Natural Language Processing – *Text analysis*

General Terms

Algorithms, Experimentation

Keywords

Joint Learning, Sentiment Classification, Emotion Classification.

1. INTRODUCTION

Sentiment and emotion classification are two popular tasks in Natural Language Processing (Pang et al., 2002; Das and Bandyopadhyay, 2009). Specifically, sentiment classification aims to predict the sentimental orientation (e.g., *positive* or *negative*) of a text while emotion classification aims to recognize basic emotions in the text, such as *expect*, *joy*, *love*, *surprise*, and *hate*. Both of them have attracted considerable interests due to their wide applications in natural language processing (Liu, 2012; Wilson, et al., 2009; Chen et al., 2010).

Although sentiment classification and emotion classification are two different tasks and separately studied in the literature, these two tasks are highly related to each other due to the strong relationship between the sentiment and emotion labels. For

example, a text with positive sentiment is more likely to be accompanied by a positive emotion, such as joy, love, and expect. Consider the following sentence, denoted as E1, from the Ren-CECps corpus (Quan and Ren, 2009). The emotion of this sentence is joy and the sentiment label is positive.

E1: *现在好了，我终于如愿以偿。*(*It's fine now. I have finally realized my goal*.)

Therefore, it is both interesting and practical to jointly learn the two tasks, i.e., sentiment and emotion classification, given the fact that each task often has a limited amount of training data. Here, joint learning means to make full use of all training data from both tasks to help each other to improve the classification performance.

Intuitively, it would appear that we could simply map the emotion labels to sentiment ones. For example, if one sample is predicted as love, it would be a positive sample. However, the emotion and sentiment labels are not always mapping. Considering another sentence from the Ren-CECps corpus:

E2: *我还是会傻傻的,忍不住替你操心,你和她会幸福吗？*(*I could not help but worry in a silly way, whether you would be happy to be with her?*)

The emotion of this sentence is love while its sentimental label is negative. Moreover, if one sample is predicted as positive, it is hard to directly map it to either joy or love. Therefore, directly mapping the emotion and sentiment labels is incapable of handling this problem.

In this paper, we propose a joint learning approach for sentiment and emotion classification. The basic idea of the approach is to annotate a small size of data where the samples are annotated with both sentiment and emotion labels. This data set, which we refer to as sharing data, is then employed to estimate the strength of statistical relationship, i.e., transformation probabilities, between the sentiment and emotion labels. Specifically, in the training phase, the two individual labeled data sets are used to train two separate classifiers. In the testing phase, one sample is classified by both classifiers and obtains the probabilities belonging to each sentiment or emotion label. Then the statistical relationship is leveraged to transfer the sentiment (or emotion) labels to the emotion (or sentiment) labels. The final label of the sample is determined by combining the original sentiment (or emotion) label and the transferred one.

The remainder of this paper is organized as follows. Section 2 overviews the related work on both sentiment and emotion classification. Section 3 proposes our approach to jointly learning sentiment and emotion classification. Section 4 presents the experimental results. Finally, Section 5 gives the conclusion and future work.

2. RELATED WORK

2.1 Sentiment Classification

Sentiment classification has become a hot research topic in NLP community and various kinds of classification methods have been proposed, such as unsupervised learning methods (Turney, 2002), supervised learning methods (Pang et al., 2002), semi-supervised learning methods (Li et al., 2010), and cross-domain classification methods (He et al., 2011). Generally, sentiment classification can be performed on four different levels: word level (Wiebe, 2005), phrase level (Wilson et al., 2009), sentence level (Kim and Hovy, 2004; Liu et al., 2005), and document level (Turney, 2002; Pang et al., 2002; Riloff et al., 2006). This paper focuses on sentence-level sentiment classification with supervised learning approaches.

2.2 Emotion Classification

Recently, emotion classification has become more and more popular in NLP (Xu et al., 2010; Purver and Battersby, 2012). One main group of related studies on this task is about emotion resource construction, such as emotion lexicon building (Xu et al., 2010) and sentence-level or document-level corpus construction (Quan and Ren, 2009). Another main group of related studies is about the supervised learning approaches on emotion classification (Alm et al., 2005; Chen et al., 2010; Purver and Battersby, 2012).

Although both sentiment and emotion classification have been widely studied, almost all previous studies consider these two tasks separately. Unlike all of them, our work focuses on a new task of how to benefit sentiment and emotion classifications from each other. To the best of our knowledge, this is the first attempt to jointly learn the two tasks.

3. OUR APPROACH

3.1 Problem Statement

In a standard supervised classification problem, we seek a predictor f that maps an input vector x to the corresponding class label y. The predictor is trained on a finite set of labeled examples $\{(X_i, Y_i)\}$ ($i=1,\ldots,n$) and aims to minimize expected error, i.e.,

$$f = \arg\min_{f \in \mathbf{H}} \sum_i^n L(f(X_i), Y_i) \qquad (1)$$

Wherein L is a prescribed loss function and H is a set of functions called the hypothesis space, which consists of functions from x to y. In sentiment classification, the input vector of one document is constructed from weights of terms. The terms $(t_1,\ldots,t_N)$ are possibly words, word n-grams, or even phrases extracted from the training data, with N being the number of terms. The output label y has a value of 1, -1 or 0 representing positive, negative or neutral respectively. In emotion classification, the same input vectors are used as the ones in sentiment classification. But the output labels are different: in emotion classification, the output label y has more values, representing different kinds of emotions, such as joy, love, expect, hate, sorrow, anger, surprise, anxiety, and neutral (without any emotion). In previous studies, these two tasks are performed individually, as shown in Figure 1.
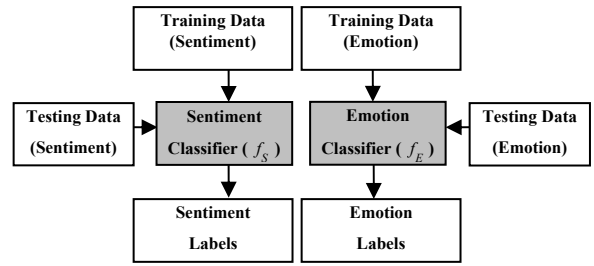

Figure 1: The framework of sentiment and emotion classifications that are performed individually

However, given the limited training data in each task, it is better to jointly learn the two tasks to benefit them from each other. Our approach to perform joint learning is to leverage both the sentiment and emotion classifier to classify the testing samples for sentiment or emotion classification, given the fact that they are sharing the same input vectors. In this way, we obtain both the emotion and sentiment labels of the samples. Then, the sentiment (or emotion) labels are transferred to the emotion (or sentiment) labels with the transformation model. Finally, the original emotion (or sentiment) labels and the transferred emotion (or sentiment) labels are combined to make the final decision. The framework of this joint learning for emotion classification is shown in Figure 2.
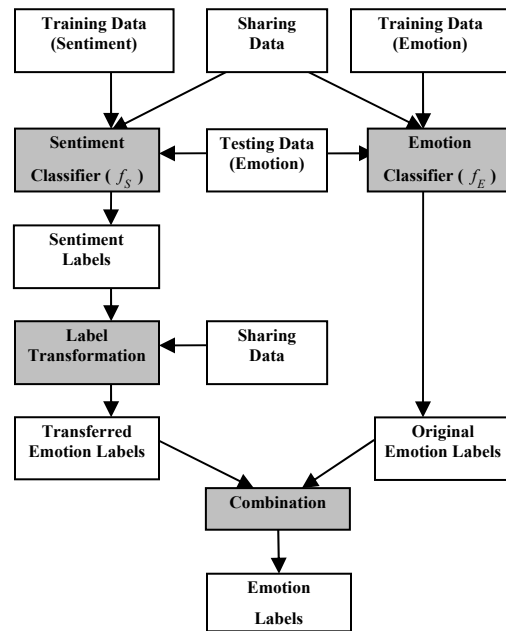

Figure 2: The framework of joint learning for emotion classification with resources from both sentiment and emotion classification

3.2 Label Transformation and Result Combination

Sentiment and emotion classification have a close relationship which makes it possible to transfer the output probability distribution from one task to another. Formally, let X denote a feature vector of a sample. The output of the sentiment classification f_S on this sample is the probability distribution $p(s_i | X)$ over the set of sentiment labels $\{s_1, s_2 \ldots s_{N_s}\}$ while the output of the emotion classifier f_E is the probability distribution

$p(e_i \mid X)$ over the set of emotion labels $\{e_1, e_2, ..., e_{N_e}\}$. In this study, the sentiment labels are positive, negative and neutral and the emotion labels are joy, hate, love, sorrow, anxiety, surprise, anger, expect, and neutral. According to Bayes theory, the output probability distribution of the sentiment or emotion classifier can be transferred as follows:

$$p_{transfer}(s_i \mid X) = \sum_{j=1}^{N_e} p(s_i \mid e_j) * p(e_j \mid X) \qquad (2)$$

$$p_{transfer}(e_i \mid X) = \sum_{j=1}^{N_s} p(e_i \mid s_j) * p(s_j \mid X) \qquad (3)$$

Wherein $p(s_i \mid e_j)$ and $p(e_i \mid s_j)$ are the statistic probabilities estimated with formula (4) and (5) from an extra data set, called sharing data $L_{Sharing}$. Laplace smoothing has been used to avoid zero probability.

$$p(s_i \mid e_j) = \frac{1 + \sum_{X \in L_{Sharing}} I(Y_S(X) = s_i, Y_E(X) = e_j)}{N_s + \sum_{X \in L_{Sharing}} I(Y_E(X) = e_j)} \qquad (4)$$

$$p(e_i \mid s_j) = \frac{1 + \sum_{X \in L_{Sharing}} I(Y_E(X) = e_i, Y_S(X) = s_j)}{N_e + \sum_{X \in L_{Sharing}} I(Y_S(X) = s_j)} \qquad (5)$$

Where $Y_S(X)$ is the sentiment label of the sample and $Y_E(X)$ is the emotion label of the sample. The function $I(x)$ is defined as:

$$I(x) = \begin{cases} 1, & x = True \\ 0, & x = False \end{cases} \qquad (6)$$

For a testing sample, both the sentiment and emotion classifiers provide classification results. The two kinds of results could be transferred to each other according to (2) or (3). Given the original result and the transferred one, the combination result is defined as:

$$p_{joint}(s_i \mid X) = (1 - \lambda) p(s_i \mid X) + \lambda p_{transfer}(s_i \mid X) \qquad (7)$$

$$p_{joint}(e_i \mid X) = (1 - \lambda) p(e_i \mid X) + \lambda p_{transfer}(e_i \mid X) \qquad (8)$$

Where $\lambda (0 \le \lambda \le 1)$ is a parameter used to balance the weights of the two kinds of results. Finally, the label of a sample is the one which has the maximum value among the joint posterior probabilities.

4. EXPERIMENTATION

4.1 Experimental Settings

Data Set: We carry out our experiments on the Ren-CECps corpus (Quan and Ren, 2009). It contains 34,603 sentences and each of them is annotated with a sentiment label and an emotion vector (sometimes contains multiple emotion labels). In our experiment, we only consider the majority emotion, that is, the emotion label with the highest value in the emotion vector.

Features: Word unigram features are used. Each sentence is treated as a bag-of-words and transformed into binary vectors encoding the presence or absence of one word feature.

Classification Algorithm: The maximum entropy (ME) classifier implemented with the public tool, Mallet Toolkits[1]. The posterior probabilities belonging to different categories are also provided in this tool.

---

[1] http://mallet.cs.umass.edu/

Implementation: From the corpus, we randomly select 6000 sentences as the testing data for both sentiment and emotion classification. Among the remaining, for sentiment classification, we randomly select 3000, 5000, and 8000 sentences as training data which take the sentiment labels only. Similarly, for emotion classification, we randomly select another 3000, 5000, 8000 sentences as training data which take the emotion labels only. Besides, we randomly select 400 sentences as sharing data. The parameter λ in formula (7) and (8) is set to 0.5.

4.2 Experimental Results

To perform sentiment and emotion classification, we implement the following approaches for comparison:

➤ **Single_Sentiment:** construct a sentiment classifier with sentiment labeled data and sharing data.
➤ **Single_Emotion:** construct an emotion classifier with emotion labeled data and sharing data.
➤ **Joint_Sentiment:** construct a sentiment classifier with all labeled data using our joint learning approach.
➤ **Joint_Emotion:** construct an emotion classifier with all labeled data using our joint learning approach.
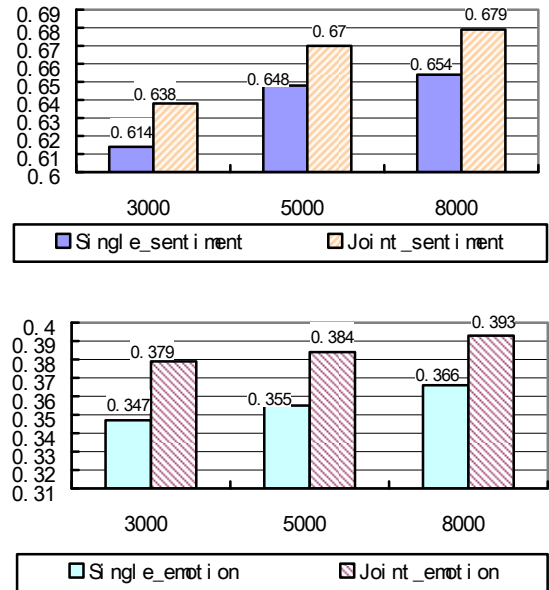
Figure 3: Accuracy performances of different classification approaches on the testing data over different sizes of labeled data

Figure 3 shows the results of these approaches when different quantities of training data are employed. From this figure, we can see that our joint-learning approach robustly outperforms single task classifiers across different sizes of labeled data on either sentiment classification or emotion classification. Averagely, they increase the accuracy from 0.638 to 0.662 in sentiment classification and the accuracy from 0.356 to 0.385 in emotion classification.

Sharing data is the key component in our approach. It is used to estimate the transformation probabilities from a sentiment (or emotion) label to an emotion (or sentiment) label. However, sharing data is a costly work that requires extra human annotation. Fortunately, our approach only needs a very small size of sharing data to estimate the transformation probabilities. Figure 4 shows

classification performances of different approaches with different sizes of sharing data. From Figure 4, we can see that 300 to 400 sentences in sharing data are enough to train a good label-transformation model. Compared to the size of training data, the size of sharing data is much smaller.

The parameter λ in formula (7) and (8) is an important parameter in our approach. To evaluate the sensitiveness of this parameter, we change it from 0 to 1 and we find that the performance is robustly good when λ is between 0.4 and 0.8. The detailed discussion of the results is omitted due to space limit.
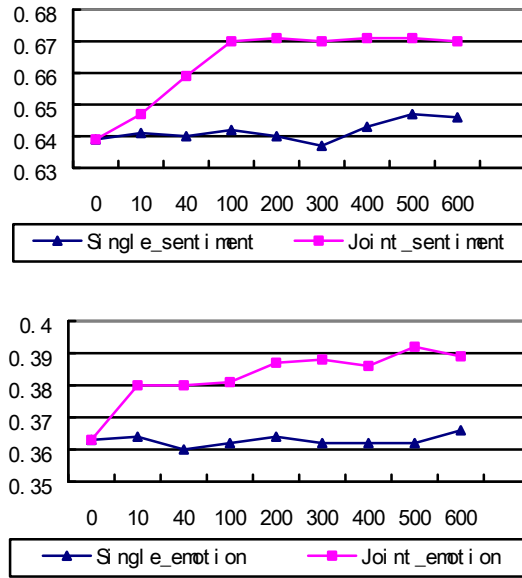

Figure 4: Classification performances of different approaches with different sizes of sharing data

5. CONCLUSION

In this paper, we propose an approach to jointly learning sentiment and emotion classification. Our approach mainly leverages an extra labeled data set to estimate the transformation probabilities between emotion and sentiment labels. Experimental results demonstrate that, compared to individual classifications, our joint learning classification can consistently achieve much better results.

In our future work, we will try to find a solution on joint learning when non sharing data is given, which is the more popular case in real applications. Moreover, we would like to adapt our joint learning approach to other NLP tasks.

6. REFERENCES

[1] Alm C., D. Roth and R. Sproat. 2005. Emotions from Text: Machine Learning for Text-based Emotion Prediction. In *Proceedings of EMNLP-05*, pp.579-586.

[2] Chen Y., S. Lee, S. Li and C. Huang. 2010. Emotion Cause Detection with Linguistic Constructions. In *Proceeding of COLING-10*, pp.179-187.

[3] Das D. and S. Bandyopadhyay. 2009. Word to Sentence Level Emotion Tagging for Bengali Blogs. In *Proceedings of ACL-09*, pp.149-152.

[4] He Y., C. Lin and H. Alani. 2011. Automatically Extracting Polarity-Bearing Topics for Cross-Domain Sentiment Classification. In *Proceeding of ACL-11*, pp.123-131.

[5] Kim S. and E. Hovy. 2006. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. In *Proceedings of the ACL Workshop on Sentiment and Subjectivity in Text*, 1-8.

[6] Li S., C. Huang, G. Zhou and S. Lee. 2010. Employing Personal/Impersonal Views in Supervised and Semi-supervised Sentiment Classification. In *Proceedings of ACL-10*, pp.414-423.

[7] Liu B. 2012. *Sentiment Analysis and Opinion Mining (Introduction and Survey)*. Morgan & Claypool Publishers, May 2012.

[8] Liu B., M. Hu, and J. Cheng.2005. Opinion Observer: Analyzing and Comparing Opinions on the Web. In *Proceedings of WWW-05*, pp.342-351.

[9] Pang B., L. Lee and S. Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. In *Proceedings of EMNLP-02*, pp.79-86.

[10] Purver M. and S. Battersby. 2012. Experimenting with Distant Supervision for Emotion Classification. In *Proceedings of EACL-12*, pp.482-491.

[11] Quan C. and F. Ren. 2009. Construction of a Blog Emotion Corpus for Chinese Emotional Expression Analysis. In *Proceedings of EMNLP-09*, pp.1446-1454.

[12] Riloff E., S. Patwardhan and J. Wiebe. 2006. Feature Subsumption for Opinion Analysis. In *Proceedings of EMNLP-06*, pp.440-448.

[13] Turney P. 2002. Thumbs up or Thumbs down? Semantic Orientation Applied to Unsupervised Classification of comments. In *Proceedings of ACL-02*, pp.417-424.

[14] Wiebe J., T. Wilson, and C.2005. Cardie. Annotating Expressions of Opinions and Emotions in Language. Language Resources and Evaluation, vol.39(2/3).

[15] Wilson T., J. Wiebe, and P. Hoffmann. 2009. Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis. Computational Linguistics, vol.35(3), pp.399-433.

[16] Xu G., X. Meng and H. Wang. 2010. Build Chinese Emotion Lexicons Using A Graph-based Algorithm and Multiple Resources. In *Proceeding of COLING-10*, pp.1209-1217.