Joint Sentiment and Emotion Classification with Integer Linear Programming

Rong Wang[1,2] Shoushan Li[1,2*] Guodong Zhou[1,2] Hanxiao Shi[3]

[1] Natural Language Processing Lab, School of Computer Science and Technology, Soochow University, China
[2] Collaborative Innovation Center of Novel Software Technology and Industrialization
[3] School of Computer Sience and Information Engineering, Zhejiang Gongshang University, China

{wangrong2022,shoushan.li}@gmail.com, gdzhou@suda.edu.cn, hxshi@mial.zjgsu.edu.cn

Abstract. As two foundational tasks in sentiment analysis, sentiment classification and emotion classification have been considered separately and studied independently in the literature. In this paper, we put forward an Integer Linear Programming (ILP)-driven joint learning approach to leveraging the relationship between these two tasks. Empirical study verifies the appropriateness and effectiveness of our proposed approach to joint sentiment and emotion classification.

Keywords: sentiment classification, emotion classification, Integer Linear Programming (ILP)

1 Introduction

During the last decade, sentiment analysis has attracted considerable attention in multiple research communities, such as natural language processing (NLP), data mining and social media (Pang and Lee, 2008; Liu, 2012). As two foundational tasks in sentiment analysis, sentiment classification is concerned with predicting coarse-grained sentimental orientations (e.g., *positive* or *negative*) towards a topic, and emotion classification is concerned with predicting fine-grained personal emotions (e.g., *happy*, *sad*, *surprise*, *fear*) expressed by a human being (Quan and Ren, 2009). Both have their value in a wide range of real-life applications, such as opinion mining and psychological analysis.

Although both sentiment classification and emotion classification have been well explored in the literature, they are normally addressed separately. This largely ignores the dependency between sentimental orientations and personal emotions. For example, if we can determine the personal emotion of the sentence in **Example 1** to be *happy*, we can easily infer the sentimental orientation of the sentence to be *positive*, while, if

* Corresponding author

we can determine the sentimental orientation of the sentence in **Example 1** to be *positive*, we can easily eliminate those very *negative* emotions, e.g., *angry* and *disgust*, although the exact emotion is hard to infer.

Example 1: *I am so happy to get this book. I really like it.*

Given the close relationship between sentimental orientations and personal emotions, it is natural to leverage preciously annotated data from one task to help another task. In this paper, we explore joint sentiment and emotion classification in better exploiting the annotated data from either task. Specifically, we design some constraints between sentimental orientations and personal emotions and perform global inference on the outputs from both the sentiment and emotion classifiers with Integer Logical Programming (ILP).

The remainder of this paper is organized as follows. Section 2 overviews related work on both sentiment classification and emotion classification. Section 3 presents our approach to joint sentiment and emotion classification. Section 4 evaluates the proposed approach. Finally, Section 6 gives the conclusion.

2 Related Work

Sentiment classification has been extensively studied in the last decade since the pioneering work by Pang et al. (2002). Earlier studies on this research issue could refer to two comprehensive surveys by Pang et al. (2008) and Liu (2012). Recent studies on sentiment classification mainly focus on the solving the sparse data problem in machine learning-based approaches to sentiment classification, such as unsupervised learning (Ou et al., 2014), semi-supervised (Zhou et al., 2013), cross-domain (Li et al., 2013a), and cross-lingual sentiment classification (Li et al., 2013b). Our work follows the same spirit but leverages the resources from a related task, i.e., emotion classification.

Emotion classification is likewise a hot research topic in the data mining and natural language processing communities. One main group of such studies is about 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 to emotion classification (Alm et al., 2005; Purver and Battersby, 2012).

Works on joint learning based on sentiment and emotion classification are rare. The only one exception we find is the work by Gao et al. (2013) which uses an extra annotated data with both sentiment and emotion annotation to estimate the transformation probabilities to help the two tasks. However, this type of annotated data is not available in most real-life applications. In contrast, in this study, our approach to joint learning is based on ILP, which needs no extra annotated data.

3 Joint Sentiment and Emotion Classification

Our basic idea to joint sentiment and emotion classification is to leverage the close relationship between sentimental orientations and personal emotions. Specifically, ILP is utilized to achieve global optimization in capturing such relationship in the outputs of sentiment and emotion classifiers via some constraints. Figure 1 illustrates our joint learning framework.

Let x be the feature vector of a testing sample. Since the text representation in both sentiment and emotion classification is sometimes the same, e.g., bag-of-words representation, the testing sample x is capable of being classified by either the sentiment classifier or the emotion classifier but with different output results.

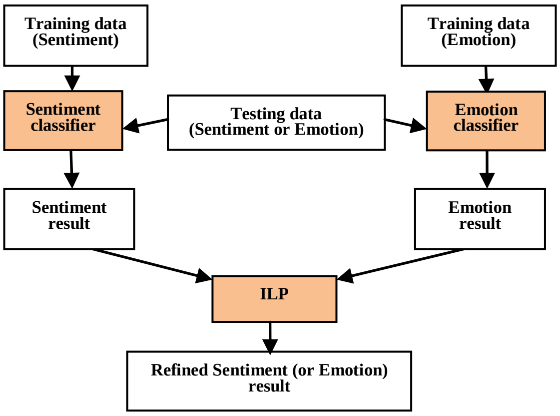

Figure 1: The framework of joint sentiment and emotion classifications with ILP

In sentiment classification, the testing sample is classified to a sentiment category, e.g., *positive* or *negative*. Suppose that the sentiment category of the *i*-th testing sample is y_i and $y_i \in \{0,1,2\}$ where 0 denotes the *neutral* category, 1 denotes the positive category and 2 denotes the negative category. In order to use binary values to represent the result, we use three different variables, i.e., $y_{0i} \in \{0,1\}$, $y_{1i} \in \{0,1\}$, $y_{2i} \in \{0,1\}$, to denote whether the *neutral*, *positive*, or *negative* category is decided by the classifier. For instance, $y_{0i} = 0$ means the classification result is not *neutral*. Therefore, the testing result is denoted by a vector, i.e., $< y_{0i}, y_{1i}, y_{2i} >$. In addition to the category labels, the statistic classifier often provides the posterior probability information, i.e.,

$$P_y =< P_{y_{0i}=1}, P_{y_{0i}=0}, P_{y_{1i}=1}, P_{y_{1i}=0}, P_{y_{2i}=1}, P_{y_{2i}=0} > \quad (1)$$

The corresponding label information is represented by

$$Y =< y_{0i}, 1-y_{0i}, y_{1i}, 1-y_{1i}, y_{2i}, 1-y_{2i} > \quad (2)$$

In emotion classification, the testing sample is classified to an emotion category, e.g., *happy* or *sad*. Suppose that the emotion category of the *i*-th testing sample is z_i and $z_i \in \{0,1,2,...,8\}$ where numbers 0-8 denote the categories of *neutral*, *joy*, *love*, *expect*, *surprise*, *anger*, *anxiety*, *hate*, *sorrow* in this study. The probability and label results are:

$$P_z =< P_{z_{0i}=1}, P_{z_{0i}=0}, P_{z_{1i}=1}, P_{z_{1i}=0}, ..., P_{z_{8i}=1}, P_{z_{8i}=0} > \quad (3)$$

and

$$Z = < z_{0i}, 1-z_{0i}, z_{1i}, 1-z_{1i}, ..., z_{8i}, 1-z_{8i} > \quad (4)$$

Our objective of joint learning is to adjust the label result given the probability result, together with some constraints. The objective function is designed to make the label result similar to the probability result as much as possible. We use *Cosine* method to measure the similarity between the label and probability vector and thus our objective becomes to maximize the following formula:

$$\textbf{Max} \quad \frac{< P_y, P_z > \cdot < Y, Z >}{\left| < P_y, P_z > \right| \left| < Y, Z > \right|} \quad (5)$$

Given the probability results from the two classifiers, the term $\left| < P_y, P_z > \right|$ is a fixed value. Additionally, we find that whatever the final label result is, the term $\left| < Y, Z > \right|$ is also a fixed value under the assumption that there is only one sentiment label and one emotion label assigning to a testing sample. Therefore, our objective function becomes:

$$\textbf{Max} \quad < P_y, P_z > \cdot < Y, Z > \quad (6)$$

This is exactly an integer linear programming problem. In detail, the objective function is:

$$\textbf{Max} \quad \sum_{k=0}^{2} P_{y_{ki}=1} \cdot y_{ki} + \sum_{k=0}^{2} P_{y_{ki}=0} \cdot (1-y_{ki}) + \sum_{l=0}^{8} P_{z_{li}=1} \cdot z_{li} + \sum_{l=0}^{8} P_{z_{li}=0} \cdot (1-z_{li}) \quad (7)$$

Subject to :

(C1) Integer constraint:

$$y_{ki} \in \{0,1\} \text{ and } z_{li} \in \{0,1\} \quad (8)$$

(C2) Single label constraint:

$$\sum_{k=0}^{2} y_{ki} = 1, \text{ and } \sum_{l=0}^{8} z_{li} = 1 \quad (9)$$

(C3) Neutral emotion constraint: When the testing sample is classified as a *neutral* sample in sentiment classification, the emotion labels of the sample must be the emotion label of *neutral*, i.e.,

$$y_{0i} = z_{0i} \quad (10)$$

(C4) Positive emotion constraint: When the testing sample is classified as a *positive* sample in sentiment classification, the emotion labels of the sample must be in the emotion label set of { *joy, love, expect, surprise* }, i.e.,

$$y_{1i} = z_{1i} + z_{2i} + z_{3i} + z_{4i} \quad (11)$$

(C5) Negative emotion constraint: When the testing sample is classified as a *negative* sample in sentiment classification, the emotion labels of the sample must be in the emotion label set of { *anger, anxiety, hate, sorrow* }, i.e.,

$$y_{2i} = z_{5i} + z_{6i} + z_{7i} + z_{8i} \quad (12)$$

4 Experimentation

Experimental Settings

- **Data Set:** Our experiment is performed on the Ren-CECps corpus (Quan and Ren, 2009), which contains 34,603 sentences. Each sentence is annotated with a sentiment label and an emotion vector (sometimes contains multiple emotion labels). In our experiment, we only consider the majority emotion.
- **Features:** Each sentence is treated as a bag-of-words and transformed into binary vectors encoding the presence or absence of word unigrams.
- **Classification Algorithm:** The maximum entropy (ME) classifier implemented with the public tool, Mallet Toolkits (http://mallet.cs.umass.edu/) is employed in all our experiments. The posterior probabilities belonging to the categories are also provided in this tool.
- **Implementation**: From the corpus, we randomly select 3500 sentences as the test data for both sentiment and emotion classification. Among the remaining data, we randomly select 2000, 3500, and 7000 sentences as the training data for sentiment classification and select another 2000, 3500, and 7000 sentences as the training data for emotion classification. The ILP is solved by the tool named lp_solve 5.5.2.0(http://web.mit.edu/lpsolve_v5520/).

Experimental Results

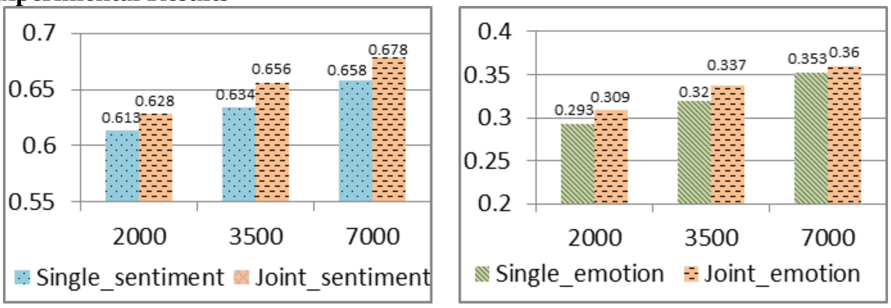

Figure 1: Accuracy performances of different classification approaches over different sizes of labeled data

In this section, we compare the following approaches to joint sentiment and emotion classification.

- **Single_Sentiment**: using only the sentiment classifier to obtain the sentiment result
- **Single_Emotion**: using only the emotion classifier to obtain the emotion result
- **Joint_Sentiment**: performing joint learning with both the sentiment and emotion classifiers to obtain the sentiment result
- **Joint_Emotion**: performing joint learning with both the sentiment and emotion classifiers to obtain the emotion result

Figure 1 shows the results of different approaches when different sizes 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. On average, they increase the accuracy from 0.635 to 0.654 in sentiment classification and the accuracy from 0.322 to 0.335 in emotion classification.

5 Conclusion

In this paper, we propose a novel approach to joint sentiment and emotion classification. Our approach mainly leverages ILP to globally infer sentiment and emotion categories from the outputs of the two classifiers. Experimental results demonstrate that, compared to individual classifiers, our joint learner consistently achieves better performance.

Acknowledgments

This research work has been partially supported by three NSFC grants, No. 61273320, No.61375073, No.61331011, one NSF grant of Zhejiang Province No.Z1110551 and Collaborative Innovation Center of Novel Software Technology and Industrialization.

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. Gao W., S. Li, S. Lee, G. Zhou, and C. Huang. Joint Learning on Sentiment and Emotion Classification. In *Proceeding of CIKM-13*, pp.1505-1508.
3. Li S., Y. Xue, Z. Wang, and G. Zhou. 2013a. Active Learning for Cross-Domain Sentiment Classification. In *Proceeding of IJCAI-13*, pp.2127-2133.
4. Li S., Y. Xue, Z. Wang, S. Lee, and C. Huang. 2013b. Data Quality Controlling for Cross-Lingual Sentiment Classification. In *Proceeding of IALP-13*, pp.125-128.
5. Liu B. 2012. *Sentiment Analysis and Opinion Mining (Introduction and Survey)*. Morgan & Claypool Publishers, May 2012.
6. Ou G., W. Chen, B. Li, T. Wang, D. Yang, and K. Wong. 2014. CLUSM: An Unsupervised Model for Microblog Sentiment Analysis Incorporating Link Information. In *Proceeding of DASFAA-14*, pp.481-494.
7. Pang B. and L. Lee. 2008. Opinion Mining and Sentiment Analysis: Foundations and Trends. *Information Retrieval*, vol.2(12): 1-135.
8. Pang B., L. Lee and S. Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. In *Proceedings of EMNLP-02*, pp.79-86.
9. Purver M. and S. Battersby. 2012. Experimenting with Distant Supervision for Emotion Classification. In *Proceedings of EACL-12*, pp.482-491.
10. 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.
11. 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.
12. Zhou S., Q. Chen, and X. Wang. 2013. Active Deep Learning Method for Semi-supervised Sentiment Classification. *Neurocomputing* 120, pp.536-546.