Two-View Label Propagation to Semi-supervised Reader Emotion Classification

Shoushan Li, Jian Xu, Dong Zhang, Guodong Zhou*
Natural Language Processing Lab
School of Computer Science and Technology, Soochow University, China
lishoushan@suda.edu.cn, jxu1017@stu.suda.edu.cn
dzhang@stu.suda.edu.cn, gdzhou@suda.edu.cn

Abstract

In the literature, various supervised learning approaches have been adopted to address the task of reader emotion classification. However, the classification performance greatly suffers when the size of the labeled data is limited. In this paper, we propose a two-view label propagation approach to semi-supervised reader emotion classification by exploiting two views, namely *source* text and *response* text in a label propagation algorithm. Specifically, our approach depends on two word-document bipartite graphs to model the relationship among the samples in the two views respectively. Besides, the two bipartite graphs are integrated by linking each *source* text sample with its corresponding *response* text sample via a length-sensitive transition probability. In this way, our two-view label propagation approach to semi-supervised reader emotion classification largely alleviates the reliance on the strong sufficiency and independence assumptions of the two views, as required in co-training. Empirical evaluation demonstrates the effectiveness of our two-view label propagation approach to semi-supervised reader emotion classification.

1 Introduction

Source **Text (*News*)**:

An earthquake of 7.0 magnitude struck China in Lushan County of Ya'an, Sichuan Province, causing serious casualties and property losses, and millions of people in distress……

Writer Emotion: *Neutral*

Reader Emotion:  (*Sadness*), (*Worry*)

Response **Text (*Comments*)**:

(1) *Be sure to cherish the golden rescue time.*

(2) *Why there is always an earthquake, so sad, wish all the best.*

(3) *I experienced this quake, all too suddenly, and I will never forget.*

Figure 1: An example of a news article, together with its writer and reader emotions

Emotion classification aims to predict the involving emotion towards a piece of text. For a particular text, there always exist two kinds of emotions, namely writer emotion and reader emotion, where the former concerns the emotion produced by the writer who writes the text and the latter concerns the

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: http://creativecommons.org/licenses/by/4.0/

* Corresponding author

Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 2647–2655, Osaka, Japan, December 11-17 2016.

emotion produced by the reader who reads the text. For instance, in Figure 1, given the news text about earthquake, the writer emotion is more likely to be *neutral* due to the professionalism of the news reporter, while the reader emotion might be *sadness*, or *worry*. Recent years have seen growing interest in reader emotion classification due to its importance in more and more real-life applications, such as content recommendation and online advertisement.

Conventional approaches to reader emotion classification conceptualize the task as a supervised learning problem and rely on a large-scale human-annotated data for model learning. Although such supervised approaches deliver reasonably good performance, the reliance on labeled data, which is normally difficult and highly expensive to obtain, presents a major obstacle to the widespread application of reader emotion classification.

To alleviate the problem above, Liu et al. (2013) originally propose a semi-supervised learning approach to reader emotion classification to improve the performance by enlarging the labeled data with automatically inferred annotations of unlabeled instances. Their basic idea mainly lies on unique characteristics in reader emotion analysis, different from the case in writer emotion analysis. That is, apart from the *source* text (e.g., *news* text), another type of text, the *response* text (e.g., *comment* text) written by the reader as a response to the *source* text, is available to help determine the reader emotion of the *source* text. For example, in Figure 1, the *comment* "*Why there is always an earthquake, so sad, wish all the best*" explicitly express reader emotion *sadness*. Therefore, the *source* text and the *response* text are casted respectively as two views in a co-training algorithm to perform semi-supervised learning.

However, the success of co-training largely depends on two strong underlying assumptions, i.e., sufficiency and independence, of the two views (Blum and Mitchell, 1998), which are actually violated in reader emotion classification when the *source* text and *response* text are utilized as two views.

On one hand, the *response* text often lacks sufficient information to correctly predict the label of an instance, since the response text tends to be short. For example, in Figure 1, if there is only one existing comment, e.g., (1) "*Be sure to cherish the golden rescue time*", the reader emotion is difficult to predict because no emotion is clearly expressed in this sentence. Even worse, as an extreme example, the *source* text (e.g., some newly posted *news*) sometimes has no *response* at all.

On the other hand, the *response* text normally depends on the *source* text, since both the *response text* and the *source* text talk about the same topics. It is really hard for them to meet the view independence assumption in co-training.

In this paper, we propose a novel semi-supervised learning approach, namely two-view label propagation (LP), to reader emotion classification. As an extension of traditional label propagation with a single view (Zhu and Ghahramani, 2002), our two-view LP approach depends on two graphs, i.e., one depicting the connections among the *source* text samples and the other depicting the connections among the *response* text samples. Besides, the two graphs are integrated by linking each *source* text sample with its corresponding *response* text sample to capture the dependence between the *source* text and the *response* text. Such a two-view LP approach thus avoids the independence assumption, as required in traditional co-training. Finally, we assign a variable weight between each *source* text sample and its *response* text sample to address the information insufficiency in the *response* text. Specifically, we design a length-sensitive linear function to calculate the transition probability between the source and response text samples.

The remainder of this paper is organized as follows. Section 2 overviews related work on emotion classification. Section 3 introduces the baseline approach to semi-supervised reader emotion classification with single-view label propagation. Section 4 presents our two-view label propagation approach to semi-supervised reader emotion classification. Section 5 empirically evaluates our approach. Finally, Section 6 gives the conclusion and future work.

2 Related Work

Among the large number of studies in sentiment analysis over the last decade (Pang et al., 2002; Turney, 2002; Alm et al., 2005; Wilson et al., 2009), only a small portion focus on emotion classification.

Besides those on emotion resource construction, such as emotion lexicon building (Xu et al., 2010; Volkova et al., 2012; Staiano and Guerini, 2014) and sentence-level or document-level corpus construction (Quan and Ren, 2009; Das and Bandyopadhyay, 2009), most of previous studies on emotion classification are devoted to designing novel classification approaches to emotion classification (Alm et al.,

Symbol	Definition
L_s	Labeled *source*-text data
L_r	Labeled *response*-text data
U_s	Unlabeled *source*-text data
U_r	Unlabeled *response*-text data
G_s	Graph of the *source*-text data
G_r	Graph of the *response*-text data
$G_{s,r}$	Joint graph of both the *resource* and *response* text data
M_s	The transition probability matrix among the *source* text data
M_r	The transition probability matrix among the *response* text data
$M_{s,r}$	The transition probability matrix among the *source* and *response*-text data

Table 1: Symbol definition

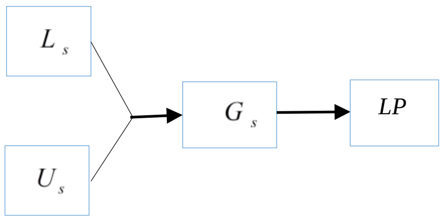

Figure 2: The framework of single-view label propagation approach to semi-supervised learning on reader emotion classification

2005; Chen et al., 2010; Purver and Battersby, 2012; Hasegawa et al., 2013; Qadir and Riloff, 2014), mainly from the supervised learning paradigm.

Compared with above studies on writer emotion classification, studies on reader emotion classification are much limited. Lin et al. (2007) first describe the task of reader emotion classification on news articles with some standard machine learning approaches. Lin et al. (2008) further exploit more features to improve the performance.

More recently, Liu et al. (2013) propose a co-training approach to semi-supervised learning on reader emotion classification by considering the *message* text and the *comment* text as two views. However, their success is much limited due to the required two strong assumptions on co-training, i.e. sufficiency and independence assumptions on the two views in co-training.

3 Single-view LP to Semi-supervised Reader Emotion Classification

In reader emotion classification, each target (e.g., a *news* article) is represented by two kinds of text, namely *source* text and *response* text. Formally, we refer the training data containing the *source* text samples as L_s and the one containing the *response* text samples as L_r. In this study, we only consider two emotion categories, i.e., *positive* and *negative* emotions. The task of semi-supervised learning on reader emotion classification is to leverage the training data L_s and L_r, together with the unlabeled data U_s and U_r, to train a classifier. For clarity, Table 1 illustrates some important symbols.

Figure 2 illustrates the framework of the LP-based semi-supervised approach to when only the view of the *source* text is utilized. Traditional label propagation (LP) is a graph-based semi-supervised learning approach with a single view (Zhu and Ghahramani, 2002). In general, a LP-based approach to semi-supervised learning consists of two main steps: graph construction to represent the relationship among the document samples and label propagation to propagate the labels of the labeled data to the unlabeled

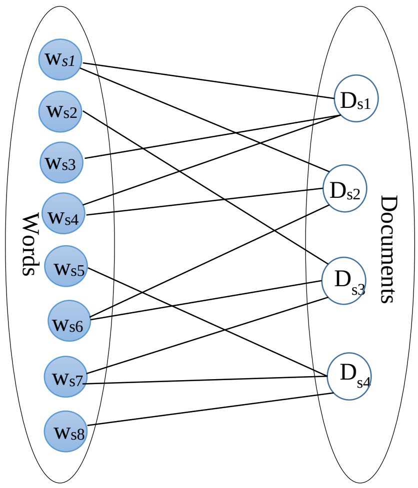

Figure 3: The word-document bipartite graph

Input:

P: The $n \times 2$ matrix, while p_{ir} represents the probability of document D_i (i=*1*...*n*) with label r (r=0,1);

M : The $n \times n$ transition probability matrix

Output:

The unlabeled data with prediction labels

Procedure:

1) Initialize P as P_0
   a) Assign each labeled sample with a fixed probability distribution (1, 0) or (0,1) according to its label r;
   b) Assign each unlabeled sample with an initial probability distribution (0.5, 0.5);
2) Loop until P converges;
   a) Propagate the labels of any vertex to nearby vertices by $P_t = M^T \cdot P_{t+1}$;
   b) Clamp the labeled data, that is, replace the probabilities of the labeled samples in P_{t+1} with their initial ones in P_0;
3) Assign each unlabeled instance with a label by computing $\arg\max_r p_{ir}$

Figure 4: The LP algorithm

data in the obtained graph.

In detail, in the first step, we adopt a word-document bipartite graph to model the relationship among the document samples due to its excellent performance in sentiment classification (Sindhwani and Melville, 2008). Figure 3 illustrates the structure of the word-document bipartite graph, in which the nodes consist of two parts: all documents and all words extracted from the documents. An undirected edge (D_i, w_k) exists if and only if document D_i contains word w_k. Let x_{ik} be the frequency of word w_k in document D_i. From the bipartite graph, the probability of walking from document D_i to word w_k can be calculated as $x_{ik} / \sum_k x_{ik}$ and the probability of walking from word w_k to document d_j can be calculated as $x_{jk} / \sum_j x_{jk}$. Thus the probability of walking from document D_i to document D_j though the word w_k can be calculated as $(x_{ik} / \sum_k x_{ik}) \cdot (x_{jk} / \sum_j x_{jk})$. When all words are considered, we get the transition probability from D_i to D_j as:

$$t_{ij} = \sum_k \frac{x_{ik}}{\sum_k x_{ik}} \cdot \frac{x_{jk}}{\sum_j x_{jk}} \quad (1)$$

and the transition probability matrix $M = \{t_{ij}\}$.

In the second step, we adopt the standard LP algorithm to perform semi-supervised learning. In detail, Figure 4 illustrates the LP algorithm (Zhu and Ghahramani, 2002), during which the probabilities of the labeled data are clamped in each loop using their initial ones and act as a force to propagate their labels to the unlabeled data.

4. Two-View LP to Semi-supervised Reader Emotion Classification

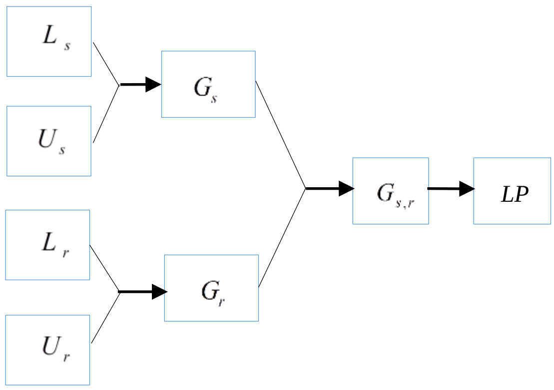

Figure 5: The framework of our two-view label propagation approach to semi-supervised learning on reader emotion classification

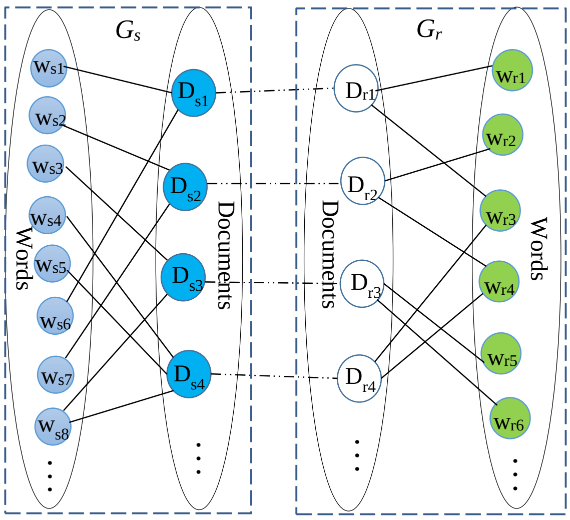

Figure 6: The joint two-view graph that contains both the *source* and *response* text sub-graphs

Figure 5 illustrates the framework of our LP-based semi-supervised approach to reader emotion classification when two views, i.e., the *source* text and the *response* text, are utilized. The graph in our approach consists of two sub-graphs, i.e., G_s and G_r , and each of them is modeled as a word-document bipartite graph. Each pair of the *source* text document and its corresponding *response* text document is connected to join the two sub-graphs together. Figure 6 illustrates the joint two-view graph of both the *source* text and *response* text data.

Basically, the transition probability between the *source* text document and its corresponding *response* text document can be set to 1, assuming that they exhibit the same category. Therefore, the transition probability matrix $M_{s,r}$ of the joint graph $G_{s,r}$ can be represented as follows:

$$M_{s,r} = \begin{bmatrix} M_s & I \\ I & M_r \end{bmatrix} \quad (2)$$

Where I is an identity matrix of dimension n, containing ones along the diagonal and zeros in all other positions; M_s and M_r are the transition probability matrixes, calculated using formula (1) in the sub-graphs, G_s and G_r, respectively.

However, when a *response* text contains too few information or even no words, it becomes a noisy sample. In this scenario, it is not appropriate to propagate the emotion label of its *source*-text sample to this noisy sample. Therefore, for the *source* text sample and its corresponding *response* text sample (e.g., D_{s1} and D_{r1} as shown in Figure 6, we design a function to measure the transition probability by taking the length of the *response* text into account, as follows:

$$t_{s_i, r_i}(z_{r_i}) = \begin{cases} 1 & z_{r_i} \geq l_{max} \\ z_{r_i} / l_{max} & z_{r_i} < l_{max} \end{cases} \tag{3}$$

Where z_{r_i} is the length of the response text document D_{r_i}, defined as the number of the words in the *response* text document; l_{max} denotes the threshold of the document length. If the length is larger than this threshold, it is given a transition probability of 1. If the length is smaller than this threshold, the longer the *response* text sample it is, the higher transition probability it has.

Accordingly, the transition probability matrix $M_{s,r}$ of the joint graph $G_{s,r}$ can be refined as follows:

$$M_{s,r} = \begin{bmatrix} M_s & I' \\ I' & M_r \end{bmatrix} \tag{4}$$

Where I' is a matrix of dimension n, containing the transition probabilities, calculated using formula (3).

5. Experimentation

We systematically evaluate our semi-supervised learning approach to reader emotion classification.

5.1 Experimental Settings

Data collection: The data is collected from Yahoo! Kimo News (http://tw.news.yahoo.com). Each *news* article and its *comments* are considered as a *source*-text sample and a *response*-text sample, respectively. Besides, each *news* article is voted with emotion tags from eight categories: *happy*, *sad*, *angry*, *meaningless*, *boring*, *heartwarming*, *worried*, and *useful*. Following Liu et al. (2013), we consider *happy* and *heartwarming* as *positive* category while *sad*, *angry*, *boring*, and *worried* as *negative* category. The emotion label of a news article is automatically derived from the votes, i.e. the *news* article with over 10 votes of *positive* (*negative*) emotions is assigned with a *positive* (or *negative*) label. Unlike Liu et al. (2013), we do not filter those news articles with less than 5 comments.

Data setting: We randomly select 1300 positive and 1300 negative *source* and *response* instances for the empirical study. Among them, 300 positive and 300 negative *source* and *response* instances are used as test data while the remaining 1000 positive and 1000 negative *source* and *response* instances are used as training data. In the training data, we select 0.5%, 1%, and 2% data as initial labeled data and the remaining data as unlabeled data respectively.

Features: Each *source* (or *response*) text 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 Mallet Toolkits (http://mallet.cs.umass.edu/).

Evaluation Measurement: The performance is evaluated using the standard accuracy measurement.

Significance test: *T*-test is used to evaluate the significance of the performance difference between two approaches (Yang and Liu, 1999).

5.2 Experimental Results on Single-view LP

In this section, we compare different approaches to semi-supervised learning on reader emotion classification when only one view is utilized. For fair comparison, we implement following approaches:

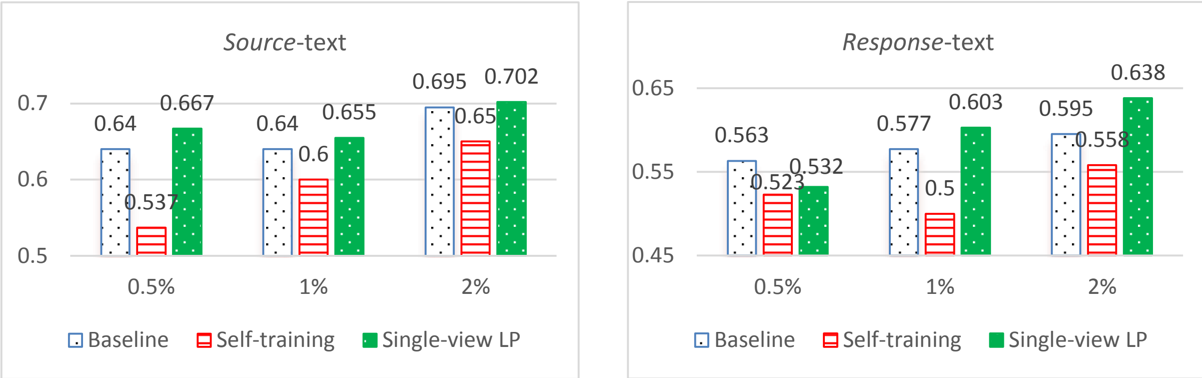

Figure 7: Performance comparison of different semi-supervised learning approaches to reader emotion classification when only one view is utilized

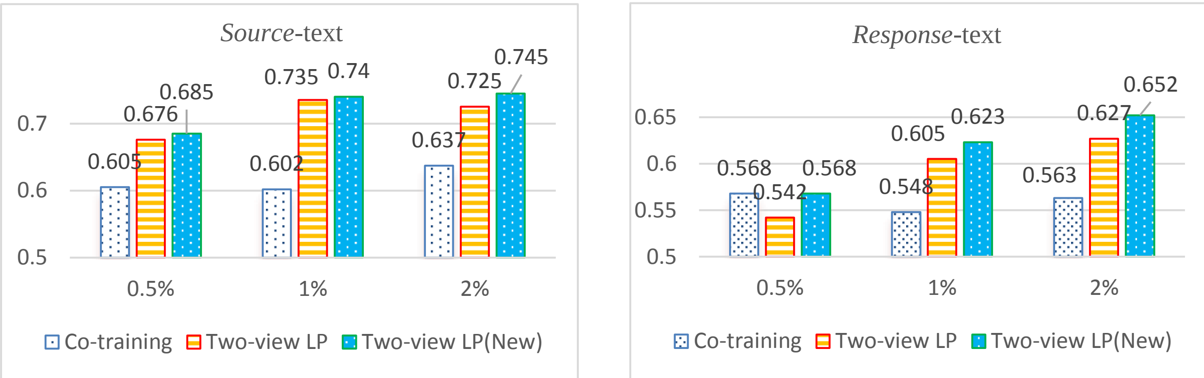

Figure 8: Performance comparison of different semi-supervised learning approaches to reader emotion classification when two views are utilized

- **Baseline**: using only labeled data to train the classifier for predicting the reader emotion of each view (No unlabeled data is used.)
- **Self-training**: using self-training, a simple bootstrapping approach, to iteratively add the high-confident unlabeled samples as automatically labeled samples in each view.
- **Single-view LP**: first using the word-document bipartite graph to model the relationship among the document samples in each view and then applying label propagation to perform semi-supervised learning as introduced in Section 3.1.

Figure 7 shows the performance of different approaches when either the *source*-text or the *response*-text view is utilized. From the figure, we can see that self-training performs dramatically worse than the single-view LP approach, even worse than the baseline approach. In general, the single-view LP approach is effective on using unlabeled data to improve the performance, although the average improvement is limited with around 2%. Besides, the single-view LP approach fails in the case when only 0.5% of the training data are used as initial labeled data. This is possibly because too few initial labeled samples make it extremely difficult to correctly bootstrap enough unlabeled samples.

5.3 Experimental Results on Two-view LP

In this section, we compare different approaches to semi-supervised learning on reader emotion classification when two views are utilized. For comparison, we implement following approaches:

- **Co-training**: using the *source*–text and *response*-text as two views in co-training, as described in Liu et al. (2013).
- **Two-view LP**: our two-view LP approach with the unique transition probability of 1 between the *source*-text sample and its *response*-text sample.
- **Two-view LP (New)**: the two-view LP approach with a variable transition probability between the *source* text sample and its corresponding *response* text sample, as calculated with formula (3). Here,

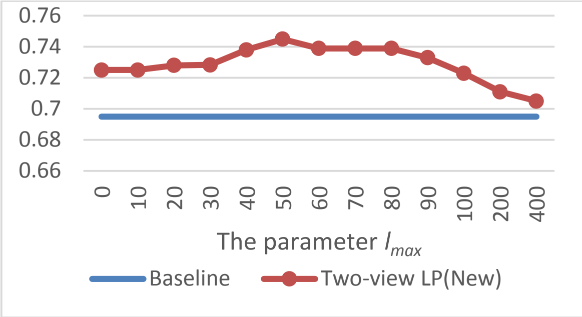

Figure 9: Sensitiveness of the performance on parameter l_{max}

parameter l_{max} is fine-tuned to be 50.

Figure 8 shows the performance of different approaches when both the *source*-text and the *response*-text views are utilized. From this figure, we can see that the two-view LP approach performs much better than co-training. Significance test shows that the improvement of two-view LP (New) over two-view LP is significant (p-value<0.05). This indicates the appropriateness of a variable transition probability between the *source* text sample and its corresponding *response* text sample.

From both Figure 7 and Figure 8, we can see that co-training fails to exploit unlabeled data to improve the performance, quite different from that in Liu et al. (2013). This is mainly due to the fact that the *response* text samples in our data contain much less comments. This makes the sufficiency assumption violated in co-training. Furthermore, we find that two-view LP significantly outperforms single-view LP (p-value<0.05) with a large margin. This verifies the effectiveness of using two views to perform semi-supervised learning on reader emotion classification.

Finally, Figure 9 shows the performance of two-view LP (New) with varying values of parameter l_{max} when testing on the *source*-text samples. Due to the space limitation, we only show the performance when using 2% training data as the initial labeled data. From this figure, we can see that our approach performs consistently well when the parameter is set from 40 to 90, which is a very broad range.

6. Conclusion

In this paper, we propose a novel approach, namely, two-view label propagation, to semi-supervised learning on reader emotion classification. Our approach consists of two main steps: (1) constructing a joint graph containing two word-document bipartite sub-graphs; (2) performing label propagation to incorporate the unlabeled data. Furthermore, we design a length-sensitive function to measure the transition probability from a *source* text sample to its responding *response* text sample. Experimental studies demonstrate that our two-view label propagation approach is capable of employing the two views and unlabeled data to improve the performance.

This work mainly focuses on reader emotion classification with only two categories, i.e., *positive* and *negative* emotions. In the future work, we will explore semi-supervised learning on reader emotion classification when more fine-grained categories, such as *happiness*, *sadness*, and *anger*, are considered. Moreover, given the wide potential of the two-view LP approach, we will explore it in other NLP tasks where two views are involved.

Acknowledgements

This research work has been partially supported by three NSFC grants, No.61273320, No.61375073, No.61331011.

References

Cecilia Ovesdotter Alm, Dan Roth and Richard Sproat. 2005. Emotions from Text: Machine Learning for Text-based Emotion Prediction. In *Proceedings of EMNLP-05*, pages 579-586.

Avrim Blum and Tom Mitchell. 1998. Combining Labeled and Unlabeled Data with Co-training. In *Proceedings of COLT-98*.

Ying Chen, Sophia Yat Mei Lee, Shoushan Li and Chu-Ren Huang. 2010. Emotion Cause Detection with Linguistic Constructions. In *Proceeding of COLING-10*, pages 179-187.

Dipankar Das and Sivaji Bandyopadhyay. 2009. Word to Sentence Level Emotion Tagging for Bengali Blogs. In *Proceedings of ACL-09*, pages 149-152.

Takayuki Hasegawa, Nobuhiro Kaji, and Naoki Yoshinaga. 2013. Predicting and Eliciting Addressee's Emotion in Online Dialogue. In *proceedings of ACL-13*, pages 964-972.

Kevin Hsin-Yih Lin, Changhua Yang and Hsin-Hsi Chen. 2007. What Emotions do News Articles Trigger in Their Readers? In *Proceeding of SIGIR-07*, poster, pages 733-734.

Kevin Hsin-Yih Lin, Changhua Yang and Hsin-Hsi Chen. 2008. Emotion Classification of Online News Articles from the Reader's Perspective. In *Proceeding of the International Conference on Web Intelligence and Intelligent Agent Technology*, pages 220-226.

Huanhuan Liu, Shoushan Li, Guodong Zhou, Chu-Ren Huang, and Peifeng Li. 2013. Joint Modeling of News Reader's and Comment Writer's Emotions. In *proceedings of ACL-13*, short paper, pages 511-515.

Bo Pang, Lillian Lee and Shivakunmar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. In *Proceedings of EMNLP-02*, pp.79-86.

Matthew Purver and Stuart Battersby. 2012. Experimenting with Distant Supervision for Emotion Classification. In *Proceedings of EACL-12*, pages 482-491.

Ashequl Qadir and Ellen Riloff. 2014. Learning Emotion Indicators from Tweets: Hashtags, Hashtag Patterns, and Phrases. In *Proceedings of EMNLP-14*, pages 1203-1209.

Changqin Quan. and Fuji Ren. 2009. Construction of a Blog Emotion Corpus for Chinese Emotional Expression Analysis. In *Proceedings of EMNLP-09*, pages 1446-1454.

Vikas Sindhwani and Prem Melville. 2008. Document-Word Co-Regularization for Semi-supervised Sentiment Analysis. In *Proceedings of ICDM-08*, pages 1025-1030.

Jacopo Staiano. and Marco Guerini. 2014. Depeche Mood: a Lexicon for Emotion Analysis from Crowd-Annotated News. In *Proceedings of ACL-14*, pages 427-433.

Peter D. Turney. 2002. Thumbs up or Thumbs down? Semantic Orientation Applied to Unsupervised Classification of comments. In *Proceedings of ACL-02*, pages 417-424.

Svitlana Volkova, William B. Dolan and Theresa Wilson. 2012. CLex: A Lexicon for Exploring Color, Concept and Emotion Associations in Language. In *Proceedings of EACL-12*, pages 306-314.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2009. Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis. *Computational Linguistics*, vol.35(3), pages 399-433.

Ge Xu, Xinfan Meng and Houfeng Wang. 2010. Build Chinese Emotion Lexicons Using A Graph-based Algorithm and Multiple Resources. In *Proceeding of COLING-10*, pages 1209-1217.

Yiming Yang and Xin Liu. 1999. A Re-Examination of Text Categorization Methods. In *Proceedings of SIGIR-99*, pages 42-49.

Xiaojin Zhu and Zoubin Ghahramani. 2002. Learning from Labeled and Unlabeled Data with Label Propagation. *CMU CALD Technical Report*. CMU-CALD-02-107.