Learning Sentence Representation for Emotion Classification on Microblogs

Duyu Tang, Bing Qin*, Ting Liu, and Zhenghua Li

Research Center for Social Computing and Information Retrieval,
Harbin Institute of Technology, Harbin, China
{dytang,qinb,tliu,lzh}@ir.hit.edu.cn

Abstract. This paper studies the emotion classification task on microblogs. Given a message, we classify its emotion as happy, sad, angry or surprise. Existing methods mostly use the bag-of-word representation or manually designed features to train supervised or distant supervision models. However, manufacturing feature engines is time-consuming and not enough to capture the complex linguistic phenomena on microblogs. In this study, to overcome the above problems, we utilize pseudo-labeled data, which is extensively explored for distant supervision learning and training language model in Twitter sentiment analysis, to learn the sentence representation through Deep Belief Network algorithm. Experimental results in the supervised learning framework show that using the pseudo-labeled data, the representation learned by Deep Belief Network outperforms the Principal Components Analysis based and Latent Dirichlet Allocation based representations. By incorporating the Deep Belief Network based representation into basic features, the performance is further improved.

Keywords: Emotion Classification, Deep Belief Network, Representation Learning, Microblogs.

1 Introduction

Users of social media such as Twitter and Weibo often express freely their opinions and emotions with others. Social media are valuable sources to mine the opinions of users. Sentiment analysis (Opinion mining) [1, 2] is a fundamental research area in natural language processing. Recently, a large number of studies have investigated the problem of sentiment analysis on social media, in particular, microblogs [3, 4]. Generally, sentiment analysis on microblogs are divided into two perspectives, target-independent [3, 5] and target-dependent sentiment analysis [4]. The difference between the two tasks is that target-dependent sentiment analysis aims to analyze the opinion of a piece of text towards an aspect. This paper studies the task of emotion classification from the perspective of target-independent sentiment analysis. Namely, given a text, we classify its emotion as happy, sad, angry or surprise.

Although previous studies [6–8] have tested a large number of learning and classification methods for opinion and emotion mining such as SVM, CRF and so on, these

* Corresponding author.

G. Zhou et al. (Eds.): NLPCC 2013, CCIS 400, pp. 212–223, 2013.

methods are mostly based on shallow representation structure, such as bag-of-word (BOW) representation [9]. The recent progress in machine learning suggests that deep learning that tries to build a deep representation structure can be more appropriate for many challenging tasks [10, 11]. In order to explore an abstract representation instead of BOW and deal with the curse of dimensionality, this paper studies the use of deep belief network (DBN) for representation learning on emotion classification. The sentence representation learned from corpus can be used as individual features or supplements to traditional features for emotion classification.

In microblogs, emoticons such as :-) :-(are widely explored as strong indicators to reflect users' opinions and emotions [12, 8, 13]. Our statistics of 10 millions random messages on Weibo [1] shows that 12% of them contain at least one emoticon. Users frequently use emoticons to express their emotion, such as the examples shown in Table 1. In these examples, the emoticons can be used to label the emotion of their corresponding plain texts [12]. For example, the emoticon in Table 1 shows a clear indicator of a happy emotion, so that the corresponding plain text *The movie is wonderful, I love it!* will be collected as a happy message. In this paper, the messages gathered via emoticons is called pseudo-labeled corpus. Although previous studies have tested the effectiveness of pseudo-labeled corpus by training distant supervision model [8] and emoticon-based language model [13] on sentiment classification task. Our preliminary experimental results with the distant supervision method on emotion classification show that the cross-validation accuracy is dissatisfied, which is just around 50%.

Table 1. Sampled Emoticon Messages from Weibo (translated examples)

Emotion type	**Content of messages**
Happy	The movie is wonderful! I love it!
Sad	My heart is broken
Angry	He pissed me off!
Surprise	OMG!! I'm shocked!!

In this paper, we take advantage of the pseudo-labeled corpus to learn the sentence representation in a DBN based framework. We have tested the approach on manually labeled corpus and experimental results in supervised learning framework show that the representation learned by DBN achieves comparable results with basic features, outperforms Principal Components Analysis and Latent Dirichlet Allocation based features, improves the basic features through feature incorporation.

This study shows that:

1. Using the pseudo-labeled data, Deep Belief Network can learn a better representation than Principal Components Analysis and Latent Dirichlet Allocation, and this can yield better results in the emotion classification task.
2. Compared with labeled data and randomly selected unlabeled data, the pseudo-labeled corpus shows positive impacts on sentence representation in the emotion classification task.

[1] `http://www.weibo.com/`

The reminder of this paper is organized as follows: Our method about representation learning for emotion classification is described in Section 2. Experimental results and analysis are reported in Section 3. Section 4 summarizes the existing work on emotion classification, Twitter sentiment analysis and deep learning for NLP. Section 5 concludes our work and presents the future work.

2 Methodology

Some steps are necessary to learn representation from pseudo-labeled corpus. In order to obtain pseudo-labeled corpus, emoticons for each category needs to be selected beforehand. For the purpose of reducing the manual work and meanwhile filtering the ambiguous emoticons, we propose to select representative emoticons based on their quality and quantity (in subsection 2.1). Subsequently, preprocessing and normalization are implemented to ensure the quality of the pseudo-labeled corpus (in subsection 2.2). Afterwards, basic features are proposed to map each message into the same dimensional feature space (in subsection 2.3) for further representation learning. Finally, deep belief network is explored to learn sentence representation by a unsupervised, greedy layer-wise algorithm (in subsection 2.4).

2.1 Emoticon Selection

Emoticons are frequently used in microblogs. In Weibo, there are 425 official emoticons[2], such as . In addition to these official emoticons, some printable characters, such as :-) and :(, are also commonly used to indicate users' emotions. However, we observe that the emotions of some emoticons are ambiguous. For example, some users use  to show their happiness, however others use it as a sad indicator. These ambiguous emoticons make the automatic annotation difficult. To guarantee the quality of the automatic annotation, not all the emoticons can be retained, and the ambiguous ones should be filtered out. Therefore, an automatic ranking strategy based on the quality and quantity of the emoticons is essential. Inspired by the work of [14], the importance of each emoticon in each emotion category is calculated as the Equation 1 shows.

$$S_i(e_j) = Acc_i(e_j) \times log_{10}(freq(e_j)) \tag{1}$$

$$Acc_i(e_j) = \frac{\sum_k co\_freq(e_j, sw_{ik})}{\sum_k \sum_I co\_freq(e_j, sw_{Ik})} \tag{2}$$

In Equation 1, the first multiplier corresponds to the quality factor and the second multiplier indicates the quantity factor. $freq(e_j)$ in the second multiplier stands for the frequency of the emoticon e_j in the corpus. In Equation 2, $co\_freq(e_j, sw_{ik})$ refers to the frequency that the emoticon e_j and the k-th emotional word sw_{ik} in the i-th emotion category co-occur within a message in the corpus. Here, Peking Emotion Lexicon (EL)[3] is used as the external lexicon resource. EL contains approximate 90 tokens with high confidence for each kind of emotions. Finally, according to the calculation results, the top ranked emoticons are selected for each emotion, as shown in Table 2.

[2] Until 2011, there are totally 425 official emoticons.
[3] EL is available at `http://icl.pku.edu.cn/icl_res/`

Table 2. Emoticons for Each Category of Emotions

Emotion type	Selected emoticons
Happy	:) :-) :D
Sad	:(:-(
Angry	
Surprise	

2.2 Pseudo Labelled Corpus Collection

Manually annotating training examples is time consuming and expensive, in this subsection, the pseudo-labeled corpus are extracted without manually labeling. As described in Section 1, pseudo labeled corpus refers to the messages with emoticons in them. In order to generate high-quality data, a lot of preprocessing and basic natural language processing work need to be done beforehand, such as word segmentation and text normalization. The implementation details of preprocessing are described as follows:

1. remove the repost part of a message to keep the pure source content edited by the users.
2. replace the official metadata to the corresponding normalization form. Specifically, replace *@username* with *REF*, *#hashtag#* with *TAG*, *http://...* with *URL*. For example, the message "*#Taylow Swift# @David I love her so much!!! http://j.mp/d5Dupr*" will be unified to "*TAG REF I love her so much!!! URL*".
3. remove duplicated messages based on the Longest Common Subsequence (LCS) algorithm. If the rate of LCS between two messages is higher than a threshold, they will be recognized as duplicated and the shorter one will be ignored.
4. remove the messages whose length are less than 10.

After preprocessing, the messages containing only one kind of emoticons will be collected as pseudo-labeled messages. That is to say, the messages containing emoticons from different emotion categories are not collected. For example, the message containing and simultaneously is ignored. Subsequently, word segmentation is implemented by the Language Technology Platform (LTP) [4].

2.3 Basic Features

Previous work [9, 15–17] has discovered some effective features for sentiment analysis on movie reviews and tweets. The commonly used features include word unigram, POS tags, polarity of a word in the sentiment lexicon, etc. Before learning the deep representation, some basic features are needed to map the messages into the same dimensional feature space. Inspired by previous studies, the basic features used in this paper are described as follows:

1. Word unigram features. To control the dimension of the feature space, only the 2000 most frequent words in the pseudo training data are considered, as done by [18].

[4] `http://ir.hit.edu.cn/demo/ltp/`

2. Punctuation features. Some punctuation sequences which can reflect emotion are muanually selected, such as "!!!", "..." and "???". These punctuation features are utilized as binary features according to whether a predifined punctuation occurs in a message.
3. Emotion lexicon features. In order to map the emotional words in a message into predefined emotion category, the external lexicon ML is introduced. Given a message, the lexicon is used to judge whether the words of each emotion exist in the message, and the corresponding feature is used as a binary feature. For example, given a message "*I am very happy today*", the word happy occurs in the lexicon's happy category, and no word exists in the lexicon's other emotion categories (sad, angry and surprise). Thus, the feature is that: *happy(1), sad(0), angry(0) and surprise(0)*. Besides, the occurrences of emotional words in the message are treated as binary features too.
4. Onomatopoeia features. In microblogs, onomatopoeia words are frequently used to express sentiment, such as "aha", "hey" etc. Therefore, an Onomatopoeia Lexicon (OL) is built manually. The onomatopoeia feature is a binary one according to whether there exists any onomatopoeia word in the given message. Similar with Emotion lexicon features, the occurrences of onomatopoeia words in the message are treated as binary features.
5. Function word features. Function words are mostly verbs that can induce a subjection expression, such as *feel*, *think*, *consider*. A Function Word Lexicon (FWL) is manually collected. The usage of FWL is similar with emotion lexicon and onomatopoeia lexicon.

After extracting the basic features, each message will be mapped into the same dimensional feature space, which will be used as the input of the visible nodes to learn the sentence representation in the following subsection.

2.4 Representation Learning for Emotion Classification

In this paper, we assume that compared with randomly selected data, the pseudo-labeled corpus is closer to the emotional dataset. Thus, the learned representation on pseudo-label corpus has much potential to improve the performance of the emotion classification model. In this subsection, we explore deep belief network (DBN) [19] for representation learning in emotion classification. The illustration of the DBN model is given in Figure 1, which is composed of three layers, and each one of the three layers stands for the Restricted Boltzmann Machine (RBM). The training procedure of DBN is greedy layer-wised, whose intuition is "re-construction" [19, 20]. The idea is to train one layer at a time, starting from lower layers, so that its training objective for the currently added layer is to reconstruct the previous layer. With unsupervised layer-wise training, each layer is trained to model the abstract distribution of the previous layer.

Restricted Boltzmann Machine [21] is proposed to model an ensemble of binary vectors as a two-layer network. Take the bottom layer in Figure 1 as an example, the observed basic features of a message correspond to the "visible" units in the layer $\mathbf{v}$ and the latent features correspond to the hidden units in the layer $\mathbf{h^1}$. A joint configuration $(\mathbf{v}, \mathbf{h})$ of the visible and the hidden units has an energy given by

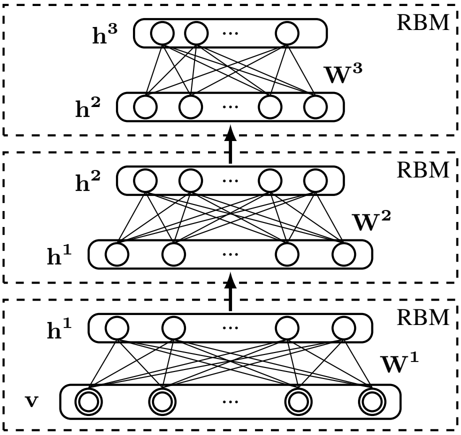

Fig. 1. The Deep Belief Network for Representation Learning

$$E(\mathbf{v},\mathbf{h}) = -\sum_{\mathbf{i}\in\mathbf{inputs}} \mathbf{b_i v_i} - \sum_{\mathbf{j}\in\mathbf{features}} \mathbf{b_j h_j} - \sum_{\mathbf{i,j}} \mathbf{v_i h_j w_{ij}} \tag{3}$$

where v_i and h_j are the binary states of the i-th node in the visible layer and the j-th node in the hidden layer, b_i and b_j are their biases respectively, and w_{ij} is the weight between them.

In the training process, each RBM performs a nonlinear transformation on its input vectors and the output vectors will be used as the input for the next layer. The sigmoid function is used to calculate the probability of each node is on. After activating the hidden units stochastically, a confabulated vector is produced. The states of the hidden units are then updated by the confabulation vector in the same way. The parameters are updated by

$$\triangle w_{ij} = \epsilon(< v_i h_j >_{data} - < v_i h_j >_{recon}) \tag{4}$$

where ϵ is the learning rate, $< v_i h_j >_{data}$ is the probability of the i-th node in $\mathbf{v}$ and the j-th node in $\mathbf{h}$ are on together when the hidden features are driven by the real data, and $< v_i h_j >_{recon}$ is the corresponding probability in the reconstruction stage.

3 Experiment

3.1 Experiment Setup

Dataset We collect 1.2 million Chinese messages from September 2009 to September 2010 using the Weibo API[5]. One million messages are randomly selected to get the pseudo-labeled corpus. After removing the duplicate messages, 20,000 emoticon data are obtained as the pseudo labeled data with 5,000 messages for each kind of emotions.

[5] http://open.weibo.com/

And the same number of messages without emoticons are randomly selected as unlabelled data. In consideration of that there are no available annotated corpus for emotion classification in Weibo, we manually annotate 5,000 messages without emoticons from the rest of 0.2 million messages. Two annotators are required to conduct the annotation, and each annotator is asked to annotate each message as happy, sad, angry, surprise or others. The inter-agreement of the annotators is 87.54%. Finally, after removing the inconsistent annotations and the messages labeled as "others", 2,875 instances are collected as the gold standard for emotion classification. There are 548 happy, 837 sad, 905 angry and 567 surprise messages in the final labeled dataset. The accuracy of cross-validation on the gold dataset is used as evaluation metric. For each type of feature, we utilize LibLinear [6] to train models for emotion classification.

Details of Lexicon Resources. In subsection 2.3, we utilize several lexicon resources, such as Emotion Lexicon (EL), Onomatopoeia Lexicon (OL) and Function Word Lexicon (FWL), to extract basic features. Figure 3 gives the detailed information about these lexicons.

Table 3. Details of Lexicon Resources

Lexicon	EL				OL	FWL
	Happy	Sad	Angry	Surprise		
Size	83	89	101	91	166	188

Architecture of the Network. In this study, different architectures are designed to check the usefulness of representation learned by DBN. For example, there are three layers in the architecture of 2,729-2,000-1,000-500, each of which corresponds to RBM with 2,729$\times$2,000, 2,000$\times$1,000, 1,000$\times$500 individually. In the bottom layer, 2,729 visible units corresponds to the basic features described in subsection 2.3 and the number of hidden units is 2,000. In the training stage, each layer is trained for 50 times greedily. Using this network, a 2,729-dimension binary vector will be represented by a 500-dimension distributed vector. In the preliminary experiment, we conduct two architectures, 2,729-2,000-1,000-500 and 2,729-1,500-750-500, to discover the influence of the architecture. Due to their close performance, we will just report the results achieved from the former architecture.

3.2 Results and Analysis

Below we first present the finding of comparing DBN based representation with the classical text-based feature and PCA or LDA based feature. Then, we compare the effectiveness of the pseudo-labeled corpus with the unlabeled data and small-scale labeled data on representation learning for emotion classification.

Comparison between Representations. In the first set of experiments, we compare the learned representation by DBN with the following methods:

[6] `http://www.csie.ntu.edu.tw/~cjlin/liblinear/`

- BOW: Bag-Of-Word representation [8] is widely used feature for opinion and emotion classification.
- BF: Basic Feature (in subsection 2.3) is re-used as a sentence representation.
- PCA: Principal Components Analysis [22] is adopted to find the directions of the greatest K variances in the dataset and represent each data by these directions.
- LDA: Latent Dirichlet Allocation [23] is utilized to map each sentence into the topic space, whose dimension is K.

In this set of experiments, pseudo-labeled corpus is used for representation learning in the PCA and DBN algorithm. To make fair comparison with DBN, K is set to 500 in the PCA algorithm. Experimental results are listed in Table 4. Each line corresponds to a kind of feature for emotion classification. For example, the first line (BOW) means that bag-of-words are used as features and the last line (BF + DBN) indicates that we use the composition of basic feature and representation learned by DBN.

Table 4. Experimental Results for Emotion Classification on Weibo

Method		Accuracy(%)
Text Feature	BOW	69.97
	BF	72.03
Learned Feature	PCA	70.54
	LDA	67.72
	DBN	**73.28**
Combined Feature	BF + PCA	72.46
	BF + LDA	70.19
	BF + DBN	**75.60**

By comparing the results of different methods, we draw the following observations:

(1) Using the pseudo-labeled data, the sentence representation learned by DBN based method achieves better than PCA and LDA based method for the task of emotion classification. In the setting of using the learned representation as feature individually, DBN-based method (73.28% in Line 4) outperforms PCA (70.54% in Line 3) by 2.74 points in accuracy. In the setting of combining basic feature with learned representation, DBN (75.60% in Line 6) outperforms PCA (72.46% in Line 5) by 3.14 points in accuracy. The same trend is observed when DBN is compared with LDA method.

(2) By concatenating the DBN based representation with basic features, the performance is further improved. The incorporation of DBN based representation (75.60%) improves the basic feature (72.03% in Line 2) by 3.57% points in accuracy. The introduce of DBN has positive impacts on the emotion classification task.

Sensitivity to Corpus. In order to further investigate the sensitivity of the DBN based method to the corpus for representation learning, in Table 5, we show the emotion classification accuracy using the learned representation as individual feature. Each line corresponds to one type of corpus used for representation learning. For example, the first line means that the small-scale labeled data is used to learn representation by DBN algorithm.

Table 5. Sensitivity of DBN to Corpus for Representation Learning

Corpus	Accuracy(%)
Labeled Data	67.31
Unlabeled Data	71.55
Pseudo-Labeled Data	**73.28**

By comparing the results of different corpus, we draw the following observations:

(1) Compared with pseudo-labeled corpus, small-scale labeled data is not sufficient for represent learning based on DBN algorithm. In Table 5, compared with using pseudo-labeled corpus (73.28% in Line 3), the accuracy achieved using the labeled data(67.31% in Line 1) decreases by 5.97 points in accuracy, which is worse that the basic feature (72.03%) by 4.72 points. The experimental results demonstrate that DBN algorithm is sensitive to the training corpus.

(2) Compared with unlabeled data, representation learned from the same number of pseudo-labeled corpus achieves better performance for emotion classification. The representation learned from pseudo-labeled corpus (73.28% in Line 3) outperforms the one learned from unlabeled data (71.55% in Line 2). The introduce of pseudo-labeled corpus has positive impacts on representation learning for emotion classification.

4 Previous Work

With the popularity of blogs and social media, sentiment analysis has become a hot point in natural language processing research community. Overall, sentiment analysis on microblogs could be viewed from two perspectives, target-independent [3, 5] and target-dependent [4] sentiment analysis. This paper studies the task of target-independent emotion classification.

4.1 Emotion Classification and Twitter Sentiment Analysis

The original attempt of sentiment analysis [9, 24] aims to classify whether a whole document expresses a positive or negative sentiment. [9] treat the sentiment classification of reviews as a special case of text categorization problem and first investigate machine learning methods. In their experiments, the best performance is achieved by SVMs with bag-of-words representation. Apart from positive and negative evaluations, some researchers aim to identify the emotion of text, such as happy, sad, angry, etc. [7] uses emoticons labelled by the blogger to collect corpus in LiveJournal. And similar with [9], SVMs is utilized to train a emotion classifier with a variety of features over 100 emotions. Mishne and [25] use a similar method to identify words and phrases in order to estimate aggregate emotion levels across a large number of blog posts. [6] combine SVMs and CRF for emotion classification at the document level. As social media become popular, Twitter sentiment analysis attracts much researcher's attention. [8] collect positive and negative data automatically with emoticons such as :-) and :-(. [12, 26] go further and use both hashtags and smileys to collect corpus. In addition, they use a KNN-like classifier for multiple emotion classification. [15] leverage three sources with noisy labels as training data and use a SVM classifier with a set of features. From a different perspective, [13] train a language model based on the manually labelled data, and then use the noisy emoticon data for smoothing. However, the majority of existing methods use the bag-of-word representation, which cannot capture the complex linguistic phenomena.

4.2 Deep Learning for NLP

The recent revival of interest deep learning, or representation learning [27], has a strong impact in the area of Natural Language Processing, such as multi-task learning [28], domain adaptation [29], parsing [30], entity disambiguation [31], etc. Majority of the existing work are based on word embedding, which means learning a distributed representation for each word [32]. In sentiment analysis, [33, 34] propose the Semi-Supervised Recursive Autoencoders for sentiment distribution prediction. [35] learn word vectors capturing semantic term-document information for document-level sentiment classification. [36] propose a deep learning approach based on Stacked Denoising Autoencoders to study the problem of domain adaptation for sentiment classification. Glorot et al verified the effectiveness of unlabelled data for domain adaptation. However, these methods mostly need labeled data or crowd intelligence. In this work, we use pseudo-labeled corpus, which is extracted automatically from microblogs, to learn the sentence representation for emotion classification on microblogs, in particular, Weibo.

5 Conclusion and Future Work

In this paper, we propose a deep learning approach that automatically learns sentence representation for emotion classification on microblogs. The sentence representation is learned leveraging pseudo labeled corpus, without any manual effort of annotating messages. Experiment reveals the importance of DBN algorithm and the usefulness of pseudo-labeled corpus in this field. By incorporating the DBN based representation into basic features, performance is further improved.

As to future work, the first plan is to learn word representation (word embedding) for emotion classification. The sentence is potential to have positive impacts on emotion classification based on a more meaningful word representation. In addition, as we observed that when pseudo-labeled data, whose size is larger than labeled data, was used, better performance was achieved. So a natural question is whether the performance will continue to increase with even more pseudo-labeled corpus.

Acknowledgments. This work was supported by National Natural Science Foundation of China (NSFC) via grant 61133012, NSFC via grant 61073126 and NSFC via grant 61273321.

We thank Jianyun Nie greatly for valuable comments and helpful suggestions, and thank Yaming Sun for refining the language of this paper, and thank Qiuhui Shi for preparing dataset. We would also thank the anonymous reviewers for their comments and suggestions.

References

1. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2(1-2), 1–135 (2008)
2. Liu, B.: Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies 5(1), 1–167 (2012)
3. Hu, X., Tang, L., Tang, J., Liu, H.: Exploiting social relations for sentiment analysis in microblogging. In: Proceedings of the Sixth WSDM, pp. 537–546. ACM (2013)
4. Jiang, L., Yu, M., Zhou, M., Liu, X., Zhao, T.: Target-dependent twitter sentiment classification. In: Proc. 49th ACL: HLT, vol. 1, pp. 151–160 (2011)
5. Kouloumpis, E., Wilson, T., Moore, J.: Twitter sentiment analysis: The good the bad and the omg! In: Fifth International AAAI Conference on Weblogs and Social Media (2011)
6. Yang, C., Lin, K., Chen, H.: Emotion classification using web blog corpora. In: IEEE/WIC/ACM International Conference on Web Intelligence, pp. 275–278. IEEE (2007)
7. Mishne, G.: Experiments with mood classification in blog posts. In: Proceedings of ACM SIGIR 2005 Workshop on Stylistic Analysis of Text for Information Access, p. 19 (2005)
8. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1–12 (2009)
9. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the Conference on EMNLP, pp. 79–86. ACL (2002)
10. Dahl, G.E., Yu, D., Deng, L., Acero, A.: Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. IEEE Transactions on Audio, Speech, and Language Processing 20(1), 30–42 (2012)
11. Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional neural networks. In: NIPS, pp. 1106–1114 (2012)
12. Read, J.: Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In: Proceedings of the ACL Student Research Workshop, pp. 43–48 (2005)
13. Liu, K., Li, W., Guo, M.: Emoticon smoothed language models for twitter sentiment analysis. In: Twenty-Sixth AAAI Conference on Artificial Intelligence (2012)
14. Li, F., Pan, S.J., Jin, O., Yang, Q., Zhu, X.: Cross-domain co-extraction of sentiment and topic lexicons. In: Proceedings of the 50th ACL, pp. 410–419. ACL (July 2012)
15. Barbosa, L., Feng, J.: Robust sentiment detection on twitter from biased and noisy data. In: Proceedings of the 23rd COLING Posters, pp. 36–44. ACL (2010)
16. Johansson, R., Moschitti, A.: Extracting opinion expressions and their polarities–exploration of pipelines and joint models. In: Proceedings of ACL, vol. 11 (2011)
17. Wiebe, J., Wilson, T., Cardie, C.: Annotating expressions of opinions and emotions in language. Language Resources and Evaluation 39(2-3), 165–210 (2005)
18. Salakhutdinov, R., Hinton, G.: Semantic hashing. International Journal of Approximate Reasoning 50(7), 969–978 (2009)
19. Hinton, G., Salakhutdinov, R.: Reducing the dimensionality of data with neural networks. Science 313(5786), 504–507 (2006)
20. Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H.: Greedy layer-wise training of deep networks. Advances in Neural Information Processing Systems 19, 153 (2007)
21. Hinton, G.: Training products of experts by minimizing contrastive divergence. Neural Computation 14(8), 1771–1800 (2002)
22. Hotelling, H.: Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology 24(6), 417 (1933)
23. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. The Journal of Machine Learning Research 3, 993–1022 (2003)

24. Turney, P.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th ACL, pp. 417–424. ACL (2002)
25. Mishne, G., De Rijke, M.: Capturing global mood levels using blog posts. In: AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs, pp. 145–152 (2006)
26. Davidov, D., Tsur, O., Rappoport, A.: Enhanced sentiment learning using twitter hashtags and smileys. In: Proceedings of the 23rd COLING: Posters, pp. 241–249. ACL (2010)
27. Bengio, Y., Courville, A., Vincent, P.: Representation learning: A review and new perspectives. arXiv preprint arXiv:1206.5538 (2012)
28. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. JMLR 12, 2493–2537 (2011)
29. Chen, M., Xu, Z., Weinberger, K., Sha, F.: Marginalized denoising autoencoders for domain adaptation. In: ICML (2012)
30. Socher, R., Bauer, J., Manning, C.D., Ng, A.Y.: Parsing with compositional vector grammars. In: ACL (2013)
31. He, Z., Liu, S., Li, M., Zhou, M., Zhang, L., Wang, H.: Learning entity representation for entity disambiguation. In: ACL (2013)
32. Turian, J., Ratinov, L., Bengio, Y.: Word representations: a simple and general method for semi-supervised learning. Urbana 51, 61801 (2010)
33. Socher, R., Pennington, J., Huang, E., Ng, A., Manning, C.: Semi-supervised recursive autoencoders for predicting sentiment distributions. In: EMNLP, pp. 151–161 (2011)
34. Socher, R., Huval, B., Manning, C.D., Ng, A.Y.: Semantic Compositionality Through Recursive Matrix-Vector Spaces. In: Proceedings of the 2012 Conference on EMNLP (2012)
35. Maas, A.L., Daly, R., Pham, P., Huang, D., Ng, A., Potts, C.: Learning word vectors for sentiment analysis. In: Proceedings of the 49th ACL, ACL 2011 (2011)
36. Glorot, X., Bordes, A., Bengio, Y.: Domain adaptation for large-scale sentiment classification: A deep learning approach. In: Proceedings of ICML (2011)