Cross-Lingual Sentiment Classification Based on Denoising Autoencoder

Huiwei Zhou, Long Chen, and Degen Huang

School of Computer Science and Technology,
Dalian University of Technology, Dalian 116024, Liaoning, China
{zhouhuiwei,huangdg}@dlut.edu.cn, chenlong.415@mail.dlut.edu.cn

Abstract. Sentiment classification system relies on high-quality emotional resources. However, these resources are imbalanced in different languages. The way of how to leverage rich labeled data of one language (source language) for the sentiment classification of resource-poor language (target language), namely cross-lingual sentiment classification (CLSC), becomes a focus topic. This paper utilizes rich English resources for Chinese sentiment classification. To eliminate the language gap between English and Chinese, this paper proposes a combination CLSC approach based on denoising autoencoder. First, two classifiers based on denoising autoencoder are learned respectively in English and Chinese views by using English corpus and English-to-Chinese corpus. Second, we classify Chinese test data and Chinese-to-English test data with the two classifiers trained in the two views. Last, the final sentiment classification results are obtained by the combination of the two results in two views. Experiments are carried out on NLP&CC 2013 CLSC dataset including book, DVD and music categories. The results show that our approach achieves the accuracy of 80.02%, which outperforms the current state-of-the-art systems.

Keywords: cross-lingual sentiment classification, combination, denoising autoencoder.

1 Introduction

Sentiment classification technique is the task of predicting sentiment polarity for a given text. It could help to mine public opinion from product reviews. Along with the rapid expansion of user-generated information, sentiment classification plays a key role in analyzing a large number of sentiment reviews on the Web.

Researches on sentiment classification have been developed rapidly. Generally, sentiment classification approaches can be divided into two categories: lexicon based approach and machine learning based approach. Lexicon based approach extracts the sentiment words in texts, and identifies the positive or negative polarities of texts based on the sentiment words' polarities in lexicon [1-2]. This method is easily implemented, and could achieve a reasonable accuracy on the foundation of an elaborate lexicon. However, it is difficult to identify the polarities of sentiment words in texts since their polarities may be changed in different context [3]. Machine learning based

C. Zong et al. (Eds.): NLPCC 2014, CCIS 496, pp. 181–192, 2014.

approach trains sentiment classifiers based on the context of sentiment words [4-6]. This approach identifies the polarities of texts more accurately than lexicon based approach and is widely used in the sentiment classification task now. However, the method heavily relies on the quality and quantity of corpora, which are considered as valuable resources in sentiment classification task.

Sentiment classification in English has been studied for a long time, and many labeled data for English sentiment classification are available on the Web. However, labeled data in different languages are very imbalanced. The lack of sentiment resources limits the research progress in some languages. In order to overcome this obstacle, cross-lingual sentiment classification (CLSC) [7-10] is proposed, which leverages resources on one language (source language) to resource-poor language (target language) for improving the performance of sentiment classification on target language.

Machine translation services are usually adopted to eliminate the gap between source language and target language in CLSC task. Wan [7] proposed a co-training approach for CLSC, which leveraged an English corpus for Chinese sentiment classification in two independent views: English view and Chinese view. To further improve the performance of co-training approach, Gui et al. [8] incorporated bilingual cross-lingual self-training and co-training approaches by estimating the confidence of each monolingual system. To reduce the impact of translation errors, Li et al. [9] selected high-quality translated samples in the source language.

These methods all adopted shallow learning algorithms, which were optimized based on limited computing units. However, shallow learning algorithms can hardly learn the kind of complicated functions that can discover intermediate representation of the input [11]. Deep learning algorithms could learn intermediate representations through multi-layer non-linear operations in the function learning [12-14]. Searching the parameter space of deep architecture is a difficult task. Bengio et al. [11] proposed an optimization principle to solve this problem, which has worked well for deep belief network (DBN) and autoencoder (AE) [15]. To further make the learned representations robust to partial destroyed input, Vincent et al. [16] raised denoising autoencoder model.

Deep learning has gained huge success in many real-world applications such as computer vision [17] and speech recognition [18]. Collobert et al. [19] applied deep learning to natural language processing (NLP), and proposed a multi-task learning system SENNA including part-of-speech (POS) tagging, chunking, named entity recognition, and semantic role labeling. Tang et al. [20] and Zhou et al. [21] studied the use of deep learning for representation learning on sentiment classification. The representation features learned from deep learning could improve the performance of sentiment classification effectively.

In CLSC task, the language gap between the original language and the translated language greatly influences the classification performance. This paper proposes a combination CLSC approach based on denoising autoencoder to decrease the effects of noisy translated examples from two aspects. On the one hand, denoising autoencoder is adopted to improve the robustness to translation noises. On the other hand, two classifiers are trained in English view and Chinese view, respectively. The final

results are obtained by combining the two classification outputs to eliminate the language gap. Though this paper leverages English corpora for Chinese sentiment classification, the proposed CLSC approach can be applied to other languages. In our systems, χ^2 (CHI) statistical method is used to select sentiment word features, and TF-IDF method is used to set feature weights. The experiments are conducted using NLP&CC 2013 CLSC dataset. The experimental results show that the proposed approach outperforms the current state-of-the-art systems.

The rest of this paper is organized as follows: Section 2 describes the structure of denoising autoencoder. Section 3 presents the combination CLSC approach in detail. Section 4 reports experimental results and analysis. Section 5 concludes our work and gives the future work.

2 Denoising Autoencoder

The traditional autoencoder maps an input vector $\mathbf{x} \in [0,1]^d$ to a hidden representation $\mathbf{y} \in [0,1]^{d'}$, through a deterministic mapping $\mathbf{y} = f_\theta(\mathbf{x}) = s(\mathbf{W}\mathbf{x} + \mathbf{b})$, parameterized by $\theta = \{\mathbf{W}, \mathbf{b}\}$. $\mathbf{W}$ is a $d' \times d$ weight matrix, $\mathbf{b}$ is a bias vector. Then, $\mathbf{y}$ is mapped back to a reconstructed vector $\mathbf{z} \in [0,1]^d$ in input space $\mathbf{z} = g_{\theta'}(\mathbf{y}) = s(\mathbf{W}'\mathbf{y} + \mathbf{b}')$ with parameters $\theta' = \{\mathbf{W}', \mathbf{b}'\}$, where $\mathbf{W}' = \mathbf{W}^T$. In the training phase, each training sample $\mathbf{x}^{(i)}$ is mapped to a latent representation $\mathbf{y}^{(i)}$ and a reconstruction $\mathbf{z}^{(i)}$. The parameters of this model are optimized by stochastic gradient descent (SGD) algorithm to minimize the average reconstruction error:

$$\theta^*, \theta'^* = \arg\min_{\theta,\theta'} \frac{1}{N}\sum_{i=1}^{n} L(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) = \arg\min_{\theta,\theta'} \frac{1}{N}\sum_{i=1}^{n} L(\mathbf{x}^{(i)}, g_{\theta'}(f_\theta(\mathbf{x}^{(i)}))) \tag{1}$$

where N is the number of training samples, L is a loss function and usually defined as reconstruction cross-entropy [18]:

$$L_H(\mathbf{x}, \mathbf{z}) = H(B_{\mathbf{x}} \parallel B_{\mathbf{z}}) = -\sum_{k=1}^{d} [\mathbf{x}_k \log \mathbf{z}_k + (1 - \mathbf{x}_k) \log(1 - \mathbf{z}_k)] \tag{2}$$

Denoising autoencoder is proposed as a modification of the autoencoder and enforces robustness to partially destroyed inputs. The training process of denoising autoencoder is shown in Fig. 1.

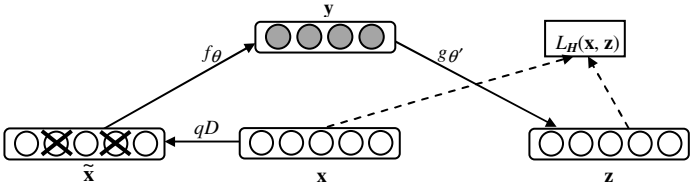


Fig. 1. Structure of denoising autoencoder

For each initial input vector $\mathbf{x}$, a partially destroyed version $\tilde{\mathbf{x}}$ is acquired by means of a stochastic mapping $\tilde{\mathbf{x}} \sim qD(\tilde{\mathbf{x}} \mid \mathbf{x})$. A fixed number νd of input components are randomly chosen, and their values are forced to 0, while the others are not changed. ν is the destruction fraction, by which the desired noise level could be adjusted. Denoising autoencoder maps $\tilde{\mathbf{x}}$ instead of $\mathbf{x}$ to a hidden representation $\mathbf{y}$ through a deterministic mapping $\mathbf{y} = f_\theta(\tilde{\mathbf{x}}) = s(\mathbf{W}\tilde{\mathbf{x}} + \mathbf{b})$, from which we reconstruct a $\mathbf{z} = g_{\theta'}(\mathbf{y}) = s(\mathbf{W}'\mathbf{y} + \mathbf{b}')$. The loss function L_H is also defined as formula (2), where $\mathbf{z}$ is a deterministic function of $\tilde{\mathbf{x}}$ rather than $\mathbf{x}$.

3 The Combination CLSC Approach Based on Denoising Autoencoder

3.1 The Combination CLSC Algorithm

The combination CLSC approach trains sentiment classifier based on English view and Chinese view, respectively. This paper utilizes the labeled English reviews to classify unlabeled Chinese reviews. The framework of the proposed approach consists of a training phase and a classification phase as shown in Fig. 2. In training phase, English labeled reviews are used to train an English classifier in English view. Meanwhile, English-to-Chinese translated labeled reviews are used to train a Chinese

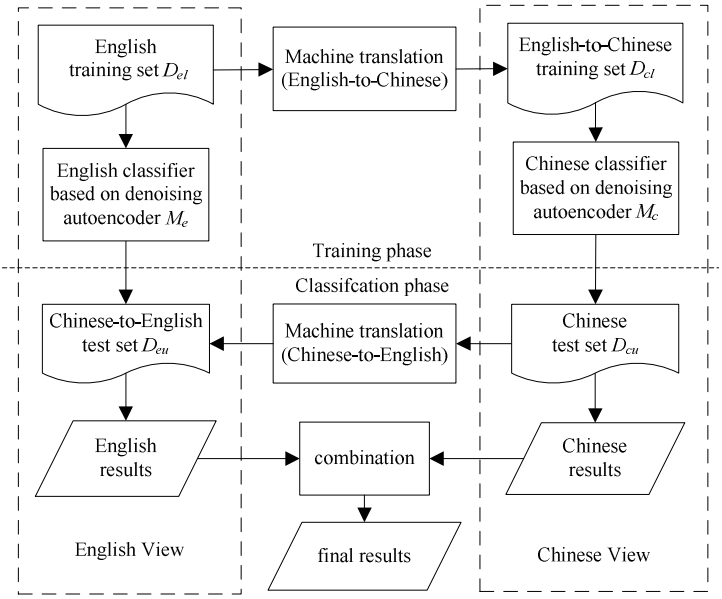


Fig. 2. Framework of the combination CLSC approach

classifier in Chinese view. In classification phase, the English classifier is applied to classify Chinese-to-English translated reviews in English view. Similarly, the Chinese classifier is applied to classify Chinese reviews in Chinese view. The two classification results from English and Chinese views are combined into the final results by comparing positive and negative possibilities for each review.

The combination CLSC algorithm is described in details in the following:

Algorithm. Combination CLSC Algorithm

Input: English training set D_{el}, and its English-to-Chinese version D_{cl};
Chinese test set D_{cu}, and its Chinese-to-English version D_{eu}.

For each review R_i:

1. Train English classifier M_e based on D_{el};
2. Use M_e to classify R_i in D_{eu};
3. Obtain the positive possibility $P_e(R_i)$ and the negative possibility $N_e(R_i)$;
4. Train Chinese classifier M_c based on D_{cl};
5. Use M_c to classify R_i in D_{cu};
6. Obtain the positive possibility $P_c(R_i)$ and the negative possibility $N_c(R_i)$;
7. Calculate the positive possibility: $P(R_i) = (P_e(R_i) + P_c(R_i))/2$;
8. Calculate the negative possibility: $N(R_i) = (N_e(R_i) + N_c(R_i))/2$;
9. If $(P(R_i) > N(R_i))$, then output positive review R_i;
10. Else output negative review R_i;

3.2 Feature Setting

(1) Sentiment Word Features Selection

Words in sentiment lexicon are too many to be all used as sentiment word features. Statistical methods are usually adopted to select effective sentiment word features. This paper investigates the influence of feature selection based on high-frequency words method and CHI statistical method respectively.

High-Frequency Words Method: This method selects the top 2000 high-frequency words as sentiment word features.

CHI Statistical Method: This method enables to measure the association between feature t_i and class C_j [22]. Chi-square value between feature t_i and class C_j is defined as follows:

$$\chi^2(t_i, C_j) = \frac{N \times (A \times D - B \times C)^2}{(A+C) \times (B+D) \times (A+B) \times (C+D)} \tag{3}$$

where *A, B, C,* and *D* denote the co-occurrence frequency between a feature t_i and a class C_j. They are corresponding to the case of (t_i, C_j), $(t_i, \overline{C_j})$, $(\overline{t_i}, C_j)$, $(\overline{t_i}, \overline{C_j})$, while *N* is the sum of them. If $A \times D - B \times C > 0$, feature t_i has a positive correlation

with C_j, otherwise a negative correlation. Therefore, the features with higher CHI-square values $\chi^2(t_i, C_j)$ will be more discriminative. This paper selects the top 2000 words with high CHI-square values as sentiment word features.

(2) Negation Features

Some sentiment words are often modified by negation words, which leads to inversion of their polarities. We take into account 14 usually used negation words in English such as "not", "without", and "none"; 5 negation words in Chinese such as "不" (no/not), "不会" (cannot), and "没有" (without). A sentiment word modified by these negation words in the window [-2, 2] expresses the opposite polarity. A negation feature is introduced to each sentiment word to represent the negative form of this word. We simply insert negation feature in front of each sentiment word in a feature vector. Meanwhile, sentiment word features remain the initial meaning. If there is no negation word in the window, the value of negation feature is set to 0. The feature vector containing negation features is extended to 4000 dimensions:

$$vector = (neg_1, sent_1, ..., neg_i, sent_i, ..., neg_{2000}, sent_{2000}) \text{ , } (i=1, 2, \ldots, 2000) \quad (4)$$

where $sent_i$ denotes sentiment word feature, and neg_i represents negation feature.

(3) The Feature Weight Calculation

The following three methods of feature weight calculation are investigated in this paper: Boolean method, Term-frequency method, and TF-IDF method.

Boolean Method: This method thinks that all feature words have the same importance to classification, and sets a feature weight to 0 or 1 only depending on whether it appears in a review.

Word Frequency Method: This method takes word frequency into account in sentiment classification. In our work, the frequency of each word in each review is calculated and normalized as feature weight.

TF-IDF Method: This method takes word frequency (TF) and inverse document frequency (IDF) into consideration [23]. The feature weight is calculated by the following formula:

$$w_{ij} = tf_{ij} \times \log \frac{N}{n_i} \quad (5)$$

where N is the total number of reviews in corpus, tf_{ij} is the frequency of the word i occurring in the review j, n_i is the number of reviews containing word i. In this way, word frequency and review frequency of word are both considered in weight calculation.

Besides, the sentiment words in the summary are more important than other parts of a review. To highlight the importance of these words, sentiment words surrounded by the "<*summary*>" tag in each review are counted twice in the frequency calculation in word frequency method and TF-IDF method.

4 Experimental Results and Analysis

4.1 Experimental Settings

The proposed approach is evaluated on NLP&CC 2013 CLSC dataset[1,2]. The dataset consists of reviews on three categories: Book, DVD and Music. Each category contains 4,000 English labeled data (ratio of positive and negative examples is 1:1), 4,000 Chinese unlabeled data.

In the experiments, *Google Translate*[3] is adopted for both English-to-Chinese translation and Chinese-to-English translation. Geniatagger[4] is used as POS tagging tool and ICTCLAS[5] is used as Chinese word segmentation tool. Denoising autoencoder is developed based on Theano system[6] [24].

The architecture of denoising autoencoder used in our experiments is 4000-500-2, which represents the number of units in input layer is 4000, in a hidden layer is 500, and in output layer is 2. The output layer is a softmax layer, where 2 units denote the possibility scores of a review being positive and negative, and their sum is 1. For layer-wise unsupervised learning, we train the weights of each layer independently with the fixed number of epochs equal to 30 and the learning rate is set to 0.1.

The performance is evaluated by the correct classification accuracy for each category, and the average accuracy of three categories, respectively. The category accuracy is defined as:

$$Accuracy_c = \frac{\#system\_correct}{4000} \tag{6}$$

where c is one of the three categories, and $\#\ system\_correct$ stands for the number of correctly classified reviews in c.

The average accuracy is defined as:

$$Accuracy = \frac{1}{3}\sum_{i=1}^{3} Accuracy_c \tag{7}$$

4.2 Evaluation on Combination CLSC Approach

In this section, the experiments evaluate the performance of our feature selection methods and combination CLSC approach. The destruction fraction ν is set to 0.1.

---

1 http://tcci.ccf.org.cn/conference/2013/dldoc/evsam03.zip
2 http://tcci.ccf.org.cn/conference/2013/dldoc/evdata03.zip
3 http://translate.google.cn/
4 http://www.nactem.ac.uk/tsujii/GENIA/tagger/
5 http://www.ictclas.org/ictclas_download.aspx
6 http://deeplearning.net/software/theano/

(1) Effect of Sentiment Word Features Selection

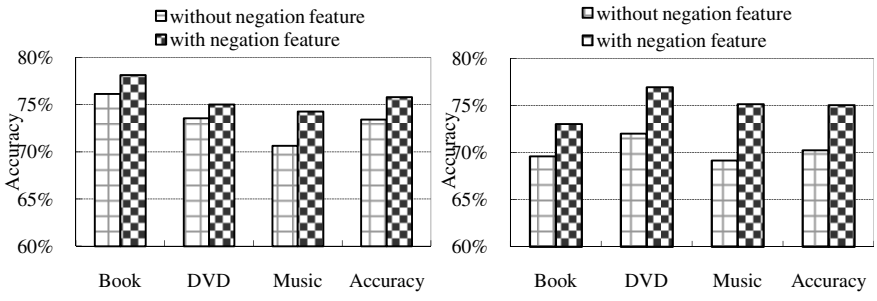
We first evaluate the classification performance in English and Chinese systems, respectively. Table 1 shows effect of sentiment word features selection methods in English and Chinese systems. As shown in Table 1, CHI statistical method outperfoms high-frequency words method in DVD and Music categories. The average accuracies of CHI statistical method are 0.46% and 0.50% higher than high-frequency words method in English and Chinese systems, respectively. From these results, we can conclude that CHI statistical method is more effective than high-frequency words method in classification task. The sentiment word features selected with CHI statistical method are used in the following experiments.

Table 1. Effect of Sentiment Word Features Selection

System	Methods	Book	DVD	Music	Accuracy
English	High-frequency	**74.53%**	75.43%	73.8%	74.58%
	CHI statistic	73.03%	**76.93%**	**75.15%**	**75.04 %(+0.46%)**
Chinese	High-frequency	**78.40%**	74.45%	73.15%	75.33%
	CHI statistic	78.15%	**75.05%**	**74.30%**	**75.83 %(+0.50%)**

(2) Effect of Negation Features

The performance of classification systems with or without negation features in Chinese and English systems are given in Fig. 3, respectively. The performance with negation features are steadly better than the performance without them. In most cases, the improvement is considerable. Negation features improve the accuracies by 2.37% (from 73.46% to 75.83%) in Chinese systems and 4.79% (from 70.25% to 75.04%) in English systems. Negation features in English systems are more effective than those in Chinese systems, which perhaps because the English negation words used in English systems are more than in Chinese systems. Negation features are employed in our following experiments.

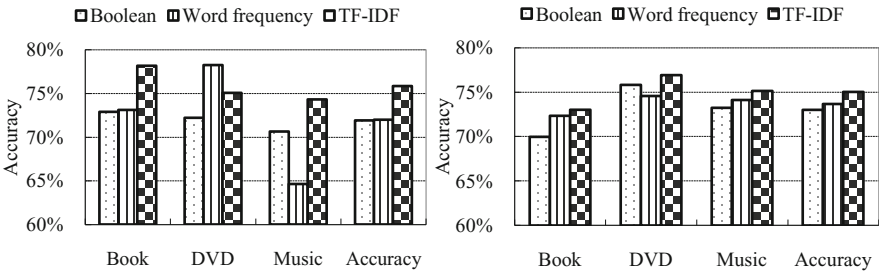


(a) Chinese systems (b) English systems

Fig. 3. Performance comparison with or without negation features

(3) Effect of Feature Weight Calculation Methods

Boolean, word frequency and TF-IDF feature weight calculation methods are compared in Fig. 4. Generally, TF-IDF method achieves the best accuracy, while Boolean method performs poorly in sentiment classification task. TF-IDF method would reflect the latent contribution of feature words to the reviews, compared with Boolean and word frequency methods. The feature setting of CHI feature selection method together with TF-IDF weight calculation method achieves the best performance in all categories, which are exploited for the following combination CLSC systems.



(a) Chinese system (b) English system

Fig. 4. Performance comparison with different weight calculation methods

(4) Performance of Combination CLSC Systems

Table 2 evaluates the combination CLSC systems which incorporate English and Chinese systems. As can be seen in Table 2, the combination CLSC systems further improve the English and Chinese systems performances. In Book, DVD and Music categories, the combination CLSC systems improve 1.53%, 1.40% and 2.93% respectively, compared to the better one of English and Chinese systems. The results illustrate that the combination CLSC system increases the possibility of a review being correctly predicted. The combination of English and Chinese systems could effectively eliminate the gap between the two languages.

Table 2. Performance of combination CLSC systems

System	Book	DVD	Music	Accuracy
English system	73.03%	76.93%	75.15%	75.04%
Chinese system	78.15%	75.05%	74.30%	75.83%
Combination system	**79.68%**	**78.33%**	**78.08%**	**78.70%**

4.3 Effect of Destruction Fraction in Denoising Autoencoders

Fig. 5 shows the accuracy curve of the CLSC systems with different destruction fractions. The destruction fraction ν varies from 0 to 0.9. When ν is set to 0, denoising autoencoders are degenerated into autoencoders. As can be seen from the figure,

denoising autoencoders outperform autoencoders from $\nu = 0.2$ to $\nu = 0.5$. We can conclude that adding noises to the training examples properly enhances the robustness to the translation noises in the CLSC task. Denoising autoencoders with $\nu = 0.2$ achieve the highest average accuracy 80.02%, which surpass the original autoencoders by 0.94%.

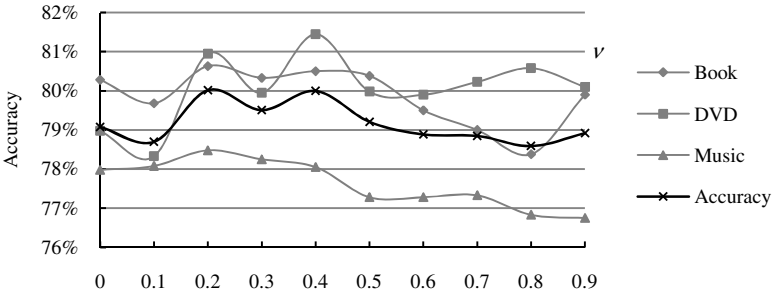


Fig. 5. Accuracy vs. Destruction fraction

4.4 Comparison with Related Work

Table 3 summarizes recent results on NLP&CC 2013 CLSC dataset. Chen et al. [10] gave different weights to sentiment words according to emotional color difference of sentiment words in subject-predicate component. In NLP&CC 2013 CLSC evaluation, they achieved the second place of accuracy. The top performer of NLP&CC 2013 CLSC share task is HLT-Hitsz system, which obtained an accuracy 77.12% by using the co-training model. They extended their model by incorporating transfer self-training model and co-training model [8], and achieved 78.89% accuracy. These methods are all based on shallow learning algorithms. We build a deep architecture to discover intermediate representation of features and achieve 80.02% accuracy.

Table 3. CLSC performance comparison on the NLP&CC 2013 Share Task test data

Team	Book	DVD	Music	Accuracy
Chen et al. [10]	77.00%	78.33%	75.95%	77.09%
HLT-Hitsz	78.50%	77.73%	75.13%	77.12%
Gui et al. [8]	78.70%	79.65%	78.30%	78.89%
Our Approach	**80.63%**	**80.95%**	**78.48%**	**80.02%**

5 Conclusion and Future Work

This paper proposes a combination CLSC approach based on denoising autoencoder to eliminate the language gap between English and Chinese. Experimental results on NLP&CC 2013 CLSC dataset show that both of the denoising autoencoder and

combination approach could improve the sentiment classification performance. In addition, we show that the feature setting of CHI feature selection method together with TF-IDF weight calculation method works well on CLSC task.

This work only combines English and Chinese systems linearly, and it's very likely that better performance could be achieved by deep combination, such as co-training and transfer learning, etc. We leave that as a future work. Meanwhile, we will select high-quality translated reviews to further reduce the impacts of translation errors.

Acknowledgements. This research is supported by Natural Science Foundation of China (No. 61272375, No. 61173100, and No. 61173101).

References

1. Turney, P.D.: Thumbs up or thumbs down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistic, pp. 417–424. Association for Computational Linguistics (2002)
2. Wei, B., Pal, C.: Cross lingual adaptation: an experiment on sentiment classifications. In: Proceedings of the ACL 2010 Conference Short Papers, pp. 258–262. Association for Computational Linguistics (2010)
3. Zhao, Y.Y., Qin, B., Liu, T.: Sentiment Analysis (in Chinese). Journal of Software 21(8), 1834–1848 (2010)
4. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment Classification using Machine Learning Techniques. In: Proceedings of the ACL-2002 Conference on Empirical Methods in Natural Language Processing, vol. 10, pp. 79–86. Association for Computational Linguistics (2002)
5. Kennedy, A., Inkpen, D.: Sentiment classification of movie reviews using contextual valence shifters. Computational Intelligence 22(2), 110–125 (2006)
6. Li, S., Xia, R., Zong, C.Q., Huang, C.R.: A framework of feature selection methods for text categorization. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th IJFNLP of the AFNLP, pp. 692–700. ACL and AFNLP (2009)
7. Wan, X.J.: Co-training for cross-lingual sentiment classification. In: Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, vol. 1, pp. 235–243. Association for Computational Linguistics (2009)
8. Gui, L., Xu, R., Xu, J., Yuan, L., Yao, Y., Zhou, J., Qiu, Q., Wang, S., Wong, K.-F., Cheung, R.: A Mixed Model for Cross Lingual Opinion Analysis. In: Zhou, G., Li, J., Zhao, D., Feng, Y. (eds.) NLPCC 2013. CCIS, vol. 400, pp. 93–104. Springer, Heidelberg (2013)
9. Li, S., Wang, R., Liu, H., Huang, C.-R.: Active Learning for Cross-Lingual Sentiment Classification. In: Zhou, G., Li, J., Zhao, D., Feng, Y. (eds.) NLPCC 2013. CCIS, vol. 400, pp. 236–246. Springer, Heidelberg (2013)
10. Chen, Q., He, Y.X., Liu, X.L., Sun, S.T., Peng, M., Li, F.: Cross-Language Sentiment Analysis Based on Parser. Acta Scientiarum Naturalium Universitatis Pekinensis 50(1), 55–60 (2014) (in Chinese)
11. Bengio, Y.: Learning deep architectures for AI. Foundations and Trends in Machine Learning 2, 1–127 (2009)

12. Bengio, Y., Delalleau, O.: On the expressive power of deep architectures. In: Kivinen, J., Szepesvári, C., Ukkonen, E., Zeugmann, T. (eds.) ALT 2011. LNCS, vol. 6925, pp. 18–36. Springer, Heidelberg (2011)
13. Bengio, Y., LeCun, Y.: Scaling learning algorithms towards AI. Large-Scale Kernel Machines 34, 1–41 (2007)
14. Hinton, G., Salakhutdinov, R.: Reducing the dimensionality of data with neural networks. Science 313(5786), 504–507 (2006)
15. Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H.: Greedy layer-wise training of deep networks. Advances in Neural Information Processing Systems 19, 153 (2007)
16. Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.: Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th International Conference on Machine Learning, pp. 1096–1103. ACM, New York (2008)
17. Farabet, C., Couprie, C., Najman, L., LeCun, Y.: Learning hierarchical features for scene labeling. IEEE Transactions on Pattern Analysis and Machine Intelligence 35(8), 1915–1929 (2013)
18. 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)
19. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. The Journal of Machine Learning Research 12, 2493–2537 (2011)
20. Tang, D., Qin, B., Liu, T., Li, Z.: Learning Sentence Representation for Emotion Classification on Microblogs. In: Zhou, G., Li, J., Zhao, D., Feng, Y. (eds.) NLPCC 2013. CCIS, vol. 400, pp. 212–223. Springer, Heidelberg (2013)
21. Zhou, S.S., Chen, Q.C., Wang, X.L.: Active deep networks for semi-supervised sentiment classification. In: COLING 2010 Proceedings of the 23rd International Conference on Computational Linguistics, pp. 1515–1523. Association for Computational Linguistics (2010)
22. Galavotti, L., Sebastiani, F., Simi, M.: Feature selection and negative evidence in automated text categorization. In: Proceedings of KDD (2000)
23. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Information processing & management 24(5), 513–523 (1988)
24. Bergstra, J., Breuleux, O., Bastien, F., et al.: Theano: a CPU and GPU math expression compiler. In: Proceedings of the Python for Scientific Computing Conference, SciPy (2010)