

Co-Training for Cross-Lingual Sentiment Classification

Xiaojun Wan

Institute of Compute Science and Technology & Key Laboratory of Computational Linguistics, MOE

Peking University, Beijing 100871, China
wanxiaojun@icst.pku.edu.cn

Abstract

The lack of Chinese sentiment corpora limits the research progress on Chinese sentiment classification. However, there are many freely available English sentiment corpora on the Web. This paper focuses on the problem of cross-lingual sentiment classification, which leverages an available English corpus for Chinese sentiment classification by using the English corpus as training data. Machine translation services are used for eliminating the language gap between the training set and test set, and English features and Chinese features are considered as two independent views of the classification problem. We propose a co-training approach to making use of unlabeled Chinese data. Experimental results show the effectiveness of the proposed approach, which can outperform the standard inductive classifiers and the transductive classifiers.

1 Introduction

Sentiment classification is the task of identifying the sentiment polarity of a given text. The sentiment polarity is usually positive or negative and the text genre is usually product review. In recent years, sentiment classification has drawn much attention in the NLP field and it has many useful applications, such as opinion mining and summarization (Liu et al., 2005; Ku et al., 2006; Titov and McDonald, 2008).

To date, a variety of corpus-based methods have been developed for sentiment classification. The methods usually rely heavily on an annotated corpus for training the sentiment classifier. The sentiment corpora are considered as the most valuable resources for the sentiment classification task. However, such resources in different languages are very imbalanced. Because most previous work focuses on English sentiment classification, many annotated corpora for English sentiment classification are freely available on the Web. However, the annotated corpora for

Chinese sentiment classification are scarce and it is not a trivial task to manually label reliable Chinese sentiment corpora. The challenge before us is how to leverage rich English corpora for Chinese sentiment classification. In this study, we focus on the problem of cross-lingual sentiment classification, which leverages only English training data for supervised sentiment classification of Chinese product reviews, without using any Chinese resources. Note that the above problem is not only defined for Chinese sentiment classification, but also for various sentiment analysis tasks in other different languages.

Though pilot studies have been performed to make use of English corpora for subjectivity classification in other languages (Mihalcea et al., 2007; Banea et al., 2008), the methods are very straightforward by directly employing an inductive classifier (e.g. SVM, NB), and the classification performance is far from satisfactory because of the language gap between the original language and the translated language.

In this study, we propose a co-training approach to improving the classification accuracy of polarity identification of Chinese product reviews. Unlabeled Chinese reviews can be fully leveraged in the proposed approach. First, machine translation services are used to translate English training reviews into Chinese reviews and also translate Chinese test reviews and additional unlabeled reviews into English reviews. Then, we can view the classification problem in two independent views: Chinese view with only Chinese features and English view with only English features. We then use the co-training approach to making full use of the two redundant views of features. The SVM classifier is adopted as the basic classifier in the proposed approach. Experimental results show that the proposed approach can outperform the baseline inductive classifiers and the more advanced transductive classifiers.

The rest of this paper is organized as follows: Section 2 introduces related work. The proposed

co-training approach is described in detail in Section 3. Section 4 shows the experimental results. Lastly we conclude this paper in Section 5.

2 Related Work

2.1 Sentiment Classification

Sentiment classification can be performed on words, sentences or documents. In this paper we focus on document sentiment classification. The methods for document sentiment classification can be generally categorized into lexicon-based and corpus-based.

Lexicon-based methods usually involve deriving a sentiment measure for text based on sentiment lexicons. Turney (2002) predicates the sentiment orientation of a review by the average semantic orientation of the phrases in the review that contain adjectives or adverbs, which is denoted as the semantic oriented method. Kim and Hovy (2004) build three models to assign a sentiment category to a given sentence by combining the individual sentiments of sentiment-bearing words. Hiroshi et al. (2004) use the technique of deep language analysis for machine translation to extract sentiment units in text documents. Kennedy and Inkpen (2006) determine the sentiment of a customer review by counting positive and negative terms and taking into account contextual valence shifters, such as negations and intensifiers. Devitt and Ahmad (2007) explore a computable metric of positive or negative polarity in financial news text.

Corpus-based methods usually consider the sentiment analysis task as a classification task and they use a labeled corpus to train a sentiment classifier. Since the work of Pang et al. (2002), various classification models and linguistic features have been proposed to improve the classification performance (Pang and Lee, 2004; Mullen and Collier, 2004; Wilson et al., 2005; Read, 2005). Most recently, McDonald et al. (2007) investigate a structured model for jointly classifying the sentiment of text at varying levels of granularity. Blitzer et al. (2007) investigate domain adaptation for sentiment classifiers, focusing on online reviews for different types of products. Andreevskaia and Bergler (2008) present a new system consisting of the ensemble of a corpus-based classifier and a lexicon-based classifier with precision-based vote weighting.

Chinese sentiment analysis has also been studied (Tsou et al., 2005; Ye et al., 2006; Li and Sun, 2007) and most such work uses similar lexicon-

based or corpus-based methods for Chinese sentiment classification.

To date, several pilot studies have been performed to leverage rich English resources for sentiment analysis in other languages. Standard Naïve Bayes and SVM classifiers have been applied for subjectivity classification in Romanian (Mihalcea et al., 2007; Banea et al., 2008), and the results show that automatic translation is a viable alternative for the construction of resources and tools for subjectivity analysis in a new target language. Wan (2008) focuses on leveraging both Chinese and English lexicons to improve Chinese sentiment analysis by using lexicon-based methods. In this study, we focus on improving the corpus-based method for cross-lingual sentiment classification of Chinese product reviews by developing novel approaches.

2.2 Cross-Domain Text Classification

Cross-domain text classification can be considered as a more general task than cross-lingual sentiment classification. In the problem of cross-domain text classification, the labeled and unlabeled data come from different domains, and their underlying distributions are often different from each other, which violates the basic assumption of traditional classification learning.

To date, many semi-supervised learning algorithms have been developed for addressing the cross-domain text classification problem by transferring knowledge across domains, including Transductive SVM (Joachims, 1999), EM(Nigam et al., 2000), EM-based Naïve Bayes classifier (Dai et al., 2007a), Topic-bridged PLSA (Xue et al., 2008), Co-Clustering based classification (Dai et al., 2007b), two-stage approach (Jiang and Zhai, 2007). DauméIII and Marcu (2006) introduce a statistical formulation of this problem in terms of a simple mixture model.

In particular, several previous studies focus on the problem of cross-lingual text classification, which can be considered as a special case of general cross-domain text classification. Bel et al. (2003) present practical and cost-effective solutions. A few novel models have been proposed to address the problem, e.g. the EM-based algorithm (Rigutini et al., 2005), the information bottleneck approach (Ling et al., 2008), the multi-lingual domain models (Gliozzo and Strapparava, 2005), etc. To the best of our knowledge, co-training has not yet been investigated for cross-domain or cross-lingual text classification.

3 The Co-Training Approach

3.1 Overview

The purpose of our approach is to make use of the annotated English corpus for sentiment polarity identification of Chinese reviews in a supervised framework, without using any Chinese resources. Given the labeled English reviews and unlabeled Chinese reviews, two straightforward methods for addressing the problem are as follows:

1) We first learn a classifier based on the labeled English reviews, and then translate Chinese reviews into English reviews. Lastly, we use the classifier to classify the translated English reviews.

2) We first translate the labeled English reviews into Chinese reviews, and then learn a classifier based on the translated Chinese reviews with labels. Lastly, we use the classifier to classify the unlabeled Chinese reviews.

The above two methods have been used in (Banea et al., 2008) for Romanian subjectivity analysis, but the experimental results are not very promising. As shown in our experiments, the above two methods do not perform well for Chinese sentiment classification, either, because the underlying distribution between the original language and the translated language are different.

In order to address the above problem, we propose to use the co-training approach to make use of some amounts of unlabeled Chinese reviews to improve the classification accuracy. The co-training approach can make full use of both the English features and the Chinese features in a unified framework. The framework of the proposed approach is illustrated in Figure 1.

The framework consists of a training phase and a classification phase. In the training phase, the input is the labeled English reviews and some amounts of unlabeled Chinese reviews¹. The labeled English reviews are translated into labeled Chinese reviews, and the unlabeled Chinese reviews are translated into unlabeled English reviews, by using machine translation services. Therefore, each review is associated with an English version and a Chinese version. The English features and the Chinese features for each review are considered two independent and redundant views of the review. The co-training algorithm is then applied to learn two classifiers

and finally the two classifiers are combined into a single sentiment classifier. In the classification phase, each unlabeled Chinese review for testing is first translated into English review, and then the learned classifier is applied to classify the review into either positive or negative.

The steps of review translation and the co-training algorithm are described in details in the next sections, respectively.

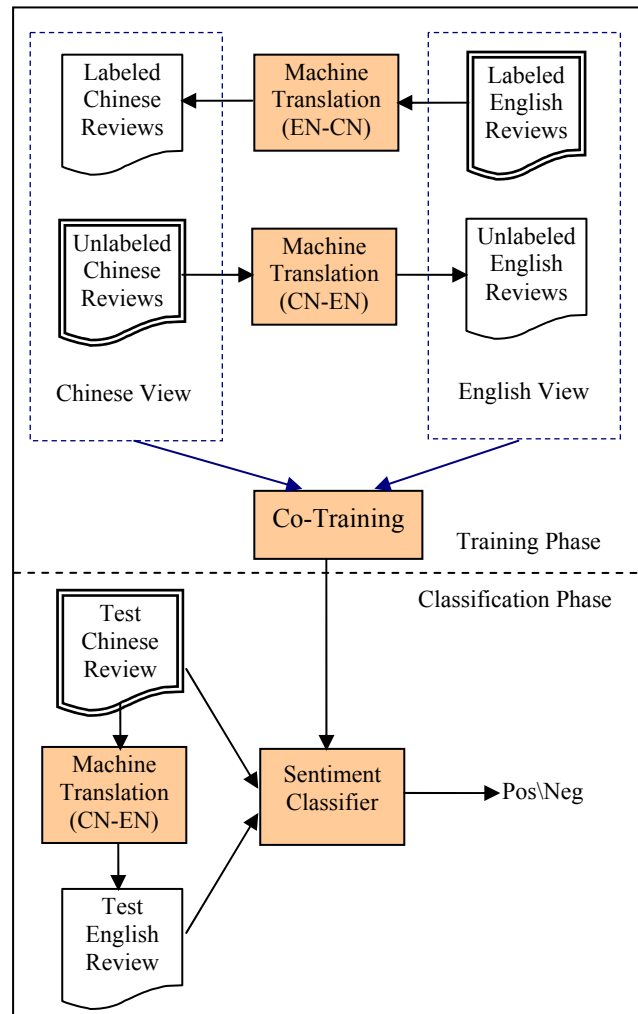


Figure 1. Framework of the proposed approach

3.2 Review Translation

In order to overcome the language gap, we must translate one language into another language. Fortunately, machine translation techniques have been well developed in the NLP field, though the translation performance is far from satisfactory. A few commercial machine translation services can be publicly accessed, e.g. *Google Translate*², *Yahoo Babel Fish*³ and *Windows Live Translate*⁴.

¹ The unlabeled Chinese reviews used for co-training do not include the unlabeled Chinese reviews for testing, i.e., the Chinese reviews for testing are blind to the training phase.

² http://translate.google.com/translate_t

³ http://babelfish.yahoo.com/translate_txt

⁴ <http://www.windowslivetranslator.com/>

In this study, we adopt *Google Translate* for both English-to-Chinese Translation and Chinese-to-English Translation, because it is one of the state-of-the-art commercial machine translation systems used today. *Google Translate* applies statistical learning techniques to build a translation model based on both monolingual text in the target language and aligned text consisting of examples of human translations between the languages.

3.3 The Co-Training Algorithm

The co-training algorithm (Blum and Mitchell, 1998) is a typical bootstrapping method, which starts with a set of labeled data, and increase the amount of annotated data using some amounts of unlabeled data in an incremental way. One important aspect of co-training is that two conditional independent views are required for co-training to work, but the independence assumption can be relaxed. Till now, co-training has been successfully applied to statistical parsing (Sarkar, 2001), reference resolution (Ng and Cardie, 2003), part of speech tagging (Clark et al., 2003), word sense disambiguation (Mihalcea, 2004) and email classification (Kiritchenko and Matwin, 2001).

In the context of cross-lingual sentiment classification, each labeled English review or unlabeled Chinese review has two views of features: English features and Chinese features. Here, a review is used to indicate both its Chinese version and its English version, until stated otherwise. The co-training algorithm is illustrated in Figure 2. In the algorithm, the class distribution in the labeled data is maintained by balancing the parameter values of p and n at each iteration.

The intuition of the co-training algorithm is that if one classifier can confidently predict the class of an example, which is very similar to some of labeled ones, it can provide one more training example for the other classifier. But, of course, if this example happens to be easy to be classified by the first classifier, it does not mean that this example will be easy to be classified by the second classifier, so the second classifier will get useful information to improve itself and vice versa (Kiritchenko and Matwin, 2001).

In the co-training algorithm, a basic classification algorithm is required to construct C_{en} and C_{cn} . Typical text classifiers include Support Vector Machine (SVM), Naïve Bayes (NB), Maximum Entropy (ME), K-Nearest Neighbor (KNN), etc. In this study, we adopt the widely-used SVM classifier (Joachims, 2002). Viewing input data

as two sets of vectors in a feature space, SVM constructs a separating hyperplane in the space by maximizing the margin between the two data sets. The English or Chinese features used in this study include both unigrams and bigrams⁵ and the feature weight is simply set to term frequency⁶. Feature selection methods (e.g. Document Frequency (DF), Information Gain (IG), and Mutual Information (MI)) can be used for dimension reduction. But we use all the features in the experiments for comparative analysis, because there is no significant performance improvement after applying the feature selection techniques in our empirical study. The output value of the SVM classifier for a review indicates the confidence level of the review's classification. Usually, the sentiment polarity of a review is indicated by the sign of the prediction value.

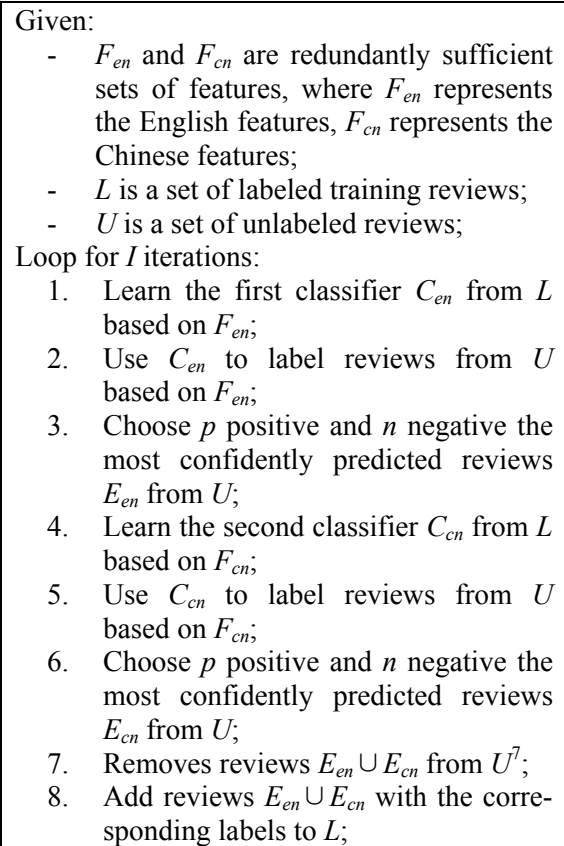


Figure 2. The co-training algorithm

In the training phase, the co-training algorithm learns two separate classifiers: C_{en} and C_{cn} .

⁵ For Chinese text, a unigram refers to a Chinese word and a bigram refers to two adjacent Chinese words.

⁶ Term frequency performs better than TFIDF by our empirical analysis.

⁷ Note that the examples with conflicting labels are not included in $E_{en} \cup E_{cn}$. In other words, if an example is in both E_{en} and E_{cn} , but the labels for the example is conflicting, the example will be excluded from $E_{en} \cup E_{cn}$.

Therefore, in the classification phase, we can obtain two prediction values for a test review. We normalize the prediction values into $[-1, 1]$ by dividing the maximum absolute value. Finally, the average of the normalized values is used as the overall prediction value of the review.

4 Empirical Evaluation

4.1 Evaluation Setup

4.1.1 Data set

The following three datasets were collected and used in the experiments:

Test Set (Labeled Chinese Reviews): In order to assess the performance of the proposed approach, we collected and labeled 886 product reviews (451 positive reviews + 435 negative reviews) from a popular Chinese IT product web site-IT168⁸. The reviews focused on such products as mp3 players, mobile phones, digital camera and laptop computers.

Training Set (Labeled English Reviews): There are many labeled English corpora available on the Web and we used the corpus constructed for multi-domain sentiment classification (Blitzer et al., 2007)⁹, because the corpus was large-scale and it was within similar domains as the test set. The dataset consisted of 8000 Amazon product reviews (4000 positive reviews + 4000 negative reviews) for four different product types: books, DVDs, electronics and kitchen appliances.

Unlabeled Set (Unlabeled Chinese Reviews): We downloaded additional 1000 Chinese product reviews from IT168 and used the reviews as the unlabeled set. Therefore, the unlabeled set and the test set were in the same domain and had similar underlying feature distributions.

Each Chinese review was translated into English review, and each English review was translated into Chinese review. Therefore, each review has two independent views: English view and Chinese view. A review is represented by both its English view and its Chinese view.

Note that the training set and the unlabeled set are used in the training phase, while the test set is blind to the training phase.

4.1.2 Evaluation Metric

We used the standard precision, recall and F-measure to measure the performance of positive and negative class, respectively, and used the

accuracy metric to measure the overall performance of the system. The metrics are defined the same as in general text categorization.

4.1.3 Baseline Methods

In the experiments, the proposed co-training approach (CoTrain) is compared with the following baseline methods:

SVM(CN): This method applies the inductive SVM with only Chinese features for sentiment classification in the Chinese view. Only English-to-Chinese translation is needed. And the unlabeled set is not used.

SVM(EN): This method applies the inductive SVM with only English features for sentiment classification in the English view. Only Chinese-to-English translation is needed. And the unlabeled set is not used.

SVM(ENCN1): This method applies the inductive SVM with both English and Chinese features for sentiment classification in the two views. Both English-to-Chinese and Chinese-to-English translations are required. And the unlabeled set is not used.

SVM(ENCN2): This method combines the results of SVM(EN) and SVM(CN) by averaging the prediction values in the same way with the co-training approach.

TSVM(CN): This method applies the transductive SVM with only Chinese features for sentiment classification in the Chinese view. Only English-to-Chinese translation is needed. And the unlabeled set is used.

TSVM(EN): This method applies the transductive SVM with only English features for sentiment classification in the English view. Only Chinese-to-English translation is needed. And the unlabeled set is used.

TSVM(ENCN1): This method applies the transductive SVM with both English and Chinese features for sentiment classification in the two views. Both English-to-Chinese and Chinese-to-English translations are required. And the unlabeled set is used.

TSVM(ENCN2): This method combines the results of TSVM(EN) and TSVM(CN) by averaging the prediction values.

Note that the first four methods are straightforward methods used in previous work, while the latter four methods are strong baselines because the transductive SVM has been widely used for improving the classification accuracy by leveraging additional unlabeled examples.

⁸ <http://www.it168.com>

⁹ <http://www.cis.upenn.edu/~mdredze/datasets/sentiment/>

4.2 Evaluation Results

4.2.1 Method Comparison

In the experiments, we first compare the proposed co-training approach ($I=40$ and $p=n=5$) with the eight baseline methods. The three parameters in the co-training approach are empirically set by considering the total number (i.e. 1000) of the unlabeled Chinese reviews. In our empirical study, the proposed approach can perform well with a wide range of parameter values, which will be shown later. Table 1 shows the comparison results.

Seen from the table, the proposed co-training approach outperforms all eight baseline methods over all metrics. Among the eight baselines, the best one is TSVM(ENCN2), which combines the results of two transductive SVM classifiers. Actually, TSVM(ENCN2) is similar to CoTrain because CoTrain also combines the results of two classifiers in the same way. However, the co-training approach can train two more effective classifiers, and the accuracy values of the component English and Chinese classifiers are 0.775 and 0.790, respectively, which are higher than the corresponding TSVM classifiers. Overall, the use of transductive learning and the combination of English and Chinese views are beneficial to the final classification accuracy, and the co-training approach is more suitable for making use of the unlabeled Chinese reviews than the transductive SVM.

4.2.2 Influences of Iteration Number (I)

Figure 3 shows the accuracy curve of the co-training approach (Combined Classifier) with different numbers of iterations. The iteration number I is varied from 1 to 80. When I is set to 1, the co-training approach is degenerated into SVM(ENCN2). The accuracy curves of the component English and Chinese classifiers learned in the co-training approach are also shown in the

figure. We can see that the proposed co-training approach can outperform the best baseline-TSVM(ENCN2) after 20 iterations. After a large number of iterations, the performance of the co-training approach decreases because noisy training examples may be selected from the remaining unlabeled set. Finally, the performance of the approach does not change any more, because the algorithm runs out of all possible examples in the unlabeled set. Fortunately, the proposed approach performs well with a wide range of iteration numbers. We can also see that the two component classifier has similar trends with the co-training approach. It is encouraging that the component Chinese classifier alone can perform better than the best baseline when the iteration number is set between 40 and 70.

4.2.3 Influences of Growth Size (p, n)

Figure 4 shows how the growth size at each iteration (p positive and n negative confident examples) influences the accuracy of the proposed co-training approach. In the above experiments, we set $p=n$, which is considered as a balanced growth. When p differs very much from n , the growth is considered as an imbalanced growth. Balanced growth of (2, 2), (5, 5), (10, 10) and (15, 15) examples and imbalanced growth of (1, 5), (5, 1) examples are compared in the figure. We can see that the performance of the co-training approach with the balanced growth can be improved after a few iterations. And the performance of the co-training approach with large p and n will more quickly become unchanged, because the approach runs out of the limited examples in the unlabeled set more quickly. However, the performance of the co-training approaches with the two imbalanced growths is always going down quite rapidly, because the labeled unbalanced examples hurt the performance badly at each iteration.

| Method | Positive | | | Negative | | | Total |
|---|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | Precision | Recall | F-measure | Precision | Recall | F-measure | Accuracy |
| SVM(CN) | 0.733 | 0.865 | 0.793 | 0.828 | 0.674 | 0.743 | 0.771 |
| SVM(EN) | 0.717 | 0.803 | 0.757 | 0.766 | 0.671 | 0.716 | 0.738 |
| SVM(ENCN1) | 0.744 | 0.820 | 0.781 | 0.792 | 0.708 | 0.748 | 0.765 |
| SVM(ENCN2) | 0.746 | 0.847 | 0.793 | 0.816 | 0.701 | 0.754 | 0.775 |
| TSVM(CN) | 0.724 | 0.878 | 0.794 | 0.838 | 0.653 | 0.734 | 0.767 |
| TSVM(EN) | 0.732 | 0.860 | 0.791 | 0.823 | 0.674 | 0.741 | 0.769 |
| TSVM(ENCN1) | 0.743 | 0.878 | 0.805 | 0.844 | 0.685 | 0.756 | 0.783 |
| TSVM(ENCN2) | 0.744 | 0.896 | 0.813 | 0.863 | 0.680 | 0.761 | 0.790 |
| CoTrain ($I=40; p=n=5$) | 0.768 | 0.905 | 0.831 | 0.879 | 0.717 | 0.790 | 0.813 |

Table 1. Comparison results

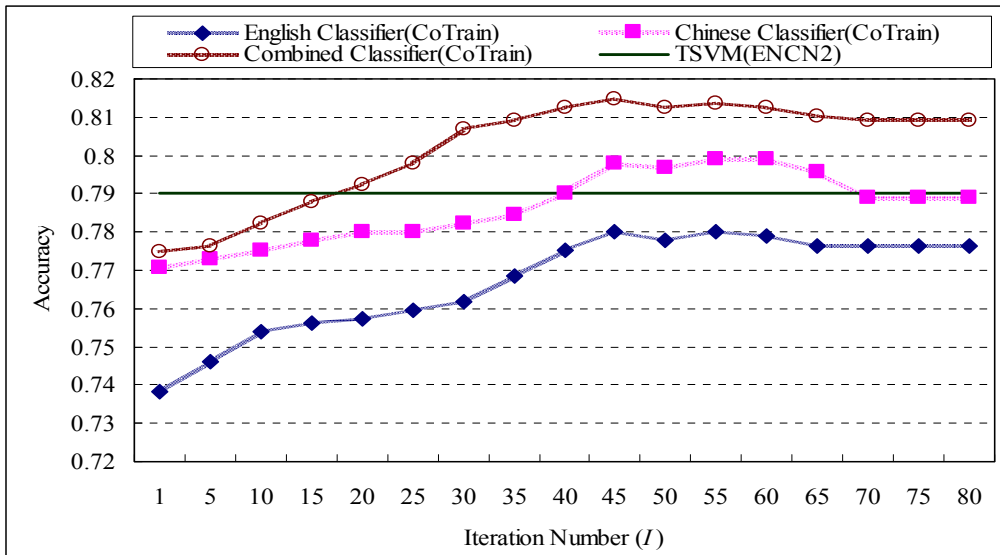


Figure 3. Accuracy vs. number of iterations for co-training ($p=n=5$)

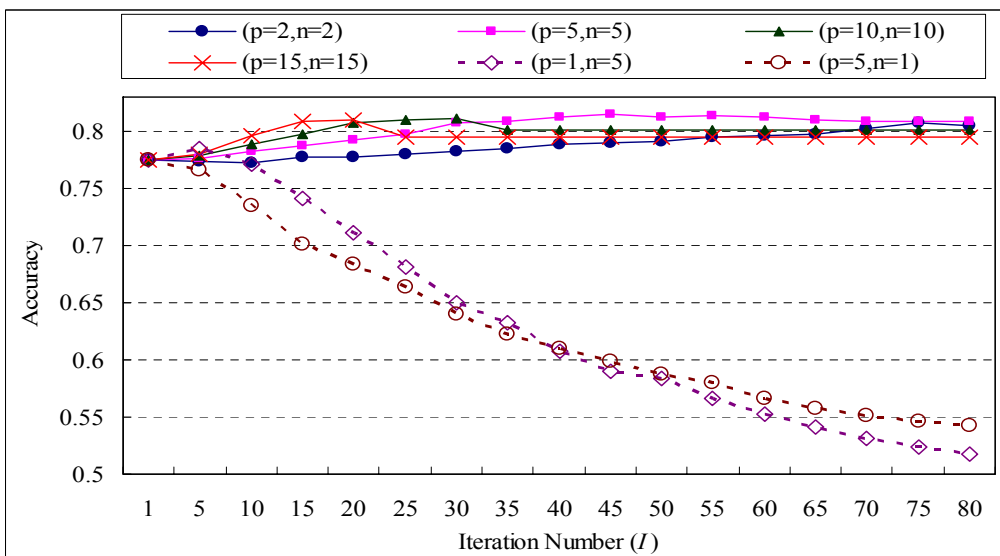


Figure 4. Accuracy vs. different (p, n) for co-training

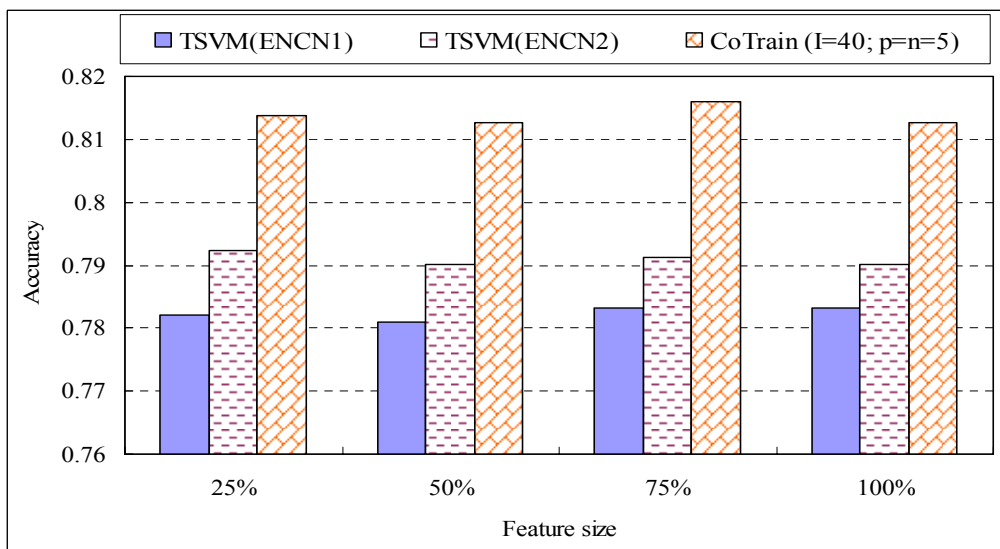


Figure 5. Influences of feature size

4.2.4 Influences of Feature Selection

In the above experiments, all features (unigram + bigram) are used. As mentioned earlier, feature selection techniques are widely used for dimension reduction. In this section, we further conduct experiments to investigate the influences of feature selection techniques on the classification results. We use the simple but effective document frequency (DF) for feature selection. Figures 6 show the comparison results of different feature sizes for the co-training approach and two strong baselines. The feature size is measured as the proportion of the selected features against the total features (i.e. 100%).

We can see from the figure that the feature selection technique has very slight influences on the classification accuracy of the methods. It can be seen that the co-training approach can always outperform the two baselines with different feature sizes. The results further demonstrate the effectiveness and robustness of the proposed co-training approach.

5 Conclusion and Future Work

In this paper, we propose to use the co-training approach to address the problem of cross-lingual sentiment classification. The experimental results show the effectiveness of the proposed approach.

In future work, we will improve the sentiment classification accuracy in the following two ways:

1) The smoothed co-training approach used in (Mihalcea, 2004) will be adopted for sentiment classification. The approach has the effect of “smoothing” the learning curves. During the bootstrapping process of smoothed co-training, the classifier at each iteration is replaced with a majority voting scheme applied to all classifiers constructed at previous iterations.

2) The feature distributions of the translated text and the natural text in the same language are still different due to the inaccuracy of the machine translation service. We will employ the structural correspondence learning (SCL) domain adaptation algorithm used in (Blitzer et al., 2007) for linking the translated text and the natural text.

Acknowledgments

This work was supported by NSFC (60873155), RFDP (20070001059), Beijing Nova Program (2008B03), National High-tech R&D Program (2008AA01Z421) and NCET (NCET-08-0006). We also thank the anonymous reviewers for their useful comments.

References

- A. Andreevskaia and S. Bergler. 2008. When specialists and generalists work together: overcoming domain dependence in sentiment tagging. In *Proceedings of ACL-08: HLT*.
- C. Banea, R. Mihalcea, J. Wiebe and S. Hassan. 2008. Multilingual subjectivity analysis using machine translation. In *Proceedings of EMNLP-2008*.
- N. Bel, C. H. A. Koster, and M. Villegas. 2003. Cross-lingual text categorization. In *Proceedings of ECDL-03*.
- J. Blitzer, M. Dredze and F. Pereira. 2007. Biographies, bollywood, boom-boxes and blenders: domain adaptation for sentiment classification. In *Proceedings of ACL-07*.
- A. Blum and T. Mitchell. 1998. Combining labeled and unlabeled data with cotraining. In *Proceedings of COLT-98*.
- S. Brody, R. Navigli and M. Lapata. 2006. Ensemble methods for unsupervised WSD. In *Proceedings of COLING-ACL-2006*.
- S. Clark, J. R. Curran, and M. Osborne. 2003. Bootstrapping POS taggers using unlabelled data. In *Proceedings of CoNLL-2003*.
- W. Dai, G.-R. Xue, Q. Yang, Y. Yu. 2007a. Transferring Naïve Bayes Classifiers for text classification. In *Proceedings of AAAI-07*.
- W. Dai, G.-R. Xue, Q. Yang, Y. Yu. 2007b. Co-clustering based classification for out-of-domain documents. In *Proceedings of KDD-07*.
- H. DauméIII and D. Marcu. 2006. Domain adaptation for statistical classifiers. *Journal of Artificial Intelligence Research*, 26:101–126.
- A. Devitt and K. Ahmad. 2007. Sentiment polarity identification in financial news: a cohesion-based approach. In *Proceedings of ACL2007*.
- T. G. Dietterich. 1997. Machine learning research: four current directions. *AI Magazine*, 18(4), 1997.
- A. Gliozzo and C. Strapparava. 2005. Cross language text categorization by acquiring multilingual domain models from comparable corpora. In *Proceedings of the ACL Workshop on Building and Using Parallel Texts*.
- K. Hiroshi, N. Tetsuya and W. Hideo. 2004. Deeper sentiment analysis using machine translation technology. In *Proceedings of COLING-04*.
- J. Jiang and C. Zhai. 2007. A two-stage approach to domain adaptation for statistical classifiers. In *Proceedings of CIKM-07*.
- T. Joachims. 1999. Transductive inference for text classification using support vector machines. In *Proceedings of ICML-99*.

- T. Joachims. 2002. Learning to classify text using support vector machines. Dissertation, Kluwer, 2002.
- A. Kennedy and D. Inkpen. 2006. Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*, 22(2):110-125.
- S.-M. Kim and E. Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of COLING-04*.
- S. Kiritchenko and S. Matwin. 2001. Email classification with co-training. In *Proceedings of the 2001 Conference of the Centre for Advanced Studies on Collaborative Research*.
- L.-W. Ku, Y.-T. Liang and H.-H. Chen. 2006. Opinion extraction, summarization and tracking in news and blog corpora. In *Proceedings of AAAI-2006*.
- J. Li and M. Sun. 2007. Experimental study on sentiment classification of Chinese review using machine learning techniques. In *Proceeding of IEEE-NLPKE-07*.
- X. Ling, W. Dai, Y. Jiang, G.-R. Xue, Q. Yang, and Y. Yu. 2008. Can Chinese Web pages be classified with English data source? In *Proceedings of WWW-08*.
- B. Liu, M. Hu and J. Cheng. 2005. Opinion observer: Analyzing and comparing opinions on the web. In *Proceedings of WWW-2005*.
- R. McDonald, K. Hannan, T. Neylon, M. Wells and J. Reynar. 2007. Structured models for fine-to-coarse sentiment analysis. In *Proceedings of ACL-07*.
- R. Mihalcea. 2004. Co-training and self-training for word sense disambiguation. In *Proceedings of CONLL-04*.
- R. Mihalcea, C. Banea and J. Wiebe. 2007. Learning multilingual subjective language via cross-lingual projections. In *Proceedings of ACL-2007*.
- T. Mullen and N. Collier. 2004. Sentiment analysis using support vector machines with diverse information sources. In *Proceedings of EMNLP-04*.
- V. Ng and C. Cardie. 2003. Weakly supervised natural language learning without redundant views. In *Proceedings of HLT-NAACL-03*.
- K. Nigam, A. K. McCallum, S. Thrun, and T. Mitchell. 2000. Text Classification from Labeled and Unlabeled Documents using EM. *Machine Learning*, 39(2-3):103-134.
- B. Pang, L. Lee and S. Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of EMNLP-02*.
- B. Pang and L. Lee. 2004. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of ACL-04*.
- J. Read. 2005. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of ACL-05*.
- L. Rigutini, M. Maggini and B. Liu. 2005. An EM based training algorithm for cross-language text categorization. In *Proceedings of WI-05*.
- A. Sarkar. 2001. Applying cotraining methods to statistical parsing. In *Proceedings of NAACL-2001*.
- I. Titov and R. McDonald. 2008. A joint model of text and aspect ratings for sentiment summarization. In *Proceedings of ACL-08:HLT*.
- B. K. Y. Tsou, R. W. M. Yuen, O. Y. Kwong, T. B. Y. La and W. L. Wong. 2005. Polarity classification of celebrity coverage in the Chinese press. In *Proceedings of International Conference on Intelligence Analysis*.
- P. Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of ACL-2002*.
- X. Wan. 2008. Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis. In *Proceedings of EMNLP-2008*.
- T. Wilson, J. Wiebe and P. Hoffmann. 2005. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *Proceedings of HLT/EMNLP-05*.
- G.-R. Xue, W. Dai, Q. Yang, Y. Yu. 2008. Topic-bridged PLSA for cross-domain text classification. In *Proceedings of SIGIR-08*.
- Q. Ye, W. Shi and Y. Li. 2006. Sentiment classification for movie reviews in Chinese by improved semantic oriented approach. In *Proceedings of 39th Hawaii International Conference on System Sciences*, 2006.