Active Learning for Cross-Domain Sentiment Classification

Shoushan Li, Yunxia Xue, Zhongqing Wang, Guodong Zhou*
Natural Language Processing Lab, School of Computer Science and Technology
Soochow University, Suzhou, 215006, China
{shoushan.li, yunxia.xue, wangzq870305}@gmail.com, gdzhou@suda.edu.cn

Abstract

In the literature, various approaches have been proposed to address the domain adaptation problem in sentiment classification (also called cross-domain sentiment classification). However, the adaptation performance normally much suffers when the data distributions in the source and target domains differ significantly. In this paper, we suggest to perform active learning for cross-domain sentiment classification by actively selecting a small amount of labeled data in the target domain. Accordingly, we propose an novel active learning approach for cross-domain sentiment classification. First, we train two individual classifiers, i.e., the source and target classifiers with the labeled data from the source and target respectively. Then, the two classifiers are employed to select informative samples with the selection strategy of Query By Committee (QBC). Third, the two classifier is combined to make the classification decision. Importantly, the two classifiers are trained by fully exploiting the unlabeled data in the target domain with the label propagation (LP) algorithm. Empirical studies demonstrate the effectiveness of our active learning approach for cross-domain sentiment classification over some strong baselines.

1 Introduction

Sentiment classification is a task of determining the sentimental orientation (e.g., positive or negative) of a given textual document towards a given topic [Pang et al., 2002; Turney, 2002]. This study has been extensively explored in multiple research communities, such as natural language processing (NLP), data mining and machine learning [Pang and Lee, 2008]. One of the main challenges for sentiment classification is the domain adaptation problem. That is, a sentiment classifier trained with the labeled data from one domain normally performs unsatisfactorily in another domain. To overcome this problem, several studies have been proposed to address the domain adaptation problem in sentiment classification by using some labeled data from the source domain and a large amount of unlabeled data from the target domain [Blitzer et al., 2007; He et al., 2011; Bollegala et al., 2011].

Although existing studies have yielded certain progress in domain adaptation on sentiment classification, the adaptation performance normally much suffers when the data distributions in the source and target domains differ significantly. In some cases, the incorporation of unlabeled data in the target domain might even adversely affect the performance in the target domain, a situation often referred to as *negative transfer* [Pan and Yang, 2010]. For example, [Blitzer et al., 2007] report that when transferring from domain '*Kitchen*' to domain '*Book*', the obtained performance by the proposed approach (68.6% in accuracy) is even lower than the performance of without using the unlabeled data in the target domain (70.9% in accuracy). The reason of the failure in such situations is that the distributions of the source and target domains become too different to make the adaptation algorithm useful. One possible solution to such dilemma is to annotate a small amount of *good* labeled data in the target domain to quickly reduce the huge difference between the two domains, which is typically an active learning paradigm. However, active learning in cross-domain sentiment classification faces some unique challenges than active learning in traditional in-domain sentiment classification.

First, one major difference is the size of the involved labeled data. That is, active learning in in-domain classification normally contains only a small amount of labeled data while active learning in cross-domain classification normally contains abundant labeled data in the source domain. The large amount of labeled data brings out novel difficulties in both the sample selection strategy and the classification algorithm for active learning in the cross-domain case.

1) Since the initial labeled data in the source domain is in a big scale, the newly-added labeled data from the target domain may become too weak to affect the selection tendency in the merged labeled data. Apparently, this is

* Corresponding author

not what an effective active learning approach is expected because the newly-added labeled data is from the same domain as the test data, i.e., the target domain, and thus should be more valuable for supervising the selection strategy.

2) Given the abundant labeled data from the source domain, the newly-added labeled data from the target domain are normally too few to quickly affect the classification decision. This makes the newly-added labeled data little helpful, especially at the beginning.

Second, another major difference is the usage of the unlabeled data. That is, in-domain classification normally employs the unlabeled data for only selecting the informative samples while cross-domain classification is encouraged to exploit the unlabeled data (from the target domain) in the classification algorithms due to the limited size of the labeled data in the target domain. Therefore, in the case of cross-domain classification, it is also challenging to design a powerful classification algorithm which could fully take advantage of the unlabeled data together with the newly-added labeled data in the target domain.

In this paper, we address above challenges in active learning for cross-domain sentiment classification.

For the first challenge, we use the newly-added labeled data from the target domain to train a separate classifier and apply it in both the sample selection strategy and the classification algorithm. Specifically, we construct two individual classifiers by training the data from the source and target domain separately (referred to as the source and target classifiers respectively). For the sample selection strategy, these two classifiers naturally drive us to adopt the active learning strategy named Query By Committee (QBC) to select the samples that disagree most in terms of the prediction labels. For the classification algorithm, we combine the source and target classifiers to make the classification decision. Besides, we notice that the diversity between the source and target classifiers is an important factor to improve the combination performance [Wang et al., 2001]. Thus, we modify the standard QBC strategy by selecting not only the disagreed samples by the source and target classifiers but also the most uncertain samples by the source classifier.

For the second challenge, we propose a label propagation (LP) -based classification algorithm, which leverages both the labeled and unlabeled data, and apply it to both the source and target classifiers. One big advantage of this approach is its effectiveness for both the source and target classifiers to exploit the unlabeled data in the target domain.

The remainder of this paper is organized as follows. Section 2 overviews the related work. Section 3 proposes our active learning approach for cross-domain sentiment classification. Section 4 evaluates our approach. Finally, Section 5 gives the conclusion and future work.

2 Related Work

This section gives an overview of the related domain adaptation work from both sentiment classification and active learning perspectives.

2.1 Domain Adaptation in Sentiment Classification

Early studies on sentiment classification mainly focus on the single-domain setting [Pang et al., 2002; Turney, 2002]. For detailed discussion on this setting, please refer to [Pang and Lee, 2008].

As for cross-domain sentiment classification, [Aue and Gammon, 2005] pioneer the studies. Although they fail to propose an effective solution, they highlight the importance and difficulty of cross-domain sentiment classification.

Subsequently, [Blitzer et al., 2007] successfully develop a domain adaptation approach, named SCL, for sentiment classification, with the main idea to bridge the knowledge between the source and target domains using some pivotal features.

More recently, [He et al., 2011] employ a topic model, called joint sentiment-topic model (JST), and [Bollegala et al., 2011] create a sentiment sensitive thesaurus, to perform cross-domain sentiment classification. Results from these studies demonstrate comparable performance to SCL.

Unlike above studies, our study pioneers active learning on cross-domain sentiment classification, which greatly improves the adaptation performance with the help of a small amount of labeled data in the target domain.

2.2 Active Learning-based Domain Adaptation

Although domain adaptation has been extensively studied in NLP and machine learning communities, such as [Shen et al. 2004], [Jiang and Zhai, 2007], [Zhu et al., 2008], and [Daumé III, 2007], there are only a few studies on active learning-based domain adaptation.

[Chan and Ng, 2007] apply active learning for word sense disambiguation in a domain adaptation setting with focus on considering an uncertainty measurement according to the available labeled data. Similarly, [Shi et al., 2008] propose a framework for active learning on the target domain (named active transferring of domain knowledge therein) with the sample selection strategy on an uncertainty measurement. More recently, [Rai et al., 2010] propose an online active learning approach on domain adaptation, still with the sample selection strategy on an uncertainty measurement.

However, all existing studies do not pay close attention to the newly-added labeled data in the target domain. In comparison, our proposed approach makes full use of the newly-added labeled data in both the sample selection strategy and the classification algorithm.

3 Active Learning for Cross-domain Sentiment Classification

In cross-domain sentiment classification, the test samples come from the target domain while the training data come from a different source domain. In the sequel, we refer the training data in the source domain to as $L_s = \{(x_i, y_i)\}_{i=1}^{n_1}$ where $x_i \in \mathbf{R}^d$ is the d dimensional input vector, and y_i is its output label. We also assume that the test samples are available and denoted as $T_T = \{(x'_i, y'_i)\}_{i=1}^{n_2}$ where $x'_i \in \mathbf{R}^d$ is the input. Let $t(x)$ and $s(x)$ be the marginal distributions of the samples from the target and source domains, respectively. In general, $t(x)$ and $s(x)$ are different. The task of cross-domain sentiment classification is to leverage the training data L_s to predict label y'_i for input x'_i in the target domain. In general, in order to reduce the gap between the source and target domains, another resource, i.e., the unlabeled data in the target domain denoted as $U_T = \{(x'_i, y'_i)\}_{i=1}^{n_3}$, is often available for the adaptation classification algorithm.

In this study, we focus on active learning in cross-domain sentiment classification. Naturally, this task contains two main steps, first selecting a small amount of "informative" samples, denoted as $L_T = \{(x'_i, y'_i)\}_{i=1}^{n_4}$, from U_T for manual annotation, and then training a classifier with the labeled data including L_s and L_T together with the unlabeled data U_T.

For clarity, some important symbols are listed in Table 1.

Symbol	Definition
L_S	Labeled source-domain data
L_T	Labeled target-domain data
f_S	The source Classifier
f_T	The target Classifier
U_T	Unlabeled target-domain data
ΔL_T	Newly-added data at each iteration
$f_{LP\text{-}S}$	The LP-based source Classifier
$f_{LP\text{-}T}$	The LP-based target Classifier
T_T	Test data in the target domain

Table 1: Symbol definition

Figure 1 and Figure 2 illustrate the main flows of sample selection and sample classification, where the former employs a QBC-based selection strategy to choose "informative" samples and the latter employs a LP algorithm to make the classification decision for a sample, both with the help of the source and target classifiers. Note that the source and target classifiers in the testing process also exploit the unlabeled data in the target domain with LP algorithm. In the following subsections, we introduce the processes of both the QBC-based sample selection and the LP-based classifiers in details.

3.1 QBC-based Selection Strategy

Query by Committee (QBC) is a group of active learning approaches that employ many copies of "hypotheses" (e.g., coming from randomized learning algorithms) to select an unlabeled example at which their classification predictions are maximally spread [Freund et al., 92]. In our setting, we first use the source classifier f_S and the target classifier f_T to collaboratively select label-disagreed samples as the selection candidates. Then, we rank the label-disagreed samples according to their uncertainty values by the source classifier. Finally, we select the top-N uncertainty samples as the newly-added data for human annotation.

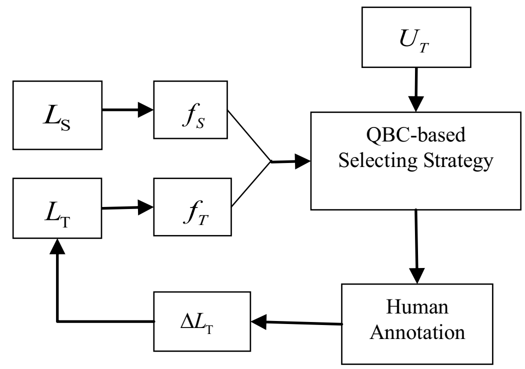

Figure 1: Sample selection in our approach

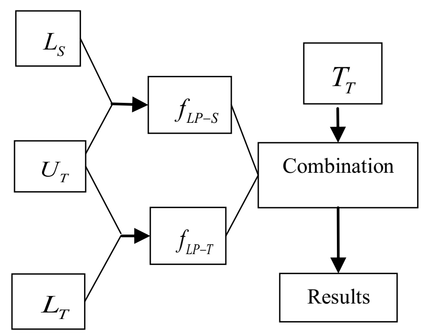

Figure 2: Sample classification in our approach

Input: Labeled source-domain data L_S
Unlabeled target-domain data U_T
Output: Automatically labeled target-domain data L_T
Procedure:
(a) Initialize $L_T = \varnothing$
(b) Train the source classifier f_S with L_S
(c) Use f_S to select top-N uncertainty samples as ΔL_T
(d) $L_T = L_T + \Delta L_T$, and $U_T = U_T - \Delta L_T$
(e) Repeat k times
e1) Train the target classifier f_T with L_T
e2) Use both f_S and f_T to select label-disagreed samples from U_T
e3) Use f_S to select top-N uncertainty samples from the label-disagreed samples as ΔL_T
e4) $L_T = L_T + \Delta L_T$, and $U_T = U_T - \Delta L_T$

Figure 3: QBC-based selection strategy in our approach

4.3 LP-based Classification Algorithm

To well incorporate the knowledge in the unlabeled data for both the source and target classifiers, we adopt a graph-based ranking approach named LP to propagate the labels from the labeled data to the unlabeled data.

The input of the LP algorithm is a graph describing the relationship among each sample pair in the labeled and unlabeled data. Here, the document-word bipartite graph is adopted due to its excellent performance in sentiment classification [Sindhwani and Melville, 2008]. In a bipartite graph, the nodes consists of two parts: documents and all words extracted from the documents. An undirected edge (d_i, w_k) exists if and only if document d_i contains word w_k. Let x_{ik} be the frequency of word w_k in document d_i. From the bipartite graph, the probability of walking from document d_i to word w_k can be calculated as $x_{ik} / \sum_k x_{ik}$ and the probability of walking from word w_k to document d_j can be calculated as $x_{jk} / \sum_j x_{jk}$. Thus the probability of walking from document d_i to document d_j though the word w_k can be calculated as $(x_{ik} / \sum_k x_{ik}) \cdot (x_{jk} / \sum_j x_{jk})$. When all words are considered, we get the transition probability from d_i to d_j as:

$$t_{ij} = \sum_k \frac{x_{ik}}{\sum_k x_{ik}} \cdot \frac{x_{jk}}{\sum_j x_{jk}}$$

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

Input:

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

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

Output:

The unlabeled data with prediction labels

Procedure:

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

Figure 4: The LP algorithm

Figure 4 illustrates the LP algorithm [Zhu and Ghahramani, 2002], during which the probabilities of the labeled data are clamped in each loop using their initial ones and act as a force to propagate their labels to the unlabeled data.

Given the above LP algorithm, it is straightforward to train a LP-based source classifier $f_{\text{LP-S}}$ and a LP-based target classifier $f_{\text{LP-T}}$.

1) Construct the document-word bipartite graph with both the labeled data and the unlabeled data (in the target domain) and get the transition probability matrix;
2) Run the LP algorithm as shown in Figure 4 to obtain the labels of unlabeled data;
3) Consider the unlabeled data with the predicted labels as pseudo-labeled data;
4) Emerge the labeled data and the pseudo-labeled data to train a classifier.

The difference between $f_{\text{LP-S}}$ and $f_{\text{LP-T}}$ lies in the domains of the labeled data. In fact, the source classifier $f_{\text{LP-S}}$ is a domain adaptation classifier while the target classifier $f_{\text{LP-T}}$ is an in-domain semi-supervised classifier.

Once the LP-based source and target classifiers are obtained, they are combined to make the prediction decision, as shown in Figure 2, using the standard Bayes rule, i.e.

$$assign \quad y \to c_j$$
$$where \;\; j = \arg\max_r (p_{Source}(c_r \mid x_i) \cdot p_{Target}(c_r \mid x'_i))$$

Where $p_{Source}(c_r \mid x_i)$ and $p_{Target}(c_r \mid x'_i)$ denote the posterior possibilities of the prediction sample belonging to the category c_r, estimated by the source and target classifiers, respectively

5 Experimentation

We systematically evaluate our active learning approach for cross-domain sentiment classification on a multi-domain dataset[2] [Blitzer et al., 2007].

5.1 Experimental Setting

Dataset: The dataset contains product reviews from four different domains: Book (B), DVD (D), Electronics (E) and Kitchen (K) appliances, each of which contains 1000 positive and 1000 negative labeled reviews. we randomly select 1600 instances from the source domain as labeled data, 1600 instances from the target domain as unlabeled data, and the remaining 400 instances from the target domain are reserved as test data. For different domains of B, D, E and K, the corresponding in-domain classifiers achieve the performance of 81.5%, 80.5%, 83.8% and 86.5% in accuracy, respectively.

Features: Each review text is treated as a bag-of-words and transformed into binary vectors encoding the presence or absence of word unigrams and bigrams.

Classification algorithm: the maximum entropy (ME) classifier implemented with the public tool, Mallet Toolkits[3].

[2] http://www.seas.upenn.edu/~mdredze/datasets/sentiment/

[3] http://mallet.cs.umass.edu/

5.2 Performance of LP-based source and target classification

Before investigating the performance of our active learning approach, we first check the effectiveness of the LP-based source classifier $f_{\text{LP-S}}$ and target classifier $f_{\text{LP-T}}$.

Since the source classifier $f_{\text{LP-S}}$ is a domain adaptation classifier, we implement the state-of-the-art SCL for domain adaptation on sentiment classification [Blitzer et al., 2007] for a comparative study. In this setting, no labeled data is available in the target domain. Table 2 shows that both the SCL and our LP-based classifier outperform the baseline with the average improvements of 1.9% and 2.9% in accuracy, respectively. However, it also shows that the adaptation performance using either SCL or LP remains unsatisfactory, with much lower performance than in-domain classification, when the source and target domains are much different, including E→B, K→B, E→D, K→D, B→E, D→E, B→K, and D→K. In the following experiments on active learning, we will focus on these eight pairs of domain adaptation.

	Baseline	SCL (Blizer *et al*., 2006)	$f_{\text{LP-S}}$
B->D	0.793	0.795	**0.798**
B->E	0.708	**0.730**	0.720
B->K	0.715	0.753	**0.770**
D->B	0.775	**0.790**	0.780
D->E	0.660	0.705	**0.768**
D->K	0.733	**0.778**	0.743
E->B	0.690	0.640	**0.715**
E->D	0.710	0.700	**0.740**
E->K	0.830	0.833	**0.835**
K->B	0.693	**0.760**	0.735
K->D	0.695	**0.745**	0.733
K->E	0.805	0.808	**0.815**

Table 2: Performance comparison between SCL and our LP-based domain adaptation

	Baseline	Personal/Impersonal (Li et al., 2010)	$f_{\text{LP-T}}$
Book	0.660	**0.720**	0.705
DVD	0.655	0.630	**0.690**
Electronic	0.740	0.760	**0.775**
Kitchen	0.735	**0.775**	0.770

Table 3: Performance comparison between the Personal/Impersonal approach and our LP-based semi-supervised classification in the target domain

Since the target classifier $f_{\text{LP-T}}$ is actually a semi-supervised classifier, we implement the state-of-the-art Personal/Impersonal approach for semi-supervised learning on sentiment classification [Li et al., 2010]. Here, we randomly select 200 samples as the initial labeled data. Table 3 shows that our LP-based target classifier achieves comparative performance to the state-of-the-art Personal/Impersonal approach.

5.3 Performance of active learning-based cross-domain sentiment classification

In this section, we systematically evaluate the performance of our active learning approach in cross-domain sentiment classification. As emphasized in Section 4, there are two main steps, i.e., the QBC-based sample selection strategy and the sample classification with classifier combination, in our approach. Accordingly, we name our approach as **QBC+Combination**. For comparison, we implement following two strong baselines,

- **Random+Combination**, which randomly selects the samples from the unlabeled data and uses the classifier combination for prediction. We performs five runs of this approach and report the average performance.
- **Uncertainty+Single**, which first actively selects the samples from the unlabeled data with uncertainty sampling, and then emerge the newly-added labeled data and the initial labeled data to train a single LP-based classifier.

In the implementation, the top 10 informative samples are selected for manual annotation in each iteration (N=10) when **QBC+Combination** and **Uncertainty+Single** are employed. Figure 5 illustrates the performances of above three approaches. From Figure 5, we can see that labeling a small amount of samples in the target domain is always helpful. For example, through actively labeling only 200 samples in the target domain, our approach significantly improves the adaptation performances, which even beat those of using in-domain classifiers trained with 1600 labeled samples.

Our approach consistently performs significantly better than either **QBC+Combination** or **Uncertainty+Single** in almost all pairs (except E->D) when the selected samples are less than 200 (p-value < 0.01). This verifies the advantage of our approach in both QBC-based sample selection and combination-based sample classification. When more than 200 samples are selected, the performances of above three approaches become similar in some cases.

Although our approach does not provide a significant improvement in E->D because the LP-based target classifier in DVD performs badly in this case, our approach still yields no worse results than the other two approaches.

6 Conclusion

In this paper, we address domain adaptation in sentiment classification when the source and target domains differ significantly. We propose a novel active learning approach for cross-domain sentiment classification by leveraging QBC-based sample selection and combination-based classifier classification. Empirical studies indicate that our

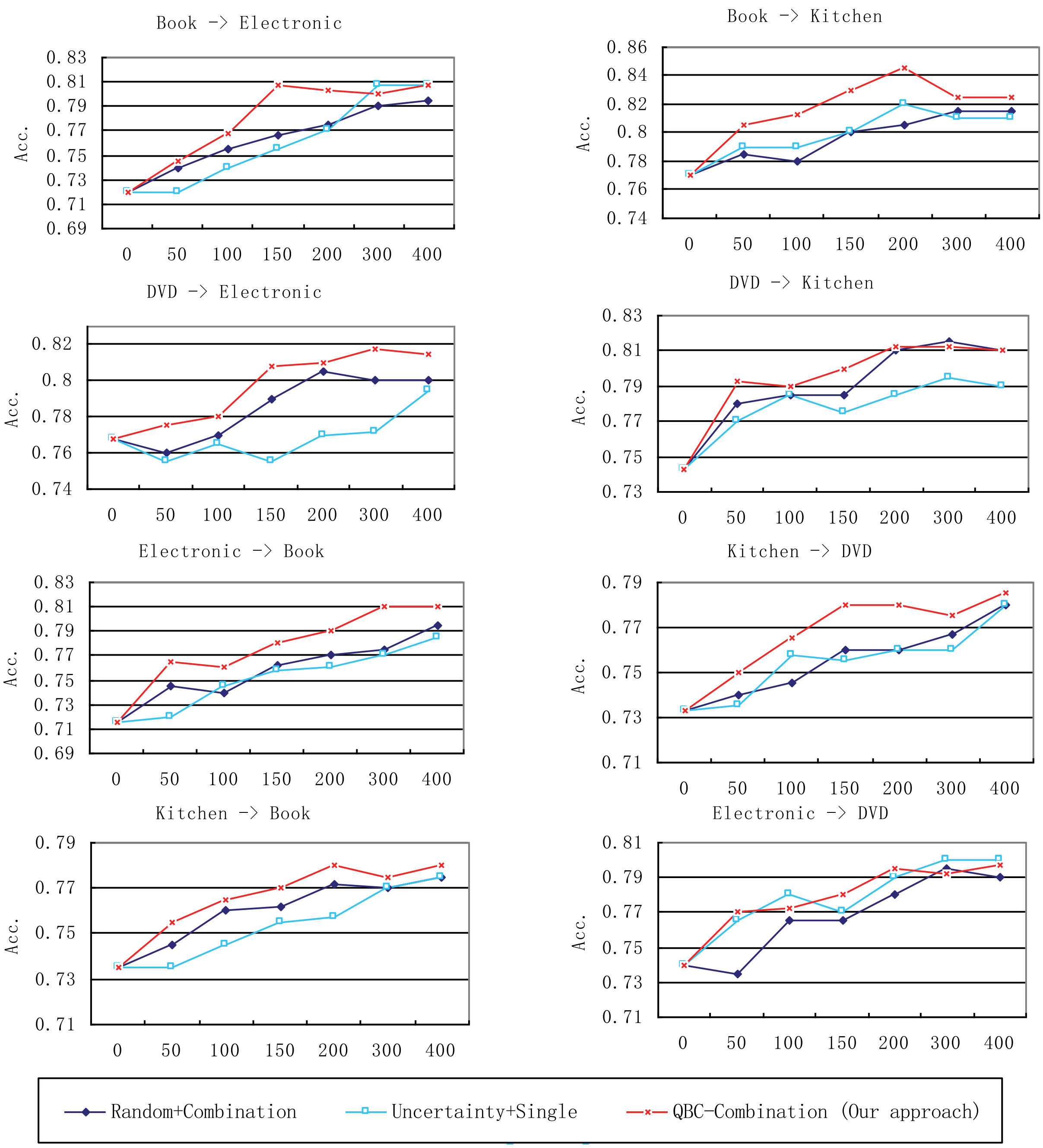

Figure 5: Impact of different numbers of selected samples in the target domain on cross-domain sentiment classification

proposed approach significantly outperforms the state-of-the-art ones. Furthermore, it shows that with only 200 labeled documents from the target domain, our approach could achieve comparable performances to those in-domain classifiers trained with 1600 labeled documents.

The research on active learning for cross-domain sentiment classification is still in its early stage. In our future work, we will exploit more effective algorithms to improve the performances of the source and target classifiers. Meanwhile, we would like to adapt our active learning approach to other cross-domain tasks in natural language processing.

Acknowledgments

The research work described in this paper has been partially supported by three NSFC grants, No.61003155, and No.60873150, one National High-tech Research and Development Program of China No.2012AA011102, Open

Projects Program of National Laboratory of Pattern Recognition, and one project supported by Zhejiang Provincial Natural Science Foundation of China, No.Y13F020030. We also thank the three anonymous reviewers for their helpful comments.

References

[Aue and Gamon, 2005] Aue A. and M. Gamon. 2005. Customizing Sentiment Classifiers to New Domains: A Case Study. *Technical report*, Microsoft Research.

[Blizer *et al*., 2007] Blitzer J., M. Dredze and F. Pereira. 2007. Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification. In *Proceedings of ACL-07*, pp.440-447.

[Bollegala *et al*., 2011] Bollegala D., D. Weir, and J. Carroll. 2011. Using Multiple Sources to Construct a Sentiment Sensitive Thesaurus for Cross-Domain Sentiment Classification. In *Proceedings of ACL-11*, pp.132-141.

[Chan and Ng, 2007] Chan Y. and H. Ng. 2007. Domain Adaptation with Active Learning for Word Sense Disambiguation. In *Proceedings of ACL-07*, pp.49-56.

[Daumé III, 2007] Daumé III H. 2007. Frustratingly Easy Domain Adaptation. In *Proceedings of ACL-07*, 256–263.

[Freund *et al*., 92] Freund Y., H. Seung, E. Shamir, and N. Tishby. Informatio *Proceeding of NIPS-92*n, Prediction, and Query by Committee. In, pp.483-490.

[He *et al*., 2011] He Y., C. Lin and H. Alani. 2011. Automatically Extracting Polarity-Bearing Topics for Cross-Domain Sentiment Classification. In *Proceeding of ACL-11*, pp.123-131.

[Jiang and Zhai, 2007] Jiang J. and C. Zhai. 2007. Instance Weighting for Domain Adaptation in NLP. In *Proceedings of ACL-07*, pp.264–271.

[Li *et al*., 2010] Li S., C. Huang, G. Zhou and S. Lee. 2010. Employing Personal/Impersonal Views in Supervised and Semi-supervised Sentiment Classification. In *Proceedings of ACL-10*, pp.414-423.

[Pan and Yang, 2010] Pan S. and Q. Yang. 2010. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*. vol.22(10), pp.1345-1359.

[Pang and Lee, 2008] Pang B. and L. Lee. 2008. Opinion Mining and Sentiment Analysis: Foundations and Trends. *Information Retrieval*, vol.2(12), pp.1-135.

[Pang *et al*., 2002] Pang B., L. Lee, and S. Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. In *Proceedings of EMNLP-02*, pp.79-86.

[Rai *et al*., 2010] Rai P., A. Saha, H. Daum´e III and S. Venkatasubramanian. 2010. Domain Adaptation Meets Active Learning. In *Proceedings of NAACL-10 Workshop on Active Learning for Natural Language Processing*, pp.27-32.

[Shen *et al*., 2004] Shen D., J. Zhang, J. Su, G. Zhou and C. Tan. 2004. Multi-Criteria-based Active Learning for Named Entity Recognition. In *Proceedings of ACL-04*, pp.589-596.

[Shi *et al*., 2008] Shi X., W. Fan and J. Ren. 2008. Actively Transfer Domain Knowledge. In *Proceedings of ECML/PKDD-08*, pp.342-357.

[Sindhwani and Melville, 2008] Sindhwani V. and P. Melville. 2008. Document-Word Co-Regularization for Semi-supervised Sentiment Analysis. In *Proceedings of ICDM-08*, pp.1025- 1030.

[Turney, 2002] Turney P. 2002. Thumbs up or Thumbs down? Semantic Orientation Applied to Unsupervised Classification of reviews. In *Proceedings of ACL-02*, pp.417-424.

[Wang and Yao, 2011]Wang S. and X. Yao. Relationships Between Diversity of Classification Ensembles and Single-Class Performance Measures. *IEEE Transactions on Knowledge and Data Engineering*, DOI: 10.1109/TKDE.2011.207.

[Zhu *et al*., 2008] Zhu J., H. Wang, T. Yao and B. Tsou. 2008. Active Learning with Sampling by Uncertainty and Density for Word Sense Disambiguation and Text Classification. In *Proceedings of COLING-08*, pp.1137-1144.

[Zhu and Ghahramani, 2002] Zhu X. and Z. Ghahramani. 2002. Learning from Labeled and Unlabeled Data with Label Propagation. *CMU CALD Technical Report*. CMU-CALD-02-107.