

Sentiment Domain Adaptation with Multiple Sources

Fangzhao Wu and Yongfeng Huang*

Tsinghua National Laboratory for Information Science and Technology
Department of Electronic Engineering
Tsinghua University, Beijing, China
wufangzhao@gmail.com, yfhuang@tsinghua.edu.cn

Abstract

Domain adaptation is an important research topic in sentiment analysis area. Existing domain adaptation methods usually transfer sentiment knowledge from only one source domain to target domain. In this paper, we propose a new domain adaptation approach which can exploit sentiment knowledge from multiple source domains. We first extract both global and domain-specific sentiment knowledge from the data of multiple source domains using multi-task learning. Then we transfer them to target domain with the help of words' sentiment polarity relations extracted from the unlabeled target domain data. The similarities between target domain and different source domains are also incorporated into the adaptation process. Experimental results on benchmark dataset show the effectiveness of our approach in improving cross-domain sentiment classification performance.

1 Introduction

Sentiment classification is a hot research topic in natural language processing field, and has many applications in both academic and industrial areas (Pang and Lee, 2008; Liu, 2012; Wu et al., 2015; Wu and Huang, 2016). Sentiment classification is widely known as a domain-dependent task (Blitzer et al., 2007; Glorot et al., 2011). The sentiment classifier trained in one domain may not perform well in another domain. This is because sentiment expressions used in different domains are usually different. For example, “boring”

and “lengthy” are frequently used to express negative sentiments in *Book* domain. However, they rarely appear in *Electronics* domain (Bollegala et al., 2011). Thus a sentiment classifier trained in *Electronics* domain cannot accurately predict their sentiments in *Book* domain. In addition, the same word may convey different sentiments in different domains. For example, in *Electronics* domain “easy” is usually used in positive reviews, e.g., “this digital camera is easy to use.” However, it is frequently used as a negative word in *Movie* domain. For instance, “the ending of this movie is easy to guess.” Thus, the sentiment classifier trained in one domain usually cannot be applied to another domain directly (Pang and Lee, 2008).

In order to tackle this problem, sentiment domain adaptation has been widely studied (Liu, 2012). For example, Blitzer et al. (2007) proposed to compute the correspondence among features from different domains using their associations with pivot features based on structural correspondence learning (*SCL*). Pan et al. (2010) proposed a spectral feature alignment (*SFA*) algorithm to align the domain-specific words from different domains in order to reduce the gap between source and target domains. However, all of these methods transfer sentiment information from only one source domain. When the source and target domains have significant difference in feature distributions, the adaptation performance will heavily decline. In some cases, the performance of sentiment domain adaptation is even worse than that without adaptation, which is usually known as *negative transfer* (Pan and Yang, 2010).

In this paper we propose a new domain adaptation approach for cross-domain sentiment classification. Our approach can exploit the sentiment information in multiple source domains to reduce the risk of negative transfer effectively. Our approach consists of two steps, i.e., *training* and

*Corresponding author.

adaptation. At the *training* stage, we extract two kinds of sentiment models, i.e., the global model and the domain-specific models, from the data of multiple source domains using multi-task learning. The global sentiment model can capture the common sentiment knowledge shared by various domains, and has better generalization performance than the sentiment model trained in a single source domain. The domain-specific sentiment model can capture the specific sentiment knowledge in each source domain. At the *adaptation* stage, we transfer both kinds of sentiment knowledge to target domain with the help of the words' sentiment graph of target domain. The sentiment graph contains words' domain-specific sentiment polarity relations extracted from the syntactic parsing results of the unlabeled data in target domain. Since sentiment transfer between similar domains is more effective than dissimilar domains, we incorporate the similarities between target domain and different source domains into the adaptation process. In order to estimate the similarity between two domains, we propose a novel domain similarity measure based on their sentiment graphs. Extensive experiments were conducted on the benchmark Amazon product review dataset. The experimental results show that our approach can improve the performance of cross-domain sentiment classification effectively.

2 Related work

Sentiment classification is widely known as a domain-dependent task, since different expressions are used to express sentiments in different domains (Blitzer et al., 2007). The sentiment classifier trained in one domain may not perform well in another domain. Since there are massive domains, it is impractical to annotate enough data for each new domain. Thus, *domain adaptation*, or so called *cross-domain sentiment classification*, which transfers the sentiment knowledge from domains with sufficient labeled data (i.e., source domain) to a new domain with no or scarce labeled data (i.e., target domain), has been widely studied. Existing domain adaptation methods mainly transfer sentiment information from only one source domain. For example, Blitzer et al. (2007) proposed a domain adaptation method based on structural correspondence learning (*SCL*). In their method, a set of pivot features are first selected according to their associations with source domain

labels. Then the correspondence among features from source and target domains is computed using their associations with pivot features. In order to reduce the gap between source and target domains, Pan et al. (2010) proposed a spectral feature alignment (*SFA*) algorithm to align the domain-specific sentiment words from different domains into clusters. He et al. (2011) proposed to extract polarity-bearing topics using joint sentiment-topic (*JST*) model to expand the feature representations of texts from both source and target domains. Li et al. (2009) proposed to transfer sentiment knowledge from source domain to target domain using nonnegative matrix factorization. A common shortcoming of above methods is that if the source and target domains have significantly different distributions of sentiment expressions, then the domain adaptation performance will heavily decline (Li et al., 2013).

Using multiple source domains in cross-domain sentiment classification has also been explored. Glorot et al. (2011) proposed a sentiment domain adaptation method based on a deep learning technique, i.e., Stacked Denoising Auto-encoders. The core idea of their method is to learn a high-level representation that can capture generic concepts using the unlabeled data from multiple domains. Yoshida et al. (2011) proposed a probabilistic generative model for cross-domain sentiment classification with multiple source and target domains. In their method, each word is assigned three attributes, i.e., the domain label, the domain dependence/independence label, and sentiment polarity. Bollegala et al. (2011) proposed to construct a sentiment sensitive thesaurus for cross-domain sentiment classification using data from multiple source domains. This thesaurus is used to expand the feature vectors for both training and classification. However, the similarities between target domain and different source domains are not considered in these methods. In addition, although unlabeled data is utilized in these methods, the useful word-level sentiment knowledge in the unlabeled target domain data is not exploited.

General-purpose multiple source domain adaptation methods have also been studied. For example, Mansour et al. (2009) proposed a distribution weighted hypothesis combination approach, and gave theoretical guarantees for it. However, this method is based on the assumption that target distribution is some mixture of source distri-

butions, which may not hold in sentiment domain adaptation scenario. Duan et al. (2009) proposed a Domain Adaptation Machine (DAM) method to learn a Least-Squares SVM classifier for target domain by leveraging the classifiers independently trained in multiple source domains. Chattopadhyay et al. (2011) explored to assign pseudo labels to unlabeled samples in the target domain using the classifiers from multiple source domains. Then target domain classifier is trained on these pseudo labeled samples. Compared with these general-purpose domain adaptation methods with multiple source domains, our approach is more suitable for sentiment domain adaptation because our approach exploits more sentiment-related characteristics and knowledge, such as the general sentiment knowledge shared by different domains and the word-level sentiment polarity relations, which is validated by experiments.

3 Sentiment Graph Extraction and Domain Similarity Measure

In this section we introduce two important components used in our sentiment domain adaptation approach, i.e., the words’ sentiment graph and domain similarity.

3.1 Sentiment Graph Extraction

Compared with labeled data, unlabeled data is usually much easier and cheaper to collect on a large scale. Although unlabeled samples are not associated with sentiment labels, they can still provide a lot of useful sentiment information for domain adaptation. For example, if “great” and “quick” are frequently used to describe the same target in the same review of *Kitchen* domain, then they probably convey the same sentiment polarity in this domain. Since “great” is a general positive word in both *Book* and *Kitchen* domains, we can infer that “quick” is also a positive word in *Kitchen* domain when transferring from *Book* domain to *Kitchen* domain.

Motivated by above observations, in this paper we propose to extract sentiment polarity relations among words from massive unlabeled data for sentiment domain adaptation. Two kinds of polarity relations are explored, i.e., sentiment coherent relation and sentiment opposite relation. The former means that two words convey the same sentiment polarity while the latter indicates opposite sentiment polarities. These polarity relations are ex-

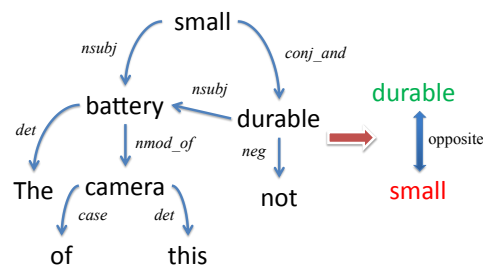


Figure 1: An illustrative example of extracting sentiment polarity relations from syntactic parsing results.

tracted from the syntactic parsing results according to manually selected rules. Two rules are used to extract sentiment coherent relations. The first one is that two words are connected by coordinating conjunctions such as “and” and “as well as”. For example, a review in *Kitchen* domain may be “it is so high-quality and professional.” Since “high-quality” and “professional” are connected by the coordinating conjunction “and”, we infer that they probably convey the same sentiment polarity. The second rule is that two words are not directly connected but are used to describe the same target in the same sentence. For example, a review in *Electronics* domain may be “It is a beautiful, durable, easy-to-use camera.” Since “beautiful”, “durable”, and “easy-to-use” are all used to describe the same camera in the same review, they tend to convey the same sentiment polarity. We also propose two rules for extracting sentiment opposite relations. The first rule is that two words are connected by adversative conjunctions such as “but” and “however”. The second rule is that two words are connected by coordinating conjunctions but there is a negation symbol before one of them. For example, a review may be “The battery of this camera is small and not durable.” We can infer that “small” and “durable” may convey opposite sentiments when they are used to describe camera battery. An illustrative example of extracting sentiment polarity relations from syntactic parsing results is shown in Fig. 1.

Based on the sentiment polarity relations among words extracted from the unlabeled data, we can build a words’ sentiment graph for each domain. The nodes of the sentiment graph represent words and the edges stand for sentiment polarity relations. We denote $\mathbf{R} \in \mathbb{R}^{D \times D}$ as the words’ sentiment graph of a specific domain. $R_{i,j}$ represents the sentiment polarity relation score between words i

and j . In this paper we define $R_{i,j}$ as $\frac{n_{i,j}^C - n_{i,j}^O}{n_{i,j}^C + n_{i,j}^O}$, where $n_{i,j}^C$ and $n_{i,j}^O$ represent the frequencies of words i and j sharing coherent and opposite sentiment polarity relations respectively in all the unlabeled samples. Thus, $R_{i,j} \in [-1, 1]$. If $R_{i,j} > 0$, then words i and j tend to convey the same sentiment polarity. Similarly, if $R_{i,j} < 0$, then these two words are more likely to convey opposite sentiments. The absolute value of $R_{i,j}$ represents the confidence of this sentiment polarity relation.

3.2 Domain Similarity

Different pairs of domains have different sentiment relatedness (Remus, 2012; Wu and Huang, 2015). Researchers have found that sentiment domain adaptation between similar domains, such as *Kitchen* and *Electronics*, is much more effective than that between dissimilar domains, such as *Kitchen* and *Book* (Blitzer et al., 2007; Pan et al., 2010). Thus, it is beneficial if we take the similarity between source and target domains into consideration when transferring sentiment knowledge.

In this paper we explore two methods to measure domain similarity. The first one is based on term distribution. The assumption behind this method is that similar domains usually share more common terms than dissimilar domains. For example, *Smart Phone* and *Digital Camera* domains share many common terms such as “screen”, “battery”, “light”, and “durable”, while the term distributions of *Digital Camera* and *Book* domains may have significant difference. Term distribution based domain similarity measures, such as \mathcal{A} -distance, have been explored in previous works (Blitzer et al., 2007). Inspired by (Remus, 2012), here we apply Jensen-Shannon divergence to measure domain similarity based on term distributions, which is more easy to compute than \mathcal{A} -distance. Denote $\mathbf{t}^m \in \mathbb{R}^{D \times 1}$ as the term distribution of domain m , where t_w^m is the probability of term w appearing in domain m . Then the similarity between domains m and n is formulated as:

$$\begin{aligned} TermSim(m, n) &= 1 - D_{JS}(\mathbf{t}^m, \mathbf{t}^n) \\ &= 1 - \frac{1}{2}(D_{KL}(\mathbf{t}^m, \bar{\mathbf{t}}) + D_{KL}(\mathbf{t}^n, \bar{\mathbf{t}})), \end{aligned} \quad (1)$$

where $\bar{\mathbf{t}} = \frac{1}{2}(\mathbf{t}^m + \mathbf{t}^n)$ is the average distribution, and $D_{KL}(\cdot, \cdot)$ is the Kullback-Leibler divergence:

$$D_{KL}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^D p_i \log_2 \left(\frac{p_i}{q_i} \right). \quad (2)$$

We can verify that $D_{JS}(\mathbf{t}^m, \mathbf{t}^n) \in [0, 1]$. Thus, the range of $TermSim(m, n)$ is also $[0, 1]$.

The term distribution based domain similarity can measure whether similar words are used in two domains. However, sharing similar terms does not necessarily mean that sentiment expressions are used similarly in these domains. For example, *CPU* and *Battery* are both related to electronics. The word “fast” is positive when used to describe *CPU*. However, it is frequently used as a negative word in *Battery* domain. For example, “this battery runs out fast.” Thus, it is more useful to measure domain similarity based on sentiment word distributions. However, although we can infer the sentiment word distributions of source domains according to labeled samples, it is difficult to compute the sentiment word distribution of target domain, since the labeled data does not exist or is very scarce in target domain.

In order to tackle this problem, in this paper we propose to estimate the similarity between two domains based on their sentiment graphs. Similar domains usually share more common sentiment words and sentiment word pairs than dissimilar domains. In addition, the polarity relation scores of a pair of words in the sentiment graphs of similar domains are also more similar. In other words, they tend to be both positive or negative in these two domains. Motivated by above observations, the domain similarity based on sentiment graph is formulated as follows:

$$SentiSim(m, n) = \frac{\sum_{w=1}^D \sum_{v \neq w} |R_{w,v}^m + R_{w,v}^n| \cdot N_{w,v}^{m \cap n}}{\sum_{w=1}^D \sum_{v \neq w} (|R_{w,v}^m| \cdot N_{w,v}^m + |R_{w,v}^n| \cdot N_{w,v}^n)}, \quad (3)$$

where $R_{w,v}^m$ is the sentiment polarity relation score between words w and v in domain m , and $N_{w,v}^m$ is its frequency in this domain. $N_{w,v}^{m \cap n} = \min\{N_{w,v}^m, N_{w,v}^n\}$. We can verify that $SentiSim(m, n) \in [0, 1]$. If two domains have more common sentiment word pairs and the polarity relation scores of these word pairs are more similar, then these two domains share higher domain similarity according to Eq. (3).

4 Sentiment Domain Adaptation with Multiple Sources

In this section we introduce our sentiment domain adaptation approach in detail. First we introduce several notations that will be used in following discussions. Assume there are M source domains.

Denote $\{\mathbf{X}^m \in \mathbb{R}^{N_m \times D}, \mathbf{y}^m \in \mathbb{R}^{N_m \times 1}\}$ as the labeled data in source domain m , where N_m is the number of labeled samples and D is the size of feature vector. $\mathbf{x}_i^m \in \mathbb{R}^{D \times 1}$ is the feature vector of the i_{th} sample in domain m , and its sentiment label is y_i^m . In this paper we focus on sentiment polarity classification, and $y_i^m \in \{+1, -1\}$. Denote $\mathbf{w} \in \mathbb{R}^{D \times 1}$ as the global sentiment model extracted from multiple source domains and $\mathbf{w}_m \in \mathbb{R}^{D \times 1}$ as the domain-specific sentiment model of source domain m . Denote $\mathbf{w}_t \in \mathbb{R}^{D \times 1}$ as the domain-specific sentiment model of target domain. Denote $f(\mathbf{x}, y, \mathbf{w})$ as the loss of classifying sample \mathbf{x} into label y under linear classification model \mathbf{w} . Our approach is flexible to the selection of loss function f , which can be square loss, logistic loss, and hinge loss. Denote $\mathbf{R}^m \in \mathbb{R}^{D \times D}$ as the sentiment graph knowledge of domain m , and $S_{m,t} \in [0, 1]$ as the similarity between source domain m and target domain.

Our sentiment domain adaptation with multiple sources approach (*SDAMS*) consists of two steps, i.e., *training* and *adaptation*. At the *training* stage, the global and domain-specific sentiment knowledge are extracted from the data of multiple source domains. And at the *adaptation* stage, these two kinds of sentiment knowledge are transferred to target domain by incorporating the sentiment graph knowledge of target domain and the similarities between target and source domains.

4.1 Training

Given the labeled data and the sentiment graph knowledge of multiple source domains, at the *training* stage, our goal is to train a robust global sentiment model to capture the general sentiment knowledge shared by various domains and a domain-specific sentiment model for each source domain. The model of the *training* process is motivated by multi-task learning (Evgeniou and Pontil, 2004; Liu et al., 2009) and is formulated as:

$$\begin{aligned} \arg \min_{\mathbf{w}, \mathbf{w}_m} \mathcal{L}(\mathbf{w}, \mathbf{w}_m) &= \sum_{m=1}^M \frac{1}{N_m} \sum_{i=1}^{N_m} f(\mathbf{x}_i^m, y_i^m, \mathbf{w} + \mathbf{w}_m) \\ &+ \alpha \sum_{m=1}^M \sum_{i=1}^D \sum_{j \neq i} R_{i,j}^m |(w_i + w_{m,i}) - (w_j + w_{m,j})| \\ &+ \lambda_1 \|\mathbf{w}\|_2^2 + \lambda_1 \sum_{m=1}^M \|\mathbf{w}_m\|_2^2 + \lambda_2 \|\mathbf{w}\|_1 + \lambda_2 \sum_{m=1}^M \|\mathbf{w}_m\|_1, \end{aligned} \quad (4)$$

where α , λ_1 , and λ_2 are nonnegative regularization coefficients. The sentiment classification model of

each source domain is decomposed into two components, i.e., a global one and a domain-specific one. The global sentiment model is shared by all source domains and is trained in these domains simultaneously. It is used to capture the general sentiment knowledge, such as the general sentiment words “great”, “worst”, “perfect” and so on. The domain-specific sentiment model is trained on the labeled data within one source domain and is used to capture the specific sentiment knowledge of this domain. For example, the domain-specific sentiment word “easy” is a positive word in *Electronics* domain but is used as a negative word in *Movie* domain.

In Eq. (4), the first term means minimizing the empirical classification loss on the labeled data of multiple source domains. In this way we incorporate the sentiment information in labeled samples into sentiment classifier learning. In the second term we incorporate the sentiment graph knowledge of each source domain. It is motivated by graph-guided fused Lasso (Chen et al., 2012). If two words have strong coherent (or opposite) sentiment polarity relations, then we constrain that their sentiment scores are more similar (or dissimilar) with each other in the final classification model. The L_1 -norm regularization terms are motivated by Lasso (Tibshirani, 1996). It can set many minor sentiment scores in the models to exact zeros. Since not all the words convey sentiments, these terms can help conduct sentiment word selection. We also incorporate the L_2 -norm regularization terms in order to improve model stability in high-dimensional problems, which is inspired by elastic net regularization (Zou and Hastie, 2003).

4.2 Adaptation

At the *adaptation* stage, we incorporate the global sentiment knowledge, the domain-specific sentiment knowledge of multiple source domains, the sentiment graph knowledge of target domain, and the domain similarities between target and source domains into a unified framework to learn an accurate sentiment classifier for target domain. The model of our *adaptation* framework is formulated as follows:

$$\begin{aligned} \arg \min_{\mathbf{w}_t} \mathcal{L}(\mathbf{w}_t) &= \sum_{m=1}^M S_{m,t} \|\mathbf{w}_m - \mathbf{w}_t\|_2^2 + \lambda_1 \|\mathbf{w}_t\|_2^2 \\ &+ \lambda_2 \|\mathbf{w}_t\|_1 + \beta \sum_{i=1}^D \sum_{j \neq i} R_{i,j}^t |(w_i + w_{t,i}) - (w_j + w_{t,j})|, \end{aligned} \quad (5)$$

where β , λ_1 , and λ_2 are nonnegative regularization coefficients. The final sentiment classifier of the target domain is a linear combination of \mathbf{w} and \mathbf{w}_t , i.e., $\mathbf{w} + \mathbf{w}_t$, where \mathbf{w} is the global sentiment model extracted from multiple source domains at the *training* stage, and \mathbf{w}_t is the domain-specific sentiment model of target domain learned at the *adaptation* stage. In the first term of Eq. (5), we transfer the knowledge in domain-specific sentiment models from multiple source domains to \mathbf{w}_t . Since the sentiment knowledge transfer between similar domains is more effective, the transfer of domain-specific sentiment knowledge is weighted by the similarities between target domain and different source domains. If target domain is more similar with source domain m than source domain n (i.e., $S_{m,t} > S_{n,t}$), then more domain-specific sentiment knowledge will be transferred to \mathbf{w}_t from \mathbf{w}_m than \mathbf{w}_n . Through the last term we incorporate the sentiment graph knowledge extracted from massive unlabeled data of target domain into the adaptation process. If two words share strong coherent (or opposite) sentiment polarity relations in the target domain, then we constrain that their sentiment scores in the sentiment classification model of target domain are more similar (or dissimilar). This term can help transfer the sentiment knowledge from source domains to target domain more effectively. For example, if we know that “great” is a positive word in the global sentiment model and there is a strong coherent polarity relation between “easy” and “great” in *Electronics* domain, then we can infer that “easy” is also a positive word in this domain.

5 Experiments

5.1 Dataset and Experimental Settings

The dataset used in our experiments is the famous Amazon product review dataset collected by Blitzer et al. (2007). It is widely used as a benchmark dataset for cross-domain sentiment classification. Four domains, i.e., *Book*, *DVD*, *Electronics* and *Kitchen*, are included in this dataset. Each domain contains 1,000 positive and 1,000 negative reviews. Besides, a large number of unlabeled reviews are provided. The detailed statistics of this dataset are shown in Table 1.

In our experiments, each domain was selected in turn as target domain, and remaining domains as source domains. In each experiment, we randomly selected N labeled samples from the

Table 1: The statistics of the dataset.

Domain	Book	DVD	Electronics	Kitchen
positive	1,000	1,000	1,000	1,000
negative	1,000	1,000	1,000	1,000
unlabeled	973,194	122,438	21,009	17,856

source domains to train sentiment models at the *training* stage. These samples were balanced among different source domains. In order to perform fair comparisons with baseline methods, following (Bollegala et al., 2011), we limited the total number of training samples, i.e., N , to 1,600. The target domain sentiment classifier was tested on all the labeled samples of target domain. Following (Blitzer et al., 2007), unigrams and bigrams were used as features. The sentiment polarity relations of bigrams were extracted by expanding the polarity relations between unigrams using modifying relations. For example, from the review “this phone is very beautiful and not expensive,” we extract not only sentiment polarity relation between “beautiful” and “expensive”, but also polarity relation between “beautiful” and “not_expensive” (coherent sentiment), and that between “very_beautiful” and “expensive” (opposite sentiment), since “very” and “not” are used to modify “beautiful” and “expensive” respectively. Classification accuracy was selected as the evaluation metric. We manually set β in Eq. (5) to 0.01. The values of α , λ_1 , and λ_2 in Eq. (4) were selected using cross validation. The optimization problems in Eq. (4) and Eq. (5) were solved using alternating direction method of multipliers (ADMM) (Boyd et al., 2011). Each experiment was repeated 10 times independently and average results were reported.

5.2 Comparison of Domain Similarity Measures

In this section, we conducted experiments to compare the effectiveness of the two kinds of domain similarity measures introduced in Section 3.2 in sentiment domain adaptation task. The experimental results are summarized in Fig. 2. The classification loss function used in our approach is hinge loss. The results of other loss functions show similar patterns.

From Fig. 2 we can see that the domain similarity measure based on sentiment graph performs consistently better than that based on term distribution in our approach. This result validates our

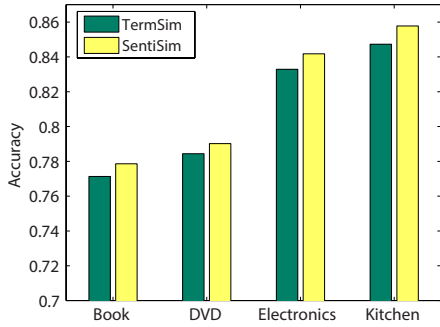


Figure 2: The performance of our approach with different kinds of domain similarity measure.

assumption in Section 3.2 that the sentiment graph based domain similarity can better model the sentiment relatedness between different domains than that based on term distribution in sentiment domain adaptation task. In all the following experiments, the sentiment graph based domain similarities were used in our approach.

5.3 Performance Evaluation

In this section we conducted experiments to evaluate the performance of our approach by comparing it with a series of baseline methods. The methods to be compared are: 1) *SCL*, domain adaptation based on structural correspondence learning (Blitzer et al., 2007); 2) *SFA*, domain adaptation based on spectral feature alignment (Pan et al., 2010); 3) *SCL-com* and *SFA-com*, adapting *SCL* and *SFA* to multiple source domain scenario by first training a cross-domain sentiment classifier in each source domain and then combining their classification results using majority voting; 4) *SST*, cross-domain sentiment classification by using multiple source domains to construct a sentiment sensitive thesaurus for feature expansion (Bollegala et al., 2011); 5) *IDDIWP*, multiple-domain sentiment analysis by identifying domain dependent/independent word polarity (Yoshida et al., 2011); 6) *DWHC*, *DAM* and *CP-MDA*, three general-purpose multiple source domain adaptation methods proposed in (Mansour et al., 2009), (Duan et al., 2009) and (Chattopadhyay et al., 2011) respectively; 7) *SDAMS-LS*, *SDAMS-SVM*, and *SDAMS-Log*, our proposed sentiment domain adaptation approaches with square loss, hinge loss, and logistic loss respectively; 8) *All-Training*, all the domains were involved in the training phase of our approach and there is no

adaptation phase. This method is introduced to provide an upper bound for the performance of our approach. The experimental results of these methods are summarized in Table 2.

Table 2: The performance of different methods.

	Book	DVD	Electronics	Kitchen
<i>SCL</i>	0.7457	0.7630	0.7893	0.8207
<i>SFA</i>	0.7598	0.7848	0.7808	0.8210
<i>SCL-com</i>	0.7523	0.7675	0.7918	0.8247
<i>SFA-com</i>	0.7629	0.7869	0.7864	0.8258
<i>SST</i>	0.7632	0.7877	0.8363	0.8518
<i>IDDIWP</i>	0.7524	0.7732	0.8167	0.8383
<i>DWHC</i>	0.7611	0.7821	0.8312	0.8478
<i>DAM</i>	0.7563	0.7756	0.8284	0.8419
<i>CP-MDA</i>	0.7597	0.7792	0.8331	0.8465
<i>SDAMS-LS</i>	0.7795	0.7880	0.8398	0.8596
<i>SDAMS-SVM</i>	0.7786	0.7902	0.8418	0.8578
<i>SDAMS-Log</i>	0.7829	0.7913	0.8406	0.8629
<i>All-Training</i>	0.7983	0.8104	0.8463	0.8683

From Table 2 we can see that our approach achieves the best performance among all the methods compared here. *SCL* and *SFA* are famous cross-domain sentiment classification methods. In these methods, the sentiment knowledge is transferred from one source domain to target domain. According to Table 2, our approach performs significantly better than them. This result indicates that the sentiment knowledge extracted from one source domain may contain heavy domain-specific bias and may be inappropriate for the target domain. Our approach can tackle this problem by extracting the global sentiment model from multiple source domains. This global model can capture the general sentiment knowledge shared by various domains and has better generalization performance. It can reduce the risk of negative transfer effectively. Our approach also outperforms *SCL-com* and *SFA-com*. In *SCL-com* and *SFA-com*, the sentiment information in different source domains is combined at the classification stage, while in our approach it is combined at the learning stage. The superior performance of our approach compared with *SCL-com* and *SFA-com* shows that our approach is a more appropriate way to exploit the sentiment knowledge in different source domains. *SST* and *IDDIWP* also utilize data from multiple source domains as our approach. But our approach can still outperform them. This is because in these methods, the similarities between target domain and different source domains are not considered. Since different domains usually

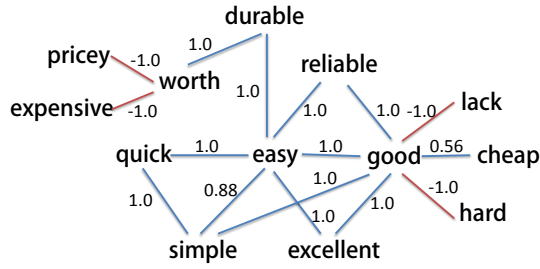


Figure 3: An illustrative example of the sentiment graph of *Electronics* domain. The value on the line represents the sentiment polarity relation score.

have different sentiment relatedness, our approach can exploit the sentiment information in multiple source domains more accurately by incorporating the similarities between target domain and each source domain into the *adaptation* process. Our approach also outperforms the state-of-the-art general-purpose multiple source domain adaptation methods, such as *DWHC*, *DAM*, and *CPMDA*. This is because our approach can exploit more sentiment-related characteristics and knowledge for sentiment domain adaptation, such as the general sentiment knowledge shared by various domains, the sentiment graph based domain similarities, and the word-level sentiment polarity relations. Thus, our approach is more suitable for sentiment domain adaptation than these general-purpose multiple source domain adaptation methods. Another observation from Table 2 is that the performance of our approach is quite close to the upper bound, i.e., *All-Training*, especially in *Electronics* and *Kitchen* domains. This result validates the effectiveness of our approach in sentiment domain adaptation.

5.4 Case Study

In this section we conducted several case studies to further explore how our sentiment domain adaptation approach works. As an illustrative example, we selected *Electronics* domain as the target domain and remaining domains as source domains. The top sentiment words in the global and domain-specific sentiment models learned from the data of multiple source domains are shown in Table 3. A subgraph of the sentiment graph extracted from the unlabeled data of target domain (*Electronics*) is shown in Fig. 3. The top words in the final domain-specific sentiment model of target domain returned by our approach are shown in Table 3.

From Table 3 we have following observations.

First, the global sentiment model extracted from multiple source domains can capture the general sentiment knowledge quite well. It contains many general sentiment words, such as “excellent”, “great”, “waste” and so on. These general sentiment words convey strong sentiment orientations. In addition, their sentiment polarities are consistent in different domains. Thus, the global sentiment model extracted from multiple source domains has good generalization ability and is more suitable for domain adaptation than the sentiment model trained in a single source domain, which may contain heavy domain-specific sentiment bias. Second, the domain-specific sentiment models can capture rich specific sentiment expressions in each source domain. For example, “easy” is a positive word in *Kitchen* domain while “return” is a negative word in this domain. Third, different domains have different domain-specific sentiment expressions. For example, “read” is frequently used as a positive word in *Book* domain, while it is a negative word in *DVD* domain. Thus, it is important to separate the global and the domain-specific sentiment knowledge. In addition, although different sentiment expressions are used in different domains, similar domains may share many common domain-specific sentiment expressions. For example, “easy” and “works” are positive words in both *Electronics* and *Kitchen* domains, and “return” and “broken” are both negative words in them. Thus, transferring the domain-specific sentiment models from similar source domains to target domain is helpful. From Fig. 3 we can see that the sentiment polarity relations in the sentiment graph extracted from massive unlabeled data are reasonable. Words with positive relation scores tend to convey similar sentiments, and words with negative relation scores usually convey opposite sentiments. In addition, this sentiment graph contains rich domain-specific sentiment information in target domain, which is useful to transfer the sentiment knowledge from multiple source domains to target domain. For example, “excellent”, “easy”, “simple”, and “quick” share the same sentiment polarity in *Electronics* domain according to Fig. 3. We can infer that “easy” is positive in this domain using the sentiment of “excellent” in the global sentiment model and the sentiment relation between “easy” and “excellent”. Then we can further infer the sentiments of the domain-specific sentiment words

Table 3: The top words in the global and domain-specific sentiment models.

Global	Positive	excellent, great, best, perfect, love, wonderful, the_best, loved, well, fantastic, enjoy, favorite
	Negative	bad, waste, boring, disappointed, worst, poor, disappointing, disappointment, terrible, poorly
Book	Positive	excellent, wonderful, easy, loved, enjoyable, life, fun, favorite, a_must, read, important, novel
	Negative	no, boring, disappointing, bad, instead, waste, little, writing, poorly, pages, unfortunately
DVD	Positive	enjoy, hope, loved, season, better_than, best, a_must, first, superman, classic, times, back
	Negative	worst, boring, bad, the_worst, terrible, waste, awful, book, horrible, dull, lame, read, hard
Kitchen	Positive	easy, great, perfect, love, works, easy_to, best, little, well, good, nice, long, durable, clean
	Negative	disappointed, back, poor, broken, too, return, off, returned, broke, waste, tried, times, doesn't
Electronics	Positive	excellent, great, perfect, best, love, easy_to, easy, little, the_best, works, good, nice, wonderful
	Negative	disappointed, poor, waste, too, bad, worst, back, broken, return, horrible, off, tried, poorly

“simple” and “quick” in target domain using the polarity of “easy” and their sentiment relations with it, even if they may be covered by no source domain.

6 Conclusion

This paper presents a sentiment domain adaptation approach which transfers the sentiment knowledge from multiple source domains to target domain. Our approach consists of two steps. First, we extract both global and domain-specific sentiment knowledge from the data of multiple source domains. Second, we transfer these two kinds of sentiment knowledge to target domain with the help of the words’ sentiment graph. We proposed to build words’ sentiment graph for target domain by extracting their sentiment polarity relations from massive unlabeled data. Besides, we proposed a novel domain similarity measure based on sentiment graphs, and incorporated the domain similarities between target and different source domains into the domain adaptation process. The experimental results on a benchmark dataset show that our approach can effectively improve the performance of cross-domain sentiment classification.

Acknowledgements

This research is supported by the Key Program of National Natural Science Foundation of China (Grant nos. U1536201 and U1405254), the National Natural Science Foundation of China (Grant no. 61472092), the National High Technology Research and Development Program of China (863 Program) (Grant no. 2015AA020101), the National Science and Technology Support Program of China (Grant no. 2014BAH41B00), and the Initiative Scientific Research Program of Tsinghua University.

References

- John Blitzer, Mark Dredze, Fernando Pereira, et al. 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL*, volume 7, pages 440–447.
- Danushka Bollegala, David Weir, and John Carroll. 2011. Using multiple sources to construct a sentiment sensitive thesaurus for cross-domain sentiment classification. In *ACL:HLT*, pages 132–141.
- Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122.
- Rita Chattopadhyay, Jieping Ye, Sethuraman Panchanathan, Wei Fan, and Ian Davidson. 2011. Multi-source domain adaptation and its application to early detection of fatigue. In *KDD*, pages 717–725. ACM.
- Xi Chen, Qihang Lin, Seyoung Kim, Jaime G Carbonell, Eric P Xing, et al. 2012. Smoothing proximal gradient method for general structured sparse regression. *The Annals of Applied Statistics*, 6(2):719–752.
- Lixin Duan, Ivor W Tsang, Dong Xu, and Tat-Seng Chua. 2009. Domain adaptation from multiple sources via auxiliary classifiers. In *ICML*, pages 289–296. ACM.
- Theodoros Evgeniou and Massimiliano Pontil. 2004. Regularized multi-task learning. In *KDD*, pages 109–117. ACM.
- Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2011. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *ICML*, pages 513–520.
- Yulan He, Chenghua Lin, and Harith Alani. 2011. Automatically extracting polarity-bearing topics for cross-domain sentiment classification. In *ACL:HLT*, pages 123–131.
- Tao Li, Vikas Sindhwani, Chris Ding, and Yi Zhang. 2009. Knowledge transformation for cross-domain sentiment classification. In *SIGIR*, pages 716–717. ACM.

- Shoushan Li, Yunxia Xue, Zhongqing Wang, and Guodong Zhou. 2013. Active learning for cross-domain sentiment classification. In *IJCAI*, pages 2127–2133.
- Jun Liu, Shuiwang Ji, and Jieping Ye. 2009. Multi-task feature learning via efficient $l_{2,1}$ -norm minimization. In *UAI*, pages 339–348.
- Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.
- Yishay Mansour, Mehryar Mohri, and Afshin Roshtamizadeh. 2009. Domain adaptation with multiple sources. In *NIPS*, pages 1041–1048.
- Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. *TKDE*, 22(10):1345–1359.
- Sinno Jialin Pan, Xiaochuan Ni, Jian-Tao Sun, Qiang Yang, and Zheng Chen. 2010. Cross-domain sentiment classification via spectral feature alignment. In *WWW*, pages 751–760. ACM.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135.
- Robert Remus. 2012. Domain adaptation using domain similarity-and domain complexity-based instance selection for cross-domain sentiment analysis. In *2012 IEEE 12th International Conference on Data Mining Workshops*, pages 717–723. IEEE.
- Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.
- Fangzhao Wu and Yongfeng Huang. 2015. Collaborative multi-domain sentiment classification. In *ICDM*, pages 459–468. IEEE.
- Fangzhao Wu and Yongfeng Huang. 2016. Personalized microblog sentiment classification via multi-task learning. In *AAAI*, pages 3059–3065.
- Fangzhao Wu, Yangqiu Song, and Yongfeng Huang. 2015. Microblog sentiment classification with contextual knowledge regularization. In *AAAI*, pages 2332–2338.
- Yasuhisa Yoshida, Tsutomu Hirao, Tomoharu Iwata, Masaaki Nagata, and Yuji Matsumoto. 2011. Transfer learning for multiple-domain sentiment analysis: identifying domain dependent/independent word polarity. In *AAAI*, pages 1286–1291.
- Hui Zou and Trevor Hastie. 2003. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.