Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Words alignment based on association rules for cross-domain sentiment classification*

Xi-bin JIA†1, Ya JIN†1, Ning LI1, Xing SU†‡1, Barry CARDIFF2, Bir BHANU3

1*Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China*
2*School of Electrical and Electronic Engineering, University College Dublin, Belfield, Dublin 4, Ireland*
3*Visualization and Intelligent Systems Laboratory, University of California, Riverside 92521, CA, USA*
†E-mail: jiaxibin@bjut.edu.cn; jinya@emails.bjut.edu.cn; xingsu@bjut.edu.cn

Received Nov. 2, 2016; Revision accepted Mar. 9, 2017; Crosschecked Feb. 15, 2018

Abstract: Automatic classification of sentiment data (e.g., reviews, blogs) has many applications in enterprise user management systems, and can help us understand people's attitudes about products or services. However, it is difficult to train an accurate sentiment classifier for different domains. One of the major reasons is that people often use different words to express the same sentiment in different domains, and we cannot easily find a direct mapping relationship between them to reduce the differences between domains. So, the accuracy of the sentiment classifier will decline sharply when we apply a classifier trained in one domain to other domains. In this paper, we propose a novel approach called words alignment based on association rules (WAAR) for cross-domain sentiment classification, which can establish an indirect mapping relationship between domain-specific words in different domains by learning the strong association rules between domain-shared words and domain-specific words in the same domain. In this way, the differences between the source domain and target domain can be reduced to some extent, and a more accurate cross-domain classifier can be trained. Experimental results on Amazon® datasets show the effectiveness of our approach on improving the performance of cross-domain sentiment classification.

Key words: Sentiment classification; Cross-domain; Association rules
https://doi.org/10.1631/FITEE.1601679 **CLC number:** TP391.1

1 Introduction

With the continuing development of information technology, ever more data is becoming available, which increases the need for automatic tools for analysis and mining. In recent years, there has been a lot of focus on automatic sentiment classification of text, which can provide useful information to customers, companies, and expert systems to make decisions (Balazs and Velasquez, 2016) by mining users' interests and attitudes toward products in large-scale product reviews. Applications include analysis of online product evaluations and comments on social media, public opinion forums, etc. Thus, the sentiment classification of text has become a hot research topic in recent years.

In many cases, supervised classification methods can perform well in sentiment classification and are widely used in analyzing the sentiments of movie reviews (Pang et al., 2002; Zhuang et al., 2006), product reviews (Dave et al., 2003), etc. However, supervised classification methods need two conditions to ensure classification accuracy. First, they need sufficient and well-labeled instances in a problem

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 61703013, 91546111, 91646201, 61672070, and 61672071), the Beijing Municipal Natural Science Foundation (No. 4152005), the Key Projects of Beijing Municipal Education Commission (Nos. KZ201610005009 and KM201810005024), and the International Cooperation Seed Grant from Beijing University of Technology of 2016 (No. 007000514116520)

ORCID: Xing SU, http://orcid.org/0000-0002-4214-3363

domain so that the model can be trained sufficiently. Second, the training data and the test data need to have the same distribution so that the test data can share the information obtained from the training data. However, it is not always easy or feasible to obtain new labeled data in a target domain, and sentiment classification is widely known as a domain-dependent problem (Blitzer et al., 2007). This means that a sentiment classifier trained in one domain with plenty of labeled data may not perform well in another domain. This is because people always tend to use different words to express the same emotion in different domains, so it is difficult to design a robust sentiment classifier that can be effectively used for different domains. Indeed, sentiment word lexicons are always different when evaluating different products or services in different domains. Each domain has many domain-specific sentiment words that are not relevant in other domains. For example, in the domain of Electronics product reviews, the words 'durable' and 'portable' are often used as positive words to express the durability and convenience of products (e.g., "These TV sets were durable and had good quality" or "The recorder is completely portable and very light in weight"). However, they are seldom used in the Books domain or Hotel domain to express emotions. So, the sentiment classifier trained in the Books domain (can be seen as the source domain) cannot accurately predict the sentiment of 'durable' and 'portable' in the Electronics domain (can be seen as the target domain). This raises a research field called cross-domain sentiment classification.

To tackle this problem, a natural solution is to train a domain-specific sentiment classifier for each target domain by using the labeled data in that domain (Pang et al., 2002). However, this natural solution is infeasible for practical applications due to an inability to accurately calculate the number of domains involved in online reviews and the huge cost required to annotate enough samples for each domain (Pan et al., 2010). In addition, although there may be some different sentiment lexicons in different domains, they can still provide knowledge of general sentiment words or domain-shared words (e.g., 'even', 'more', and 'good'). Thus, learning the relationships between domain-specific words and domain-shared words is also useful in cross-domain sentiment classifier training. In addition, large-scale unlabeled reviews in a domain are usually much easier and cheaper to get than labeled reviews, and can provide useful sentiment knowledge for selecting domain-shared words and learning the associations of domain-specific words from these domain-shared words. For example, if 'more' is a domain-shared word and 'perfect' and 'support' are domain-specific words in the Electronics domain and the Books domain, respectively, two association rules can be '(more ⇒perfect) = 0.083 445 95' and '(more ⇒support) = 0.083 057 851', which are learned from unlabeled reviews in the Electronics domain and the Books domain, respectively. Because the domain-specific words 'perfect' and 'support' have strong associations with domain-shared word 'more' in two different domains, they probably have a sentiment relationship in different domains and can make up a word pair that can reduce the difference between two domains. If we can find all word pairs in the source domain and target domain by domain-shared words, then we can reduce the difference between the source and target domains. Thus, in addition to the labeled reviews of the source domain, many unlabeled reviews in the source domain and target domain can be used to create a mapping between the domains for cross-domain sentiment classifier training.

Motivated by the above observations, we propose a new approach called words alignment based on association rules (WAAR) by establishing an indirect mapping relationship between domain-specific words via domain-shared words. The sentiment words that express the same emotion in different domains may be related to each other, but it is not feasible to establish a direct mapping relationship between them. Therefore, our approach aims to establish an indirect mapping relationship between domain-specific words via the domain-shared words (Pan et al., 2010). Specifically, two kinds of sentiment information extracted from the source domain and the target domain are used in our approach to establish an indirect mapping relationship that can be used to train a cross-domain sentiment classifier. The first kind of information is domain-specific and domain-shared sentiment knowledge in two domains, which can be extracted from unlabeled reviews to avoid the high cost of manual annotation. The second kind of information is the direct mapping between domain-specific words and domain-shared words in the same domain, which are mined from the reviews of two domains, respectively. We propose a weighted direct graph-based evaluation model for correlation to illustrate the mapping relationship between domain-specific words and domain-shared words. Based on the constructed mapping graph, the domain-specific words achieve alignment by establishing an indirect mapping relationship between domain-specific words via domain-shared words. An effective cross-domain sentiment classifier is trained using meaningful features, is constructed using the indirect mapping relationship, and can reveal similar syntactic relations between domain-specific words in different domains. The experimental results show that our approach can eliminate some of the differences between domains and improve the performance of sentiment classification.

The main contributions of this paper are summarized as follows:

1. We borrow a strategy modified by Pan et al. (2010) to extract the domain-specific words and domain-shared words from unlabeled reviews for each domain. In addition, we propose an approach by using the association rule learning method based on the Apriori algorithm (Agrawal and Srikant, 1994) to learn the direct mapping relationship between domain-specific words and domain-shared words in the same domain.

2. We propose a weighted direct graph-based evaluation model for correlation to establish an indirect mapping relationship between domain-specific words in different domains via domain-shared words.

3. We evaluate the WAAR approach using extensive experiments on benchmark sentiment datasets and the experimental results validate the effectiveness of our approach.

2 Related work

The aim of sentiment polarity recognition is to automatically predict the sentiment polarity (e.g., positive and negative) of a piece of text. Many machine learning algorithms have been proposed for opinion-oriented information retrieval (also known as opinion mining and sentiment analysis), including unsupervised learning (Turney, 2002), supervised learning (Pang et al., 2002), graph-based semi-supervised learning (Goldberg and Zhu, 2006), and matrix-based decomposition (Li et al., 2009; Zhou et al., 2015). However, these studies focus mainly on the single-domain problem, and sentiment polarity recognition of text is widely known as a domain-dependent task. Cross-domain sentiment analysis was proposed to alleviate the need for repeated training of single-domain sentiment classifiers by using available training data from a source domain, along with little or no training data from the target domain to train the target classifier. Three major approaches to cross-domain sentiment analysis are reviewed in this section: instance adaptation, feature adaptation, and model adaptation.

There are several studies that have proposed instance adaptation for cross-domain classification (Zadrozny, 2004; Jiang and Zhai, 2007; Schölkopf et al., 2007). The basic idea of instance adaptation approaches is to resample source domain instances to equalize the distribution of each sample and train a classifier for the target domain based on resampled source domain datasets. However, even though the methods based on instance adaptation are simple, they do not work when there are tremendous differences between source domains and target domains.

The basic idea of feature adaptation approaches is to transform the data representations of source domains into target domains so that they present the same joint distribution of observations and labels. Blitzer et al. (2007) proposed the structural correspondence learning (SCL) algorithm to exploit domain adaptation techniques for sentiment classification. SCL was inspired by the alternating structural optimization (ASO) multitask learning algorithm, which was proposed by Ando and Zhang (2005). The main idea of SCL is to achieve feature alignment in different domains by choosing a set of pivot features and modeling the correlations between 'pivot features' and other features (called 'non-pivot features'). Pan et al. (2010) proposed a spectral feature alignment (SFA) algorithm to find an alignment between domain-specific and domain-independent features by performing spectral clustering based on a bipartite graph, which is constructed based on their cooccurring relationship between domain-specific and domain-independent features. Glorot et al. (2011) proposed a sentiment-domain adaptation method based on a deep learning technique named stacked denoising auto-encoders (SDA) to learn a high-level representation that can capture generic concepts using the unlabeled data from multiple domains to achieve abstract feature alignment in different domains. However, the SDA algorithm training with gradient descent or other optimization algorithms is slow and highly depends on the initial values. To solve this problem, Chen et al. (2012) proposed the mSDA algorithm, which preserves strong feature learning capabilities without using the optimization algorithm to learn the parameters.

Wu et al. (2010b) proposed an iterative algorithm based on model adaptation that integrates opinion orientations of reviews into a graph-ranking algorithm for cross-domain sentiment analysis. The algorithm can compensate for a shortage of labeled reviews in target domains and use an iterative method to update the target domain model for text sentiment analysis. Wu et al. (2010a) also proposed a method that is based on the random-walk model by simultaneously using reviews and sentiment words from both the source and target domains for cross-domain sentiment analysis.

There are several other ways to overcome the feature divergence problem by creating a common sentiment-sensitive distributional thesaurus for every possible domain. Bollegala et al. (2013) proposed a cross-domain sentiment classifier by grouping different words expressing the same sentiment into one thesaurus and to some extent solve the feature mismatch problem in cross-domain sentiment classification. The method is effective in multiple domains, but it does not fit a single domain. Pantelis et al. (2018) created a tool called DidaxTo to extract domain-oriented sentiment words to be used in an unsupervised classification approach to discover patterns. Li et al. (2012) proposed a method named topic correlation analysis (TCA) for cross-domain text classification, which solves the feature divergence problem in view of topics.

Reviews above show that extracting sentiment-sensitive words from the source domain and target domain and establishing a bridge to align the sentiment polarity of domain-specific words in different domains is a good way to use a sentiment classifier trained in one domain to the others. However, current word-alignment solutions based on the co-occurrence matrix are not good at revealing relationships, so classification accuracy is still not sufficient. Therefore, we propose an unsupervised approach to achieve alignment of domain-specific words by using association rules to explore more accurate relationships and improve the performance of cross-domain sentiment classification.

3 Problem description

For ease of explanation, in this study, we explain only the problem of aligning domain-specific words from one labeled domain (the source domain) to one unlabeled domain (the target domain) in detail. The problem of achieving alignment of domain-specific words from multiple source domains is regarded as a future work. We denote the source domain as D_s and the target domain as D_t. The number of labeled reviews in the source domain is expressed as n_s, and the pair of instances in the source domain $(X_{\mathrm{s}_i}, Y_{\mathrm{s}_i})$ means the i^th sentence instance and its corresponding sentiment label Y_{s_i}, where $Y_{\mathrm{s}_i} \in \{+1, -1\}$, and $+1$ and -1 are sentiment labels. There are only n_t unlabeled reviews in the target domain, and each review is denoted as X_{t_i}:

$$D_\mathrm{s} = \{(X_{\mathrm{s}_i}, Y_{\mathrm{s}_i})\}_{i=1}^{n_\mathrm{s}}, \tag{1}$$

$$D_\mathrm{t} = \{(X_{\mathrm{t}_i})\}_{i=1}^{n_\mathrm{t}}. \tag{2}$$

The task of a cross-domain sentiment classifier is to determine the sentiment polarity of unlabeled datasets from D_t based on the given source domain information D_s. However, the source domain datasets and target domain datasets follow different distributions, because different words are used in different domains to express sentiment or describe characteristics, such as the examples in Table 1. So, to train an accurate classifier to predict the sentiment polarity of unlabeled data from D_t, there are three main subtasks that concern the alignment of domain-specific sentiment words in different domains: (1) identify the domain-shared words; (2) discover the association rules between domain-specific words and domain-shared words; (3) align the domain-specific words. Previously, a unified vocabulary set V for all domains needed to be extracted from all datasets and $|V| = m$, and single words or n-grams can be included in the vocabulary set V as features to represent sentiment data. In the first subtask, we aim to find the domain-shared words that are frequently used and have similar sentiment polarity in both D_s and D_t. These domain-shared words can be used as a bridge to establish a direct mapping relationship with domain-specific words in each domain. Then, an indirect mapping relationship between domain-specific words in different domains will be established via the domain-shared words. In the second subtask, we aim to obtain strong association rules between domain-shared words and domain-specific words: we use $r\,(\mathrm{item}_{\mathrm{sh}} \Rightarrow \mathrm{item}_{\mathrm{sp}})$ to establish a direct mapping relationship between domain-shared words and domain-specific words in the same domain; $\mathrm{item}_{\mathrm{sh}}$ and $\mathrm{item}_{\mathrm{sp}}$ denote the domain-shared words and domain-specific words, respectively. The association rules are used to establish an indirect mapping relationship between domain-specific words from different domains. In the third subtask, we aim to align domain-specific words from both domains by counting the number of domain-shared words that all have strong association rules between domain-specific words from different domains. The difference between domain-specific words from different domains can be reduced by establishing an indirect mapping relationship between domain-specific words from different domains via domain-shared words.

4 Words alignment based on association rules

In this section, we describe our algorithm to adapt association rules mining techniques to establish an indirect mapping relationship between domain-specific words from different domains for cross-domain sentiment classification. The detailed flow of the algorithm is shown in Fig. 1. The raw features of reviews for text classification problems are either single words or n-grams that represent sentiment data. In the following subsections, we describe how to identify domain-shared words, discover the rules, and align the domain-specific words from different domains.

4.1 Identifying domain-shared words

Pan et al. (2010) explored many possible methods that can be used to identify likely domain-shared words. We assume that if the words have higher correlation in every domain, they are more likely to be treated as domain-shared features. Here, we denote the judging process function as $\phi_{\mathrm{shared}}(\cdot)$. Following the principle of the judging process function, m domain-shared words are selected from the vocabulary set V by using the supervised selection criteria. Pan et al. (2010) used a decision criterion modified from the mutual information criterion to select domain-shared words, by computing mutual information to measure the dependence between words from the vocabulary set V and domains from the domain variable set D. We also use this strategy to measure the possibility of a word x being a domain-shared word by counting its occurrence frequency in every domain. The evaluation function of the possibility of a word being a domain-shared word is calculated as follows:

$$\phi_{\mathrm{shared}}(x;D)=\sum_{d\in D}p(x,d)\log_2\left(\frac{p(x,d)}{p(x)p(d)}\right), \quad (3)$$

where x is the word being assessed, D is the domain variable set, and d is an element of D. The joint probability $p(x,d)$ denotes the probability of the co-occurrence of x and d in the given sentiment reviews. $p(x)$ and $p(d)$ denote the marginal probabilities of x and d, respectively.

In the two-domain problem, $\phi_{\mathrm{shared}}(x;D_{\mathrm{s}})$ and $\phi_{\mathrm{shared}}(x;D_{\mathrm{t}})$ are computed for all words in the entire vocabulary set V. The l words that have the smaller values of $\phi_{\mathrm{shared}}(x;D)$ are treated as the set of domain-shared words; the remaining $m-l$ words are deemed domain-specific words. Here, l is an empirical value that represents the number of domain-shared words.

4.2 Discovering the rules

After selecting the domain-shared words based on the above strategies, we can identify the domain-specific words. Then, we need to obtain a direct mapping relationship between the domain-shared words and domain-specific words by using association rules. For this purpose, we use the Apriori algorithm proposed by Agrawal and Srikant (1994), which is often used to determine association rules between items in large datasets. For better understanding, we also construct a directed graph to present the discovered relationships between all words.

As its name implies, the Apriori algorithm uses prior knowledge of the k^{th} itemset to search the $(k+1)^{\mathrm{th}}$ itemset with the candidate sets and the least support s. In this study, we need only 2 itemsets to reflect the mapping relationship between a domain-shared word and a domain-specific word. The pseudo-code description of the Apriori algorithm to discover the 2 itemsets and mine the strong association rules is given in Algorithm 1, where the min-support and min-confidence are empirical values.

The support is defined as the probability of items being included in $\{\mathrm{item}_i^1\}_{i=1}^{n_1}$ and $\{\mathrm{item}_j^2\}_{j=1}^{n_2}$,

Table 1 Positive (+) and negative (−) sentiment reviews in the Books and the Electronics domains

	Books	Electronics
−	The book is filled with **typos** and **grammatical** errors, too **easy** for me; It's bad; What a disappointment it was!	Bad investment!!! I'm really disappointed all I have a **charger** that doesn't work.
+	This is an excellent novel, worth to **read**.	They were comfortable, affordable, **lightweight**, **durable**, and had good **sound** quality.
+	It's a very good book, with beautiful pictures.	**Easy** to use, nice (in appearance).

Boldfaced text indicates domain-specific words, which are much more frequent in one domain than in the other. Underlined words are domain-shared words, which are frequently used in both domains and have the same sentiment polarity

and it is calculated as follows:

$$\text{support(item)} = P(\text{item}). \tag{4}$$

The confidence is defined as the conditional probability of item_{sp} and item_{sh}, calculated as follows:

$$\text{confidence}(\text{item}_{\text{sh}} \Rightarrow \text{item}_{\text{sp}}) = P(\text{item}_{\text{sp}} \mid \text{item}_{\text{sh}}). \tag{5}$$

In this study, to have a better understanding of the relationships among words based on learned strong association rules, we construct a weighted directed graph $G = \{V_{\text{item}_{\text{sh}}} \cup V_{\text{item}_{\text{sp}}}, E\}$ to find the relationships between the domain-specific words in D_{s} and D_{t} (Fig. 2). The directed graph G contains the union of domain-shared vertex set $V_{\text{item}_{\text{sh}}}$ and domain-specific vertex set $V_{\text{item}_{\text{sp}}}$, where each vertex is simply a word from the corresponding set. In this study, we use the confidence value of rules to denote the correlation between item_{sh} and item_{sp}. The weight of an edge that connects item_{sh} and item_{sp} in the graph expresses the correlation between item_{sh} and item_{sp}. The specific computing formula of the relativity is shown in Eq. (5). To visualize the extent of correlation, we use the distance along the vertical direction to represent the weight. In Fig. 2, the set of domain-shared words is denoted as $V_{\text{item}_{\text{sh}}} = \{\text{SH}^1, \text{SH}^2\}$, and the set of domain-specific words is denoted as $V_{\text{item}_{\text{sp}}} = V_{\text{sp}_{\text{s}}} \cup V_{\text{sp}_{\text{t}}}$, where $V_{\text{sp}_{\text{s}}} = \{\text{SP}_{\text{s}}^1, \text{SP}_{\text{s}}^2, \text{SP}_{\text{s}}^3, \text{SP}_{\text{s}}^4, \text{SP}_{\text{s}}^5\}$ and $V_{\text{sp}_{\text{t}}} = \{\text{SP}_{\text{t}}^6, \text{SP}_{\text{t}}^7, \text{SP}_{\text{t}}^8, \text{SP}_{\text{t}}^9\}$ are the sets of source and target domain-specific words, respectively. We also find that the correlation between SH^1 and SP_{s}^1 is smaller than the correlation between SH^1 and SP_{s}^4.

4.3 Aligning domain-specific words

According to the graph spectral theory (Chung, 1997), two vertices would be considered similar or have strong correlation if they are connected by multiple common vertices. In this study, we employ this assumption to realize the domain-specific words alignment; i.e., if two domain-specific words are related with multiple domain-shared words, the two domain-specific words are considered to be correlated.

Using the vertices in Fig. 2, the basic flow of aligning the domain-specific words in different domains is described as follows:

1. We find that there are some vertices in Fig. 2

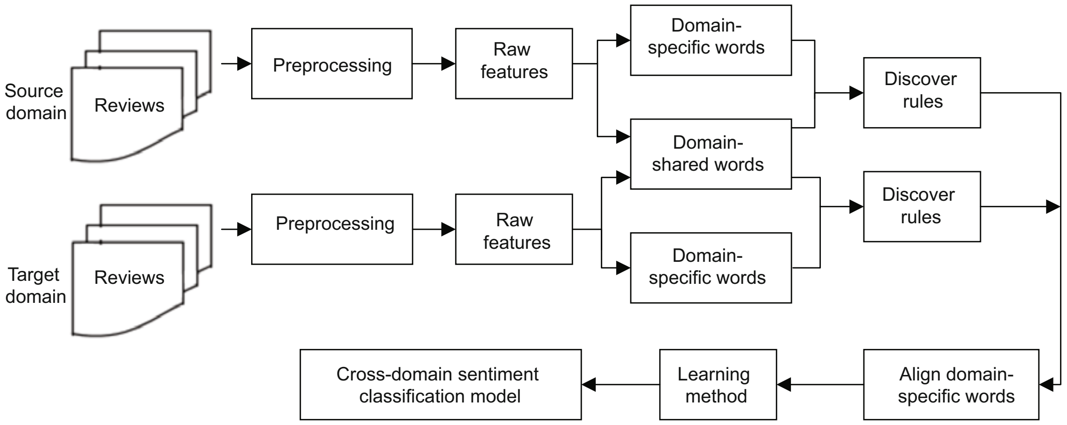

Fig. 1 Overview of the proposed method

Algorithm 1 Apriori algorithm

```
Input: reviews of $D_\mathrm{s}$ and $D_\mathrm{t}$; domain-shared words set $V_{\mathrm{item}_\mathrm{sh}}$; domain-specific words set $V_{\mathrm{item}_\mathrm{sp}}$; min-support and min-confidence.
1. Find the frequent 1 itemset based on the min-support:
       $\{\mathrm{item}_i^1\}_{i=1}^{n_1} = \{\mathrm{item}_1^1, \mathrm{item}_2^1, \ldots, \mathrm{item}_{n_1}^1\}$,
   where $n_1$ is the number of items in 1 itemset, and $\mathrm{item}_i^1$ is an item in $V_{\mathrm{item}_\mathrm{sh}}$ or $V_{\mathrm{item}_\mathrm{sp}}$. The frequency of items included in $\{\mathrm{item}_i^1\}_{i=1}^{n_1}$ must satisfy the minimal support that is noted as min-support. The formula for support is shown in Eq. (4).
2. Find the frequent 2 itemsets with the help of the min-support and 1 itemset and delete the frequent 2 itemsets that contain only domain-shared words or domain-specific words:
       $\{\mathrm{item}_j^2\}_{j=1}^{n_2} = \{(\mathrm{item_{sh}}^1, \mathrm{item_{sp}}^1)_j\}_1^{n_2}$,
   where $n_2$ is the number of items in the 2 itemsets and $\mathrm{item}_j^2$ denotes the association rule between domain-shared word $\mathrm{item}_\mathrm{sh}^1$ and domain-specific word $\mathrm{item}_\mathrm{sp}^1$ in the same domain. The frequency of items included in $\{\mathrm{item}_j^2\}_{j=1}^{n_2}$ must satisfy the min-support. The formula of support is shown in Eq. (4).
3. Obtain the strong association rules between domain-shared and domain-specific words for two domains based on the min-confidence:
       $\mathrm{rule}_j = r\left((\mathrm{item_{sh}} \Rightarrow \mathrm{item_{sp}})_j\right)$,
   where the minimal confidence is noted as min-confidence, and the strong association rule is built when the conditional probability of $\mathrm{item_{sp}}$ and $\mathrm{item_{sh}}$ is larger than min-confidence. The formula of confidence is shown in Eq. (5).
4. Finally, the meaningful rule set learned by the Apriori algorithm is denoted as RS:
       $\mathrm{RS} = \{\mathrm{rule}_j\}_{j=1}^{n_\mathrm{r}}$,
   where $n_r$ is the number of rules in the set of strong association rules and $\mathrm{rule}_j$ is a rule for the $j^\mathrm{th}$ word pair:
       $(\mathrm{item_{sh}} \Rightarrow \mathrm{item_{sp}})_j$.
Output: strong association rules between domain-shared and domain-specific words are denoted as $r\,(\mathrm{item_{sh}} \Rightarrow \mathrm{item_{sp}})$.
```

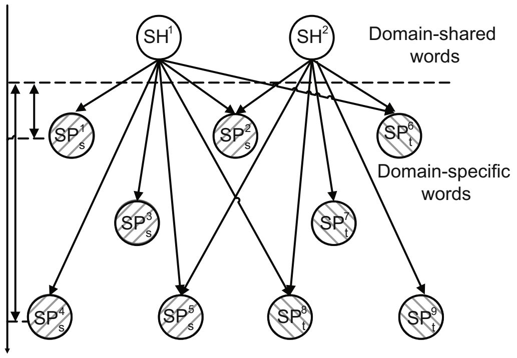

Fig. 2 Directed graph example of domain-shared and domain-specific words, where the blank circles and the shadow circles represent domain-shared and domain-specific words, respectively

The directed links between circles (one-way arrow) express the strong association between domain-shared and domain-specific words. The directed links in the vertical direction (two-way arrow) are used to visualize the extent of correlation

that have approximately equal vertical distances with domain-shared words SH^1 and SH^2, respectively. We assume that the values of vertical distance are similar when the difference of vertical distances from domain-specific words (i.e., sp_s^a) to domain-shared words (i.e., SH^i) is no more than an empirical parameter ε, i.e., $r\,(\mathrm{SH}^i \Rightarrow \mathrm{sp}_\mathrm{s}^a) \approx r\,(\mathrm{SH}^i \Rightarrow \mathrm{SP}_\mathrm{t}^b)$, if $|r\,(\mathrm{SH}^i \Rightarrow \mathrm{sp}_\mathrm{s}^a) - r\,(\mathrm{SH}^i \Rightarrow \mathrm{SP}_\mathrm{t}^b)| \leq \varepsilon$. We can find these similar relations in Fig. 2:

$$r(\mathrm{SH}^1 \Rightarrow \mathrm{SP}_\mathrm{s}^1) \approx r(\mathrm{SH}^1 \Rightarrow \mathrm{SP}_\mathrm{s}^2) \approx r(\mathrm{SH}^1 \Rightarrow \mathrm{SP}_\mathrm{t}^6), \tag{6}$$

$$r(\mathrm{SH}^1 \Rightarrow \mathrm{SP}_\mathrm{s}^4) \approx r(\mathrm{SH}^1 \Rightarrow \mathrm{SP}_\mathrm{s}^5) \approx r(\mathrm{SH}^1 \Rightarrow \mathrm{SP}_\mathrm{t}^8), \tag{7}$$

$$r(\mathrm{SH}^2 \Rightarrow \mathrm{SP}_\mathrm{s}^2) \approx r(\mathrm{SH}^2 \Rightarrow \mathrm{SP}_\mathrm{t}^6), \tag{8}$$

$$r(\mathrm{SH}^2 \Rightarrow \mathrm{SP}_\mathrm{s}^5) \approx r(\mathrm{SH}^2 \Rightarrow \mathrm{SP}_\mathrm{t}^8) \approx r(\mathrm{SH}^2 \Rightarrow \mathrm{SP}_\mathrm{t}^9), \tag{9}$$

where SP_s^a and SP_t^b are domain-specific words from D_s and D_t, respectively.

2. We need to identify word pairs. In the previous stage, we found some similar relations:

$$r(\mathrm{SH}^1 \Rightarrow \mathrm{SP}_\mathrm{s}^2) \approx r\,(\mathrm{SH}^1 \Rightarrow \mathrm{SP}_\mathrm{t}^6), \tag{10}$$

$$r(\mathrm{SH}^2 \Rightarrow \mathrm{SP}_\mathrm{s}^2) \approx r\,(\mathrm{SH}^2 \Rightarrow \mathrm{SP}_\mathrm{t}^6). \tag{11}$$

In these similar relations, both of the domain-specific words SP_s^2 and SP_t^6 have a direct mapping relationship with the domain-shared words SH^1 and SH^2. So, we conclude that couple $= \{\mathrm{SP}_\mathrm{s}^2, \mathrm{SP}_\mathrm{t}^6\}$ is a word pair. After repeating this operation, k word pairs $\{\mathrm{couple}_i\}_{i=1}^k$ are obtained.

3. We need to align the domain-specific words. We assume that each word pair obtained above belongs to a cluster and all words in a single cluster

have the same features. In this way, most of the domain-specific words from D_{s} and D_{t} are clustered in corresponding clusters, which will create an indirect mapping relationship between domain-specific words from different domains. So, for each sentiment sentence $\boldsymbol{X}_i$, where $\boldsymbol{X}_i \in \{D_{\mathrm{s}}, D_{\mathrm{t}}\}$, the new representation is described as

$$\widetilde{\boldsymbol{X}_i} = [\mathrm{SH}^1, \ldots, \mathrm{SH}^l, \mathrm{couple}_1, \ldots, \mathrm{couple}_k], \qquad (12)$$

where $\boldsymbol{X}_i \in \mathbb{R}^{1\times m}$ and $\widetilde{\boldsymbol{X}_i} \in \mathbb{R}^{1\times(l+k)}$. The feature dimension of $\widetilde{\boldsymbol{X}_i}$ is $l+k$. Each of the feature dimensions is a binary value. A word from $\boldsymbol{X}_i$ that appears in $\widetilde{\boldsymbol{X}_i}$ has a feature value 1; otherwise, the feature value is 0.

4.4 Algorithm summary and time complexity analysis

The whole WAAR approach process for cross-domain sentiment classification is summarized and presented in Algorithm 2.

The time complexity of our approach depends mainly on the computation of aligning the domain-specific words in two domains, including the selection of domain-shared words, the learning of strong association rules, and the establishment of an indirect mapping relationship. The computation is based on the number of domain-shared words l, the number of 1 itemset n_1, and the numbers of domain-specific words in two domains m_{s} and m_{t}. The WAAR time complexity is $O(n^2 + m_{\mathrm{s}} + l \cdot m_{\mathrm{t}} + m_{\mathrm{s}} \cdot m_{\mathrm{t}}) \approx O(n^2)$. The time complexity of the SFA algorithm (Pan et al., 2010) is $O(n^3)$ (Ding et al., 2014), which depends mainly on the spectral clustering on a bipartite graph. So, we can conclude that our approach is superior to SFA in terms of time cost and has better performance in a big data stream than SFA.

Algorithm 2 Words alignment based on association rules for cross-domain sentiment classification

```
Input: labeled source domain data $D_s = \{(X_{s_i}, Y_{s_i})\}_{i=1}^{n_s}$; unlabeled target domain data
    $D_t = \{(X_{t_i})\}_{i=1}^{n_t}$; number of domain-shared words $l$; number of word pairs $k$;
    min-support and min-confidence.
1. According to the procedure in Section 4.1 and Eq. (3), select $l$ domain-shared words from sets $D_s$ and $D_t$.
    The remaining $m-l$ words are deemed domain-specific words.
2. Discover the rules set RS using the Apriori algorithm in Section 4.2:
        $\mathrm{RS} = \{\mathrm{rule}_j\}_{j=1}^{n_r}$,
    where $n_r$ is the size of RS and rule = ($\mathrm{item_{sh}} \Rightarrow \mathrm{item_{SP}}$).
3. Construct a directed graph $G$ based on the learned rules in set RS. In $G$, vertices denote domain-shared
    words and domain-specific words, edges express the correlation, and the relative weight $r$ is equal to the
    confidence value of the rule.
4. Use the similarity of mined rules to find $k$ word pairs of domain-specific words from $D_s$ and $D_t$. The
    standard to distinguish is $|r\,(\mathrm{SH}_i \Rightarrow \mathrm{SP}_s^a) - r\,(\mathrm{SH}_i \Rightarrow \mathrm{SP}_t^b)| \le \varepsilon$, where $i \in (0, l)$ and $\varepsilon$ is an
    empirical value. Then the domain-specific words $\mathrm{SP}_s^a$ and $\mathrm{SP}_t^b$ can form a word pair.
5. The new feature representation of $X_i$ is described as $\widetilde{X_i} = [\mathrm{SH}^1, \ldots, \mathrm{SH}^l, \mathrm{couple}_1, \ldots, \mathrm{couple}_k]$,
    where $\widetilde{X_i} \in \mathbb{R}^{1\times(l+k)}$, and the feature value of dimension is 1 or 0.
6. Return the cross-domain sentiment classifier support vector machine trained on $\{(\widetilde{X}_{s_i}, Y_{s_i})\}_{i=1}^{n_s}$.
Output: cross-domain sentiment classifier $f\colon X \to Y$.
```

5 Experiment and analysis

In this section, we will introduce our experiments and show the effectiveness of our WAAR approach for cross-domain sentiment classification.

5.1 Datasets

We evaluate our approach on the benchmark datasets of Blitzer et al. (2007), which include customer reviews of Amazon® products. The datasets include four product domains: Books (B), DvDs (D), Electronics (E), and Kitchen (K). Each review has been assigned a sentiment label, +1 (positive review) or −1 (negative review), based on the rating scores given by the reviewers. Each domain has 1000 positive reviews, 1000 negative reviews, and thousands of unlabeled reviews. The details of the datasets are shown in Table 2. In these datasets, we will make full use of reviews in the four domains to create 12 cross-domain sentiment classification tasks to test the performance of our approach: D→B, E→B, K→B, B→D, E→D, K→D, B→E, D→E, K→E, B→K, D→K, E→K, where the domain at the left of an arrow represents the source domain and the domain at the right of an arrow represents the target domain.

In this study, each review is expressed as a

Table 2 Detailed description of Amazon® datasets used for experiments

Domain	Number of reviews			
	Positive	Negative	Non-labeled	All
Books	1000	1000	4465	6465
DvDs	1000	1000	3586	5586
Electronics	1000	1000	5681	7681
Kitchen	1000	1000	5945	7945

feature set that is composed of single words or bigrams and converted into a vector that is 0 or 1. The detailed preprocessing corresponds to the settings in Blitzer et al. (2007) and Pan et al. (2010). To reduce the computational cost, we filter out the stop words and the words that appear fewer than three times in both domains. We select only the top 5000 words that have higher information gain (Yang and Pedersen, 1997) in the vocabulary set V as the representation of input vectors.

5.2 Procedure

To verify the effectiveness of our proposed approach, we make the comparison with several reference algorithms:

1. NoTransf (Pan et al., 2010), a classifier, such as support vector machine, which is directly trained with data in D_s and tested on data in D_t.

2. SCL (Blitzer et al., 2007), which adopts structural correspondence learning for sentiment classification.

3. SFA (Pan et al., 2010), which aligns words from source domains and target domains to bridge the gaps between them. Our approach is largely based on this algorithm.

For all the above cross-domain sentiment classification tasks, each classifier is trained with all available labeled data in the source domain and tested on all the data in the target domain. For example, in the task of D→B, the classifier is trained with 2000 labeled reviews in the D domain and tested on 2000 reviews in the B domain. We initialize parameters in our experiments as follows: $l = 600$, min-support $=$ 0.014, min-confidence $= 0.08$, and $\varepsilon = 0.005$. In this study, we use the transfer ratio to calculate the mean transfer error in all tasks to evaluate the performance of the cross-domain sentiment classifier. The transfer ratio $\rho(D_\mathrm{s}, D_\mathrm{t})$ and transfer loss $t(D_\mathrm{s}, D_\mathrm{t})$ were mentioned in Glorot et al. (2011):

$$\rho(D_\mathrm{s}, D_\mathrm{t}) = \frac{1}{n} \sum_{(D_\mathrm{s}, D_\mathrm{t})_{D_\mathrm{s} \neq D_\mathrm{t}}} \frac{e(D_\mathrm{s}, D_\mathrm{t})}{e_\mathrm{b}(D_\mathrm{t}, D_\mathrm{t})}, \tag{13}$$

$$t(D_\mathrm{s}, D_\mathrm{t}) = e(D_\mathrm{s}, D_\mathrm{t}) - e_\mathrm{b}(D_\mathrm{t}, D_\mathrm{t}), \tag{14}$$

where $n = 12$ is the number of cross-domain sentiment classification tasks. The definitions of $e(D_\mathrm{s}, D_\mathrm{t})$ and $e_\mathrm{b}(D_\mathrm{t}, D_\mathrm{t})$ were mentioned by Glorot et al. (2011).

5.3 Results

Table 3, Figs. 3, and 4 show a comparison of our approach with four reference algorithms on the accuracy of cross-domain sentiment classification, transfer loss, and transfer ratio, respectively.

Table 3 shows that the NoTransf method is less accurate than the other algorithms in all tasks. This implies that knowledge transfer between domains can improve the accuracy of cross-domain sentiment recognition. The accuracy of WAAR and SFA is higher than SCL in 11 tasks, except the task of B→E. This implies that the method of aligning domain-specific words with the help of domain-shared words (also called 'domain-independent words' in Pan et al. (2010)) can reduce the difference between domains. Compared with SFA, WAAR performs poorer in tasks E→B, B→K, D→K, and E→K, because there are some domain-specific words that cannot be perfectly aligned in domains using an association rule algorithm when compared with SFA. In these cases, SFA can cause more informative models to align domain-specific words by feature extraction. However, we also find that the accuracy of WAAR in two tasks (K→B and K→E) is higher than SFA and SCL by only 0.1% and 0.01%, respectively. This means that the accuracy of our approach is equal to those of SFA and SCL in some cases when there is less difference between domains, depending on the domain difference. As compared with SFA and SCL, WAAR wins in seven tasks, and the accuracy is improved by 0.01%–2.06%, with 0.7% on average. We can see that an association rule algorithm such as the Apriori algorithm can also be used to learn the indirect mapping relationship between domain-specific words in different domains via domain-shared words. However, the improved accuracy of the cross-domain sentiment classification of WAAR is lower in some domains (K→E and K→B). A possible reason is that

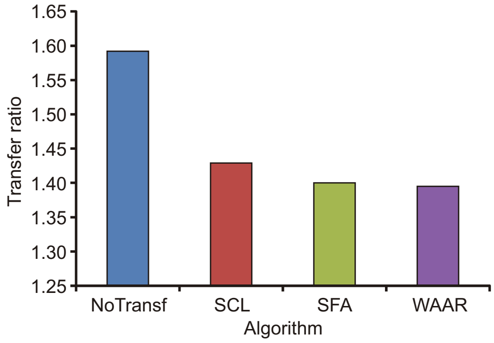

Fig. 4 Transfer ratio of algorithms on Amazon® datasets

there are some words whose sentiment orientations are different in different domains; for example, 'easy' is a positive word in the Kitchen domain and a negative word in the Books domain, as shown in Table 1. So, the accuracy of a cross-domain sentiment classifier decreases when words with sentiment orientation divergence in different domains are selected as domain-shared words.

Fig. 3 shows the transfer loss of all cross-domain sentiment classification tasks using four methods. As shown in Fig. 3, the best transfer is achieved by our approach in 8 of 12 tasks; SFA has better transfer loss when Electronics is the source domain. We also find that one task has a negative transfer loss for SCL, SFA, and WAAR methods when the source domain is Kitchen and the target domain is Electronics, which means that a classifier trained with a different domain can outperform a classifier trained with the target domain.

Fig. 4 shows the transfer ratio of four methods on Amazon® datasets. We find that our approach has a lower transfer ratio than other methods, so we can conclude that the domain-specific words in different domains can be aligned by our approach to some extent. Therefore, we conclude that making full use of the relationship of domain-shared words and domain-specific words can reduce the difference between domains and improve the performance of cross-domain sentiment classification.

Table 3 Accuracy of algorithms on 12 cross-domain sentiment classification tasks using the the Amazon® datasets

Task	Accuracy (%)			
	NoTransf	SCL	SFA	WAAR
B→D	76.80	78.50	80.54	**81.25**
E→D	71.25	75.25	75.50	**75.80**
K→D	73.05	76.65	76.70	**76.90**
D→B	73.35	78.26	77.54	**79.60**
E→B	72.42	75.02	**75.40**	73.50
K→B	71.80	72.78	74.20	**74.30**
B→E	71.28	**75.22**	72.10	73.50
D→E	72.69	74.20	76.04	**77.60**
K→E	83.02	85.04	85.02	**85.05**
B→K	74.45	77.08	**78.02**	76.90
D→K	75.02	78.94	**79.50**	77.85
E→K	85.02	85.06	**85.95**	85.03

Best results are shown in bold

5.4 Parameter sensitivity exploration

As described in Algorithm 2, there are four empirical parameters in our algorithm, which are the number of shared words l, the minimum support min-support, the minimum confidence min-

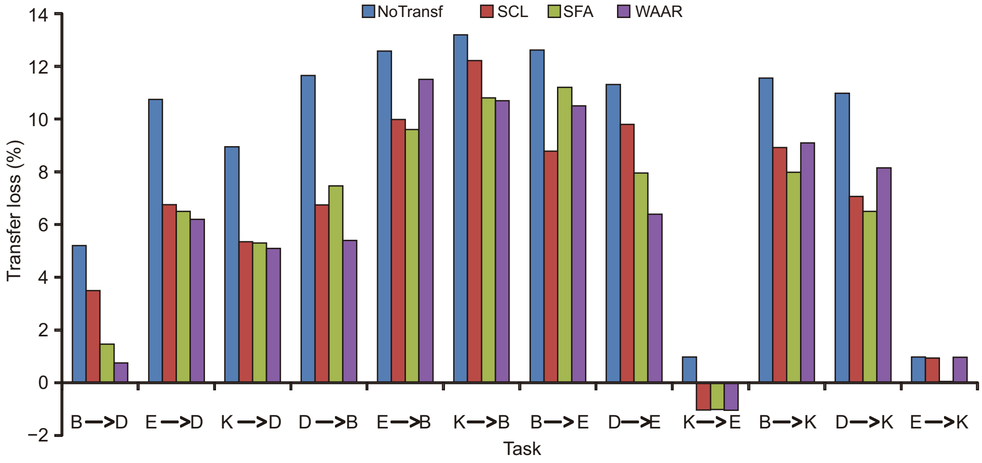

Fig. 3 Transfer loss of algorithms on Amazon® datasets for 12 cross-domain sentiment classification tasks

References to color refer to the online version of this figure. B: Books; D: DvDs; E: Electronics; K: Kitchen confidence, and the threshold ε for estimating the correlation between two words.

In this subsection, we discuss how the four parameters influence the accuracy of the cross-domain sentiment classifier based on all cross-domain sentiment classification tasks on WAAR. We will fix the values of the other parameters when testing the influence of one parameter on the accuracy of cross-domain sentiment classification.

In the first experiment, we test the accuracy of WAAR on the number of domain-shared words parameter l. In this experiment, we set $s = 0.014$, $c = 0.08$, and $\varepsilon = 0.005$. The values of l vary from 400 to 1100 (Figs. 5a and 5b). We can see that the ideal l is approximately located in the range $[500, 700]$ and in this range, the accuracy of the algorithm is stable.

In the second experiment, we test the accuracy of WAAR on the minimum support parameter min-support. In this experiment, we set $l = 600$, $c = 0.08$, and $\varepsilon = 0.005$. The values of min-support vary from 0.002 to 0.02 (Figs. 5c and 5d). We can see that the ideal min-support is approximately located in the range $[0.008, 0.016]$ and in this range, the accuracy of the algorithm is stable.

In the third experiment, we test the accuracy of WAAR on the minimum confidence parameter min-confidence. In this experiment, we set $l = 600$, $s = 0.014$, and $\varepsilon = 0.005$. The values of min-confidence are varied from 0.02 to 0.2 (Figs. 5e and 5f). We can see that the ideal min-confidence is approximately located at the range $[0.06, 0.12]$ and in this range, the accuracy of the algorithm is stable.

In the last experiment, we test the accuracy of WAAR on the threshold parameter ε. In this experiment, we set $l = 600$, $s = 0.014$, and $c = 0.08$. The values of ε vary from 0.001 to 0.01 (Figs. 5g and 5h). We can see that the ideal ε is approximately located in the range $[0.003, 0.005]$ and in this range, the accuracy of the algorithm is stable.

6 Conclusions

In this study, we propose a novel approach to align domain-specific words in different domains for cross-domain sentiment classification, which is called words alignment based on association rules (WAAR). In our approach, we first choose domain-shared words and domain-specific words in two domains. We then establish a direct mapping relationship between domain-shared words and domain-specific words in the same domain using an association rule algorithm such as the Apriori algorithm. Finally, we establish an indirect mapping relationship between domain-specific words in different domains to achieve alignment of domain-specific words from the source domain to the target domain, with the help of domain-shared words. In this way, the differences of two domains can be reduced to some extent, which is helpful for training an accurate cross-domain sentiment classifier. By analyzing experimental results and time complexity, the effectiveness of our approach is verified.

From the results of our experiments, we also find that the improved accuracy of our approach for cross-domain sentiment classification is lower in some domains, possibly because the sentiment orientations of some selected domain-shared words are quite different in different domains. In the future, we plan to solve the problem of sentiment orientation divergence in domain-shared words. In addition, we plan to align domain-specific words in different domains based on graph spectral theory to improve the accuracy of the cross-domain sentiment classifier. Finally, we plan to use our approach to solve sentiment classification problems from multiple source domains.

References

Agrawal R, Srikant R, 1994. Fast algorithms for mining association rules. 20th Int Conf on Very Large Data Bases, p.487-499.

Ando R, Zhang T, 2005. A framework for learning predictive structures from multiple tasks and unlabeled data. *J Mach Learn Res*, 6:1817-1853.

Balazs J, Velasquez J, 2016. Opinion mining and information fusion: a survey. *Inform Fus*, 27(C):95-110. https://doi.org/10.1016/j.inffus.2015.06.002

Blitzer J, Dredze M, Pereira F, 2007. Biographies, bollywood, boom-boxes, and blenders: domain adaptation for sentiment classification. 45th Annual Meeting of the Association of Computational Linguistics, p.440-447. https://doi.org/10.1109/IRPS.2011.5784441

Bollegala D, Weir D, Carroll J, 2013. Cross-domain sentiment classification using a sentiment sensitive thesaurus. *IEEE Trans Knowl Data Eng*, 25(8):1719-1731. https://doi.org/10.1109/TKDE.2012.103

Chen M, Xu Z, Kilian Q, et al., 2012. Marginalized denoising autoencoders for domain adaptation. https://arxiv.org/abs/1206.4683

Chung FRK, 1997. Spectral Graph Theory. Co-publication of the AMS and CBMS.

Dave K, Lawrence S, Pennock D, 2003. Mining the peanut gallery: opinion extraction and semantic classification

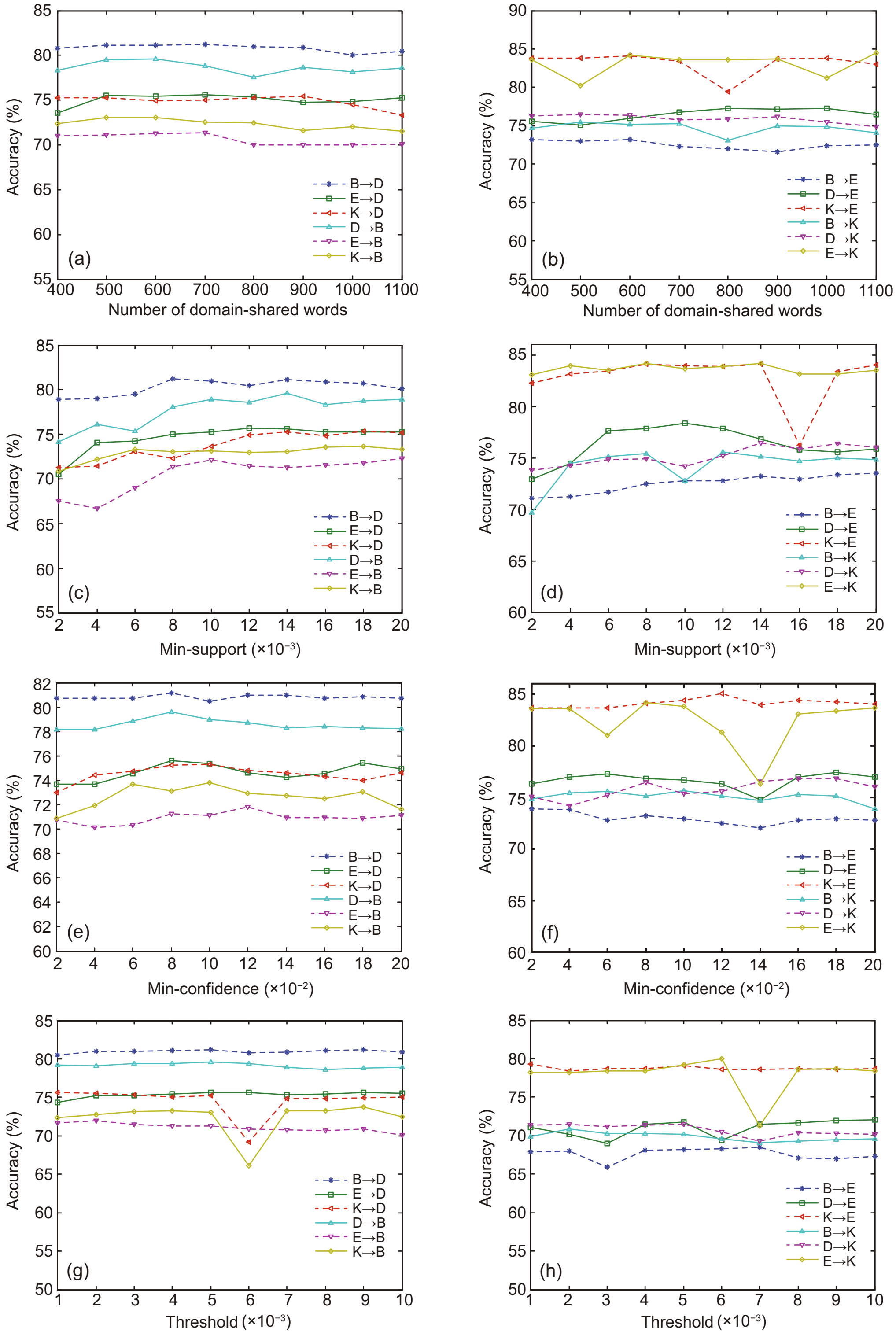

Fig. 5 Accuracy of words alignment based on association rules (WAAR) under four varying parameters of all cross-domain sentiment classification tasks: (a) and (b) accuracy for different values of l on 12 subtasks; (c) and (d) accuracy for different values of min-support on 12 subtasks; (e) and (f) accuracy for different values of min-confidence on 12 subtasks; (g) and (h) accuracy for different values of ε on 12 subtasks

of product reviews. 12th Int Conf on World Wide Web, p.519-528. https://doi.org/10.1145/775152.775226

Ding S, Jia H, Shi Z, 2014. Spectral clustering algorithm based on adaptive Nystrom sampling for big data analysis. *J Softw*, 25(9):2037-2049.

Glorot X, Bordes A, Bengio Y, 2011. Domain adaptation for large-scale sentiment classification: a deep learning approach. 28th Int Conf on Machine Learning, p.513-520.

Goldberg A, Zhu X, 2006. Seeing stars when there are not many stars: graph-based semi-supervised learning for sentiment categorization. 1st Workshop on Graph Based Methods for Natural Language Processing, p.45-52. https://doi.org/10.3115/1654758.1654769

Jiang J, Zhai C, 2007. Instance weighting for domain adaptation in NLP. 45th Annual Meeting of the Association of Computational Linguistics, p.264-271. https://doi.org/10.1145/1273496.1273558

Li L, Jin X, Long M, 2012. Topic correlation analysis for cross-domain text classification. AAAI Conf on Artificial Intelligence, p.998-1004.

Li T, Zhang Y, Sindhwani V, 2009. A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge. Joint Conf of the 47th Annual Meeting of the ACL and the 4th Int Joint Conf on Natural Language Processing, p.244-252. https://doi.org/10.3115/1687878.1687914

Pan S, Ni X, Sun J, et al., 2010. Cross-domain sentiment classification via spectral feature alignment. 19th Int Conf on World Wide Web, p.751-760. https://doi.org/10.1145/1772690.1772767

Pang B, Lee L, Vaithyanathan S, 2002. Thumbs up? Sentiment classification using machine learning techniques. ACL-02 Conf on Empirical Methods in Natural Language Processing, p.79-86. https://doi.org/10.3115/1118693.1118704

Pantelis A, Ioannis K, Ioannis K, et al., 2018. Learning patterns for discovering domain oriented opinion words. *Knowl Inform Syst*, 55(1):45-77. https://doi.org/10.1007/s10115-017-1072-y

Schölkopf B, Platt J, Hofmann T, 2007. Correcting sample selection bias by unlabeled data. Advances in Neural Information Processing Systems, p.601-608.

Turney PD, 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. 40th Annual Meeting on Association for Computational Linguistics, p.417-424. https://doi.org/10.3115/1073083.1073153

Wu Q, Tan S, Xu H, et al., 2010a. Cross-domain opinion analysis based on random-walk model. *J Comput Res Dev*, 47(12):2123-2131.

Wu Q, Tan S, Zhang G, et al., 2010b. Research on cross-domain opinion analysis. *J Chin Inform Process*, 24(1):77-83.

Yang Y, Pedersen J, 1997. A comparative study on feature selection in text categorization. 14th Int Conf on Machine Learning, p.412-420.

Zadrozny B, 2004. Learning and evaluating classifiers under sample selection bias. 21st Int Conf on Machine Learning, p.114-121. https://doi.org/10.1145/1015330.1015425

Zhou G, Zhou Y, Guo X, et al., 2015. Cross-domain sentiment classification via topical correspondence transfer. *Neurocomputing*, 159:298-305. https://doi.org/10.1016/j.neucom.2014.12.006

Zhuang L, Jing F, Zhu X, 2006. Movie review mining and summarization. 15th ACM Int Conf on Information and Knowledge Management, p.43-50. https://doi.org/10.1145/1183614.1183625