



Original Article

Entropy based classifier for cross-domain opinion mining

Jyoti S. Deshmukh^a, Amiya Kumar Tripathy^{b,*}^a Department of Computer Engineering, PAHER University, Udaipur, India^b Department of Computer Engineering, Don Bosco Institute of Technology, Mumbai, India

ARTICLE INFO

Article history:

Received 30 August 2016

Revised 11 February 2017

Accepted 20 March 2017

Available online 22 March 2017

Keywords:

Data mining

Opinion mining

Knowledge discovery

Expert systems

Information systems

Machine learning

ABSTRACT

In recent years, the growth of social network has increased the interest of people in analyzing reviews and opinions for products before they buy them. Consequently, this has given rise to the domain adaptation as a prominent area of research in sentiment analysis. A classifier trained from one domain often gives poor results on data from another domain. Expression of sentiment is different in every domain. The labeling cost of each domain separately is very high as well as time consuming. Therefore, this study has proposed an approach that extracts and classifies opinion words from one domain called source domain and predicts opinion words of another domain called target domain using a semi-supervised approach, which combines modified maximum entropy and bipartite graph clustering. A comparison of opinion classification on reviews on four different product domains is presented. The results demonstrate that the proposed method performs relatively well in comparison to the other methods. Comparison of SentiWordNet of domain-specific and domain-independent words reveals that on an average 72.6% and 88.4% words, respectively, are correctly classified.

© 2017 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Opinionated text has created a new area of research in text analysis. Traditionally, fact and information-centric view of text was expanded to enable sentiment-aware applications. Nowadays, increased use of the Internet and online activities like ticket booking, online transactions, e-commerce, social media communications, blogging, etc. has led to the need for the extraction, transformation and analysis of huge amount of information. Therefore, new approaches need to be applied to analyze and summarize the information [14].

Organizations take the review of product given by users seriously, as it adversely affects the sales of the product. Consequently, organizations take the effort to respond to the reviews, as well as monitor the effectiveness of its advertising campaigns. In this regard, sentiment analysis, a popular method, is used to extract and analyze sentiments [5,4].

Opinion mining is constantly growing due to the availability of views, opinions and experiences about a product/service online, as people are shedding their inhibition to express their opinions online. However, automatic detection and analysis of opinions about products, brands, political issues, etc. is a daunting task. Opinion mining involves three chief elements: feature and feature-of relations, opinion expressions and the related opinion attributes (e.g., polarity), and feature-opinion relations. An opinion lexicon is a list of opinion expressions or a set of adjectives, which are used to indicate opinion/sentiment polarity like positive, negative and neutral. This lexicon arises from synonyms in the WordNet, while antonyms are used to expand lexicon in the form of graphs. Such a dictionary-based approach has been used to partially disambiguate the results of parts of speech tagger. Further, fuzzy logic is used to determine opinion boundaries and to adopt syntactic parsing to learn and infer propagation rules between opinions and features [24,13].

Medhat et al. [18] conducted a survey on sentiment algorithms and its applications and found that sentiment classification and feature selection are more prominent areas in recent research. They also reported that Support vector machine and Naïve Bayes algorithms are the generally used algorithms to classify sentiments, and English is the language used in many resources like WordNet. Opinions and reviews given on social networking sites are used to generate datasets for the experiments.

The WordNet is a generalized lexicon and cannot be used for sentiment analysis; therefore, a need arose for the development

* Corresponding author at: School of Science, Edith Cowan University, Perth, Australia.

E-mail addresses: jyoja2007@gmail.com (J.S. Deshmukh), amiya@dbit.in (A.K. Tripathy).

Peer review under responsibility of King Saud University.



Production and hosting by Elsevier

of sentiment lexicon. SentiWordNet evolved out of WordNet was created as a lexical resource for opinion mining. It assigns to each synset of WordNet three sentiment scores: positive, negative and neutral [19,11].

Manufacturers, as well as consumers, require opinion mining tools to collect opinions about a certain product. The opinion analysis tools can be used by manufacturers to decide a marketing strategy for estimating production rate. On the other hand, consumers can use these tools to make decision on buying a new product or take a trip to vacation locations, or select hotel, etc.

Labeled opinions are used to analyze the classifier. Practically, labeled opinions for every domain is not possible, as it delimited by time and cost, while domain adaptation or transfer learning could be used to circumvent this limitation. In this paper, we propose the approach of domain adaptable lexicon which predicts the polarity of lexicon of one domain using a set of labeled lexicon of another domain using a modified entropy algorithm. This algorithm uses enhanced entropy with modified increment quantity instead of traditional entropy algorithm. Dataset of different types of products containing textual reviews has been used for evaluation. Multiple experiments were carried out to analyze the algorithm using accuracy and F-measure. We designed the approach in two phases: (i) preprocessing of dataset and (ii) applying classifier and clustering on dataset.

The rest of the paper is structured as follows. In Section 2, we describe the related work on domain adaptation approaches. In Section 3, we introduce our new improved entropy based semi-supervised approach. In Section 4, we evaluate our approach using cross-domain sentiment classification tasks, and compare it with other baseline methods. Finally, in Section 5 we draw conclusions on the proposed approach and set directions for future work.

2. Related work

The text documents containing opinions or sentiments were classified based on their polarity, i.e. whether a document is written with a positive approach or a negative approach. Although machine learning approach uses a word's polarity as a feature, the polarity of some words cannot be determined without domain knowledge. Hence, the reusability of learned result of a domain is essential. Transfer learning, also known as domain adaptation, can be used to address this challenge. Transfer learning utilizes the results learned in a source domain to solve a similar problem in another target domain [22]. Approaches used to classify single and cross-domain polarity opinions are usually a bag of words, n-grams or lexical resource-based classifiers.

The main aim of domain adaptation is to transfer knowledge across domains or tasks. Tagging the opinion word and building a classifier is time consuming and expensive, as opinions are domain dependent. Normally, users express their opinions specific to a particular domain. An opinion classifier trained in one domain may not work well when directly applied to another domain due to mismatch between domain-specific words. Thus, domain adaptation algorithms are extremely desirable to reduce domain dependency and labeling costs. Sentiment classification problem are considered as a feature expansion problem, in which related features are appended to reduce mismatch of features between the two domains. To overcome this problem, sentiment-sensitive thesaurus, which contains different words and their orientation in different domains, has been created. Bollegala et al. [7] used labeled, as well as unlabeled data, for evaluation. The results suggested that method performs significantly well compared to baseline.

To overcome domain adaptation issue, various adaptation methods have been proposed in the past, e.g., ensemble of classifiers. Combination of various feature sets and classification tech-

niques yielded in the ensemble framework was proposed by Xia et al. [26]. They used two types of feature sets, namely, Parts-of-speech information and Word-relations and Naïve Bayes, Maximum Entropy and Support Vector Machines classifiers. For better accuracy, ensemble approaches like fixed combination, weighted combination and Meta-classifier combination, were applied. Li et al. [29] proposed active learning in which source and target classifiers were trained separately. Using Query By Committee (QBC) selection strategy, informative samples were selected, and classification decision were made by combining classifiers. Label propagation was used to train both classifiers. The result demonstrated that significantly outperformed the baseline methods.

Most often, opinions are given in the natural language. One major issue with natural language is the ambiguity of words. Fersini et al. [10] applied Bayesian ensemble model in which uncertainty and reliability was taken care. Greedy approach was used for classifier selection, while gold standard datasets were used for experimental analysis. However, classification performance is frequently affected by the polarity shift problem. Polarity shifters are words and phrases that can change sentiment orientation of texts. Xia et al. [28] addressed this issue using three-stage models which include detection of polarity shift, removal of polarity shifts and sentiment classification. Onan et al. [2] proposed the weight based ensemble classifier, in which weighted voting scheme was used to assign weight to classifier. As a base learner Bayesian logistic regression, Naïve Bayes, linear discriminant analysis, logistic regression and Support vector machine are used. A different type of experimental analysis shows better result than conventional ensemble learning. Da Silva et al. [8] used classifier ensembles formed by different classifier which is applicable to find products on the web. Augustyniak et al. [16] demonstrated Twitter dataset to have good accuracy only for positive and negative queries. They found that Bag of Words with ensemble classifier performs better than supervised approach.

Identification of feature and weighting is an important step in opinion mining. Khan et al. [12] proposed a new approach that identified features and assigned term label using SentiWordNet. In this method, point wise mutual information and chi square approaches were used to select features to SentiWordNet that were weighted. Support vector machine was used as classifier. Experimental evaluation on benchmark dataset shows effectiveness of approach.

Social networking sites contains text data in long format as well as short messages with symbols, emoticons etc. Opinion detection in long reviews is easy than short reviews, as short reviews contain fewer features, and more symbols, idioms etc. hence difficult to extract opinion. Lochter et al. [15] proposed ensemble approach to tackle this issue. This approach used text normalization methods to improve the quality of features. The features thus filtered and enhanced served as the input for machine learning algorithms. Proposed framework was evaluated using real and non-coded datasets and concluded that this approach was superior to other methods with a 99.9% confidence level. However, this approach was suggested to be expensive for offline processes due to higher cost of computing power. Hence, parallelization of this process has been stated as future work by the authors.

Sparseness is another issue in short text data. Word co-occurrence and context information approaches are generally used for solving sparseness issue. These approaches are less efficient. To address this problem, Chutao et al. [32] considered probability distribution of terms as the weight of terms.

Similar to ensemble classifiers, graph-based methodology are also used for domain adaptation. Dhillon et al. [25] proposed the graph-based domain adaptation method. Similarity graphs were constructed between features from all domains, if these features were similar then it demonstrated the presence of edge between

them. All labeled features were used in metric-learning algorithms. Graph was constructed using data-dependent metric and the weight was calculated for each edge. Experimental results demonstrated the reduction of classification error.

Pan and Yang [22] and Wang and Shi [31] focused on bridging the gap between domain-specific and domain-independent lexicons, as these approaches do not work well when applied to two extremely different domains. Singh and Husain [30] presented different datasets used in sentiment analysis as well as classification and clustering methods. This review reveals that same method is not applicable for all domains. From the analysis of the literature, it can be summarized that Naïve Bayes works well for or text classification, clustering for consumer services, and SVM for biological review and analysis.

Training and testing data from same feature space and same distribution has been reported to give good results for machine learning algorithms. Estimating the effect of distribution changes through statistical models is reported to be very expensive, as it has to be rebuilt from scratch. In many real world applications, it is expensive or impractical to recollect the needed training data and rebuild the models. In such cases, domain adaptation or transfer learning between task domains would be desirable.

To overcome the problem of feature distribution variance across domains, Xiao and Guo [21] proposed a feature space independent semi-supervised kernel matching method, based on a Hilbert-Schmidt Independence Criterion. Two kernel matrices were created over the instances in the source domain and the instances in the target domain. Each labeled instance in the target domain was definitely mapped into a source instance with the same class label through prediction function. Evaluation of the proposed method performed on Amazon product reviews and Reuters' multilingual newswire stories showed reduction in human annotation efforts.

The Lexicon based approach works with the polarities of the opinion-oriented words and relies on a lexicon. A collection of known terms that contribute to the sentiment of a text is called sentiment lexicon. Many open source lexicons are available which serve as a database for extracting the polarity values of opinion words. But these generic polarity lexicons reflect the most generic sentiment of opinion words. An opinion word need not express the same sentiment everywhere, i.e., opinion words could be context-dependent or domain-specific. The word “small” in “room is too small” indicates a negative opinion, whereas in “small screen size” as seen in the mobile domain indicates a positive opinion.

The variation of opinion found for the same word in different domains restricts the usage of generic lexicons as it generalizes the polarity of a word. Therefore, lexicons with updated polarity values that can give polarity of a same word in different domains using same lexicon database will have to be built. The proposed work attempts in building such an enhanced polarity lexicon using the maximum entropy algorithm with modification being made in increment quantity which helps in refining the classification granularity from document to word level. The knowledge gained from one domain is used to predict and classify the polarity of opinion words from another domain, resulting in an improved lexicon using semi-supervised approach. The common words from all domains having same polarity orientation are treated as domain-independent words and remaining as domain-specific words.

3. Proposed framework

3.1. Generic processes

Most of the existing research regarding opinion mining is domain dependent, which limits the scope of the application as

well as the generalization of the information. Hence, generalized domain adaptable algorithms are needed for the automatic identification and classification of opinion lexicons.

Domain adaptability is a major issue in sentiment analysis or opinion mining, which has been addressed in the proposed framework. There are many resources and training corpora available in English with proven results. A proposed model will be trained from a training dataset, which will be used for sentiment classification. SentiWordNet resource will be used for this research as it is a publicly available for opinion lexicons with polarity.

An opinion lexicon is one or more words with positive or negative orientation. Lexicons are used when no training data are available because the training data contain prior knowledge about the sentiment of a feature. It is a vital component of unsupervised sentiment classification methods. The construction of a large sized lexicon is an expensive and time-consuming task. Hence, building automated methods that influence existing resources to expand existing lexicons are needed.

Domain adaptation of sentiment models from a domain with sufficient labeled data to a new domain with less labeled data is a challenge that requires new and efficient algorithms to solve it. The proposed system has constructed a domain adaptable lexicon which can adapt seamlessly from one domain to another. The expected outcome would be a set of lexicons with polarities for different domains with development of robust model.

Labeled set of documents from source and labeled or unlabeled documents from target domain is taken as input (Fig. 1). Preprocessing is done to eliminate unnecessary words called as stop words. The irrelevant data would be eliminated by this process.

Most of the English sentences include words like “a, an, of, the, I, it, you, etc.” Such words do not carry any particular meaning. Information extraction from natural language can be done effectively and clearly by avoiding those words which occur frequently. To

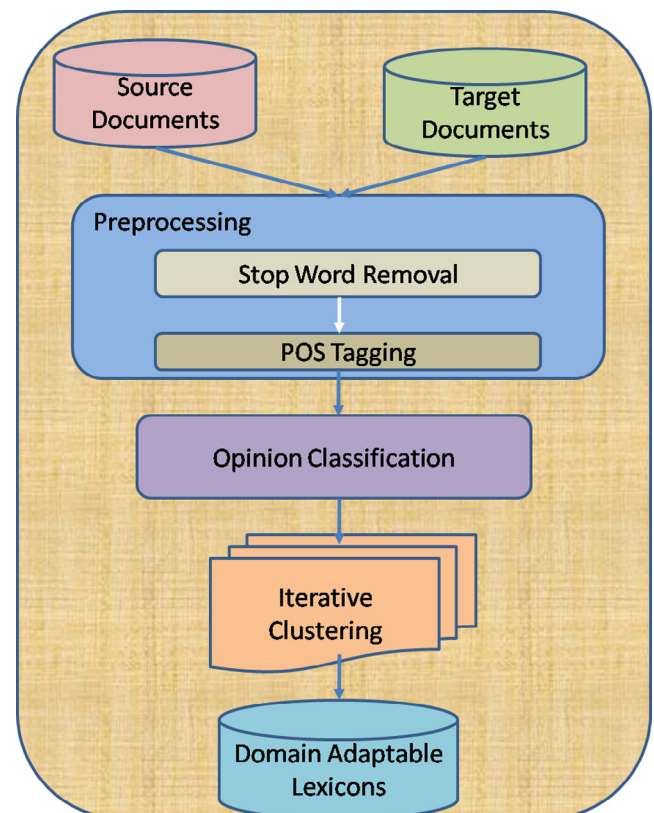


Fig. 1. Workflow of proposed system.

remove stop words from sentences, a text file that consists of list of English stop words is used.

After the removal stop word, parts of speech like noun, adjective, adverb, verb, etc. are extracted using the parser. Parsing is a vital step as it gives opinion words as an output. Sentence parsing involves assigning different parts of speech tags to a given text. This process is known as Part-Of-Speech (POS) tagging. For information extraction, POS tagging is important because each category plays a specific role within a sentence. Nouns give names to objects, or entities from reviews. An adjective describes opinion. Also, some verbs and adverbs can play an important role as an adjective.

Examples:.

the/DT battery/NN life/NN on/IN the/DT iphone/JJ 4S/CD is/VBZ
amazing/JJ
this/DT phone/NN is/VBZ very/RB slow/JJ

In pre-processing step, text review is first divided into sentences. Stanford parser is used to generate the POS tagging of each word present in the sentence [9], as it is essential to find general language patterns.

Adjectives and adverbs are good indicators of opinion, hence are extracted from each review. Some verbs are also considered as opinion, e.g., like, love, recommend, etc. Two consecutive words, i.e., adverb-verb, adverb-adjective also extracted from processed tagged reviews as a verb alone does not indicate opinion. Nouns are not considered in framework.

All tagged words after POS tagging phase tagged words are classified using an algorithm explained in Section 3.2.

3.2. Algorithm

Classification of opinions can be done using a modified maximum entropy algorithm. The increment quantity is modified according to the importance of the measure of words as specified in Eq. (3). The maximum entropy classifier is closely related to the Naïve Bayes classifier, except that it uses a search-based optimization to find weights for the features that maximize the likelihood of the training data. It can handle mixture of boolean, integer, and real-valued features [17]. It is also used when the conditional independence of the features cannot be assumed, i.e., in problems like text classification where features are words and are not independent [1].

The main aim of the study is to construct a stochastic model that accurately represents the behavior of the random process. Let d be document in a dataset; w , the word present in document; and c , the class.

1. For each word w and class $c \in C$, a joint feature $p(w, c) = f(w, c) = N$ is defined, where N is the number of times that w occurs in a document in class c . (N could also be boolean, registering presence vs. absence.)
2. Empirical distribution is used to build the statistical model of the random process, which distributes text to specific class.

$$f_i(d, c) = \int_0^1 \text{if } c = c_i \& d \text{ contains } w_k \text{ otherwise} \quad (1)$$

Above indicator function called as feature. Via iterative optimization, assign a weight to each joint feature so as to maximize the log-likelihood of the training data.

3. The probability of class c given a document d and weights λ is

$$P(c|d, \lambda) \stackrel{\text{def}}{=} \frac{\exp \sum_i \lambda_i f_i(c, d)}{\sum_{c' \in C} \exp \sum_i \lambda_i f_i(c', d)} \quad (2)$$

where $\lambda = \lambda_i + \delta_i$ (λ_i is calculated by an iterative scaling algorithm) δ_i is increment quantity

As the granularity of classification is refined from document-level to word level, the increment quantity (δ_i) is modified. The modified quantity (δm_i) is defined as

$$\delta m_i = \frac{1}{M} \log \left(\sum_i^k idf_i \right), \quad (3)$$

where $M = \max \sum_i^k f_i(d, c)$.

Eq. (3) calculates inverse document frequency of each word, which is a popular measure of words' importance. It is defined as the logarithmic ratio of the number of documents in a collection to the number of documents containing the given word. This suggests that uncommon words have higher idf_i and common function words have lower idf_i , where idf_i is inverse document frequency. This is useful to measure the words ability to discriminate between documents. M is the sum of all features in training instance. Feature value is taken as $tfidf_i$.

Algorithm works on word level. POS tagged words are extracted from preprocessing steps are used. Each word acts as feature. Feature value $f_i(d, c)$ is calculated as $tfidf_i$ of each word. As per Eq. (3) inverse document frequency of each POS tagged word is calculated. Algorithm is executed for total number of features provided in input dataset. According to this weight words are classified into two categories. Classified words are having POS tag, polarity tag, and weight value. From this list, common and uncommon words are picked and used for bipartite graph clustering explained below.

Probability distribution of class c is calculated based on term frequency. Classification process predicts the polarity of the target domains lexicon from source domain. The clustering algorithm is applied on classified word lists and documents until it reaches convergence. All extracted words from source domain are tagged, and weight is calculated for each word using mutual information available for words. Target words are extracted and compared with the source. If they match then they will be categorized as domain-independent, otherwise domain-specific. Domain-independent words are from both source and target domains; whereas, domain-specific are from target domain only. A graph is constructed between domain dependent and domain-independent words. Co-occurrence relationship between these words represents edge. Occurrence of domain-specific word along with domain-independent word means that both a related to each other and assign edge. Using domain-independent words weight is assigned to domain-specific words and classified accordingly. Each file from target domain is assigned score on which basis it is classified as positive or negative. Each word has weight assigned to it. Summation of weights of words in each sentence gives score to sentence. Then addition of all sentence score is nothing but score of file. On the basis of this, file is classified.

Clustering helps in reducing mismatch between domain-specific words of source and target domains. Two sets of lexicons are extracted as an output with polarity which is compared with SentiWordNet (Fig. 2).

4. Result and discussions

4.1. Experiment1

The dataset from John et al. [6] was used for experiments. It contains a collection of product reviews from Amazon.com. This dataset contains three types of files positive, negative and unlabeled in XML format. Each line in form of: feature:<count>... fea

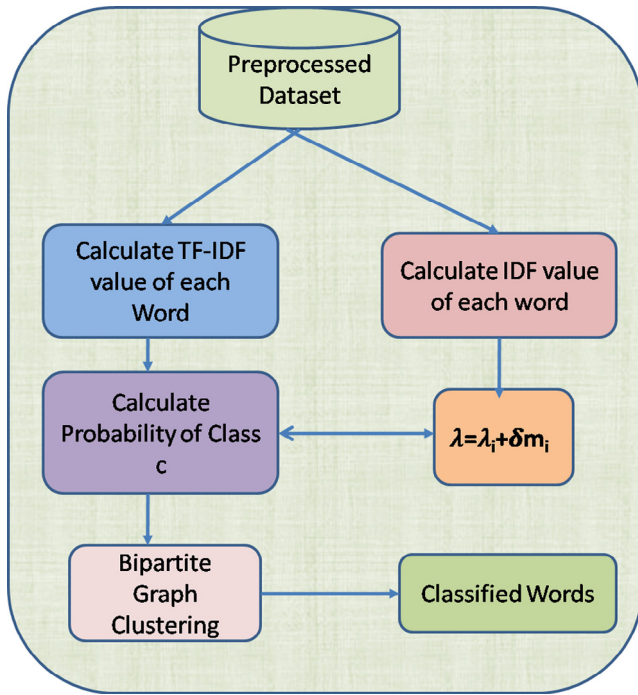


Fig. 2. Flow of proposed algorithm.

ture: <count>#label#: <label>, e.g., old_boy:1 i_am:1 the: 1 boy_had:1 so_i:1 #label#: negative. These files were extracted using XML file splitter and reviews were converted into text file. The dataset contains 1000 positive files and 1000 negative files for each domain. The reviews are about four product domains: Books (B), DVDs (D), Electronics (E) and Kitchen appliances (K) and are written in English language. For experiment, labeled dataset of 1000 positive and 1000 negative files was used. An instance in each domain is recorded in Table 1. Except book domain other domains had more number of positive instances.

From this dataset, 12 cross-domain sentiment classification tasks were constructed: $B \rightarrow D$, $B \rightarrow E$, $B \rightarrow K$, $D \rightarrow B$, $D \rightarrow E$, $D \rightarrow K$, $E \rightarrow B$, $E \rightarrow D$, $E \rightarrow K$, $K \rightarrow B$, $K \rightarrow D$, $K \rightarrow E$, where the word before an arrow corresponds to the source domain and the word after an arrow corresponds to the target domain.

From Table 2, it is evident that Book and DVD, if considered as a source domain, achieve a good compatibility with electronics and kitchen domain, which is considered as target domain. Besides, electronic and kitchen are compatible domains.

Baseline methods use in this study are Feature Ensemble plus Sample selection (SS-FE) [27], Spectral feature alignment (SFA) [23], and Supervised word clustering (SWC) [20]. SFA achieved was between 72.5% and 86.75%, SS-FE was between 72.94% and 84.87% and SWC was between 72.11% and 85.33%, whereas accuracy of proposed algorithm was between 70% and 88.35%.

Only the DVD, electronics, and kitchen were considered as the source domain, while book, kitchen and DVD as a target domain, producing comparatively less accurate results than the baseline method (Fig. 3). There are two key points in proposed framework:

Table 1
Negative and positive instances for multi-domain dataset.

Domain Name	Negative Instances	Positive Instances
Book	73,500	72,794
DVD	66,126	76,759
Electronics	43,806	44,321
Kitchen appliances	36,106	36,733

Table 2

Comparative analysis of accuracy of proposed method and baseline methods.

Source \rightarrow Target	Accuracy (%) Proposed Method	Accuracy (%) SS-FE	Accuracy (%) SFA	Accuracy (%) SWC
$B \rightarrow D$	82.45	79.10	82.55	81.66
$B \rightarrow E$	78	74.24	72	77.04
$B \rightarrow K$	78.65	78.07	78	82.26
$D \rightarrow B$	74.35	80.38	77	79.95
$D \rightarrow E$	79.78	77.07	77	76.98
$D \rightarrow K$	84.21	77.82	81	82.13
$E \rightarrow B$	82.15	72.86	75.5	72.11
$E \rightarrow D$	87.8	74.60	77	73.81
$E \rightarrow K$	81.44	84.87	87.1	85.33
$K \rightarrow B$	81.05	72.94	74	75.78
$K \rightarrow D$	70	75.70	77	76.88
$K \rightarrow E$	88.35	82.93	84.6	84.78

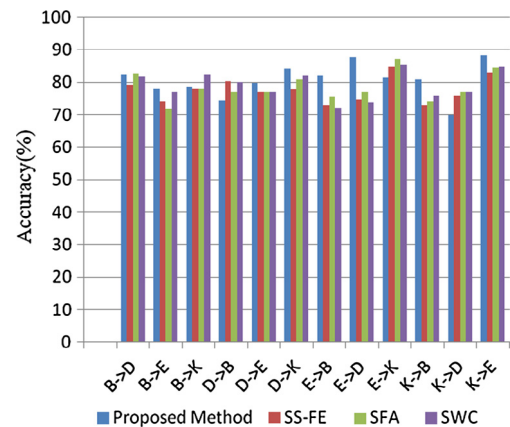


Fig. 3. Accuracy analysis.

first it classifies the words and documents, and then clusters them. After classification step, opinionated words are extracted with weight value as well as polarity. These weights are very important factor as it increases the importance of discriminative terms. Proposed approach also identifies domain-independent and specific features which are used for clustering. All classified words clustered again leads to acceptable results. One of the reasons for low accuracy for some domain is imbalance of class labels and the presence of word disambiguation.

Accuracy is used as an evaluation measure. Accuracy is the proportion of correctly classified examples to the total number of examples; on the other hand, error rate refers to incorrectly classified examples to correctly classified examples. F-measure or precision and recall can be used as evaluation measures.

F-measure is only defined in terms of true positive (TP), false positive (FP) and false negative (FN), while true negative (TN) is not considered. Accuracy and F-measure is compared for proposed approach which shows that, in general, F-measure is similar to accuracy. But only single class is considered in F-measure as positive class (Fig. 4). On the other hand, when calculating accuracy equal weight is given to both the classes.

Classified words are used to find domain-independent and domain-specific words from the respective domains. Domain-independent and domain-specific words are compared to the SentiWordNet [3], in order to find out how many words match with them (Tables 3–6). From the tables, it has been observed that on an average 72.6% domain-specific words are correctly classified for different domains; while 88.4% words from domain-independent word list are correctly classified. Domain-independent words typically occur in every domain; hence, matching percentage is more than the matching percentage of domain-

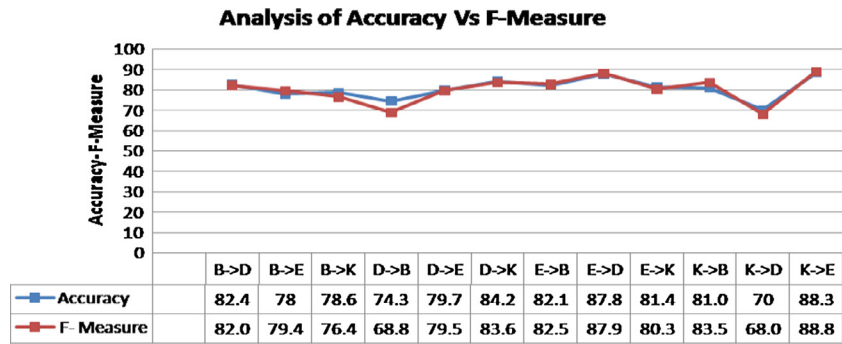


Fig. 4. Analysis of accuracy vs F-measure.

Table 3
Comparison of domain-specific and domain-independent words against SentiWordNet considering book (B) as a source domain.

Domains	Domain-specific words	Domain-independent words	No. of words matching SentiWordNet	
			Domain-specific words	Domain-independent words
B → D	11,503	9744	8934	8972
B → E	5250	4796	4077	4395
B → K	4200	4325	3262	3979

Table 4
Comparison of domain-specific and domain-independent words against SentiWordNet considering DVD(D) as a source domain.

Domains	Domain-specific words	Domain-independent words	No. of words matching SentiWordNet	
			Domain-specific words	Domain-independent words
D → B	11,130	9744	8644	9009
D → E	5238	4781	4068	4418
D → K	4205	4320	3266	3980

Table 5
Comparison of domain-specific and domain-independent words against SentiWordNet considering Electronics (E) as a source domain.

Domains	Domain-specific words	Domain-independent words	No. of words matching SentiWordNet	
			Domain-specific words	Domain-independent words
E → B	16,105	4769	12,508	4430
E → D	16,466	4781	12,789	4462
E → K	4802	3723	3729	3454

Table 6
Comparison of domain-specific and domain-independent words against SentiWordNet considering Kitchen (K) as a source domain.

Domains	Domain-specific words	Domain-independent words	No. of words matching SentiWordNet	
			Domain-specific words	Domain-independent words
K → B	16,549	4325	12,853	3980
K → D	16,927	4320	13,147	3987
K → E	6296	3723	4890	3449

specific words. As SentiWordNe has generalized opinion lexicons, percentage of matching domain-specific words is relatively less.

4.2. Experiment 2

To evaluate accuracy of proposed algorithm for unlabeled target dataset, we have performed 12 cross-domain tasks with labeled source dataset and unlabeled target dataset. Fig. 5 illustrates the accuracy analysis. Accuracy achieved by proposed method for unlabeled target lies between 65.65% and 98.0%; whereas, labeled target achieves accuracy between 70.0% and 88.35%. Performance of classifier, therefore, increases significantly. It shows that elec-

tronics, kitchen and DVD domains are compatible with each other due to their more similar features. In general, the features of these three domains are more or less similar. Also kitchen as source and electronics as target and vice versa gives better accuracy as both domains share more common features. Kitchen appliances domain also shares electronics appliances; hence, major information of both domains is similar. But kitchen as source and DVD as target does not give good results for both labeled and unlabeled dataset, because kitchen and DVD are not similar to each other, as that of kitchen and electronics. Usually, if two domains are more similar then a number of features transferred from source to target are also more because source data is used as training dataset and tar-

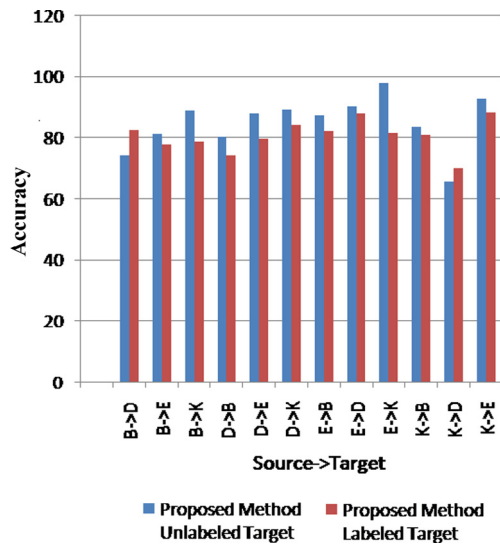


Fig. 5. Accuracy analysis of unlabeled and labeled target.

get features are derived from it. The impact of unlabeled target data is more than that of labeled target data. It states that unlabeled target data can be used for the accuracy gain, as well as reduce the annotation cost significantly.

4.3. Experiment 3

Generally, Naïve Bayes and SVM algorithms are used for text classification. Experiment 3 was conducted to evaluate proposed framework. using Rapid Miner 5.3.015 software for Naïve Bayes and SVM algorithm. The software contains text mining plug-in which converts non-structured textual data into structured format for further analysis. This study adopted implemented linear SVM as most of the text classification problems are linearly separable. Further, as text classification contains large number of features, linear kernel was found to be better suited for this purpose. Results were obtained from Blitz dataset for four different domains of unigrams and for applying word frequency in document and in entire corpus.

Naïve Bayes classifier is a probabilistic classifier based on probability models that incorporate strong independence assumptions among the features. Independence assumption of features is a subtle issue with Naïve Bayes. If certain feature and class label value do not occur together then the frequency-based probability estimate will become zero. When all the probabilities are multiplied, the answer will be zero and this affects the posterior probability estimate. Hence, it provides poor accuracy compared to the SVM

for some domains. The result shows that the proposed algorithm was better than both SVM and Naïve Bayes.

Figs. 6 and 7 provides accuracy and F-measure analysis for each of 12 cross-domain sentiment classification tasks on Amazon product reviews. For this study, 1000 positive and 1000 negative reviews were taken as the source and target domains. Comparison of proposed approach with baseline approaches shows that domain adaptation from book as source domain to the DVD as target domain rather than kitchen as source to DVD as target domain is more feasible. It also shows that relatedness between source and target domain reviews are more important factors for the effectiveness of domain adaptation.

Electronics as a source domain is more compatible with every domain. This suggests that more number of features are relevant in both source and target domains. The proposed approach provides the highest accuracy for electronics as compared to baseline as well as SVM and Naïve Bayes. Compared to accuracy, F-measure results are improved but these are only related with positive documents. Results show that the proposed approach have better accuracy and F-measure. Further, Naïve Bayes provides better results than SVM. Some of the major drawback of Naïve Bayes is assumption of independent attributes and difficulty in interpretations of SVM results. In proposed framework, maximum entropy that is used provides a natural mechanism of multiclass classification. The results are better as increment quantity focuses on term frequency and inverse document frequency. It concentrates on features and its presence which is not focused in Naïve Bayes or SVM.

Random Trees area collection of individual decision trees, in which each tree is generated from different samples and subsets of the training data. Classification of dataset based on random sub selection of training samples result in many decision trees, hence this method is called Random trees. Each tree can be voted to make final decision.

An experiment was carried using Rapid Miner 5.3.015 software. Results are recorded in Fig. 8, which shows the comparison between the proposed approach and the Random tree. The proposed approach gives better accuracy than the Random tree. The random trees classifier takes the input feature vector, classifies it with every tree in the forest, and produces the class label that received the majority of “votes” as output. Using bootstrap approach, training sets are generated. Vectors are randomly selected, hence some vectors will occur more than once or some will be absent. All variables are not used to find the best split.

In contrast, the proposed approach works at word level where each word acts as a feature. Later, importance of word is analyzed using term frequency and inverse document frequency of each word consequently producing better accuracy. Using Random tree higher accuracy achieved in Electronics as source and kitchen as target domain wherein the proposed approach it is reverse way.

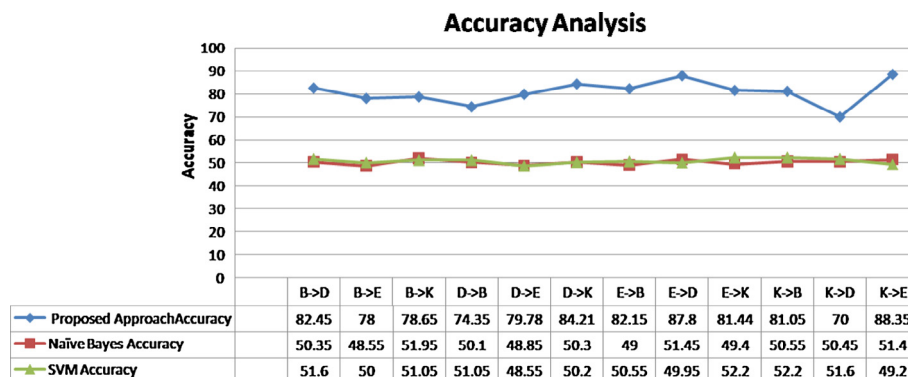


Fig. 6. Accuracy analysis for 12 cross-domain classification tasks.

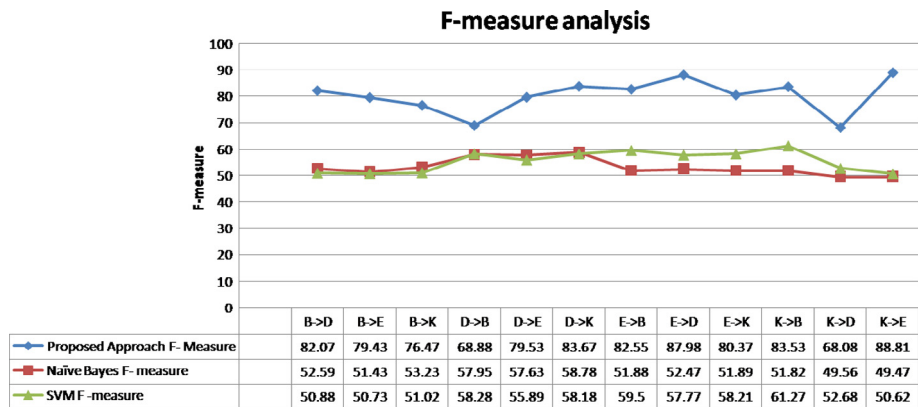


Fig. 7. F-measure analysis for 12 cross-domain classification tasks.

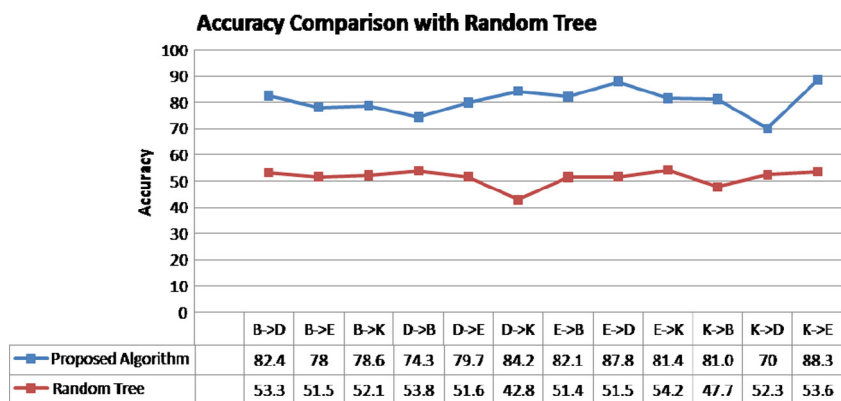


Fig. 8. Accuracy comparison with Random Tree.

Therefore, it is clear that electronics and kitchen are more compatible domains as they are having similar features.

4.4. Experiment 4

For testing the model, Amazon's balanced 6cats dataset collected by Mark Drezde and processed by Richard Johansson in 2012 has been used for this study. The review collection is divided

into six topic subdirectories: Books, Camera, DVD, Health, Music and Software. The document in each topic directory is divided into positive and negative subdirectories. From this dataset, 30 classification tasks were constructed. For 1000 positive and 1000 negative files of each domain, the accuracy was achieved between 70% and 97% (Fig. 9).

Highest accuracy achieved in software as source and camera as target domain. $B \rightarrow C$, $C \rightarrow B$, $D \rightarrow M$, $H \rightarrow S$, $M \rightarrow B$, $S \rightarrow C$ clas-

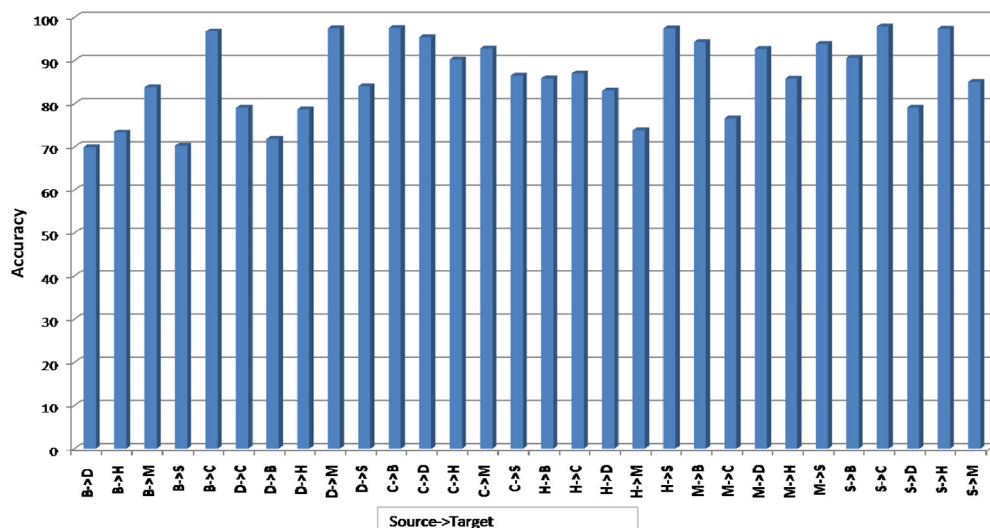


Fig. 9. Accuracy analysis for Amazon balanced 6 cats dataset.

Table 7

Domain-independent, Domain-specific and SWN matched words for Book as source domain.

Domains	Domain-specific words	Domain-independent words	SWN matched words
B → C	3942	4459	4533
B → D	12,044	9896	10,358
B → H	4356	4468	4444
B → M	10,304	7438	7910
B → S	4712	5222	5356

Table 8

Domain-Independent, Domain-Specific and SWN matched words for DVD as source domain.

Domains	Domain-specific words	Domain-independent words	SWN matched words
D → B	10,519	9896	10,259
D → C	3770	4631	4670
D → H	4306	4518	4481
D → M	9305	8437	8883
D → S	4741	5193	5342

Table 9

Domain-independent, Domain-specific and SWN matched words for Camera as source domain.

Domains	Domain-specific words	Domain-independent words	SWN matched words
C → B	15,956	4459	5049
C → D	17,309	4631	5264
C → H	5186	3638	3652
C → M	13,845	3897	4410
C → S	6029	3905	4142

sification tasks are giving high accuracy. From the results, it was found that book, camera and music domains are having more common features.

Tables 7–9 present domain-specific, domain-independent and matching words with SentiWordNet. Domain-independent words are from both source and target domains. Domain-specific are only from target domains. With SentiWordNet matching percent is average 55.7%.

4.5. Discussion

For our experiments, two different datasets were used. Each dataset consisted of labeled positive and negative text review documents. All the results from above sections reveal that the proposed approach gives better accuracy than baseline methods. Word is important entity as it indicates sentiment or opinion of object. The proposed framework is based on modified entropy classifier. Opinionated words are extracted based on the term frequency and inverse document frequency. Increment quantity is modified as granularity refined from document to word level which shows drastic difference between traditional maximum entropy and modified entropy. Bipartite graph clustering is applied on classified data which has enhanced the results.

As compared to baseline methods, moderate accuracy was achieved by the proposed method. Relatedness between source and target domain is important factor in domain adaptation. Also for unlabeled target dataset better accuracy was achieved. It means that it can significantly reduce the annotation cost also. F-measure is also taken as evaluation measure which shows better results of proposed framework over base line methods. But it does not consider the true negative features. Some domains are not compatible

to each other or relatedness between these domains is less, hence the results are lower.

5. Conclusions

Opinion mining is a popular research area; yet, researchers have mainly focussed on domain adaptation. This work addressed the major issue of domain adaptation. In this work, semi-supervised approach was used which holds maximum entropy classifier with modified increment value and bipartite clustering. Labeled as well as unlabeled set of lexicons from different domains were collected from Amazon are used for experimental analysis of the proposed approach. Pre-processing step was used to remove noise from dataset. Each word from dataset is tagged for parts of speech using the Stanford parser. This tagged data is used by classifier which is based on features $tfidf_i$ and idf_i , value which is useful to measure the words' ability to discriminate between documents. Domain-specific and domain-independent lexicons are used for clustering. Classified lexicons are compared with SentiWordNet 3.0 to find matching percentage as SentiWordNet is publicly available lexicon resource.

This work was able to produce relatively good results for some of the domain, and it was able to handle only two classes with an acceptable accuracy. Domain-specific and domain-independent words compared to SentiWordNet 3.0 shows average matching percent 68.25%. The experimental results of proposed approach have shown a significant increase in accuracy for different domains over baseline approach as the proposed framework emphasizes on granularity of the word. This is the major change in classifier in comparison to traditional approach. Importance of each word that has more impact on results of classifier was classified. Testing of approach carried on Amazon cat6 dataset, which shows a significant improvement in accuracy ranging from 3 to 6 points compared to dataset from Blitzer. In comparison to SVM and Navie Bayes, we have proposed an algorithm that could provide better accuracy. It shows that relatedness between domains is a major factor for effectiveness of domain adaptation.

In the proposed system, bipartite graph clustering was used to reduce the mismatch between domain specific words of source domain and target domain. Domain-independent words were used to cluster domain-specific words from source and target domains. To train classifier for target domain, clustering was used as it reduced the gap between domain-specific words of different domains. Future studies can be taken up to determine the co-clustering of words and documents from different domains. The proposed system focuses on only words, in future non-word features like the age of document, the recommendation counts of document can be considered. At present, framework considers only unigrams and reviews are in English language. Also in future this work can be extended for other languages as well as n-grams.

References

- [1] Abinash Tripathy, Ankit Agrawal, Santanu Kumar Rath, Classification of sentiment reviews using n-gram machine learning approach, in: Expert Systems with Applications, vol. 57, 2016, pp. 117–126.
- [2] Aytug Onan, Serdar Korukoglu, Hasan Bulut, A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification, in: Elsevier Expert Systems With Applications, vol. 62, 2016, pp. 1–16.
- [3] Baccianella Stefano, Esuli Andrea, Sebastiani Fabrizio, SentiWordNet 3.0: An Enhance Lexical Resource for Sentiment Analysis and Opinion Mining, in: Proceedings of the 7th Language Resources and Evaluation Conference (LREC 2010), Valletta, Malta, May 17–23, 2010, pp. 2200–2204.
- [4] A. Bermingham, M. Conway, L. McInerney, N. O'Hare, A. Smeaton, Combining Social network analysis and sentiment analysis to explore the potential for online radicalisation, in: Proc. of Int'l Conf. on Advances in Social Network Analysis and Mining, Athens, Greece, July 20–22, 2009, pp. 231–236.

- [5] Bing Liu, *Sentiment Analysis & Opinion Mining*, Kindle Edition., Morgan & Claypool Publishers, 2012.
- [6] John Blitzer, Mark Dredze, Fernando Pereira, Biographies, Bollywood, boom-boxes and blenders: domain adaptation for sentiment classification, in: Association of Proceedings of the 45th Annual Meeting of the Computational Linguistics (ACL), Prague, Czech Republic, June 2007, pp. 440–447.
- [7] D. Bollegala, D. Weir, J. Carroll, Cross-domain sentiment classification using a sentiment sensitive thesaurus, in: Knowledge and Data Engineering, IEEE Transactions, vol. 25(8), 2013, pp. 1719–1731.
- [8] N.F. Da Silva, E.R. Hruschka, E.R. Hruschka, Tweet sentiment analysis with classifier ensembles, in: Elsevier Decision Support Systems, vol. 66, 2014, pp. 170–179.
- [9] M.C. De Marneffe, B. MacCartney, C.D. Manning, Generating typed dependency parses from phrase structure parses, in: Proceedings of LREC, vol. 6, 2006, pp. 449–454.
- [10] E. Fersini, E. Messina, F.A. Pozzi, Sentiment analysis: Bayesian Ensemble Learning, Elsevier Decision Support Syst. 68 (2014) 26–38.
- [11] A. Esuli, F. Sebastiani, Senti-WordNet: A Publicly Available Lexical Resource for Opinion Mining, in: Proc. of the 05th Conf. on Language Resources and Evaluation, Genova Italy, May 22–28, 2006, pp. 417–422. Available at <<http://sentwordnet.isti.cnr.it/>>.
- [12] Farhan Hassan Khan, Usman Qamar, Saba Bashir, S WIMS: semi-supervised subjective feature weighting and intelligent model selection for sentiment analysis, Knowledge-Based Syst. 100 (2016) 97–111.
- [13] Haiqing Zhang, Aicha Sekhari, Yacine Ouzrout, Abdelaziz Bouras, Jointly identifying opinion mining elements and fuzzy measurement of opinion intensity to analyze product features, Eng. Appl. Artificial Intelligence 47 (2016) 122–139.
- [14] Kumar Ravi, Vadlamani Ravi, A survey on opinion mining and sentiment analysis: tasks, approaches and applications, Knowledge Based Syst. 89 (2015) 14–46.
- [15] J.V. Lochter, R.F. Zanetti, D. Reller, T.A. Almeida, ShortText opinion detection using ensemble of classifiers and semantic indexing, Expert Syst. Appl. 62 (2016) 243–249.
- [16] Łukasz Augustyniak, Tomasz Kajdanowicz, Piotr Szymński, Włodzimierz Tułgłowicz, Przemysław Kazienko, Reda Alhajj, Bolesław Szymanski, Simpler is better? lexicon-based ensemble sentiment classification beats supervised methods, in: International Workshop on Curbing Collusive Cyber-gossips in Social Networks (C3-2014), August 17, 2014 Proc. IEEE/ACM Int. Conf. Advances in Social Network Analysis and Mining, ASONAM, Beijing, China, 2014.
- [17] McCallum Andrew, Freitag Dayne, Pereira Fernando, Maximum entropy markov models for information extraction and segmentation, in: 17th International Conf. on Machine Learning, Stanford University, June 29–July 2, 2000.
- [18] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: a survey, Ain Shams Eng. J. 5 (4) (2014) 1093–1113.
- [19] G.A. Miller, WordNet: a lexical database for English, Commun. ACM 38 (11) (1995) 39–41.
- [20] Min Xiao, Feipeng Zhao, Yuhong Guo, Learning latent word representations for domain adaptation using supervised word clustering, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 152–162, Seattle, Washington, USA, 18–21 October, 2013.
- [21] Min Xiao, Yuhong Guo, Feature space independent semi-supervised domain adaptation via kernel matching, IEEE Trans. Pattern Anal. Mach. Intelligence 37 (1) (2015) 52–66.
- [22] S. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowledge Eng. 22 (10) (2009) 1345–1359.
- [23] Pan SinnoJialin, Xiaochuan Ni, Jian-Tao Sun, Qiang Yang, Zheng Chen, Cross-domain sentiment classification via spectral feature alignment, in: Proceedings of the 19th International World Wide Web Conference, ACM, Raleigh, USA, April 26–30, 2010.
- [24] B. Pang, L. Lee, Opinion mining and sentiment analysis, Found. Trends Inform. Retrieval 2 (1–2) (2008) 1–135.
- [25] Paramveer S. Dhillon, Partha Talukdar, Koby Crammer, Metric Learning for Graph-Based Domain Adaptation, University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-12-17, January 2012.
- [26] Rui Xia, Chengqing Zong, Shoushan Li, Ensemble of feature sets and classification algorithms for sentiment classification, in: Information Sciences, vol. 181(6), 15 March 2011, pp. 1138–1152.
- [27] Rui Xia, Chengqing Zong, Xuelei Hu, E. Cambria, Feature ensemble plus sample selection: domain adaptation for sentiment classification, IEEE Intelligent Syst. 28 (3) (2013) 10–18.
- [28] Rui Xia, Feng Xu, Jianfei Yu, Yong Qi, Erik Cambria, Polarity shift detection, elimination and ensemble: a three-stage model for document-level sentiment analysis, Inform. Process. Manage. 52 (2016) 36–45.
- [29] Shoushan Li, Yunxia Xue, Zhongqing Wang, Guodong Zhou, Active learning for cross-domain sentiment classification, in: Proceedings of IJCAI'13 the Twenty-Third International Joint Conference on Artificial Intelligence, 2013, pp. 2127–2133.
- [30] P.K. Singh, M.S. Husain, Methodological study of opinion mining and sentiment analysis techniques, Int. J. Soft Comput. 5 (1) (2014) 11.
- [31] M. Wang, H. Shi, Research on sentiment analysis technology and polarity computation of sentiment words, in: Proc. of Int'l Conf. on Progress in Informatics and Computing, Shanghai, vol. 1, December 10–12, 2010, pp. 331–334.
- [32] Zheng Chutao, Liu Cheng, Wong Hau-San, Iterative term weighting for short text data, in: Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, 9–12 October 2015.