Pattern Recognition Letters 56 (2015) 38–44

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Efficient extraction of domain specific sentiment lexicon with active learning ☆

Sungrae Park, Wonsung Lee, Il-Chul Moon*

Department of Industrial and Systems Engineering, KAIST, 291 Daehak-ro, Daejeon 305-701, Republic of Korea

ARTICLE INFO

Article history:
Received 18 July 2014
Available online 11 February 2015

Keywords:
Sentiment analysis
Active learning
Sentiment lexicon

ABSTRACT

Recent research indicates that a sentiment lexicon focusing on a specific domain leads to better sentiment analyses compared to a general-purpose sentiment lexicon, such as SentiWordNet. In spite of this potential improvement, the cost of building a domain-specific sentiment lexicon hinders its wider and more practical applications. To compensate for this difficulty, we propose extracting a sentiment lexicon from a domain-specific corpus by annotating an intelligently selected subset of documents in the corpus. Specifically, the subset is selected by an active learner with initializations from diverse text analytics, i.e. latent Dirichlet allocation and our proposed lexicon coverage algorithm. This active learning produces a better domain-specific sentiment lexicon which results in a higher accuracy of the sentiment classification. Subsequently, we evaluate extracted sentiment lexicons by observing (1) the increased F1 measure in sentiment classifications and (2) the increased similarity to the sentiment lexicon with the full annotation. We expect that this contribution will enable more accurate sentiment classification by domain-specific sentiment lexicons with less sentiment tagging efforts.

1. Introduction

Recent research on big data analytics has developed and identified major analytics problems, such as sentiment analysis. While the sentiment analysis on product reviews and political debates on the Web is not a new research problem, social networking services (e.g., Twitter and Facebook) and blogospheres are now producing enormous amount of datasets to be analyzed from the sentiment analysis. The difference between the amount of datasets from the old-fashioned Web and modern social media is just one facet of the increasing difficulties in the sentiment analysis. For example, as more Internet users produce documents with a certain sentiment, more sentiment words are introduced to represent a specific sentiment. Furthermore, the new sentiment words are either (1) new words; (2) words with completely different meanings, but that are used to show a certain sentiment in a specific domain; or (3) traditional sentiment words that have been collected and analyzed by researchers. Arising of new sentiment words is particularly rapid in the big data era, so we need a new approach to resolve this problem. While we limit the sentiment analysis to the sentiment identification area, capturing new sentiment words is one of the keys to improving the result of the analysis. Either supervised classification or unsupervised clustering approaches utilize a set of sentiment keywords, or a sentiment lexicon. One of the most popular sentiment lexicons is SentiWordNet [1]. The challenge of SentiWordNet is its slow adaptability to the constantly piling documents and their embedded sentiment words. Moreover, SentiWordNet points out which words show which sentiment without regarding the context of the words. These domain-specific sentiment words turn out to be a key determinant in identifying a sentiment from the past experiments, i.e. Ref. [2]. A solution to the inadaptability and context awareness is a dynamic generation of sentiment lexicons based upon a specific domain. This dynamic generation fundamentally relies on how to capture new sentiment words.

This papers aims to solve two problems in developing domain-specific sentiment lexicons. The first problem is extracting a sentiment lexicon from document level annotations. We present a simple probabilistic algorithm utilizing Bayesian estimation. This algorithm turns the sentiment annotation at the document level into the annotation at the word level. The second problem is efficiently selecting the document to annotate to minimize the document annotation effort. In detail, we sample documents from a sentiment corpus with active learning techniques. As the active learning optimizes the decision boundary by querying the most important data point in determining the boundary, the active learners find the most significant document that determines the sentiment classification boundary. Moreover, the performance of active learning is often influenced by the initial

☆ This paper has been recommended for acceptance by Jie Zou.

* Corresponding author. Tel.: +82 42 350 3118; fax: +82 42 350 3110.
E-mail address: icmoon@kaist.ac.kr (I. Moon).

http://dx.doi.org/10.1016/j.patrec.2015.01.004
0167-8655/

documents for training, so we utilize diverse machine learning algorithms to find the best initialization. Solutions to the two problems collectively reduce the cost of building a domain-specific sentiment lexicon. Our evaluation indicates that the sentiment classification with our lexicon made statistically significant improvement on F1 scores when more than 8% of documents are annotated compared to the case using SentiWordNet and 15% of documents are annotated compared to the case using unsupervised methods.

2. Previous research

Before looking more closely at domain-specific lexicon, we briefly introduce general sentiment lexicons. Numerous studies of developing a sentiment lexicon depend on the concept of general sentiment. The general lexicons, such as General Inquirer [3], OpinionFinder's Subjectivity Lexcion [4], and SentiWordNet [5], provide sentiment polarity of a word in general contexts; however, these well-defined lexicons are not adaptable to all possible domains. While there are researches utilizing the general sentiment lexicon, often such works utilize a sentiment analysis algorithm that is optimized to the target domain or venue.

A sentiment lexicon capturing better domain characteristics improves the quality of sentiment analyses, such as sentiment classification [6] and opinion retrieval [7]. In several early works [8], human annotators classified sentence-level sentiments [9]. However, this manual annotation of all words in documents is expensive, as well as has high divergence by the annotators. To tackle this problem, a lexicon induction with machine learning techniques has been studied extensively [10].

Previous methods extracting domain-specific sentiment lexicon categorize two approaches as corpora-based and network-based methods. Corpora-based approaches extract sentiment words by exploring practical usage text data. Especially previous researches proposed rule-based algorithm using the co-occurrence pattern with general sentiment lexicon. Zagibalov and Carroll [11] proposed a method of automatic seed word selection using the co-occurrence pattern of negations and adverbials. Turney [12] calculated the pointwise mutual information (PMI) with GSL to classify whether a sentiment polarity of a given phrase is positive or negative. Kanayama and Nasukawa [13] proposed clause-level sentiment analysis to identify domain-specific sentiment atoms using parse tree of each sentence and GSL information. Qiu et al. [14] suggested double propagation algorithm utilizing dependency information of each text to find domain specific opinion words.

The network-based approaches used lexical relations defined in WordNet to find new opinion words from pre-defined sentiment words. Kim and Hovy [15] proposed a simple method to add all synonyms of a polar word with the same polarity and its antonyms with opposite polarity. Kamps [16] found sentiments of adjectives in WordNet by measuring relative distance of each term from pre-defined sentiment words. Rao and Ravichandran [17] proposed a label propagation to detect the polarity of a word in WordNet. These approaches cannot be easily applied when lexical relations are not accessible. Since our proposed approach can be categorized as corpura-based method, we only compared our model with corpura-based approaches in the experiment.

Active learning, a.k.a. query learning, is a technique to support a learner to create a decision boundary more efficiently with the small number of training instances [18]. There are three major variants in the active learning techniques: membership query synthesis, stream-based selective sampling, and pool-based sampling. The pool-based sampling means that a learner has access to a pool of totally unlabeled data and, in turn, selects which to label. This approach is very analogous to our setting because we have a pool of corpus to annotate. In this reason, we chose the pool-based sampling approach and utilized the simple margin rule [19] to sample the pool.

At the initial phase of applying active learning to the classification, we need to select an initial set for the first learning batch, as the selecting rule does not have any basis to operate at the first time. This initial set critically influences the performance of active learning, as there have been many studies on how to decide the initial training set. Nguyen and Smeulders [20] verified that active learning could be more efficient with pre-clustering information. They clustered total instances in a pool and used it as a feature for active learning and classification. Kang et al. [21] proposed a method of selecting an initial training set for active learning using K-means clustering. They assumed that instances in the same cluster have the same label and that the best representatives of clusters help the learner to study much more information.

3. Sentiment lexicon extraction

Sentiment lexicon is a set of words indicating a certain sentiment aspect. Most research on sentiment lexicon building utilizes general sentiment lexicons to improve the quality of the extracted sentiment lexicon that is domain-specific. On the contrary, this paper proposes to extract a sentiment lexicon by utilizing human annotations at the document level. Our proposed model, which we named sentiment lexicon extraction (SLE), is a generative probabilistic model that includes two key assumptions. One assumption is that a word v has a probability distribution W_v over the number of sentiments S that an annotator defines as such:

$$W_v \sim (W_{v,1}, \ldots, W_{v,S}), \quad W_{v,S} \geq \forall S, \quad \sum_{s=1}^{S} W_{v,S} = 1 \tag{1}$$

For example, if there are polarity sentiments which are positive and negative, good has high positive probability but low negative probability, bad has low positive probability but high negative probability, neutral which looks no sentiment word has positive and negative probability close to 0.5. The other assumption is that a sentiment of a document d is defined as DS_d, and it is represented as the product of sentiment probabilities of words in the document d:

$$P(DS_d = s|W) \propto \prod_{n=1}^{N_d} W_{(n)_d,S}, \tag{2}$$

where W represents all sentiment probabilities of words in a sentiment corpus D. N_d indicates the number of words in the document d. $W_{(n)_d,s}$ means a probability that $W_{(n)_d}$ is assigned to the sentiment s. Here, $(n)_d$ is nth word index in document d. This assumption indicates that a document with words bearing a high probability of a certain sentiment is more likely to have the same sentiment at the document level. According to this equation, neutral words that have similar degrees of positive and negative probability do not influence much to the document level sentiment because such words will equally contribute to the sentiment likelihood of the document. However, the words with a biased sentiment probability distribution would contribute to the document sentiment differently. Another important point is that $W_{(n)_d,s}$ is ultimately derived from the Bayesian estimation with a prior. This will prevent $W_{(n)_d,s}$ from being zero, so the whole probability becomes zero.

Initially, we tested maximum likelihood estimator (MLE) of W_v. Our test shows that MLE is very susceptible to errors when we have few or no annotations on a certain word, which is very likely to happen when we annotate only a subset of a large corpus. In order to infer these distributions, we applied Bayesian Estimation of W_v (SLE-BE).

Fig. 1 illustrates SLE-BE in the plate diagram. The diagram indicates that the prior information α of each word is also introduced. In this case, we assume that the word-sentiment probability distribution is a Dirichlet distribution whose parameter is α, $W_v \sim \mathrm{Dir}(\alpha)$. According to the model structure in Fig. 1, the sentiment lexicon building given

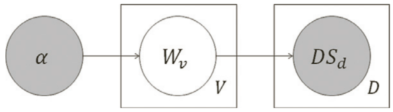

Fig. 1. Plate diagram of sentiment lexicon extraction-Bayesian estimation, or SLE-BE, model.

annotated documents is inferencing the probability distribution of W_v given D. The following is the extended description of the inference.

$$
\begin{aligned}
P(W|D) &\propto P(D|W)P(W) = \prod_{d=1}^{D} P(DS_d|W) \times \prod_{v=1}^{V} P(W_v|\alpha) \\
&\propto \prod_{v=1}^{V}\prod_{s=1}^{S} W_{v,s}^{n_{v,s}} \times \prod_{v=1}^{V} \frac{\Gamma(\sum_{s=1}^{S}\alpha_s)}{\prod_{s=1}^{S}\Gamma(\alpha_s))} \prod_{s=1}^{S} W_{v,s}^{\alpha_s - 1} \\
&= \prod_{v=1}^{V} \frac{\Gamma(\sum_{s=1}^{S}\alpha_s)}{\prod_{s=1}^{S}\Gamma(\alpha_s))} \prod_{s=1}^{S} W_{v,s}^{n_{v,s}+\alpha_s-1} \propto \prod_{v=1}^{V} \mathrm{Dir}(\{n_{v,s}+\alpha_s\}) \quad (3)
\end{aligned}
$$

In the above proof, $n_{v,s}$ is the count of word v in documents that are labeled as sentiment s. As the results of the inference, the posterior distribution of W_v is derivated as follows:

$$\therefore W_v|D \sim \mathrm{Dir}(\{n_{v,s}+\alpha_s\}) \quad (4)$$

The below is the expectation on the distribution over the conditional probability of a word in a certain sentiment.

$$E(W_{v,s}|D) = \frac{n_{v,s}+\alpha_s}{\sum_{s=1}^{S}(n_{v,s}+\alpha_s)} \quad (5)$$

The point estimation of $W_{v,s}$ represents an intuition that the number of appearances in documents of a certain sentiment determines the sentiment probability of the word v in the same sentiment. This result corresponds with word sentimentality analysis of Potts [22], and the only difference is the existence of prior parameter α. This difference originates from the difference of the maximum likelihood estimation (MLE) and Bayesian estimation (BE). Setting of α makes significant differences of the estimated sentiment distribution. This difference comes from two dimensions. First, the sum amount of α controls the noise sensitivity in the data, and this control is particularly important in getting stable estimation results. Second, the biased distribution of α provides the prior knowledge on the sentiment of a corpus, as all the Bayesian approaches do. SLE-BE infers the probability distribution of words over sentiments. While this is a probabilistic representation of a sentiment lexicon, many sentiment classification algorithms require a definite lexicon, as a list of keywords without probabilities. This requirement leads us to define a threshold to cut the keyword list that is ordered by probability.

4. Active learning to estimate sentiment of documents

This paper aims to extract sentiment lexicon using our proposed method, SLE-BE. However, SLE-BE requires a set of sentiment-labeled documents to learn the Bayesian estimator. In this process, labeling all documents would be too expensive given the potential amount of documents. To reduce this inefficient labeling effort, this paper adopted the active learning to get the most salient input document set for learning the SLE-BE. The active learning is a closed loop of sampling critical data points efficiently in order to significantly influence the learners decision boundary. That is, an active learner might have a better performance with a subset of train instances than a learner trained with the same size of instances which are randomly selected.

Our detailed approach of the active learning is the pool-based active learning, which assumes that a learner has access to a large pool of unlabeled instances from the beginning of the process. This paper focuses on applying active learning to a SVM sentiment classifier. SVM is a generally used, high-performance classification algorithm in the sentiment analysis field [23]. Our active learning of SVM incorporates the simple margin rule [19] to select a document to annotate. Additionally, while the active learning of SVM can be either inductive or transductive, we chose to use the inductive learning because of the long training time required in the transductive case.

Whereas these settings of active learning with SVM are standard in the previous research, i.e. Ref. [24], our contribution lies in improving the selection of the initial dataset to train. The most popular pre-clustering method for selecting initial set is the K-means algorithm [20]. Here, we add two alternatives to the K-means algorithm. First, we test LDA to cluster the documents. Second, we present our lexicon coverage analysis (LCA) algorithm that does not cluster the documents to find the set of initial datasets from the centroids, but, rather, incrementally expands the coverage of unique words in the sentiment corpus.

4.1. *K*-means algorithm

The K-means algorithm groups a collection of instances into K clusters so as to minimize the sum of the squared distances to

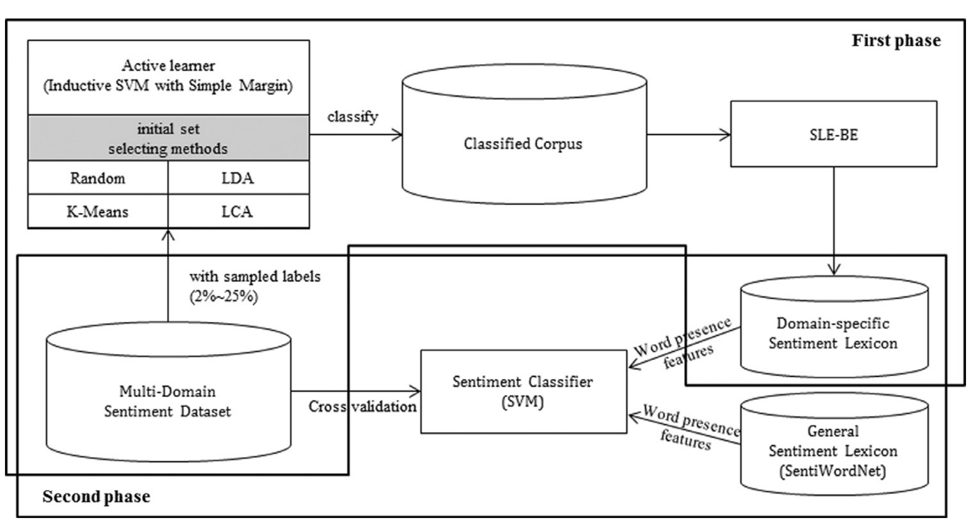

Fig. 2. Illustration of experimental framework.

the cluster centers. It can be implemented as a simple EM procedure to create clusters. To obtain the initial set required for active learning, the instances that are closest to the cluster centroids are selected. We set the size of initial set as the number of clusters and pick the closest one to the cluster centroid of each cluster.

4.2. Latent Dirichlet allocation

Probabilistic topic models have increased the capability of analyzing a large corpus, and the most popular model is latent Dirichlet allocation, or LDA. LDA is generative graphical model using a Bayesian network. By learning LDA, the conditional probabilities of topic appearance in each document are inferred and these can are used for soft clustering features. We set the size of initial set as the number of topics and pick the most probable one to each topic. The parameters in LDA are approximated by the Gibbs sampling. Further details of the sampling procedure on LDA can be found in Ref. [25].

4.3. Lexicon coverage analysis

Given that the initial set is critical in the performance of the active learning, we hypothesize that covering the unique words maximally is not always equal to finding every centroid from the clustering algorithms. If the clustering result is not optimally distributed across the space of unique words, it might be better to greedily increase the word coverage from the perspective of the sentiment lexicon extraction. To approach the initial dataset problem from the lexicon coverage view, we propose an algorithm, lexicon coverage analysis (LCA), for finding maximum lexicon in a corpus (see Algorithm 1). First, we find a document with the maximum number of words. After the selection of a document, we iteratively add the next document that has the minimum Jaccard similarity between the words in a document and the words that have been covered.

Data: 1. A pool of unlabeled documents $D = \{D_1, \ldots, D_{|D|}\}$ where each document is a set of words: $\forall D_i = \{W_{i,1}, \ldots, W_{i,N_i}\}$, 2. The size of next documents to annotate, k
Result: A set of documents $MDC = \{D_1, \ldots, D_k\}$
begin
  $md \leftarrow \arg\max_{i \in |D|} |D_i|$;
  $MDCW \leftarrow D_{md}$;
  $MDC \leftarrow \{D_{md}\}$;
  while $|MDC| < k$ **do**
    $cd \leftarrow \arg\min_{i \in |D|, D_i \notin MDC} JaccardSimilarity(MDCW, D_i)$;
    $MDCW \leftarrow MDCW \cup D_{cd}$;
    $MDC \leftarrow MDC \cup \{D_{cd}\}$;
  end
  return MDC
end

Algorithm 1: Algorithm for lexicon coverage analysis (LCA).

5. Experimental framework

Our experiment involves multiple machine learning algorithms and datasets, so this section displays the experiment settings in detail (see Fig. 2). The first phase of the experiment is extracting sentiment lexicons through active learning and the SLE-BE algorithm. The input dataset provides a partially annotated corpus from 2% to 25% by the experiment setting, and the annotation is only attached to the documents queried by the active learners, which contain either random initialization, K-means initialization, LDA initialization, or LCA initialization. The active learner results in a corpus with fully sentiment-classified documents by only using the partial annotations. Then, this classified corpus becomes the input to the sentiment lexicon extraction model (SLE-BE), and the SLE-BE produces domain-specific

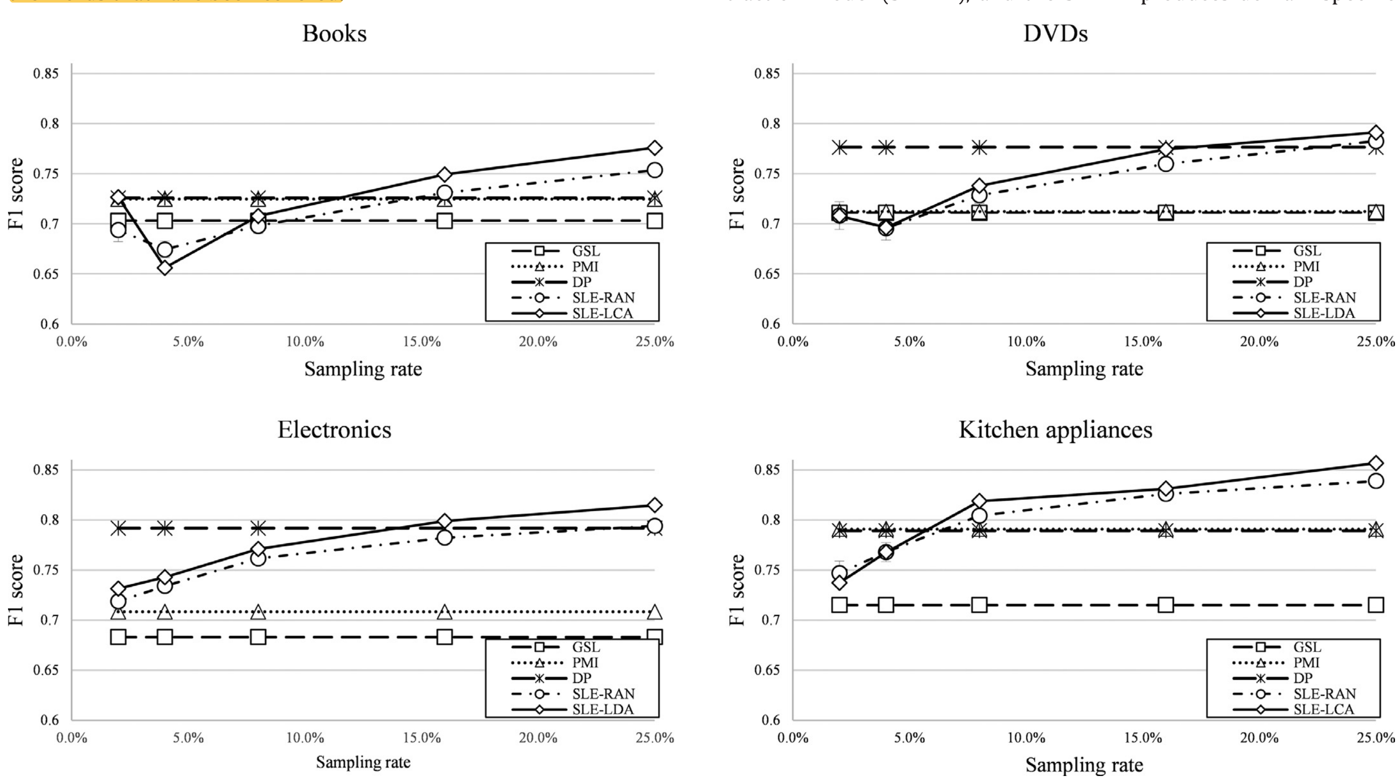

Fig. 3. Comparison of F1 classification scores using the domain-specific sentiment lexicons and the general sentiment lexicon by sampling rates; general sentiment lexicon (GSL), pointwise mutual information (PMI), double propagation (DP), SLE case from random sampled documents (SLE-RAN) and SLE case from active learned documents with the best performed initial set building algorithms, latent Dirichlet allocation (SLE-LDA) and lexicon coverage analysis (SLE-LCA).

sentiment lexicons by thresholds. The second phase of the experiment evaluates the performance of the domain-specific sentiment lexicons. We trained a simple SVM for the sentiment classification, and its features include the word presences in the sentiment lexicon. We designed this evaluation procedure by following the general feature selection and evaluation methodology in the previous works, i.e. Ref. [26]. We tested two cases: one case using the domain-specific sentiment lexicon and another using general sentiment lexicon, SentiWordNet, in this experiment. Five-fold cross-validations were performed to the two cases, and their F1 scores have been compared. In terms of parameters in the algorithms and the models, we calibrated the cutting threshold of SLE-BE to produce a sentiment lexicon with the same size of corresponding words with SentiWordNet at each domain. Also, the prior value α in SLE-BE is set to be three. We replicated the experiment 30 times at each experiment setting.

6. Results

Our algorithms and models are applied to the standard sentiment corpus, the multi-domain sentiment dataset. In addition, the general sentiment words extracted from SentiWordNet are used to compare with the performance of the domain-specific lexicon. The evaluation is performed at two stages. First, we demonstrate that domain-specific sentiment lexicons from our approach capture more salient information than general sentiment lexicon with higher sentiment classification accuracy. Additionally, we compared our approaches with previous works, PMI from [12] and double propagation (DP) from Ref. [14]. Second, we show the evolution of domain-specific sentiment lexicons, and we observe the increasing similarity between the sentiment lexicon from full annotations and the lexicon from samplings.

6.1. Dataset description

The multi-domain sentiment dataset (MDSD) contains product reviews in a number of products from Amazon.com. We selected its preprocessed version from the dataset provider, and our domains were limited to books, DVDs, electronics, and kitchen appliances reviews [27]. Each original dataset consists of 2000 sentiment-labeled documents with 1000 positive and 1000 negative documents, and we used

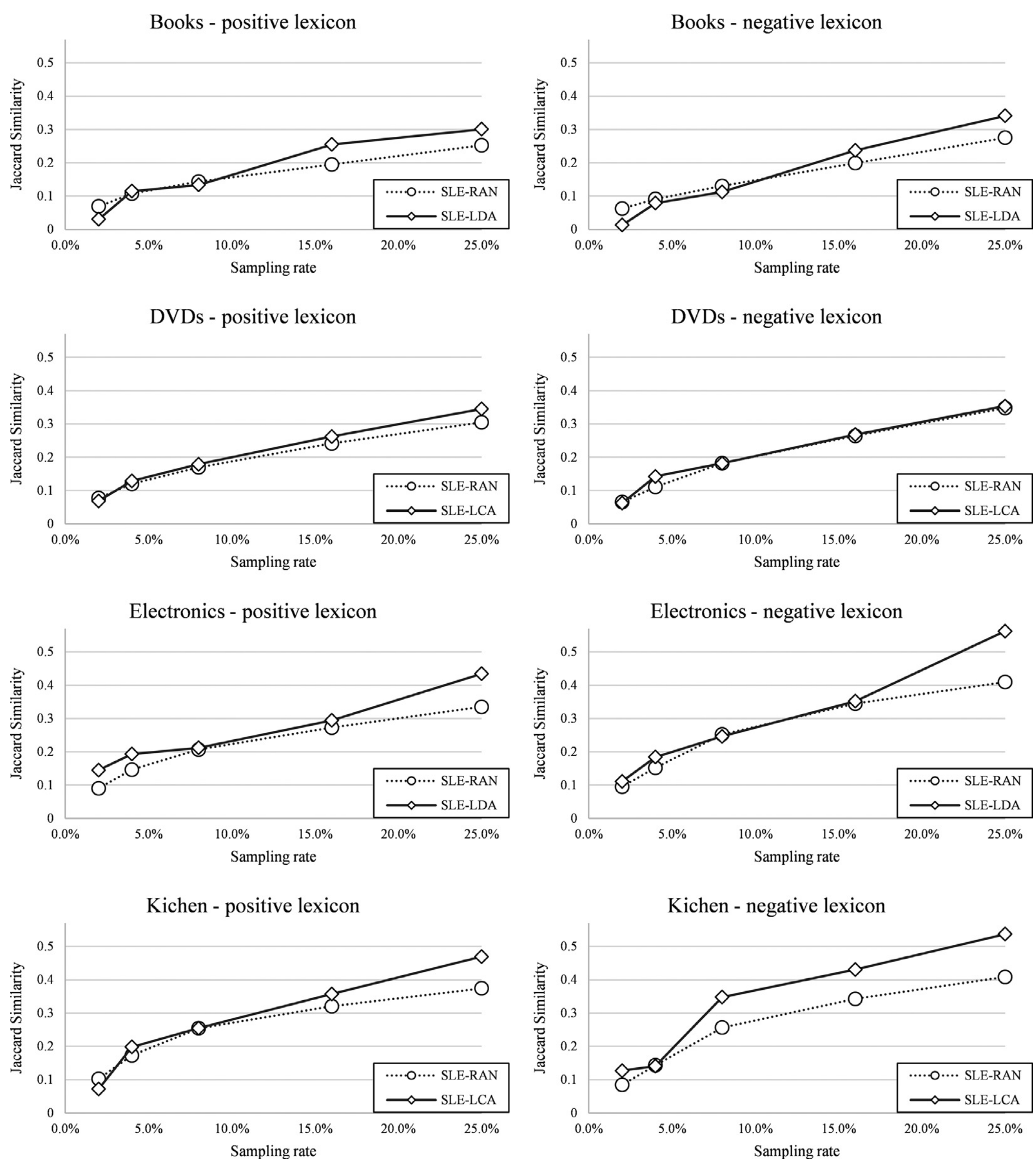

Fig. 4. Jaccard similarity by sampling rates; the similarity between the domain-specific sentiment lexicon from full annotations and the lexicon from sampled annotations, (1) SLE of active learning cases with the best performed initial set building algorithm, latent Dirichlet allocation (SLE-LDA) and lexicon coverage analysis (SLE-LCA), and (2) SLE of random sampled documents (SLE-RAN).

Table 1
Descriptive statistics of the multi-domain sentiment dataset utilized in the evaluation.

Domain	Average number of words in a document	Total number of unique words	Number of words in positive SentiWordNet	Number of words in negative SentiWordNet
Books	91.6	5915	305(5.16%)	232(3.92%)
DVDs	90.4	5731	277(4.83%)	251(4.38%)
Electronics	61.3	3375	142(4.24%)	105(3.11%)
Kitchen appliances	53	3015	133(4.41%)	93(3.08%)

the whole dataset in each domain. Additionally, we filtered words whose document frequency was either less than three or greater than 1000. Table 1 represents the description of preprocessed dataset. A corpus of each domain contains words which correspond with general sentiment words in SentiWordNet which we chose strong and normal positive (negative) words to represent general sentiment lexicon.

6.2. Performance of domain-specific sentiment lexicon

The performance of the proposed domain-specific sentiment lexicons is compared to the general sentiment lexicon and the existing domain-specific lexicon building algorithms in Fig. 3. Three flat lines indicate the performance of (1) the general sentiment lexicon, (2) PMI based domain-specific lexicon, and (3) DP based domain-specific lexicon because it does not depend on the sampling rate which indicates labeling ratio during active learning. The figure shows that all the existing methods are inferior than the active learning based methods when the sampling rates become high enough. Specifically, the domains on books and kitchen appliance show significantly higher accuracy when we use the active learning.

6.3. Analysis of lexicon evolution

This section investigates the evolution of the domain-specific sentiment lexicon as the sampling rate increases. The evolution is measured by the Jaccard similarity between the domain-specific sentiment lexicon with full annotations and the lexicon with sampled annotations. The higher similarity at the low sampling rate reflects the cost saving with fewer annotations. Fig. 4 illustrates the changes of the Jaccard similarity in four domains. As the sampling rate increases, the partial annotation gets closer to the complete annotation. We showed two curves: the curve from the active learning with the optimal initialization and the other curve from random-sampling method. As expected, the active learning with the optimal initialization outperforms the other case in increasing the similarity.

7. Conclusion

We demonstrate the effectiveness of our proposed method by evaluating the extracted lexicon against SentiWordNet and existing algorithms to build domain specific lexicons. The sentiment classification with the active learning provides significantly better F1 scores than the baselines. The proposed sentiment lexicon extraction algorithm and the active learning approach are generally applicable to N-gram cases. We expect that this work would provide better feature selections in the sentiment classifications and better lexicon building in the domain-specific sentiment analysis.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) and funded by the Ministry of Science, ICT & Future Planning (2012R1A1A1044575).

References

[1] Y. Dang, Y. Zhang, H. Chen, A lexicon-enhanced method for sentiment classification: an experiment on online product reviews, IEEE Intell. Syst. 25(4) (2010) 46–53.
[2] Y. Lu, M. Castellanos, U. Dayal, C. Zhai, Automatic construction of a context-aware sentiment lexicon: an optimization approach, in: Proceedings of the 20th International Conference on World Wide Web, ACM, 2011, pp. 347–356.
[3] P.J. Stone, D.C. Dunphy, M.S. Smith, The general inquirer: a computer approach to content analysis. Cambridge, MA: MIT Press, 1966.
[4] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity in phrase-level sentiment analysis, in: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2005, pp. 347–354.
[5] S. Baccianella, A. Esuli, F. Sebastiani, Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, in: Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta, May 2010.
[6] R. Xie, C. Li, Lexicon construction: a topic model approach, in: 2012 International Conference on Systems and Informatics (ICSAI), IEEE, 2012, pp. 2299–2303.
[7] S.H. Na, Y. Lee, S.H. Nam, J.H. Lee, Improving opinion retrieval based on query-specific sentiment lexicon, in: Advances in Information Retrieval, Springer, 2009, pp. 734–738.
[8] R.F. Bruce, J.M. Wiebe, Recognizing subjectivity: a case study in manual tagging, Nat. Lang. Eng. 5(2) (1999) 187–205.
[9] D. Ghazi, D. Inkpen, S. Szpakowicz, Prior and contextual emotion of words in sentential context, Comput. Speech Lang. 28(1) (2014) 76–92.
[10] B. Pang, L. Lee, Opinion mining and sentiment analysis, Found. Trends Inform. Retrieval 2(1–2) (2008) 1–135.
[11] T. Zagibalov, J. Carroll, Automatic seed word selection for unsupervised sentiment classification of Chinese text, in: Proceedings of the 22nd International Conference on Computational Linguistics, vol. 1, Association for Computational Linguistics, 2008, pp. 1073–1080.
[12] P.D. Turney, Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews, in: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, 2002, pp. 417–424.
[13] H. Kanayama, T. Nasukawa, Fully automatic lexicon expansion for domain-oriented sentiment analysis, in: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2006, pp. 355–363.
[14] G. Qiu, B. Liu, J. Bu, C. Chen, Opinion word expansion and target extraction through double propagation, Comput. Linguist. 37(1) (2011) 9–27.
[15] S.M. Kim, E. Hovy, Identifying and analyzing judgment opinions, in: Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Association for Computational Linguistics, 2006, pp. 200–207.
[16] J. KAMPS, Using wordnet to measure semantic orientation of adjectives, in: Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), 2004, pp. 1115–1118.
[17] D. Rao, D. Ravichandran, Semi-supervised polarity lexicon induction, in: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, 2009, pp. 675–682.
[18] J. Baldridge, M. Osborne, Active learning and the total cost of annotation, in: EMNLP, 2004, pp. 9–16.
[19] S. Tong, D. Koller, Support vector machine active learning with applications to text classification, J. Mach. Learn. Res. 2 (2002) 45–66.
[20] H.T. Nguyen, A. Smeulders, Active learning using pre-clustering, in: Proceedings of the Twenty-First International Conference on Machine Learning, ACM, 2004, p. 79.
[21] J. Kang, K.R. Ryu, H.C. Kwon, Using cluster-based sampling to select initial training set for active learning in text classification, in: Proceedings of 8th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD 2004), Sydney, Australia, May 26–28, 2004, Springer, 2004, pp. 384–388.
[22] C. Potts, On the negativity of negation, in: Proceedings of SALT, vol. 20, 2011, pp. 636–659.

[23] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques, in: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, vol. 10, Association for Computational Linguistics, 2002, pp. 79–86.

[24] S. Dasgupta, V. Ng, Mine the easy, classify the hard: a semi-supervised approach to automatic sentiment classification, in: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL, vol. 2; Proceedings of the 4th International Joint Conference on Natural Language Processing of the AFNLP, vol. 2, Association for Computational Linguistics, 2009, pp. 701–709.

[25] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.

[26] B. Agarwal, N. Mittal, Optimal feature selection for sentiment analysis, in: Computational Linguistics and Intelligent Text Processing, Springer, 2013, pp. 13–24.

[27] J. Blitzer, M. Dredze, F. Pereira, Biographies, bollywood, boom-boxes and blenders: domain adaptation for sentiment classification, in: Annual Meeting-Association For Computational Linguistics, vol. 45, 2007, p. 440.