Expert Systems With Applications 116 (2019) 285–298

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Automatic construction of target-specific sentiment lexicon

Sixing Wu[a,*], Fangzhao Wu[b], Yue Chang[c], Chuhan Wu[a], Yongfeng Huang[a]

[a] Department of Electronic Engineering, Tsinghua University, Beijing, China
[b] Microsoft Research Asia, China
[c] Department of Telecommunications Engineering with Management, Beijing University of Posts and Telecommunications, Beijing, China

ARTICLE INFO

Article history:
Received 14 June 2018
Revised 21 August 2018
Accepted 10 September 2018
Available online 13 September 2018

Keywords:
Opinion mining
Sentiment lexicon
Sentiment analysis
Opinion target extraction

ABSTRACT

Sentiment lexicon plays an important role in sentiment analysis system. In most existing sentiment lexica, each sentiment word or phrase is given a sentiment label or score. However, a sentiment word may express different sentiment orientations describing different targets. It's beneficial but challenging to incorporate knowledge of opinion targets into sentiment lexicon. In this paper we propose an automatic approach to construct a target-specific sentiment lexicon, in which each term is an opinion pair consisting of an opinion target and an opinion word. The approach solves two principle problems in construction process, namely, opinion target extraction and opinion pair sentiment classification. An unsupervised algorithm is proposed to extract opinion pairs in high quality. Both semantic feature and syntactic feature are incorporated in the algorithm, to extract opinion pairs containing correct opinion targets. A group of opinion pairs are generated and a framework is proposed to classify their sentiment polarities. Knowledge of available resources including general-purpose sentiment lexicon and thesaurus, and context knowledge including syntactic relations and sentiment information in sentences, are extracted and integrated in a unified framework to calculate sentiment scores of opinion pairs. Experimental results on product reviews datasets in different domains prove the effectiveness of our method in target-specific sentiment lexicon construction, which can improve performances of opinion target extraction and opinion pair sentiment classification. In addition, our lexicon also achieves better performance in target-level sentiment classification compared with several general-purpose sentiment lexicons.

1. Introduction

Sentiment lexicon, which contains a list of words or phrases as well as their sentiment labels or scores, is the most important semantic resource in sentiment analysis (Dey, Jenamani, & Thakkar, 2018; Liu & Zhang, 2012). There have been a variety of approaches to constructing sentiment lexica manually (Hu & Liu, 2004; Wilson, Wiebe, & Hoffmann, 2005) or automatically (Neviarouskaya, Prendinger, & Ishizuka, 2009; Tan & Wu, 2011; Tang, Wei, Qin, Zhou, & Liu, 2014; Wu, Huang, Song, & Liu, 2016). A great number of domain-specific sentiment words change their semantic orientations across domains, e.g., "easy" usually conveys positive sentiment in *Electronics* domain (e.g., The system is easy to install.), while expresses negative sentiment in *Movie* domain (e.g., The ending of the movie is easy to guess.). As a result, constructing domain-specific sentiment lexica automatically has become a hot research topic.

However, semantic orientations of sentiment words could change even in a specific domain, which are called context-dependent sentiment words (Liu, 2015). For instance, "high" is positive when describing "definition" of a screen but has negative polarity when modifying "price" of a product. Such phenomenon shows that both the opinion words and targets are important to indicate sentiments. In most sentiment lexica which assign a polarity or score to each sentiment word or phrase, information of opinion target is ignored, leading to the absence of the most important context-dependent knowledge.

Existing domain-specific sentiment lexicon construction methods cannot be applied directly to construct target-specific sentiment lexicon because of two reasons. First, opinion targets, which are usually entities or attributes of entities (Liu, 2015), play an important part in target-specific sentiment lexicon. However, opinion targets are different across domains. For example, "computer" and "phone" are targets in *Electronics* domain, while "knife" and "spoon" are targets in *Kitchen* domain. It's labor-intensive to manually label opinion targets in each domain. Therefore, it's an

* Corresponding author.
E-mail addresses: wu-sx15@mails.tsinghua.edu.cn (S. Wu), wufangzhao@gmail.com (F. Wu), changyue@bupt.edu.cn (Y. Chang), wuch15@mails.tsinghua.edu.cn (C. Wu), yfhuang@mail.tsinghua.edu.cn (Y. Huang).

https://doi.org/10.1016/j.eswa.2018.09.024
0957-4174/

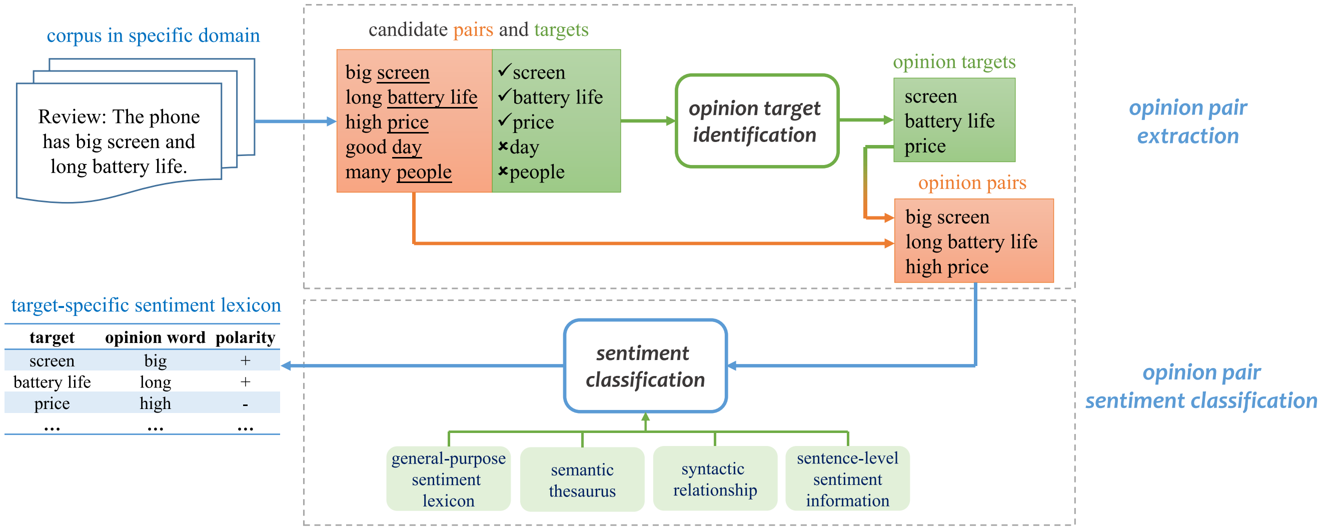


Fig. 1. Overall framework of target-specific sentiment lexicon construction method.

important issue to extract opinion targets automatically. Second, sentiment polarities of some opinion words depend on opinion targets of them. For example, “low risk” is positive while “low quality” is negative. Thus it’s also important to extract pairs of opinion words and targets, and classify their sentiment polarities. Most domain-specific sentiment lexicon construction methods are incapable to solve these two problems. As a result, an automatic approach to build a sentiment lexicon incorporating information of opinion target is beneficial to sentiment analysis.

There are some proposed methods focusing on context-dependent sentiment word polarity detection or target-dependent sentiment lexicon construction (Ding, Liu, & Yu, 2008; Fahrni & Klenner, 2008; Lu, Castellanos, Dayal, & Zhai, 2011; Zhang, Zhang, Zhang, Liu, & Ma, 2014). Although effective, there are still some limitations of previous studies. Firstly, some of the methods supposed that the opinion targets are manually defined (Lu et al., 2011; Zhang et al., 2014), which makes it inconvenient for their methods to be applied to other domains. Some others used existing tools such as Wikipedia’s category system[1], which may ignore opinion targets not in the system. Some methods were proposed only to extract opinion targets based on syntactic patterns or frequencies (Hu & Liu, 2004; Liu, Liu, Zhang, Kim, & Gao, 2016; Qiu, Liu, Bu, & Chen, 2011), which could ignore opinion targets in low frequencies or not in specific syntactic patterns, and extract incorrect targets in high frequencies. Secondly, opinion pairs are not classified in high accuracy. Some of previous methods used linguistic rules such as conjunctions (Ding et al., 2008; Fahrni & Klenner, 2008) to propagate sentiment information, which couldn’t classify opinion pairs not connected by conjunctions. Other methods made use of information of available resources, such as general-purpose sentiment lexica and reviews sentiment scores (Lu et al., 2011; Zhang et al., 2014). However, labeled reviews are not available in many cases. In addition, sentiment label of a review indicates overall semantic orientation of the document, while there may exist opinion pairs with opposite sentiment polarities. As a result, the applicabilities and performances of previous methods are still unsatisfactory.

In order to solve the above limitations, in this paper we propose a framework to construct a target-specific sentiment lexicon automatically. The overall framework of our construction method is illustrated in Fig. 1. Each item in our lexicon is a pair of *(opinion word, opinion target)* with a sentiment label. Our method solves two main problems in construction process, i.e., opinion target extraction and sentiment classification of opinion pairs. We propose an automatic method to identify opinion targets from candidate nouns and noun phrases. An one-class SVM classifier is trained based on word embeddings of a set of seed opinion targets in high accuracy, through which semantic features are incorporated, and thus infrequent opinion targets can be identified correctly. In addition, a rule-based bootstrapping method is proposed to expand the group of opinion targets iteratively to improve the coverage.

To solve the problem of inaccuracy of opinion pair sentiment classification caused by sparsity of opinion pairs and insufficiency of prior knowledge, we propose a unified framework fusing four kinds of sentiment information to calculate sentiment scores of opinion pairs. The first is lexicon information extracted from general-purpose sentiment lexicon. The second is semantic thesaurus information generated from existing thesaurus, in which words are linked to each other according to semantic relations (Miller, 1995). The third is syntactic information extracted by dependency parser. The fourth is sentence-level sentiment information, extracted with an unsupervised method to detect sentiment polarities of sentences containing opinion pairs. Four kinds of information are integrated in a unified framework to calculate sentiment scores of opinion pairs through solving an optimization problem. Our method is motivated by previous work (Lu et al., 2011), but incorporates more information and is not in need of labeled reviews, and thus could be applied to other domains effectively.

We conduct experiments on product reviews in different domains to evaluate the effectiveness of our method. Experimental results have proved that our opinion target extraction method outperforms other unsupervised baselines in achieving both high accuracy and recall. Moreover, our sentiment classification method is able to predict sentiment polarities of opinion pairs in high accuracy. In addition, performances of different lexicons on target-level sentiment classification are evaluated, and our target-specific sentiment lexicon has the best performance compared with several general-purpose sentiment lexicons.

The major contributions of our work are as follows:

- We propose an unsupervised opinion target extraction method, which utilizes syntactic relations and semantic similarities to identify opinion targets in high accuracy and recall.

[1] https://en.wikipedia.org/wiki/Portal:Contents/Categories

- We propose an effective method to extract and fuse different sources of information to classify sentiment polarities of opinion pairs.
- We propose an overall framework to construct target-specific sentiment lexica in different domains automatically, through integrating our proposed methods of opinion target extraction, opinion pair extraction and opinion pair sentiment classification.

The rest of the paper is organized as follows. We introduce the related work in Section 2. The methods of opinion pairs extraction and opinion pairs sentiment classification are presented in Section 3 and Section 4 respectively. In Section 5, we report the experimental results. A conclusion is made in Section 6.

2. Related work

2.1. Sentiment lexicon construction

Sentiment lexicon plays an important role in sentiment analysis, therefore construction of a high-quality sentiment lexicon has attracted much research interests (Fernández-Gavilanes, Juncal-Martínez, García-Méndez, Costa-Montenegro, & González-Castaño, 2018; Neviarouskaya et al., 2009; Qiu, Liu, Bu, & Chen, 2009; Tang et al., 2014). As the most straightforward and reliable way, manual approaches have created a large number of trustworthy sentiment lexica which are widely utilized in sentiment analysis systems, such as MPQA (Wilson et al., 2005), LIWC (Pennebaker, Francis, & Booth, 2001), Hu and Liu's Lexicon (Hu & Liu, 2004), VADER (Gilbert, 2014) and so on. In spite of the high accuracy, manually constructed sentiment lexica suffer from a relatively low coverage. In addition, the manually constructing process is time-consuming and labor-intensive.

In order to automatically create a sentiment lexicon in high quality, a substantial amount of approaches were proposed. Most of them manually collected a small group of seed words (e.g. good, bad, etc.), and made use of heterogeneous sources of information in corpus to propagate sentiment information from seed words, such as syntactic rules (Qiu et al., 2011; Wang & Wang, 2008; Zhuang, Jing, & Zhu, 2006), morphological relations (Hatzivassiloglou & McKeown, 1997; Huang, Niu, & Shi, 2014), co-occurrences (Kanayama & Nasukawa, 2006; Tan & Wu, 2011; Turney, 2002) and so on. Qiu et al. (2009, 2011) proposed to extract opinion words and targets based on seed words and a group of syntactic rules. For example, an adjective will be assigned positive polarity if connected to another positive sentiment word with conjunction "and". Huang et al. (2014) exploited contextual and morphological relations between sentiment terms to measure their semantic associations, and constructed a similarity graph to propagate sentiment labels to domain-specific sentiment words. Fernández-Gavilanes et al. (2018) extracted sentiment information in sentences in an unsupervised way, considering both sentiment polarities of opinion words and syntactic features of words, to create an emoji lexicon.

Previous methods have been proved effective in assigning sentiment labels or scores to sentiment terms. Most of them take sentiment words as sentiment terms. However, semantic orientations of some sentiment words could be variational even in one domain. For example, in *Electronics* domain, "long battery life" conveys positive sentiment while "long response time" expresses negative sentiment, and similar words include "high", "small", "quiet" and so on. Thus it's still limited to construct sentiment lexicon in absence of opinion target information. In order to handle this problem, some methods were proposed to construct context-dependent or target-dependent sentiment lexica. Ding et al. (2008) dealt with context-dependent opinion words based on intra-sentence and inter-sentence conjunction rules. They assigned sentiment polarities to opinion words according to specific opinion targets. However, their method was just applied to determine sentiment polarities of opinion targets, instead of creating a sentiment lexicon. In addition, their method was rule-based, which may cause inaccuracy in sentences with informal syntactic structures. Lu et al. (2011) constructed a context-aware sentiment lexicon in which each term is a pair of opinion word and opinion target. They determined sentiment polarities of opinion pairs through solving an optimization problem, fusing different sources of information. However, they assumed that opinion targets were manually selected, which makes it inconvenient for their method to be applied to other domains. Fahrni and Klenner (2008) identified the targets by querying Wikipedia's category system and determined the semantic orientations of opinion pairs depending on conjunctions. However, opinion targets in corpus may not always be in Wikipedia's categories. In addition, their sentiment detection method relied on conjunctions, which may ignore opinion pairs not connected to other opinion words by conjunctions.

Our research combines opinion target extraction and sentiment lexicon construction to generate a target-specific sentiment lexicon, which differs from existing methods in two phases. First, we propose an unsupervised method to identify opinion targets automatically, which solves the difficulty of labeling opinion targets manually in different domains. Second, we propose a framework to incorporate different kinds of information in texts to classify sentiment polarities of opinion pairs, which achieves better performance than existing methods.

2.2. Opinion target extraction

Opinion target extraction is an important task in sentiment analysis and has been widely studied. According to reliance on labeled data, existing approaches can be categorized into supervised (Jakob & Gurevych, 2010; Liu, Xu, Liu, & Zhao, 2013; Liu, Joty, & Meng, 2015; Yang & Cardie, 2013; Yin et al., 2016; YING, Yu, & Jiang, 2017) and unsupervised (Hu & Liu, 2004; Liu et al., 2016; Popescu & Etzioni, 2007; Qiu et al., 2011).

In most supervised approaches, the extraction task was viewed as a sequence labeling problem, to predict whether each word or phrase in a sentence is an opinion target or not (Liu et al., 2013). Jakob and Gurevych (2010) and Yang and Cardie (2013) implemented Conditional Random Field (CRF) to solve the problem using multiple linguistic features. Different neural network models have been proposed to automatically learn features for opinion targets extraction from training data (Liu et al., 2015; Yin et al., 2016; YING et al., 2017). Supervised approaches rely heavily on large amount of labeled data, which is unavailable in most cases. Our research focuses on unsupervised opinion target extraction.

Various unsupervised opinion target extraction methods have been proposed, including frequency-based methods, syntax-based methods and topic modeling methods. Frequency-based methods depend on the observation that frequent nouns or noun phrases have higher probabilities to be opinion targets (Hu & Liu, 2004; Popescu & Etzioni, 2007). Hu and Liu (2004) and Popescu and Etzioni (2007) regarded nouns and noun phrases with high appearances as opinion targets, and pruned redundant nouns. Although effective and powerful, frequency-based methods may mistakenly identify nouns with high frequencies as opinion targets, such as "weather" and "day". In addition, they may ignore opinion targets with low frequencies. Syntax-based methods extract opinion targets according to their syntactic relations. Qiu et al. (2011) proposed a bootstrapping method extracting opinion words and opinion targets based on a list of syntactic rules. Liu et al. (2016) proposed two recommendation rules to improve the performance of syntactic method in previous research (Qiu et al., 2011). However, some incorrect opinion targets could be extracted as well. For example, it's an important rule that if a noun is modified by a sentiment word, it will be extracted as an opinion target. In the sentence "I will spend the boring weekend on watching the new TV", "weekend" is modified by a negative word "boring", but is not an opinion target.

Topic models have also been applied in opinion target extraction. Different from frequency-based and syntax-based methods, topic models could be used to categorize opinion targets with similar meaning into a same aspect (Brody & Elhadad, 2010; Chen, Mukherjee, & Liu, 2014; He, Lee, Ng, & Dahlmeier, 2017; Mukherjee & Liu, 2012). As a result, they mainly focus on the topic coherences, and precisions of representative words in different aspects, which are different from our task. However, representative words in each aspect are usually important opinion targets, thus topic models could be effective methods of opinion target extraction. Similar with frequency-based methods, infrequent opinion targets are usually ignored by topic models, which leads to the low recall of extraction.

Different from above opinion target extraction approaches, our method trains a classifier to identify opinion targets from candidate nouns and noun phrases. Word vectors trained from unlabeled corpus are used to construct word features, which are not influenced by the frequencies of words. As a result, our method can extract opinion targets with low frequencies, and reduce mistakes caused by incorrect frequent nouns. In addition, we propose a rule-based extraction method to expand the group of opinion targets iteratively based on syntactic relations. Experimental results have proved the effectiveness of our method.

3. Opinion pair extraction

Each element in our target-specific sentiment lexicon is a pair consisting of an opinion target and an opinion word. Thus, the construction of our sentiment lexicon contains two main steps. First, extract opinion pairs from a domain-specific corpus. Second, assign sentiment polarities to opinion pairs. The extraction method will be discussed in this section.

Opinion targets in our lexicon are entities or their aspects, such as features or attributes, as described in previous research (Ding et al., 2008; Liu, 2015; Lu et al., 2011). If an opinion word is used to express sentiment upon an opinion target in the corpus, they form an opinion pair in the lexicon. Most of opinion targets are nouns or noun phrases. However, not all nouns described by opinion words should be extracted as opinion targets. For example, "day" could be frequently described by sentiment words such as "good" in *Electronics* domain, but is not an opinion target. Thus, we extract opinion pairs in a two-step process. First, a group of candidate opinion pairs are extracted based on dependency grammar. Second, a group of opinion targets are identified from candidates. Only candidate opinion pairs containing opinion targets are assigned sentiment labels and put into the lexicon. An overall illustration of the opinion pair extraction method is shown in Fig. 2, with detailed explanation in the following subsections.

3.1. Preprocessing

Preprocessing is a significant process in target-specific sentiment lexicon construction, because of the importance of syntactic information in opinion pairs extraction. In our method, the preprocessing consists of three steps. Firstly, each sentence in reviews is analyzed by a dependency parser to extract dependency relations between words. Stanford Parser[2] is used. Secondly, parts of speech (POS) of words are assigned, with the help of Stanford POS Tagger. Thirdly, Stanford Lemmatizer is applied to nouns after the parsing step, in order to aggregate nouns and their plural forms, to reduce the sparsity caused by different forms of a same opinion target. Details of making use of syntactic information in opinion pair extraction are explained in the following subsections.

[2] https://nlp.stanford.edu/software/lex-parser.shtml

3.2. Candidate opinion pair generation

Sentiment words are usually used to describe opinion targets. Thus, syntactic relations are important information for opinion pairs extraction. Instead of taking the target's nearest adjective as sentiment word (Hu & Liu, 2004), syntactic dependencies are used in our method. For example, syntactic dependency of the sentence "The phone has a beautiful screen" can be represented as Fig. 3:

The adjective "beautiful" is used to describe the noun "screen", so "screen" and "beautiful" are extracted as opinion target and opinion word respectively.

Motivated by previous works (Ding et al., 2008; Hu & Liu, 2004; Qiu et al., 2011), opinion words are assumed to be adjectives and opinion targets are considered to be nouns or noun phrases in our lexicon. We extract dependency relations between adjectives and nouns based on three rules, which have been proved effective in extracting opinion words and opinion targets by Qiu et al. (2011):

1. If an adjective and a noun are related directly, they are extracted as opinion word and opinion target. (e.g. The screen is big. $screen \xrightarrow{nsubj} big \xRightarrow{rule1} big \leftrightarrow screen$).
2. If an adjective and a noun are both related to a third word, they are extracted as opinion word and opinion target. (e.g. The price gets high. $price \xrightarrow{nsubj} gets \xleftarrow{xcomp} high \xRightarrow{rule2} high \leftrightarrow price$).
3. If two nouns are connected with a conjunction and one of them is related to an adjective, both of them are extracted as opinion targets with the adjective as opinion word. Similarly, two coordinate adjectives with one of which related to a noun will be extracted in the same way. (e.g. The keyboard and mouse are comfortable. $keyboard \xrightarrow{nsubj} comfortable, keyboard \xrightarrow{conj} mouse \xRightarrow{rule3} comfortable \leftrightarrow keyboard, comfortable \leftrightarrow mouse$).

A set of candidate pairs are generated with the proposed rules, consisting of candidate opinion targets and opinion words. However, these rules will identify all nouns and noun phrases related to adjectives as opinion targets, while some of them are not. Therefore, we propose a method to select opinion targets from all candidate opinion targets.

3.3. Opinion target identification

Without prior knowledge of specific domain, it's challenging to distinguish opinion targets from non-targets. Low coverage of frequency-based methods and low accuracy of syntax-based methods show the limitation of frequency and syntax information. Neither of them considers semantic knowledge, which is significant in opinion target extraction. For example, in *Electronics* domain, "screen" is a frequent opinion target, as well as "definition" should be an opinion target, because it has strong semantic relation with "screen", although its frequency is low. "Weather" shouldn't be an opinion target, because it's not related to *Electronics* domain, although it's described by opinion words sometimes. Thus, we propose an algorithm to extract opinion targets incorporating semantic knowledge and syntactic knowledge. A set of seed opinion targets are extracted, to train a classifier to identify opinion targets from candidates, followed by a syntax-based extraction method to expand the group of opinion targets iteratively. The algorithm is introduced in Algorithm 1, which is explained in detail in the following.

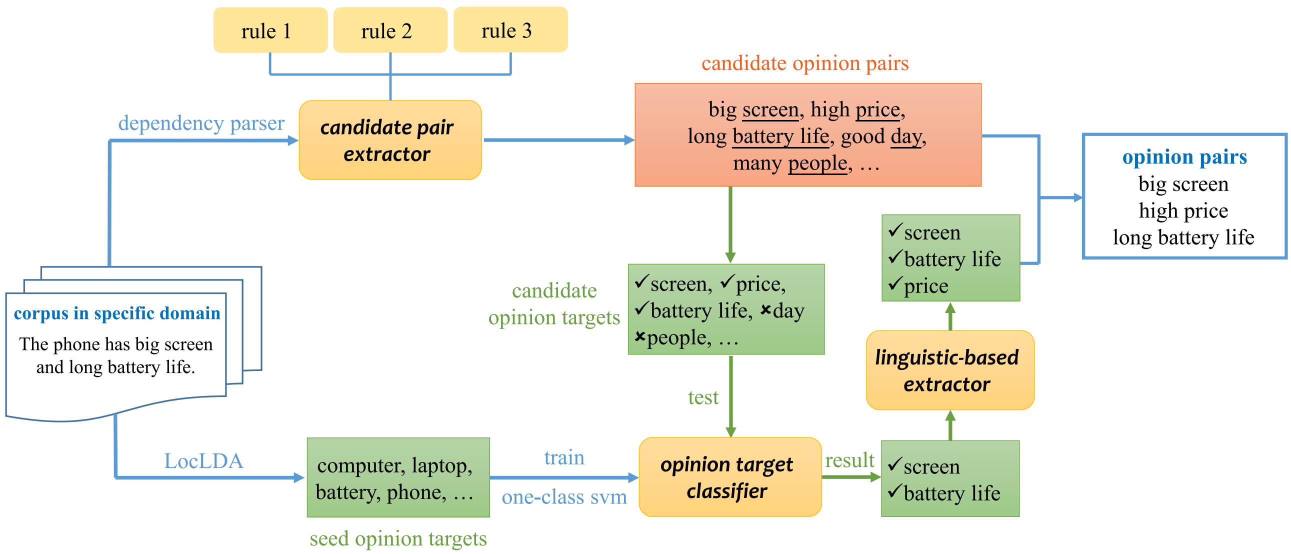


Fig. 2. Overall framework of opinion pair extraction method.

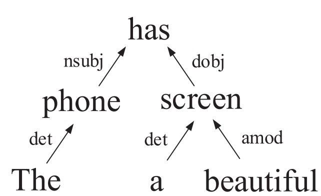


Fig. 3. Dependency tree of the sentence “The phone has a beautiful screen”.

Algorithm 1 Opinion pairs extraction algorithm.

Input:
Product review dataset **D**, dependency parser **P**.
Output:
A group of opinion targets T and opinion pairs OP.
Initialize T and OP as empty sets.
Parse sentences in **D** with **P**.
Extract candidate opinion pairs OP and candidate opinion targets C_T based on parsing results.
Train word embeddings $v_1, v_2, \ldots, v_n$ of words in **D**.
Extract a group of seed opinion words S_T with topic model based method LocLDA.
Train an one-class SVM on S_T with features constructed based on $v_1, v_2, \ldots, v_n$.
Identify a group of opinion targets T_{SVM} from C_T with one-class SVM.
Set $T = S_T + T_{SVM}$.
Extract new opinion targets N_T with rule-based extraction method.
while $N_T \neq \varnothing$ **do**
  Set $T = T + N_T$.
  Extract new opinion targets N_T with rule-based extraction method.
end while
Filter candidate opinion pairs in OP which do not contain opinion targets in T.
return Opinion targets T, and a group of opinion pairs OP.

Although suffers from incapability of extracting infrequent aspects accurately, topic model has high precision in extracting top representative words in each aspect (Brody & Elhadad, 2010; Chen et al., 2014; Mukherjee & Liu, 2012). Therefore, in our method, a small group of seed opinion targets are extracted by topic model.

We employ LocLDA (Brody & Elhadad, 2010) to reviews, which is an aspect extraction method designed for review texts based on standard implementation of LDA. It's proved able to effectively extract aspects and representative words in aspects for review texts through treat each sentence as a document, in order to extract local topics rather than global topics in reviews. The number of aspects are set to t and top n ranking words in each aspect are extracted as representative words.

Top representative words are strongly correlated to aspects of review texts, and have high probabilities to be opinion targets in corresponding domain. But there could be global topic words identified as representative words, such as “quality”, “product”. Although they are opinion targets as well, they are not domain-specific, which could have negative influence on training of the classifier. Therefore, when there are more than one domain, a word will be extracted as opinion target when it is identified as representative word only in one domain. In this way, a small group of seed opinion targets S_T are generated.

To expand the set of seed opinion targets S_T, we use S_T as training data to learn an one-class SVM (Schölkopf, Platt, Shawe-Taylor, Smola, & Williamson, 2001), to classify each word in candidate sets C_T into target or non-target. The motivation of applying one-class SVM lies in the difficulty to obtain negative samples, and easy access to a small group of positive samples. As mentioned before, if a candidate target has high semantic similarities with seed opinion words, it has high probability to be an opinion target as well. Therefore, the one-class SVM classifier is trained to learn semantic features of seed opinion targets. Candidate targets which have similar semantic meanings with seed opinion targets will be classified as in-class data points (targets), and others as outliers (non-targets). Because the classifier is trained based on semantic features, it's not influenced by frequency and syntactic patterns, thus is able to extract infrequent opinion targets.

Word vector has shown great performances in capturing semantic word relationships and measuring similarities between words (Liu et al., 2016; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). Semantically related words usually have close word vectors according to cosine distance. As a result, we construct the features of S_T and C_T based on word vectors. Instead of using publicly available resources of word vectors, we train our own word vectors with review corpus in each domain to generate domain-specific word representations. The features are constructed based on the following observations:

- If a word has a high semantic similarity with an opinion target, it is more probably an opinion target as well.

For example, if “laptop” is an opinion target, “computer” should be an opinion target as well. However, there could be mistakes in the opinion targets extracted by topic model. If a negative sample exists in the group of seed opinion targets, more incorrect targets could be extracted. Therefore, a revised observation is made:

- If a word is an opinion target, it should be semantically related to a group of opinion targets.

In reviews of a domain, domain-specific opinion targets are usually semantically correlated to each other. For example, “mouse”, “keyboard” and “laptop” are related, same as “software”, “application” and “program”. If a non-target is extracted incorrectly, it won’t influence the final result as long as most of the seed opinion targets are correct.

Based on the observations proposed, we represent features of each target in seed sets S_T and candidate sets C_T with a 2-dim vector. The number of words in S_T is defined as M, and number of words in C_T is defined as N. So, s_i is the ith opinion target in S_T and $\boldsymbol{f_{si}}$ is the feature of s_i. Respectively, c_i is the ith candidate target in C_T and $\boldsymbol{f_{ci}}$ is the feature of c_i. v_w represents the vector of the word w. The feature of s_i is constructed as follows:

$$\boldsymbol{f_{si}} = [f_{si1}, f_{si2}], \tag{1}$$

$$f_{si1} = \max_{j \neq i} \frac{v_{s_i}^T v_{s_j}}{\|v_{s_i}\| \left\| v_{s_j} \right\|}, \tag{2}$$

$$f_{si2} = \frac{1}{k} \sum_{j=1}^{k} \frac{v_{s_i}^T v_{sim_j}}{\|v_{s_i}\| \left\| v_{sim_j} \right\|}, \tag{3}$$

where $sim_1, sim_2, \ldots, sim_k$ are the top-k words in S_T whose vectors have the largest similarities with the vector of s_i, except s_i itself.

The feature of candidate target c_i is constructed as follows:

$$\boldsymbol{f_{ci}} = [f_{ci1}, f_{ci2}], \tag{4}$$

$$f_{ci1} = \max_{j} \frac{v_{c_i}^T v_{s_j}}{\|v_{c_i}\| \left\| v_{s_j} \right\|}, \tag{5}$$

$$f_{ci2} = \frac{1}{k} \sum_{j=1}^{k} \frac{v_{c_i}^T v_{sim_j}}{\|v_{c_i}\| \left\| v_{sim_j} \right\|}, \tag{6}$$

where $sim_1, sim_2, \ldots, sim_k$ are the top-k words in S_T whose vectors have the largest similarities with the vector of c_i.

There are some noun phrases in the candidate targets, whose vectors are sum of vectors of words in the phrase.

An one-class SVM (Schölkopf et al., 2001) is trained with S_T as the training set, based on the constructed features. The candidate targets are classified with the SVM into positive (targets) or negative (non-targets). Candidate targets labeled positive by the classifier will be added into the set of opinion targets.

However, some general opinion targets might be ignored. Because most of opinion targets in the training set are domain-specific, which are not semantically related to general opinion targets. For example, “price” is an important opinion target, but has low similarities with domain-specific targets in *Electronics* domain. To handle this problem, we propose a simple but effective extraction rule:

- A candidate target will be identified as opinion target if it is conjuncted with an opinion target in a sentence, and they are both modified by sentiment words.

Some examples are shown in Table 1. The rule is based on the observation that in product reviews, people usually use sentiment words to describe opinion targets. Moreover, different aspects of products are usually connected by conjunctions. Because rule-based extraction method could introduce mistakes into the result, we set a strict extraction rule to maintain high accuracy. The rule-based extraction is performed iteratively to expand the group of opinion targets, until no more opinion targets are extracted. We prove effectiveness of the extraction method in the experiment section.

The overall algorithm is summarized in Algorithm 1. Opinion targets are extracted through the proposed algorithm. If the noun of a candidate pair is an opinion target, it is identified as opinion pair. For example, “big screen” is identified as an opinion pair in *Electronics* domain, while “good weather” is not. Next, we describe the method to determine sentiment polarities of opinion pairs.

4. Sentiment polarity detection of opinion pairs

Sentiment polarity of opinion pair is determined by both opinion target and opinion word. However, it’s difficult to classify sentiment polarities of opinion pairs, because frequencies of opinion pairs are much lower than opinion words. On the one hand, texts may use only opinion words instead of opinion pairs to express sentiments. On the other hand, opinion pairs with a same opinion word are different with each other, if they have different opinion targets. Low frequencies of opinion pairs cause the problem of insufficiency of sentiment information, which makes it difficult to determine sentiment polarities of opinion pairs accurately. As a result, we propose a framework fusing heterogeneous sources of information to calculate sentiment scores of opinion pairs, i.e., general-purpose sentiment lexicon information, thesaurus information, syntactic information and sentence-level sentiment information. Overall framework of our method is illustrated in Fig. 4.

Our work is motivated by previous work (Lu et al., 2011), which combined different sources of information in a unified framework. However, their method needs sentiment labels at the document level in specific domain, which is difficult to obtain in each domain. Our framework incorporates more comprehensive knowledge in a different way, in which no domain-specific sentiment label is required.

We describe each kind of information in detail first, and propose the overall framework, which assigns a sentiment score s to each opinion pair.

4.1. General-purpose sentiment lexicon information

In general-purpose sentiment lexica, sentiment polarities are only assigned to general sentiment words, which have consistent semantic orientations. For example, “great” always conveys positive semantic orientation, thus it is more likely for pairs which have “great” as their opinion words to have positive sentiment polarities. Therefore, if an opinion pair has a general-purpose sentiment word as opinion word, its sentiment polarity should be the same as the opinion word.

4.2. Thesaurus information

Thesaurus connects words according to semantic relations, in which synonyms and antonyms are grouped together. We use WordNet[3] as the thesaurus in our method, which groups words

[3] https://wordnet.princeton.edu/

Table 1
Examples of extraction rules in opinion target extraction.

Opinion targets	Sentence	Extracted targets
Sound	Sound is good and the sensitivity is great.	Sensitivity
Microphone	The microphone and recording unit are quite good.	Recording unit

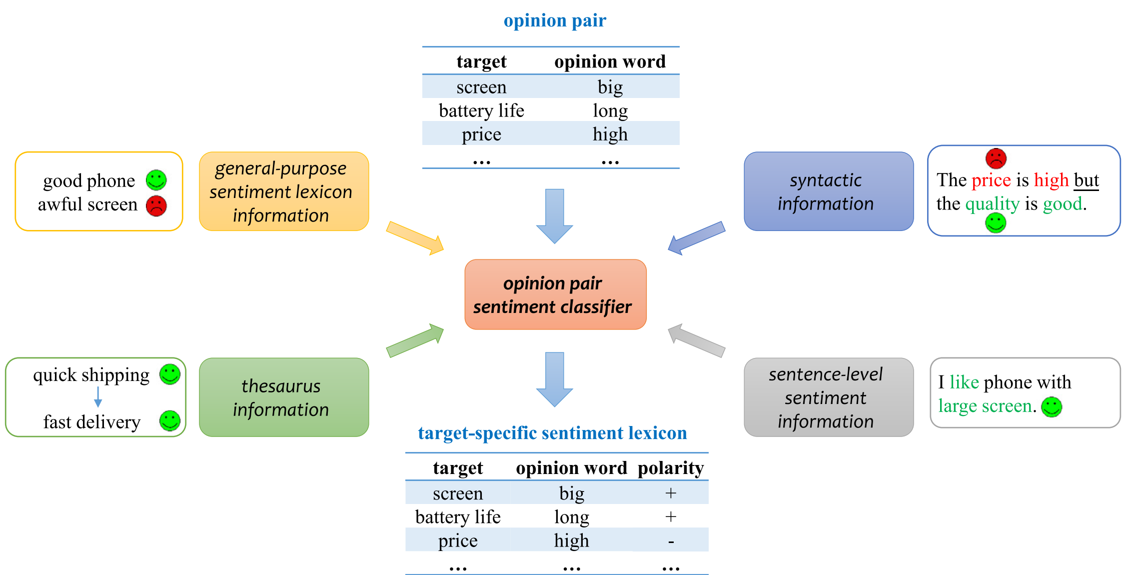


Fig. 4. Overall framework of opinion pair sentiment classification method.

into sets of cognitive synonyms, and adjectives are also organized in terms antonym. For example, the relationship between "big" and "small" is antonym in thesaurus, and "big" and "huge" are synonyms. It has been proved to be useful to find new sentiment words in many previous works (Baccianella, Esuli, & Sebastiani, 2010; Kamps, Marx, Mokken, De Rijke et al., 2004). It is assumed that synonyms have same sentiment polarities and antonyms have opposite sentiment polarities. However, semantic relation between two opinion pairs is influenced by both opinion targets and opinion words.

We assume that if two pairs' opinion targets are synonyms and opinion words are synonyms respectively, they tend to have the same sentiment polarity. On the other hand, if two pairs' opinion targets are synonyms and opinion words are antonyms, they tend to have opposite sentiment polarities. For example, if "fast delivery" is classified into positive polarity, "quick shipping" should be positive as well, and "slow shipping" should be negative.

4.3. Syntactic information

Syntactic information is helpful to extract contextual sentiment relations between opinion words. Coordinate conjunctions such as "and" usually connect words with consistent sentiment polarities, while words connected by adversative conjunctions such as "but" usually have opposite sentiment polarities. It's more complex to extract syntactic relations between opinion pairs. We define two cases where syntactic information can be extracted. First, the opinion targets or the opinion words are the same. For example, "the screen is large and bright", and "both the keyboard and mouse are comfortable". Second, the opinion targets and opinion words are both different. For example, "the screen is large but the battery capacity is small". In these cases, the connected opinion pairs have high probabilities to have same or opposite sentiment polarities.

To denote the consistency and reverse of sentiment orientations extracted by conjunctions, two matrices, $\mathbf{C} \in \mathbb{R}^{n \times n}$ and $\mathbf{A} \in \mathbb{R}^{n \times n}$, are used. $C_{ij} = \frac{S^c_{ij}}{S^c_{ij}+S^a_{ij}+\alpha}$, where S^c_{ij} and S^a_{ij} denote the times the ith and jth opinion pairs connected by coordinate and adversative conjunctions, respectively. α is used for smoothing, which we set to 5. Similarly, $A_{ij} = \frac{S^a_{ij}}{S^c_{ij}+S^a_{ij}+\alpha}$. If C_{ij} (or A_{ij}) is large, the ith and jth opinion pairs are likely to have same (or opposite) sentiment orientations.

Although syntactic information is useful, it can only extract sentiment information when conjunctions are used. Thus, we introduce another method to extract contextual sentiment information.

4.4. Sentence-level sentiment information

Conjunction contains short-term contextual sentiment information in most cases. However, there are other ways to express sentiments, in which no conjunction is used and the sentence structures could be more complex. For example, in sentence "I like the phone I bought yesterday, because of its large screen, which has high definition.", there are two opinion pairs, i.e. "large screen" and "high definition", none of which have general sentiment words. But we can infer their sentiment polarities from the word "like", although it has a long distance from the opinion pairs. As a result, a method is proposed to capture long-term contextual sentiment information, which we call sentence-level sentiment information.

For each sentence in review corpus which contains one or more opinion pairs, its sentiment polarity will be classified with general-purpose sentiment lexicon based on the following rules. (1) If all general sentiment words in a sentence convey the same sentiment orientation, and no negations or adversative conjunctions appear, we assign sentence the same sentiment polarity as sentiment words in it. (2) If there's only one adversative conjunction in a sentence, the sentence will be split by the conjunction. Each part is viewed as an independent sentence, and (1) is applied to both of them. If the first part has sentiment words with consistent sentiment polarity, and the second has no general sentiment word, then

Table 2
Examples of rules in inferring sentence-level sentiment polarity.

	Example	Opinion pair	Polarity
First rule	It's a great advantage that this phone has a big screen.	Big screen	Positive
Second rule	I like the T-shirt, but the delivery is quite slow.	Slow delivery	Negative

it will be assigned opposite sentiment polarity to the first part, and vice versa. Examples of the two rules are given in Table 2. The sentiment polarities of sentences are used to calculate sentiment scores of opinion pairs in the framework.

4.5. Overall framework

There are four kinds of information considered to classify sentiment polarities of opinion pairs. We propose a unified framework to integrate different kinds of sentiment information. We formulate the construction problem as an optimization problem, to calculate sentiment scores of opinion pairs.

We first introduce notations which will be used in the following. $\mathbf{s} \in \mathbb{R}^{n\times 1}$ denotes sentiment scores of opinion pairs, where n is the number of opinion pairs and s_i is the sentiment score of the ith opinion pair. $\mathbf{g} \in \mathbb{R}^{n\times 1}$ denotes general-purpose sentiment lexicon information. $g_i = 1$ (or -1) if the ith opinion pair has a positive (or negative) general sentiment word as opinion word, otherwise $g_i = 0$.

$\mathbf{M} \in \mathbb{R}^{n\times n}$ and $\mathbf{N} \in \mathbb{R}^{n\times n}$ denote consistency and reverse of sentiment orientations between opinion pairs. $M_{ij} = T^s_{ij} + C_{ij}$ indicates consistency between the ith and jth opinion pairs. $T^s_{ij} = 1$ if they are synonyms according to thesaurus, otherwise $T^s_{ij} = 0$. Similarly, $N_{ij} = T^a_{ij} + A_{ij}$, where $T^a_{ij} = 1$ if two opinion pairs are antonyms according to thesaurus. $\mathbf{C}$ and $\mathbf{A}$ indicate syntactic information extracted from corpus discussed in Section 4.3. If value of M_{ij} (or N_{ij}) is large, the ith and jth opinion pairs have high probability to have same (or opposite) sentiment polarities.

$\mathbf{d} \in \mathbb{R}^{m\times 1}$ denotes sentiment polarities of the sentences. Only sentences which meet the rules in Section 4.4 are selected. $d_i = 1$ if the ith sentence has positive polarity, otherwise $d_i = -1$. $X \in \mathbb{R}^{m\times n}$ shows the appearances of opinion pairs in sentences. If the ith opinion pair appears in the jth sentence, X_{ij} is set to $\frac{1}{k}$, where k is the number of pairs in the sentence, otherwise $X_{ij} = 0$.

According to the notations mentioned above, the overall framework is formulated as follow:

$$\begin{aligned}\arg\min_{s}\ \mathcal{L}(s) = {} & \frac{1}{\|\mathbf{g}\|_1}\sum_{i=1}^{n} s_i g_i + \frac{1}{\|\mathbf{M}\|_1}\sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij}(s_i - s_j)^2 \\ & + \frac{1}{\|\mathbf{N}\|_1}\sum_{i=1}^{n}\sum_{j=1}^{n} N_{ij}(s_i + s_j)^2 \\ & + \frac{1}{\|\mathbf{d}\|_1}\sum_{i=1}^{m}(d_i - \sum_{i=1}^{n} X_{ij}s_j)^2, \end{aligned} \tag{7}$$

In Eq. (7), minimizing the term $\sum_{i=1}^{n} s_i g_i$ means that if the opinion word of an opinion pair is a general sentiment word, its sentiment score should be the same polarity with its opinion word. By minimizing the term $\sum_{j=1}^{n} M_{ij}(s_i - s_j)^2$, two opinion pairs are given close sentiment scores if connected by coordinate conjunctions, or they are synonyms according to thesaurus. Similarly, minimizing term $\sum_{j=1}^{n} N_{ij}(s_i + s_j)^2$ means that if two opinion pairs are connected by adversative conjunctions, or they are antonyms according to thesaurus, their sentiment scores should have different signs. Finally, opinion pairs will be given positive polarities if they appear frequently in positive sentences, and vice versa, by minimizing the term $\sum_{i=1}^{m}(d_i - \sum_{i=1}^{n} X_{ij}s_j)^2$. In addition, $\|\mathbf{g}\|_1$, $\|\mathbf{M}\|_1$, $\|\mathbf{N}\|_1$ and $\|\mathbf{d}\|_1$ are used to normalize different constraints, to balance influences of different sources of information. It can be verified that Eq. (7) is a convex optimization problem, and can be solved by gradient descent.

Table 3
Statistics of dataset.

Number of	Electronics	Kitchen	Sports	Apparel
Labeled reviews	2000	2000	2000	2000
Unlabeled reviews	21009	17856	3728	7252

5. Experiment

In this section, we introduce the dataset first, then describe the evaluation metrics. After that, we present the evaluation results of our target-specific sentiment lexicon construction method in various domains compared with different baseline methods. Because our method solves two main problems in target-specific sentiment lexicon construction, i.e., opinion target extraction and opinion pair sentiment classification, we evaluate our method in two phases. Firstly, the target extraction method is evaluated. Secondly, the effectiveness of sentiment polarity classification method is evaluated. Our target-specific sentiment lexicon construction method is represented as *TSSL* in the following.

5.1. Description of dataset

There are no existing dataset to evaluate methods of target-specific sentiment lexicon construction. Therefore, we create a dataset to evaluate our method based on a publicly available product review dataset,[4] which have been used in previous research (Blitzer, Dredze, & Pereira, 2007). The dataset contains reviews from a large variety of domains. We choose reviews from 4 domains in our experiment, which are *Electronics, Kitchen, Sports* and *Apparel*. The statistics of the dataset are shown in Table 3.

In each domain, there are 1000 positive and 1000 negative reviews, as well as a large amount of unlabeled data. Word vectors are trained using all reviews in each domain. The open source tool word2vec[5] is employed. Our method contains two parts, but the dataset only has labels of review-level sentiment polarities, which are not useful to evaluate performances of algorithms in the whole construction process. As a result, we annotate the dataset manually to get finer-grained information, which are used to evaluate our method more comprehensively. The annotation method is described as follows.

Because there are thousands of reviews in each domain, and it's hard to label all opinion pairs manually, we annotate opinion pairs extracted from labeled reviews (1000 positive and 1000 negative) in each domain as the test set. Opinion target extraction method and opinion pair sentiment classification method are evaluated according to the test set. A group of candidate opinion pairs are extracted based on the extraction method described in Section 3.2, consisting of candidate targets and candidate opinion words. The candidate targets are manually annotated. Domain-specific targets

[4] http://www.cs.jhu.edu/mdredze/datasets/sentiment/
[5] https://code.google.com/archive/p/word2vec/

Table 4
Statistics of candidate opinion pairs and opinion targets.

Number of	Electronics	Kitchen	Sports	Apparel
Candidate opinion pairs	1545	1184	491	555
Candidate targets	422	377	242	180
Opinion pairs	471	316	131	190
Opinion targets	224	172	102	77

such as “computer” in *Electronics* domain, and general targets such as “price”, are labeled as opinion targets. Irrelevant nouns such as “day”, “minute” and “one” are labeled as non-targets. After that, sentiment polarities of opinion pairs containing opinion targets are manually annotated as “positive”, “negative” and ”neutral”, and opinion pairs in the first two classes are put into the target-specific sentiment lexicon.

In a word, two fine-grained datasets are created based on the extracted candidate opinion pairs from 2000 reviews in each domain. Three annotators participate in dataset creation. Each sample will be given a label if two or more annotators agrees on it. Otherwise the sample will be discarded, which doesn’t happen in our experiment. These two fine-grained datasets are used to evaluate our method in the following.

5.2. Evaluation of opinion target identification

This section describes the results of opinion target extraction. We extract all candidate opinion pairs of reviews in each domain according to extraction method proposed in Section 3.2. Only pairs which occur more than 3 times are kept. Statistics of candidate opinion pairs are shown in Table 4.

We evaluate methods according to precision, recall and F-measure, the same as previous works (Liu et al., 2015; Liu et al., 2016; Qiu et al., 2011). Precision is the ratio of correctly identified opinion targets to all opinion targets identified by the method. Recall is the ratio of correctly identified opinion targets to all annotated opinion targets. F-measure is the harmonic mean of precision and recall, defined as follows:

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}, F = \frac{2 * P * R}{P + R} \tag{8}$$

Where *TP* means number of opinion targets identified correctly, *FP* means number of non-targets which are classified as opinion targets, and *FN* means number of opinion targets classified as non-targets.

In our method, LocLDA is applied to generate a small set of seed opinion targets. Parameters of LDA are set (i.e., $\alpha = 0.1$, $\beta = 0.1$, 2000 iterations, number of topic is 14), the same as previous work (Brody & Elhadad, 2010), and top 10 ranking words in each topic are extracted as representative topic words. Candidate targets will be identified as seed opinion targets if they are representative topic words. An one-class SVM (Schölkopf et al., 2001) is trained with seed opinion targets. μ in SVM is set to 0.5. Influences of parameters will be discussed at the end of this section.

We conduct experiments to compare our target extraction methods with several baseline methods. The first method is a frequency-based approach (Hu & Liu, 2004), which extracts frequent nouns or noun phrases as features (opinion targets) according to a threshold of minimum support, following by several pruning methods to remove incorrect features. In our experiment, the threshold of minimum support is set to 0.5%. Opinion targets with low frequencies will be filtered. The second method is *Double Propagation (DP)* (Qiu et al., 2011), which extracts opinion targets based on a set of syntactic relations between words in dependency trees, and thus is a rule-based approach. The third method is *LocLDA* (Brody & Elhadad, 2010), which are employed in TSSL to generate seed opinion targets. We employ *LocLDA* by itself, without the following steps in TSSL, as the third baseline. Top-k representative topic words are identified as opinion targets, where k is set to 5, 10, and 15. The comparison results are shown in Table 5.

We can see from Table 5 that our opinion target extraction method achieves the best performances in the balance of precision and recall in different domains. Hu & Liu’s method suffers from low precision, because it mistakenly extracts some frequent nouns as opinion targets, and ignores infrequent opinion targets. Although DP method achieves the highest recall, its precision is unsatisfactory because rule-based method could extract meaningless nouns modified by adjectives. Our method outperforms Hu & Liu’s method and DP in precision significantly, because we use a set of seed opinion targets in high accuracy to generate a larger amount of opinion targets. The expansion is based on semantic similarity, which doesn’t influenced by the negative effects of frequency and syntactic rules.

As to LocLDA, its precision is high when extracts only a small number of words in each topic. However, the precision decreases with the increase of representative topic words. Because some opinion targets in low frequencies have low probabilities in all topics, while irrelevant nouns in high frequencies could be extracted. Compared to LocLDA, our method makes use of semantic and linguistic information, both of which are effective in extracting infrequent opinion targets. Compared with the above baselines, our method achieves both high accuracy and recall, thus improves the F-measure significantly.

In order to analyze the effects of different submodules in the whole process of target extraction, we conduct experiments on evaluation of each submodule. The experimental results are shown in Table 6. First, we evaluate the small group of seed opinion targets extracted. Then, the expanded group of opinion targets based on one-class SVM are evaluated. Finally, we evaluate the performances of extraction rules. It can be seen from Table 6 that the seed extraction module generates a set of seed opinion targets in high precision but low recall. The one-class SVM trained with the features of seed opinion targets expands the group of opinion targets. Finally, the rule-based extraction method identifies a large number of opinion targets and improves the recall significantly, but makes some mistakes, causing the decrease of accuracy. The F-measure achieves the highest with the combination of three submodules, which shows that all of the extraction steps can make contribution to identify opinion targets, and our method combines them effectively.

There are two most important parameters in target extraction algorithm, i.e. the number of representative words k in each topic, and μ in one-class SVM. k influences the number of seed opinion targets, and μ influences the learning result of one-class SVM. We set k from 3 to 20, to analyze influence of the number of topic words. As can be seen from Fig. 5, although precision, recall and F-measure change with k, the influence is not obvious. Because with the increase of k, a larger number of seed opinion targets are extracted, which could introduce incorrect targets at the same time. The performance of one-class SVM will be better with a larger number of correct training samples, but worse with a larger number of incorrect training samples.

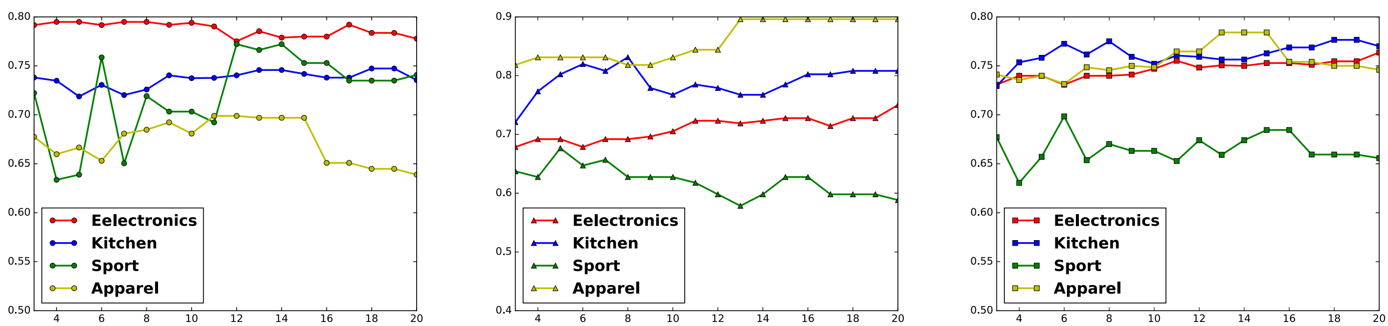
Although the relationship between k and performance of target extraction method is obscure, we can see from Fig. 5 that F-measure maintains at a high level. The satisfactory result derives from the effectiveness of the second module (one-class SVM) and the third module (syntactic rules). When k is small, one-class SVM can still identify many opinion targets because targets in a specific domain are usually semantically related, which could be extracted even with a small set of seed opinion targets. Moreover, the proposed syntactic rules can extract a large number of opinion targets iteratively, even in lack of seed opinion targets. Therefore, the performance of our method is not influenced by k significantly, which

Table 5
Performances of different methods on opinion target extraction.

	Electronics			Kitchen			Sports			Apparel		
	P	R	F	P	R	F	P	R	F	P	R	F
Hu & Liu	0.5977	0.6986	0.6442	0.5000	0.7143	0.5882	0.4000	0.6139	0.4844	0.4715	0.7838	0.5888
DP	0.4912	**0.8904**	0.6331	0.4306	**0.9226**	0.5872	0.3727	**0.8119**	0.5109	0.3681	0.8108	0.5063
LocLDA(5)	**0.8980**	0.2009	0.3283	**0.7895**	0.2679	0.4001	0.6250	0.2475	0.3546	0.6279	0.3649	0.4616
LocLDA(10)	0.7722	0.2785	0.4094	0.7717	0.4226	0.5461	0.6140	0.3465	0.4430	0.6364	0.4730	0.5427
LocLDA(20)	0.7132	0.4201	0.5287	0.6889	0.5536	0.6139	0.5000	0.4158	0.4541	0.5714	0.5405	0.5467
TSSL	0.7939	0.7054	**0.7470**	0.7374	0.7674	**0.7521**	**0.7033**	0.6275	**0.6632**	**0.6809**	**0.8311**	**0.7485**

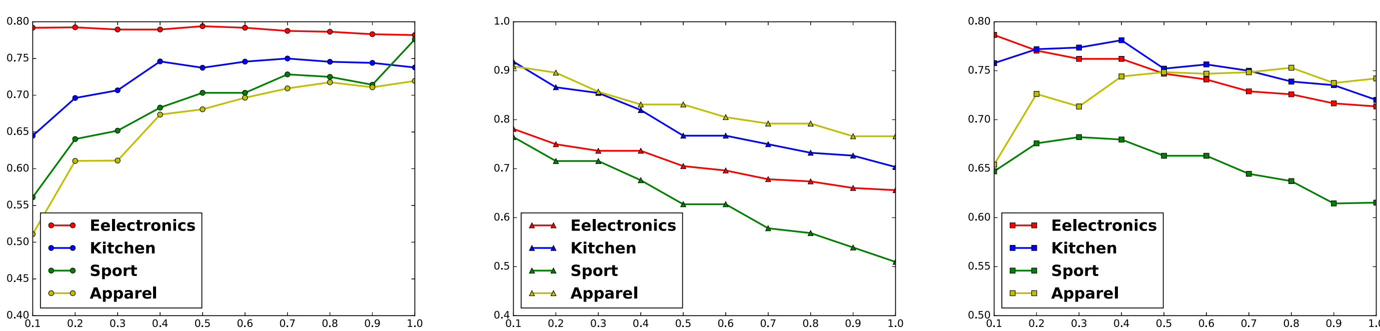
Table 6
Analysis of different submodules in opinion target extraction.

	Electronics			Kitchen			Sports			Apparel		
	P	R	F	P	R	F	P	R	F	P	R	F
seed	0.9394	0.1429	0.2481	**0.9444**	0.1977	0.3270	**0.8000**	0.1176	0.2051	**0.7500**	0.1558	0.2580
+one-class SVM	**0.9524**	0.2679	0.4182	0.8871	0.3198	0.4701	0.6428	0.2647	0.3750	0.6905	0.3766	0.4874
+rules	0.7939	**0.7054**	**0.7470**	0.7374	**0.7674**	**0.7521**	0.7033	**0.6275**	**0.6632**	0.6809	**0.8311**	**0.7485**



(a) Precision (b) Recall (c) F-measure

Fig. 5. Performances of target extraction for different k in different domains.



(a) precision (b) recall (c) recall

Fig. 6. Performances of target extraction for different μ in different domains.

is an advantage because it makes our method have stable results in different domains, with no need for tuning of k.

Fig. 6 shows influence of μ on performance of our target identification method. μ is set in the range of (0.1, 1), with the interval of 0.1. As can be seen from Fig. 6, the precision of our method decreases, while the recall increases, with the increase of μ. This phenomenon repeats in all domains. The reason lies in the effect of μ on one-class SVM. μ controls the amount of outliers. When μ is small, most of samples in training samples are classified as positive. As a result, the precision of our method is relatively low, because most of seed opinion targets are treated as positive samples, with some incorrect targets. With the increase of μ, the precision raises because only the most probable words are identified as positive samples, which are in high accuracy, while the recall decreases because some opinion targets may be ignored.

μ could be set to different values according to the requirement. If a group of opinion targets in high precision are required, then μ should be set to a large value. If it's more important to extract as more opinion targets as possible, then a small value of μ should be set. In our following experiment, we set μ to 0.5, where the precision and recall are both satisfactory.

There is a smoothing parameter α in Section 4.3 to compute matrices of syntactic information. We set α to 5 in our experiment. Because we observe that performance of target extraction is not influenced when α is set to 3 to 10. If α is too large, the impact of syntactic information will be too small. If α is too small, influence

Table 7
Statistics of candidate opinion pairs and opinion targets.

Number of	Electronics	Kitchen	Sports	Apparel
All opinion pairs	471	316	131	190
Positive opinion pairs	405	275	112	166
Negative opinion pairs	66	41	19	24

of syntactic information will be too large, which could introduce errors made by syntactic information.

5.3. Evaluation of sentiment classification of opinion pairs

This section describes the experimental results of sentiment classification of opinion pairs. The dataset described in Section 5.1 is used, which contains opinion pairs with positive or negative labels. A more detailed information of opinion pairs are given in Table 7.

It can be seen from Table 7 that positive and negative opinion pairs are unbalanced, mainly because there are much more positive reviews than negative reviews in the unlabeled dataset.

We evaluate methods according to their precision, recall and F-measure, in common with previous work (Lu et al., 2011). Some methods may ignore opinion pairs without general opinion words. For example, "high quality" may neither be classified into positive nor negative polarity. As a result, recall is an important evaluation, indicating the ability of a method to classify opinion pairs with implicit opinion words. There are two ways to calculate precision, recall and F-measure, i.e., micro-averaging and macro-averaging (Rossi, de Andrade Lopes, & Rezende, 2016; Sokolova & Lapalme, 2009). While micro-averaging considers sum of all terms, it may be dominated by classes with larger number of samples (Rossi et al., 2016). Thus, we use both micro-averaging F-measure and macro-averaging F-measure to give a more comprehensive evaluation of different methods in opinion pair sentiment classification. Micro-averaging and macro-averaging precision and recall are calculated as follows:

$$P_{micro} = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N}(TP_i + FP_i)}, R_{micro} = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N}(TP_i + FN_i)},$$
$$F_{micro} = \frac{2 * P_{micro} * R_{micro}}{P_{micro} + R_{micro}} \quad (9)$$

Where TP_i indicates number of terms classified correctly in class i, FP_i indicates number of terms classified into class i incorrectly, and FN_i indicates number of terms with label i which are classified into other classes. Macro-averaging precision and recall are calculated through averaging precision and recall in all classes.

$$P_{macro} = \frac{\sum_{i=1}^{N} P_i}{N}, R_{macro} = \frac{\sum_{i=1}^{N} R_i}{N}, F_{macro} = \frac{2 * P_{macro} * R_{macro}}{P_{macro} + R_{macro}} \quad (10)$$

Several baselines are compared with our method. Firstly, several publicly available general-purpose sentiment lexica (Baccianella et al., 2010; Hu & Liu, 2004; Wilson et al., 2005) are used to classify sentiment polarities of opinion pairs. If the opinion word of an opinion pair is contained in a sentiment lexicon, then the opinion pair is given the same sentiment polarity with its opinion word. The second baseline is a rule-based method to identify domain-specific sentiment words (Qiu et al., 2011). If an opinion word is identified as positive or negative in specific domain, the opinion pair is given the same polarity. The third baseline is a target-specific sentiment polarity detection method (Fahrni & Klenner, 2008), which classifies polarities of adjectives according to specific targets based on two syntactic rules. The experimental results are shown in Tables 8 and 9.

As can be seen from Tables 8 and 9, our method achieves the best performances in construction of target-specific sentiment lexicon in different domains, with the highest recall and f-score. General-purpose sentiment lexicon has highest precision sometimes, because it contains general-purpose sentiment words which have consistent sentiment polarities without regard to opinion targets. Hu & Liu's sentiment lexicon has the highest precision in our dataset among the three kinds of publicly available sentiment lexica. However, recalls of general-purpose sentiment lexica are low because they ignore sentiment words whose sentiment polarities change with opinion targets. Fahrni & Klenner's method calculates sentiment scores of opinion pairs based on syntactic rules, which achieves higher recall compared to general-purpose sentiment lexica. Nevertheless, syntactic rules of their method require two opinion words modify a same opinion target in a sentence, and connected by a conjunction. Such syntactic patterns occur infrequently in texts, thus the improvement in recall is limited. Compared with existing methods, our target-specific sentiment lexicon construction method fuses different sources of information in classifying sentiment polarities of opinion pairs. As can be seen from Tables 8 and 9, our method improves recall and f-score significantly, because different sources of information all play a part in correctly classifying sentiment polarities of opinion pairs without general sentiment words. Meanwhile, the precision maintains at a high level, because various sources of information are combined in one framework to reduce the error.

There are several sources of information fused in the overall framework. To analyze the influences of different sources of information, we conduct experiments by removing one kind of information at a time, and evaluate the performances of sentiment classification of opinion pairs. The experimental results are shown in Tables 10 and 11.

It can be seen from Tables 10 and 11 that removing any sources of information will bring negative influence to the final performance. Among four kinds of information, the general-purpose sentiment lexicon has the greatest impact on sentiment classification of opinion pairs, because general sentiment words appear frequently, and most of them express consistent sentiment polarities describing different opinion targets, which makes the general-purpose sentiment lexicon a useful and important prior knowledge. In addition, when sentence-level sentiment information is dropped, the recall decreases dramatically, which suggests that sentence-level sentiment information is more important in identifying new opinion pairs, i.e., opinion pairs without general opinion words. On the other hand, some sentence-level sentiment information could ignore infrequent opinion pairs which don't appear in positive or negative sentences, while syntactic and thesaurus information are not influenced by frequencies. It can be seen from the result that our framework achieves the best performances in different domains when all sources of information are integrated, which proves the effectiveness of extracted information from different sources.

5.4. Evaluation of target-specific sentiment lexicon in target-level sentiment classification

In previous subsections, experiments are conducted to evaluate effectiveness of our method to construct target-specific sentiment lexicon. In this section, we compare our lexicon with several existing general-purpose sentiment lexicons in target-level sentiment classification to evaluate its performance in application.

We conduct experiments on a target-level sentiment analysis dataset constructed with book reviews by Álvarez-López et al. (2017). The dataset contains 2977 tagged sentences from 300 reviews of 40 different books. Targets in sentences are tagged, and sentiment polarities of targets are labeled as well. Statistics of the dataset are shown in Table 12. A target-specific sentiment lexicon in *Book* domain is constructed with our method

Table 8
Performances of different methods on opinion pair sentiment classification in micro-averaging.

	Electronics			Kitchen			Sports			Apparel		
	P	R	F	P	R	F	P	R	F	P	R	F
MPQA	0.9670	0.8089	0.8809	0.9494	0.7722	0.8517	0.9558	0.8244	0.8852	0.9568	0.8158	0.8807
Hu & Liu	**0.9872**	0.8217	0.8969	**0.9884**	0.8101	0.8904	0.9820	0.8321	0.9008	**0.9755**	0.8368	0.9009
SentiWordnet	0.9004	0.8450	0.8718	0.8557	0.8070	0.8306	0.9098	0.8473	0.8775	0.9091	0.8421	0.8743
Fahrni&Klenner	0.9802	0.8429	0.9063	0.9817	0.8481	0.9100	0.9825	0.8550	0.9143	0.9649	0.8684	0.9141
TSSL	0.9864	**0.9256**	**0.9551**	0.9829	**0.9082**	**0.9441**	**0.9917**	**0.9160**	**0.9524**	0.9630	**0.9579**	**0.9604**

Table 9
Performances of different methods on opinion pair sentiment classification in macro-averaging.

	Electronics			Kitchen			Sports			Apparel		
	P	R	F	P	R	F	P	R	F	P	R	F
MPQA	0.8984	0.7938	0.8429	0.8333	0.7134	0.7687	0.8529	0.7444	0.7950	0.9109	0.7698	0.8345
Hu & Liu	0.9537	0.7885	0.8633	**0.9444**	0.7145	0.8135	0.9375	0.7925	0.8589	**0.9619**	0.7819	0.8626
SentiWordnet	0.7897	0.8528	0.8200	0.7168	0.7749	0.7447	0.7942	0.8015	0.7978	0.7995	0.8205	0.8099
Fahrni&Klenner	0.9372	0.8072	0.8673	0.9265	0.7363	0.8205	0.9375	0.8059	0.8667	0.9342	0.8000	0.8619
TSSL	**0.9623**	**0.8744**	**0.9162**	0.9440	**0.8643**	**0.9024**	**0.9737**	**0.9290**	**0.9508**	0.9109	**0.9224**	**0.9166**

Table 10
Influences of different sources of information in opinion pair sentiment classification.

	Electronics			Kitchen			Sports			Apparel		
	P	R	F	P	R	F	P	R	F	P	R	F
No lexicon	0.9568	0.7049	0.8117	0.9679	0.6677	0.7903	0.9873	0.5954	0.7428	0.9359	0.7684	0.8439
No syntactic	0.9862	0.9087	0.9459	0.9858	0.8766	0.9280	0.9829	0.8779	0.9274	**0.9779**	0.9316	0.9542
No thesaurus	0.9749	0.9087	0.9407	0.9658	0.8924	0.9276	0.9913	0.8702	0.9268	0.9333	0.8842	0.9081
No sentence	**0.9874**	0.8344	0.9045	**0.9928**	0.8671	0.9257	0.9835	0.9084	0.9444	0.9290	0.8947	0.9115
All	0.9864	**0.9256**	**0.9551**	0.9829	**0.9082**	**0.9441**	**0.9917**	**0.9160**	**0.9524**	0.9630	**0.9579**	**0.9604**

Table 11
Influences of different sources of information in opinion pair sentiment classification.

	Electronics			Kitchen			Sports			Apparel		
	P	R	F	P	R	F	P	R	F	P	R	F
No lexicon	0.8699	0.6065	0.7147	0.9296	0.6119	0.7380	0.9375	0.5012	0.6532	0.7716	0.5823	0.6637
No syntactic	0.9563	0.8708	0.9116	0.9657	0.8253	0.8900	0.9474	0.8849	0.9151	**0.9668**	0.8896	**0.9265**
No thesaurus	0.9270	**0.8772**	0.9014	0.9065	0.8137	0.8576	0.9667	0.8149	0.8843	0.8388	0.8446	0.8417
No sentence	0.9537	0.7959	0.8677	**0.9677**	0.7991	0.8754	0.9500	0.9246	0.9371	0.8205	0.8328	0.8266
All	**0.9623**	0.8744	**0.9162**	0.9440	**0.8643**	**0.9024**	**0.9737**	**0.9290**	**0.9508**	0.9109	**0.9224**	0.9166

Table 12
Statistics of target-level sentiment analysis dataset.

Number of	Positive	Neutral	Negative	All
All targets	1083	2220	530	3833
Explicit targets	956	2164	384	3504
Implicit targets	127	56	146	329
With opinion words	373	346	120	839

on a large number of unlabeled reviews in *Book* domain, and is used to classify target-level sentiment polarity in the dataset.

The dataset contains implicit targets as well. For example, in sentence “The story is worked around how the sisters cope with other people, and each other”, the implicit target is “plot”. In our experiment, we only use sentences with explicit targets to evaluate our target-specific sentiment lexicon. Because our lexicon focuses on cases where opinion targets are known. If the opinion target is implicit, an implicit target extraction step has to be performed before the lexicon is used. In this section, we concentrate on comparison among our lexicon and other general-purpose sentiment lexicons.

In addition, to get more detailed evaluation, apart from using the whole dataset, we perform experiments on a specific subset of the dataset as well. It’s because sentiment orientations are not always expressed with adjectives describing opinion targets. For example, in sentence “The book will appeal to many people”, “book” is a positive opinion target without adjective descriptive word. However, our target-specific sentiment lexicon mainly focuses on opinion targets described by opinion words. As a result, we evaluate performances of different methods on sentences in which targets are related to at least one adjective.

In our experiment, the sentiment classification method is motivated by Qiu et al. (2011), which extracts relations between opinion words and opinion targets based on dependency relations, and propagates sentiment information in relations. Relations of opinion targets and opinion words are extracted based on the rules proposed in Section 3.2. If a target is described by an opinion word, and both of them are contained in an opinion pair in our lexicon, the target will be given the same sentiment polarity as the opinion pair. If they are not contained in our lexicon, but the opinion word is in general-purpose sentiment lexicon, the target will be given the same sentiment polarity as the opinion word. Otherwise, the sentiment polarity of target will be classified as neutral. The influences of negations are considered by examining words in the surrounding window.

To evaluate the effectiveness of different lexicons, precision, recall and F-score in macro-averaging are used as the evaluation

Table 13
Comparison with baselines on 3-class target-level sentiment classification.

	All targets			Targets with opinion words		
	P	R	F	P	R	F
Sentence-level	0.4686	**0.5286**	0.4968	0.4805	0.5294	0.5038
MPQA	0.6133	0.4448	0.5156	0.6024	0.6038	0.6031
Hu & Liu	0.6236	0.4344	0.5121	0.6053	0.5884	0.5967
SentiWordnet	0.5120	0.4314	0.4683	0.4886	0.5098	0.4990
TSSL	**0.6249**	0.4487	**0.5224**	**0.6158**	**0.6168**	**0.6163**

Table 14
Examples of new opinion words.

Target	Opinion word	Polarity
Word	Descriptive	+1
Book	Informative	+1
Character	Thin	−1
Plot	Easy	−1

metrics. Sentiment polarities of opinion targets are classified into positive, negative or neutral.

Experiments of some baseline approaches are conducted for comparison. The first is a sentence-level sentiment classification method (Velikovich, Blair-Goldensohn, Hannan, & McDonald, 2010). Sentiment polarities of sentences are assigned directly to opinion targets in them, to evaluate the effect of dependency relation extraction in target-level sentiment classification. The second to the fourth are several publicly available general-purpose sentiment lexicons, also used in Section 5.3. If the opinion word of an opinion target is contained in general-purpose sentiment lexicon, the target will be given the same sentiment polarity as the opinion word, taking negations into consideration.

The performances of different baseline approaches are shown in Table 13, from which we can observe that assigning sentiment polarities of sentences to opinion targets in them will lead to low precision. The reason lies in diversity and inconsistency of sentiment expressions in texts. A sentence could contain several opinion targets with different sentiment polarities. Moreover, unsupervised sentiment classification method depends on opinion words which appear in sentences, while opinion words may not be reliable to detect sentiment polarities of sentences sometimes, especially in *Book* domain. A number of reviews in *Book* domain mention plots in books. For example, in sentence “The book is about a man who is having a dark time”, “dark” is a negative opinion word, while sentiment polarity of the sentence should be neutral. Assigning sentiment polarities of sentences to opinion targets will classify neutral opinion targets into positive or negative polarities. As a result, dependency relations are useful in unsupervised target-level sentiment classification.

It can be seen from Table 13 that our target-specific sentiment lexicon (TSSL) achieves the best performance in target-level sentiment classification, compared with several general-purpose sentiment lexicons. Both precision and recall are higher, mainly because our method identifies some target-specific or domain-specific opinion words, thus is able to correctly classify sentiment polarities of more opinion targets. Some of the examples are shown in Table 14.

In Table 13, lexicons have higher recall when applied to targets described with adjectives, because more targets with positive or negative sentiments are correctly classified. The precision is not high because of the difficulty of sentiment classification in *Book* domain. A number of opinion words appear in reviews to describe plots of books, which are not used to express opinions upon targets. Thus, some targets with neutral polarities could be classified into positive or neutral because of their related opinion words. To get a more detailed observation, we perform evaluations on sentences in which opinion targets are positive or negative, to remove influences of neutral samples. The results are shown in Table 15.

Table 15
Comparison with baselines on target-level sentiment classification in different classes.

	3-class			2-class		
	P	R	F	P	R	F
Sentence-level	0.5142	0.5521	0.5325	0.7794	**0.6593**	**0.7144**
MPQA	0.6024	0.6038	0.6031	0.8587	0.5705	0.6856
Hu & Liu	0.6053	0.5884	0.5967	0.8278	0.5086	0.6301
SentiWordnet	0.4886	0.5098	0.4990	0.7254	0.5986	0.6559
TSSL	**0.6158**	**0.6168**	**0.6163**	**0.8868**	0.5861	0.7058

It can be observed that our target-specific sentiment lexicon achieves high precision if opinion targets are positive or negative in sentences. Assigning sentiment polarities of sentences to opinion targets has higher recall at the sacrifice of precision. On the other hand, it also inspires our future work to extract more useful sentiment information in target-level sentiment classification. Our target-specific sentiment lexicon improves the precision significantly, and achieves higher F-score compared with several general-purpose sentiment lexicons.

6. Conclusion

This paper presents an automatic approach to construct a target-specific sentiment lexicon, in which each term is a pair of opinion word combined with an opinion target. Our method extracts a group of opinion pairs, and classifies sentiment polarities of them. An unsupervised extraction algorithm is proposed, which extracts candidate pairs based on syntactic relations. To prune irrelevant candidate pairs containing non-targets, we propose an unsupervised opinion target extraction algorithm. The algorithm combines both semantic features and syntactic features, through training an one-class SVM based on semantic features constructed with word embeddings, and expanding the group of opinion targets iteratively with a syntax-based method. It's been proved able to identify opinion targets in both high accuracy and coverage. The extracted opinion targets are used to prune candidate pairs, to generate opinion pairs in high quality. To solve the problem of insufficiency of prior sentiment knowledge and assign accurate sentiment labels to diverse opinion pairs, we fuse different sources of information extracted from available resources and contexts in a unified framework. We calculate sentiment scores of opinion pairs through solving an optimization problem, and classify opinion pairs into three sentiment categories.

Experimental results on product reviews in different domains show that our method effectively improves the performances of opinion target extraction and sentiment classification of opinion pairs, compared with several baseline methods. It is proved to achieve both high accuracy and recall in opinion target extraction, and able to accurately classify opinion pairs without general sentiment words. In addition, the whole process of our target-specific sentiment lexicon construction needs no human annotation, and thus can be easily and effectively applied to different domains.

Acknowledgments

The authors thank the reviewers for their insightful comments and suggestions which help a lot for improvements of the work. This research is supported by the National Key Research and Development Program of China (No.2016YFB0800402) and National Natural Science Foundation of China (Grant nos. U1536201 and U1705261).

References

Álvarez-López, T., Fernández-Gavilanes, M., Costa-Montenegro, E., Juncal–Martínez, J., García-Méndez, S., & Bellot, P. (2017). A book reviews dataset for aspect based sentiment analysis. In *Proceedings of the 8th language & technology conference: Human language technologies as a challenge for computer science and linguistics (LTC 2017)* (pp. 49–53).

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the seventh international conference on language resources and evaluation (LREC'10)* (pp. 2200–2204).

Blitzer, J., Dredze, M., & Pereira, F. (2007). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of the 45th annual meeting of the association of computational linguistics* (pp. 440–447).

Brody, S., & Elhadad, N. (2010). An unsupervised aspect-sentiment model for online reviews. In *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics* (pp. 804–812). Association for Computational Linguistics.

Chen, Z., Mukherjee, A., & Liu, B. (2014). Aspect extraction with automated prior knowledge learning. In *Proceedings of the 52nd annual meeting of the association for computational linguistics* (pp. 347–358).

Dey, A., Jenamani, M., & Thakkar, J. J. (2018). Senti-n-gram: An n-gram lexicon for sentiment analysis. *Expert Systems with Applications, 103*, 92–105.

Ding, X., Liu, B., & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 international conference on web search and data mining* (pp. 231–240). ACM.

Fahrni, A., & Klenner, M. (2008). Old wine or warm beer: Target-specific sentiment analysis of adjectives. In *Proceedings of the symposium on affective language in human and machine, AISB* (pp. 60–63).

Fernández-Gavilanes, M., Juncal-Martínez, J., García-Méndez, S., Costa-Montenegro, E., & González-Castaño, F. J. (2018). Creating emoji lexica from unsupervised sentiment analysis of their descriptions. *Expert Systems with Applications, 103*, 74–91.

Gilbert, C. H. E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Eighth international conference on weblogs and social media (ICWSM-14)*.

Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for computational linguistics* (pp. 174–181). Association for Computational Linguistics.

He, R., Lee, W. S., Ng, H. T., & Dahlmeier, D. (2017). An unsupervised neural attention model for aspect extraction. In *Proceedings of the 55th annual meeting of the association for computational linguistics* (pp. 388–397).

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 168–177). ACM.

Huang, S., Niu, Z., & Shi, C. (2014). Automatic construction of domain-specific sentiment lexicon based on constrained label propagation. *Knowledge-Based Systems, 56*, 191–200.

Jakob, N., & Gurevych, I. (2010). Extracting opinion targets in a single-and cross-domain setting with conditional random fields. In *Proceedings of the 2010 conference on empirical methods in natural language processing* (pp. 1035–1045). Association for Computational Linguistics.

Kamps, J., Marx, M., Mokken, R. J., De Rijke, M., et al. (2004). Using wordnet to measure semantic orientations of adjectives. In *Proceedings of the 4th international conference on language resources and evaluation (LREC 2004)* (pp. 1115–1118). Citeseer.

Kanayama, H., & Nasukawa, T. (2006). Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of the 2006 conference on empirical methods in natural language processing* (pp. 355–363). Association for Computational Linguistics.

Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In *Mining text data* (pp. 415–463). Springer.

Liu, K., Xu, H. L., Liu, Y., & Zhao, J. (2013). Opinion target extraction using partially-supervised word alignment model. In *Proceedings of the twenty-third international joint conference on artificial intelligence* (pp. 2134–2140).

Liu, P., Joty, S., & Meng, H. (2015). Fine-grained opinion mining with recurrent neural networks and word embeddings. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1433–1443).

Liu, Q., Liu, B., Zhang, Y., Kim, D. S., & Gao, Z. (2016). Improving opinion aspect extraction using semantic similarity and aspect associations. In *Thirtieth AAAI conference on artificial intelligence* (pp. 2986–2992).

Lu, Y., Castellanos, M., Dayal, U., & Zhai, C. (2011). Automatic construction of a context-aware sentiment lexicon: An optimization approach. In *Proceedings of the 20th international conference on world wide web* (pp. 347–356). ACM.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).

Miller, G. A. (1995). Wordnet: A lexical database for english. *Communications of the ACM, 38*(11), 39–41.

Mukherjee, A., & Liu, B. (2012). Aspect extraction through semi-supervised modeling. In *Proceedings of the 50th annual meeting of the association for computational linguistics* (pp. 339–348). Association for Computational Linguistics.

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2009). Sentiful: Generating a reliable lexicon for sentiment analysis. In *3rd international conference on affective computing and intelligent interaction and workshops* (pp. 1–6). IEEE.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001*. Mahwah, NJ: Erlbaum.

Popescu, A.-M., & Etzioni, O. (2007). Extracting product features and opinions from reviews. In *Natural language processing and text mining* (pp. 9–28). Springer.

Qiu, G., Liu, B., Bu, J., & Chen, C. (2009). Expanding domain sentiment lexicon through double propagation. In *Proceedings of the 21st international jont conference on artifical intelligence* (pp. 1199–1204).

Qiu, G., Liu, B., Bu, J., & Chen, C. (2011). Opinion word expansion and target extraction through double propagation. *Computational Linguistics, 37*(1), 9–27.

Rossi, R. G., de Andrade Lopes, A., & Rezende, S. O. (2016). Optimization and label propagation in bipartite heterogeneous networks to improve transductive classification of texts. *Information Processing & Management, 52*(2), 217–257.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation, 13*(7), 1443–1471.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437.

Tan, S., & Wu, Q. (2011). A random walk algorithm for automatic construction of domain-oriented sentiment lexicon. *Expert Systems with Applications, 38*(10), 12094–12100.

Tang, D., Wei, F., Qin, B., Zhou, M., & Liu, T. (2014). Building large-scale twitter-specific sentiment lexicon: A representation learning approach. In *Proceedings of coling 2014, the 25th international conference on computational linguistics: Technical papers* (pp. 172–182).

Turney, P. D. (2002). 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* (pp. 417–424). Association for Computational Linguistics.

Velikovich, L., Blair-Goldensohn, S., Hannan, K., & McDonald, R. (2010). The viability of web-derived polarity lexicons. In *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics* (pp. 777–785). Association for Computational Linguistics.

Wang, B., & Wang, H. (2008). Bootstrapping both product features and opinion words from chinese customer reviews with cross-inducing. In *Proceedings of the third international joint conference on natural language processing*.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 347–354). Association for Computational Linguistics.

Wu, F., Huang, Y., Song, Y., & Liu, S. (2016). Towards building a high-quality microblog-specific chinese sentiment lexicon. *Decision Support Systems, 87*, 39–49.

Yang, B., & Cardie, C. (2013). Joint inference for fine-grained opinion extraction. In *Proceedings of the 51st annual meeting of the association for computational linguistics* (pp. 1640–1649).

Yin, Y., Wei, F., Dong, L., Xu, K., Zhang, M., & Zhou, M. (2016). Unsupervised word and dependency path embeddings for aspect term extraction. In *Proceedings of the twenty-fifth international joint conference on artificial intelligence* (pp. 2979–2985).

Ying, D., Yu, J., & Jiang, J. (2017). Recurrent neural networks with auxiliary labels for cross-domain opinion target extraction. *Thirty-first AAAI conference on artificial intelligence*.

Zhang, Y., Zhang, H., Zhang, M., Liu, Y., & Ma, S. (2014). Do users rate or review?: Boost phrase-level sentiment labeling with review-level sentiment classification. In *Proceedings of the 37th international ACM SIGIR conference on research & development in information retrieval* (pp. 1027–1030). ACM.

Zhuang, L., Jing, F., & Zhu, X.-Y. (2006). Movie review mining and summarization. In *Proceedings of the 15th ACM international conference on information and knowledge management* (pp. 43–50). ACM.