Cogn Comput (2015) 7:369–380
DOI 10.1007/s12559-014-9298-4

Word Polarity Disambiguation Using Bayesian Model and Opinion-Level Features

Yunqing Xia · Erik Cambria · Amir Hussain · Huan Zhao

Received: 16 July 2014 / Accepted: 17 July 2014 / Published online: 2 August 2014

Abstract Contextual polarity ambiguity is an important problem in sentiment analysis. Many opinion keywords carry varying polarities in different contexts, posing huge challenges for sentiment analysis research. Previous work on contextual polarity disambiguation makes use of term-level context, such as words and patterns, and resolves the polarity with a range of rule-based, statistics-based or machine learning methods. The major shortcoming of these methods lies in that the term-level features sometimes are ineffective in resolving the polarity. In this work, opinion-level context is explored, in which intra-opinion features and inter-opinion features are finely defined. To enable effective use of opinion-level features, the Bayesian model is adopted to resolve the polarity in a probabilistic manner. Experiments with the Opinmine corpus demonstrate that opinion-level features can make a significant contribution in word polarity disambiguation in four domains.

Keywords Sentiment disambiguation · Bayesian model · Sentiment analysis · Opinion-level features

Y. Xia (✉) · H. Zhao
Department of Computer Science, TNList, Tsinghua University, Beijing 100084, China
e-mail: yqxia@tsinghua.edu.cn

H. Zhao
e-mail: byzhaohuan@gmail.com

E. Cambria
School of Computer Engineering, Nanyang Technological University, Singapore 639798, Singapore
e-mail: cambria@ntu.edu.sg

A. Hussain
Division of Computing Science and Maths, School of Natural Science, University of Stirling, Stirling FK9 4LA, Scotland, UK
e-mail: ahu@cs.stir.ac.uk

Introduction

Opinion mining research has made significant progress in the past decade. Many systems have been developed, some of which also commercialized to achieve either document-level sentiment analysis or fine-grained opinion mining [2, 4]. Challenging issues in sentiment analysis are many. Much of the research attention is given to subjectivity detection and sentiment classification. For the case where corpora are available on product reviews, some research efforts have been made on opinion holder/target extraction.

Turney and Littman [18] claimed that word polarity ambiguity is an unavoidable challenge. Unfortunately, not much research work is attracted until a relevant SemEval-2010 task was conducted on disambiguating sentiment-ambiguous adjectives (DSAA) [20]. For the first time, the task organizers provided 2,917 test sentences to 17 participant systems to disambiguate sentiment polarity of 14 Chinese adjectives. Today, sentiment analysis research steps into the era of finely grained aspect-based opinion mining, in which sentiment keywords play a vital role as features in machine learning methods or as key elements in rule-based methods. Resolving sentiment polarity of a word now becomes a necessity.

Previous work including systems in SemEval-2010 DSAA task concentrates on adjectives in news. Two limitations are worth noting. Firstly, adjectives are just part of the sentiment-ambiguous words. There are many nouns and verbs which give different sentiment polarity in different context. Ignoring these words is fatal to opinion mining systems. Secondly, sentiment-ambiguous words appear much more frequently in reviews created by social media and e-commerce. The major battlefield is reviews. In this work, we conduct research in reviews and address sentiment-ambiguous words of all types. We first define the problem and then summarize our proposal and contributions.

Problem Definition

Similar to the well-known word sense disambiguation (WSD) task, word polarity disambiguation (WPD) aims at resolving polarity of the sentiment-ambiguous words in a specific context. Formally, we address the WPD problem as follows:

Given a sentiment-ambiguous word w and a certain context Ω which is usually a sentence, WPD attempts to predict a deterministic polarity. With a probabilistic model, the process is formalized as:

$$l^* = \operatorname*{argmax}_{l \in \{1,-1\}} P(l|w,\Omega), \tag{1}$$

where l denotes a polarity which is usually positive (1) or negative (-1).

Two questions will be addressed in this paper:

Q1: *What should be considered as effective context in* Ω?
Q2: *How is the probability* $P(l|w,\Omega)$ *accurately calculated*?

Question Q1 is answered by exploring features for word polarity disambiguation, while question Q2 is a mathematical problem which calculates the probability that the word is assigned a polarity label.

Our Proposal and Contributions

In this work, we make use of two types of features with opinion-level context in word polarity disambiguation: intra-opinion features and inter-opinion features. The intra-opinion features are mainly opinion components, i.e., opinion targets, modifying words and opinion indicating ngrams (more details in "Intra-opinion Features" section). These features can determine polarity of the sentiment-ambiguous words if they do exist. When some of the features are missing, we then rely on inter-opinion features to assist our calculation. We consider conjunctions that connect two sentences (or sub-sentences). Regarding word polarity disambiguation, we adopt a Bayesian model, which calculates polarity probability of a given word within a given context based on posterior distributions that are estimated from training/development data.

The following contributions are worth noting: First, it is observed in this work that the opinion-level context is effective in resolving polarity ambiguity of sentiment words. Second, a Bayesian model can be designed to calculate probability that the given sentiment word carries certain sentiment polarity. An assumption of independence is made among the features that influence the model. Finally, substantial experiments have been conducted that justify usefulness of the features in the opinion-level context and the effectiveness of the proposed Bayesian model.

Paper Structure

The rest of this paper is organized as follows: We summarize related work in "Related Work" section. In "Opinion-Level Features" section, we define features employed in our work. In "The Bayesian Model" section, we present the Bayesian model that computes polarity probability with these features. We report the simulation experiments carried out, along with discussion in "Evaluation" section. The paper is finally concluded in "Conclusions" section.

Related Work

Contextual polarity disambiguation is related to general sentiment analysis and opinion mining. Next, we present an overview of recent related work reported on sentiment polarity disambiguation.

Sentiment Analysis and Opinion Mining

The sentiment analysis concept was initially formulated by classifying text into positive, neutral and negative as an orientation determination problem [9]. Later, Turney [17] proposed to classify positive and negative reviews by Thumbs Up and Thumbs Down. Further, SentiWordNet [8] and SenticNet [3] are compiled. Sentiment lexicon was proved necessary and important, which led to the appearance of sentiment lexicons for other languages, including Chinese [10] and Japanese [16]. Use of sentiment lexicon is accepted as indispensable in sentiment analysis. Some researches use sentiment keywords directly as features in machine learning algorithms for sentiment classification. Others use statistics based on sentiment keywords [23] and linguistic patterns [14]. Since many sentiment keywords present different polarity in different contexts, further natural language processing (NLP) techniques have become the popular approach for polarity classification [5].

In fine-grained opinion mining on product reviews, the use of sentiment lexicon is a must. One sentiment word may imply two opinion polarities, while one sentence is too short and leads to the sparse data problem in sentiment classification. However, using sentiment words directly in fine-grained opinion mining poses a serious polarity ambiguity challenge.

Word Polarity Disambiguation

Turney and Littman [18] claim that sentiment-ambiguous words cannot be avoided easily in a real-world application.

In their work, they designed a point-wise mutual information (PMI)-based algorithm to calculate sentiment orientation of the sentiment word within a review corpus. However unfortunately, word polarity disambiguation has not attracted much research attention. Yi et al. [27] used a lexicon and manually developed patterns to classify contextual polarity. Though the patterns were high-quality and yielded quite high precision over the set of expressions, the recall was rather low. Wilson et al. [19] proposed to recognize contextual polarity of all instances of words from a large lexicon of subjectivity clues that appeared in the corpus. The lexicon included not only adjectives, but nouns, verbs, adverbs and even modals. Popescu and Etzioni [13] use relaxation labeling algorithm to recognize the contextual polarity of words that were at the heads of select opinion phrases. Features were used to represent conjunctions and dependency relations between polarity words. And expressions were limited either to those that targeted specific items of interest, such as products and product features, or to tuples of adjectives and nouns. Ding et al. [6] adopted a holistic lexicon-based approach to resolve the ambiguity problem by exploiting external information and evidences in other sentences and other reviews. Qiu et al. [15] proposed to combine lexicon-based methods and corpus-based methods to first determine the sentence polarity. They simply assigned the sentence polarity to the polarity of the sentiment polarity words in the sentence. Wu and Wen [21] proposed a knowledge-based unsupervised method to automatically disambiguate dynamic sentiment-ambiguous adjectives with a search engine. Interestingly, pattern-based and character-based methods were exploited to infer sentiment expectation of nouns, which is in turn applied to determine the polarity of adjectives.

The SemEval-2010 task on disambiguating sentiment-ambiguous adjectives [20] was an important event that advanced research on word polarity disambiguation. For the first time, sentiment words and test dataset (in Chinese) were provided by the organizers. To disambiguate the sentiment-ambiguous adjectives, Yang and Liu [26] made use of heuristic rules; Xu et al. [25] exploited various heuristics such as collocation and word similarity; Balahur and Montoyo [1] explored methods based on classification, search and rules; Lu and Tsou [11] combined maximum entropy and lexicon. Pak and Paroubek [12] adopted a classifier trained with tweet annotations. We summarize four limitations of the SemEval-2010 systems. First, the systems target adjectives while noun and verb can also play ambiguously in sentiment expression. Second, the popular methods are still knowledge-based though some preliminary machine learning methods have been explored. Third, the systems handle news report. Lastly, all the systems are built on term level. All terms are viewed equally in word polarity disambiguation. In handling reviews, the systems inevitably suffer the serious sparse data problem since reviews are normally too short to match any rules.

Difference of this work lies in three aspects: First, we deal with reviews which are rather different from news reports in nature. Second, we explore probabilistic models, which have never been employed in word polarity disambiguation. To train our models, we collect a large volume of raw reviews as development data to address issues caused by the small training data. Third, we build our method on opinion level. That is, terms in our method are classified into different opinion elements, such as opinion target and modifying constituent. We design the Bayesian model to exploit and quantify their contribution in word polarity disambiguation.

Opinion-Level Features

In this work, features for word polarity disambiguation are defined on opinion level. Generally speaking, we consider two groups of features in this work. First, we consider *intra-opinion features*, which are actually opinion elements. We can safely assume that opinion elements make a different contribution to word polarity disambiguation. Then, we consider *inter-opinion features*, which are helpful in connecting two opinions, say conjunction words. In this section, we describe the two types of features with examples.

Intra-opinion Features

The intra-opinion features refer to opinion elements which help to present the opinion thoroughly. Practically, we observe the elements in the aforementioned 6 tuples one by one to assess their contributions to word polarity disambiguation.

Xia et al. [22] made effective use of sentiment units in sentiment analysis. Their key observation was that in subjective text, a sentiment unit contains sentiment word, modifying word and negation word. We further investigate and advance this theory to model the concept of an opinion unit which is represented with a 6 tuples $<h, t, w, m, n, I>$, in which h represents opinion holder, t opinion target, w sentiment word (i.e., to be disambiguated in terms of polarity), m modifying word(s), n negation word and I a set of indicative words. In this work, we view these elements as candidates of features for word polarity disambiguation. The following opinion elements are found useful:

- Opinion target
- Modifying word
- Indicative words

Table 1 Example reviews containing opinion targets (where S. word refers to the sentiment word)

Review	S. word	Opinion target	Polarity of the S. word
(1) (Performance is greatly improved)	(Improve)	(Performance)	Positive
(2) (Price is increased too much)	(Increase)	(Price)	Negative

Table 2 Example reviews containing a modifying word

Review	S. word	Modifying word	Polarity of the S. word
(1) (Truly too small)	(Small)	(Too)	Negative
(2) (Irresistible small)	(Small)	(Irresistibly)	Positive

We exclude opinion holder as our feature in this work. This is because reviews are usually created in social network or e-commerce Web sites; thus, author information is contained (i.e., the registered user name). So when one deals with reviews, h is usually ignored. Thus, we obtain a revised opinion unit with 5 tuples $<t, w, m, n, I>$. Note that social analysis based on user information can be useful to sentiment analysis, but we focus on textual content for word polarity disambiguation in this work. The intra-opinion features are described as follows:

1. Opinion target

Our study shows that opinion target plays an important role in word polarity disambiguation. This observation is obvious. Let us first examine the example reviews containing opinion target in Table 1.

We notice that within the reviews in Table reftab:target, polarity of the sentiment words depends highly on the opinion targets. For example, the sentiment word (*improve*) gives a positive sentiment with the opinion target (*performance*) in comment (1) and a negative sentiment with the opinion target (*price*) in comment (2). This law works perfectly for all the four examples.

However, less than 50 percent online reviews contain opinion target. In many reviews, opinion target is missing or out-of-vocabulary. In these cases, we need to define extra features to resolve the polarity ambiguity.

2. Modifying word

For presentation convenience, we first define the modifying word. We term a word as modifying word if it syntactically modifies the sentiment word. In practice, we make use of dependency parser to recognize the modifying word.

In cases where the opinion target is not found in reviews, we use the modifying word as a feature in this work. Table 2 gives some example reviews in which opinion target is not mentioned.

In Table 2, we find that the modifying words are also indicative in word polarity disambiguation. On reading these reviews, one can infer the sentiment polarity subconsciously even without the opinion target. For example, in the review (*Irresistible small*), the modifying word (*irresistibly*) can clearly indicate that the polarity of word (*small*) is favorable. In our corpus, we find 178 unique modifying words with the sentiment word (*small*). This shows that the modifying words are significant for polarity disambiguation.

Table 3 Example reviews containing some indicative words

Review	S. word	Indicative words	Polarity of the S. word
(1) (Seeing such a small mobile phone is truly excited)	(Small)	(Truly), (excited)	Positive
(2) (Holding such a small mobile phone makes me hesitating for a while)	(Small)	(Hesitate), (for a while)	Negative

3. Indicative words

Besides opinion target and modifying word, some words in reviews can also be indicative. We define an indicative word as the word that still helps resolve polarity ambiguity without being an opinion target or modifying word. Again, we first give some example reviews in Table 3, in which the indicative words are contained.

It can be seen from Table 3 that opinions are very flexibly given in reviews. From a linguistics point of view, indicative words are still modifying constituents. However, in most cases, it is difficult to recognize the modifying words. We collectively term these as indicative words. For example, in review in Table 3, the indicative words include (*truly*), (*excited*), (*vague*) and (*keyboard*). Using these words as features to disambiguate word polarity can be promising. In review (1), one can easily figure out a positive polarity of the sentiment word (*small*) when indicative words(*truly*) and (*excited*) are given. Next, looking at review (2), one can see that it is not difficult to infer a negative sentiment with indicative words (*hesitate*) and (*for a while*). We can thus safely conclude that indicative words are indeed important in resolving polarity ambiguity. We also notice that a single indicative word cannot function effectively on its own. For example in review (1), (*truly*) as an indicative word cannot give a clear tendency on sentiment polarity. It must be combined with (*excited*) to indicate a positive sentiment. Instead, if it is combined with (*disappointed*), the implied polarity then

Table 4 Example review sentences containing a correlative word

Review	S. word	Correlative word	Polarity of the S. word
(1) (Quality is good, but it is big)	(Big)	(But)	Negative
(2) (Screen resolution is high, and it is big)	(Big)	(And)	Positive

becomes the opposite. Thus, in this work, we consider word ngrams in the word polarity disambiguation method. For example, we consider (*truly*) (*excited*) and (*vague*) (*keyboard*) as ngram features in our work.

Inter-opinion Features

Opinions are not independent of each other. For example, two opinions can appear in one sentence. Two sentences can be connected by a correlation word. Sometimes, the window can be much larger. We define inter-opinion features as the information that connects two or more opinions in the following three types of context:

- *Complex sentence*: In a complex sentence, clauses are connected with some correlative words. Opinions appearing in different clauses can provide some information for word polarity disambiguation.
- *Discourse*: In a discourse, two adjacent sentences are connected by a conjunction word. Opinions appearing in different sentences also hold certain relations, which can help resolving polarity of the sentiment word.
- *Application*: In many applications, positive (Pro) opinions and negative (Con) opinions are given separately in two paragraphs (i.e., Pro and Con), respectively. Thus, opinions in the Pro paragraph are uniformly positive ones, while those in the Con paragraph are uniformly negative ones. Thus, we take the pro–con scenario features into consideration in word polarity disambiguation.

Next, the inter-opinion features are elaborated as follows:

1. Sentence-level opinion correlation

In some complex sentences, a correlative word connects two sub-sentences. Two examples are given in Table 4.

We can see in Table 4 that the correlative word is useful in word polarity disambiguation. For example, in review (1), the correlative word (*but*) connects two opposite opinions. With this correlative word, we are able to figure out polarity of the latter opinion based on the polarity of the former. In this way, polarity ambiguity of (*big*) can be resolved. Such an assumption is pretty safe in complex sentences.

Table 5 Six types of correlative words

Type	Correlative word	Example	S. word	Polarity of the S. word
Coordinative	NIL/(also)	(Screen is clear, also big)	(Big)	Positive
Successive	(And)	(Screen resolution is high, and it is big)	(Big)	Positive
Adversative	(But)	(Quality is good, but it is big)	(Big)	Negative
Concessive	(Even if)	(I won t consider even if it is bigger)	(Big)	Positive
Hypothetical	(If)	(It would be better if the screen is bigger)	(Big)	Negative
Conditional	(As long as)	(I will consider as long as it is bigger)	(Big)	Negative

Table 6 Example review discourses containing correlative word

Discourse	Correlative word	S. word	Polarity of the S. word
(1) (Many merits: screen is clear and bright. Speaker is loud)	NIL	(Big)	Positive
(2) (Screen is clear and bright. But speaker is loud and frightens me)	(But)	(Big)	Negative

In grammar, the correlative word can be a conjunction or an adverb, which is enumerable. However, the correlative word cannot work alone in word polarity disambiguation. We notice that to resolve the polarity, we must consider the correlated opinion. For the review (1) in Table 4, polarity of the correlated opinion is positive. Considering the correlative word (*but*), we know polarity of the current opinion is negative. As no negation is used in the current opinion, we can safely infer that polarity of (*big*) is negative.

In this work, we consider six types of correlative words (see Table 5). Note that there are more types of correlative words, and we select the six as they give clear indication in word polarity disambiguation.

Note that NIL in Table 5 represents that the correlative word can be omitted in the coordinative complex sentence.

2. Discourse-level opinion correlation

Discourse reflects how narration transfers from one topic to another in sentences. Similarly, opinions may transfer in reviews from one polarity to another. Compared to a sentence, a discourse usually contains multiple sentences. In word polarity disambiguation, the discourse-level opinion correlation is similar to sentence-level correlation except for the learning window. Table 6 gives two discourses in

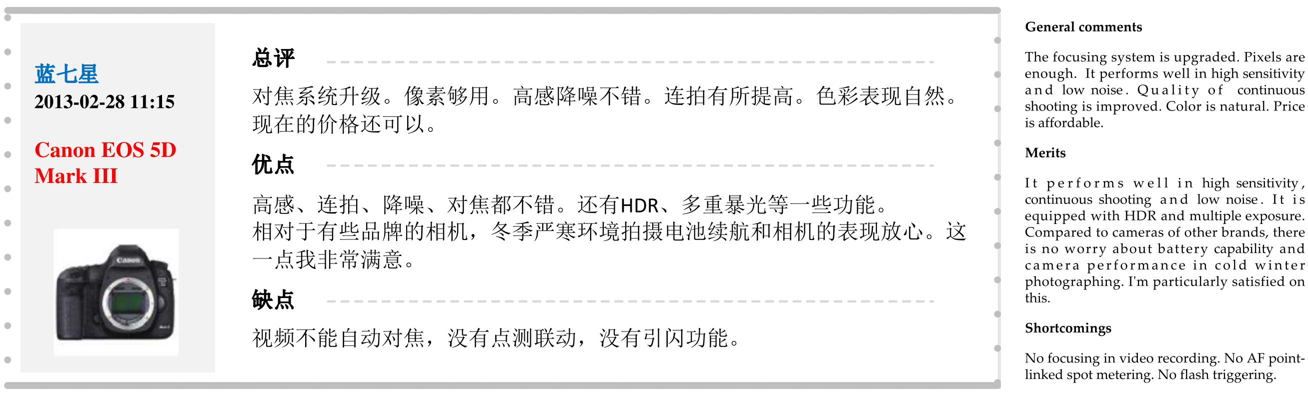

Fig. 1 Example of Pro-Con style review from www.xitek.com, in which reviews are organized in three fields: general, merits (Pro) and shortcomings (Con)

which a correlative word connects two opinions into sentences.

We usually detect a sentence boundary with punctuation marks such as period (i.e., in Chinese). Thus, the discourses in Table 6 all contain three sentences. We make an assumption that in a discourse, the polarity of the opinions will not change unless a correlative word appears to alter the polarity. While this assumption is not ideal for applying to complex sentences, we have found that in many cases when discourses are encountered, the assumption can yield substantial gains.

Similar to the sentence-level opinion correlation, we represent the discourse-level opinion correlation again with a 3 tuples. The window is enlarged to discourse, in this case.

3. Application-level opinion correlation

In many online e-commerce review sites, reviews are structured so as to guide users to input Pro reviews and Con ones separately. Figure 1 gives an example.

Figure 1 shows that opinions in the Merit field are all positive and those in the Shortcomings filed are all negative. This provides us trustworthy evidence in word polarity disambiguation. Note that application-level features only work on those structured reviews. For such free reviews, the application-level features are ignored.

The Bayesian Model

In general, we can safely assume that polarity of a sentiment word can be determined by certain observable context Ω_w in the review. In this work, we only consider two opposite polarity values: positive and negative, represented by 1 and -1, respectively. We propose to resolve polarity of w, i.e., $\varrho_w(\varrho_w \in \{1,-1\})$, within the context Ω_w with a Bayesian model as follows:

$$\varrho_w^* = \underset{\varrho_w \in \{1,-1\}}{\operatorname{argmax}} \; p(\varrho_w|\Omega_w), \tag{2}$$

where $p(\varrho_w|\Omega_w)$ is further calculated based on Bayes rule:

$$p(\varrho_w|\Omega_w) = \frac{p(\varrho_w)p(\Omega_w|\varrho_w)}{p(\Omega_w)}. \tag{3}$$

As $p(\Omega_w)$ is a constant, Eq. 2 can be further revised as follows:

$$\varrho_w^* = \underset{\varrho_w \in \{1,-1\}}{\operatorname{argmax}} \; p(\varrho_w)p(\Omega_w|\varrho_w). \tag{4}$$

In what follows, we describe how different contexts (Ω_w) work in word polarity disambiguation.

The Term-Based Model

The assumption underlying the term-based model is traditional. That is, polarity of the sentiment keyword can be resolved with term-level context. The typical term-level features are ngrams. Letting g_w^i represent one ngram, and $\Omega_w = \{g_w^1, g_w^2, \ldots, g_w^K\}$ represent the context where K denotes number of features, we revise Eq. 4 as follows:

$$\varrho_w^* = \underset{\varrho_w \in \{1,-1\}}{\operatorname{argmax}} \; p(\varrho_w)p(g_w^1, g_w^2, \ldots, g_w^K|\varrho_w). \tag{5}$$

Applying the independence assumption, we further revise Eq. 5 as follows:

$$\varrho_w^* = \underset{\varrho_w \in \{1,-1\}}{\operatorname{argmax}} \; p(\varrho_w)\prod_{i=1}^{K} p(g_w^i|\varrho_w). \tag{6}$$

With a training corpus, we using maximum likelihood estimation (MLE) to estimate $p(\varrho_w)$ and $p(g_w^i|\varrho_w)$.

Two major drawbacks are worth noting in the term-based model. First, Eq. 5 indicates that all the ngrams are used as features in word polarity disambiguation. In fact, many of them are not effective. This inevitably introduces noise in the calculation. Second, Eq. 6 indicates that term-level features are deemed independent of each other. This is usually not true in reviews. As discussed in "Opinion-

Level Features" section, elements of opinion can help resolving word polarity in different ways. Even opinion in a context can influence polarity of the sentiment word. In what follows, we describe the opinion-level model.

The Opinion-Based Model

We make a new assumption on word polarity disambiguation: Polarity of the sentiment keyword depends highly on opinion-level context. The underlying motivation is that since polarity is part of an opinion, polarity of a sentiment word should be more precisely resolved with opinion-level context. Again, we start from Eq. 4 to incorporate opinion-level features step by step, as below.

1. Polarity disambiguation with intra-opinion features

Based on the analysis in "Opinion-Level Features" section, we use the opinion target t_w, modifying word m_w and indicative words I_w as intra-opinion features. We consider an opinion-level context $\Omega_w = \{t_w, m_w, I_w\}$ for sentiment word w. We first design an opinion-level Bayesian model using the intra-opinion features: t_w, m_w and I_w. Equation 4 is first revised as follows.

$$\varrho_w^* = \underset{\varrho_w \in \{1,-1\}}{\operatorname{argmax}} \; p(\varrho_w) p(t_w, m_w, I_w | \varrho_w). \tag{7}$$

Applying the independence assumption, we further obtain:

$$\varrho_w^* = \underset{\varrho_w \in \{1,-1\}}{\operatorname{argmax}} \; p(\varrho_w) p(t|\varrho_w) p(m|\varrho_w) p(I|\varrho_w). \tag{8}$$

We model the indicative words with ngrams $I_w = \{g_I^1, g_I^2, \ldots, g_I^L\}$. Note that ngram of I_w is different from the term-level ngram as the former considers only non-opinion indicative words. Next, Eq. 8 is revised as follows.

$$\varrho_w^* = \underset{\varrho_w \in \{1,-1\}}{\operatorname{argmax}} \; p(\varrho_w) p(t|\varrho_w) p(m|\varrho_w) \prod_{j=1}^{L} p(g_I^j|\varrho_w). \tag{9}$$

In Eq. 9, $p(t|\varrho_w), p(m|\varrho_w)$ and $p(g_I^j|\varrho_w)\{j = 1, \ldots, L\}$ are all estimated with a training corpus.

In cases where some of the opinion-level features are not explicitly given, we set $p(t|\varrho_w) = 1, p(m|\varrho_w) = 1$ and $p(I|\varrho_w) = 1$ accordingly. For the extreme case in which all opinion-level features are missing, Eq. 9 becomes:

$$\varrho_w^* = \underset{\varrho_w \in \{1,-1\}}{\operatorname{argmax}} \; p(\varrho_w), \tag{10}$$

which indicates that polarity of the sentiment word in this case is determined randomly by the polarity distribution $p(\varrho_w)$. As no author intends to confuse the review readers, this case scarcely happens for the ambiguous sentiment words except that inter-opinion features are given, which are discussed next.

2. Polarity disambiguation with inter-opinion features

Though the inter-opinion features influence polarity of the ambiguous sentiment word linguistically at three levels (i.e., sentence, discourse and application), the essential evidence is correlation between opinions. We represent the correlation with a 3 tuples: $< O_w, O_r, c_w >$, in which O_w denotes the current opinion containing the ambiguous sentiment word w, O_r the correlative opinion and c_w the correlative word that connects the two opinions. Polarity (pol) of the current opinion O_w is determined as follows.

$$\mathrm{pol(O_w) = pol(O_r)} \; \underset{\mathrm{f} \in \{1,-1\}}{\operatorname{argmax}} \; \mathrm{p(f|c_w)} \tag{11}$$

in which $p(f|c_w)$ represents the probability that correlative word c_w implies a function of either coordinating (1) or reversing (-1). They are both estimated using the training corpus.

In our work, $\mathrm{pol(O_w)}$ is calculated as follows.

$$\mathrm{pol(O_w)} = \varrho_\mathrm{w} \; (-1)^{|\mathrm{N}|} \tag{12}$$

in which $|\mathrm{N}|$ represents the number of occurrences of negation words. Thus

$$\varrho_w = \mathrm{pol(O_r)} \; (-1)^{|\mathrm{N}|} \; \underset{\mathrm{f} \in \{1,-1\}}{\operatorname{argmax}} \; \mathrm{p(f|c_w)} \tag{13}$$

Equation 13 can be used to handle the context at sentence level and discourse level. As for the application-level context in which opinions in a whole paragraph hold the same polarity, we perform preprocessing so as to detect polarity of the paragraph pol(para) and replace $f(c_w)$ with it. Thus, we obtain the following:

$$\varrho_w = \mathrm{pol(O_r)} \; (-1)^{|\mathrm{N}|} \; \mathrm{pol(para)} \tag{14}$$

According to Eqs. 13 and 14, calculation of $\mathrm{pol(O_r)}$ is a prerequisite to word polarity disambiguation. In most cases, polarity of an opinion can be computed with intra-opinion context. But for those unresolvable ones, we assign $\mathrm{pol(O_r)} = 0$ to discard the inter-opinion features in word polarity disambiguation. For the cases in which no correlative word is found in the context, we assign a coordinating function by default. That is, $p(1|c_w)$ is assigned 1 and $p(-1|c_w)$ assigned 0. This is safe for most cases.

3. The unified model

Observations show that intra-opinion features play an essential role in word sentiment polarity. They help in resolving polarity of most sentiment words. As a supplement, the inter-opinion features play a secondary role. They can help in confirming a good prediction and improving confidence. We let ϱ_w^Θ and ϱ_w^Φ represent polarity resolved with intra-opinion features and inter-opinion features, respectively. According to the above law, ϱ_w^Θ gives an initial orientation, while ϱ_w^{Φ} further enhances the orientation. We propose the following model to unify the two models:

$$\varrho_w = \text{sign}\left(\varrho_w^{\Theta}|\varrho_w^{\Theta} + \varrho_w^{\Phi}|\right) \tag{15}$$

Next, we discuss the effectiveness of the proposed unified model in two scenarios:

- Scenario #1: Intra-opinion features and inter-opinion features make same predictions.

According to Eq. 15, if $\varrho_w^{\Theta} = 1$ and $\varrho_w^{\Phi} = 1$, thus $\varrho_w = \text{sign}(2) = 1$; if $\varrho_w^{\Theta} = -1$ and $\varrho_w^{\Phi} = -1$, thus $\varrho_w = sign(-2) = -1$. In this scenario, ϱ_w^{Φ} confirms and enhances ϱ_w^{Θ}.

- Scenario #2: Intra-opinion features and inter-opinion features make conflict predictions.

According to Eq. 15, if $\varrho_w^{\Theta} = 1$ and $\varrho_w^{\Phi} = -1$, thus $\varrho_w = \text{sign}(0) = 0$; if $\varrho_w^{\Theta} = -1$ and $\varrho_w^{\Phi} = 1$, thus $\varrho_w = \text{sign}(0) = 0$. In this scenario, ϱ_w^{Φ} corrects ϱ_w^{Θ}.

4. Parameter estimation

The following parameters should be estimated with the training corpus:

- $p(t|\varrho_w)$: The probability of the opinion target being used in opinions regarding polarity (1: positive; -1: negative). The opinion target t should appear in the training corpus.
- $p(m|\varrho_w)$: The probability of the modifying word m being used in opinions regarding polarity (1: positive; -1: negative). The modifying word m should appear in the training corpus.
- $p(g_I^j|\varrho_w)\{j = 1, \ldots, L\}$: The probability of the indicative ngram g_I^j being used in opinions regarding polarity (1: positive; -1: negative). For zero-frequency cases, we apply the add-one smoothing technique.
- $p(f|c_w)$: The probability of the correlative word c_w being used in reviews regarding functions (1: coordinating; -1: reversing). The correlative word word c_w should appear in the training corpus.

In this work, we use Opinmine corpus [24] for parameter estimation. Details of the Opinmine corpus are given in "Setup" section.

Evaluation

Setup

The Polarity-Ambiguous Sentiment Words

The sentiment words used in this evaluation are automatically extracted from the Opinmine corpus with the following steps employed in each domain:

1. Sort the sentiment words with count of occurrences of the sentiment keywords appearing in the Opinmine corpus;
2. Delete the words from the list which are judged as holding a unique polarity in all reviews.
3. Select top 20 sentiment-ambiguous sentiment words for evaluation.

The 20 sentiment words selected from the four domains are presented in Tables 7, 8, 9 and 10. We find most of these words are included in the SemEval-2010 Task on disambiguating sentiment-ambiguous adjectives [20].

Note that keywords appear for different number of times in the four domains. So the 20 selected sentiment-ambiguous keywords are slightly different in each of the four domains. The common keywords are obvious.

Training/Test Corpus

We use the Opinmine opinion corpus [24] as training/test corpus in this evaluation. Four domains are involved in the second version: *digital camera*, *mobile phone*, *hotel* and *restaurant*. Statistics of the Opinmine corpus v2 are given in Table 11.

As this work is focused on polarity disambiguation, we only use reviews that contain the aforementioned sentiment-ambiguous keywords. We conduct experiments in every domain separately. As opinion annotations are less than 10K in each domain, we adopted the fivefold cross-validation approach in the experiments.

The Opinmine corpus contains annotations in four domains. As we do not investigate the cross-domain method, we conduct experiments in the four domains, separately.

Evaluation Metrics

The goal of the proposed method is to determine positive or negative polarity of a sentiment keyword in a given context. We thus naturally adopt accuracy as a measure of evaluation, with accuracy defined as the proportion of the correctly determined reviews within all test reviews.

Experiment 1: Different Methods for Word Polarity Disambiguation

In this experiment, we intend to compare our proposed Bayesian model-based method against the following state-of-the-art methods for word polarity disambiguation:

- *Pattern-based method* (PTN): Patterns are finely designed based on words in sentiment lexicon [27] and applied in word polarity disambiguation. In this

Table 7 The 20 polarity-ambiguous sentiment keywords in *mobile phone* domain (POS represents part of speech)

Word	POS	Word	POS	Word	POS	Word	POS
(*Many*)	Adj	(*Big*)	Adj	(*Improve*)	Verb	(*Decrease*)	Verb
(*Prominent*)	Adj	(*Low*)	Adj	(*High*)	Adj	(*Sensitive*)	Adj
(*Quick*)	Adj	(*Thin*)	Adj	(*Heavy*)	Adj	(*Decrease*)	Verb
(*Small*)	Adj	(*Light*)	Adj	(*Increase*)	Verb	(*Miracle*)	Noun
(*Simple*)	Adj	(*Serious*)	Adj	(*Drop*)	Verb	(*Surprise*)	Verb

Table 8 The 20 polarity-ambiguous sentiment keywords in *digital camera* domain

Word	POS	Word	POS	Word	POS	Word	POS
(*Many*)	Adj	(*Big*)	Adj	(*Improve*)	Verb	(*Small*)	Adj
(*High*)	Adj	(*Low*)	Adj	(*Light*)	Adj	(*Miracle*)	Noun
(*Quick*)	Adj	(*Little*)	Adj	(*Improve*)	Verb	(*decrease*)	Verb
(*Decrease*)	Verb	(*Prominent*)	Adj	(*Increase*)	Verb	(*Sensitive*)	Adj
(*Simple*)	Adj	(*Heavy*)	Adj	(*Drop*)	Verb	(*Surprise*)	Verb

Table 9 The 20 polarity-ambiguous sentiment keywords in *hotel* domain

Word	POS	Word	POS	Word	POS	Word	POS
(*Big*)	Adj	(*Special*)	Adj	(*Serious*)	Adj	(*Surprise*)	Verb
(*High*)	Adj	(*Little*)	Adj	(*Increase*)	Verb	(*Drop*)	Verb
(*Soft*)	Adj	(*Long*)	Adj	(*Low*)	Adj	(*Heavy*)	Adj
(*Simple*)	Adj	(*Quick*)	Adj	(*Improve*)	Verb	(*Interesting*)	Adj
(*Honest*)	Adj	(*Prominent*)	Adj	(*Decrease*)	Verb	(*Surprise*)	Noun

Table 10 The 20 polarity-ambiguous sentiment keywords in *restaurant* domain

Word	POS	Word	POS	Word	POS	Word	POS
(*Big*)	Adj	(*Hot*)	Adj	(*Light*)	Adj	(*Interesting*)	Adj
(*Honest*)	Adj	(*Low*)	Adj	(*Increase*)	Verb	(*Surprise*)	Noun
(*High*)	Adj	(*Impressive*)	Adj	(*Hard*)	Adj	(*Long*)	Adj
(*Many*)	Adj	(*Little*)	Adj	(*Thick*)	Adj	(*Quick*)	Adj
(*Small*)	Adj	(*Special*)	Adj	(*Hot*)	Adj	(*Prominent*)	Adj

Table 11 Statistics of Opinmine corpus v2

Domain	# of reviews	# of opinions	# of unique sentiment words
Mobile phone	1,200	6,034	1,437
Digital camera	1,200	9,706	1,689
Hotel	1,200	5,566	1,500
Restaurant	1,200	4,869	1,192

implementation, we use the HowNet [7] sentiment lexicon to handle Chinese reviews.

- *PMI-based statistical method* (PMI): Point-wise mutual information (PMI) is used to calculate sentiment orientation of sentiment words within the review corpus [18]. The starting seeds used in this work are translated to Chinese in order to handle Chinese word polarity disambiguation.
- *Machine learning method* (ML): A polarity classifier is trained on the Opinmine corpus and then applied to predict polarity of a sentiment word in a given context [?].
- *Bayesian model-based method* (Bayes): Our proposed method based on the Bayesian model is trained on the Opinmine corpus to first obtain the model parameters. These are then used to calculate the probability that a sentiment word is assigned a certain polarity (in accordance with Eq. 15).

Experimental results are presented in Table 12.

We can see from Table 12 that the word polarity disambiguation method based on Bayesian model outperforms the baseline methods significantly in the four domains. Firstly, the pattern-based method performs worse (i.e.,

Table 12 Experimental results of different word polarity disambiguation methods

Domain	PTN	PMI	ML	Bayes
Mobile phone	0.749	0.761	0.781	0.829
Digital camera	0.733	0.751	0.764	0.807
Hotel	0.704	0.733	0.753	0.792
Restaurant	0.721	0.755	0.772	0.804
Average	0.727	0.75	0.768	0.808

−1 % on average) than our method in this experiment. We find this is largely because of the limited coverage of the hand-compiled patterns and lexicon. Reviews are too flexibly given on social media or e-commerce sites for the patterns to handle. Secondly, the PMI-based method is also less effective (i.e., −5.8 % on average) than our method. Study shows that the PMI equation relies on a much bigger corpus to produce reasonable statistics. Thus, we believe that a bigger corpus may improve the PMI-based method. At last, the machine learning method is inferior to our method by 4.0 % on average. Similar to the PMI-based method, the machine learning method may rely on more training data, which is rather difficult to obtain. As a comparison, our method achieves around 80 % on accuracy with the small training corpus. From the encouraging results, we do see a significant performance gain when the Bayesian model is used, which justifies the advantage of the Bayesian model in word polarity disambiguation. Meanwhile, we believe the outperformance is due partially to the opinion-level features as the baseline methods all use term in matching or learning. This leads to a conclusion that opinion-level features are more effective than the term-level features in word polarity disambiguation.

Experiment 2: Different Bayesian Models for Word Polarity Disambiguation

In this experiment, we intend to investigate how features influence the Bayesian model in word polarity disambiguation. The following Bayesian models are implemented:

- *Term-based Bayesian model* (TMB): A Bayesian model using term-level features and determining polarity of a sentiment word with Eq. 6.
- *Opinion-based Bayesian model* (OPB#1): A Bayesian model using intra-opinion features and determining polarity of a sentiment word with Eq. 9.
- *Opinion-based Bayesian model* (OPB#2): A Bayesian model using both intra-opinion features and inter-opinion features, and determining polarity of a sentiment word with Eq. 15. Note that the inter-opinion features usually play as a secondary role in word polarity disambiguation. We do not implement the opinion-based Bayesian model with merely inter-opinion features.

Table 13 Experimental results of different Bayesian models

Domain	TMB	OPB#1	OPB#2
Mobile phone	0.728	0.804	0.829
Digital camera	0.705	0.782	0.807
Hotel	0.689	0.761	0.792
Restaurant	0.701	0.785	0.804
Average	0.746	0.783	0.808

Experimental results are presented in Table 13.

It can be seen from Table 13 that, with the same Bayesian model, opinion-level features significantly outperform term-level features significantly in each of the four domains. Firstly, the two opinion-level methods outperform the term-level method by 6.2 % in OPB#1 and by 3.7 % in OPB#2, on average. This proves that opinion-level features are more effective than term-level features for word polarity disambiguation. Secondly, it can be seen that inter-opinion features improve the opinion-level method consistently by 2.5 % on average. This indicates that the inter-opinion features are indeed effective for word polarity disambiguation.

Per-word performance analysis

To observe how the proposed Bayesian model performs on individual words, we present accuracy values of three Bayesian model-based methods on sentiment words in the four domains, in Fig. 2.

Figure 2 shows that on all sentiment words, the contribution of opinion-level features is consistent. This further confirms the aforementioned two conclusions. However, the contribution of individual words varies. For example, OPB#1 improves TMB the most on sentiment word (*serious*) by 12.7 %, while the least on (*thin*) by 2.4 %. OPB#2 improves TMB the most on (*increase*) by 14.6 %, while least on (*simple*) by 5.2 %. Two reasons are worth noting. Firstly, polarity of some words can be well resolved with term-level features (e.g., (*thin*)). So opinion-level features can actually improve very little. But for other cases, e.g., (*serious*), the term-level features are rather hopeless, whereas opinion-level features can work pretty effectively. This leads to a significant improvement. Secondly, volume of training data for different words varies. As the corpus is naturally collected, it is very common for some words to appear more frequently than others. This leads to various accuracy levels.

We also notice in Fig. 2 that OPB#2 improves OPB#1 most on (*improve*) by 5.5 % while least on (*high*) by 1 %. In fact, OPB#2 improves OPB#1 from 75.2 to 80.7 % on (*improve*). We find in many reviews the polarity of (*improve*) cannot be easily resolved with the intra-opinion

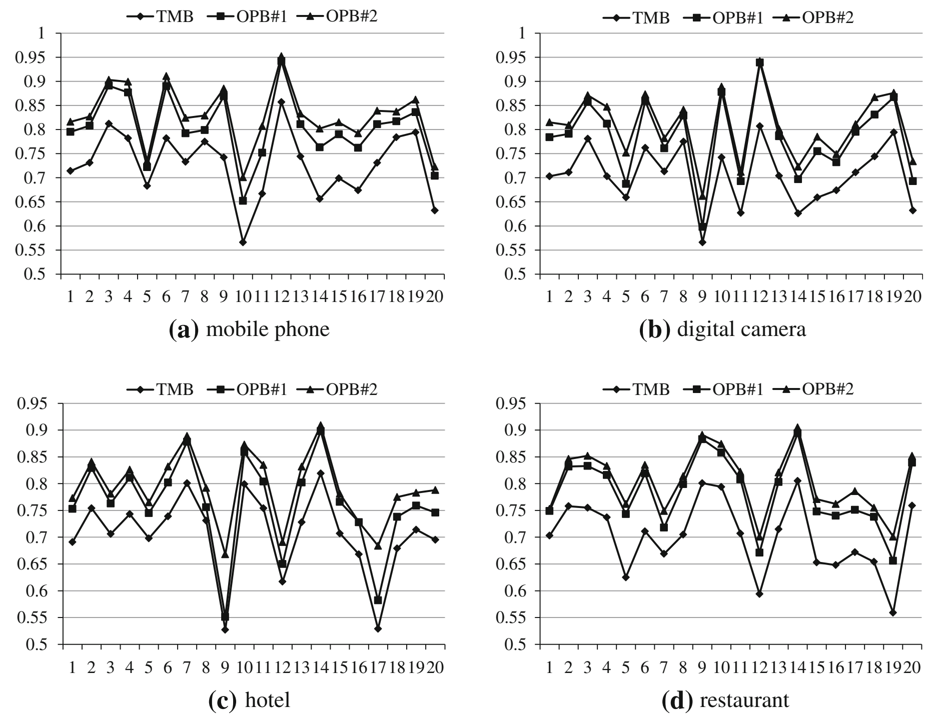

Fig. 2 Per-word performance analysis of Bayesian model-based methods in four domains

features alone. As a comparison, OPB#2 improves OPB#1 from 94.2 to 95.2 % on (*high*). This is because intra-opinion features are powerful enough in resolving polarity of (*high*). Thus, there remains very little space for the inter-opinion features to play a further role. From the above observation, we conclude that inter-opinion features can effectively complement intra-opinion features in word polarity disambiguation.

Conclusions

Contextual polarity ambiguity is an important problem in sentiment analysis. In this work, we study this problem with reviews. In contrast to previous work which makes use of term-level features, we propose to resolve the polarity ambiguity with opinion-level features. Specifically, we investigate intra-opinion features such as opinion target, modifying word and indicative words, as well as inter-opinion features such as correlative words in sentence, discourse and application. We adopt the Bayesian model and deal with the word polarity disambiguation task in a probabilistic manner. Experiments using the Opinmine corpus show that opinion-level features can make a significant contribution in word polarity disambiguation across the four domains.

Acknowledgments This paper is supported by NSFC (61272233) and the Royal Society of Edinburgh. We are grateful to the reviewers for their invaluable comments.

References

1. Balahur A, Montoyo A. Opal: applying opinion mining techniques for the disambiguation of sentiment ambiguous adjectives in semeval-2 task 18. Proceedings of the 5th international workshop on semantic evaluation., SemEval'10. Stroudsburg, PA, USA: Association for Computational Linguistics; 2010; p. 444–7.
2. Cambria E, Hussain A. Sentic computing: techniques, tools, and applications. Dordrecht: Springer; 2012.
3. Cambria E, Olsher D, Rajagopal D. SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In: AAAI. Quebec City; 2014. p. 1515–21.
4. Cambria E, Schuller B, Xia Y, Havasi C. New avenues in opinion mining and sentiment analysis. IEEE Intell Syst. 2013;28(2):15–21.
5. Cambria E, White B. Jumping NLP curves: a review of natural language processing research. IEEE Comput Intell Mag. 2014;9(2):48–57.
6. Ding X, Liu B, Yu PS. A holistic lexicon-based approach to opinion mining. Proceedings of the 2008 international conference on web search and data mining., WSDM'08. New York, NY, USA: ACM; 2008. p. 231–40.
7. Dong Z, Dong Q. Hownet and the computation of meaning. River Edge, NJ, USA: World Scientific Publishing Co.; 2006.
8. Esuli A, Sebastiani F. Sentiwordnet: a publicly available lexical resource for opinion mining. In: Proceedings of the 5th conference on language resources and evaluation (LREC06), pp. 417–422 (2006).
9. Hatzivassiloglou V, McKeown KR. Predicting the semantic orientation of adjectives. 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., ACL'98. Stroudsburg, PA, USA: Association for Computational Linguistics; 1997. p. 174–81.
10. He Y, Alani H, Zhou D: Exploring english lexicon knowledge for chinese sentiment analysis. In: CIPS-SIGHAN Joint conference on Chinese language processing (2010).

11. Lu B, Tsou BK. Cityu-dac: Disambiguating sentiment-ambiguous adjectives within context. Proceedings of the 5th international workshop on semantic evaluation., SemEval'10. Stroudsburg, PA, USA: Association for Computational Linguistics; 2010. p. 292–5.
12. Pak A, Paroubek P. Twitter based system: using twitter for disambiguating sentiment ambiguous adjectives. Proceedings of the 5th international workshop on semantic evaluation., SemEval'10. Stroudsburg, PA, USA: Association for Computational Linguistics; 2010. p. 436–9.
13. Popescu AM, Etzioni O. Extracting product features and opinions from reviews. Proceedings of the conference on human language technology and empirical methods in natural language processing., HLT'05. Stroudsburg, PA, USA: Association for Computational Linguistics; 2005. p. 339–46.
14. Poria S, Cambria E, Winterstein G, Huang GB. Sentic patterns: dependency-based rules for concept-level sentiment analysis. Knowledge-Based Systems (2014).
15. Qiu L, Zhang W, Hu C, Zhao K. Selc: a self-supervised model for sentiment classification. In: CIKM, pp. 929–936 (2009).
16. Torii Y, Das D, Bandyopadhyay S, Okumura M. Developing japanese wordnet affect for analyzing emotions. Proceedings of the 2Nd workshop on computational approaches to subjectivity and sentiment analysis., WASSA'11. Stroudsburg, PA, USA: Association for Computational Linguistics; 2011. p. 80–6.
17. Turney PD. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. Proceedings of the 40th annual meeting on Association for Computational Linguistics., ACL'02. Stroudsburg, PA, USA: Association for Computational Linguistics; 2002. p. 417–24.
18. Turney PD, Littman ML. Measuring praise and criticism: inference of semantic orientation from association. ACM Trans Inf Syst. 2003;21(4):315–46.
19. Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis. Proceedings of the conference on human language technology and empirical methods in natural language processing., HLT'05. Stroudsburg, PA, USA: Association for Computational Linguistics; 2005. p. 347–54.
20. Wu Y, Jin P. Semeval-2010 task 18: disambiguating sentiment ambiguous adjectives. Lang Resour Eval. 2013;47(3):743–55.
21. Wu Y, Wen M. Disambiguating dynamic sentiment ambiguous adjectives. Proceedings of the 23rd international conference on computational linguistics., COLING'10. Stroudsburg, PA, USA: Association for Computational Linguistics; 2010. p. 1191–9.
22. Xia Y, Hao B, Wong KF. Opinion target network and bootstrapping method for chinese opinion target extraction. Proceedings of the 5th Asia information retrieval symposium on information retrieval technology., AIRS'09. Berlin, Heidelberg: Springer; 2009. p. 339–50.
23. Xia Y, Wang L, Wong KF, Xu M. Sentiment vector space model for lyric-based song sentiment classification. Proceedings of the 46th Annual Meeting of the association for computational linguistics on human language technologies: short papers., ACL-Short'08. Stroudsburg, PA, USA: Association for Computational Linguistics; 2008. p. 133–6.
24. Xu R, Xia Y, Wong KF, Li W. Opinion annotation in on-line chinese product reviews. In: LREC (2008).
25. Xu R, Xu J, Kit C. Hitsz_cityu. Combine collocation, context words and neighboring sentence sentiment in sentiment adjectives disambiguation. Proceedings of the 5th international workshop on semantic evaluation., SemEval'10. Stroudsburg, PA, USA: Association for Computational Linguistics; 2010. p. 448–51.
26. Yang SC, Liu MJ. Ysc-dsaa: an approach to disambiguate sentiment ambiguous adjectives based on saaol. Proceedings of the 5th international workshop on semantic evaluation., SemEval'10. Stroudsburg, PA, USA: Association for computational linguistics; 2010. p. 440–3.
27. Yi J, Nasukawa T, Bunescu R, Niblack W. Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques. In: Proceedings of the third IEEE international conference on data mining, ICDM'03, p. 427. IEEE Computer Society, Washington, DC, USA (2003).