Sentiment Analysis of Chinese Documents: From Sentence to Document Level

Changli Zhang
College of Computer Science and Technology, Jilin University, China, and Artillery Command College of Shenyang, China. E-mail:shenyangzcl@163.com

Daniel Zeng
MIS Department, University of Arizona and the Institute of Automation, Chinese Academy of Sciences, Beijing, China. E-mail: zeng@email.arizona.edu

Jiexun Li
College of Information Science and Technology, Drexel University, Philadelphia, PA. E-mail: jiexun.li@ischool.drexel.edu

Fei-Yue Wang
Institute of Automation, Chinese Academy of Sciences, Beijing, China. E-mail: feiyue@gmail.com

Wanli Zuo
College of Computer Science and Technology, Jilin University, China. E-mail: wanli@jlu.edu.cn

User-generated content on the Web has become an extremely valuable source for mining and analyzing user opinions on any topic. Recent years have seen an increasing body of work investigating methods to recognize favorable and unfavorable sentiments toward specific subjects from online text. However, most of these efforts focus on English and there have been very few studies on sentiment analysis of Chinese content. This paper aims to address the unique challenges posed by Chinese sentiment analysis. We propose a rule-based approach including two phases: (1) determining each sentence's sentiment based on word dependency, and (2) aggregating sentences to predict the document sentiment. We report the results of an experimental study comparing our approach with three machine learning-based approaches using two sets of Chinese articles. These results illustrate the effectiveness of our proposed method and its advantages against learning-based approaches.

Introduction

With the advance in Internet technologies, especially the wide adoption of Web 2.0 applications, online communities such as blogs, wikis, online forums, social-networking groups, etc., are becoming increasingly popular, producing rich and rapidly growing user-generated contents. As one application, the Web has become an increasingly popular communication platform for people to learn others' opinions and express their own. For instance, many E-commerce Web sites such as Amazon.com and Epinions.com encourage users to submit product reviews, either "thumbs-up (positive)" or "thumbs-down (negative)," and present the customer-generated reviews alongside with product information, hoping to provide support for the customers to make purchasing decisions. In another example, many public health-related topics, such as the legalization of euthanasia and AIDS patients' life quality, are being debated globally in many online forums and blogs, often with passion. The opinions and sentiments extracted from these online discussions can be of substantial help to policy making.

Regardless of the topic domain, a common characteristic of these online discussions or articles is that they often represent the author's certain sentiment orientation toward a particular topic. Given the large amounts of articles posted online, recent years have seen a lot of research efforts to develop techniques that can automatically mine opinions and analyze sentiments in articles. In effect, many opinion miners have been developed and adopted in application areas such as business intelligence and government intelligence. Specifically, product reviews often have strong sentiment orientation and

Received May 13, 2009; revised July 16, 2009; accepted July 20, 2009

 • Published online 2 September 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.21206

are important for companies to improve their products and services. Product reviews from E-Commerce sites (e.g., Amazon.com) and review sites (e.g., rottentomatoes.com) are popular test-beds for sentiment analysis (Mishne & Glance, 2006). Sentiment analysis techniques are also applied in government intelligence to monitor online communications for public opinion tracking or security surveillance (Laver, Benoit, et al., 2003; Efron, 2004; Mullen & Malouf, 2006; Thomas, Pang, et al., 2006; Bansal, Cardie, et al., 2008).

Most of the existing studies on sentiment analysis focus on articles written in English. However, another major language, Chinese, used by a large portion of the world's population as well as Internet users, has not attracted sufficient attention in the previous sentiment analysis literature. As compared to English, Chinese has several distinct linguistic characteristics, which pose several unique challenges in mining the sentiment of Chinese articles: (1) Chinese does not segment words by spaces in sentences; (2) The use of various adverbs in Chinese can lead to a higher degree of subtlety and ambiguity in sentences; (3) Chinese also shows a larger variety of word sense and syntactic dependency in sentences than English.

Due to such unique characteristics of Chinese, existing sentiment analysis techniques developed for English may not provide satisfactory performance for articles in Chinese. This research aims at developing a new technique to determine the overall sentiment orientation of Chinese articles. Our proposed technique first determines each sentence's sentiment polarity based on the occurring sentiment words and their syntactic context, and predicts the document's sentiment polarity by aggregating all sentences' polarities. The remainder of the paper is organized as follows. In the next section we review the existing literature on sentiment analysis. In the third section we introduce our proposed approach to analyzing sentiment of Chinese articles in detail. We have conducted comparative experiments and present the results in the fourth section. Last, we conclude the paper with a summary of our work and future directions.

Literature Review

Sentiment analysis techniques have been extensively used in areas such as text filtering, public opinion tracking, customer relationship management, and so on (Nasukawa & Yi, 2003; Liu, Hu, et al., 2005; Popescu & Etzioni, 2005; Riloff, Patwardhan, et al., 2006). Previous studies on sentiment analysis vary in their levels of granularity, analytical techniques, and languages. In this section we review related work from these three aspects.

Level of Granularity

Due to the different objectives of applications, previous studies tackled the sentiment analysis problem at different levels of granularity, from document level to sentence level. Some early attempts of sentiment analysis predicted the polarity of an article simply based on the occurrence of subjective words in the entire document. For example, Pang et al. (2002) classified articles' sentiments by adopting a standard bag-of-features framework, in which features are unigrams and bigrams of words. Turney (2002) proposed an unsupervised learning algorithm called PMI-IR (pointwise mutual information and information retrieval) to predict the semantic orientation of an article by calculating the similarity of its contained phrases to two reference words, "excellent" and "poor." It computes word similarity based on word co-occurrence in Web pages, and therefore its performance largely relies on the search engine adopted. Some recent studies also took into account the spread, density, and intensity of polar lexical terms to improve the performance for sentiment classification (Tsou, Yuen, et al., 2005). However, these document-level analytical approaches rarely modify the polarity of individual subjective words by considering their contexts; as such, the derived overall polarity assessment is often not accurate.

To address this problem, a more thorough analysis at the sentence level has been explored. Yu and Hatzivassiloglou (2003) developed several sentence-level classifiers to detect sentences of opinions and predicted the sentiment polarity of sentences based on their likelihood ratio scores. Although their approach analyzes sentiment at the sentence level, the technique ignores the syntactic context in polarity assessment, which is similar to the document-level analysis. More sophisticated sentence-level sentiment analysis techniques should consider the contextual features of subjective words, including negation and intensity modifiers, to better predict sentence polarity. For example, Meena and Prabhakar (2007) represented sentences as dependency trees and considered the effect of conjunctions to analyze the sentence structures for sentiment classification.

In many applications the goal of sentence-level analysis is to provide better prediction of the document's overall polarity. There are two approaches to make fine-to-coarse (i.e., sentence-to-document) sentiment analysis: a cascaded approach and a joint approach. A cascaded approach first computes the polarity of each sentence and then inputs the result into a document-level classifier. For example, Mao and Lebanon (2006) used a sequential conditional random fields (CRF) model to measure polarity on the sentence level in order to determine the "sentiment flow" of reviews. Pang and Lee (2004) used a global minimum-cut inference algorithm to identify each sentence as subjective or objective. The top identified subjective sentences are then input into a document-level polarity classifier with improved performance compared to approaches that consider all sentences in the document. In contrast, a joint approach combines sentence-level and document-level analysis in a single joint model to determine the overall polarity. For example, McDonald et al. (2007) proposed a CRF-based structured model for sentence-document sentiment analysis. This integrated model is claimed to be simple and computationally traceable, and it outperforms polarity classifiers trained in isolation as in the cascaded approach. However, the model requires large training datasets at both sentence and document level, which are often very costly and time-consuming.

Analytical Techniques

In terms of the types of analytical techniques, existing sentiment analysis approaches can be categorized into rule-based and learning-based approaches. Rule-based approaches often require an expert-defined dictionary of subjective words and predict the polarity of a sentence or document by analyzing such words' occurring patterns in text. For example, Wiebe et al. (2004) provided a lexicon source of subjectivity clues such as verbs, adjectives, and nouns with their polarity (positive, negative, or neutral) and strength (strong or weak) annotated. Such a lexicon can only define the prior polarity of a word and yet the actual polarity of a word can be modified by its context in a sentence. Several approaches have been proposed to determine the sentiment orientation of words by considering its context such as modifiers and word dependencies. For example, Hatzivassiloglou and McKeown (1997) predicted the semantic orientation of adjectives by analyzing them in pairs conjoined by "and," "or," "but," "either-or," or "neither-nor." Yuen et al. (2004) proposed an approach to deriving words' semantic polarity based on morpheme. Knowledge sources such as WordNet have also been used to measure adjectives' semantic polarity (Kamps, Marx, et al., 2004). Rules for sentiment analyses are usually formed as sequence patterns of certain linguistic features. For example, Turney (2002) summarized several sequence patterns of POS tags (e.g., RB/RBR/RBS+JJ+NN/NNS, JJ+NN/NNS+Anything RB+VB+Anything) for sentiment analysis. Similarly, Nasukawa and Yi (2003) used patterns such as "subjective verbs + target" and "subjective adjectives + target" and defined some transfer verbs (e.g., "get" and "feel") in the patterns as indicators of sentiment expressions. Popescu and Etzioni (2005) defined 10 types of rules to extract opinion patterns. Each of these rules represents a certain pattern including a feature (target) and subjective words. These rule-based approaches analyze the sentence sentiment by matching a set of predefined linguistic patterns as templates.

Unlike rule-based approaches that require predefining rules, machine-learning approaches can construct predictive models by learning from labeled training datasets. For example, Hu and Liu (2006) developed an approach to extracting option features from product reviews based on linguistic patterns called Class Sequential Rules (CSR), which can be mined from a set of labeled training sequences of words and part-of-speech (POS) tags. Pang et al. (2002) represented reviews as a bag of unigram/bigram features and applied three machine-learning methods to predict their sentiment. However, they found that, for sentiment classification, machine-learning algorithms did not perform as well as traditional topic categorization tasks. Besides, learning-based sentiment classification requires sufficiently large training datasets with positive and negative examples manually labeled, which is often very costly and time-consuming (Turney, 2002). Wiebe and Riloff (2005) proposed a rule-based method to automate the process of training data annotation. However, the rule-based method did not consider terms' context and thus could not produce reliable and high-quality training data for machine learning.

Language

Due to the early advance and prevalence of online communications in English-speaking communities, most sentiment analysis studies have focused on English and achieved success in many applications. However, Chinese sentiment analysis remains an area that has not attracted sufficient investigation efforts. The unique linguistic characteristics of Chinese pose several technical challenges for Chinese sentiment analysis. Unlike most Western languages such as English, Chinese does not segment words by spaces in sentences. Therefore, word segmentation is often required as an additional step in Chinese language processing. Moreover, Chinese contains various adverbs and the use of these adverbs can lead to subtlety and ambiguity in sentences. English mainly uses suffixes (-er and -est) to express comparative and superlative words, whereas Chinese uses various degree adverbs such as "更/more' and "最/most." As such, determining the sentiment polarity of Chinese sentences is a more difficult task, especially when multiple adverbs and subjectivity clues appear in one sentence. A document that contains several positive words may actually indicate a strong negative tone, and vice versa.

To the best of our knowledge, there are only a limited number of studies that focus on Chinese sentiment analysis (e.g., Tsou, Yuen, et al., 2005; Yuen, Chan, et al., 2004; Ye, Shi, et al., 2006; Su, Xu, et al., 2008). Yuen et al. (2004) presented a method based on Turney and Littman (2003) to determine the semantic polarity of a Chinese word based on its morpheme and its statistical association with strongly polarized words. In a later work, Tsou et al. (2005) extended this method by considering the spread, density, and intensity of polar lexical terms to improve sentiment classification. Su et al. (2008) proposed an approach to mining opinions on different product features from Chinese reviews. This approach does not consider words' context when predicting an opinion word's polarity. A word's context in a sentence can be represented as its dependencies with other words. Ye et al. (2006) proposed an approach that considers binary dependencies between words for sentiment analysis of Chinese articles. This approach extracts word-word phrases that match one of five POS patterns (e.g., adjective-noun) and predicts the polarity of each phrase by counting its co-occurrences with certain reference words pair (RWP) (e.g., "excellent" and "poor") on the Web. Chinese sentences often contain complex dependency relations among words and they should be taken into consideration for sentiment analysis. Moreover, most existing Chinese sentiment analysis approaches (e.g., Abbasi, Chen, et al., 2008) are based on machine learning. These approaches can achieve good performance but require

DOI: 10.1002/asi

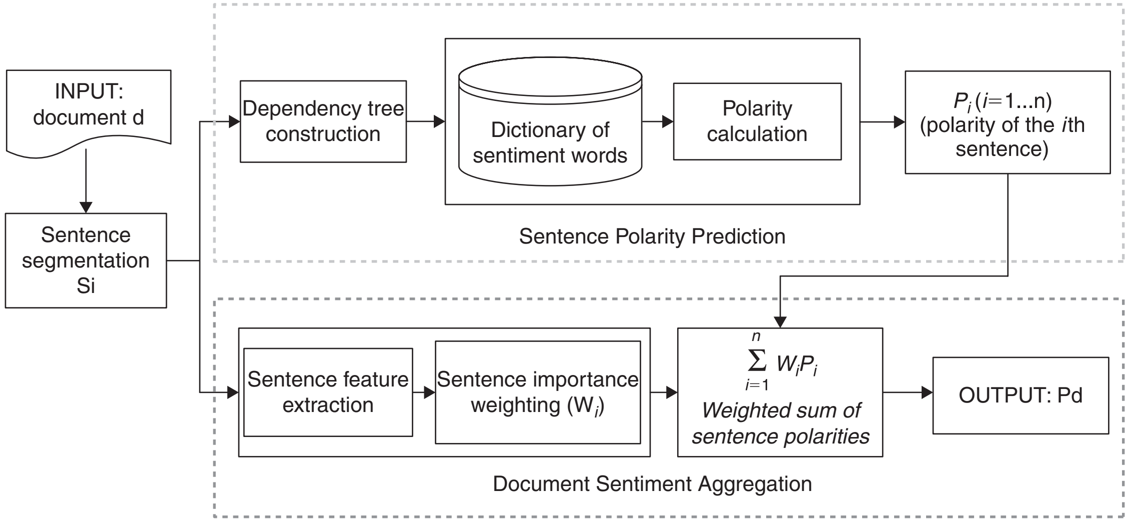

FIG. 1. Rule-based sentiment analysis framework.

a high cost of data annotation. Therefore, in order to overcome these limitations, our study aims at developing a new approach for Chinese sentiment analysis with high accuracy and at no cost of data annotation.

An Approach to Chinese Sentiment Analysis

Kim and Hovy (2004) described an opinion as a quadruple [Topic, Holder, Claim, Sentiment], in which the Holder makes a Claim about the Topic, and often associates with certain Sentiment, such as good or bad. Sentiments always involve the Holder's emotions or desires, and may be presented explicitly or implicitly in many cases. In this paper, we aim to address the following question: Given a topic and a relevant article in Chinese, how do we analyze the claims made by the holder and determine the sentiment polarity of the article?

Overall Method

In order to tackle the unique challenges posed by Chinese, we propose a rule-based approach for sentiment analysis of Chinese documents. Figure 1 shows the overall architecture of our proposed approach, including two major steps: sentence sentiment analysis and document sentiment aggregation. Mindful of the subtlety of Chinese expression, we first decomposed a document into its constituting sentences and determine the sentiment polarity of each sentence. As opposed to other document-level analytical approaches (Turney, 2002; Ye et al., 2006), we treated sentences as atomic units for semantic analysis. Next, the polarity scores of all sentences were synthesized to compute the overall polarity of the entire document. The importance of a sentence to a document can be represented as weight in the overall polarity computation. The thematic sentences should be assigned a greater weight than others in a document. We formalized the problem as follows.

Given a document d containing sentences $\{s_1, \ldots, s_n\}$ as input, the system is to calculate the polarity score pi of each sentence si and determine the sentiment polarity P^d of d:

$$P^d = \sum_i w_i p_i$$

where w_i is the weight of sentence s_i. If $P^d > 0$, the document shows positive sentiment; otherwise, it shows negative sentiment.

Sentence Sentiment Analysis

In the above section we reviewed several methods for sentence sentiment analysis (e.g., Yu & Hatzivassiloglou, 2003; Meena & Prabhakar, 2007; Wiebe & Riloff, 2005; Wiebe, Bruce, et al., 1999). In this paper we propose a new method of sentence sentiment analysis based on a sentiment word lexicon and sentence syntactic structures.

Subjective sentence identification. A document can be segmented into multiple sentences, from which the overall polarity of the document can be inferred. However, not all the sentences are subjective and show sentiment polarity. We screened out the subjective sentences based on the occurrence of subjective words for further analysis. The remaining sentences are disregarded because they do not help infer the polarity of the document. Specifically, we adopt a dictionary of Chinese subjective words summarized by HowNet (Dong and Dong, 2003). HowNet is an online common-sense knowledge system revealing interconceptual relations and interattribute relations of concepts as connoted in Chinese and English bilingual lexicons. It has been used in many studies of Chinese NLP (Veale, 2005; Yan, Bracewell, et al., 2007; Zhu, Zheng, et al., 2008). HowNet provides a comprehensive dictionary of Chinese subjective words, including: (1) 3,730 positive opinionated words (e.g., 漂亮/pretty) and 3,116 negative opinionated words (e.g., 丑陋/ugly); (2) 836 positive affective words

DOI: 10.1002/asi

TABLE 1. Types of dependency relations between words in Chinese.

Relation type	Description	Example	Dependency
ATT(attribute)	定中关系	这部相机/this camera	这部←相机
COO(coordinate)	并列关系	美 丽 大 方/beautiful and decent	美丽→大方
CMP(complement)	动补结构	好得很/quite good	好→得很
MT(mood-tense)	语态结构	好极了/extremely good	好→极了
DE	"的"字结构	精彩的/wonderful	精彩←的
ADV(adverbial)	状中结构	很好/very good	很←好
SBV(subject-verb)	主谓关系	我赞成/I agree	我←赞成
VOB(verb-object)	动宾关系	是错误/is a mistake	是→错误

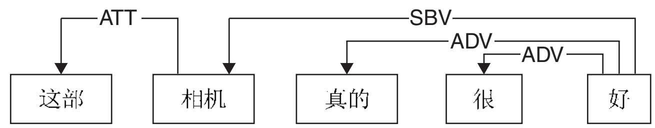

FIG. 2. Dependency relations of an example sentence (这部相机真的很好).

(e.g., 爱慕/love) and 1,254 negative affective words (e.g., 悲伤/sad); and (3) 219 degree adverbs (e.g., 很/very). Besides, negative adverbs are recognized by morpheme-based and rule-based method (e.g., 否/not + suffix, 不/no + suffix, 勿/do not + suffix, 毋/not + verb morpheme, words such as 禁止/forbid). HowNet also quantifies the degree adverbs according to their intensity. For example, "非常" has a degree of 2 and "很" has a degree of 1.5, though they have the same meaning as "very" in English. We will make use of these degrees in predicting sentences' sentiment polarity.

Dependency tree construction. Due to the subtlety of Chinese language expression, simply predicting the polarity of a sentence based on the subjective words it contains is not sufficient. More thorough analysis of the syntactic structure is necessary to determine the sentiment polarity. In this study we used an open-source Chinese dependency grammar parser, called Harbin Institute of Technology Information Retrieval Lab Language Technology Platform (HIT-IR LTP), developed by the Harbin Institute of Technology to convert each sentence into a dependency tree. We selected this parser because of its high accuracy for parsing Chinese sentences (Wanxiang Che et al., 2008).

In a dependency tree, each node represents a word and the tree is composed of multiple binary relations among words. Each relation has a word as a parent (or head) and the other one as a child (or modifier). Each word has one and only one parent, while a word may have multiple children. In Chinese, the dependency relations between two words can be categorized into eight types (Table 1).

Figure 2 shows an example of a dependency tree constructed by HIT-IR LTP for the sentence: "这部相机真的很好" ("This camera is really very good"). A relation denoted with an arrow starts from the parent and directs to the child. The tag on an arrow denotes the relation type. In this example, words "相机/camera," "真的/really," and "很/very" are the child nodes of the subjective word "好/good." The words "真的/really" and "很/very" modify "好/good" directly.

Sentence polarity prediction. Given a sentence consisting of subjective words, to determine the polarity is still a nontrivial task. A word's sentiment polarity may have the following three forms: prior polarity, modified polarity, and dynamic polarity (Wilson, Wiebe, et al., 2005).

Prior polarity is the general polarity of the subjective word. There are different ways to obtain a subjective word's prior polarity. One simple approach is to obtain polarity information from a certain dictionary of subjective words. This dictionary can be predefined based on domain knowledge or sentiment-bearing adjectives extracted from sources such as WordNet. Some studies estimated the prior polarities of words using an information retrieval method. In this study, we used HowNet as a dictionary of Chinese subjective words.

The actual polarity of a subjective word is not always the same as its prior polarity. Modified polarity refers to the situation where some surrounding modifier words may change the intensity or even the polarity of the subjective words. Such modifier words include negative words and degree adverbs. Specifically, negative words can reverse the polarity of sentiment words or opinion words and degree adverbs can strengthen or weaken the word intensity in the sentence. Not taking into account such modified polarity may lead to incorrect prediction of sentiment polarity.

Dynamic polarity refers to subjective words whose polarity is object- or context-dependent. For example, an "unpredictable" camera indicates a negative opinion of the camera, whereas a movie with an "unpredictable" plot sounds positive to moviegoers. Dynamic polarity is more difficult to deal with and is not considered in our current study.

In order to determine the polarity of a subjective word, we took into account its dependent modifiers in the sentence. Each word has its prior polarity defined in a subjective word dictionary. When appearing in a sentence represented as a dependency tree, a word may or may not have a number of children that can modify its polarity. By studying the linguistic patterns of Chinese sentences, we summarized several rules for polarity modification based on two factors: (1) the

DOI: 10.1002/asi

type of dependency relations between a subjective word and its children, and (2) the type of the children, i.e., negation or degree adverbs. The heuristics of these rules are summarized as follows:

- If a word has a VOB (e.g., 是→错误), COO (e.g., 美丽→大方), or DE (e.g., 精彩←的) dependency with a child, its polarity is determined by the child's polarity (see Table 1 for definitions).
- If a word has a CMP (e.g., 好→得很) or MT (e.g., 好→极了) dependency with a child word and the child word is a degree adverb, the polarity of the parent word is modified by an operation named intensification, i.e., multiplying the intensity of the child.
- Polarity modification is more complicated if the dependency is ADV or ATT, In this case, the parent word may have multiple children, either negative words or degree adverbs or both. In Chinese, the combination of a degree adverb and a negative word in different sequential order can have a different effect on modifying the sentiment polarity.
  - If the word's closest child is a degree adverb with no negation next to it (e.g., 很←高兴), its polarity is modified by intensification, i.e., the prior polarity multiplied by the intensity of the adverb.
  - If the word's closest child is a negative word with no degree adverb next to it (e.g., 不←高兴), its polarity is modified by an operation named negation, i.e., the opposite of the prior polarity divided by two.
  - If the word has two consecutive children in the order of a degree adverb and a negative word (e.g., 很–不←高兴), the polarity of the sentiment word is modified by first applying the operations negation followed by intensification. The modified polarity is equal to the opposite of the prior polarity divided by two and multiplied by the intensity.
  - If the word has two consecutive children in the order of a negative word and a degree adverb (e.g., 不–很←高兴), the negation operation is not applied to the sentiment word but the degree adverb by weakening the intensity by 1/2 and then intensification is performed on the prior polarity using the modified intensity. The modified polarity is equal to the prior polarity times the intensity divided by two.

We implemented these word polarity prediction rules as follows:

w: a word
C(w): {c1, c2, ...} a list of w's children, order by the distance to w
ci: the ith child of w in C(w)
|C(w)|: the cardinality of C(w), i.e., the number of w's children
d(w, c): the dependency relation from w to c (w → c)
I(c): the intensity score of a degree adverb c
Q(w): a queue of child nodes ?
s: a sentence
p: the polarity of a word
p0: the prior polarity of a word
P: the polarity of a sentence
root: the root word of a dependency tree

An algorithm of word polarity prediction:

```
Procedure WordPolarity(w, s)
p = p0 //denote the prior polarity of the word
FOR i = 1 TO |C(w)|
  IF d(w, ci) is VOB, COO, or DE THEN p = WordPolarity
(ci, s) END IF
  IF d(w, ci) is CMP or MT THEN p = p ∗ I(ci) END IF
  IF d(w, ci) is ADV or ATT THEN Enqueue(Q, ci)
END IF
END FOR
WHILE(Q(w) ≠ Φ)
  IF(Q(w) ≠ Φ) THEN m1 = Dequeue(Q) END IF
  IF(Q(w) ≠ Φ) THEN m2 = Dequeue(Q) END IF
  IF(m1 is a degree adverb && m2 is a negative word)
THEN p = −p ∗ I(m1)/2 END IF
  IF(m1 is a negative word && m2 is a degree adverb)
THEN p = p ∗ I(m1)/2 END IF
  IF(m1 is a degree adverb && m2 is null) THEN p = p ∗
I(m1) END IF
  IF(m1 is a negative word && m2 is null) THEN
p = −p/2 END IF
END WHILE
IF (p = 0) THEN
  FOR i = 1 TO |C(w)|
    p = p + WordPolarity(ci, s)
  END FOR
END IF
RETURN p
```

A sentence is composed of its constituting words. Therefore, the polarity of a sentence P can be determined by its root node in the dependency tree. We designed an algorithm for sentence polarity prediction that computes each word's polarity score starting from the root node and in a recursive fashion:

```
Procedure SentencePolarity(s)
P = WordPolarity(root, s)
IF P ≠ 0 THEN
  RETURN P
ELSE
  FOR i = 1 to |C(root)|
    P = P + WordPolarity(ci, s)
  END FOR
  RETURN P
END IF
```

The following examples illustrate the computation of polarity in different situations, and Dependency Relations of these examples are shown in Figure 3a–e. The prior polarity of "高兴(happy)" is +1. The intensity of degree adverb "很(very)" is 1.5.

a) 我高兴。I am happy.
   Modified polarity = prior polarity = +1
b) 我很高兴。I am very happy.
   Modified polarity = prior polarity × intensity = +1 × 1.5 = +1.5
c) 我不高兴。I am not happy.
   Modified polarity = prior polarity × (−1)/2 = +1 × (−1)/2 = −0.5
d) 我很不高兴。I am very unhappy.
   Modified polarity = prior polarity × (−1)/2 × intensity = (+1) × (−1/2) × (1.5) = −0.75

DOI: 10.1002/asi

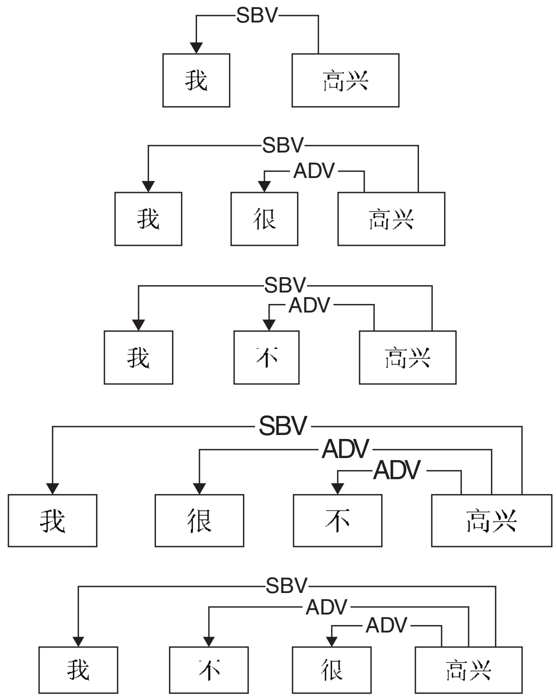

FIG. 3. Some examples of dependency relationships in sentences.

e) 我不很高兴。I am not very happy.
Dependency relations:
Modified polarity = prior polarity × intensity × 1/2 = (+1) × (1.5) × (1/2) = +0.75

Document Sentiment Aggregation

In the second section we reviewed several methods for document-level sentiment analysis. However, some of these methods (e.g., Pang et al., 2002; Turney, 2002; Tsou et al., 2005) are based on the occurrence of sentiment words without considering their context in a sentence, while some learning-based approaches require manual annotation of a training dataset (e.g., Pang & Lee, 2004; Mao & Lebanon, 2006; McDonald et al., 2007). Unlike these methods, we took a cascaded approach to predict the overall polarity of a document by aggregating polarity scores of individual sentences. Noting that sentences vary in their importance in a document, we assigned sentences different weights to adjust their contribution to the overall polarity. To measure a sentence's importance in a document, we defined five domain-independent features: the position of the sentence (w_p), the term weights in the sentence (w_t), the similarity between the sentence and the title (w_h), the occurrence of keywords (w_k), and the first-person mode (w_f). These five features are defined as follows:

- Position of the sentence (w_p)

Let document D = $\{s_1, s_2, \ldots, s_N\}$ be a set of sentences, where si is the ith sentence. The position of a sentence in a document can indicate its importance. The beginning and the ending sentences are often thematic sentences and considered more important than those in the middle. Therefore, we assigned higher weights to sentences at the two ends of the document. The position feature (w_p) of a sentence is defined as:

$$w_p(s_i) = \frac{1}{\min(i, N - i + 1)},$$

where $1 \le i \le n$.

- Term weight (w_t)

The second feature (w_t) awards sentences that contain more important terms. The occurrence of terms in a sentence can indicate its significance in a document. We modified the definitions of term frequency and document frequency in the traditional vector space model (Salton, Wong, et al., 1975) and defined two sentence-level measures: term frequency *tf(t, s)*, i.e., the frequency of term *t* in sentence *s*, and sentence frequency *sf(t)*, i.e., frequency of sentences that contains word *t* in the document. The term frequency–inverse document frequency *(tf-idf)* weight is a weight often used in information retrieval and text mining. This weight is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. Following the idea of the *tf-idf* weighing scheme, we define a term frequency–inverse sentence frequency *(tf-isf)* weight to measure the importance of a term to a sentence. The importance increases proportionally to the number of times a word appears in the sentence but is offset by the frequency of this word's appearance in the document. The term weight of a sentence is defined as the sum of *tf-isf* scores of all sentences:

$$w_t(s_i) = \sum_{t \in S_i} tf(t, s_i) \cdot \log \frac{N}{sf(t)}.$$

- The similarity between the sentence and the headline (w_h)

The third feature (w_h) measures the similarity between a sentence and the headline. Because the headline is often a good summary of the document, a sentence that is similar to the headline h should contribute more to the document. The headline feature is calculated using a cosine similarity function:

$$w_h(s_i) = \frac{s_i \cdot h}{\|s_i\| \|h\|},$$

where s_i and h are represented as *tf-isf* vectors, $\|\cdot\|$ denotes the dimensionality of the vector.

- The occurrence of keywords in the sentence (w_k)

The fourth feature (w_k) measures the total number of keywords that occur in the sentence. The frequency of keywords in the sentence also indicates the relevance and importance of the sentence to the document. If the keywords are unknown, we can use the keyword extraction approach (Matsuo & Ishizuka, 2004) to identify keywords in the document. The algorithm uses the probability distribution of co-occurrence between a term and the frequent terms to extract keywords in a single document, based on the Word Co-Occurrence Statistical Information. This approach is especially applicable to domain-independent keyword extraction. The keyword

DOI: 10.1002/asi

feature of a sentence s_i is defined as follows:

$$w_k(s_i) = \sum_{t \in s_i} keyword(t)$$

where *keyword(t)* equals 1 if term t is a keyword of the document and equals 0 otherwise.

- The first-person mode (w_f)

The fifth feature (w_f) denotes whether the sentence is in the first-person mode. Sentences in first-person mode, indicated by pronouns such as 我/I and 我们/we, tend to have the same polarity as the entire document. The first-person feature of sentence si is defined as:

$$w_f(s_i) = \begin{cases} 1 & \text{if } s_i \text{ contains a first-person pronoun} \\ 0 & \text{otherwise} \end{cases}$$

All the five features can be normalized to the same scale of [0, 1]. Furthermore, we aggregated the five features into one importance score by the following formula:

$$w_i = \lambda_p w_p(s_i) + \lambda_t w_t(s_i) + \lambda_h w_h(s_i) + \lambda_k w_k(s_i) + \lambda_f w_f(s_i),$$

Where the five parameters λ_p, λ_t, λ_h, λ_k, and λ_f sum up to 1 and they determine the effect of each feature on the total weight for a sentence. We labeled a relatively small training dataset manually to tune the values of these five parameters. For example, we chose 20 documents randomly from news, blogs, and forums that cover various domains, such as politics, economics, and so on. We then manually labeled the important sentences in those documents and used them as a training dataset so as to tune the five parameters. These parameters can be further fine-tuned depending on the application domain after they were initialized based on this training dataset. In our experiments, λ_p, λ_t, λ_h, λ_k, and λ_f were set to 0.15, 0.1, 0.5, 0.1, 0.15 , respectively.

Having calculated the weights and polarity scores of all sentences, we can compute the weighted sum polarity of the entire document:

$$P^d = \sum_{i=1}^{n} w_i p_i,$$

where p_i denotes the polarity score of sentence s_i.

Experiments

In this study we conducted several experiments to examine our proposed approach for sentiment analysis using two test-beds of Chinese articles.

Dataset Description

Product reviews on E-commerce Websites are the most popular test-bed for sentiment analysis. In addition, social and health issues such as euthanasia and AIDS also invoke dramatically conflicting sentiments between people. Debates on such issues can be easily found on the Web and can be a good corpus for sentiment analysis as well. We collected two datasets as the test-beds for examining our approach.

TABLE 2. Description of the AmazonCN dataset.

Category	Positive	Neutral	Negative	Unknown	Total
Book	105,834	23,257	15,109	11,566	155,766
Music	177,568	9,602	7,887	53,375	248,432
Movie	20,150	4,704	4,403	13,202	42,459
Electrical Appliance	3,358	1,287	1,024	170	5,839
Digital product	2,884	1,052	929	85	4,950
Camera	596	272	188	20	1,076
Total	310,390	40,174	29,540	78,418	458,522

The first test-bed contains articles about euthanasia-related discussions collected from various Websites. The second test-bed contains products reviews of six product categories from Amazon.cn, one of the largest Chinese E-commerce Websites.

Euthanasia dataset. We extracted 851 Chinese articles on "euthanasia" from several Webpages, blogs, and online forums, including public heath Websites (e.g., http://www.39.net/nursing/anlesi), news Websites (e.g., http://news.sina.com.cn/view/als/), law-related Websites (e.g., http://www.dffy.com), information Websites (e.g., http://www.ynet.com,), etc. We manually reviewed and divided them into 502 positive articles and 349 negative articles. Because the numbers of positive and negative instances are relatively balanced in this dataset, all the articles were used in the experiments to test sentiment analysis approaches. Standard 10-fold cross-validation was chosen for evaluation.

AmazonCN dataset. In total, we collected 458,522 reviews from six categories, specifically 155,766 book reviews, 248,432 music reviews, 42,459 movie reviews, 5,839 electrical appliance reviews, 4,950 digital product reviews, and 1,076 camera reviews. Each Amazon product review usually has a rating assigned by the reviewer. We considered the rating as the sentiment polarity of the review. If the rating is above 3, the review is positive; if the rating is below 3, it is negative; if the rating equals 3 it is neutral. Some reviews have no rating and therefore their polarity is unknown. Table 2 describes the AmazonCN dataset.

The distribution of positive and negative instances (310,390 vs. 29,540) in the AmazonCN dataset is very unbalanced, which may affect the machine-learning performance. To balance the distribution of the two classes, we randomly selected up to 200 positive and 200 negative reviews from each of the six product categories as the training dataset. From the remaining reviews, up to 500 positive and 500 negative reviews were randomly selected from each category as the test set for evaluation.

Baseline Methods

In this study we compared our proposed rule-based sentiment analysis approach against learning-based approaches as baselines for sentiment classification of Chinese articles.

Learning-based approaches formulate sentiment analysis as a binary classification problem, which classifies each document into one of two categories of sentiment polarities: positive and negative (Pang et al., 2002; Turney, 2002; Wiebe & Riloff, 2005; Abbasi et al., 2008). Features are critical to such supervised learning tasks. N-gram features are usually used in traditional text classification tasks. For sentiment classification, not only lexical features but also semantic features need to be taken into consideration. In our experiments, we chose the following three types of features for sentiment polarity classification of Chinese reviews.

- Bag-of-words features: Bag-of-words features represent a document by an unordered collection of words. Unlike most Western languages like English, Chinese does not separate words by spaces in sentences. We used a popular Chinese natural language processing (NLP tool, Institute of Computing Technology, Chinese Lexical Analysis System [ICTCLAS, http://www.ictclas.org/]) for word segmentation.
- Word/POS features: POS tags are syntactic features commonly used in natural language processing. The same tool ICTCLAS was also used for POS tagging in our testbeds. In our experiments, we regard each word/POS pair as one feature, such as "有 益(wholesome)/JJ(adjective)" and "最高法院(supreme court)/NN(noun)."
- Appraisal features: Appraisal phrases are phrases that indicate a certain opinion toward a topic or an object. Specifically, we define an appraisal phrase as a triad, ⟨S, D, N⟩, in which S stands for a subjective word, D stands for a degree modifier, and N stands for a negative modifier. For example, from a sentence "我很高兴/I am very happy," an appraisal triad ⟨高兴/happy, 很/very, null⟩; is extracted. Similarly, from a sentence "我不高兴/I am not happy," an appraisal triad ⟨高兴/happy, null, 不/not⟩ is extracted. We developed a program to extract such triads from each sentence.

Once the features were defined and extracted from the documents, we used three popular supervised learning algorithms, Support Vector Machines (SVM) (Joachims, 1998, 1999), Naïve Bayes (Lewis, 1998; McCallum & Nigam, 1998), and Decision Trees (Quinlan, 1996; Mitchell, 1997), to train sentiment classification models. These algorithms were chosen because of their reported good performance in many classification applications (Masuyama & Nakagawa, 2002; Yu al., 2002; Wang, Hodges, et al., 2003; Dai, Xue, et al., 2007; Sharma & Jain, 2007). Specifically, we used the C4.5 algorithm as a representative of decision tree learning and a popular package called LIBSVM for SVM learning.

Evaluation Metrics

In our experiments, standard evaluation metrics for classification (i.e., accuracy, precision, recall, and F-measures) were used to evaluate the performance of sentiment analysis approaches (Pang et al., 2002; Turney, 2002; Ye et al., 2006).

In particular, accuracy measures the overall classification correctness.

$$\text{accuracy} = \frac{\text{\# of all correctly identified instances}}{\text{total \# of instances}}$$

Precision, recall, and F-measure indicate the correctness for each class. Specifically, precision indicates the correctness of identified relations and recall indicates the completeness of identified relations. F-measure is the harmonic mean of precision and recall.

$$\text{precision}(i) = \frac{\text{\# of correctly identified instances for class } i}{\text{total \# of instances identified as class } i},$$

$$\text{recall}(i) = \frac{\text{\# of correctly identified instances for class } i}{\text{total \# of instances in class } i}, \text{ and}$$

$$\mathrm{F}_\beta(i) = \frac{(\beta^2 + 1) \times \text{precision}(i) \times \text{recall}(i)}{\beta^2\text{precision}(i) + \text{recall}(i)},$$

where β is a parameter allowing different weighting of precision and recall. We set $\beta = 1$ to get the harmonic mean of the two measures.

Experimental Results

In order to evaluate our rule-based approach against the learning-based approaches for Chinese sentiment analysis, we designed and conducted the following three experiments.

Experiment I: Comparison of features for sentence semantic aggregation. Our proposed rule-based sentiment analysis approach requires aggregating sentiment polarity scores of all sentences based on five sentence importance features to predict the document's polarity. The overall importance of a sentence in a document is measured by a weighted sum of the five features.

Experiment I aims at examining these five features' contributions to the prediction of document sentiment polarity. First, by assuming sentences in a document are equally important, we set each $w(s_i) = 1$ and the document sentiment polarity is the mean of all sentences' polarity scores, which we refer to as the unweighted method. Furthermore, we tested the effect of each individual feature by assigning its parameter $\lambda = 1$ and others equal to 0. In addition, we used a weighted method by adjusting the five parameters λ's and calculating the aggregated polarity score for each sentence. Specifically, we randomly sampled a small subset of 20 articles from the entire corpus that covers various domains for parameter tuning purposes. We manually labeled the sentiment polarity of all the sentences in these articles. By comparing different parameter settings based on each feature's contribution, we found the optimal setting ($\lambda_p = 0.15$, $\lambda_t = 0.1$, $\lambda_h = 0.5$, $\lambda_k = 0.1$, and $\lambda_f = 0.15$) for the tuning set. We also used this parameter setting for other datasets in our other experiments. The sentiment classification accuracies on the two datasets are shown in Figure 4.

From the results of the Euthanasia dataset, we found that the unweighted sentiment aggregation method achieved the lowest accuracy, whereas taking into account the sentence importance measured by the five features can improve the accuracy of document sentiment classification. In particular, feature w_h, i.e., the similarity between the sentence and the

DOI: 10.1002/asi

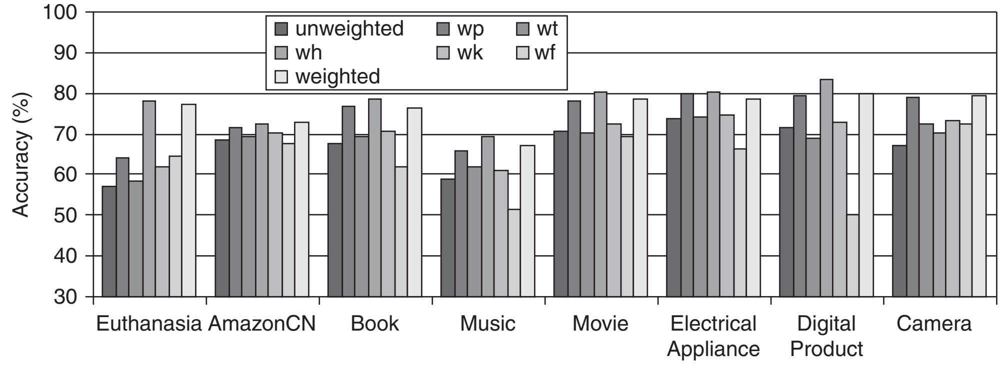

FIG. 4. Accuracies of the rule-based approach with different parameter settings.

title, was shown to be the most effective indicator of sentence importance. The weighted method achieved the second-highest prediction accuracy (77.2%), only slightly lower than w_h (78.14%). From the AmazonCN dataset we found similar results, except that the weighted method achieved the highest accuracy (75.93%). Similar results were found for each category of AmazonCN. For AmazonCN the effects of sentence importance features seemed less obvious as compared to those in the Euthanasia dataset. This is mainly due to the shorter length and fewer sentences in the review articles. Our experimental results showed that, in general, our chosen values for parameters λ's were consistent with the contributions of the five features for document sentiment aggregation on different datasets. This parameter-setting approach based on a small tuning set is expected to perform well for rule-based sentiment analysis in other domains. In the rest of our experiments we used the weighted method for document-level sentiment aggregation.

Experiment II: Feature selection for the learning-based approaches. In our experiments we chose three machine-learning approaches as baselines and three types of features were extracted. In order to raise the bar of our baseline methods, Experiment II focused on evaluating the predictive powers of the three feature types and finding the best feature subset for sentiment classification. For each of the learning algorithms, SVM, Naïve Bayes, and Decision Trees, we used seven different combinations of feature types to train a classification model. For each feature-technique pair we selected the most relevant features based on information gain (IG) and conducted standard 10-fold cross-validations to estimate the classification performance. The classification accuracies of learning-based approaches using different feature subsets are summarized in Tables 3 and 4 and Figures 5 and 6.

The results show that, for each feature subset, SVM always outperformed the other two algorithms in terms of classification accuracy. This is consistent with the outstanding performance of SVM reported in many studies (Yang & Liu; Joachims, 1998; Dumais & Chen, 2000; Pang et al., 2002). In addition, we also found that, for both SVM and Naïve Bayes classification, models based on the feature set of Word/POS tags and appraisal phrases achieved the highest accuracies. For decision trees, this feature subset also achieved one of the highest accuracies. Therefore, we used this feature subset of Word/POS tags and appraisal phrases in our later experiments.

TABLE 3. Accuracies of learning-based approaches on the euthanasia dataset.

Feature set	SVM	Naïve Bayes	Decision tree
Bag-of-words	80.24%	66.28%	71.45%
Word/POS	81.29%	65.57%	69.92%
Appraisal phrases	79.41%	68.39%	74.50%
Bag-of-words + Word/POS	80.24%	64.04%	70.27%
Bag-of-words + appraisal phrases	82.47%	67.57%	**76.50%**
Word/POS + appraisal phrases	**83.88%**	**68.86%**	75.21%
All features	81.29%	67.33%	75.09%

TABLE 4. Accuracies of learning-based approaches on the AmazonCN dataset.

Feature set	SVM	Naïve Bayes	Decision tree
Bag-of-words	79.57%	73.16%	**70.32%**
Word/POS	79.34%	71.05%	69.26%
Appraisal phrases	79.41%	73.16%	68.17%
Bag-of-words + Word/POS	79.77%	71.20%	70.27%
Bag-of-words + appraisal phrases	79.88%	72.63%	69.68%
Word/POS + appraisal phrases	**79.97%**	**73.53%**	**70.32%**
All features	79.48%	71.78%	71.18%

Experiment III: Comparison of the rule-based and learning-based approaches. Based on the results of Experiments I and II, we adjusted the settings of rule-based and the learning-based approaches to achieve their best performances. Next, Experiment III focused on comparing the performance of these two types of approaches for sentiment classification. Table 5 and Figure 7 show the accuracies of the rule-based and learning-based approaches for predicting the sentiment polarities on thetwo datasets. Numbers in bold represent the highest accuracies for the each dataset. We can see that the rule-based approach outperformed the Naïve Bayes and Decision Tree classifiers for most datasets, and it outperformed the SVM for book, music, movie, and digital product

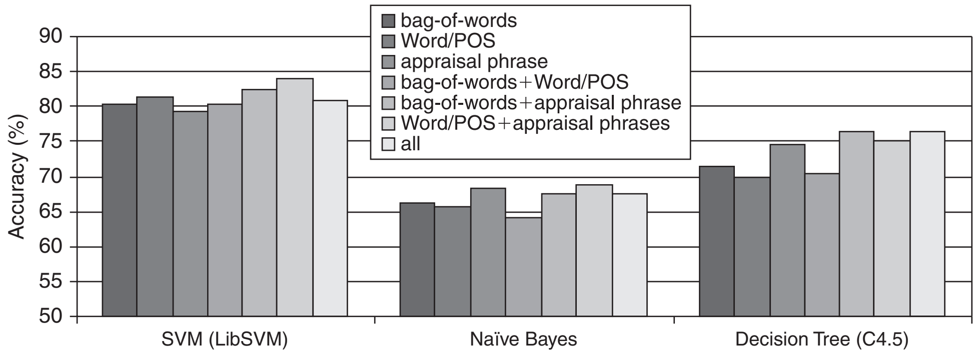

FIG. 5. Accuracies of learning-based approaches on the Euthanasia dataset.

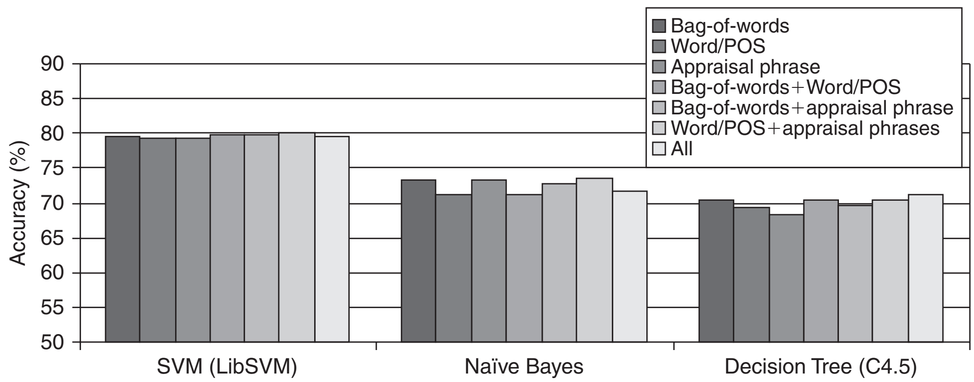

FIG. 6. Accuracies of learning-based approaches on the AmazonCN dataset.

TABLE 5. Accuracies of sentiment analyses on the two datasets.

	Learning-based			
Datasets	SVM	Naïve Bayes	Decision tree	Rule-based
Euthanasia	83.88%	68.86%	75.21%	77.20%
AmazonCN	79.78%	73.53%	70.32%	75.93%
Book	71.50%	64.30%	61.70%	76.44%
Music	63.70%	59.50%	57.20%	67.78%
Movie	74.60%	65.20%	60.40%	78.02%
Electrical appliance	78.70%	73.30%	67.80%	77.06%
Digital product	79.00%	73.30%	66.40%	78.61%
Camera	71.32%	66.08%	67.22%	79.61%
Average	75.31%	68.01%	65.78%	**76.33%**

in the AmazonCN dataset. The macro average of classification accuracy of the rule-based approach was 76.33%, higher than the three learning-based approaches: SVM (75.31%), Naïve Bayes (68.01%), and Decision Trees (65.78%). Furthermore, pairwise t-tests between these approaches show that the rule-based approach significantly outperformed both Naïve Bayes ($p < .001$) and Decision Trees ($p < .001$) but not SVM ($p = .582$).

Table 6 shows more detailed evaluation results of the rule-based sentiment analysis approach, including the precision, recall, F-measure, and accuracy for the two sentiment polarities. Due to the limited space, we only present the detailed results of learning-based approach using SVM in Table 7. The macro precision, recall, F1, and accuracy for

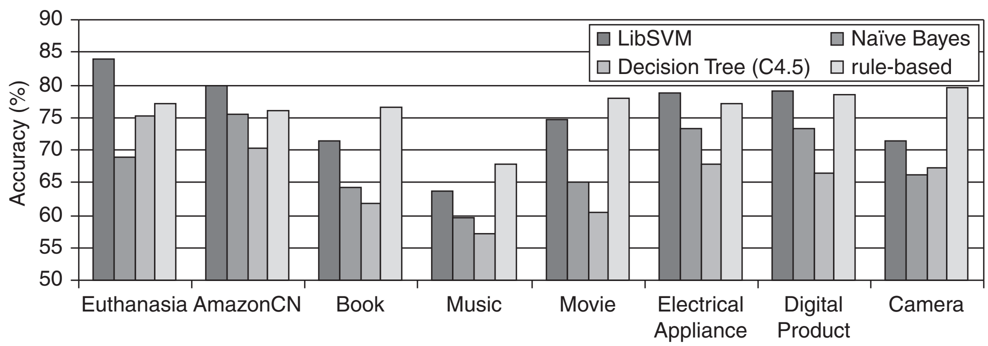

FIG. 7. Accuracies of sentiment analyses on the eight datasets.

DOI: 10.1002/asi

TABLE 6. Performance of the rule-based sentiment analysis approach.

Dataset	Polarity	Precision	Recall	F1	Accuracy
Euthanasia	Positive	76.27%	89.64%	82.42%	77.20%
	Negative	80.86%	59.31%	68.43%	
AmazonCN	Positive	71.53%	89.21%	79.40%	75.93%
	Negative	84.05%	61.55%	71.06%	
Book reviews	Positive	71.34%	90.68%	79.85%	76.44%
	Negative	86.11%	61.32%	71.63%	
Music reviews	Positive	62.80%	92.31%	74.75%	67.78%
	Negative	83.48%	41.56%	55.49%	
Movie reviews	Positive	72.88%	89.40%	80.30%	78.02%
	Negative	86.22%	66.60%	75.15%	
Electrical appliance reviews	Positive	71.04%	91.93%	80.14%	77.06%
	Negative	88.32%	61.97%	72.84%	
Digital product reviews	Positive	72.73%	92.18%	81.31%	78.61%
	Negative	89.05%	64.78%	75.00%	
Camera reviews	Positive	85.52%	86.30%	85.91%	79.61%
	Negative	63.86%	62.35%	63.10%	
Average	—	77.88%	75.07%	74.80%	76.33%

TABLE 7. Performance of the learning-based sentiment analysis approach using SVM.

Dataset	Polarity	Precision	Recall	F1	Accuracy
Euthanasia	Positive	83.73%	90.24%	86.86%	83.88%
	Negative	84.14%	74.71%	79.15%	
AmazonCN	Positive	80.63%	79.97%	80.30%	79.98%
	Negative	78.89%	79.58%	79.23%	
Book reviews	Positive	70.18%	72.00%	71.08%	70.70%
	Negative	71.25%	69.40%	70.31%	
Music reviews	Positive	61.63%	72.60%	66.67%	63.70%
	Negative	66.67%	54.80%	60.15%	
Movie reviews	Positive	74.80%	74.20%	74.50%	74.60%
	Negative	74.40%	75.00%	74.70%	
Electrical appliance reviews	Positive	76.14%	83.60%	79.69%	78.70%
	Negative	81.82%	73.80%	77.60%	
Digital product reviews	Positive	76.95%	82.80%	79.77%	79.00%
	Negative	81.39%	75.20%	78.17%	
Camera reviews	Positive	90.37%	67.74%	77.43%	71.32%
	Negative	48.56%	80.85%	60.68%	
Average	—	75.10%	75.41%	74.77%	75.24%

the rule-based approach were 77.88%, 75.07%, 74.80%, and 76.33%, while those for the learning-based approach were 75.10%, 75.41%, 74.77%, and 75.24%, respectively. The results show that the rule-based approach achieved similar performance as the learning-based approach using SVM in terms of average precision, F1, and accuracy. In particular, for book, music, movie, and camera reviews the rule-based approach outperformed SVM for sentiment analysis in terms of accuracy.

As shown in the experimental results, our proposed rule-based approach outperformed learning-based approaches, Naïve Bayes classifier, and decision trees, and achieved similar performance as compared to SVM. This is because, unlike learning-based approaches, the rule-based approach can capture the word dependencies and identify subtle differences of sentiment expressed in Chinese sentences. Furthermore, the rule-based approach considers the importance of sentences when aggregating sentence polarity scores to determine the overall document polarity, while it is ignored in the learning-based approaches. Although the SVM learning approach achieved similar or even higher performance than our approach, it requires training the classification model on a manually annotated dataset. By contrast, the rule-based approach is much more portable and adaptable to various topic domains since it does not require the manual annotation of large amounts of training data. The rule-based approach is shown to be effective for sentiment analysis in both public health and E-commerce domains. It can also be easily applied to other domains such as politics and economy with no need for model training.

From the experiment results, we found that the rule-based approach performed better on the euthanasia dataset than on the AmazonCN dataset. There might be two reasons for this. First, the product reviews in the AmazonCN dataset tend to be more informal. Therefore, some words used could not be found in the sentiment word dictionary and the dependency trees constructed are less accurate. Second, articles that discuss euthanasia are generally longer than product reviews from AmazonCN. The rule-based approach weighs the importance of sentences when aggregating their sentiment polarity scores. The five sentence importance features are more effective for longer articles that contain more sentences.

Conclusion and Future Directions

In this study we focused on predicting sentiment polarity of Chinese articles in different topic domains, including articles discussing public health issues and product reviews in E-commerce Websites. The main contribution of this study is that we proposed a rule-based semantic analysis approach, which considers the word dependency structures in Chinese sentences and the importance of sentences to predict the sentiment polarity of a Chinese article. Using a dataset of euthanasia-related articles and one of online product reviews, we conducted comparative experiments between our proposed rule-based approach and three learning-based approaches. The rule-based approach showed good performance and, importantly, it does not need manually labeled training instances, which are required for learning-based approaches.

As shown in our experiments, there is still a lot of room for performance improvement on our sentiment analysis approaches. One promising line of future research is to refine rules in the sentiment polarity prediction algorithm to further improve the performance. Besides, we can also use learning-based approaches to help create new rules that can be incorporated in the algorithm. Furthermore, the sentiment analysis approach will be applied to a larger dataset and other domains to examine their effectiveness.

DOI: 10.1002/asi

Acknowledgments

This work is supported in part by NNSFC #70890084 and #60621001; MOST 2006AA010106; CAS #2F05N01, #2F07C01, and #2F08N03; and NSF #IIS-0839990 and #IIS-0428241. We thank Hengmin Zhou and Chao Chang from the Institute of Automation for collecting and preprocessing the AmazonCN dataset.

References

Abbasi, A., Chen, H., & Salem, A. et al. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. ACM Trans. Information Systems, 26(3), 1–34.

Anonymous. HIT-IR LTP. Retrieved from http://ir.hit.edu.cn/phpwebsite/index.php?module=pagemaster&PAGE_user_op=view_page&PAGE_id=147&MMN_position=52:48.

Bansal, M., Cardie, C., & Lee, L. (2008). The power of negative thinking: Exploiting label disagreement in the min-cut classification framework. In Proceedings of the International Conference on Computational Linguistics (COLING 2008) Companion volume: Posters (pp.15–18). East Stroudsburg, PA: Association for Computational Linguistics.

Dai, W., Xue, G.-R., Yang, Q., & Yu, Y. (2007). Transferring Naive Bayes classifiers for text classification. In R. C. Holte & A. Howe (Chairs), Proceedings of the 22nd Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence (pp. 540–545). Menlo Park, CA: AAAI Press.

Dong, Z., & Dong, Q. (2003). HowNet—a hybrid language and knowledge resource. In Proceedings of International Conference on Natural Language Processing and Knowledge Engineering (pp. 820–824). Los Alamitos, CA: IEEE Press.

Dumais, S.T., & Chen, H. (2000). Hierarchical classification of Web content. In Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2000) (pp. 256–263). New York: ACM Press.

Efron, M. (2004). Cultural orientation: classifying subjective documents by cociation [sic] analysis. In Proceedings of the AAAI Fall Symposium on Style and Meaning in Language, Art, Music, and Design (pp. 41–48). Menlo Park, CA: AAAI Press.

Hatzivassiloglou, V., & McKeown, K.R. (1997). Predicting the semantic orientation of adjectives. In Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics (pp. 174–181). East Stroudsbourg, PA: Association for Computational Linguistics.

Hu, M., & Liu, B. (2006, March). Opinion feature extraction using class sequential rules. Paper presented at AAAI-CAAW-06, the Spring Symposia on Computational Approaches to Analyzing Weblogs, Stanford, CA.

Joachims, T. (1998). Text categorization with suport vector machines: learning with many relevant features. In Proceedings of the Tenth European Conference on Machine Learning (ECML). Lecture Notes in Computer Science, 1398, 137–142.

Joachims, T. (1999). Making large-scale support vector machine learning practical. Advances in kernel methods: Support vector machines (pp. 169–184). Cambridge, MA: MIT Press.

Kamps, J., Marx, M., Mokken, R., & de Rijke, M. (2004). Using WordNet to measure semantic orientation of adjectives. In Proceedings of 4th International Conference on Language Resources and Evaluation IV (LREC-04) (pp. 1115–1118). Paris: Evaluations and Language Resources Distribution Agency.

Kim, S.-M., & Hovy, E. (2004). Determining the sentiment of opinions. In Proceedings of the International Conference on Computational Linguistics (COLING 2004) (p.1367). East Stroudsburg, PA: Association for Computational Linguistics.

Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. American Political Science Review, 97(2), 311–331.

Lewis, D.D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. In Proceedings of the Tenth European Conference on Machine Learning (ECML). Lecture Notes in Computer Science, 1398, 4–15.

Li, J., Fan, Q.N., & Zhang, K. (2007). Keyword extraction based on tf/idf for Chinese news document. Wuhan University Journal of Natural Sciences, 12(5), 917–921.

Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: analyzing and comparing opinions on the Web. In Proceedings of the 14th International Conference on World Wide Web (WWW '05) (pp. 342–351). New York: ACM Press.

Mao, Y., & Lebanon, G. (2006). Isotonic conditional random fields and local sentiment flow. In Proceedings of the 2000 Neural Information Processing Systems Conference (NIPS) (pp. 961–968). Cambridge, MA: MIT Press.

Masuyama, T., & Nakagawa, H. (2002). Applying cascaded feature selection to SVM text categorization. In Proceedings of the 13th International Workshop on Database and Expert Systems Applications (DEXA '02). Lecture Notes in Computer Science, 2453, 241–245.

Matsuo, Y., & Ishizuka, M. (2004). Keyword extraction from a single document using word co-occurrence statistical information. International Journal on Artificial Intelligence Tools, 13(1), 157–169.

McCallum, A., & Nigam, K. (1998). A comparison of event models for Naive Bayes text classification.

McDonald, R., Hannan, K., Neylon, T., Wells, M., & Reynar, J. (2007). Structured models for fine-to-coarse sentiment analysis. In Proceedings of the Association for Computational Linguistics (ACL) (pp. 432–439). East Stroudsburg, PA: Association for Computational Linguistics.

Meena, A., & Prabhakar, T.V. (2007). Sentence level sentiment analysis in the presence of conjuncts using linguistic analysis. In Proceeedings of the European Conference on IR Research (ECIR). Lecture Notes in Computer Science, 4425, 573–580.

Mishne, G., & Glance, N. (2006, March). Predicting movie sales from blogger sentiment. Paper presented at the AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW 2006), Stanford, CA.

Mitchell, T. (1997). Decision tree learning. New York: McGraw-Hill.

Mullen, T., & Malouf, R. (2006, March). A preliminary investigation into sentiment analysis of informal political discourse. In Proceedings of the AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW 2006) (pp. 159–162), Menlo Park, CA: AAAI Press.

Nasukawa, T., & Yi, J. (2003). Sentiment analysis: capturing favorability using natural language processing. In Proceedings of the 2nd International Conference on Knowledge Capture (K-Cap '03) (pp. 70–77). New York: ACM Press.

Pang, B., & Lee, L. (2004). A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the Association for Computational Linguistics (ACL) (pp. 271–278). East Stroudsburg, PA: Association for Computational Linguistics.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 79–86). East Stroudsburg, PA: Association for Computational Linguistics.

Popescu, A.-M., & Etzioni, O. (2005). Extracting product features and opinions from reviews. In Proceedings of the Human Language Technology Conference/Conference on Empirical Methods in Natural Language Processing (HLT-EMNLP '05) (pp. 339–346). East Stroudsburg, PA: Association for Computational Linguistics.

Quinlan, J.R. (1996). Learning decision tree classifiers. ACM Computer Survey, 28(1), 71–72.

Riloff, E., Patwardhan, S., & Wiebe, J. (2006). Feature subsumption for opinion analysis. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 440–448). East Stroudsburg, PA: Association for Computational Linguistics.

Salton, G., Wong, A., & Yang, C.S. (1975). A vector space model for automatic indexing. Communications of the ACM, 18, 613–620.

Sharma, A., & Jain, R. (2007). Applying decision tree for automatic classification of agricultural web documents. In Proceedings of the Third Indian International Conference on Artificial Intelligence (IICAI-07) (pp. 1525–1532). London: Springer-Verlag.

Su, Q., Xu, X., Guo, H., Guo, Z., Wu, X., Zhang, X., Swen, B., & Su, Z. (2008). Hidden sentiment association in Chinese Web opinion mining. In Proceeding of the 17th International Conference on World Wide Web (WWW '08) (pp. 959–968). New York: ACM Press.

Thomas, M., Pang, B., & Lee, L. (2006). Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 327–335). East Stroudsburg, PA: Association for Computational Linguistics.

Tsou, B.K.Y., Yuen, R.W.M., Kwong, O.Y., Lai, T.B.Y., & Wong, W.L. (2005, May). Polarity classification of celebrity coverage in the Chinese press. Paper presented at the International Conference on Intelligence Analysis, Vienna, VA.

Turney, P. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In Proceedings of the Association for Computational Linguistics (ACL) (pp. 417–424). East Stroudsburg, PA: Association for Computational Linguistics.

Turney, P.D., & Littman, M.L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. ACM Transactions on Information Systems, 21(4), 315–346.

Veale, T. (2005). Analogy as functional recategorization: Abstraction with HowNet semantics. In Proceedings of the Second International Joint Conference on Natural Language Processing (IJCNLP-05). Lecture Notes in Computer Science, 3651, 326–333.

Wang, Y., Hodges, J., & Tang, B. (2003). Classification of web documents using a Naive Bayes method. In Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence (ICTAI '03) (pp. 560–564). Los Alamitos, CA: IEEE Press.

Wanxiang Che, Z.L., Hu, Y., Li, Y., Qin, B., Liu, T., & Li, S. (2008). A cascaded syntactic and semantic dependency parsing system. In Proceedings of the Twelfth Conference on Computational Natural Language Learning (CONLL-2008) (pp. 238–242). East Stroudsburg, PA: Association for Computational Linguistics.

Wiebe, J., Bruce, R., & O'Hara, T.P. (1999). Development and use of a gold-standard data set for subjectivity classifications. In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics (pp. 246–253). East Stroudsburg, PA: Association for Computational Linguistics.

Wiebe, J., & Riloff, E. (2005). Creating subjective and objective sentence classifiers from unannotated texts. In Proceedings of the Sixth International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2005). Lecture Notes in Computer Science, 3406, 486–497.

Wiebe, J., Wilson, T., Bruce, R., Bell, M., & Martin, M. (2004). Learning subjective language. Computer Linguistics, 30(3), 277–308.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the Human Language Technology Conference/Conference on Empirical Methods in Natural Language Processing (HLT-EMNLP '05). East Stroudsburg, PA: Association for Computational Linguistics. Retrieved August 28, 2009, from http://www.cs.pitt.edu/~wiebe/pubs/papers/emnlp05polarity.pdf

Yan, J., Bracewell, D.B., Ren, F., & Kuroiwa, S. (2007). Semi-automatic construction of an emotion ontology using HowNet. Artificial Intelligence and Pattern Recognition, 17–21.

Yang, Y., & Liu X. A re-examination of text categorization methods. 42–49.

Ye, Q., Shi, W., & Li, Y. (2006). Sentiment classification for movie reviews in Chinese by improved semantic oriented approach. In HICSS '06: Proceedings of the 39th Annual Hawaii International Conference on System Sciences (p. 53.2).

Yu, H., Han, J., & Chang, K. (2002). PEBL: Positive example based learning for Web page classification using SVM. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD) (pp. 239–248). New York: ACM Press.

Yu, H., & Hatzivassiloglou, V. (2003). Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. Paper presented at the Conference on Empirical Methods in Natural Language Processing (EMNLP).

Yuen, R.W.M., Chan, T.Y.W., Lai, T.B.Y., Kwong, O.Y., & T'sou, B.K.Y. (2004). Morpheme-based derivation of bipolar semantic orientation of Chinese words. In Proceedings of the 20th International Conference on Computational Linguistics (COLING '04) (p. 1008). East Stroudsburg, PA: Association for Computational Linguistics.

Zhu, H., Zheng, D., et al. (2008). Research on query translation disambiguation for CLIR based on HowNet. In Proceedings of the Ninth International Conference for Young Computer Scientists (ICYCS) (1677–1682). Los Alamitos, CA: IEEE Press.

DOI: 10.1002/asi