Learning Chinese Polarity Lexicons by Integration of Graph Models and Morphological Features

Bin Lu[1], Yan Song[1], Xing Zhang[1], and Benjamin K. Tsou[1,2]

[1] Department of Chinese, Translation & Linguistics, City University of Hong Kong
Tat Chee Avenue, Kowloon, Hong Kong
[2] Research Centre on Linguistics and Language Information Sciences,
Hong Kong Institute of Education, Tai Po, New Territories, Hong Kong
lubin2010@gmail.com, {yansong,zxing2}@student.cityu.edu.hk,
btsou@ied.edu.hk

Abstract. This paper presents a novel way to learn Chinese polarity lexicons by using both external relations and internal formation of Chinese words, i.e. by integrating two kinds of different but complementary models: graph models and morphological feature-based models. The polarity detection is first treated as a semi-supervised learning in a graph, and then machine learning is used based on morphological features of Chinese words. The results show that the the integration of morphological feature-based models and graph models significantly outperforms the baselines.

Keywords: Polarity Lexicon Induction, Graph Models, Chinese Morphology.

1 Introduction

In recent years, sentiment analysis, which mines opinions from large-scale subjective information available on the Web such as news, blogs, reviews and tweets, has attracted much attention [7, 15]. It can be used for a wide variety of applications, such as opinion retrieval, product recommendation, political polling and so on.

In such applications, polarity lexicons consisting of either positive or negative words/phrases are usually important resources for practical systems. They can be constructed by different approaches, including manual construction; using lexical resources such as WordNet to induce positive/negative words [2]; or learning sentiment-bearing words from large-scale corpora, such as news corpora [2] or even the Web [15].

Graph models have been recently used in sentiment analysis for various tasks, such as polarity lexicon induction, ranking the word senses by polarity properties, or document-level sentiment analysis. However, most work has been done based on either on WordNet or English documents. Although these methods can be applied on Chinese, Chinese has its own special characteristics, i.e. Chinese words are composed of characters or morphemes, the smallest meaning blocks. In Chinese, each morpheme has its own meaning, and the polarity of a Chinese word is influenced or even determined by the polarities of its component morphemes. Ku et al. [5, 6] proposed the

P.-J. Cheng et al. (Eds.): AIRS 2010, LNCS 6458, pp. 466–477, 2010.

character-based methods to use sentiment scores of Chinese characters to compute the sentiment scores of Chinese words.

However, either of these two kinds of models, namely graph models and character-based models, is not enough to well tackle the problem by themselves alone. On the other hand, these two kinds of models are complementary to each other, since they models respectively the external relations and internal structures of one Chinese word: word graphs encode the external relations of one word with others in either lexical resources or real texts, while morphological features denote the internal formation or structure of individual Chinese words.

We first build word graphs based on lexical sources, and then induce more positive/negative words from seed words by using semi-supervised graph models. We also propose to use machine learning for polarity classification of Chinese words based on morphological features. Then, we integrate the external relations and internal structures of Chinese words (i.e. graph models and morphological feature-based models) under different strategies. The experiments show that our integrated approach achieves significantly better performance than the baselines.

The rest of the paper is organized as follows. Sec. 2 introduced related work. In Sec. 3, we describe our method for learning Chinese polarity lexicon. Sec. 4 gives the experiments, followed by the discussion in Sec. 5. Finally, we conclude in Sec. 6.

2 Related Work

Some related works have tackled the automated determination of term polarity based on either corpora or lexical resources such as WordNet. Hatzivassiloglou and McKeown [2] learned polarity of adjectives by exploiting co-occurence of conjoined adjectives. Turney and Littman [14] used two statistical methods, namely PMI-IR and LSA, to calculate the polarity of individual terms by calculating mutual information between words and seed words via search engines or corpora. Kamps and Marx [4] proposed the WordNet-based method to compute the word polarity by calculating the semantic distance between words and seed words *good* and *bad*.

2.1 Graph Models for Polarity Lexicon Induction

Recently, graph models have also been tried on polarity lexicon induction. Esuli and Sebastiani [2] presents an application of PageRank to rank WordNet synsets in terms of how strongly they possess a given semantic property, e.g. positivity and negativity. Rao and Ravichandran [8] treated polarity detection as a semi-supervised Label Propagation problem in a graph. Their results indicate that Label Propagation improves significantly over the baseline and other semi-supervised learning methods like Mincuts and Randomized Mincuts for this task.

Velikovich et al. [15] described a new graph propagation framework by constructing large polarity lexicons from lexical graphs built from the web, and they built an English lexicon that is significantly larger than those previously studied. They evaluated the lexicon, both qualitatively and quantitatively, and show that it provides superior performance to previously studied lexicons.

2.2 Chinese Polarity Lexicon Induction

For Chinese, Yuen et al. [16] proposed a method, based on [14], to compute the polarity of Chinese words by using their co-occurrence with Chinese morphemes. It was noted that morphemes are much less numerous than words, and that also a small number of fundamental morphemes may be used to get great advantage. Zhu et al. [18] tried to compute the polarity of Chinese words using the semantic distance or similarity between words and seeds in HowNet[1] based on [14].

Ku et al. [5] measured sentiment degrees of Chinese words by averaging the sentiment scores of the composing characters, called the bag-of-character (BOC) method. The sentiment score of each character is calculated by using the observation probabilities of the character in positive and negative seed words. Ku et al. [6] further considered the internal morphological structures of Chinese words for opinion analysis on words. Chinese words were classified into eight morphological types by the proposed classifiers, and then heuristic scoring rules were manually defined for each morphological type based on the character scores obtained by the BOC method.

2.3 Analysis of the Two Kinds of Models

Graph models and the morpheme-based or character-based models provide different perspectives of Chinese words, and have different characteristics. Word graph encode the external relations of one word with others, while morphological features represent the internal formation or structures of Chinese words. Graph models would need external resources, such as thesauri, lexical resources, or large corpora to construct word graphs, while the character-based methods [5, 6] can assign an opinion score to an arbitrary word without any thesauri or large corpora. However, the character-based methods could have the following problems:

(1) The polarities of many Chinese words cannot directly be derived from its component characters, such as 泡汤 (fail), 仓皇 (in panic), 蓄意 (malicious), etc. For example, 泡汤 (fail) is composed of 泡 (soak) and 汤 (soup), and none of these two characters have salient polarity, but the whole word is negative;

(2) A character may have many possible senses with different polarities, but the character-based methods only compute one polarity score for each character. For instance, the character 动 has many senses in HowNet: a) SelfMoveInManner|方式性自移 or alter|改变, e.g. the 动 in 动荡 (turmoil) and 动乱 (unrest) is negative; b) excite|感动, e.g. the 动 in 动人 (making you feel emotional or sympathetic) and 动听 (pleasant to the ears) is positive; c) use|利用, e.g. the 动 in 动用 (utilize) is neutral;

(3) To cover most Chinese characters, the character-based methods will need a large amount of training data, i.e. Chinese words annotated with polarities.

The problem of graph models could be the need of large-scale lexical resources or corpora to construct the word graphs and to achieve good performance, and even with such large-scale resources it sometimes cannot cover the words concerning. But they could more easily adapt to different domains and compute domain-dependent polarity score based on different corpora. Meanwhile, once the word graphs are constructed, graph models can do semi-supervised learning with a small number of seed words.

[1] http://www.keenage.com/html/e_index.html

From the above analysis, we can know that the two kinds of model have different advantages. Thus it would be very attractive to integrate graph models and the morpheme-based or character-based methods to get better performance.

3 Learning Chinese Polarity Lexicons with Graph Models and Morphological Features

3.1 Graph Models for Polarity Lexicon Induction

Let $(x_1, y_1) \dots (x_l, y_l)$ be labeled words or phrase, where $Y_L = \{y_1 \dots y_l\}$ are the polarity labels, i.e. positive, negative or neutral. Let $(x_{l+1}, y_{l+1}) \dots (x_{l+u}, y_{l+u})$ be unlabeled words where $Y_U = \{ y_{l+1} \dots y_{l+u} \}$ are unobserved, usually $l<<u$. A graph is built where the nodes are all words, both labeled and unlabeled, and the edge between nodes i, j is weighted so that the closer the nodes are, the larger the weights w_{ij}. Intuitively, words that are close should have similar labels, and thus the labels of a node could be propagated to all nodes through the edges. We assume as input an undirected edge weighted graph $G = (V, E)$, where $w_{ij} \in [0, 1]$ is the weight of edge $(v_i, v_j) \in E$. We also assume as input two sets of seed words, denoted P for the positive seed set and N for the negative seed set. After constructing the graph, we can use Label Propagation or PageRank to derive Y_U from X and Y_L based on the graph.

Label Propagation (LP) is an iterative algorithm for classification or regression [17], where each node takes on the weighted average of its neighbor's values from the previous iteration. The result is that nodes with many paths to seeds get high polarities due to the influence from their neighbors. LP is known to have many desirable properties including convergence, a well defined objective function (minimize squared error between values of adjacent nodes), and an equivalence to computing random walks through graphs. We use LP taking a form slightly different from the algorithms in [8] and [15] by adding another sets of seed words, denoted T for the neutral seed sets. The neutral words are manually chosen from the top 200 most frequent words in the LIVAC[2] corpus, and many of them are monosyllabic, include 的 (of), 在 (at), 一 (one), 与 (and), 他 (he), etc. The neutral words are used as *stop* words to prevent polarity propagate into them and also to prevent flow from passing through them into other related words.

PageRank is also a random walk model [1], but used for ranking problem. It allows the random walk to "jump" to its initial state with a nonzero probability. PageRank can be used to get two ranked lists respectively for positive and negative words, and we can normalize the scores for each ranked list, and then use the score in the positive ranking minus that in the negative ranking to get the final score for each word. If the final score is positive, then the word can be classified as positive; otherwise, negative. By this means, we are actually using PageRank for classification. Although Label Propagation and PageRank originally were proposed for different tasks, namely classification and ranking, respectively, they actually are closely related and have theoretical connection [8].

[2] http://www.livac.org

3.1.1 Building Word Graphs

The word graph could also be built by different means from a wide variety of resources, such as news corpora [2], WordNet [8], and the Web documents [15]. In this paper, we use Tongyici Cilin (Cilin) [12] and a combined bilingual lexicon to construct word graphs. All the entries in Cilin are organized in a hierarchical tree, and the vocabulary is divided into different categories, i.e. 12 large categories and 1, 400 subcategories. There are some synonym groups within each subcategory, and the words in the same group either have the same or similar meaning or have high relevance. The total number of synonym groups in Tongyici Cilin is 13,440. Following are two synonym group examples:

Ed03A01={好, 优, 精, 良, 帅, 妙, 良好, 优秀, 优异, 精彩, …}

{good, excellent, superior, fine, handsome, brilliant, all right, excellent, outstanding, wonderful, …}

Ed03B01={坏, 差, 次, 软, 浅, 破, 不好, 不良, 不行, 差劲, …}

{bad, bad, inferior, weak, shallow, rubbishy, not good, not fine, poor, bad, …}

Another lexical resource we used for word graph construction is a combined bilingual lexicon. The idea behind is that a word in one language could be translated into different words in another language. For example, *beautiful* could be translated into 漂亮, 优美, or 美丽 in Chinese, while *ugly* could be translated into 丑, 丑陋 or 难看. In such cases, the different translations of the same word could be seen as synonyms. We combine three bilingual lexicons as the final bilingual lexicon: namely, LDC_CE_DIC2.0[3] constructed by LDC, bilingual terms in HowNet and the bilingual lexicon in Champollion[4]. In total, there are about 251K bilingual entries in the combined dictionary. By using the English words as the pivot, we get 45,448 synonym groups.

In our constructed word graphs, the nodes are all words, the edge between nodes indicates a synonym relation and each edge has a weight w of 1 initially. We assume there are n nodes in the graph which could be represented as a $n \times n$ transition matrix T derived by row-normalizing edge weights. Then, Labe Propagation and PageRank can be run on the constructed matrix.

3.2 Polarity Lexicon Induction with Morphological Features

As words are the basic building blocks for texts, most researches on sentiment analysis in English have been done based on words. However, when it comes to Chinese, the situation is rather different. The majority of Chinese words in a corpus are disyllabic or polysyllabic, where each syllable is normally represented by a single logograph, or usually a morpheme. The meaning of the predominant polysyllabic words can be seen as derived from the meanings of its component morphemes, which are considered to be the smallest meaningful linguistic unit[5].

---

[3] http://projects.ldc.upenn.edu/Chinese/LDC_ch.htm

[4] http://sourceforge.net/projects/champollion/

[5] For simplicity, we consider morphemes to be monosyllabic and represented by single character in the following discussion.

In the BOC method [5], the opinion score of a word is determined by the combination of the observation probabilities of its composite characters in positive and negative words. In our implementation of the BOC method, we make a small modification by considering negation markers, such as 无 (*no*), 不 (*no*), 没 (*no*). For the calculation of the sentiment score of a character *c*, if a negation marker *neg* occurs before some other characters, the characters following the *neg* would be considered as occurring in a word with an opposite polarity. For instance, when computing the frequency of the character 好 (good) as a positive or negative character, our modified method would consider the 好 in the negative word 不好 (not good) as a positive occurrence because of the negation marker 不 before the character好; while in the original BOC method, it would be considered it as a negative occurrence because it occurs in a negative word. Negation markers are processed with the similar method for calculating sentiment scores of test words by the scores of its component characters.

One problem of the BOC method is that it only assigns a sentiment value for each character without considering character contexts, and cannot easily integrate other possibly useful features, such as bigrams of characters, POS, position information of characters in the word, into the model. Therefore, we propose to learn word polarity using machine learning by integrating more morphological features in addition to its component characters as basic features. The polarity lexicon induction is considered a classification problem and we use machine learning to solve it by using morphological features in Chinese words. The feature templates for the classification model are shown in Table 1.

The POS of each Chinese word is obtained by querying the POS of that word with most senses in HowNet. The features mined for each Chinese word with templates are converted into a vector in which each dimension has the weight of 1.

Table 1. Features used in classification models

Description	Feature Templates	Example Features for 美丽 (beautiful)
Character Unigrams	$\{c_i\}$, $1 \le i \le l$	{美, 丽}
Character Bigrams	$\{c_{i-1}c_i\}$, $2 \le i \le l$	{美丽}
Word POS	*POS*	ADJ
Character Unigrams with Position	$\{i\_c_i\}$, $1 \le i \le l$	{1_美, 2_丽}

3.3 Integrating Graph Models and Morphological Features

Since graph models and morphological features provide two individual and independent perspectives (i.e. external and internal) of Chinese words, we propose to integrate them to achieve better performance. After obtaining different classifiers based on graph models and morphological features, we could exploit different ensemble methods to combine the results of individual classifiers. According to theoretical analysis [8], the effectiveness of ensemble learning is determined by the diversity of its component classifiers, i.e. each classifier need to be as unique as possible, particularly with respect to misclassified instances. The different classifiers built based on graph models and morphological features could satisfy this diversity requirement.

Let $F=\{f_k(x)|1\leq k\leq p\}$ be the polarity values given by classifiers, where p is the number of classifiers, and $f_k(x) \in [-1, 1]$. We exploit the following ensemble methods for deriving a new value from the individual values:

(1) Average: It is the most intuitive combination method and the new value is the average of the values in F.

(2) Weighted Average: This combination method improves the average method by associating each individual value with a weight, indicating the relative confidence in the value. The weights are experimented to be set in the following two ways:

F1-Weighted Average: The weight of $f_k(x)$ is set to the Micro-F1 of the individual classifier obtained on the development data.

Pre-Weighted Average: The weight of $f_k(x)$ is set to the Micro-Precision of the individual classifier obtained on the development data.

(3) Majority Voting: This combination method relies on the final polarity tags given by each classifier, instead of the exact polarity values. A word can obtain p polarity tags based on the individual analysis results in the p classifiers. The polarity tag receiving more votes is chosen as the final polarity tag of the Chinese word.

(4) SVM Meta-classifier: Motivated by the supervised hierarchical learning, we also propose to use SVM to automatically adjust the weights for each component classifier. It is similar to the re-ranking process with two-layer models: the output values given by the individual low-level classifiers are fed into a machine learning framework (namely SVM) as features, and thus a weight model for individual classifiers is learned from the training and development data. By this strategy, we actually use a two-level classification model with a higher-level meta-classifier to learn the corresponding weights for the individual lower-level classifiers.

4 Experiments and Evaluation

Two manually constructed polarity lexicons are used as gold standard for evaluation: *The Lexicon of Chinese Positive Words* [13] consisting of 5,045 positive words, and *The Lexicon of Chinese Negative Words* [19] consisting of 3,498 negative words. Thus, we have 8,543 words marked with polarity as the gold standard. The entries in the gold standard are randomly split into 6 folds: the first fold as the development set, and the remaining ones for 5-fold cross validation (4 folds for training and 1 fold for testing). The bag-of-character method [5] and Label Propagation [8] are used as baselines.

We use the standard precision (Pre), recall (Rec), and F-measure (F1) to measure the performance of positive and negative class, respectively, and used the MacroF1 and MicroF1 to measure the overall performance. The metrics are defined the same as in general text categorization. The SVM_{light} package is used for training and testing, with all parameters set to their default values. The evaluations are shown as follows.

4.1 Experiments with Graph Models

In this section, we evaluate the performance of PageRank and Label Propagation (LP) on the word graphs built from two resources, namely Tongyici Cilin (Cilin) and the bilingual lexicon (BiLex) introduced in Sec. 3.2.1, and the graph built from their combination (Cilin+BiLex). The residual probability of PageRank is set to 0.85. Since we do not have annotated ranking data of Chinese polarity lexicon to evaluate PageRank, we use the converted classification results introduced in Sec. 3.1 for the evaluation. We run both algorithms for 10 iterations, and show the results in Table 2.

Table 2. Results of graph models (in %)

		Positive			Negative			Total	
		Pre	Rec	F1	Pre	Rec	F1	MacroF1	MicroF1
Cilin	PageRank	92.83	60.37	73.15	92.89	59.89	72.79	72.97	73.24
	LP	93.22	60.47	73.35	93.17	60.23	73.13	73.24	72.99
BiLex	PageRank	84.10	40.73	54.89	93.19	32.57	48.24	51.56	52.30
	LP	84.48	40.62	54.86	92.90	32.91	48.58	51.72	52.40
Cilin+ BiLex	PageRank	89.65	67.86	77.24	95.21	62.66	75.54	76.39	76.55
	LP	89.93	67.60	77.17	94.75	63.04	75.67	76.42	76.56

From Table 2, we can observe that: the graph models show better performance on the word graph built from the combination of Cilin and BiLex than on the graphs built from either of the resources; and the graph models show better performance on Cilin than on BiLex. Meanwhile, PageRank and Label Propagation show similar performance, and the differences are not so remarkable for the word graphs built from Cilin, BiLex or their combination.

4.2 Experiments with Models of Morphological Features

In this section, we investigate the performances of different models with morphological features, including the BOC (Ku) method [5], our modified BOC method with negation processing introduced in Sec. 3.2, and our proposed SVM models with different kinds of features introduced in Sec. 3.2. The SVM-All method uses all the features introduced in Table 1. The results are shown in Table 3.

The SVM models outperform the BOC models. Although the improvement of about 1% from BOC to SVM-ALL seems not remarkable, the t-test shows that the MacroF1 and MicroF1 differences are statistically significant at the 90% level, and the difference between BOC and SVM-Uni+Bi, or between SVM-All and BOC is statistically significant at the 95% level. By adding bigrams of characters, unigrams with position, and POS into SVM, we can improve by about 0.5% compared SVM with only unigrams. Meanwhile, our modified BOC method achieves slightly better result than the original BOC method [5].

We also tried to integrate the features of morphological types in [6] into our morphological feature-based SVM model. Since we do not have the words annotated with morphological types, we just use the unsupervised heuristic rules to compute the morphological type for each Chinese word, and then use the heuristic rules in [6] or integrate it into the SVM model. But it did not improve the performance.

Table 3. Results of models with morphological features (in %)

	Positive			Negative			Total	
	Pre	Rec	F1	Pre	Rec	F1	MacroF1	MicroF1
BOC (Ku)	92.62	89.35	90.95	88.51	83.56	85.96	88.45	88.92
BOC	92.93	89.54	91.20	88.85	83.68	86.18	88.69	89.16
SVM-Unigram	88.13	95.11	91.48	92.01	81.54	86.44	88.96	89.54
SVM-Unigram +Bigram	88.39	95.30	91.70	92.34	81.95	86.82	89.26	89.83
SVM-All	88.69	95.25	91.85	92.32	82.49	87.12	89.48	90.02

4.3 Experiments on Integration

In this section, we investigate the performance of the integration of graph models and morphological features. Different classifier combinations are tried based on the SVM strategy introduced in Sec. 3.3. Since the graph built from the combination of two lexical resources show better performance than the graph built from the individual resources, we use only the combination graph for the graph models in this section. The development data are used to adjust the parameters of each model. The results are shown in Table 4. The LPBOC method denotes the BOC method based on the positive and negative word lists generated by the LP model.

Table 4. Integration results (in %)

	Positive			Negative			Total	
	Pre	Rec	F1	Pre	Rec	F1	MacroF1	MicroF1
BOC+SVM-ALL	92.32	92.90	92.60	89.65	88.88	89.25	90.93	91.25
LP+BOC	93.39	95.94	94.64	93.87	90.20	91.99	93.32	93.58
LP+SVM-ALL	91.49	94.52	92.98	92.16	88.02	90.03	91.50	91.76
LP+BOC+ SVM-ALL	94.43	96.02	95.22	94.07	91.86	92.94	94.08	94.30
LP+ SVM-ALL +BOC+LPBOC	95.17	95.99	95.58	94.11	92.97	93.53	94.56	94.75

All of the ensembles in Table 4 significantly outperform the baselines: the bag-of-character method in Table 3, and Label Propagation in Table 2, which shows that graph models and models with morphological features have their own evidences for polarity classification, and thus the integration of models could significantly improve performance.

The best performance is obtained by the integration of all the four methods: LP, ML-ALL, BOC and LPBOC, i.e. it improves MacroF1 to 94.56% from 88.69% of BOC or 76.42% of LP, and improves MicroF1 to 94.75% from 89.16% of BOC and 76.56% of LP, which are both significant improvements. Even without graph models, we can also significantly improve performance by integrating BOC and SVM-ALL. The integration of LP with BOC shows better performance than that of LP with SVM-ALL, but the integration of these three methods outperforms the integration of any two methods.

We then investigate the other ensemble methods introduced in Sec. 3.3 to integrate LP, BOC, SVM-ALL, and LPBOC. Table 5 gives the comparison results. We can see that all the ensemble methods outperform the constituent individual method, while SVM performs the best, followed by the precision-weighted average. The results further demonstrate that 1) the good effectiveness of the ensemble combination of individual analysis results for Chinese word polarity classification, 2) the SVM strategy seems to be able to find better weights compared with other simpler combination methods.

Table 5. Ensemble results for "**LP+BOC+SVM-ALL+LPBOC**" (in %)

	Positive			Negative			Total	
	Pre	Rec	F1	Pre	Rec	F1	MacroF1	MicroF1
Average	95.91	95.91	95.91	91.04	91.04	91.04	93.48	93.91
F1-Weighted Average	93.87	95.87	94.86	93.84	91.01	92.40	93.63	93.87
Pre-Weighted Average	93.99	95.99	94.97	94.01	91.19	92.58	93.77	94.01
Majority Voting	94.43	93.64	94.03	95.10	86.03	90.34	92.18	92.56
SVM	95.17	95.99	95.58	94.11	92.97	93.53	94.56	94.75

5 Discussion

In this section, we investigate the influence of two factors on the models for Chinese polarity lexicon induction, i.e. the iteration number for graph models and the size of training data. The micro-precision, micro-recall and micro-F1 are reported in this section. Fig. 1 shows the influence of iteration numbers of Label Propagation on the combined graph built from Tongyici Cilin (Cilin) and the combination of Cilin and BiLex (Com). We can observe that the precisions of Label Propagation for Cilin and Com show little difference, both above 90%, but the recalls with Com are much higher with those with Cilin, and thus the F1s are much higher with Com consequently.

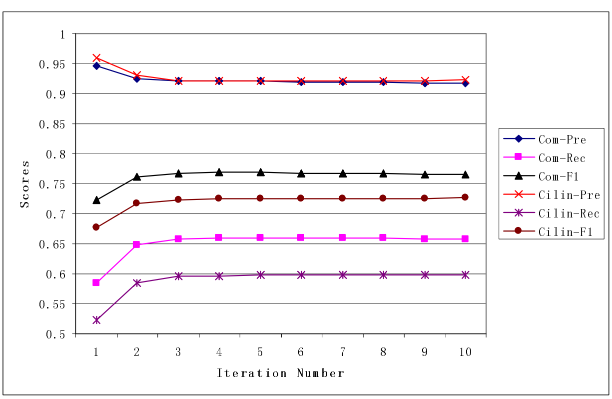

Fig. 1. Influence of iteration numbers on Label Propagation

Fig. 2 and 3 show the influence of the sizes of training data (i.e. the number of training words) on Label Propagation (LP), the BOC method and the SVM Meta-classifier based integration of LP+BOC+SVM-ALL+LPBOC (Integration) introduced in Sec. 4.3. From Fig. 2, we can see that (1) The precision of BOC improves steadily with more training data, from 75% of 100 training words to above 90% of 5K+ words; while the precision of LP remains quite high (i.e. always above 86%), even with only 100 seed words; (2) The recall of BOC improves even much faster the precision when the training data increases, from 23% of 100 training words to above 87% of 5K+ words; while the recall of LP improves much slowly. From Fig. 3, we can observe that the SVM-based integration of the four methods has been always significantly outperforming the individual methods of BOC and LP, and even with only 100 seed/training data, we can achieve 82% MicroF1.

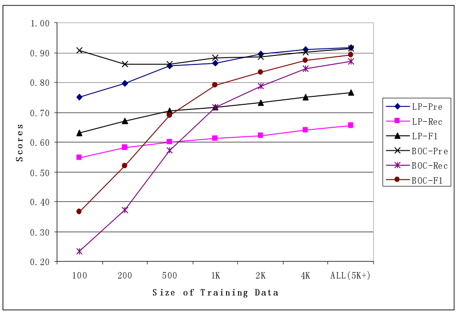

Fig. 2. Influence of training data sizes on LP and BOC

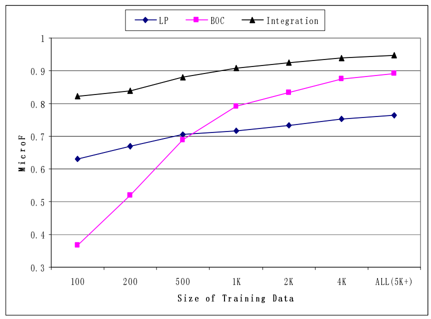

Fig. 3. Influence of training data sizes

To summarize Fig. 2 and 3, when the training data is small, LP outperforms BOC, but when the training data becomes large, BOC outperforms LP inversely. However, no matter how much training data, the integration of graph models and morphological features could significantly improve the performance compared to individual methods.

6 Conclusion and Future Work

This paper proposes a novel approach to integrate both internal structures and external relations of Chinese words for polarity lexicon induction via graph models and morphological features. The polarity detection is first treated as a semi-supervised learning in a graph, machine learning, namely Support Vector Machine (SVM), is used based on morphological features, and then we integrate morphological features with the graph models to further improve the performance. The results show that the integration could significantly improve the performance.

In future work, more resources could be explored to further improve the results, especially large-scale corpora or even the Web. Since a word could have different senses with different polarities, we are also interested in classifying the polarity of word senses in Chinese, instead of only the word level. Meanwhile, evaluation of the ranking problem of word polarities could be another direction.

Acknowledgements. We wish to thank Prof. Jingbo Zhu from Northeastern University, China, Dr. Preslav Nakov from National University of Singapore, and anonymous reviewers for their valuable comments, as well as the Harbin Institute of Technology's IR Lab for their sharing of the extended version of Tongyici Cilin.

References

1. Brin, S., Page, L.: The Anatomy of a Large-scale Hypertextual Web Search Engine. Computer Networks and ISDN Systems 30(1-7), 107–117 (1998)
2. Esuli, A., Sebastiani, F.: PageRanking WordNet SynSet: An application to opinion mining. In: Proceedings of ACL, pp. 424–431 (2007)
3. Hatzivassiloglou, V., McKeown, K.: Predicting the Semantic Orientation of Adjectives. In: Proceedings of ACL1997. pp. 174–181 (1997)
4. Kamps, J., Marx, M.: Words with Attitude. In: Proceedings of the First International Conference on Global WordNet, pp. 332–341 (2002)
5. Ku, L.W., Chen, H.H.: Mining Opinions from the Web: Beyond Relevance Retrieval. Journal of American Society for Information Science and Technology, Special Issue on Mining Web Resources for Enhancing Information Retrieval 58(12), 1838–1850 (2007)
6. Ku, L.W., Huang, T.H., Chen, H.H.: Using Morphological and Syntactic Structures for Chinese Opinion Analysis. In: Proceedings of EMNLP, pp. 1260–1269 (2009)
7. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of EMNLP, pp. 79–86 (2002)
8. Polikar, R.: Ensemble Based Systems in Decision Making. IEEE Circuits and Systems Magazine 6(3), 21–45 (2006)
9. Rao, D., Ravichandran, D.: Semi-supervised Polarity Lexicon Induction. In: Proceedings of EACL, pp. 675–682 (2009)
10. Rao, D., Yarowsky, D.: Ranking and Semi-supervised Classification on Large Scale Graphs Using Map-Reduce. In: Proceedings of Textgraphs-4, pp. 58–69 (2009)
11. Riloff, E., Wiebe, J.: Learning Extraction Patterns for Subjective Expressions. In: Proceedings of EMNLP, pp. 105–112 (2003)
12. Mei, J., Zhu, Y., Gao, Y., Yin, H.: Tongyici Cilin(2nd version). Shanghai CiShu Press (1996) (in Chinese)
13. Shi, J., Zhu, Y.: The Lexicon of Chinese Positive Words. Sichuan Lexicon Press (2006) (in Chinese)
14. Turney, P.D., Littman, M.L.: Measuring Praise and Criticism: Inference of semantic orientation from association. ACM Trans. On Information Systems 21(4), 315–346 (2003)
15. Velikovich, L., Blair-Goldensohn, S., Hannan, K., McDonald, R.: The Viability of Web-derived Polarity Lexicons. In: Proceedings of NAACL, pp. 777–785 (2010)
16. Yuen, R.W.M., Chan, T.Y.W., Lai, T.B.Y., Kwong, O.Y., Tsou, B.K.Y.: Morpheme-based Derivation of Bipolar Semantic Orientation of Chinese Words. In: Proceedings of COLING, pp. 1008–1014 (2004)
17. Zhu, X., Ghahramani, Z.: Learning from labeled and unlabeled data with Label Propagation. Technical Report CMU-CALD-02-107, CarnegieMellon University (2002)
18. Zhu, Y., Min, J., Zhou, Y., Huang, X., Wu, L.: Semantic Orientation Computing Based on HowNet. Journal of Chinese Information Processing 20(1), 14–20 (2006) (in Chinese)
19. Zhu, L., Zhu, Y.: The Lexicon of Chinese Negative Words. Sichuan Lexicon Press (2006) (in Chinese)