International Journal of Hybrid Information Technology
Vol.8, No.1 (2015), pp.237-248
http://dx.doi.org/10.14257/ijhit.2015.8.1.21

Research on Domain-independent Opinion Target Extraction

Sun Yongmei[1,*] and Huo Hua[1, 2]

[1]*Information Engineering College, Henan University of Science and Technology, Luoyang, Henan, 471023, China*
[2]*School of Software, Henan University of Science and Technology, Luoyang, Henan, 471023, China*
[*]*Corresponding author: sym_ly @163.com*

Abstract

Opinion Target Extraction is one of the important tasks for text sentiment analysis, which has attracted much attention from many researchers. For this task, we proposed an M-Score algorithm utilized in the model which realized the domain-independent opinion target extraction function. This algorithm is derived from the Pointwise Mutual Information algorithm, but the difference is that it doesn't need any manual seeds collection or any web searching engines, which reduces the manual participation and easy to be transplanted. This model starts with document preprocessing, effective opinion sentences extraction and candidate opinion target extraction by employing Conditional Random Fields Model with feature templates. Next, the M-Score algorithm is employed to extract seed set, and the bootstrapping approach is invoked to process the candidate opinion targets. Finally, the model uses word frequency and the Noun pruning algorithm to filter the opinion targets, and then obtains the final opinion targets for output. The experimental results show that the M-score method performs better than Pointwise Mutual Information algorithm in precision and recall.

Keywords: *text sentiment analysis; opinion target extraction; Conditional Random Fields; domain-independent*

1. Introduction

With the rapid development of the Internet, the network has become one of the important parts of people's work, study and entertainment activities. Especially the web2.0 technology thrives and popularized, the Internet users can create internet information instead of obtain information from the traditional web only, which realized the two-way communication of internet users and the web, so it provides a new platform for people to obtain information, post opinions and exchange feelings. Thus, there are more and more subjective texts posted by the users on the Internet, these texts could be users' comments on any products, or people's opinions on news topics. The information of these texts contains people's various emotions and emotional tendencies, such as love, hate, criticism, praise and so on. With the increasing number of internet users, the text information flooded and it have significant effect on people's life and their attitudes to the society, so extract and analyze the subjective texts information is very important and necessary. For example, the product manufactures want to know the opinions of customers about their products, the potential customers want to know the general opinions of previous customers about some product features to help them making better decisions, and the rulers of the country want to know the public's attitudes and ideas towards some social issues. How to use computers to obtain and analyze the information of these texts has become a hot topic for researchers, and text sentiment analysis technology arises under this background.

The text sentiment analysis, as a new research field, has highly research values and practical values, thus it has been attracting much attention from many scholars [1-8].

ISSN: 1738-9968 IJHIT

According to the difference of tasks, the text sentiment analysis can be roughly divided into three tasks, namely "sentiment information extraction", "sentiment classification" and "sentiment information retrieval and induction". Sentiment information extraction, as one of the important tasks of sentiment analysis, includes opinion target extraction, opinion words extraction and polarity discrimination and opinion holder extraction.

In this paper, we focus on opinion target extraction of Chinese consumer reviews about electronic products. Opinion target extraction, as one of the important tasks of information extraction, is critical to sentiment analysis, and its extracting accuracy directly affects the accuracy of text sentiment analysis. The opinion target is the object that the opinion words modified or the people talk about. For example, in the laptop computer product reviews, the opinion target can be one kind of products (ThinkPad) or the attribute of the produce (the screen resolution). Opinion target extraction also called topic extraction or feature extraction. For this task, we proposed an M-Score algorithm to solve the domain limitation problem in the opinion target extraction research.

The rest of this paper is structured as follows: section 2 introduces the related works of opinion target extraction; section 3 describes the details of our proposed approach; section 4 gives the experimental setup and the results analysis; finally, we give a conclusion of our work.

2. Related Works

From the view of sentiment analysis status, many researchers devoted their research to opinion target extraction [9-17]; they employed different methods for this research. Some researchers used the method based on rules/template to extract opinion targets. Yi *et al.* [18] used three progressive limited levels of part-of-speech to extract real opinion targets from candidate opinion targets. Hu and Liu [19] used association rule mining based on the Apriori algorithm to extract opinion targets; they distinguished high frequency opinion targets based on the co-occurrence of the opinion targets, and employed pruning rules to improve the accuracy and coverage. Popescu *et al.* [20] proposed a method by computing a Point-wise Mutual Information score of a noun, and then the Bayesian classification was employed to extract product features. Xu *et al.* [21] developed a method based on heuristic rules for NTCIR-8 tasks of opinion holder extraction and opinion target extraction. The core of this method is constructing rules and using pattern matching to extract opinion targets. Rule/template based method uses pattern matching to extract opinion targets, and these rules/templates established in the process are easy to understand. But it is difficult to guarantee the systematic and logic of these rules/templates, and they have a higher domain-related property which is hard to be transplanted.

Some researchers use natural language processing approach extracting opinion targets. Liu *et al.* [22] got the candidate opinion targets by the syntactic analysis result, and then employed Point-wise Mutual Information algorithm and noun pruning rules to filter the candidate opinion targets. Wang *et al.* [23] developed a method based on syntactic analysis and dependency parsing. They used likelihood testing method to calculate the relevance degree of the candidate opinion targets to the topic, and then ordered the candidate opinion targets and filtered the irrelevant opinion targets to the topic. This method processes sentences by syntactic analysis and semantic role labeling, and it has achieved better performance in the status that the experiment contains a large amount of text, but less useful when the sentences lack syntactic structures or the structures are very completed.

At present, many scholars are keen to the machine learning method. According to the degree of automation of the machine learning model, the method of machine learning can be divided into supervised-based machine learning method or semi-supervised-based machine learning method and unsupervised-based machine learning method. Supervised or semi-supervised-based machine learning method need to label the corpus or label part

of the corpus beforehand, and then train the labeled corpus with specific rules to obtain the available model. Kobayashi *et al.* [24] applied a semi-automatic cycle method to extract product features. Cheng X's [25] research adopted ontology-based extraction method. First of all, they used the semi-automatic method to construct the automobile domain ontology based on the existing resources, and then combined the information extraction engine with the named entity recognition technology to identify the topic in the field of automobile. Instead, unsupervised-based machine learning method doesn't need any labeled corpus. Generally, it uses clustering, bootstrapping and Transmission to extract information. Song *et al.* [26] put forward an unsupervised method without the dependency of external resource in the automobile field. In their paper, they used fuzzy matching algorithm and pruning algorithm to extract opinion targets, and then bootstrapping approach was adopted for the identification of product features from candidate features. Jin w *et al.* [27] utilized the bootstrapping method realized the semi-automatic tagging corpus, and Lexical - HMM classifier is used for the extraction of "product feature entity" and "entity", this paper showed that their method is better than rule-based method.

Supervised-based or semi-supervised-based machine learning method has a higher accuracy compared with the unsupervised-based method, but it takes a lot of manpower resources to label the corpus and time consuming, and its performance completely determined by the quality and quantity of the training sample. Relatively, unsupervised-based machine learning method doesn't need to label the corpus, so it reduces the manual labeling efforts which not only becomes much easier but also gets a good cross-domain performance.

There are some problems of existing opinion target extraction in Chinese: Interdisciplinary ability of extraction is not strong and the system migration is inconvenient; the extraction algorithm doesn't consider comprehensively; the system's learning ability is not strong and so on. As far as possible to solve these problems, we proposed a domain-independent opinion target extraction algorithm. We will give the detail description next.

3. Domain-independent Opinion Target Extraction Model

Figure 1 gives the architecture overview of our proposed domain-independent model for opinion target extraction. The general steps of the model as follows:

Firstly, word segmentation, part-of speech tagging and syntax analysis were done to process the sentences of the document, and then effective opinion sentences are extracted by using Polarity Lexicon. Secondly, the conditional random fields model was employed to train these effective opinion sentences and extract the candidate opinion targets, in the process of training combined the feature template and the results of syntax analysis. Thirdly, the proposed method of M-Score algorithm is used for candidate opinion targets domain-relevant processing to remove the domain-redundant targets. Finally, the candidate opinion targets filtration is done for final results output. Below, we will discuss each step of the proposed model in details.

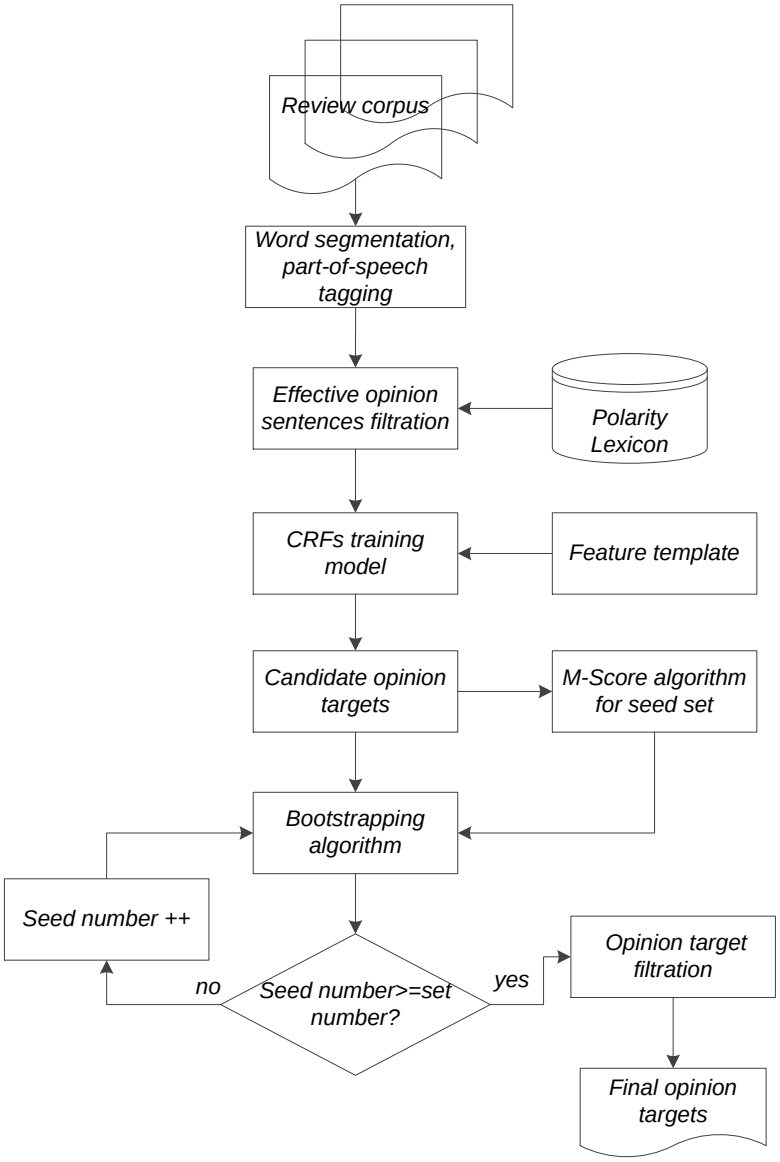

Figure 1. The Proposed Model for Opinion Target Extraction

3.1. Document Preprocessing

One of the important differences between Chinese and English is that Chinese needs to word segmentation. In English language, it is easy to distinguish word by a space or punctuation. But in the Chinese, it needs to word segmentation for the text stream. As word is the smallest semantic unit in Chinese sentences, so the first step of the model is sentences processing. For each sentence, word segmentation and part-of-speech tagging are utilized. Part-of-speech tagging (POS) is the task of assigning parts of speech to each word, such as adjectives, nouns, adverbs and so on. We use the ICTCLAS of Institute of Computing Technology Chinese Academy of Sciences for Chinese word segmentation and part-of-speech tagging. An example is shown in Figure 2. In a large number of review sentences, there are many objective sentences, we are not concerned with these sentences, because these sentences contain no opinion words and give no sentiment tendency, and they are also time-consuming in the experimental process. Thus, we use the polarity lexicon to filter these redundant sentences, just carry out the extraction task from effective

opinion sentences. Polarity lexicon is the lexicon contains many opinion words, and sometimes it can include the degree of polarity of opinion words.

Sentence: 外观比网上的图片漂亮，就是硬盘才 5400 转每分钟，不是很快，要是换成 7200 转每分钟就更好了。(the appearance is much beautiful than on-line picture, but the hard disk just has 5400 revolutions per minute, not so fast, it will be much better if it changes to 7200 revolutions per minute)

After POS: 外观/n 比/p 网上/s 的/ude1 图片/n 漂亮/a ，/wd 就是/v 硬盘/n 才/d 5400/m 转/v 每分钟/n ，/wd 不/d 是/vshi 很/d 快/a ，/wd 要/v 是/vshi 换/v 成/v 7200/m 转/v 每分钟/n 就/d 更/d 好/a 了/y 。/wj

3.2. Candidate Opinion Target Generation

As the Conditional Random Fields (CRFs) model has the better capability to capture the context information, it was first put forward by Lafferty in 2001. Then it has been widely used in the natural language processing by many researchers, such as word segmentation, part-of-speech tagging and named entity recognition. Especially in recent years, it has been successfully employed in opinion target extraction task. CRFs are undirected conditional probability model, that is a kind of statistical model used to tag and segment serialized data. Assume that X is the input set of random variables through the observation labeled sequence, Y is the output set of random variables by the model prediction and corresponding to the labeled sequence. The calculation formula of CRFs is:

$$p(y \mid x) = \frac{1}{Z(x)} \prod_{i \in N} \varphi(y_i, x_i) \tag{1}$$

Where $Z(x)$ is the normalization factor, $\phi_i(y_i, x_i)$ is the potential function and defined as:

$$\varphi_i(y_i, x_i) = \exp(\sum_k \lambda_k f_k(y_i, x_i)) \tag{2}$$

In this paper, we use Conditional Random Fields to extract candidate opinion targets. After the preprocessing of the document review corpus, we employed the Conditional Random Fields model combined with feature template and dependency parsing for candidate opinion targets extraction.

Table 1 gives the description of word features, part-of speech features and dependency parsing features used in the process of CRFs training.

Table 1. Summery of Features

Features	description
Word	the string feature of current word
Part-of-speech	part of speech tagging feature of current word
Dependency parsing	the interdependency feature between word and word

3.3. The Proposed m-Score Algorithm

The candidate opinion target set extracted by Conditional Random Field model must be includes some nouns or noun phrases which irrelevant to the field that customers are not interested. Take the following sentence for example, “The pixel of this mobile phone is very good, but the attitude is poor”. We can extract the opinion target “pixel” and “courier”, but for this sentence what customers concerned is the pixel of the mobile phone not the courier, so the noun “courier” should be removed from the candidate opinion target set. For filtering the redundant words irrelevant to the domain, many researchers

use Pointwise Mutual Information (PMI) algorithm to calculate the mutual information of the current opinion target and the domain words.

First, we give a brief introduction of PMI. PMI is a measure of association used in information theory and statistic. Turney P D. [29] first used PMI in his paper, and then it has been popularly applied in data mining and natural language processing. PMI defined as:

$$PMI(w_i, w_j) = \log \frac{p(w_i, w_j)}{p(w_i)\, p(w_j)} \tag{3}$$

Where $p(w_i, w_j)$ is the probability of co-occurrence of word w_i and word w_j in one sentence, $p(w_i)$ and $p(w_j)$ are the probability of occurrence of word w_i and word w_j in the sentences of corpus respectively.

In the process of PMI calculation, an artificial seed set related to the domain is needed in opinion target extraction task. In the construction process of artificial seed set, it is hard to avoid manual omission and words misunderstanding.

Considering most of the candidate opinion targets are domain-related, we proposed a domain-independent opinion targets extraction method, which is M-Score algorithm. This algorithm needs no person participation to construct domain-related seed set, and it can be directly selected the seed words from the candidate opinion targets. So it has no domain limitation, and available to various fields and essay to be transplanted. The thought of M-Score approach is derived from PMI, but it considers the current word frequency and internal relation information between words at the same time. We will give a detail description of the algorithm next. The M-Score calculation formula is as follows:

$$M - Score(a) = p(a) / N \cdot \sum_k \frac{1}{\log_2[(p(a, b_k) + 1) / (p(a) \cdot p(b_k))]} \tag{4}$$

Where a denotes the current opinion target, k is the words number of seed set, b_k is the kth opinion target of the seed set, $p(a)$ and $p(b_k)$ denotes the number of sentences contain word a and b_k respectively in the review corpus, $p(a, b_k)$ is the co-occurrence sentences number that contain opinion target a and b_k together, N denotes the total sentences number of the review corpus. We add 1 in the formula to avoid that the logarithm value be zero. When the M-Score algorithm is used to generate the initial seed set from the candidate opinion targets, b_k is the kth opinion target of the candidate opinion targets, then we use the seed set to deal with the rest of the candidate opinion targets.

3.4. Bootstrapping Algorithm

After getting a number of seed words, we need to employ the bootstrapping algorithm to calculate the M-Score value of the rest of the candidate opinion targets. The basic idea of bootstrapping algorithm is iterative calculating the candidate opinion targets M-Score values with the seed set, selecting the one with maximum M-score value adds to the seed set, the iteration stops until the seed number equals to the set number. The set number is the opinion target number we will extract from the review. The process of bootstrapping algorithm is described as follows:

Algorithm: M-Score-based iterative bootstrapping algorithm
Input: seed set (SFeatures list), candidate opinion target set (CFeatures list), Set number
Process:

```
While (Seed number< Set number)
        For each Feature in CFeatures
           M=M-Score (Feature);
```


If M>max then max=M;
Add Feature with maximum values to SFeatures list;
Seed number=Seed number+1;
End for
End while

Output: SFeatures list

The specific steps of opinion target extraction are as follows:

S1: preprocess the review corpus, get the sentence set with word segmentation and POS tagging;

S2: syntactic analysis is employed to get the word dependency relation;

S3: use polarity lexicon extracts effective opinion sentences and delete the objective sentences;

S4: train the opinion sentences by CRFs with feature template, get the candidate opinion target set, and call it I_1;

S5: M-Score algorithm is employed to extract a certain number of seed words from I_1, the set of seed words marked as S , and the rest of opinion targets set called I_2;

S6: call bootstrapping algorithm, calculate the M-Score value of each current opinion target a and b_i $(a \in I_2, b_i \in S)$, and the current seed number value add 1, then compare the current seed number with the set number; if the current seed number less than the set number, then put a into S , and delete it from I_2, otherwise, stop the while loop; repeat this step until the current seed number equals to the set number;

S7: the final set of S as the final opinion target set output.

3.5. Opinion Target Filter

After the above steps of opinion target extraction, the results still contain some repeat or non-key opinion targets, so we need to further process these redundant ones. In this process, we use word frequency filtering technique which is minimum p-support pruning algorithm for opinion targets filter.

Word frequency filtering technique aims to filter the very small frequency words. Generally, people are not concerned about these words. Take the following opinion sentence into consideration: "A mobile phone's performance is really good, the pixel is very high, it doesn't like my previous phone B, its pixel is very bad."). We can extract one of the opinion target "phone B" from this sentence , but what customers concerned is all the features of phone A, so "B phone" is the non-key target, and it will have a very low frequency in the corpus, we can use the word frequency filtering technique to filter out this one. Even if occasionally filtered out several real opinion targets, it doesn't affect the performance of the system for a large size of data set. Li *et al.* [30] analyzed the neighboring rules pruning algorithm and the minimum p-support pruning algorithm. According to their research results, the minimum p-support pruning algorithm achieved better performance. So we adopt this method for opinion targets filtering. Let's give a brief description of minimum p-support firstly. For example, "battery" as an opinion target, "battery life" and "battery capacity" are the opinion targets at the same time, and they both contain the word "battery". Assume there are 20 sentences contain the noun "battery" in the review corpus, and there are 7 and 8 sentences that contain "battery life" and "battery capacity" respectively, then the number of the noun "battery" appears alone in the sentences is 5, then we said that the minimum p-support of "battery" is 5. In short, the minimum p-support is the number of sentences that contain the noun or noun phrase alone but don't contain its superset.

4. Experiments and Analysis

4.1. Corpus and Evaluation Measures

In the experiments, the dataset consists of three domains of online product reviews; they are mobile phones, digital cameras and laptop computers. The detail of the dataset can be seen in Table 2.

We use precision (P), recall (R) and F-score to evaluate the experiment results, and the calculation formulas based on Table 3. The calculate formulas as follow:

$$P = \frac{TP}{TP + FP} \quad (5), \qquad R = \frac{TP}{TP + FN} \quad (6), \qquad F = \frac{2 \cdot P \cdot R}{P + R} \quad (7)$$

Table 2. Datasets of the Experiments

	Mobile phones	Digital cameras	Laptop computers	Total
Number of review texts	112	96	104	312
Number of sentences	1721	1256	1342	4319
Number of product features	1731	1012	1521	4264

Table 3. Evaluation Criterion

Evaluation parameters	Correct opinion targets	Wrong opinion targets
Extracted opinion targets	TP	FP
Non-extracted opinion targets	FN	TN

4.2. Experiment Analysis

The proposed M-Score algorithm is an unsupervised method, the seed words are extracted from candidate opinion targets. The key problem of the seed set extraction process is finding an appropriate number of the seed words. For this problem, we experimented different numbers (*i.e.* 15, 20, 25, 30, 35, 40) to observe their effects on the experimental results. Figure 3 shows the change curve of different seed number and the F-score. From this figure, we can see that 30 is an appropriate number for the mobile phone domain and the digital camera domain, and 35 is an appropriate number for laptop computer domain. In order to achieve the unity of the experiments, we chose the number 30 and extract the first 30 number of opinion targets with larger M-Score value as the seeds in the experiments. In the while loop iteration, the stop set number is 70.

In order to give more accurate results, we adopt 3-fold cross-validation method in the experiments. The datasets are randomly divided into three equal sub sets; two of them are randomly selected for training set, the rest one for testing set. The final result is the average of the three experiments. Firstly, we employed the Base Line experiment to show the M-score algorithm's effectiveness. The Base Line experiment is that we just filter the candidate opinion targets, after the filtration process output the final extraction results. The results can be seen in Table 4. Secondly, we compared our proposed method with traditional PMI algorithm and the result is showed in Table 5.

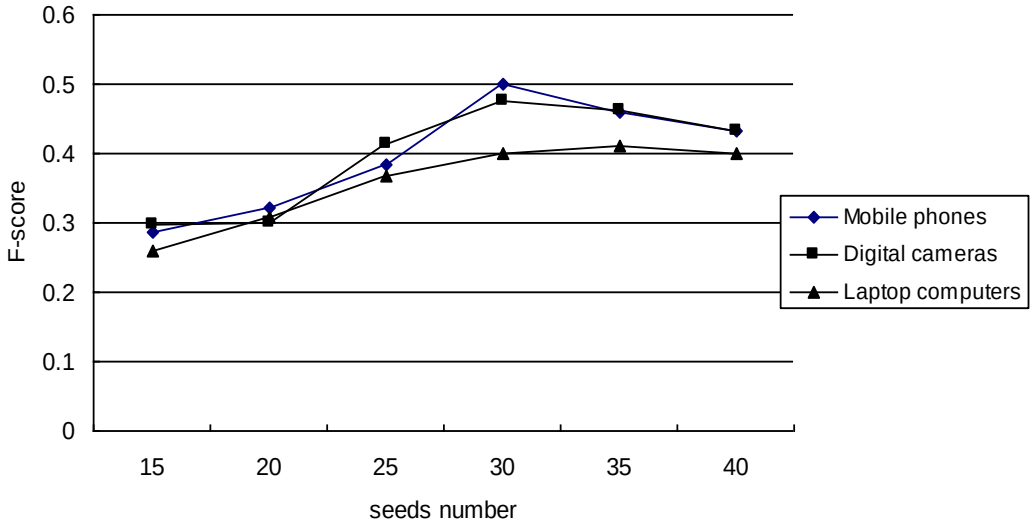

Figure 3. Different Seeds Number and the F-score Change Curve

Table 4. Base Line Results

Products	Base Line method			M-score algorithm method		
	P	R	F	P	R	F
Mobile phones	0.3021	0.4101	0.3479	0.4931	0.5061	0.4995
Digital cameras	0.2679	0.3800	0.3143	0.4610	0.4912	0.4756
Laptop computers	0.2514	0.3108	0.2780	0.4035	0.3968	0.4001
Average	0.2738	0.3670	0.3134	0.4525	0.4647	0.4584

It is clearly observed from Table 4 that the Base Line method's precision, recall and F-score values are all lower than the M-score algorithm method in three domains. One explanation of this result is that the Base Line method didn't deal with the domain-related processing after the filtration process of the candidate opinion targets, so the result still contains some opinion targets unrelated to the domain which are not the correct ones.

Table 5. Comparison Results of the M-score Algorithm with PMI Algorithm

Products	PMI algorithm			M-score algorithm		
	P	R	F	P	R	F
Mobile phones	0.4472	0.4810	0.4635	0.4931	0.5061	0.4995
Digital cameras	0.3919	0.4006	0.3962	0.4610	0.4912	0.4756
Laptop computers	0.3879	0.4566	0.5200	0.4035	0.3968	0.4001
Average	0.4090	0.4461	0.4599	0.4525	0.4647	0.4584

From Table 5 we can see that the M-score algorithm performed better in precision and recall than traditional PMI algorithm, The F-score of M-score algorithm exceeds PMI

algorithm in the mobile phone domain and digital camera domain, but lower than PMI in the laptop computer domain. The reason of the result is that when using PMI algorithm, it needs some artificial seeds, because of the human intervention, the process of seed collection maybe not comprehensive or error, and that is inevitable. Thus, its experimental precision is lower than M-score algorithm as the consequence. On the contrary, M-Score method needs no human intervention for seeds collection, it directly collects the seed set from candidate targets, and its experimental precision dependent on the number of seed words, proper seed number can gain a good result. Besides, in order to obtain more accurate results, we can adjust the seed number according to the size and domain of the dataset.

In order to promote the development of view information retrieval, extraction, tendency analysis and the construction of Chinese tendency analysis corpus, the Chinese Information Society of Information Retrieval Committee sponsored the Chinese Opinion Analysis Evaluation in 2008 (COAE 2008). The main aim of COAE 2008 is promoting the construction of Chinese tendency analysis lexicon, improving the technology of Chinese subjective and objective analysis and opinion target extraction and developing view retrieval technique. Over the following years, many researchers take this evaluation as the results reference. In the experiment, we also employed the average evaluation results of COAE2008 evaluation task 3 (Chinese text tendency related factors extraction, COAE2008-3) in comparison to the proposed method. See Table 6.

Table 6. Comparison Results of M-score Method and COAE2008-3

Evaluation	Mobile phones	Digital cameras	Laptop computers	Average of M-score algorithm	Average of COAE2008-3
P	0.4931	0.4610	0.4035	0.4525	0.3798
R	0.5061	0.4912	0.3968	0.4647	0.4172
F	0.4995	0.4756	0.4001	0.4584	0.3976

From Table 6, it is notable that the performance of the proposed model in precision, recall and F-score all performed better than the average evaluation results of COAE2008 evaluation task 3, and it well demonstrated the effectiveness and practicality of M-Score method.

5. Conclusions

In this paper, we studied the extraction of opinion targets from customer reviews, and proposed the M-Score algorithm based on PMI method. This algorithm is the further improvement of PMI algorithm. The M-score based model is domain-independent and easy to be transplanted. This model has achieved better performance and the experimental results proved the validity of this method.

Though our proposed method has yielded good results, the performance of the model needs to be further improved, and the precision and recall of extraction need to further improve, too. What's more, this paper only deal with the opinion targets extraction task, but made no tendency analysis and summary of opinion targets, so in our further work we plan to further improve the accuracy of opinion targets and give the tendency analysis of these targets and make a reference summarized report for users at the same time.

Acknowledgements

This research is supported by the National Natural Science Foundation of China (No. 60743008) and the Project Sponsored by the Scientific Research Foundation for the

Returned Overseas Chinese Scholars, State Education Ministry of China. We thank the Institute of Computing Technology Chinese Academy of Science for providing us the word segmentation system.

References

[1] Y. V. Hatzivassiloglou and K R. McKeown, "Predicting the semantic orientation of adjectives", Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, (**1997**) July12-15, Madrid, Spain.
[2] R M. Tong, "An operational system for detecting and tracking opinions in on-line discussion", Proc of ACM SIGIR Workshop on Opera-tional Text Classification, (**2001**) September 21-23, New Orleans, USA.
[3] B. Liu, M.-q. Hu and J.-s. Cheng, "Opinion observer: analyzing and comparing opinions on the Web Proc. of the14th International Conference on World Wide Web", (**2005**) May18-20, New York, USA.
[4] K. Dave, S. Lawrence and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews", Proceedings of the 12th international conference on World Wide Web", ACM, (**2003**) May 20-24, Budapest, Hungary.
[5] B. Pang and L. Lee, "Opinion mining and sentiment analysis", Foundations and trends in information retrieval, vol. 2, nos. 1-2, (**2008**), pp. 1-135.
[6] W. Jin, H. H. Ho and R. K. Srihari, "OpinionMiner: a novel machine learning system for web opinion mining and Extraction", Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, (**2009**), Paris, France.
[7] Y. Y. Zhao, B. Qin and T. Liu, "Sentiment analysis [J]", Journal of Software, vol. 21, no. 8, (**2010**), pp. 1834-1848.
[8] W. Wei, Y. Xiang and Q. Chen, "Survey on Chinese text sentiment analysis", Journal of Computer Applications, vol. 31, no. 12, (**2012**), pp. 3321-3323.
[9] H. Liu, Y. Zhao, B. Qin and T. Liu, "Comment Target Extraction and Sentiment Classification", Journal of Chinese Information Processing, vol. 24, no. 1, (**2010**), pp. 84-88.
[10] J. Zhu, H. Wang and M. Zhu, "Aspect-based opinion polling from customer reviews", Affective Computing, IEEE Transactions on, vol. 2, no. 1, (**2011**), pp. 37-49.
[11] I. Chetviorkin and N. Loukachevitch, "Extraction of Domain-specific Opinion Words for Similar Domains", Proceedings of the Workshop on Information Extraction and Knowledge Acquisition held in conjunction with RANLP (**2010**) October 26-30, Toronto, Canada.
[12] T. Ma and X. Wan, "Opinion target extraction in Chinese news comments", Proceedings of the 23rd International Conference on Computational Linguistics: Posters. Association for Computational Linguistics, (**2010**), pp. 782-790.
[13] F. Xianghua, L. Guo and G. Yanyan, "Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon", Knowledge-Based Systems, vol. 37, (**2013**), pp. 186-195.
[14] A. Duric and F. Song, "Feature selection for sentiment analysis based on content and syntax models", Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, Association for Computational Linguistics, (**2011**), pp. 96-103.
[15] C. Quan and F. Ren, "Unsupervised product feature extraction for feature-oriented opinion determination", Information Sciences, vol. 272, (**2014**), pp. 16-28.
[16] L. Chen, L. Qi and F. Wang, "Comparison of feature-level learning methods for mining online consumer reviews", Expert Systems with Applications, vol. 39, no. 10, (**2012**), pp. 9588-9601.
[17] F. He, Y. He, N. Liu, J. b. Liu and M. Peng, "A Micro-blogging Short Text Oriented Multi-class Feature Extraction Method of Fine-grained Sentiment Analysis", Acta Scientiarum Naturalium Universitatis Pekinensis, vol. 1 (**2014**), pp. 48-54.
[18] J. Yi, IBM Almaden Res. Center San Jose CA USA, T. Nasukawa, R. Bunescu and W. Niblack, "Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques", Third IEEE International Conference on Data Ming: ICDM 2003, (**2003**) November 19-22; Melbourne, Florida.
[19] M. Hu and B. Liu, "Mining opinion features in customer reviews", AAAI, (**2004**) March 22-24, Stanford university in Palo Alto, California.
[20] A. M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews", Natural language processing and text mining, (**2007**), London, UK.
[21] R. Xu and C. Kit, "Incorporating feature-based and similarity-based opinion mining--ctl in ntcir-8 moat", Proceedings of the 8th NTCIR Workshop, (**2010**) June 15-18, Tokyo, Japan.
[22] H. Liu, Y. Zhao, B. Qin and T. Liu, "Comment Target Extraction and Sentiment Classification", Journal of Chinese information processing, vol. 1, (**2010**), pp. 84-88.
[23] W. Wang and C. Meng, "Opinion Object Extraction Based on the Syntax Analysis and Dependence Analysis. Computer Systems &Applications", vol. 20, no. 8, (**2011**), pp. 52-57.

[24] N. Kobayashi, K. Inui and Y. Matsumoto, "Collecting evaluative expressions for opinion extraction", Natural Language Processing–IJCNLP 2004. (2005), Berlin Heidelberg, Germany.
[25] X. Cheng, "Automatic Topic Term Detection and Sentiment Classification for Opinion Mining", Saarbrücken, Germany: The University of Saarland, (**2007**).
[26] X. Song, S. Wang and H. Li, "Research on Comment Target Recognition for Specific Domain Products", Journal of Chinese information processing, vol. 1, (**2010**), pp. 89-93.
[27] W. Jin, H. H. Ho and R. K. Srihari, "OpinionMiner: a novel machine learning system for web opinion mining and extraction", Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, (**2009**); Paris, France.
[28] S. Hui and S. Nan-sheng, "Comment Object Extraction Based on Pattern Maching and Semi-supervised Learning", Computer Engineering, vol. 39, no. 10, (**2013**), pp. 221-226.
[29] P. D. Turney, "Thumbs up or thumbs down?", semantic orientation applied to unsupervised classification of Reviews, Proceedings of the 40th annual meeting on association for computational linguistics, Association for Computational Linguistics, (**2002**) July 7-12, University of Pennsylvania, America.
[30] S. Li and Q. S. Li, "Research on Pruning Algorithm of Product Feature Mining in Chinese Review", Computer Engineering, vol. 37, no. 23, (**2011**), pp. 43- 45.

Authors

Sun Yongmei, she is a postgraduate student of Information Engineering College, Henan University of Science and Technology. Her research interests in intelligent information processing and data mining.

Huo Hua, he is a professor with a Ph.D. in Information Engineering College and School of Software, Henan University of Science and Technology. His research interests in intelligent information processing and video semantic extraction.

