Research on Sentiment Dictionary Construction Method for Chinese Book Reviews (Postprint)

Authors: Guo Shunli Zhang Xiangxian

Date: 2017-10-11T00:00:00+00:00

Abstract

[Objective] To investigate methods for constructing sentiment lexicons for Chinese book reviews to facilitate sentiment analysis of user book reviews.

[Method] Drawing upon relevant studies, user emotions were categorized into 7 types. The collected corpus was segmented, and combined with a foundational sentiment lexicon to derive a sentiment word set for Chinese book reviews, from which seed words for each emotion category were selected. An improved SO-PMI algorithm and a Tongyici Cilin (Synonym Forest) expansion approach were employed to determine the sentiment categories of words. Actual book reviews were utilized as corpus for experimental validation.

[Results] A method for constructing sentiment lexicons for Chinese book reviews is proposed, achieving average precision, average recall, and mean F1-score of 0.90, 0.83, and 0.85, respectively.

[Limitations] The corpus size is small, and the sample scope exhibits certain limitations.

Conclusion Experimental results demonstrate that the proposed method exhibits high effectiveness and reliability, and can effectively perform sentiment analysis of user book reviews.

Full Text

Preamble

ChinaXiv Cooperative Journal, Issue 267, 2016, Issue 2
Method for Building Sentiment Dictionary for Chinese Book Reviews
Guo Shunli, Zhang Xiangxian
(School of Management, Jilin University, Changchun 130022, China)

Abstract

[**Objective**] This study explores methods for constructing a sentiment dictionary for Chinese book reviews to enable sentiment analysis of user-generated book reviews. [**Methods**] We divided user emotions into seven categories based on relevant research, segmented the collected corpus, and combined it with a basic sentiment dictionary to obtain a sentiment word set for Chinese book reviews. Seed words for each category were selected, and an improved SO-PMI algorithm combined with synonym expansion was used to determine the sentiment categories of words. Actual book reviews served as the corpus for experimental verification. [**Results**] The proposed method for building a Chinese book review sentiment dictionary achieved average precision, recall, and F1 scores of 0.90, 0.83, and 0.85, respectively. [**Limitations**] The corpus size is relatively small, and the sample scope has certain limitations. [**Conclusions**] Experimental results demonstrate that the proposed method is highly effective and reliable for conducting sentiment analysis of user book reviews.

Keywords: Chinese book reviews; Sentiment dictionary; Seed word; Sentiment classification; SO-PMI algorithm

Classification Number: G353

Introduction

With the development of the Internet, an increasing number of users express personal opinions on current events, products, and policies through online social platforms, generating user reviews that contain substantial sentiment words reflecting individual emotions. In the commercial domain, user reviews represent evaluations of products and services after purchase or experience, embodying users' personal sentiment information typically used for business feedback. User book reviews specifically refer to comments or introductions about books —texts that analyze a book' s form and content in a factual and insightful manner to explore its ideological, academic, knowledge-based, and artistic qualities, thereby establishing an information exchange channel among authors, readers, and publishers. These reviews express users' evaluations of book content and personal sentiment perspectives after reading, reflecting their emotional information about books. Utilizing user book reviews for sentiment analysis can better mine user behavior and provide recommendations for book publishing and other users' reading choices. Consequently, how to quickly and accurately conduct sentiment analysis on large volumes of user book reviews has become an important research topic.

Sentiment analysis of user book reviews requires a sentiment dictionary. Currently, no comprehensive large-scale Chinese sentiment dictionary for book reviews exists in China. A Chinese book review sentiment dictionary is a prerequisite for such analysis, and automatically extracting sentiment words from large volumes of user book reviews to construct such a dictionary has become an urgent problem to solve.

2. Research Status

Foreign researchers have primarily built English sentiment dictionaries based on WordNet. Turney achieved good results in unsupervised sentiment analysis by improving the PMI-IR algorithm. Subasic et al. manually constructed an emotion category-related dictionary that marked word intensity (the strength of emotional expression) and centrality (the degree of relevance to the category). Domestic research on Chinese sentiment dictionary construction has also yielded achievements, with many studies building domain-specific dictionaries based on HowNet and *Tongyici Cilin* (Extended). For instance, Liu Weiping et al. proposed a method for calculating emotional weights of Chinese sentiment words based on Chinese word similarity computation, constructing a Chinese basic sentiment dictionary using a sentiment word set as the benchmark. Research on microblog sentiment dictionary construction is relatively extensive, with scholars employing various methods from different perspectives. Li Yu integrated basic sentiment dictionaries, function word dictionaries, emoticon sentiment dictionaries, and internet slang sentiment dictionaries to create a microblog sentiment dictionary. Gui Bin et al. proposed an automatic sentiment dictionary construction method based on microblog emoticons. Other researchers have constructed domain-specific sentiment dictionaries, such as Zhou Yongmei et al., who proposed a news comment sentiment dictionary construction method based on graph ranking models, and Jiang Shengyi et al., who built a Chinese music domain sentiment dictionary using an improved Hevner emotion ring model, semantic resources from HowNet, and lyric text corpora crawled from the internet. Additional domain-specific dictionaries include those for hotel reviews, microblog product reviews, and movie reviews.

Our investigation reveals that sentiment research for the Chinese book review domain is scarce, with related sentiment dictionary research nearly blank. Chinese book reviews differ from other user reviews in that they employ many words with strong literary and academic qualities—some sentiment words rarely appear in other domains and possess distinct professional characteristics (e.g., “surprised,” “brilliant writing”). Additionally, Chinese book reviews follow fixed writing formats with certain normative features. Therefore, sentiment research from other domains cannot be effectively applied to Chinese book review sentiment analysis, making this research valuable and significant. This paper proposes a method for constructing a Chinese book review sentiment dictionary to facilitate subsequent sentiment analysis.

4. Research on Chinese Book Review Sentiment Dictionary Construction Based on SO-PMI Algorithm

The basic process of our proposed Chinese book review sentiment dictionary construction method is shown in Figure 1 [Figure 1: see original paper].

1. Referencing the sentiment classification method in literature [13], we divide Chinese book review sentiments into seven major categories.

2. Using the ROST CM6 segmentation tool [14], we segment the collected Chinese book review corpus and perform word frequency statistics. After comparative analysis with basic sentiment dictionaries, we comprehensively obtain the Chinese book review sentiment word set.
3. We generate seven categories of seed sentiment words for Chinese book reviews. Based on the sentiment word set, we query the intensity of sentiment words in the sentiment lexicon ontology, combine this with word frequencies in the sentiment word set, and use manual screening to determine seed words.
4. Using an improved SO-PMI algorithm, we classify sentiment words in the Chinese book review sentiment word set (excluding seed words) into sentiment categories. The *Tongyici Cilin* (Extended) [15] expansion method solves data sparsity problems. We obtain the relevance degree of each sentiment word and categorize it into one of the seven Chinese book review sentiment categories based on relevance magnitude.
5. The Chinese book review sentiment dictionary is obtained.

4.1. Sentiment Classification for Chinese Book Reviews

Chinese book reviews reflect readers' emotions, and different book types associate with different sentiment categories. Therefore, it is necessary to classify user emotions expressed in Chinese book reviews. Currently, psychology has no universally accepted standard for emotion classification, with researchers dividing emotions into 4, 6, 8, 10, or even more than 20 categories. Referencing the user emotion classification method in literature [13] and integrating existing sentiment vocabulary resources, we divide Chinese book review user emotions into seven categories: joy, good, anger, sorrow, fear, disgust, and surprise.

4.2. Generation of Chinese Book Review Sentiment Word Set

Internet user book reviews contain many new internet words and slang that existing sentiment dictionaries cannot cover. Therefore, we construct user book review data by integrating existing sentiment dictionaries, sentiment words from user book review corpora, and internet slang. The main sentiment dictionary resources used in constructing the Chinese book review sentiment word set include the Emotional Lexicon Ontology from Dalian University of Technology' s Information Retrieval Laboratory [13], HowNet Sentiment Dictionary [16], NTUSD Chinese Sentiment Polarity Dictionary from National Taiwan University [17], and an Internet sentiment dictionary constructed through manual collection.

We used the GooSeeker web crawler [18] to crawl book reviews from Douban Books, covering literature, popular culture, lifestyle, economics, and technology categories. After preprocessing to remove neutral comments, spam, and convert emoticons, we collected over 255,000 user book reviews on 8,500 books. Using the ROST CM6 segmentation tool for word segmentation and frequency statistics, we manually screened and removed low-frequency words to form Chinese book review word set WordSet1. We merged sentiment words from the Emo-

tional Lexicon Ontology, HowNet Sentiment Dictionary, NTUSD, and internet sentiment dictionary to form basic sentiment dictionary word set WordSet2. By comparing WordSet1 and WordSet2 and taking their intersection, we formed Chinese book review sentiment word set WordSet, which contains 881 sentiment words.

4.3. Selection of Seed Sentiment Words

This study employs an improved SO-PMI algorithm for word sentiment orientation judgment, requiring seed word selection. Seed words are those with obvious, strong, and representative sentiment attitudes. Based on WordSet from Section 4.2, we queried these words' intensity in the sentiment ontology, selecting words with maximum intensity and high corpus frequency as candidate seed sentiment words. For example, the word "reunion" has an intensity of 9 (strongest) in the sentiment ontology, belongs to the joy subcategory and joy major category, and appears 4,782 times in the Chinese book review corpus, making it a candidate seed word for joy. Through this selection process, we obtained 191 seed sentiment words across seven categories, forming seed word set S, as shown in Table 1 .

4.4. Sentiment Category Determination Based on Improved SO-PMI Algorithm

Word sentiment category judgment methods mainly include HowNet-based semantic similarity calculation and SO-PMI-based sentiment orientation calculation [19-20]. Since user book reviews contain numerous internet neologisms (e.g., "awesome," "positive energy," "screwed") that lack sememes in HowNet, making similarity calculation impossible, the HowNet-based method is unsuitable for judging sentiment categories of some words in Chinese book reviews. Therefore, we propose an improved SO-PMI-based word sentiment category discrimination method that uses mutual information to calculate the association between unknown words and various seed words for sentiment category judgment.

The basic PMI formula is:

$$PMI(word_1, word_2) = \log \frac{P(word_1, word_2)}{P(word_1)P(word_2)}$$

Where $P(word_1, word_2)$ represents the co-occurrence probability of $word_1$ and $word_2$, $P(word_1)$ represents the occurrence probability of $word_1$ in the corpus, and $P(word_2)$ represents the occurrence probability of $word_2$.

Using word occurrence counts to replace probabilities, and considering that the distance between two co-occurring words is inversely proportional to their association strength (i.e., closer words have stronger association; distant words have weaker association), we incorporate this into sentiment orientation analysis: closer words are more likely to share sentiment orientation. We measure inter-word distance by character count, using the minimum distance between two

words in the same review as their co-occurrence distance d, calculated as:

$$d = \min |d|$$

Where d represents the co-occurrence distance between two words, x_d represents the character count from the review start to the last character of the preceding word, and y_d represents the character count from the review start to the first character of the following word.

We incorporate this co-occurrence distance d into the mutual information calculation formula, transforming formula (1) into formula (3):

$$PMI(word_1, word_2) = \log \frac{N \cdot hit(word_1, word_2) \cdot d}{hit(word_1) \cdot hit(word_2)}$$

Where N represents the total word count in the corpus, $hit(word_1, word_2)$ represents the total co-occurrence count of $word_1$ and $word_2$ within reviews of the same book, and $hit(word)$ represents the word frequency count.

The SO-PMI value between unknown sentiment word $xword$ and seed word set S_i for emotion category i is the sum of PMI values between $xword$ and each seed word in S_i, as shown in formula (4):

$$SO\_PMI(word, S_i) = \sum_{a=1}^{M_i} PMI(word, s_a)$$

Where s_a represents the a-th sentiment word in emotion category i' s seed word set S_i, and $hit(word, s_a)$ represents the co-occurrence count of $xword$ and s_a within reviews of the same book.

Since each emotion category has a different number of seed words, an unknown sentiment word might not be most related to a category but could have the highest total relevance simply because that category has more seed words and thus more accumulation terms. To avoid this bias, we incorporate M_i (the number of seed words in category i's seed set S_i) into the formula. The improved algorithm is:

Algorithm for Sentiment Category Classification Using Improved SO-PMI

Input: WordSetX (sentiment words to be classified), S (seed word sets), synonym set SameWord, N (total word count), M_i $(1 \leq i \leq 7)$.

Output: Sentiment category classification for unknown words in WordSetX.

1. Extract unknown sentiment word $xword$ from WordSetX, where seed word sets for each of the 7 emotion categories are S_i $(1 \leq i \leq 7)$.
2. Calculate the association degree $SO\_PMI(word, S_i)$ between $xword$ and each of the 7 emotion categories using formula (4).

3. Rank the calculated $SO\_PMI(word, S_i)$ values in descending order and select the maximum as the basis for determining $xword$'s emotion category 归属.
4. If the maximum $SO\_PMI(word, S_i)$ is not 0, proceed to step 6; if it is 0, use SameWord to expand $xword$' s synonym set WordSameB, calculate the number of synonyms B, then compute the affiliation degree $SO\_PMI(WordSame, S_i)$ for each word in WordSameB with the 7 emotion categories.
5. Return to step 3 to determine $xword$' s emotion category. If $SO\_PMI(WordSame, S_i)$ remains 0, further expand $xword$' s synonym set WordSameB using step 4' s method, with a maximum of three expansion iterations. If the maximum $SO\_PMI(WordSame, S_i)$ remains 0, the word' s sentiment intensity is considered weak and it is removed directly.
6. Output the sentiment category classification result for $xword$.
7. Algorithm ends.

To address data sparsity issues caused by corpus scale and varying user writing habits (where different users may express the same meaning with different words, and even the same user often uses synonyms), we expand candidate words using *Tongyici Cilin* (Extended) when calculating sentiment polarity. This mitigates data sparsity from low-frequency words. Experiments show that excessive expansion iterations cause semantic loss, so we set the maximum expansion iterations to three.

5. Experimental Validation

To verify the effectiveness of our Chinese book review sentiment dictionary construction method, we conducted experimental validation from two aspects: word sentiment category determination accuracy and classification performance using the constructed dictionary. Using the GooSeeker crawler [18], we collected book reviews for 100 books from Douban, totaling over 15,000 reviews to form a corpus. After preprocessing and data cleaning, we selected 5,000 reviews for experiments and manually annotated their 7-category sentiment classifications. We manually annotated the sentiment categories of 690 Chinese book review sentiment words (excluding seed words) obtained in Section 4.2.

5.1. Accuracy Verification Experiment

Through querying the sentiment ontology intensity table and manual judgment, we obtained manual classification results for 690 sentiment words. We then classified these words using both the original SO-PMI algorithm and our improved SO-PMI algorithm from Section 4.4. The results are shown in Table 2 .

Table 2. Distribution of Sentiment Words Across 7 Categories

Category	Joy	Good	Anger	Sorrow	Fear	Disgust	Surprise
SO-PMI Algorithm Classification	49	204	46	52	47	182	46
SO-PMI Correct Classification	42	182	38	45	40	160	40
Improved SO-PMI Classification	49	204	46	52	47	182	46
Improved SO-PMI Correct Classification	45	188	41	47	42	166	42

This experiment uses precision (P), recall (R), and F1-score (F1) as evaluation metrics to compare the performance of the improved SO-PMI algorithm with the original. Results are calculated and shown in Table 3 , with comparative visualizations created using SPSS 19.0 shown in Figure 2 [Figure 2: see original paper].

Table 3. Performance Evaluation of SO-PMI Algorithm for 7 Emotion Categories

Metric	Joy	Good	Anger	Sorrow	Fear	Disgust	Surprise	Overall
SO-PMI Precision	0.86	0.89	0.83	0.87	0.85	0.88	0.87	0.87
SO-PMI Recall	0.82	0.85	0.76	0.81	0.78	0.82	0.80	0.81
SO-PMI F1	0.84	0.87	0.79	0.84	0.81	0.85	0.83	0.83

Metric	Joy	Good	Anger	Sorrow	Fear	Disgust	Surprise	Overall
Improved SO-PMI Precision	0.92	0.92	0.89	0.90	0.89	0.91	0.91	0.90
Improved SO-PMI Recall	0.88	0.88	0.82	0.85	0.82	0.85	0.84	0.85
Improved SO-PMI F1	0.90	0.90	0.85	0.87	0.85	0.88	0.87	0.86

As shown in Tables 2-3 and Figure 2, the improved SO-PMI algorithm achieved average precision, recall, and F1 values of 0.89, 0.85, and 0.86, respectively—all higher than the original SO-PMI algorithm. Thus, the improved algorithm demonstrates better performance and higher accuracy and usability.

For example, when classifying the word "value" in the review: "This book has certain value, highly recommended. The author uses tragic endings to explain missing childhood, resonating deeply with readers. I didn' t cry while reading—am I particularly cold-blooded? But it' s better than I expected," our method from Section 4.4 correctly classifies "value" under the "good" category. Without considering seed word count differences and co-occurrence distance d, the association degree for "sorrow" would exceed that for "good," causing classification errors. Incorporating d and seed word count M improves algorithm accuracy.

Synonym expansion using *Tongyici Cilin* also solves data sparsity. For the comment "Gave up halfway, too bitter, some things I can' t agree with," when judging "too bitter," the short text contains no seed words from the 7 categories. Expanding "give up" via synonyms yields: {甩掉, 放任, 遗弃, 放手, 舍弃, 吐弃, 丢弃, 抛却, 屏弃, 抛弃, 罢休, 甩手, 松手, 停止, 牺牲, 摒弃, 摈弃, 唾弃, 废弃, 放胆, 撒手}. Using this synonym set enables classification of "too bitter" as "sorrow," thereby solving data sparsity and improving judgment accuracy.

Results also show that seed word quantity affects the improved SO-PMI method. Categories "good," "disgust," and "sorrow" with more seed words achieve relatively higher precision, recall, and F1 values. However, our method overall outperforms the original SO-PMI algorithm. Additionally, the number of words in the sentiment dictionary affects classification performance—more words yield higher recall. Short text segmentation and sentiment feature extraction in Chinese book reviews also impact results.

5.2. Classification Experiment

The purpose of building a Chinese book review sentiment dictionary is to use it for sentiment analysis. We validated the dictionary' s effectiveness through comparative experiments using the collected 100-book corpus. After segmenting the 5,000 reviews and extracting sentiment words, we used the method from Section 4.4 for sentiment category judgment to achieve classification. Table 4 shows the manual annotation results and classification results using our dictionary.

Table 4. Sentiment Classification Statistics for 5,000 Book Reviews

Category	Joy	Good	Anger	Sorrow	Fear	Disgust	Surprise
Manual Classification	437	1,384	406	489	476	1,113	138
Dictionary Classification	402	1,280	371	445	432	1,020	127

This experiment also uses precision, recall, and F1-score as evaluation metrics, with results shown in Table 5 .

Table 5. Performance Evaluation of Chinese Book Review Sentiment Dictionary

Metric	Joy	Good	Anger	Sorrow	Fear	Disgust	Surprise	Overall
Precision	0.91	0.90	0.89	0.91	0.90	0.90	0.89	0.90
Recall	0.83	0.84	0.82	0.83	0.82	0.84	0.82	0.83
F1	0.87	0.87	0.85	0.87	0.86	0.87	0.85	0.85

The dictionary achieved average precision of 0.90, recall of 0.83, and F1 of 0.85, demonstrating good feasibility and accuracy for book review sentiment classification.

Conclusion

This study proposes a domain-oriented sentiment dictionary construction method for Chinese book reviews, dividing user emotions into seven categories and developing an improved SO-PMI algorithm to discriminate sentiment categories of domain-specific words. Comparative experiments verify the method' s good accuracy and reliability. This approach can also be extended to other domains.

Limitations include: errors in short text segmentation and word frequency statistics for Chinese book reviews; influence of adverbs and conjunctions on sentiment judgment; and impacts of seed word quantity selection and corpus scale on

word categorization. Future research will focus on rational seed word selection and corpus expansion.

References

[1] Book Reviews [EB/OL]. [2015-03-03]. http://baike.baidu.com/view/978454.htm.

[2] Andreevskaia A, Bergler S. Mining WordNet for a Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses [C]. In: Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics. 2006: 209-216.

[3] Turney P D. Thumbs up or Thumbs down?: Semantic Orientation Applied to Unsupervised Classification of Reviews [C]. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. 2002: 417-424.

[4] Subasic P, Huettner A. Affect Analysis of Text Using Fuzzy Semantic Typing [C]. In: Proceedings of the 9th IEEE International Conference on Fuzzy Systems. IEEE, 2001.

[5] Liu Weiping, Zhu Yanhui, Li Chunliang, et al. Research on Building Chinese Basic Semantic Lexicon [J]. Journal of Computer Application, 2009, 29(10): 2875-2877.

[6] Li Yu. Microblog Emotional Dictionary Built and Application on Sentiment Analysis of Microblog [D]. Zhengzhou: Zhengzhou University, 2014.

[7] Gui Bin, Yang Xiaoping, Zhang Zhongxia, et al. Research on Building Lexicon for Sentiment Analysis Based on the Chinese Microblogging Smiley [J]. Transactions of Beijing Institute of Technology, 2014, 34(5): 537-541.

[8] Zhou Yongmei, Yang Aimin, Yang Jianeng. Construction Method of Sentiment Lexicon for New Reviews [J]. Computer Science, 2014, 41(8): 67-69, 80.

[9] Jiang Shengyi, Yang Yao, Liao Jingxin. Research of Building Chinese Musical Emotional Lexicon and Emotional Classification [J]. Computer Engineering and Applications, 2014, 50(24): 118-121, 163.

[10] Yang A M, Lin J H, Zhou Y M, et al. Research on Building a Chinese Sentiment Lexicon Based on SO-PMI [J]. Applied Mechanics and Materials, 2013, 263-266: 1688-1693.

[11] Yu Zhenzhi. Research on the Key Technologies of Chinese Online Product Review' s Sentiment Analysis [D]. Hangzhou: Hangzhou Dianzi University, 2011.

[12] Li Ming. Emotion Classification for Weibo Movie Reviews [D]. Kunming: Yunnan Finance University, 2014.

[13] Xu Linhong, Lin Hongfei, Pan Yu, et al. Constructing the Affective Lexicon Ontology [J]. Journal of the China Society for Scientific and Technical Information, 2008, 27(2): 180-185.

[14] Virtual Learning Team of Wuhan University. ROST CM6 [EB/OL]. [2015-03-05]. http://download.csdn.net/download/sdaqdahai/5488041.

[15] HIT-CIR Tongyici Cilin (Extended) [EB/OL]. [2015-03-05]. http://www.datatang.com/data/42306/.

[16] HowNet. HowNet Sentiment Analysis Using Word Set: Beta [EB/OL]. [2015-03-03]. http://www.keenage.com/download/sentiment.rar.

[17] Chinese Emotion Words Dictionary (NTUS) [EB/OL]. [2015-03-08]. http://www.datatang.com/data/44317.

[18] Ji Souke GooSeeker Web Spiders [EB/OL]. [2015-09-05]. http://www.gooseeker.com/pro/product.html.

[19] Zhu Yanlan, Min Jin, Zhou Yaqian, et al. Semantic Orientation Computing Based on HowNet [J]. Journal of Chinese Information Processing, 2006, 20(1): 14-20.

[20] Du Rui. Rearch on Sentiment Classification for Chinese Microblog Text [D]. Changsha: Hunan University of Technology, 2014.

Author Contributions

Guo Shunli: Data acquisition and analysis, experiment implementation, manuscript drafting.
Zhang Xiangxian: Research concept, study design, final manuscript revision.

Conflict of Interest Statement

All authors declare no conflict of interest.

Supporting Data

Supporting data is self-archived by the authors and available upon request via email: ssguoshunli@sina.com.

[1] Guo Shunli, Zhang Xiangxian. 8500-book review dataset.rar. Over 255,000 user book reviews on 8,500 books.
[2] Guo Shunli, Zhang Xiangxian. Chinese book review word set.rar. 881 sentiment words.
[3] Guo Shunli, Zhang Xiangxian. Manual sentiment word classification results.rar. Manual classification of 690 sentiment words.
[4] Guo Shunli, Zhang Xiangxian. Sentiment word SO-PMI classification results.rar. SO-PMI classification results for 690 sentiment words.
[5] Guo Shunli, Zhang Xiangxian. Improved SO-PMI classification results.rar. Improved SO-PMI classification results for 690 sentiment words.
[6] Guo Shunli, Zhang Xiangxian. 100-book corpus.rar. 5,000 reviews on 100 books.
[7] Guo Shunli, Zhang Xiangxian. 100-book review manual classification results.rar. Manual classification of 5,000 reviews on 100 books.

[8] Guo Shunli, Zhang Xiangxian. 100-book review SO-PMI classification results.rar. SO-PMI classification results for 5,000 reviews on 100 books.
[9] Guo Shunli, Zhang Xiangxian. 100-book improved SO-PMI classification results.rar. Improved SO-PMI classification results for 5,000 reviews on 100 books.

Received: September 11, 2015
Revised: October 21, 2015

Note: Figure translations are in progress. See original paper for figures.