Relationship between Reviews Polarities, Helpfulness, Stars and Sales Rankings of Products: A Case Study in Books

Qingqing Zhou[1, 2] and Chengzhi Zhang[1, 3, *]

[1] Department of Information Management, Nanjing University of Science and Technology, Nanjing, China
[2] Alibaba Research Center for Complex Sciences, Hangzhou Normal University, Hangzhou, China
[3] Jiangsu Key Laboratory of Data Engineering and Knowledge Service (Nanjing University), Nanjing, China
breeze7zhou@gmail.com, zhangcz@njust.edu.cn

Abstract. To help customers, especially the customers without explicit purchasing motivation, to obtain valuable information of products via E-commerce websites, it is useful to predict sales rankings of the products. This paper focuses on this problem by finding relationship between reviews, star level and sales rankings of products. We combine various factors with the information of helpfulness and conducting correlation analysis between sales rankings and our combinations to find the most correlative combinations, namely the optimal combinations. We use three domains of books from Amazon.cn to conduct experiments. The main findings show that helpfulness is really useful to predict book sales rankings. Different domains of books have different optimal combinations. In addition, in consideration of helpfulness, the combination of number of positive reviews, score of review stars and score of frequent aspects is the most correlative combination. In this paper, although reviews on Amazon.cn are written in Chinese, our method is language independent.

Keywords: "sales ranking", "sentiment analysis", "helpfulness of reviews".

1 Introduction

With the development of E-commerce, people are more likely to buy products online. It is time-consuming for customers to understand products more deeply and choose the favorite ones. For customers who have no explicit purchasing motivation, sales ranking is a good index for their choices. However, there only partly products are in the list of sales ranking, new products are always out of it.

Many existing researches use reviews of products to predict sales rankings, few of them combine reviews and stars of the products together [1]. In this paper, we present

* Corresponding author.

H.-Y. Huang et al. (Eds.): SMP 2014, CCIS 489, pp. 175–186, 2014.

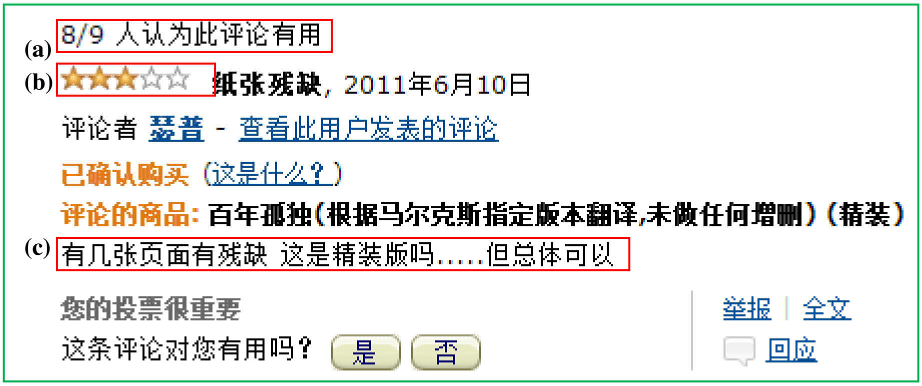

Fig. 1. Helpfulness of reviews from www.amazon.cn

three main factor combination methods with the information of helpfulness and conduct correlation analysis with sales rankings of books, so as to find the most correlative combinations, namely the optimal combinations. Helpfulness means assessing the quality of reviews by other users, namely judge whether the reviews are effective for the purchase decision [2-3]. An example of the books 'One Hundred Years of Solitude' is shown in Figure 1. There are eight of nine people who think the review is helpful; this book has 3 stars and the user thinks this book is generally nice, but may not be hardcover.

We aim at finding an effective combination for all domains and the specific combination for every domain, so as to help customers to find their favorite books. According to the experimental results, we can get the following conclusions: firstly, **WH (with helpfulness)** conclusion, which means that helpfulness is really useful to predict sales rankings of books; Secondly, **DOC (domain optimal combination)** conclusion, which shows that different domains of books have different domain optimal combinations; lastly, **OOC (overall optimal combination)** conclusion, which means that in the consideration of helpfulness, the combination of number of positive reviews, score of review stars and score of frequent aspects is the most correlative combination. All of these findings might be valuable information for customers to make the effective purchasing decisions.

The remain or rest of this paper is organized as follows. Section 2 reviews related works. Data collection and annotation are introduced in Section 3. Section 4 presents our methodology. Experimental results are provided in Section 5. The last part is about the conclusion and the future work.

2 Related Works

Two types of works are related to our study: sentiment analysis and sales forecast.

Sentiment analysis is to identify the attitudes of users by mining reviews. In this paper we focus on document-level and aspect-level sentiment analysis. Document-level sentiment analysis is to predict whether the whole document expresses a positive sentiment or a negative one [4]. Many researches have been done by supervised [5-7] and unsupervised learning methods [8-9]. Rather than gathering isolated opinions about a whole item, users generally prefer to compare specific features of different products, so it is important to conduct fine-grained aspect-level sentiment analysis [10]. Methods for sentiment analysis at this level are various. Methods like LDA models, sentiment lexicons are often used for aspect-level sentiment analysis [11-12]. In this paper, we use statistical methods to conduct document-level sentiment analysis, and lexical affinity methods for aspect-level sentiment analysis.

There are also many related works about sales forecast. Chang & Lai proposed a hybrid system to combine the self-organizing map of neural network with case-based reasoning method, for sales forecast of new released books [13]. Tanaka used high correlations between short-term and long-term accumulated sales within similar products groups to present a new forecasting model for new-released products [14]. Bollen's results indicated that the accuracy of DJIA (Dow Jones Industrial Average) predictions can be significantly improved by the inclusion of specific public mood dimensions [15]. Lee et al. developed and compared the performance of three sales forecasting models for the forecasting of fresh food sales, and the research results reveal that Logistic Regression performs better than the other methods [16]. Yu et al. conducted a case study in the movie domain to predict sales performance by analyzing the large volume of online reviews [17].

In this paper, we propose three main factor combination methods and conduct correlation analysis with book sales rankings, so as to find relationships between reviews polarities, helpfulness, stars and sales rankings of products.

3 Data

3.1 Data Collection

We collected sales rankings of three domains of books in the first half of 2013 from Amazon, including Literature[1], Social Science[2] and Economic Management books[3]. We chose top 50 books of each domain to conduct analysis. In total, we have collected 92,595 book reviews, including 60,903 literature book reviews, 17,476 social science book reviews and 14,216 economic management book reviews. The corpora cover reviews, stars and helpfulness of the books. The detail information is shown in Table 1.

[1] http://www.amazon.cn/gp/feature.html/ref=br_lf_m_353738_pglink_1?ie=UTF8&docId=353738&plgroup=1&plpage=1

[2] http://www.amazon.cn/gp/feature.html/ref=br_lf_m_353748_pglink_1?ie=UTF8&docId=353748&plgroup=1&plpage=1

[3] http://www.amazon.cn/gp/feature.html/ref=br_lf_m_353758_pglink_1?ie=UTF8&docId=353758&plgroup=1&plpage=1

3.2 Data Annotation

In order to construct the training set, we tagged part of reviews manually. We have tagged 5, 000 reviews manually. Among them, 2, 500 reviews express a positive feeling towards the entity and 2, 500 reviews express a negative one. For the convenience and reliability of the further comparison, we conduct cross validation on the training set to test the performance. We employed SVM as the classifier. Specifically, we used the LibSVM[4] to conduct experiments with 5-fold cross validation and present evaluation results in Table 2. From Table 2 we can find that the performance of reviews annotation is excellent. Therefore, it is trustable to use it as training data to conduct sentiment analysis on the whole corpus.

Table 1. Samples of data collection

Domains	Books	Reviews	helpfulness	Stars
Literature	Insight	The whole book is like a novice worked out in a short period of time, bad writing, and unclear thinking.	45 / 48	1
Social Science	On China	Careful packaging. It is Content that matters. A good book.	3 / 3	5
Economic Management	Rich Dad, Poor Dad	Sorry. I know a lot of people like it, but I really don't love it.	0 / 1	2

Table 2. Cross-validation performance of the reviews annotation

Metrics	Recall	Precision	F1 value
Scores	0.9805	0.9756	0.9780

4 Methodology

4.1 Framework

We conducted correlation analysis of our combinations and sales rankings on book of three domains by combining the information of book reviews, review stars and review helpfulness. We proposed three main ranking schemes, each of them includes three parts: **without helpfulness**, which means that we would not take the information about helpfulness into consideration when we compute the book scores and sort them; **with helpfulness**, which means that the information about helpfulness would be taken into consideration; **product ranking**, we multiplied the book scores that we got from the above two steps and then sorted them. We conducted correlation analysis between our rankings and sales rankings, so as to find the optimal combinations. The details are shown in Figure 2.

[4] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

4.2 Factor Combinations

In order to carry out the correlation analysis, we proposed 7 factor combination methods, which can be divided into three categories: combination 1, 2 and 3. The details are show in Table 3, and the calculations of factors are shown in Table 4.

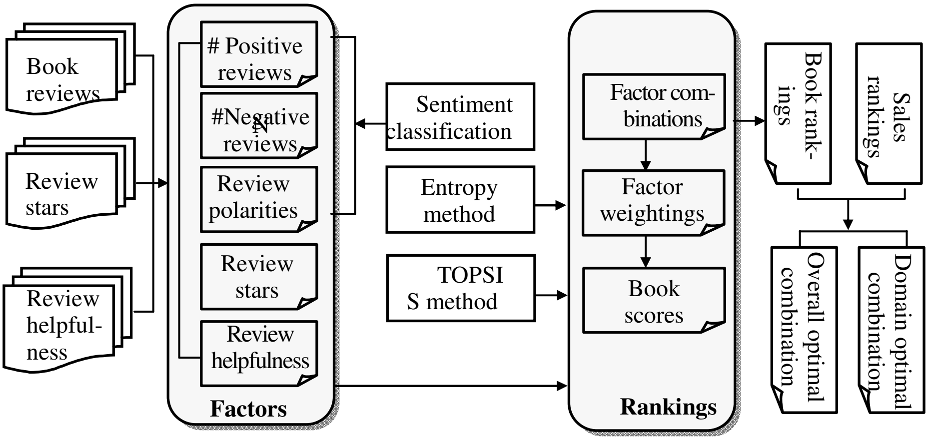

Fig. 2. Framework of optimal combinations selection

Table 3. Factor combination methods

Factors	Combination 1	Combination 2			Combination 3		
		2(a)	2(b)	2(c)	3(a)	3(b)	3(c)
#Positive reviews	○	○		○			○
#Negative reviews	○	○	○			○	
Score of review sentiment polarities	○		○	○	○		
Score of review stars	○	○	○	○	○	○	○
Score of frequent aspects	○	○	○	○	○	○	○

Table 4. Calculation of factors

Factors	Without helpfulness	With helpfulness
Score of review sentiment polarities	$\mathrm{scorep}(B) = \sum_{i=1}^{N} sp(i)/N$	$\mathrm{scorep}(B) = \sum_{i=1}^{N} sp(i) * he(i)/N$
Score of review stars	$\mathrm{scores}(B) = \sum_{i=1}^{N} st(i)/N$	$\mathrm{scores}(B) = \sum_{i=1}^{N} st(i) * he(i)/N$
Score of frequent aspects	$\mathrm{scorea}(B) = \sum_{i=1}^{m} (\sum_{i=1}^{N} po(i)/n)/m$	$\mathrm{scorea}(B) = \sum_{i=1}^{m} (\sum_{i=1}^{N} po(i) * he(i)/n)/m$

In Table 4, scorep(B) means score of review sentiment polarity of book B, $sp(i)$ means the sentiment polarity of review i, if it is a positive review, $sp(i)$ equals to +1, else it equals to -1; N denotes the number of reviews of book B, $he(i)$ donates score of helpfulness of review i. scores(B) donates score of review stars of book B; $st(i)$ means the star of review i, it ranges from 1 to 5. scorea(B) means score of aspects of book B; m denotes the number of frequent aspects, n denotes the review number of aspects A, $po(i)$ means the aspect sentiment classification in review i, if it is a positive review, $po(i)$ equal to 1, else it equal to -1.

4.3 Key Technologies

4.3.1 Sentiment Classification

For document-level sentiment classification, we used linear SVM as the classification model. In the preprocessing step, we chose CHI as feature selection method and TF-IDF as feature weighting method.

For aspect-level sentiment classification, we extract aspects of products by LDA method. Specifically, we use gensim [5] to identify frequent aspects. For aspect sentiment classification, sentiment polarity of aspect A in a review can be calculated via formula (1) [11].

$$\text{Score(A)} = \sum_{i=1}^{n} \frac{w_i.SO}{dis(w_i, A)} \tag{1}$$

where w_i denotes a sentiment word, n means number of sentiment words in review s, and $dis(w_i, A)$ denotes the distance between aspect A and sentiment word w_i. $w_i.SO$ is the sentiment score of the word w_i. If word w_i is a positive word, $w_i.SO$ equals to +1, else it equals to -1. If $Score(A) > 0$, the sentiment polarity of aspect A in the review s is positive, else it is negative.

4.3.2 Factor Weighting Calculation

We use the entropy method to calculate factor weightings [18].

(1) Normalization

We calculate the proportion of object i in factor j, it is computed by Eq.(2)

$$P_{ij} = X_{ij}/\sum_{i=1}^{n} X_{ij}, (i = 1,2, \dots, n, j = 1,2, \dots, m \tag{2}$$

where, X_{ij} denotes value of object i in factor j; n means the numbers of books (here, it equal to 50, the same below); m means the numbers of factors.

(2) Factors entropies

$$e_j = -\frac{1}{\ln(n)} \sum_{i=1}^{n} P_{ij} \ln(P_{ij}) \tag{3}$$

where, e_j denotes entropy of factor j.

(3) Factor weightings

$$w_j = \frac{1-e_j}{m-\sum_{j=1}^{m} e_j} \tag{4}$$

where, w_j denotes weighting of factor j; m means the numbers of factors.

---

[5] http://radimrehurek.com/gensim/

4.2.3 Book Score Calculation

We use the TOPSIS method to calculate book scores [19].

(1) Weighted factors

$$P_{ij} = w_j * p_{ij} \tag{5}$$

where, P_{ij} denotes value of weighted factors j of object i; p_{ij} means proportion of object i in factor j; w_j means weighting of factor j.

(2) Identification of ideal points

$$\mathrm{PIP}_j = \max(P_{ij}),\ (\mathrm{i} = 1,2,\ldots,\mathrm{n}) \tag{6}$$
$$\mathrm{NIP}_j = \min(P_{ij}),\ (\mathrm{i} = 1,2,\ldots,\mathrm{n}) \tag{7}$$

where, PIP_j denotes positive ideal point of factor j; n means the numbers of books (it equals to 50, the same below); NIP_j denotes negative ideal point of factor j.

(3) Distances of each book to the positive and negative ideal points

$$DP_i = \sqrt{\sum_{j=1}^{m}(P_{ij} - PIP_j)^2},\ (i = 1,2,\ldots,n) \tag{8}$$

$$DN_i = \sqrt{\sum_{j=1}^{m}(P_{ij} - NIP_j)^2},\ (i = 1,2,\ldots,n) \tag{9}$$

where, DP_i denotes distances of book i to the positive ideal points; NP_i denotes distances of book i to the negative ideal points; *m* means the numbers of factors.

(4) Score of each book

$$\mathrm{score}(\mathrm{b}_i) = NP_i/(DP_i + NP_i),\ (i = 1,2,\ldots,n) \tag{10}$$

where, $\mathrm{score}(\mathrm{b_i})$ denotes score of book i.

5 Experiments

5.1 Overall Optimal Combination

5.1.1 Correlation Analysis on Combination 1

The results of correlation analysis on the combination 1 are shown in Table 5. From Table 5 we can find that, for Literature and Social Science, sales rankings and all the three rankings have significant correlations at the level of 0.01 (bilateral). Among them with helpfulness rankings have the biggest correlation coefficients. It means that this kind of ranking is more useful to predict sales rankings. However, for Economic Management, there is no significant correlation between sales ranking and our three rankings. All these analyses above show that combination 1 is not useful enough for all domains.

5.1.2 Correlation Analysis on Combination 2

The results of correlation analysis on combination 2(a) are shown in Table 6. The correlation results in Table 6 are similar to combination 1. So we can get the conclusion that combination 2(a) is not useful enough for all domains.

The results of correlation analysis on combination 2(b) are shown in Table 7. From Table 7 we can find that, for Literature and Social Science, sales rankings and all the three rankings have significant correlations at the level of 0.01 (bilateral). However, for Economic Management, there is no significant correlation between sales ranking and our three rankings. All these analyses above show that combination 2(b) is not useful enough for all domains.

The results of correlation analysis on combination 2(c) are shown in Table 8. From Table 8 we can find that, the correlation results are similar to combination 2(b). So combination 2(c) is not useful enough for all domains.

Table 5. Correlation analysis on combination one

Domains	Without Helpfulness	With Helpfulness	Product
Literature	**0.348****	**0.372****	**0.360****
Social Science	**0.381****	**0.391****	**0.389****
Economic Management	0.183	0.183	0.183

Table 6. Correlation analysis on combination 2(a)

Domains	Without Helpfulness	With Helpfulness	Product
Literature	**0.303****	**0.372****	**0.365****
Social Science	**0.372****	**0.392****	**0.382****
Economic Management	0.197	0.182	0.191

Table 7. Correlation analysis on combination 2(b)

Domains	Without Helpfulness	With Helpfulness	Product
Literature	**0.332***	**0.282***	**0.335***
Social Science	**0.361****	**0.372****	**0.369****
Economic Management	0.237	0.207	0.225

Table 8. Correlation analysis on combination 2(c)

Domains	Without Helpfulness	With Helpfulness	Product
Literature	**0.405****	**0.382****	**0.389****
Social Science	**0.384****	**0.399****	**.395****
Economic Management	0.241	0.223	0.224

5.1.3 Correlation Analysis on Combination 3

The results of correlation analysis on combination 3(a) are shown in Table 9. From Table 9 we can find that, for Literature and Social Science, sales rankings and last two rankings have significant correlations at the level of 0.01 (bilateral). However, for

Economic Management, there is no significant correlation between sales ranking and our three rankings. All these analyses above show that combination 3(a) is not useful enough for all domains.

The results of correlation analysis on combination 3(b) are shown in Table 10. From Table 10 we can find that, for Literature, only with helpfulness ranking and sales ranking have a significant correlation; For Social Science, sales ranking and all the three rankings have significant correlations at the level of 0.01 (bilateral), among them with helpfulness ranking have the biggest correlation coefficient. However, for Economic Management, there is no significant correlation between sales ranking and our three rankings. All these analyses above show that combination 3(b) is not useful enough for all domains.

The results of correlation analysis on combination 3(c) are shown in Table 11. From Table 11 we can find that, for Literature, sales ranking and all the three rankings have significant correlations at the level of 0.01 (bilateral); For Social Science, sales ranking and last two rankings have significant correlations, and with helpfulness ranking have bigger correlation coefficient; For Economic Management, only product ranking and sales ranking have significant correlation. All the analysis above shows that product ranking in combination 3(c) is useful enough for all domains.

Table 9. Correlation analysis on combination 3(a)

Domains	Without Helpfulness	With Helpfulness	Product
Literature	0.256	**0.345****	**0.448****
Social Science	0.104	**0.384****	**0.328****
Economic Management	0.083	0.093	0.073

Table 10. correlation analysis on combination 3(b)

Domains	Without Helpfulness	With Helpfulness	Product
Literature	0.256	**0.372****	0.274
Social Science	**0.354****	**0.374****	**0.371****
Economic Management	0.236	-0.030	0.141

Table 11. Correlation analysis on combination 3(c)

Domains	Without Helpfulness	With Helpfulness	Product
Literature	**0.385****	**0.372****	**0.372****
Social Science	0.181	**0.401****	**0.372****
Economic Management	0.240	0.240	**0.368****

From the analysis above, we can draw the **OOC (overall optimal combination) conclusion** that combination 3(c) is the most useful combination, namely, in the consideration of helpfulness, the combination of numbers of positive reviews, score of review stars and frequent aspects is the most correlative combination.

5.2 Domain Optimal Combination

(1) Domain optimal combination of Literature books

We conducted correlation coefficients of three main combinations about Literature books and the results are shown in Figures 3. From Figure 3 we can find that the biggest correlation coefficient belongs to product ranking in combination 3(a), followed by combination 2(c) and 3(c). In addition, with helpfulness rankings are the highest of our proposed rankings in three of the combinations and product rankings are the highest in two of the combinations, which means that the information of helpfulness is useful to predict sales rankings of Literature books.

(2) Domain optimal combination of Social Science books

The results of correlation coefficients about Social Science books are shown in Figure 4. From Figure 4 we can find that the biggest correlation coefficient belongs to with helpfulness ranking in combination 3(c), followed by 2(c) and 2(a). With helpfulness rankings are the highest of our three rankings in all of the combinations, which means that helpfulness is useful to predict sales rankings of Social Science books.

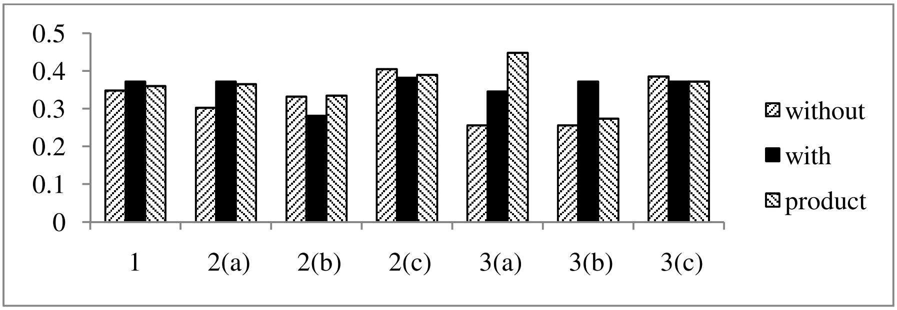

Fig. 3. Correlation coefficients of Literature books

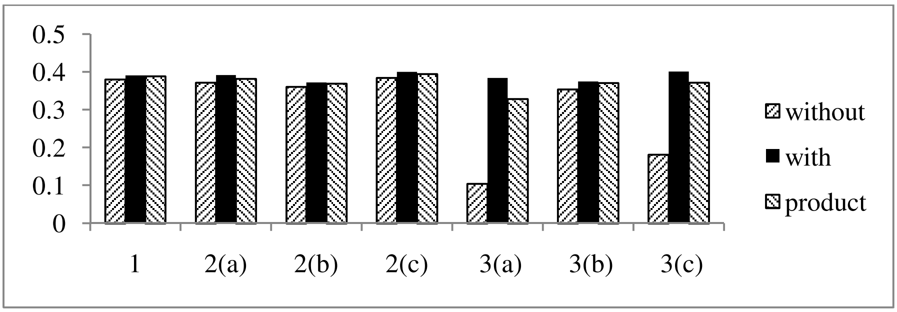

Fig. 4. Correlation coefficients of Social Science books

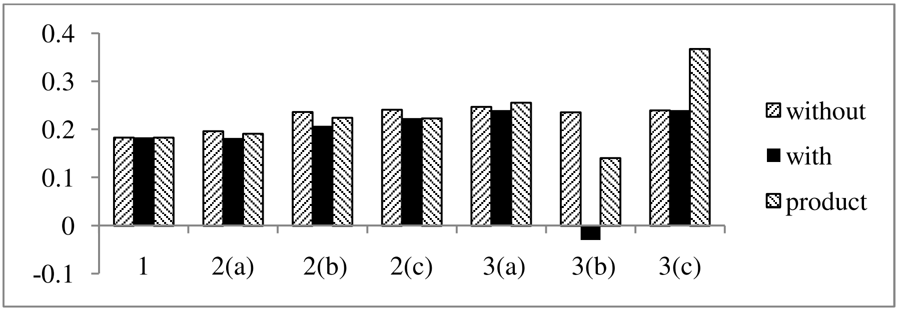

Fig. 5. Correlation coefficients of Social Science books

(3) Domain optimal combination of Economic Management books

For Economic Management, only product ranking in combination 3(c) has significant correlation with sales ranking, which also proved that helpfulness is useful to help predict sales rankings of Economic Management books.

From the analysis above, we can draw **DOC (domain optimal combination) conclusion** that different domains of books have different domain optimal combinations. For Literature books, the domain optimal combination is combination 3(a), while for Social Science and Economic Management books, combination 3(c) is the domain optimal combination. We can draw **WH (with helpfulness) conclusion** that helpfulness is really useful to predict sales rankings of books.

6 Conclusions and Future Works

In this paper, we proposed three main factor combination methods and chose the optimal ones via correlation analyses between our combinations and book sales rankings. Three main conclusions can be drawn according to our above mentioned analysis:

(1) **WH conclusion**: the information of helpfulness is really useful to predict or help predict sales rankings of books.
(2) **DOC conclusion**: different domains of books have different domain optimal combinations.
(3) **OOC conclusion**: in the consideration of helpfulness, the combination of numbers of positive reviews, score of review stars and score of frequent aspects is the most correlative combination.

The data in this paper is in Chinese, however our method for classification and correlation analysis is language independent. According to the three conclusions, we may predict sales rankings of books and provide effective purchasing suggestions for customers. In the future works, we will consider more languages of book and more types of products in the future. In addition, we will filter the untrusted reviews more efficiently.

Acknowledgments. This work is supported by Major Projects of National Social Science Fund (13&ZD174), National Social Science Fund Project (No.14BTQ033) and the Opening Foundation of Alibaba Research Center for Complex Sciences, Hangzhou Normal University (No. PD12001003002003).

References

1. Dellarocas, C., Zhang, X.M., Awad, N.F.: Exploring the value of online product reviews in forecasting sales: The case of motion pictures. Journal of Interactive Marketing 21(4), 23–45 (2007)
2. Yin, D., Bond, S.D., Zhang, H.: Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. Mis Quarterly 38(2), 539–560 (2014)
3. Liu, J., Cao, Y., Lin, C.Y., et al.: Low-Quality Product Review Detection in Opinion Summarization. In: Proceedings of EMNLP-CoNLL, pp. 334–342 (2007)
4. Liu, B.: Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies 5(1), 1–167 (2012)
5. Mullen, T., Collier, N.: Sentiment Analysis using Support Vector Machines with Diverse Information Sources. In: Proceedings of EMNLP, pp. 412–418 (2004)
6. Xia, R., Zong, C., Li, S.: Ensemble of feature sets and classification algorithms for sentiment classification. Information Sciences 181(6), 1138–1152 (2011)
7. Li, S., Lee, S.Y.M., Chen, Y., Huang, C.-R., Zhou, G.: Sentiment classification and polarity shifting. In: Proceedings of the 23rd International Conference on Computational Linguistics, pp. 635–643 (2010)
8. Turney, P.D.: Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 417–424 (2002)
9. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Computational Linguistics 37(2), 267–307 (2011)
10. Cambria, E., Hussain, A., Havasi, C., Eckl, C., Munro, J.: Towards crowd validation of the UK national health service. In: Proceedings of WebSci, pp. 1–5 (2010)
11. Ding, X., Liu, B., Yu, P.S.: A holistic lexicon-based approach to opinion mining. In: Proceedings of the International Conference on Web Search and Data Mining, pp. 231–240 (2008)
12. Moghaddam, S., Ester, M.: On the design of LDA models for aspect-based opinion mining. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, pp. 803–812 (2012)
13. Chang, P.C., Lai, C.Y.: A hybrid system combining self-organizing maps with case-based reasoning in wholesaler's new-release book forecasting. Expert Systems with Applications 29(1), 183–192 (2005)
14. Tanaka, K.: A sales forecasting model for new-released and nonlinear sales trend products. Expert Systems with Applications 37(11), 7387–7393 (2010)
15. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. Journal of Computational Science 2(1), 1–8 (2011)
16. Lee, W.I., Chen, C.W., Chen, K.H., et al.: A comparative study on the forecast of fresh food sales using logistic regression, moving average and BPNN methods. Journal of Marine Science and Technology 20(2), 142–152 (2012)
17. Yu, X., Liu, Y., Huang, X., et al.: Mining online reviews for predicting sales performance: A case study in the movie domain. IEEE Transactions on Knowledge and Data Engineering 24(4), 720–734 (2012)
18. Hongzhan, N., Lü Pan, Q.Y., Yao, X.: Comprehensive fuzzy evaluation for transmission network planning scheme based on entropy weight method. Power System Technology 33(11), 60–64 (2009)
19. Hwang, C.L., Yoon, K.: Multiple Attribute Decision Making: Methods and Applications, A State of the Art Survey. Springer, New York (1981)