

# Mining Online Customer Reviews for Products Aspect-Based Ranking

Chonghui Guo<sup>1,2</sup>(✉), Zhonglian Du<sup>1</sup>, and Xinyue Kou<sup>1</sup>

<sup>1</sup> Institute of Systems Engineering, Dalian University of Technology,  
Dalian, China

dlutguo@dlut.edu.cn, duzhonglian@mail.dlut.edu.cn

<sup>2</sup> State Key Laboratory of Software Architecture (Neusoft Corporation),  
Shenyang, China

**Abstract.** Massive online reviews contain a lot of useful information that can not only provide purchasing decision support for consumers, but also allow producers and suppliers to understand the competitive market. This paper proposes a new aspect-based online reviews mining method, which combines both textual data and numerical data. Firstly, the probability distribution of topics and words is constructed by LDA topic model. With word cloud images, the keywords are visualized and corresponding relationship between LDA topics and product reviews is analyzed. The weight of each aspect is calculated based on the probability distribution of documents and topics. Then, the dictionary-based approach is used to calculate the objective sentiment values of the product. The subjective sentiment tendency from different consumers because of their different individual needs are also taken into consideration. Finally, the directed graph model is constructed and the importance of each node is calculated by improved PageRank algorithm. The experimental results illustrate the feasibility of proposed mining method, which not only makes full use of massive online reviews, but also considers individual needs of consumers. It provides a new research idea for online customer review mining and personalized recommendation.

**Keywords:** Online review mining · LDA topic model · Improved PageRank algorithm · Personalized recommendation

## 1 Introduction

With the rapid development of web 2.0, the role of Internet users has changed greatly from previous information recipients to information producers. Specifically, the development of online shopping platforms has attracted more and more consumers to share their shopping experience and product feedback on the Internet. This information is presented in forms of semi-structured or unstructured data, such as text, numbers, images, etc., providing more objective references for other potential consumers. A survey found that 63% of online shoppers consistently search and read reviews before making a purchasing decision. Among them, 64% spend 10 min reading reviews, whereas 33% spend 30 min or more. Although many scholars have adopted content-based recommendation, collaborative filtering and knowledge-based recommendation methods for

user recommendation research, there are relatively few user recommendations based on online review information [1, 2]. At the very beginning, scholars used numerical scoring information from online reviews to sort products. Mary et al. selected 8 million product reviews and 1.5 million merchant reviews to evaluate quality and rank different products. She found that the average value has a better performance [3]. However, due to fuzziness of the numerical scores, it can't well reflect users' sentiment expression of products' unique characteristics. That is, products with same scores can't be distinguished from each other [4]. As valuable user-generated contents, review texts can better reflect users' emotional expression. Khan et al. proposed a new text sentiment analysis method (eSAP) for decision support systems and carried out experiments with data sets from seven different fields to verify the effectiveness of the method [5]. In order to solve the problem of unbalanced sentiment classification in Chinese product review, Tian et al. proposed a new method based on topic sentence to improve the accuracy of classification [6].

However, text reviews can't cover comprehensive information, most of which are written only based on a certain aspect or some aspects of products. Recently, models which integrate numerical scores and review texts have attracted a lot of attention [7, 8]. From the aspect of product evaluation method, the fusion of different data is pretty simple and the weight is mostly determined by subjective decision or weighted average method. Yang et al. proposed a method based on multi-source heterogeneous information. However, only online reviews are taken into consideration and subjective opinions of consumers are neglected while calculating weights. Zhang et al. proposed a method to calculate weights based on the validity of comments and importance of time, of which the former one is decided by the votes from other consumers and the latter one is decided by the time this review was published. However, while calculating the overall score of the product, the same weights are given to different aspects which cannot reflect their different importance for consumers [10]. Some other methods, such as intuitionistic fuzzy set theory, TOPSIS are also used to do products ranking [11, 12].

There are different types of online reviews, such as numeric ratings and text descriptions. Such heterogeneous information brings customers more complexity to make purchase decisions. Zhang et al. used the directed graph model to evaluate relative advantages of products, but there was no analysis of node weights in the network [13]. The node represents comprehensive sentiment value of the product. The direction of the edge represents relative advantage between two products. The weight of the side represents the value of relative advantage.

In this paper, a directed graph model is proposed to integrate such rich and heterogeneous information. The modified PageRank method is used to calculate the value of each node and to rank different products. The rest of this paper is organized as follows. In Sect. 2, a method to mine and integrate text and numerical information is proposed. A directed graph model is constructed and the importance of each node is calculated by improved PageRank algorithm. In Sect. 3, an experimental study based on automobile online reviews is taken to illustrate the feasibility of proposed method. Finally, in Sect. 4, contribution of this paper and future work is summarized.

## 2 Research Methods

In order to make full use of user-generated contents for more effective ranking of different products, this paper selects aspect-based online text reviews and data information to do analysis. The directed graph model is used for information fusion, and improved PageRank method is deployed to find the node value, which refers to the final scores of different products. The research framework is shown in Fig. 1.

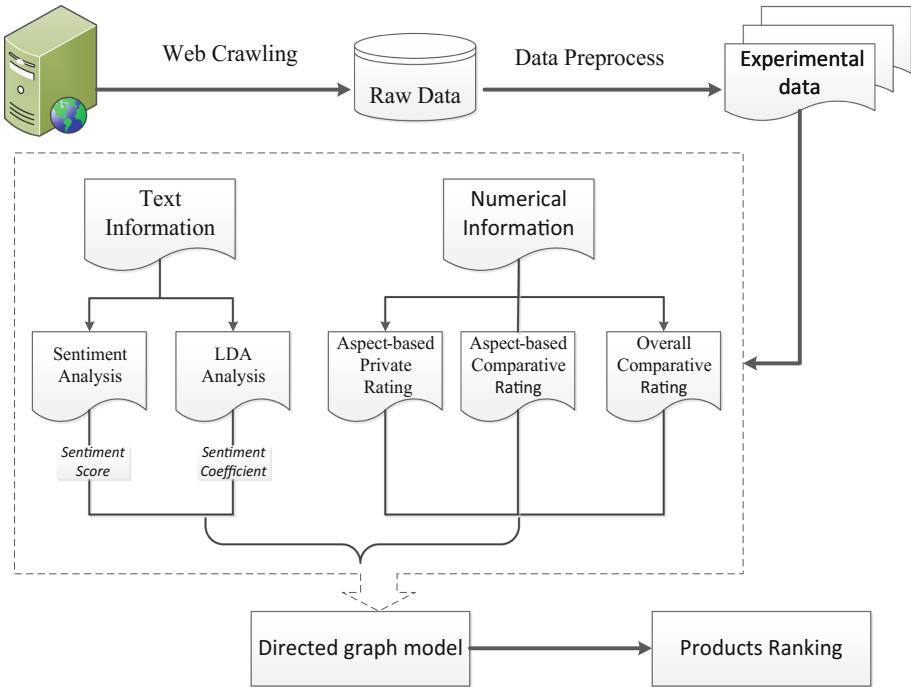


Fig. 1. Research framework

### 2.1 Text Information Processing Method

In this section, we first formulate the problem of aspect-based products ranking. Then we describe the method of sentiment intensity calculation and aspect-based weight determination.

#### 2.1.1 Sentiment Intensity Calculation Method

A large number of online aspect-based reviews are crawled from the third-party website. The set  $P = \{P_1, P_2, \dots, P_i, \dots, P_I\}$  denotes the set of  $I$  alternative products, where  $P_i$  is the  $i$ th alternative product. The original corpus is denoted as  $C_0 = \{r_{i1}, r_{i2}, \dots, r_{ij}, \dots, r_{iN}\}$ , where  $r_{ij}$  is the  $j$ th online review of alternative product  $P_i$ . Firstly, the original

corpus needs to be preprocessed, including word segmentation and stop words removing. The result is a word set denoted as  $W_{ij} = \{w_{ij1}, w_{ij2}, \dots, w_{ijm}, \dots, w_{ijM}\}$ , where  $w_{ijm}$  means the  $m$ th word segmentation result in the  $j$ th online review of alternative product  $P_i$ .

In this paper, the sentiment word lexicon established by the information retrieval team of Dalian University of Technology is used as the basis of text analysis [14]. The sentiment word lexicon is denoted by  $SW_0 = \{sw_1, sw_2, \dots, sw_q, \dots, sw_Q\}$ , where  $sw_q$  denotes the  $q$ th sentiment word. The sentiment intensity of sentiment word is denoted by  $SV_0 = \{sv_1, sv_2, \dots, sv_q, \dots, sv_Q\}$ , where  $sv_q$  denotes the sentiment intensity of the  $q$ th sentiment word in the lexicon. Let  $W_{ij}^k = \{W_{ij}^1, W_{ij}^2, \dots, W_{ij}^k, \dots, W_{ij}^n\}$ , where  $W_{ij}^k$  denotes the word segmentation result of the  $j$ th online review concerning the  $k$ th aspect of product  $P_i$ . Let  $S_{ij}^k = \{S_{ij}^1, S_{ij}^2, \dots, S_{ij}^k, \dots, S_{ij}^n\}$ , where  $S_{ij}^k$  denotes the sentiment intensity of the  $k$ th aspect in the  $j$ th online review of alternative product  $P_i$ . The sentiment intensity calculation process is given below.

Step 1: If the  $i$ th word of  $W_{ij}^k$  appears in  $SW_0$  and its corresponding segment word is  $sw_i$ , then the sentiment intensity and sentiment polarity are denoted as  $sv_i$  and  $sp_i$  respectively.

Step 2: Take the location of  $W_{ij}^{kn}$  as center and two characters as window size to determine whether there is any negative words within this range. If there is not, the modified sentiment intensity is calculated by  $sv'_i - sv_i = 0$ ; otherwise, it is calculated by  $sv'_i + sv_i = 0$ .

Step 3: Determine the sentiment polarity  $sp_i$ . If  $sp_i = 1$ , which means the review is positive, then  $S_{ij}^k = S_{ij}^k + sv'_i$ ; otherwise,  $S_{ij}^k = S_{ij}^k - sv'_i$ .

In this paper, we assume that weights of reviews from different customers for the  $k$ th aspect of the  $i$ th product is equal, that is,  $S_{ij}^k$  and  $S_{ij+1}^k$  are of no individual differences. Therefore, the sentiment intensity of  $k$ th aspect about alternative product  $P_i$  can be calculated as:

$$S_i^k = \frac{1}{n} \sum_{j=1}^n S_{ij}^k \tag{1}$$

### 2.1.2 Aspect-Based Weight Determination Method

In order to obtain the overall sentiment intensity of product  $P_i$ , we need to calculate the weight of each aspect. LDA (Latent Dirichlet Allocation) is a typical topic model, and it is a generative directed graph model. It is mainly used to deal with discrete data, and has a wide range of applications in information retrieval, natural language processing and other fields.

The relationship of LDA variables is described in Fig. 2, where hollow circles represent hidden variables and solid circles represent observable variables. Only word frequency  $w_{t,n}$  of word  $n$  in document  $t$  is observable, it depends on topic of word  $n$  in

document  $t$   $Z_{t,n}$  and word frequency corresponding to topic  $k$   $\beta_k$ . Meanwhile,  $Z_{t,n}$  depends on the topic distribution  $\theta_t$  and  $\theta_t$  depends on the parameter  $\alpha$  of Dirichlet Allocation, while  $\beta_k$  depends on parameter  $\eta$ . Accordingly, the probability distribution of LDA is given as follows:

$$p(W, z, \beta, \theta|\alpha, \eta) = \prod_{t=1}^T p(\theta_t|\alpha) \prod_{i=1}^K p(\beta_k|\eta) (\prod_{n=1}^N P(w_{t,n}|z_{t,n}, \beta_k)P(z_{t,n}|\theta_t)) \quad (2)$$

where  $p(\theta_t|\alpha) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_k \theta_{t,k}^{\alpha_k-1}$  and  $p(\beta_k|\eta)$  usually obey  $K$ -dimensional and  $N$ -dimensional Dirichlet Allocation with parameters  $\alpha$  and  $\eta$ . The aspect-based weight calculation process is given below.

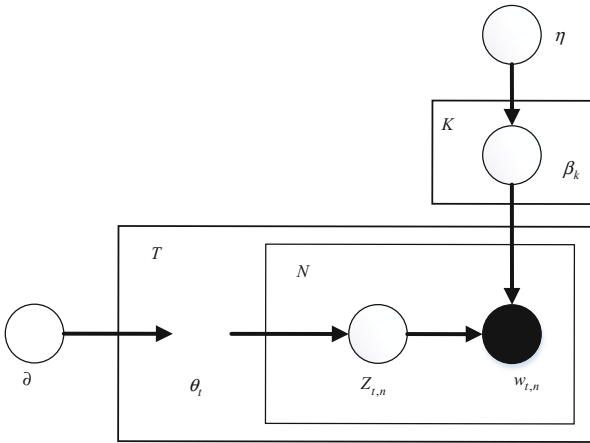


Fig. 2. LDA model

Step 1: With prior knowledge, parameters  $K$ , number of topics, can be determined. The probability matrix of subject and word can be obtained from  $\beta_k$ , based on which the word cloud for each subject can be generated. Using text data visualization, we can display each subject more intuitively and summarize realistic meaning of each subject with consideration of realistic background;

Step 2: From  $\theta_t$ , we can know the probability matrix of document and topic, and then determine the proportion  $\alpha\theta_t$  of each topic in a document. Let  $\alpha_k$  be the weight of each topic:

$$\alpha_k = \frac{1}{T} \sum_{t=1}^T \alpha\theta_t \quad (3)$$

Step 3: Compute comprehensive sentiment value of a product  $TS_i$ . It is consisted of two parts: one is the objective sentiment value based on consumer's product reviews, i.e.,

$$S_i = \sum_{k=1}^k \alpha_k S_i^k \quad (4)$$

The other one is subjective sentiment value based on the intention of consumers who have the intension to purchase, i.e.,

$$S'_i = \sum_{k=1}^k \beta_k S_i^k \quad (5)$$

Where  $\beta_k$  represents personal preference for all aspects of the product. Therefore, the comprehensive sentiment value of a product  $P_i$  can be computed as:

$$TS_i = \delta S_i + (1 - \delta) S'_i \quad (6)$$

## 2.2 Numerical Information Processing Method

Numerical information is adopted by many scholars because of its easy access and intuitive comment form. Numerical value represents the satisfaction degree of the consumer to the product. This paper will analyze it from two aspects: numerical rating based on itself and the numerical rating based on the comparison.

### a. The numerical rating based on itself

In this paper, we assume the weight from different customers for the  $k$ th aspect of the product is equal.  $R_{ij}^k$  denotes the numerical rating concerning  $j$ th aspect in the  $k$ th online review of alternative product  $P_i$ . So, the comprehensive numerical rating of product  $P_i$  concerning the  $j$ th aspect can be computed as:

$$R_{ij} = \frac{1}{k} \sum_{k=1}^K R_{ij}^k \quad (7)$$

In summary, the comprehensive numerical rating of product  $P_i$  can be computed based on the topic weight  $\alpha_k$ .

$$TR_i = \sum_{k=1}^K \alpha_k R_{ij} \quad (8)$$

### b. The numerical rating based on the comparison

Nowadays, many third-party platforms have launched comparison plates for homogeneous products. Let  $CR_{ij}$  be comparative rating concerning the  $j$ th aspect of

alternative product  $P_i$ . Let  $PR_i$  be the superiority of product  $P_i$  over other same level products. If  $P_i$  is superior to other products in the  $j$ th aspect, then  $PR_{ij} > 0$ . If it is inferior to other products,  $PR_{ij} < 0$ . If they are same in the  $j$ th aspect,  $PR_{ij} = 0$ . So, the relative comparative superiority of product  $P_i$  over  $P_m$  can be computed as:

$$Q_{im} = \frac{1}{J} \frac{\sum_{j=1}^J CR_{ij}}{\sum_{j=1}^J CR_{mj}} \tag{9}$$

### 2.3 Information Fusion Network Construction Method

Construct a directed graph model  $G(V, E, Q^V, Q^E)$ . Node  $V$  denotes the product. Edge  $E$  denotes the directed connection between products, and  $Q^V$  denotes weight of node  $V$ , which refers to the comprehensive sentiment value  $TS_i$ .  $Q^E$  denotes weight of edge  $E$ , that is, the relative comparative superiority  $q_{im}$ , which can be computed as:

$$q_{im} = \frac{|Q^i - Q^m|}{\min(Q^i, Q^m)} \tag{10}$$

If  $q_{im} > 1$ , the edge is directed from  $P_m$  to  $P_i$ . If  $q_{im} < 1$ , the direction is reversed. If  $q_{im} = 1$ , there is no connection between these two nodes.

There are many ways to calculate the importance of network nodes. The classical search engine page ranking algorithm PageRank is based on the idea that pages from high-quality web pages must be quality web pages [15]. Similarly, one product with comparative advantage from another high evaluation product will surely be of high

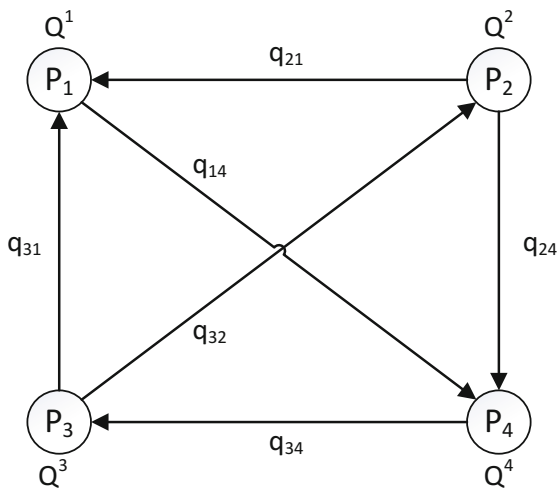


Fig. 3. Directed graph example

evaluation. We use improved PageRank algorithm to calculate importance of each node. Assume that there are  $n$  nodes in a directed graph,  $n = 1, 2, 3, \dots, N$ , the number of output edges of node  $n$  is denoted as  $L(n)$ . Therefore, the importance of node  $A$  can be computed as:

$$Q^A = Q^A + \frac{Q^i q^{iA}}{L(i)} \quad (11)$$

Where the node  $i$  means all nodes that have a directed chain to node  $A$ . Figure 3 shows the importance of the nodes. Therefore, the importance of the node 1 can be computed as:

$$Q^1 = Q^1 + \frac{Q^2 q^{21}}{2} + \frac{Q^3 q^{31}}{2} \quad (12)$$

### 3 Experimental Study

With the rapid development of China's economy, people's living standards have been improved significantly. Automobile has become a necessity in people's life. It is shown that the third-party consumer reviews platforms have a better reputation and recognition [16].

#### 3.1 Data Collection and Preprocessing

In order to verify the practicality and effectiveness of proposed method, two third-party consumer review platforms, autohome (<http://www.autohome.com.cn/>) and Pcauto Network (<http://www.pcauto.com.cn/>) are selected to do online reviews analysis. According to products' level and price, we choose compact SUV automobiles with prices between 100,000 and 300,000 yuan. We choose four brands automobiles, Mazda, Honda, Trumpchi and Roewe, and then use web crawler to obtain the four kinds of products' data. The detailed data is shown in Table 1.

**Table 1.** Data set

	Total number of aspect-based reviews	Total number of comprehensive reviews	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$
$P_1$	2338	307	297	280	287	298	301	294	291	290
$P_2$	2140	282	264	263	260	273	274	269	276	261
$P_3$	1384	179	175	171	174	172	172	174	176	170
$P_4$	2381	316	297	296	302	294	301	293	300	298

$P_1$ : Mazda CX-5 2015, 2.5L (Deluxe Edition automatic four-wheel drive);  $P_2$ : Trumpchi GS4 2016, (Deluxe Edition 235T G-DCT);  $P_3$ : Roewe RX5 2016 (30T automatic four-wheel drive version of the Internet);  $P_4$ : Honda CR-V 2015, 2.4L (Deluxe Edition two-wheel drive).  $f_1$ : Appearance;  $f_2$ : Car interior;  $f_3$ : Space;  $f_4$ : Configuration;  $f_5$ : Power;  $f_6$ : Cross-country;  $f_7$ : Oil consumption;  $f_8$ : comfort.



To preprocess the text data, we apply Stanford Parser to segment Chinese text. An external dictionary is added to the word segmentation to avoid ambiguities because of separation of field words. The external dictionary includes 2401 words selected from automobile field words in Sogou cell lexicon.

### 3.2 Data Experiment

The comprehensive reviews are decomposed into aspect-based review corpus, based on which sentiment analysis is carried out. The sentiment intensity values of product  $P_i$  based on different aspects  $S_{ij}^k$  are obtained. And then the sentiment value of product  $P_i$  towards the  $k$ th aspect  $S_i^k$  can be calculated by  $S_i^k = \frac{1}{n} \sum_{j=1}^n S_{ij}^k$ . The sentiment value based on aspect of the four kinds of products are calculated and the result is shown in Table 2.

**Table 2.** Aspect-based sentiment values of different products

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$
$P_1$	4.77	3.37	2.29	2.84	3.78	7.00	3.16	2.51
$P_2$	4.95	2.91	2.97	2.22	3.13	5.18	2.57	1.94
$P_3$	6.80	5.88	4.33	3.97	3.70	4.99	2.92	2.68
$P_4$	4.57	2.48	3.00	2.63	3.90	4.02	3.55	2.68

According to the method proposed in this paper, the preprocessed text review data is used to determine the weight based on LDA topic model. Take Mazda review data as an example. Given the  $K = 8$ , then the document and topic probability matrix  $\theta_i$  can be obtained. Meanwhile, with LDA topic model, the distribution of keywords under each topic can be determined and then value of word frequency  $f$  can be set. Let  $f = 40$ , 163 keywords are kept, from which a  $8 \times 163$  topic-keywords probability matrixes  $\beta_k$  can be generated.

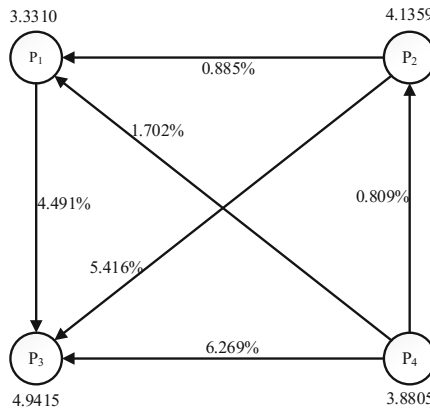
$$\beta_k = \begin{matrix} & & \text{Word1} & \text{Word2} & \dots & \text{Word163} \\ \text{Topic1} & \left[ \begin{array}{cccc} 0.0034 & 0.0146 & \dots & 0.0023 \\ 0.0043 & 0.0092 & \dots & 0.0022 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \text{Topic8} & \left[ \begin{array}{cccc} 0.0047 & 0.0104 & \dots & 0.0026 \end{array} \right. \end{array} \right.$$

With  $\beta_k$ , we can make word cloud to visualize display the subject words, from which we can get the realistic meaning of LDA topic in automobile reviews. By Eq. (4), the objective sentiment values of different products based on online reviews can be obtained. The results are shown in Table 3.

**Table 3.** Topic-product correspondence and sentiment value analysis

	Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7	Topic8	
$P_1$	$f_1$	$f_2$	$f_7$	$f_6$	$f_3$	$f_5$	$f_8$	$f_4$	
	$\alpha_k$	0.012	0.305	0.119	0.317	0.021	0.017	0.074	0.135
	$S_i^k$	4.772	2.071	3.160	4.997	2.292	3.788	2.508	2.844
	$S_i$	0.0573	0.6318	0.3760	1.5840	0.0481	0.0644	0.1856	0.3839
$P_2$	$f_3$	$f_6$	$f_4$	$f_8$	$f_1$	$f_7$	$f_2$	$f_5$	
	$\alpha_k$	0.088	0.041	0.031	0.18	0.048	0.246	0.338	0.028
	$S_i^k$	2.968	5.184	2.794	3.943	4.950	3.574	4.915	3.128
	$S_i$	0.2612	0.2126	0.0866	0.7098	0.2376	0.8793	1.6612	0.0876
$P_3$	$f_4$	$f_7$	$f_5$	$f_8$	$f_6$	$f_2$	$f_3$	$f_1$	
	$\alpha_k$	0.023	0.057	0.009	0.065	0.047	0.461	0.281	0.046
	$S_i^k$	3.972	2.916	3.704	2.682	4.994	5.883	4.330	6.804
	$S_i$	0.0914	0.1662	0.0333	0.1743	0.2347	2.7119	1.2166	0.3130
$P_4$	$f_1$	$f_6$	$f_2$	$f_4$	$f_3$	$f_7$	$f_5$	$f_8$	
	$\alpha_k$	0.025	0.102	0.326	0.241	0.017	0.164	0.044	0.081
	$S_i^k$	4.557	4.016	4.481	3.627	2.997	3.551	3.899	2.684
	$S_i$	0.1139	0.4096	1.4608	0.8740	0.0509	0.5823	0.1715	0.2174

Based on analysis in Sect. 2.3, we can construct the directed graph model after calculating the weights of each network node, the direction and weight of each edge. The result is shown in Fig. 4.

**Fig. 4.** Directed graph model of four products

Based on the improved PageRank algorithm, the final scores for different products can be obtained. By Eq. (10), the final scores of the four products are 3.3713, 4.1464, 5.2842 and 3.8805. So, the final ranking result is  $P_3 > P_2 > P_4 > P_1$ .

Competition car ranking information provided by Pacific automotive network shows that the actual ranking result is  $P_3 > P_1 > P_2 > P_4$ . It means the result calculated by proposed method in this paper is relatively consistent with the performance of each product in the competition ranking. However, the proposed method is not totally consistent with the actual result. It is because the text review information is taken into consideration in the method proposed in this paper, which also verifies the effectiveness of the proposed method.

## 4 Conclusion

In this paper, a new method of mining aspect-based online reviews is proposed. Firstly, different from other studies with single data, this paper uses both numerical data and review text data. Secondly, LDA topic model in text mining field is used to solve probability distribution of document-topic and topic-keyword, based on which the word cloud about the topic is obtained. Through the visualization of text data, each topic's content is shown. The virtual topic of LDA topic model is endowed with practical significance and the weights of aspects are calculated. With full use of massive online review knowledge, the customers' objective sentiment values are finally digged out.

At the same time, considering that different consumers have different preferences towards different aspects of products, their subjective sentiment orientation weights are taken into account. The probability of text and of personal preference are integrated to calculate the comprehensive sentiment value of the product, which makes the result more persuasive and reliable. The proposed algorithm of this paper provides a new theoretical method for aspect-based online text review mining. What's more, developing decision support systems for personalized recommendation is also of great practical significance.

**Acknowledgements.** This work was supported in part by the Natural Science Foundation of China [Grant Numbers 71771034, 71031002] and the Open Program of State Key Laboratory of Software Architecture [Item number SKLSAOP1703].

## References

1. Yang, X., Liang, C., Zhao, M., Wang, H., Ding, H., Liu, Y., Li, Y., Zhang, J.: Collaborative filtering-based recommendation of online social voting. *IEEE Trans. Comput. Soc. Syst.* 1–13 (2017)
2. Park, Y., Park, S., Jung, W., Lee, S.G.: Reversed CF: a fast collaborative filtering algorithm using a k-nearest neighbor graph. *Expert Syst. Appl.* **42**, 4022–4028 (2015)
3. Mary, M., Natalie, G., Zach, R.: Star quality: aggregating reviews to rank products and merchants. In: International Conference on Weblogs and Social Media, ICWSM 2010, Washington, D.C., USA, May 2010
4. Ghose, A., Ipeirotis, P.G.: Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics. *IEEE Trans. Knowl. Data Eng.* **23**, 1498–1512 (2011)

5. Khan, F.H., Qamar, U., Bashir, S.: eSAP: a decision support framework for enhanced sentiment analysis and polarity classification. *Inf. Sci.* **367–368**, 862–873 (2016)
6. Tian, F., Wu, F., Chao, K.M., Zheng, Q., Shah, N., Lan, T., Yue, J.: A topic sentence-based instance transfer method for imbalanced sentiment classification of Chinese product reviews. *Electron. Commer. Res. Appl.* **16**, 66–76 (2016)
7. Hu, N., Koh, N.S., Reddy, S.K.: Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. *Decis. Support Syst.* **57**, 42–53 (2014)
8. Najmi, E., Hashmi, K., Malik, Z., Rezgui, A., Khan, H.U.: CAPRA: a comprehensive approach to product ranking using customer reviews. *Computing* **97**, 843–867 (2015)
9. Yang, X., Yang, G., Wu, J.: Integrating rich and heterogeneous information to design a ranking system for multiple products. *Decis. Support Syst.* **84**, 117–133 (2016)
10. Zhang, K., Cheng, Y., Liao, W.K., Choudhary, A.: Mining millions of reviews: a technique to rank products based on importance of reviews. In: *International Conference on Electronic Commerce*, pp. 1–8 (2011)
11. Liu, Y., Bi, J.W., Fan, Z.P.: Ranking products through online reviews: a method based on sentiment analysis technique and intuitionistic fuzzy set theory. *Inf. Fusion* **36**, 149–161 (2017)
12. Chen, K., Kou, G., Shang, J., Chen, Y.: Visualizing market structure through online product reviews: integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches. *Electron. Commer. Res. Appl.* **14**, 58–74 (2015)
13. Zhang, Z., Guo, C., Goes, P.: Product comparison networks for competitive analysis of online word-of-mouth. *ACM Trans. Manag. Inf. Syst.* **3**, 1–22 (2013)
14. Xu, L.H., Lin, H.F., Pan, Y., et al.: The construction of emotional lexical ontology. *J. Inf. Sci.* **27**, 180–185 (2008)
15. Page, L.: The PageRank citation ranking: bringing order to the web. In: *Stanford Digital Libraries Working Paper*, vol. 9, pp. 1–14 (1998)
16. Gu, B., Park, J., Konana, P.: Research note: the impact of external word-of-mouth sources on retailer sales of high-involvement products. *Inf. Syst. Res.* **23**, 182–196 (2012)