MANAGEMENT SCIENCE AND ENGINEERING
Vol. **5**, No. **2**, **2011**, pp.10-15
www.cscanada.org

ISSN 1913-0341 [Print]
ISSN 1913-035X [Online]
www.cscanada.net

An Original Search Results Ranking Algorithm based on C2C Platform*

ZENG Ziming[1,*]

ZHANG Xinping[2]

Abstract: C2C E-Commerce has been growing at a very high speed in China in recent years. Consumers usually access commodities information with the help of search tools .Thus, how to rank the search results in a rational order is a crucial issue. In this paper, the factors influencing search results ranking are first analyzed, then an original search results ranking algorithm based on Vector Space Model (VSM) is proposed. An experiment is carried out to testify its validity. The experiment results show that consumers can enjoy a wonderful search experience with the presented algorithm.
Key words: C2C; Consumer Behavior; Results Ranking Algorithm; Vector Space Model (VSM)

INTRODUCTION

With the development of Internet, C2C E-Commerce has been growing at a very high speed in China in recent years. Consumers prefer to go shopping on the C2C platform such as Taobao.com and Paipai.com. C2C E-Commerce has been accepted by more and more people in various circles. One report published by B2B Research Center of China shows that the transaction volume of China C2C E-Commerce market will reach up to 388.3 billion RMB in 2011.

There is no doubt that Internet has provided different shopping experience in various ways to consumers, as there are much more benefits for consumers to purchase online. They can access more information than ever before with the help of search engine and other exploration tools on the Internet. However, when consumers purchase on C2C platform, consumers' behavior have many differences from traditional purchase activities. For example, the effect of negative online consumer reviews on product attitude is a crucial factor for consumers to make purchase decision. Furthermore, sales amount, postage and popularity are also very important factors that affect consumers' purchase decision.

When purchasing on C2C websites, almost all consumers access commodities information with the help of search engine. Search tools are very useful on C2C websites. However, there are tremendous amount search results returned by the search tools. Consumers tend to look at only the top part of the rank results list in order to find relevant information. Thus, how to rank the search results in a rational order is a crucial issue. The objective of this paper is to gain a better understanding of influence factors that affect

* **The work was supported by "the Fundamental Research Funds for the Central Universities" (No: 09ZZKY096), and supported by "Research on Information Aggregation and Visualization Modeling of new E-commerce Search Engine", MOE Project of Humanities and Social Science in Chinese University(No:08JC870011)，and also supported by "the Post-70s Scholars Academic Development Program of Wuhan University"**.
[1] School of Information Management, Wuhan University, Wuhan, 430072, China
† Corresponding Author. Email: zmzeng1977@163.com
[2] School of Information Management, Wuhan University, Wuhan, 430072, China
‡Received April 1, 2011; accepted May 2, 2011

consumers' purchase decision and propose an original search results ranking algorithm based on C2C platform.

The remainder of this paper is structured as follows. In section 2, we present a description of related works on consumers' online purchase behavior and search results ranking algorithms. Section 3 discusses the search results ranking influence factors based on C2C platform. Section 4 describes an original search results ranking algorithm and then an experiment is carried out to show the advancement of presented algorithm. This paper is concluded in section 5.

1. RELATED WORKS

The research area we have identified above is most relevant to the research domain of consumers' online purchase behavior and search results ranking algorithms. At the present time, the research on consumers' online purchase behavior has mainly focused on consumers' online shopping influence factors and consumers' purchase decision-making model etc. Yan & Dai proposed a conceptual model of consumer's online shopping decision making model and tested its validity with empirical data (2009). Chen presents a conceptual framework to access consumer acceptance of e-commerce websites, and it stresses on the roles of contingency factors and pinpoints potential areas for consumer search behavior (2005). Some researchers examined some special influence factors during online shopping. Lee et al. have done some research on the effect of negative online consumer reviews on product attitude. They investigated how negative online consumer reviews have a negative effect on consumer attitude (2008). Yeh et al. have done some research on Emotional and Rational Purchasing Behavior for Online Shopping (2007).

In search results ranking algorithms' field, many researchers have done much research in recent years. Most of the research focused mainly focuses on the improvement of present ranking algorithms such as PageRank, Hits etc. Besides, some researchers paid much attention to ranking search results with inferring relevance and intents of queries from user logs etc. It places more emphasis on search results ranking algorithm based on user preference. Bansal et al. presented an approximation algorithm that can provide highly ranked relevant search results based on user logs (2010). Guo proposed a similarity computing approach based on VSM with the improved TF-IDF algorithms (2008).

2. ANALYSIS OF SEARCH RESULTS RANKING INFLUENCE FACTORS BASED ON C2C PLATFORM

There are many influence factors which influence online shoppers' purchase decision when shopping on C2C websites. The factors, which influence the buying intentions of online consumers, are different to those that influence traditional consumers (Chen & Li, 2007). When consumers purchase online, shoppers' psychological thought are quiet different from traditional purchase activities. Traditional shoppers' behavior theories can't explain their online shopping behavior effectively. Analysis results of personal online shopping influence factors has many advantages.C2C website operators can achieve better marketing objectives with better designed commodity ranking strategies and more creative websites. In addition, it will improve shoppers' acceptance on shopping.

When shoppers submit a query on C2C websites, they return list of search results for shoppers. Among these results, shoppers have to make a decision to click which item. If we rank these search results based on shoppers' click possibility, it may greatly enhance shoppers' acceptance of commodities. In this paper, four elements are supposed to influence shoppers' click possibility which including keywords, commodity, website and environment. Figure 1 shows the influence factors of search results ranking based on C2C platform.

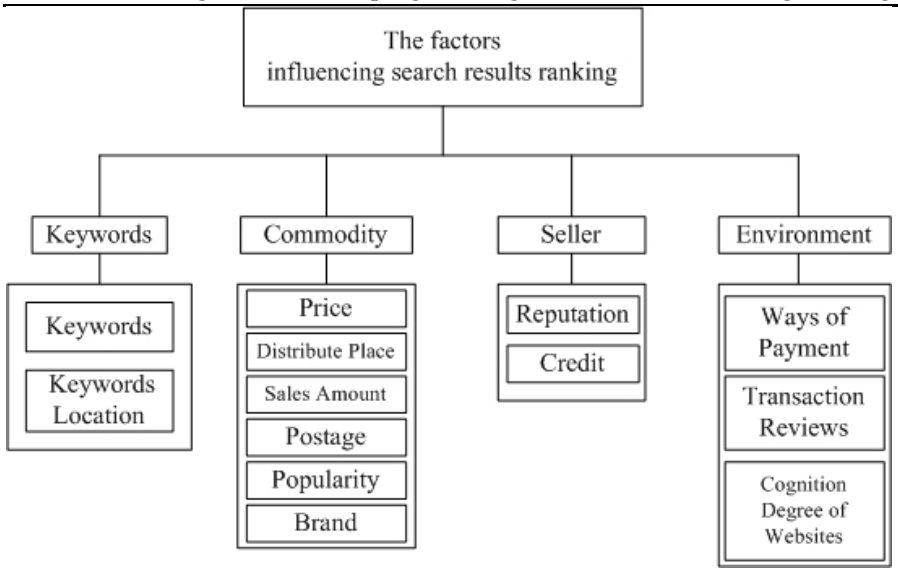

Figure 1. The Influence Factors of Search Results Ranking

The most important factor is shoppers' search keywords, and it plays an essential role regarding click possibility of search results. Search keywords and keywords' location consist of this factor. The second important factor is commodity which is the object of online shopping. Commodity itself influence online shoppers' perception and will influence their online shopping decision. Price level, distribute place of commodity, sales amount, postage，popularity and brand consist of this factor. The third factor is the seller. Seller is an important factor influencing shoppers' click possibility. Sellers' reputation and sellers' credit consist of this factor. The last factor is environment of online shopping on C2C websites. Means of payment, convenience of transaction on the website, transaction reviews, and cognition degree of websites consist of this factor.

3. AN ORIGINAL SEARCH RESULTS RANKING ALGORITHM

In this section, we classify the factors that influence search results ranking into two parts by the principle of utility maximization. Then an original search results ranking algorithm based on C2C platform is presented. An experiment is also carried out to testify its validity and advancement.

3.1 Classification of Search Results Ranking Influence Factors

According to the classic economics theory, consumers will follow the principle of utility maximization in the decision-making process, thus we can treat all the external factors and internal factors that influence decision-making into two parts that is gain and loss. In this paper, we classify the factors into positive factors and negative factors. Consumers tend to reduce the negative factors rather than to maximize positive factors when making purchase decisions. Here, we propos a supposed formula shown below.

$$\text{Click possibility (CP)} = \frac{\text{Perceived gain (PG)}}{\text{Perceived loss(PL)}} \qquad (1)$$

Perceived gain is the positive factors, while perceived loss is the negative factors. As the same as search results ranking of personal online search engine, shoppers can make decisions to click the item or not according to the supposed model. The search items are ranked from high to low by the value of CP. However, this model can not be calculated. Search results ranking is still complex. In next part, we propose an original ranking algorithm to solve this issue.

3.2 Ranking Algorithm based on Vector Space Model (VSM)

As mentioned above, we classify these factors into positive factors and negative factors by following the principle of utility maximization. However, this model can not be calculated. The vector space model is a perfect way to solve this issue. Suppose that each search results' page contains all the positive and negative factors mentioned above, here we call the factors terms. Moreover, suppose that the terms are independent. Then, each search results page can be expressed in a vector, that simplify the complexity relationship of the words and make the model have calculability.

When shoppers search on C2C websites, system returns the results set. Each search results item is composed by terms (positive factors and negative factors), which are independency. Now, suppose that any search results item ***i*** is expressed in a vector R_i.

$$R_i = \{ t_1, t_2, t_3, \ldots\ldots, t_n \} \quad (1 \leqq i \leqq n);$$

Where t_n is one of the terms of search results item *i*, each dimension corresponds to a separate influence factor. In order to concentrate on the main issue, there is a hypothesis below.

Hypothesis 1: Any search results item is composed only by the positive factors and negative factors that have been set.

Each search results item R_i, its every dimension will be given a weight, each weight is a measure of the importance of corresponding terms (factors).Here, and we express it as follows.

$$R_i = \{ w_{i1}, w_{i2}, w_{i3}, \ldots\ldots, w_{in} \} \quad (1 \leqq i \leqq n);$$

In a general way, we can get the similarity by calculate the two vectors' cosine. However, the search results items are tremendous and the efficiency of the calculation is low. Thus, we propose an improved approach. As for shoppers, they expect the terms corresponding to the positive factors have a higher weight, while the terms corresponding to the negative factors have a lower weight. Thus, in order to express clearly, there is another hypothesis below.

Hypothesis 2: Each term of the search results item has a weight, and we suppose it is a numerical value from 1 to 10. Each weight is a measure of the importance of corresponding factors.

Theoretically speaking, shoppers expect the positive factors' weight in search results item are the highest, it means that the terms corresponding to positive factors have the weight value 10.On the contrary, shoppers expect the terms corresponding to negative factors have the weight value 1.Thus, shoppers expect the ideal search results item's weight vector R_t is:

$$R_t = \{10,10,10,\ldots,1,1,1\ldots\ldots\ldots\};$$

The number of value 10 is the same as the number of positive factors, while the number of value 1 is the same as the number of negative factors. The similarity between any search results item and shoppers' ideal search results item is:

$$Sim(r_i, r_t) = \frac{\sum_{j=1}^{n} r_{ij} r_{tj}}{\sqrt{\sum_{j=1}^{n} (r_{ij})^2 \sum_{j=1}^{n} (r_{tj})^2}} \quad (1 \leqq i \leqq t \leqq j \leqq n); \quad (2)$$

We make every search results item compare with the shoppers' ideal search results item. The value of $Sim(r_i, r_t)$ is a value from 0 to 1.Then we rank the search results items according to the similarity from high to low.

Owing to each shopper's different criterion on the ideal search results item, this algorithm can rank the search results with user preference. It can realize users' search personalization.

3.3 Experiment

Finally, an experiment is carried out on Taobao.com to show the effectiveness of the algorithm presented above, and it also testifies that the search results can be ranked with user personalization. We test it with surveys, from both online volunteers and offline paper surveys. In order to simplify the experiment, we choose only 8 influence factors which include keywords, price, sales amount, credit, brand, and ways of payment, postage and transaction comments. So each search results item is a set of these factors. In this paper, we choose two samples to testify the algorithm.

As for user A, in his criterion, the ideal R_t = {10, 10, 1, 10, 1, 1, 1, 10}. Suppose that he searches the commodity "MP3" on Taobao.com, we choose the top 7 search result items to analyze. The weight and the similarity are shown in table 1.

Table 1

Experiment Results of User A

	Key-Words	Price	Sales Amount	Credit	Brand	Ways of Payment	Postage	Transaction Comments	Similarity
1	10	8	5	7	3	4	2	9	0.9441
2	8	9	4	7	5	6	3	10	0.9137
3	10	7	4	9	5	3	2	8	0.9441
4	8	7	6	9	4	2	1	10	0.9374
5	6	7	4	5	3	2	1	10	0.9313
6	7	8	2	9	3	4	5	10	0.9441
7	10	6	1	4	3	2	5	8	0.9066

We can conclude that the top 7 items should be ranked as follows: 1, 3, 6, 4, 5, 2, 7.With the presented algorithm, the shoppers can get the search results personalization. Due to the search results items we chose are all the top 7 items, the value of similarity have little difference. There is another user B which is a volunteer online, in her criterion, the ideal R_t = {10, 10, 10, 1, 10, 1, 1, 1}. This time we choose the search results item randomly. The weight and the similarity are shown in table 2

Table 2

Experiment Results of User B

	Key-Words	Price	Sales Amount	Credit	Brand	Ways of Payment	Postage	Transaction Comments	Similarity
1(3)	10	8	9	4	6	3	2	1	0.9592
2(5)	7	8	6	5	10	4	2	3	0.9260
3(8)	6	7	5	8	10	3	1	9	0.7838
4(15)	8	4	3	9	7	6	5	1	0.7153
5(21)	10	9	8	5	7	6	3	2	0.9233
6(34)	6	5	7	4	3	8	9	10	0.6151
7(83)	9	10	7	4	8	3	2	1	0.9674

The numbers in the bracket in the fist list means the real search results items index on Taobao.com. We can conclude that these 7 items should be ranked as follows: 7(83), 1(3), 2(5), 5(21), 3(8), 4(15), 6(34).In the experiment, we find that user B prefers to item 21 and item 83 though they are not in the top rank list.

This experiment has testified the validity of the algorithm proposed above, and it also shows that the search results ranked based on user preference.

CONCLUSION

This paper analyzes the search results ranking influence factors based on C2C platform. We classify these factors into positive factors and negative factors by following the principle of utility maximization. An original search results ranking algorithm based on VSM is proposed, and the search results can be ranked in

a rational order with the presented algorithm. Consumers can enjoy a wonderful search experience when purchase on C2C websites.

However, this paper has many limitations. The amount of samples in the experiment is a little small comparing to other empirical researches. In the future research, more data and tests should be undertaken to testify the algorithm's advancement.

REFERENCES

Bansal Nikhil, Jain Kamal, et al. (2010) . Approximation Algorithms for Diversified Search Ranking. *Lecture Notes in Computer Science*, *6*, 273-284.

Chen Hui & Li Yuanzhi. (2007). An Empirical Research of Factors Influencing the Decision-Making of Chinese Online Shoppers. *IFIP International Federation for Information Processing*, Volume 251, Integration and Innovation Orient to E-Society, *1*, 202-210.

Chen Rui. (2005). Modeling of User Acceptance of Consumer E-Commerce Website. *Lecture Notes in Computer Science*, *3806*, 454-462.

Guo Qinglin. (2008). The Similarity Computing of Documents Based on VSM. *Network-Based Information Systems*, *5186*, 142-148.

Lee Jumin, Do-Hyung Park, Ingoo Han. (2008). The effect of negative online consumer reviews on product attitude: An information processing view. *Electronic Commerce Research and Applications*, *7*(3), 341-352.

Yan Xiangbin & Dai Shiliang(2009). Consumer's Online Shopping Influence factors and Decision-making model, *Lecture Notes in Business Information Processing*, *36*, Value Creation in E-Business Management, Part 2, 89-102.

Yeh Lifen, Wang Eric Min-Yang, et al. (2007). A Study of Emotional and Rational Purchasing Behavior for Online Shopping. *Online Communities and Social Computing*, *4564*, 222-227.