WHICEB 2019 Proceedings

Wuhan International Conference on e-Business

Summer 6-26-2019

An Optimal Method For Product Selection By Using Online Ratings And Considering Search Costs

Zhouyang Liang
School of Management, Guangdong University of Technology, Guangzhou, 510520, China, liangzy@gdut.edu.cn

Meihua Zuo
Huizhou University, Huizhou, 516000, China;School of Management, Guangdong University of Technology, Guangzhou, 510520, China, zuomeihua123@126.com

Yonghua Zuo
School of Mathematics and Information Science, Jiangxi Normal University, Nanchang, 330022, China, youghua.zuo@foxmail.com

Hui Zhu
School of Management, Guangzhou University, Guangzhou, 510006, China, zhuhui@gzhu.edu.cn

Mengting Fan
School of Management, Guangdong University of Technology, Guangzhou, 510520, China, 1534825724@qq.com

See next page for additional authors

Authors

Zhouyang Liang, Meihua Zuo, Yonghua Zuo, Hui Zhu, Mengting Fan, and Mingjun Zhan

An Optimal Method For Product Selection By Using Online Ratings And Considering Search Costs

Zhouyang Liang[1], Meihua Zuo[2,1], Yonghua Zuo[3], Hui Zhu[4], Mengting Fan[1], Mingjun Zhan[1]*
[1]School of Management, Guangdong University of Technology, Guangzhou, 510520, China
[2]Huizhou University, Huizhou, 516000, China
[3]School of Mathematics and Information Science, Jiangxi Normal University, Nanchang, 330022, China
[4]School of Management, Guangzhou University, Guangzhou, 510006, China

Abstract: With the collecting and publishing data about consumers purchasing and browsing products at the platform of online, this data prodives new ways to better understand the consumers search behavior before purchase. How to base on consumers online search behavior and simutaneously consider offline experience costs is worth studying. An optimal method based on the utility of the attribute of product is proposed. The proposed method follows steps below. Firstly, based on the multi-attribute utility theory, the overall utility of product is calculated by using ratings data. Secondly, the overall utility is combined into the original sequential search model to find the optimal selection strategy. Thirdly, the candidate product sets arranged in descending order of the reservation utilities are finally obtained. Finally, taking the online ratings data provided by a comprehensive automobile website as an example, lastly the proposed method is simulated and compared with other method. The result shows that the proposed method is feasible and effective.

Keywords: sequential search model, multi-attribute utility theory, product selection strategy, earch cost

1. INTRODUCTION

With the rapid development of the Internet, most online platforms provide a rating function (comment, word of mouth) for product or services. Such as JD, Taobao, Douban, Automobile Home and so on. Many scholars are studying how to use online rating or evaluation information to provide quality services to consumers. Some scholars have found that online rating information can effectively improve customer repurchase rate and loyalty [1-2], using rating information to shorten the brand relevance of consumers' purchasing decision-making process [3]. Research shows that taking reviews as a source of decision information can reduce the risk associated with purchasing decisions. Studies have shown that online rating information or comment information plays an important role in purchasing decisions [4-5]. Therefore, it is a worthwhile research to support consumers to make product selection decisions based on online ratings.

Product selection process is an important part of the purchase decision making and it has a direct impact on the purchase results. This process can be seen as an optimal sequential search with different expectations and uncertainties. Consumers decide to stop searching, which may be attributed to the inefficiencies or high search costs of next search [6-7].

Search cost is the time and effort that consumers spend on searching for goods. Search theory believes that lower search costs leading to an increase in search behaviors. Assuming the search cost is zero, the rational

* Corresponding authors. Email: liangzy@gdut.edu.cn (Zhouyang Liang), zuomeihua123@126.com (Meihua Zuo), youghua.zuo@foxmail.com (Yonghua Zuo), zhuhui@gzhu.edu.cn (Hui Zhu), 1534825724@qq.com (Mengting Fan), zhanmj_gdut@163.com (Mingjun Zhan)

consumer will search for all the options and choose the one with the most greater real utility. In contrast, when the search cost is infinite, then the consumer will not search for any of the alternatives.

For example, a consumer plans to buy a car. In general, he or she will obtain the relevant information through the professional online car information platform, and choose N models of cars as an alternative. The costs of this initial information filtering is generally considered to be relatively low, even near to zero. However, making purchase decisions in N alternatives, consumers also need to conduct more in-depth search surveys on these alternative vehicles. Because it is only to test-drive to understand the true effect . In theory, the car that was test-drove, then consumers can choose the one that suits them best according to their preferences. However, in the actual situation, test-driving all alternative set are not the best strategy due to various costs (such as traffic expenses, time and so on).

Therefore, consumers usually develop an action plan: First, all the alternative cars based on the information they know are sorted (depending on the strategy, such as random sorting, sort by expected utility). Then, test drive the first of the alternative cars. Every time customer test-drive a car, there is a certain cost to pay. If customer think that this car is more suitable than all the remaining vehicles, this one can be chose. Otherwise, continue to spend a certain amount of time to test-drive next car. At this time, the consumer can compare the testing-drive with the test-drove, and can choose the one with the highest utility or continue to test-drive. This cycle finish until the purchase decision is made. Then, assuming that the consumer is completely rational and able to evaluate the cost of searching for each commodity, whether there is an optimal test drive strategy to maximize the utility of the consumer is a problem worth studying.

2. REVIEW OF THE LITERATURE

when consumes search for product or service information online, a similar problem arises: which strategy can maximize the utility during sequential search process. Researchers have studied the methods of product ranking based on online rating data [8-9]. Li et al (2014) researched the problem of ordering product based on online rating information, and proposed an online product ranking method based on social network analysis. The method solves the problem of inconsistent online rating information rating group. Liu et al (2017) used sentiment analysis and intuitionistic fuzzy sets to sort the items after processing the online comment information. Most of the existing research do not study how to select product based on online rating information from the perspective of behavioral economics and rarely considering consumers' search cost while searching for product information. According to economic theory, search cost is a key factor in determining the scope of consumer search activities. Therefore, when the expected marginal profit of the search is lower than the marginal cost, the search should be stopped [10-12]. Most research focused on identifying search costs and models [13] . Developing a structured approach that uses integrated data to estimate the distribution of consumer search costs [14] . Using comprehensive data and estimated a search model (search under a fixed alternative) in the automotive market environment [15] . Using behavioral data viewed and purchased by users to analyze which classic search model is more in line with observed data patterns [16] . Model based on user clickstream data, including product attribute information and user purchase information. However, few studies have used rating information to model user search behavior. This paper present a novel method to model the decision-making process of consumer product selection based on rating information, and provide decision-making basis for consumers to choose product.

3. MODEL BUILDING AND DEVELOPMEN

3.1 Utility and search costs

A multi-attribute utility function based on utility theory, which is a model that can fully describe consumers' purchasing decisions based on multiple attributes [17-18].

Assuming that each item has *l* attributes, and the score of *j* item attribute k is x_{jk} , then item *j* can be represented as a vector $J=(x_{j1},x_{j2},\cdots,x_{jl})$ composed of *l* pieces of x_{jk} .

Suppose x_{jk} has *v* discretized attribute rating values r, where by the rating set $R=(r_{k1},r_{k2}\cdots,r_{kv})$ constituting attribute value k is discrete. Most of the the product rating data is discrete. For a small number of continuous data, it can be graded to facilitate discretization.

According to the multi-attribute utility theory, the utility function can be expressed as follows:

$$u_{ij}=\sum_{k=1}^{l}w_{ik}\cdot u_{ijk}(x_{jk}) \tag{1}$$

u_{ij} is the utility of product *j* to user *i*, and w_{ik} is the preference weight of user i's sensitivity and satisfies to product attributes.

$$\sum_{k=1}^{l}w_{ik}=1;0\le w_{ik}\le 1 \tag{2}$$

$u_{ijk}(x_{jk})$ refers to the utility of product *j's* attribute K to user *i*, and the calculation formula is as follows:

$$u_{ijk}(x_{jk})=a_{ijk}+b_{ijk}x_{jk} \tag{3}$$

This paper specifies a linear utility function to represent consumer's preference for an attribute to improve computational efficiency , but it is not affecting the ranking results. In this paper, the rating data of each attribute is within the interval [1, 5]. So it can be specified that the rating who's value is 1 corresponds to the lowest utility 1, and the the rating who's value is 5 corresponds to the highest utility 5. Therefore, the scale parameters in equation (4)can be determined, namely a_{ijk}=0 and b_{ijk}=1.

Consumers have to pay a certain search cost for each additional commodity. Search cost can be interpreted as the time and effort spent on searching. Consumers with stronger time constraints have higher search cost levels[19]. In the context of buying a car, search cost is interpreted as the time cost for identifying and evaluating a candidate. Consumers can evaluate search costs based on their distance from the 4S store and their income level.

3.2 Probability distribution of utility

We assume the distribution of utility of each attribute is truncated normal distribution. It's probability density function as follow:

$$f\left(u_{ijk};\mu,\sigma,a,b\right)=\frac{\frac{1}{\sigma}\phi\left(\frac{u_{ijk}-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)} \tag{5}$$

Where $\phi(\cdot)$ is standard normal distribution. $\Phi(\cdot)$ is cumulative distribution function of standard normal distribution. From formula (6), the cumulative density distribution of the utility can be calculated:

$$F(u_{ij})=\int_{-\infty}^{\infty}f\left(u_{ij}\right)du_{ij}=\int_{-\infty}^{\infty}f\left(\sum_{k=1}^{l}w_{ik}\cdot x_{jk}\right)dx_{ij} \tag{7}$$

w_{ik} is the preference of user i for the attribute k of product j, $F(u_{ij})$ is the cumulative density distribution of the product j.

3.3 Utility-based sequential search model

In the sequential search process, the consumer decides to stop or continue searching after searching for one commodity. The optimal Sequential Search theory holds that the search is stopped when the marginal revenue of the continued search is less than the marginal cost. The utility value of product j for consumer i is recorded as u_{ij} . At any stage of the search process, u_i^* is defined as the highest value of the product searched so far. When searching for product j, the consumer's expected marginal benefit is:

$$\mathrm{M}(u_i^*) = \int_{u^*}^{\infty} (u_{ij} - u_i^*) f\left(u_{ij}\right) du_{ij} \tag{8}$$

Where *f*(*) is the probability distribution of u_{ij} . The marginal benefit is the expected utility value of product j. Assuming that the marginal benefit is higher than u_i^* , the probability that u_{ij} exceeds u_i^* is greatly increased.

Please note that the revenue from the search only depends on the ranking of utility values above u_i^* . Utility values below u_i^* are at the left end of the utility distribution and can be rearranged arbitrarily without affecting the search or selection.

Given the current best selection, the consumer's goal is to maximize the difference between the expected utility value and the search cost incurred after searching a set of products. At the individual level, these products include the following two characteristics: the product-specific search cost c_{ij}, and the product-specific uncertainty *f*(*). It means that, if there is at least one product j, the premise of the consumer continuing to search is as follows:

$$c_{ij} < \mathrm{M}_{ij}\left(u_i^*\right) \tag{9}$$

That is, the expected marginal benefit of continuing to search is greater than the marginal cost c_{ij} .

The optimal Sequential Search strategy can be normalized to the following steps. First, divide the product collection into $S_i \cup \overline{S}_i$, S_i contains all the products that have been searched. $\overline{S}_i$ contains all unsearched products. Suppose u_i^* is the maximum utility value of all products in set S_i ,

$$u_i^* = \max_{j \in S_i} \left\{\hat{u}_{ij}\right\} \tag{10}$$

$\hat{u}_{ij}$ is the true value of u_{ij} . For convenience, the product with a utility value of u_i^* is called a candidate. At any stage of the search process, the state of the system is given by $(\overline{S}_i, u_i^*)$. Defining the function value $H(\overline{S}_i, u_i^*)$ is the expected value (calculated discount) that follows the optimal search strategy as it moves forward from the current state. This function value must satisfy the following Bellman equation [12]:

$$H(\overline{S}_i, u_i^*) = \max\left\{u_i^*, \max_{j \in \overline{S}}\left\{\beta_i\left[H\left(\overline{S}_i - \{j\}, u_i^*\right)\int_{-\infty}^{u_i^*} dF\left(u_{ij}\right) + \int_{u^*}^{\infty} H\left(\overline{S}_i - \{j\}, u_i^*\right) dF\left(u_{ij}\right)\right] - c_{ij}\right\}\right\} \tag{11}$$

where $F(\cdot)$ is the cumulative density function of u_{ij} (CDF), $E(\cdot)$ is the expected utility of u_{ij} .

Equation (12) shows that starting from state $(\overline{S}_i, u_i^*)$, the consumer can choose to terminate the search and purchase the candidate for u_i^* , or choose to search for any $j \in \overline{S}_i$. When the latter is chosen, the consumer attempts to maximize the expected value of product j, and that is:

$$H(\overline{S}_i - \{j\}, u_i^*)\int_{-\infty}^{u_i^*} dF(u_i^*) + \int_{u_i^*}^{\infty} H(\overline{S}_i - \{j\}, u_i^*) dF(u_{ij}) \tag{13}$$

Because the single session involved in the search is performed in a short period of time (within a few days), the discount rate β_i can be set to 1.

For now, some important assumptions can be made for the search model. First, the model is a full information model in which consumers are assumed to have a complete understanding of the product and its attribute values. This allows the consumer to form u_{ij} for all products prior to the search and use them in the sequential search process to calculate the reservation utility value. Second, consumers can estimate their own search cost c_{ij} based on the information they have.

3.4 Optimal strategy

The solution to the above dynamic programming is to continue searching until a utility value u_i^* greater than a certain limit is found, which in turn depends on how many products remain in the unsearched collection. This limit depends on the size of the reservation utility value.

Each consumer i has a reservation utility z_{ij} for each product *j*. To define this concept, assume that the consumer has found that the reservation utility value of an product is the same whether she continues to search or stop searching. In other words, the reservation utility value z_{ij} satisfies the following equation:

$$c_{ij} = \mathrm{M}_{ij}(z_{ij}) = \int_{z_{ij}}^{\infty} (u_{ij} - z_{ij}) dF(u_{ij}) \tag{14}$$

Therefore, the formula for the reservation utility value is:

$$z_{ij} = \mathrm{M}_{ij}^{-1}(c_{ij}) \tag{15}$$

The optimal search strategy, namely the problem of solving the consumer maximization equation (5) (for example Weitzman(1979)) has three components: (i) Search rule: it determines the order in which the sequences are searched; (ii) Stop rule: it determines the length of the search sequence; (iii) Selection rule.

(i) Search rule: Calculate all reservation utility values z_{ij} and sort them in descending order. If you're searching for an commodity, you should search for products that have the highest utility value z_{ij} in the products that have not yet been searched.

(ii) Stop rule: If the maximum utility value u_i^* obtained so far is greater than the biggest reservation utility value among the commodity that have not been searched,then stopped the search.

(iii) Selection rule: Once the search stops, select the candidate commodity u_i^* in S_i.

It is assumed that the above-mentioned optimal selection rule and stopping rule are derived based on the knowledge obtained by searching for one product without affecting the cognition of other products, that is, the process of assuming that the consumer does not exist in the search process. The current literature on consumer behavior suggests that this is a reasonable assumption for consumers in the search process [20].

4. EXAMPLES AND SIMULATION

Take car buying as an example to illustrate the product selection strategy proposed in this paper. Suppose a consumer wants to buy a compact SUV. There are four alternative models in the alternative (A, B, C , D) . The favorite five attributes are: space, fuel consumption, power, control, cost performance.And the attribute preference is given as w=(0.4,0.2,0.2,0.1,0.1), Assuming that the cost of each test-drive is c (in combination with the time it takes to arrives the each 4S store, consumer can estimate his own search cost c).

A crawler program is designed by python and is used to download the word-of-mouth score data for the alternative models from the Automobile Home. For the score data of the five attributes, it can be determined that the number of consumers participating in the online evaluation of the four alternative models is 4896, 5066, 4977, 3987. Then, according to the equations (1) and (4), the utility probability distribution *f(u)* of each candidate vehicle model can be calculated.

After $f(u)$ is obtained, the reservation utility of the alternative product can be calculated according to formula (12). The calculation result is shown in Table 1. The formula for calculating the expected net utility is as follows:

$$NetU_{ij} = E(u_{ij}) - c_{ij} \tag{16}$$

Table 1. Expected utility of each alternative vehicle and reservation utility (considering C)

		A	B	C	D
(C=0)	Reservation utility	5	5	5	5
	Expected net utility	4.772	4.447	4.424	4.139
(C=0.1)	Reservation utility	4.744	4.674	4.673	4.453
	Expected net utility	4.672	4.347	4.324	4.039
(C=0.5)	Reservation utility	4.286	4.029	4.025	3.758
	Expected net utility	4.273	3.947	3.924	3.639
(C=1)	Reservation utility	3.775	3.465	3.451	3.17
	Expected net utility	3.744	3.447	3.424	3.139

Note: The order of reservation utility and expected utility is exactly the same because the distribution of rating data for each vehicle is similar.

The selection process for the above four vehicles is now simulated according to different selection strategies. In this paper, we use R to simulate the selection process. The first round of simulation simulates the selection process 2000 times, then each round of simulation adds 2000 times in turn. There are 200 rounds of simulations. So ,the last round of simulations is 200,000 times. We use the average net utility in each round of simulation to evaluate the actual effects of the two strategies, the calculation formula is as follows:

$$AverageNetU = \frac{1}{n}\sum_{t=1}^{n}\left(u_t^* - \sum_{j\in S} c_{tj}\right) \quad , \quad t = 1,2,\ldots,n \tag{17}$$

Where the number of simulations per round is n, u_t^* is the final choice obtained from the t-th simulation, The average net utility of the two strategies at different search costs can be seen in Figure 1 .

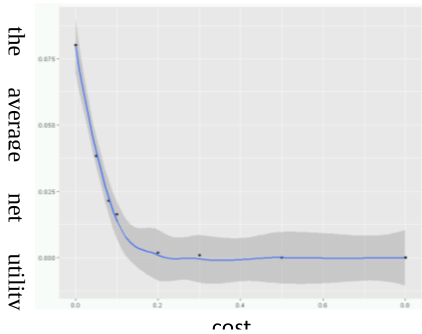

Figure 1. Search cost, average net utility difference between the two strategies

It can be seen from the simulation results that as the number of simulations increases, the average net utility converges to a certain value. At the same time, the average net utility of presented method is higher than the expected net utility method, and the lower the search cost is, the more obvious the benefit of proposed strategy is. As can be seen from Figure 1, as the search cost increases, the average net utility difference will become smaller and smaller. Based on the data of this example, the cost c starts from 0.2, and the average net utility difference gradually approaches zero. That is, when the cost-utility ratio C/U is greater than 4%, the effect of using the two strategies is the same, because the search cost at this time has begun to approach the marginal benefit, so the consumer loses the willingness to search for the next commodity. This also leads to the same depth of search for both methods, and the resulting utility is indiscriminate (because the actual utility of the search depends on the utility distribution of products). According to that, the merchant can minimize consumer's search cost to enhance the attraction to the consumer and increase the probability of the consumer searching for the product. Using the product selection strategy proposed in this paper, we can increase the maximum consumer surplus.

5. CONCLUSION

This paper examines the issue of product selection considering consumer's search costs and their preferences. Since consumers need to spend time and effort to evaluate and purchase goods, rational consumers seeking to maximize utility need an optimal selection strategy. We use discretized online attribute rating data to build a utility-based optimal sequential search model. From a new perspective, the problem of product selection in reality is solved, and the strategy calculation process is simple and easy to program.

The results show that the sequential search strategy in descending order of reservation utility is the best choice for consumers. In particular, reservation utility is primarily related to the distribution of higher utility in the utility distribution. That is to say, when consumers choose goods, they should give preference to low-probability and high-return products, because it is easier to reduce the number of searches and stop the search as soon as possible.

There are still some aspects that can be studied in depth in the future. Firstly, consumers in reality are often not completely rational. What kind of characteristics do irrational consumers have in the process of product selection? This is a problem worthy of study. Secondly, in this model, the proposed optimal search model does not have the learning ability, but in reality, consumers have learned and updated the distribution of product information in real time during the search process. So in the future research, the product selection strategy considering consumer's learning ability is also a valuable research direction.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71671048 and the National Social Science Fund of China under Grant 17BJL025.

REFERENCES

[1] Gauri D K, Bhatnagar A, Rao R. Role of word of mouth in online store loyalty[J]. Communications of the Acm. 2008, 51(3): 89-91.

[2] Cheung C M K, Thadani D R. The impact of electronic word-of-mouth communication: A literature analysis and integrative model[J]. Decision Support Systems. 2012, 54(1): 461-470.

[3] Hu Y, Li W, Lu Q. Developing Evaluation Model of Topical Term for Document-Level Sentiment Classification[J]. 2008, 5351: 175-186.

[4] Qu Z, Zhang H, Li H. Determinants of online merchant rating: Content analysis of consumer comments about Yahoo merchants[J]. Decision Support Systems. 2009, 46(1): 440-449.

[5] Park I, Cho J, Rao H R. The effect of pre- and post-service performance on consumer evaluation of online retailers[M]. Elsevier Science Publishers B. V., 2012: 415-426.

[6] Sorensen A T. Equilibrium Price Dispersion in Retail Markets for Prescription Drugs[J]. Journal of Political Economy. 2000, 108(4): 833-850.

[7] Koulayev S. Search for differentiated products: identification and estimation[J]. Rand Journal of Economics. 2014, 45(3): 553-575.

[8] 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[J]. Information Fusion. 2017, 36: 149-161.

[9] Li Y, Wu C, Luo P. Rating online commodities by considering consumers' purchasing networks[J]. Management Decision. 2014, 52(10): 2002-2020.

[10] Ratchford B T. Cost-Benefit Models for Explaining Consumer Choice and Information Seeking Behavior[J]. Management Science. 1982, 28(2): 197-212.

[11] Rothschild M. Searching for the Lowest Price When the Distribution of Prices Is Unknown[J]. Uncertainty in Economics. 1978, 82(4): 689-711.

[12] Weitzman M L. Optimal Search for the Best Alternative[J]. Econometrica. 1979, 47(3): 641-654.

[13] Hong H, Shum M. Using price distributions to estimate search costs[J]. Rand Journal of Economics. 2006, 37(2): 257-275.

[14] Moragagonzález J L, Sándor Z, Wildenbeest M R. Consumer Search and Prices in the Automobile Market[J]. Social Science Electronic Publishing. 2009.

[15] Santos B D L, Hortaçsu A, Wildenbeest M R. Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior[J]. American Economic Review. 2012, 102(6): 2955-2980.

[16] Kim J B, Albuquerque P, Bronnenberg B J. Online Demand Under Limited Consumer Search[J]. Social Science Electronic Publishing. 2010, 29(6): 1001-1023.

[17] Manouselis N, Costopoulou C. Analysis and Classification of Multi-Criteria Recommender Systems[M]. Kluwer Academic Publishers, 2007: 415-441.

[18] Feng Deng, Yongan Zhang. Collaborative filtering recommendation system based on multi-attribute utility. Journal of Computer Applications.2015(07): 1988-1992.

[19] Roberts J H, Lattin J M. Development and testing of a model of consideration set composition.[J]. Journal of Marketing Research. 1991, 28(4): 429-440.

[20] Moorthy S B T R. Consumer information search revisited: Theory and empirical analysis[J]. Consumer Res. 1997, 4(23): 263-277.