Research on Chinese well-known e-commerce enterprises' innovation ability based on real comment

Shufang Li*

School of Accounting and Finance,
Zhejiang Vocational College of Commerce,
Hangzhou 310053, China
Email: lishufang77@163.com
*Corresponding author

Changbing Jiang

School of Management Science and Engineering,
Contemporary Business and Trade Research Center,
Zhejiang Gongshang University,
Hangzhou 310018, China
Email: johncabin@zjgsu.edu.cn

Liang Li

School of Management Science and Engineering,
Zhejiang Gongshang University,
Hangzhou 310018, China
Email: 371952360@qq.com

Abstract: Based on the real comments from Jingdong Mall, T-mall and other e-commerce platforms as for the data resource, according to analysis to the customers' online shopping comments' vocabularies, obtained 11 indicators which are constituted e vocabularies, obtained 11 indicators which are constituted e-commerce platform customer satisfaction index system and the satisfaction indicator value, using projection pursuit model (PPM) and genetic algorithms to make a further calculation to satisfaction indicators. The results show that: which has the greatest impact on e-commerce platforms' overall satisfaction is the product reputation and popularity; Jingdong Mall has the best overall satisfaction among Jingdong Mall, T-mall, and Amazon this three e-commerce platforms. Meanwhile, each ecommerce platforms has its own superiority and own superiority and weakness.

Keywords: comment; word frequency statistics; satisfaction indicator; projection pursuit; genetic algorithms.

Reference to this paper should be made as follows: Li, S., Jiang, C. and Li, L. (2020) 'Research on Chinese well-known e-commerce enterprises' innovation ability based on real comment', *Int. J. Computing Science and Mathematics*, Vol. 11, No. 1, pp.54–62.

Biographical notes: Shufang Li is a Professor of School of Accounting and Finance, Zhejiang Vocational College of Commerce, Hangzhou, China. Her current research focuses on e-commerce public opinion supervision and business data processing.

Changbing Jiang received his PhD in Business Management from Zhejiang Gongshang University in 2016. He is currently an Assistant Professor at the Zhejiang Gongshang University, Hangzhou, China. His mainly engaged in supply chain management, teaching, researching and training of logistics system modelling and simulation.

Liang Li is a graduate student of Zhejiang Gongshang University. His main research interests are the application of cellular automata in logistics network optimisation.

1 Introduction

In recent years, the development of e-commerce is flourishing, even has the national online shopping tide, online shopping comments show the more and more obvious and important role in people's normal life, it not only helps seller to analyse the key indicators of the product and services, but also provide the important reference comments for the customers who want to buy the same products, analysis of mass comments is beneficial to the management and development of e-commerce, it has high research value.

Many scholars have studied it. Cui et al. (2017) present a new variant of bat algorithm combined with centroid strategy is introduced. Three different centroid strategies with six different designs are introduced. In addition, the velocity inertia-free update equation is also provided. Zhang et al. (2017) propose a hybrid multi-objective CS (HMOCS) for solving multi-objective optimization problems (MOPs). Wang et al. (2017) present a kind of high performance computing approaches, evolutionary multi-objective optimisation (EMO) algorithms, is used to deal with these MOPs. A floorplanning case study is presented to demonstrate the feasibility of our proposed approach. Cai et al. (2018) propose a triangle-flipping strategy to update the velocity of bats. Three different triangle-flipping strategies with five different designs are introduced. Cai et al. (2016) optimal forage strategy is designed to guide the search direction for each bat and a random disturbance strategy is also employed to extend the global search pattern.

Current researches of satisfaction, mostly using the questionnaires to obtain data, however the questionnaires has a number of uncontrollable factors, the quality of the data often cannot be guaranteed. This article use the network information extraction tool to extract more than five hundred thousand comments information from each e-commerce platform, meanwhile taking advantage of Chinese information processing technology to analysis the comments language, thus get the 11 indicators which constitutes the customer satisfaction of e-commerce platform.

In generally we mostly use the single indicator comparative analysis, multi-index analysis, survey, etc. to study the satisfaction. These methods are mainly one by one compare and evaluate the indicators of satisfaction, thus have a larger one-sidedness and anthropogenic factor, cannot fully and objectively reflect the overall level of satisfaction. Satisfaction indicators are numerous, complicated, with the typical high-dimensional, non-linear characteristics. To solve this problem, this article will creatively combine the Projection pursuit (PP) theory and satisfaction research.

PP is a new exploratory data analysis methods, which combine the statistics, applied mathematics and computer technology for handling high-dimensional, non-linear data, it

can effectively overcome the difficulty of 'dimension curse'. Analysis the features of satisfaction indicator, basically meet the applicable conditions of PP theory, therefore, we believe PP can be applied to satisfaction evaluation.

2 Construction of customer satisfaction indicator based on real comments

2.1 Initial customer satisfaction index

Combined with results of previous studies on the factors affecting the e-commerce enterprise customer satisfaction and customer satisfaction index system, this article constructs six two stage index which including site satisfaction, marketing planning satisfaction, product satisfaction, logistics satisfaction, pay satisfaction and after-sales satisfaction and initial customer satisfaction Index which has eighteen three stage index includes: web response speeds, product variety, product information, comprehensive, web design layout, discount promotions, advertising, specialty plates, product performance, product reputation and visibility, product price, product appearance, distribution methods diversity, logistics speed, the quality of packaging, logistics secure payment diversity, payment security mechanism to return, advisory services, complaints process.

2.2 Research objects and data sources

This study chooses three e-commerce platforms including Jingdong Mall, T-mall, Amazon, which have a high influence in e-commerce, based on real user reviews from e-commerce platform as a data source. Among them, Jingdong Mall is Chinese domestic comprehensive online store, which transforms from the 3C digital, T-mall is the largest Chinese integrated online store, Amazon is a global integrated e-commerce platform, has the advantage in books audio and video products, these platforms are representative and have the true and reliable source of data.

2.3 Data acquisition and processing

2.3.1 Data acquisition

Collection tools used in the study is called 'Network miner (Soukey Miner) Data Acquisition Software Professional Edition V2012SP2', to the consumer electronic business platform Jingdong Mall, T-mall, Amazon Reviews area comments as a data acquisition target, through the development of collection rules specify automatic extraction comment on the page. As result, it collected 247,866 online reviews from Jingdong Mall comments area, collected 200,001 online reviews from T-mall comments area, and collected 155,138 online reviews from Amazon comments area.

2.3.2 Data processing

This paper randomly collected 10,000 comments from Jingdong, T-Mall, Amazon this three e-commerce platform. After loading the word frequency statistics tool in EXCEL, we count the frequency of these 10,000 sample data, and count former 200-frequency

descriptors. After comparison with the initial customer satisfaction index system, we can get the correspondence between the index and descriptive words, show in Table 1.

Table 1 The correspondence between the index and descriptive words

Indicator	*Description*
Speed of website response	
Product variety	Much; little; rich
Comprehensive product information	Clear; not clear
Layout of web design	
Discount and promotions	Discounts; no discounts; promotion; events; no activities; cheap; special deals, etc.
Advertisement	
Features section	
Product performance	Good; not good; satisfied; not satisfied; worthy; unworthy; not bad; bad; nothing to tell; have some problems; love; like; dislike; great; awesome; easy to use; hard to use, etc.
Product reputation and popularity	Famous; famous brand; liar; poor; classic; dependence; king; trust; unworthy, etc.
Product price	Affordable; not affordable; worth; worthwhile; justice; expensive; cheap; awesome; not awesome; high; low; economic; suitable; kind; just-so-so, etc.
Product exterior	Delicate; not delicate; good-looking; bad-looking; beautiful; not so beautiful; like; dislike; ok; not ok; generous; not bad; ugly; unsightly, etc.
Delivery diversity	Fast; slow; timely; late; speed; not arrive; few days; long time, etc
Speed of logistics	
Packaging quality	Well-packaging; good-looking; damaged; completed; intact; broken; incompleted; defective, etc.
Logistics security	Kind; good person; touched; bad attitude; responsible;
Courier Attitude	Irresponsible; ferocious, etc
Payment diversity	
Payment security	
Return mechanism	Agree; disagree; changeable; not changeable, etc.
Consultation service	Good customer service; bad customer service; bad attitude; good service; friendly; ferocious
Complaints process	Complaints; dissatisfied; responded; trouble; deal with; ignore

From the correspondence in the above table, we can find that in the initial customer satisfaction index system, there has many indicators which could not find the corresponding description words in a real review, based on the above study, we get customer satisfaction index system based on real comments, which is shown in Table 2.

After obtaining customer satisfaction indicators and its corresponding description word, we use Chinese Word Frequency Statistics Tool Ver 1.0.1, which is written by penguin, to do a further word frequency statistics about the comments which is classified by indicators.

Table 2 customer satisfaction index based on real comments

Primary indicators	*Secondary indicators*	*Tertiary indicators*
E-commerce customer satisfaction	Marketing planning satisfaction	Discount and promotions
	Product satisfaction	Product performance
	Product reputation and popularity	
	Product price	
	Product appearance	
Logistics satisfaction	logistics speed	
	Packaging quality	
	Logistics security	
Aftermarket satisfaction	Return mechanism	
	Consultation service	
	Complaint handling	

In the process of analysing the comments data, we can find many descriptive words and evaluation index are not one to one correspondence (Niu and You, 2013), there are some conditions which includes a descriptive word corresponds to a plurality of indicators as well as an indicator includes a variety of descriptive words, this is due to the casual comments from the customers, so it may use more than one word to describe one indicator. We analyse the word frequency about descriptive words which are used to describe the same indicator; the all descriptors for the same indicator are added to obtain the total frequency of the indicator.

Table 3 E-commerce platform indicators satisfaction

		Jingdong	*T-mall*	*Amazon*
Website and marketing satisfaction	Discount and promotions	0.999707	0.999778	0.9986987
Product satisfaction	Product performance	0.9825014	0.937968	0.95365
	Product reputation and popularity	0.98776	0.994716	0.8807282
	Product price	0.897567	0.558722	0.91659
	Product appearance	0.999822	0.99992	0.99517
Logistics satisfaction	Logistics speed	0.89875	0.803946	0.761242
	Packaging quality	0.81841	0.996708	0.798063
	Courier attitude	0.94098	0.6	0.785714
Aftermarket satisfaction	Return mechanism	0.864865	0.868381	0.8571429
	Consultation service	0.75	0.71212	0.66667
	Complaint handling	0.855161	0.691356	0.923246

The comment data is positive but also negative, which means the corresponding descriptive words of indicators are also positive and negative. The number of the frequency of positive description plus the number of the frequency of negative description are as the total number frequency of this indicator, meanwhile, the proportion of the positive description frequency in the total, which is the satisfaction value of this indicator. Take the third indicators of package quality which is under the secondary

indicators of logistics satisfaction as an example, through the word frequency statistics of Jingdong Mall's comments, we can get the high-frequency description words which is used to describe Jingdong Mall's packaging quality and it's frequency number are as following: well package (13,972 times), attractive (1,546 times), damage (679 times), completed (39 times), intact (378 times), bad (1,638 times), poor(906 times), crack (309 times), incompleted (0), defective (2 times), etc., which the number of the frequency of positive description words is 15,935 times, the number of the frequency of negative description words is 3,534 times, therefore, Jingdong Mall customer satisfaction with the quality of packaging is 0.8185.

Based on the above analysis, we process the comments data collected from Jingdong Mall, T-mall, Amazon, the results are shown in Table 3.

3 PP and index weight

3.1 Overview of PP

PP (Lou and Qiao, 2013) is a class of new statistical method which is used to analyse and process high-dimensional data, especially non-normal data, the basic idea is projecting the high-dimensional data onto the low-dimensional subspace, then find out a projection which reflect the original high-dimensional data structure characteristics, to research and analysis high-dimensional data. Different with the traditional 'assumes-simulation-forecast' modelling methods, PP is a exploratory data analysis methods, using the traditional 'examine data – computer simulation – forecast' modelling ideas. It does not make any assumptions about the model, only analyse assumptions about the model, only analyse data structure, is a new way of thinking. Genetic algorithm is a class of robust search algorithms which can be used for optimisation of complex systems. Because the calculation traditional PP computing technique is quite large, to some extent, it limits the widespread application of PP method. But the PP modelling based on genetic algorithms can effectively solve can effectively solve the problems of application of PP method which need a huge calculation. Therefore, based on the genetic algorithm and PP establish e-commerce platform customer satisfaction evaluation model:

1 Establish projection. Assume that the j^{th} index data of i^{th} e-commerce platform is x $x_{ij}(i = 1, 2, \ldots, m; j = n)$, in order to eliminate the effect of each dimension indicators, we make the standardised treatment of x_{ij} to get to x_{ij}^*. Assume that $\vec{a} = (a_1, a_2, \ldots, a_i, \ldots, a_m)$ is the projection direction, by PP, we could project x_{ij}^* into $\vec{a}$, then get the one-dimensional projection value Z_j of sample j:

$$Z_j = \sum_{i=1}^{m} x_{ij}^* \vec{a}_i \tag{1}$$

2 Establish projection objective function. In calculating the projection values, it requires projection value Z_j collects as much variation information of x_{ij}^* as possible, which is the standard deviation S_z of Z_j should be as bigger possible. Meanwhile, the intra-class density D_z should also should also be as bigger as possible, in order to

make significant clustering. This paper let projection objective function $f(a)$ be as following:

$$f(a) = S_Z D_z \tag{2}$$

3 Optimise projection direction. Projection objective function $f(a)$ changes with the projection direction of the optimise the direction change of $\vec{a}$, when the projection index function has its maximum value, corresponding the direction of $\vec{a}$ is the optimal projection direction which could best reflect the characteristics of the data. Therefore, we could find the best projection direction by solving the maximum projection objective function. In this paper, the best projection direction is Satisfaction Index weight distribution, which is:

$$\begin{aligned} &\max \quad f(a) = S_Z D_z \\ &\text{s.t.} \quad \sum_{i=1}^{m} a_i^2 = 1 \qquad a_i \in [-1, 1] \end{aligned} \tag{3}$$

Firstly, randomly generates p set of initial unit projection direction $\vec{a}$, according to formula (1) to calculate the value of integrated projection Z_j for e-commerce supplier's satisfaction. Secondly, calculate the standard deviation S_Z and intra-class density D_Z, According to the formula (2), calculate the projection objective function. Thirdly, using the selection of genetic algorithm, crossover and mutation algorithms, to generate p set of new projection directions $\vec{a}$. Fourthly, calculate Z_j, the standard deviation S_Z and intra-class density D_Z, according to the formula (2), to calculate the projection objective function $f(a)$, according to the principle that the larger the value of $f(a)$ the more excellent, to keep the projection direction $\vec{a}$, when $f(a)$ is in maximum value. Fifthly, repeat the third step and the fourth step, it stops only until reach the disposed cycling conditions, get the optimal projection direction.

3.2 Satisfaction index weights

According to the foregoing data in accordance with relevant indicators, establish the PP model, using the genetic algorithm for the optimal projection direction of the e-commerce enterprises satisfaction indicator, the result is: $\vec{a}$ = (0.369699224, 0.168945331, 0.417933997, 0.023395588, 0.377061673, 0.299764822, 0.31340669, 0.090239854, 0.409392559, 0.386257752, 0.066068319). The results are shown in Table 4.

Analysis of the results above, we can find satisfaction with the well-known e-commerce platform of the following features.

1 Overall, satisfaction index value of each e-commerce platforms are over high, even considerable parts are more than 90%, this is a very significant phenomenon. However, from the partial perspective, there are some indicators values are too low. For example, T-mall price of the product satisfaction reached only about 56%, Amazon's consulting service satisfaction reached only 66.7%, these all need to be improved.

2 In the transverse view, we could draw a conclusion from the value of Z_j, Jingdong overall satisfaction ranks the first of three, the next is T-mall, the last one is Amazon. Meanwhile, due to the different focus of these three e-commerce platforms, there is a huge difference in the distribution of indicators satisfaction.

3 In the vertical view, compare the value of a_i, we could rank the satisfaction indicators by importance, it is as following: product reputation and popularity, return mechanism, consultation service, product appearance, discount and promotions, logistics speed, product performance, courier attitude, complaint handling, product price.

Table 4 The empirical results PP model

Satisfaction index	*Jingdong*	*T-mall*	*Amazon*	a_i
Discount and promotions	0.99970715	0.99977832	0.998698764	0.369699224
Product performance	0.982501378	0.937967682	0.953649935	0.168945331
Product reputation and popularity	0.987759253	0.994715984	0.880728186	0.417933997
Product price	0.897566985	0.558722158	0.916590284	0.023395588
Product appearance	0.999822018	0.99991894	0.995168767	0.377061673
Logistics speed	0.898754542	0.803945522	0.761241544	0.299764822
Packaging quality	0.818480662	0.996707819	0.798063158	0.31340669
Courier attitude	0.940978078	0.6	0.785714286	0.090239854
Return mechanism	0.864864865	0.868381241	0.857142857	0.409392559
Consultation service	0.75	0.712121212	0.666666667	0.386257752
Complaint handling	0.855160866	0.691355744	0.923245614	0.066068319
Z_j	2.65749945	2.617672164	2.513728154	
$f(a)$	20.93			

4 Conclusions

In this paper, it use the network information acquisition and processing tool to collect and process the comment data of China's leading e-commerce platforms. Thereby constructing a customer satisfaction index system based on real customer reviews of electronic business platform and gets the satisfaction of each index value, while using PP model and genetic algorithm satisfaction index value for each further calculation processing. In this process, we draw the following conclusions and recommendations:

1 Customer satisfaction index derived from the true evaluation has the certain degree of influence on the overall satisfaction of e-commerce platform, therefore the resulting satisfaction index for each e- a strong guiding significance, and each e-commerce platform can start to improve their customer satisfaction by considered from these indicators.

2 Different e-commerce platforms have different strengths and shortcomings, when it is working to improve customer satisfaction, it should focus on different points, such as T-mall can focus on the pricing strategy, in order to improve their satisfaction with price; Amazon, can focus on their packaging quality.

Acknowledgements

This research was supported by Zhejiang Provincial Natural Science Foundation of China (LY15G020007, LY15G020001); National Natural Science Foundation of China (50575083).

References

Cai, X., Gao, X-z. and Xue, Y. (2016) 'Improved bat algorithm with optimal forage strategy and random disturbance strategy', *International Journal of Bio-Inspired Computation*, Vol. 8, No. 4, pp.205–214.

Cai, X., Wang, H., Cui, Z., Cai, J., Xue, Y. and Wang, L. (2018) 'Bat algorithm with triangle-flipping strategy for numerical optimization', *International Journal of Machine Learning and Cybernetics*, Vol. 9, No. 2, pp.199–215.

Cui, Z., Cao, Y., Cai, X., Cai, J. and Chen, J. (2017) 'Optimal LEACH protocol with modified bat algorithm for big data sensing systems in internet of things', *Journal of Parallel and Distributed Computing*, DOI:10.1016/j.jpdc.2017.12.014.

Wang, G., Cai, X., Cui, Z., Min, G. and Chen, J. (2017) 'High performance computing for cyber physical social systems by using evolutionary multi-objective optimization algorithm', *IEEE Transactions on Emerging Topics in Computing*, DOI:10.1109/TETC.2017.2703784.

Zhang, M., Wang, H., Cui, Z. and Chen, J. (2017) 'Hybrid multi-objective cuckoo search with dynamical local search', *Memetic Computing*, Vol. 10, No. 2, pp.199–208.