Principle and Applications of Big Data Consumer Behavior Analysis

Zilin Xu*

Shanghai Nanhui High School, Shanghai, China

*Corresponding author: doliver79865@student.napavalley.edu

Abstract. As a matter of fact, with the rapid development of computation ability as well as machine learning approaches, the usage of big data has long been a theme that has attracted researchers from different fields. To be specific, it includes practical applications in the fields of economy, education, as well as market development. With this in mind, in this article, the author explores the actual use of big data in terms of consumer behavior analysis through different research methods of big data to understand big data. The principle, utilized models as well as procedure for big data analysis will be demonstrated in detail. At the same time, the current limitations of different models will be discussed. Finally, the study puts forward the problems faced by big data and the views on future big data development. Overall, these results shed light on guiding further exploration of consumer analysis based on the state-of-art bigdata analysis scenarios.

Keywords: Big data; market; customer behavior.

1. Introduction

Big data has played a huge role in collecting and analyzing consumer behavior. Big data technology can achieve the past four aspects in the following four aspects that are impossible to analyze consumer behavior in the past: The full record of the Internet, understand the complete traces; search for data, reflect the attention and needs; the data of social media, judge personality and type; online payment, display actual purchase behavior [1].

The processing method of big data has a batch data processing system (batch data mining can get specific meanings, make wise decisions, and eventually make effective measures to achieve business purposes. This is the urgency of dealing with big data. It is suitable for the scenes that are not high in real time, but the accuracy and comprehensive requirements of data are calculated before storage and then calculated); Streaming data processing system (the goal is to shorten the processing time of PB data from minutes to seconds); interactive data processing (its data processing method is also flexible, intuitive, and intuitive. The characteristics of easy control, at the same time, the processing time is also through human -machine dialogue to make the system interact with the operator. The operator issued a request to enter the data through the dialogue, and guide the operator to gradually complete the required operation as required. The system will provide corresponding data or prompt information until the final processing result is obtained); graphic data processing system (it has become a hot spot for research on various disciplines because of its structural characteristics. Due to the close connection between the data processing system, the data processing system needs to perform a series of operations on the points and edges in the figure. These functions include the storage, query map, the shortest path query, keyword query, diagram mode mining, and classification of graph data, including the storage, query, and other functions of graph data. data processing). These functions include the storage, query map, the shortest path query, keyword query, diagram mode mining, and classification of graph data, including the storage, query, and other functions of graph data. data processing). Psychologist Luce and Statisticalist Tukey further developed the method of common analysis and applied it to the field of social sciences in the field of investigation. Wind and Jain successfully used this method in the business field [2]. The production and development of joint analysis, the same cluster analysis and multi-dimensional standard analysis (Multi-Development analysis), thanks to the rapid development of consumer behavior [3]. By using data to determine target consumer groups with similar requirements, attitudes, and interests, cluster analysis can solve the problems of market

segmentation. Brand diagnosis adopts multi -dimensional scale analysis method to build cognitive and preference maps [4, 5].

2. Application in Economic Models

The author collects information by issuing questionnaires and the data of previous data research. A total of 150 cases were recovered in this questionnaire, and 138 valid questionnaires were recovered, and the recovery ratio was 92%. Data processing is mainly descriptive statistical analysis and interpretation of related measurement economic models. The main survey is shown in Table 1.

Table 1. Feature of sample

Item	**Standard of statistics**	**Quantity**	**Ratio**
Gender of consumer	Male	94	68.1
	Female	44	31.1
Range of age	20~30	36	26.1
	31~40	44	31.9
	41~50	39	28.3
	51~60	14	10.1
	Above 61	5	3.6
Level of education	Primary school	12	8.7
	Junior high school	23	16.7
	High school	37	26.7
	Associate degree or above	66	47.7
Income per month	Under 5RMB	9	6.5
	501~1000RMB	28	20.3
	1001~2000RMB	61	44.2
	2001~3000RMB	23	16.7
	Above 3001RMB	17	12.3
Occupation	Catering service industry	8	5.8
	Medical hygiene	5	3.6
	Governmental agencies	32	23.2
	Educational research institutions	5	3.6
	Other	88	63.8
Status of marriage	Unmarried	35	25.4
	Married	103	74.6
Place of dwelling	Countryside	55	39.9
	City	83	60.1

The main group of consumer green pork can be inferred as urban consumers through the proportion of survey objects. Consumer behavior theory believes that television is still the main medium for consumers to understand environmental pork, even if consumers purchased the medium in recent years, they have replaced TV in recent years. It should be noted that 24.6%of those who understand the environmental protection knowledge of pork through the Internet. The main group of consumer green pork can be inferred as urban consumers through the proportion of survey objects. Consumer behavior theory believes that television is still the main medium for consumers to understand environmental pork, even if consumers purchased the medium in recent years, they have replaced TV in recent years. It should be noted that 24.6%of those who understand the environmental protection knowledge of pork through the Internet.

Table 2. Consuming frequency of pollution-free pork of consumer

Consuming frequency	Almost every day	Every other day	Once every three days	Once every five days	Once a week
Quantity	90	30	8	4	6
Percentage	65.3	21.7	5.8	2.9	4.3

Table 3. The quantity of pollution free pork per time

Consuming quantity	Under 50g	50~100g	100~150g	150~200g	Above 200g
Quantity	18	39	32	36	13
Percentage	13	28.3	23.2	26.1	9.4

Table 4. The place of purchasing pollution free pork of consumer

Item	Special counter in agricultural market fair	Big supermarkets	Food store of pork	Sales point in suburb	Total
Quantity	72	20	11	35	138
Percentage	52.2	14.5	7.9	25.4	100.0

According to the survey data analysis (Table 2, Table 3 and Table 4), consumers have a ratio of 65.3%of pork consumption almost every day, the percentage of pork consumption the next day reached 21.7%, and the two increased to 87%. It shows that consumer environmentally friendly pork has become an indispensable part of consumers' daily diet, and with the advent of the era of well -off life of residents. And 49.3%of consumers spending 100g200g each time. Consumers spend a lot of environmentally friendly pork, which accounts for almost half of this proportion. 52.2%of consumers choose to buy at counters in the large -scale farmer's market, which indicates that consumers generally buy food such as vegetables and other foods without buying environmentally friendly pork. The survey results show that the safety of pork, the convenience of purchasing, the rationality of the price, the attitude of service, and the shopping environment. Consumers will give priority to the following factors when choosing to buy pork locations. At present, the procurement ratio of large supermarkets is 14.5%. It is expected that more and more consumers who choose to buy environmentally friendly pork in large supermarkets in the future will be more and more, because there are more agricultural products entering the supermarket. With the continuous acceleration of urban and rural integration, the proportion of urban residents who choose to buy pork at a fixed -point sales point of pork reaches 25.4%, and it is expected to continue to increase.

In consumer practice, there are many factors that affect consumer willingness. Comprehensive related research and close combination of actual surveys, this article selects information satisfaction, monthly income, cultural level, age, gender and other factors that are closely related to environmental pork consumption. analyze. In the model, information satisfaction, age, cultural level, per capita monthly income and gender as an interpretation variable, regarding the willingness to pay for environmental pork as a variable. So one can analyze ordinary pork and environmentally friendly pork through the logit model:

$$\text{In}\left[\frac{p(Environmentally\ friendly\ pork)}{p(Ordinary\ pork)}\right] = \alpha + \sum_{k=1}^{k} \beta_k x_k \quad (1)$$

Among them: αis the constant item, β_k is the regression coefficient, and x_k is the influencing factor of the willingness to pay for environmental pork. The R-Squared of this model is 0.9635, F = 711.367, which shows that the model can fit the sample data well. According to the results of the Logit analysis of Table 5, the following judgments can be made: the WALD statistics of information satisfaction variables are significant at 0.01, indicating that information satisfaction is an important reason for consumers to pay for environmental pork. Consumers generally attach great importance to the quality and safety of pork, but the consumer's understanding of environmentally friendly pork is

not enough. Although more than 70%of consumers know the existence of green signs, only 30%of the meaning of environmentally friendly pork can really accurately understand the meaning of environmentally friendly pork. Pork without feeding, no veterinary drugs, and water not injection cannot be identified carefully when purchasing. Some consumers have misunderstandings about environmental pork standards. People have bias for environmental protection pork, because the quality and safety of food are not guaranteed. Most consumers have a strong willingness to buy environmentally friendly pork, showing that they attach great importance to the safety of pork consumption and are willing to pay a higher price for the purchase of environmentally friendly pork. Compared with the price of ordinary pork, about 70%of consumers can accept 10%to 20%of prices. It mainly includes information satisfaction, income level, cultural level, age and other factors that affect consumers' willingness to buy environmentally friendly pork. Among them, the greatest influence on consumer payment willingness is information satisfaction, income level, and cultural level [6].

Table 5. Description of explanatory variable of the model

Name of variables	Definition of variable	Average of sample	Standard deviation of simple
Satisfying level of information	Satisfaction with the provision of pork quality and safety information by governments, enterprises, etc.: satisfied(1) ~ not satisfied(3)	2.19	1.155
Age	20~30 years old=1, 30~40 years old=2,40~50 years old=3, 50~60 years old=4, above 60 years old=5	2.21	1.088
Level of education	Primary school=1, Junior high school=2, High school=3, Associate degree or above=4	3.11	2.036
Income per month	Under 5RMB=1, 501~1000RMB=2, 1001~2000RMB=3, 2001~3000RMB=4, Above 3001RMB=5	3.45	1.522
Sex	Male=1, Female=0	0.73	0.21

3. Application in CRM

In the research report, the author focused on the potential impact of the key success factor (CSF) of big data (BD) on customer relationship management (CRM). The report is centered on the concept and guides the concept of the author's framework construction to determine the key area of CRM. The purpose is to solve the inspiration of CSF in the development and management of the CRM plan. This framework consists of technology, data and insights, projects and organizations -it is developed based on the five CRM perspectives described by Zablah, Bellenger, and Johnston. And in the six main steps in the six stages of in -depth literature review: (i) identify keywords and search string; (ii) definition of time window; (iii) citation database selection; (IV) subject/theme selection; (V) Refine the results through the multi -stage reading process; (VI) review related papers.

Table 6 shows the scheme that guides the author's literature review the first four steps. Excluding CSFs outside the keywords enables us to consider a broader paper set to determine the factors.The author created five logical nodes based on Table 6 for each category and distributed the paper, which includes the review references for each macro category.Through Bur logical combination keywords, if the distance between them does not exceed N words, the software can highlight their common appearance in the text text. Based on the existing BD literature analysis of the key evidence generated in the previous steps. Therefore, in order to explain the influence of BD on CRM CSFS, and five theoretical propositions were proposed and discussed. As far as the logic of the dissertation is concerned, the number of propositions is consistent with the number of CRM macro categories, and the CRM macro category originated from five CRM inputs. Those who are in. In order to obtain the

proposition from the literature review, the author was randomly divided into two pairs, avoiding the same coupling of the second step.

Table 6. Parameters for CRM.

PARAMETER	DETAIL	JUSTIFICATION
KEYWORDS	Customer Relationship Management' in Title, Abstract, and Keywords OR CRM in Title OR CRM in Keywords.	One tends to explore the entire CRM theme and reasonably extract explicit and implicit SF from the paper. Therefore, although the purpose of the review was to determine CSF, the keyword CF was not considered.
TIME WINDOW	Jan 2003~Feb 2016	The logical continuity of Ngai's (2005) CRM review has a time window from 1992 to 2002.
TYPE OF ARTICLE	Journal (J), reviews, Conference (C)	Because journals are considered the best scientific source; According to Webster and Watson (2002), selected meeting minutes should be reviewed, especially those noted for their quality.
SOURCES	Scopus	Due to its better flexibility, it is favored by ISI Science Network (WoS). All main results are also found on WoS.
DISCIPLINES	Business, Management and Accounting; Computer Science, Engineering; Social Sciences: Decision Sciences; Economics, Econometrics and Finance.	Select disciplines based on their alignment with research objectives.

Each couple reviewed the evidence of 20 CSF independently and summarized their possible contact with BD literature. The results of the two teams were compared and refined until consensus was reached. As another result, the review shows that scientific literature is ignored, and the relationship between CRM and other business processes may be underestimated. CSF is connected with possible use of non -alternative and semi -structured data, especially in terms of social media data, without considering the clear relationship between CSF and BD of the classic CRM. According to analysis, one proposed five propositions about the five potential effects of the BD method on the five types of CRM CSF. These propositions are proposed by dealing with the corresponding CSF [7].

4. Limitations and Prospects

In such a large amount of data, some of them is useless. In most cases, useful ones is even a small part of them. As the amount of data continues to increase, more and more redundant and spam data will be increasing, which will be faster than data information. In this way, it is difficult for us to judge its authenticity and practicality. The timeliness of big data is very short, often new data will continue to replace the old data. Therefore, in order to ensure the accuracy of the data, the real -time collection, real -time processing, and real -time distribution of data are required. This is difficult to do in many times, thereby limiting the application of big data to a fairly extent [8, 9].

Big data is very likely to involve personal privacy, commercial secrets, public rights and national information security in the process of collection and use. Therefore, the problem of security and legitimacy constitutes one of the restrictive factors for the value of big data. Privacy issues often cause public concerns. Just as Baidu once sold multiple stickers for short -term commercial interests, it publicized public privacy, and was pointed out by Qianfu. However, if it does not violate the privacy of individuals and the country, big data does not have relatively accuracy, so how to maintain the balance between the two becomes a question worth thinking about. From this point of view, one has

to face these issues in the future. Therefore, one needs to measure the complexity of big data and explore new operating modes, and study how to ensure the accuracy of data when ensuring personal privacy [10].

5. Conclusion

To sum up, the advent of the big data era has brought great changes to the lives. When it brings us a lot of convenience, it also has a huge problem. In fact, this study focuses on studying the role of big data on customer behavior and starting from different market aspects to explore the actual application of big data. Overall, big data promotes the collection and circulation of information, but it is necessary to continue to explore solutions in terms of timeliness and privacy. Overall

References

[1] Lu T. 50 Years of Consumer Behavior: Evolution and Disruption. Foreign Economics and Management, 2017, 39(6): 23-38.

[2] Green P, Srinivasan V. Conjoint analysis in consumer research: issues and outlook. Journal of Consumer Research, 1978, 5: 103~123

[3] Carroll D, Arabie P, Chaturvedi A. Multidimensional Scaling and Clustering and their Applications. Advances in Marketing Research: Progress and Prospects, 2002.

[4] Sun X, Chen Y. The Joint Analysis Method in Consumer Behavior Research. Consumer Analysis, 2005, 17.

[5] Cheng X Q, Jin X L, Wang Y Z, GUO J F, Zhang T Y, Li G J. Survey on big data system and analytic technology. Ruan Jian Xue Bao/Journal of Software, 2014, 25(9): 1240-1252.

[6] Ren D, Qi Y. Consumer Behavior Analysis of Environmentally Friendly Pork. Journal of Sichuan Agricultural University, 2017, 27(03): 371-376.

[7] Zerbino P, Aloini D, Dulmin R, et al. Big data-enabled customer relationship management: A holistic approach. Information Processing & Management, 2018, 54(5): 818-846.

[8] Roma P G, Reed D D, DiGennaro Reed F D, Hursh S R. Progress of and prospects for hypothetical purchase task questionnaires in consumer behavior analysis and public policy. The Behavior Analyst, 2017, 40: 329-342.

[9] Vogel K M. Big data and biodefense: Prospects and pitfalls." Defense Against Biological Attacks, 2019, 1: 297-315.

[10] Kraljevski I, Constanze T, Matthias W. Limits and Prospects of Big Data and Small Data Approaches in AI Applications. KI-Kritik/AI Critique Volume 2023, 4: 115.