

Research on big data marketing method based on statistical machine learning algorithm

Yuran Wang

university of liverpool

sgywa159@liverpool.ac.uk

Abstract: In the innovation and development of human science and technology, from the simple relationship between buying and selling to today's huge and complex online trading, not only reflects the main content of the progress of The Times, but also indirectly proves the diversified characteristics of marketing methods and technological means. In the marketing market at home and abroad, database marketing, with high efficiency and low cost, has been paid more attention to by the marketing staff, and gradually become the focus of the marketing market. As database marketing belongs to the foreknowledge and ferryman in the increasingly competitive market environment, it is more inclined to analyze simple data before application in practice. Therefore, in contrast, the database with machine learning algorithm as the core, marketing is the main content of marketing research and discussion. Based on the analysis of big data marketing methods and application effects in recent years, this paper mainly discusses the main content of statistical machine learning algorithms, and takes a Tmall flagship store as the research target to systematically understand the database marketing content with machine learning algorithms as the core. The final results show that the flagship store is mainly based on the database marketing method of model prediction, and the effective measures of how to improve the marketing effect with the help of model prediction are put forward.

Keywords: Machine learning algorithm; Big data; Marketing methods; Model prediction; Database Marketing

1. Introduction

Since the 1990s, database marketing has been applied to marketing management activities by various enterprises while rapidly adapting to the environment of modern information society. Especially after entering the era of big data, database marketing has been paid more attention by new retail and e-commerce industries. The rich consumer behavior data contained in it provides a basis for database marketing research, and the machine learning algorithm and data mining technology provide a basis for e-shop, which can be reasonably used in almost every link of all marketing activities. However, from the perspective of overall development, due to the increasing amount of network information and data, information overload will affect the efficiency and quality of consumers' access to the platform. At this time, businesses should quickly find more effective strategies to accurately locate the target consumer groups, so as to cope with the competitive challenges brought by market development. One of the most representative is the database marketing application strategy. Compared with database marketing in the traditional sense, database marketing based on machine learning algorithms can help enterprises and consumers build a stable relationship, narrow the cognitive differences between them, and provide more precise decision-making direction and development strategy. In the innovation and development of network technology and e-commerce platforms, most e-commerce platform stores have the same type of products, leading to the dilution of business in the market. In order to occupy a powerful position in the increasingly competitive market environment, electric stores should learn to accurately identify the target groups in the huge consumer groups, so as to improve their marketing efficiency.[1.2.3]

Enterprise development and e-commerce profit are inseparable from the precise positioning of target groups, but because there are many types of goods and services on e-commerce platform, it brings more challenges to such precise database marketing. Therefore, database marketing has become a major issue for domestic and foreign researchers and enterprises. As the core content of

the big data revolution, consumer analytics is conducive to the rapid grasp of consumer behavior data. However, the pioneer of relevant research is developed countries, and relevant scholars have applied data mining technology and machine learning algorithms to marketing and other fields. However, China should combine relevant theoretical guidance and practical experience to provide effective reference for the innovation and development of enterprise e-commerce platform. For example, Knott et al. determined the relationship between the positioning accuracy of target customers and corporate earnings based on empirical analysis when they learned about retail business banks, which proved that in the process of database marketing, in order to further improve the marketing performance of enterprises, the most important thing is to ensure the positioning accuracy of target customers. Chapelle et al. proposed the use of logistic regression model to estimate click-through rate of advertisements and constructed the corresponding machine learning framework when solving the problem of click-through rate estimation. Amy et al. proposed an individual level dynamic model by learning the behavior data of users adding products to the shopping cart in real time, which is mainly used to understand the user's choice intention, so as to dig into its commercial value. The model structure was determined and analyzed in the practical study, mainly testing the application effect of different network marketing methods in different scenarios, and converting corresponding web pages through various materials, so as to reduce the structure rate and improve the conversion rate.[4.5]

Compared with foreign research results, domestic research on database marketing is in the initial stage. User data analysis is gradually transformed from simple data to machine learning algorithms, and a few industries are popularized to other industries. For example, Wang Qi used computer technology to collect a large amount of consumer information, such as the degree of product preference, commodity purchase price, consumption frequency, and so on, and integrated processing and analysis with the industry information, and finally formulated the enterprise marketing strategy scientifically according to the analysis results. Ou Pengyi et al. used data mining technology to accurately identify the responsive customers among dormant customers, and established a variety of algorithm models based on logistic regression and undersampling support vector machine under the condition of fixed wake up cost and data imbalance characteristics, and selected 50,000 target dormant customers for model prediction. The optimal E-mail delivery strategy is proposed. On the basis of understanding the accumulated experience of big data marketing in recent years, this paper analyzes the big data marketing method with machine learning algorithm as the core combined with empirical cases.[7.8]

2. Methods

2.1 Big Data Analytics

The definition of big data starts from two aspects: on the one hand, it refers to the data that cannot be loaded into the computer memory; on the other hand, it refers to the use of traditional computer technology and software and hardware to perceive and collect data within the tolerable time of data. From the perspective of current technology application, the characteristics of big data from the original diversity, volume, speed gradually developed into virtualization, value, efficiency, etc., only by integrating the data and computer technology together, can we fundamentally improve the level of data processing. The technical structure of big data is shown in Figure 1 below:[9.10]

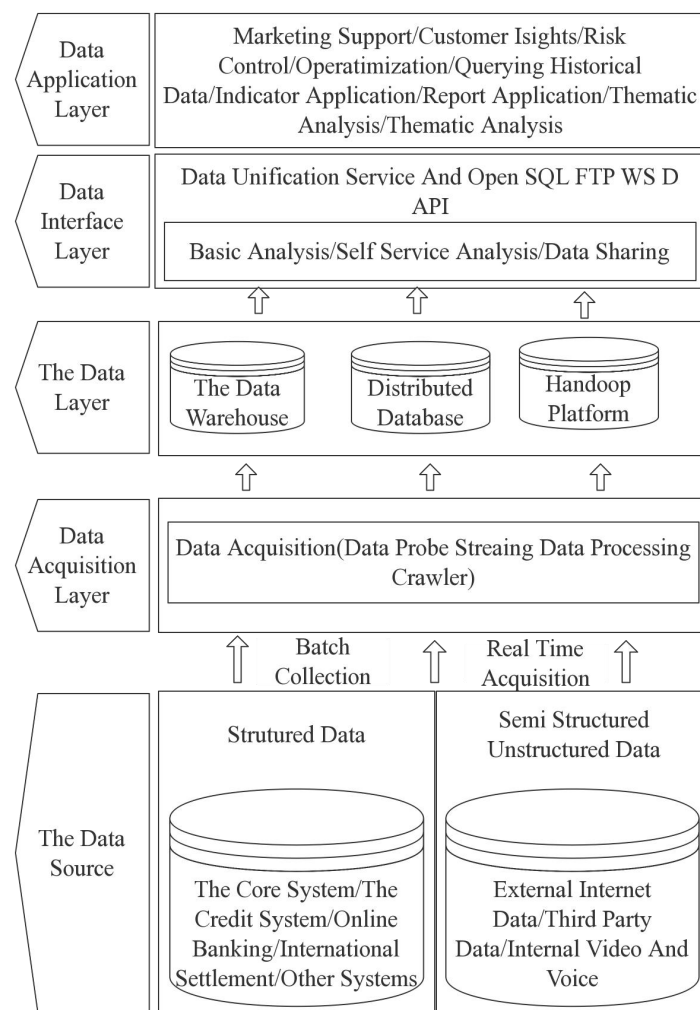


Figure 1 Structure diagram of big data technology

2.2 Research direction of machine learning

First, algorithm C 4.5. Is currently the most widely used in machine learning algorithms of the content, it belongs to the classification of the decision tree algorithm, is based on the algorithm are improved, the mapping of the target attribute, every node on the decision tree represents a goal, m diameter on behalf of all the target attribute value, each different decision tree is a tree structure, The target attributes can be classified and processed according to the branch type. At the same time, the data can be segmented and tested according to the database.

Secondly, k-means algorithm. In machine learning, after data analysis and sorting, the data with high similarity is found in the machine for collection. As long as the similarity between the data can be accurately analyzed, the effectiveness of machine learning can be fundamentally guaranteed. This algorithm is regarded as a typical modified iterative dynamic clustering algorithm, which is very fast in practice. It can be used to calculate and analyze K-average algorithm several times, and then select the optimal result according to the actual situation.

Finally, support vector machine algorithm. In the big data environment, the amount of data information is large, and the actual processing work is tedious, so the support vector machine algorithm can be used for operation. In essence, this kind of algorithm belongs to supervised learning algorithm, which can learn the complexity of samples, gradually improve its generalization ability, and minimize the error and maximize the edge region.

2.3 Database Marketing

Combined with the analysis of the operating procedures shown in Figure 2 below, it can be seen that database marketing can not only comprehensively monitor quality and effectively manage it, but also treat information technology as a development medium and reach a new height in step-by-step practice testing. Nowadays, when excluding the influencing factors of the market, database marketing can abstract and quantify the problems in the market, so as to make more accurate classification and positioning, and develop efficient marketing programs. At the same time, database marketing and enterprise consumers to build a sense of intimacy and trust, so the current enterprise marketing personnel should start from the perspective of human, in-depth discussion of the relationship between human and machine, data and market, from which to build a personalized marketing system.

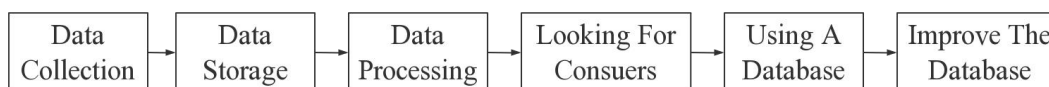


Figure 2 Database marketing operation procedure diagram

2.4 Prediction model

In this paper, we study a certain Tmall flagship store, in 2019, 618 during the activity, on the basis of the consumer purchase behavior data system to sort out the relevant portraits of consumer information, transaction information, browse the log, product data, construct the prediction model with machine learning algorithms for the core, and in data cleaning, remove duplicate data and error correcting technology, In this way, the data information is guaranteed to be perfect and accurate, which provides technical support for effective training and learning model. At the same time, the characteristics of variable selection as difficult point of model building, characteristics of the variable combination effect directly affects the performance of the application of classification algorithms, therefore, this paper studies from the perspective of the consumer, this paper puts forward the consumers to interact, to purchase goods diversity, taobao platform activity, buy goods carefully degrees, the sensitivity of the promotional activities, such as demographic attributes selected characteristics in many aspects. In addition, when output prediction model results, the confusion matrix shown in Table 1 below should be constructed, where TP represents the prediction of positive samples and the actual result of positive samples. FN stands for the result that the predicted negative samples are actually positive samples; FP represents the result that the prediction is positive and the sample is actually negative. TN stands for the result that the prediction is a negative sample and the actual result is a negative sample:

Table 1 Confusion matrix

	Is kind of sample	Negative samples	The total number of samples
Prediction is class	TP(true positive)	FP(false positive)	A(predicted to be positive)
Predict negative class	FN(false positive)	TN(true positive)	B(predicted to be negative)
Predicting samples	C (actually positive class)	D(actually negative class)	E

The correct rate is calculated based on the above table, and the specific formula is as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

3. Result analysis

After data preprocessing, received a total of 203516 data, divided them into training set and test set, using the preconfigured training data collection using logistic regression, random forests, support vector machine (SVM), the GBDT model training, can be combined with the confusion matrix and the model of evaluation index to judge the applicability of machine learning algorithm model, This shows the application effect of various models. Among them, Table 2 represents the results of logistic regression model test data set, Table 3 represents the results of random forest model test data set, Table 4 represents the results of support vector machine model test data set, Table 5GBDT model test data set:

Table 2 Logistic regression model test data set results

	Correct	Accuracy	The recall rate	F1 value	AUC
Logistic regression L1	0.8000	0.7062	0.4457	0.5465	0.8381
Logistic regression L2	0.8565	0.8365	0.5860	0.6892	0.9028

Table 3 Test data set results of random forest model

	Correct	Accuracy	The recall rate	F1 value	AUC
Random Forest 1	0.7438	0.7314	0.0523	0.0976	0.7821
Random Forest 2	0.7854	0.7177	0.1838	0.2927	0.7794
Random Forest 3	0.6970	0.6718	0.2433	0.3572	0.7345

Table 4 Test data set results of support vector machine mode

	Correct	Accuracy	The recall rate	F1 value	AUC
Support Vector Machine 1	0.7943	0.6547	0.2821	0.3943	0.6431
Support Vector Machine 2	0.7686	0.6292	0.3210	0.4251	0.6572
Support Vector Machine 3	0.7111	0.6267	0.3983	0.4871	0.6687

Table 5 Results of GBDT model test data set

	Correct	Accuracy	The recall rate	F1 value	AUC
GBDT1	0.8200	0.7199	0.5026	0.5919	0.8328
GBDT2	0.8920	0.8305	0.7503	0.7884	0.9430
GBDT3	0.8969	0.8426	0.7638	0.8013	0.9483
GBDT4	0.8255	0.7900	0.4758	0.5939	0.8267

Combined with the above research results, it is shown that the test data set results of GBDT model are optimal. Therefore, in this study of a Tmall flagship store, it is necessary to continue to optimize other database marketing models, pay attention to auxiliary model prediction to improve marketing efficiency, and take machine learning algorithm as a breakthrough for practical discussion, so as to propose more valuable marketing measures.

4. Conclusion

In summary, based on the systematic understanding of the research status of big data marketing, the purchase prediction model is constructed according to the machine learning algorithm, and the application effects of the four machine learning algorithms proposed in this paper are systematically studied in the operation steps of data acquisition, preprocessing, selection of feature variables, model construction and so on. The final results show that the GBDT model has the best test data set. Therefore, when promoting big data marketing, we should continue to study big data marketing mode with machine learning algorithm as the core, so as to provide effective basis for the development of Chinese e-commerce enterprises.

Reference

- [1] Zhaohua Ji , Xiaohua Zhang, Xinhui Yan. Research and discussion on machine learning algorithm based on big data technology [J]. 2021(2020-15):24-25.
- [2] Qian Li , Shuo Wang , Jixue Sui. Research on evaluation method of campus network students' online behavior based on machine learning algorithm [J]. Science and Technology Innovation and Application, 2021, 11(33):6.
- [3] Yan Wang, Haizeng Zhang, Xinhua Hu , et al. Dynamic Energy Consumption Prediction Model of Thermal Station based on big data Pattern recognition machine learning algorithm [J]. Guizhou Electric Power Technology, 2020, 023(004):47-53.
- [4] YanWang , Haizeng Zhang, Xinhua Hu, et al. Dynamic Energy Consumption Prediction Model of Thermal Station based on big data Pattern recognition machine learning algorithm [J]. Electric Power Big Data, 2020, 23(4):7.
- [5] Mingzhuang Cai, XiangrenKong . Swarm Intelligence Big Data Optimization Based on Machine Learning and Hash Algorithm: A Case study of Vehicle Precision detection [J]. Electronic World, 2020(2):2.
- [6] Yinghui Pan. Research on machine learning algorithm under big data [J]. Computer Knowledge and Technology: Academic Edition, 2020, 16(32):3.
- [7] Suoping Zeng, Yanmei Lv, Jian Tian. Evolution Path and Application scenario of Intelligent Algorithm in urban form Optimization Research: Based on Citespace Knowledge Graph analysis [J]. Urban Issues, 2022(4):10.
- [8] Changfeng Fu. Research on Online Purchase behavior Prediction Based on Machine Learning Fusion Algorithm [J]. Journal of Beijing Institute of Printing, 2020, 28(S01):3.
- [9] Zhen Li, Haiyan Liu, Ce Liu, et al. Research on K-means clustering optimization algorithm based on machine learning [J]. Data Mining, 2022, 12(1):7.
- [10] Zhaohua Ji, Xiaohua Zhang, Xinhui Yan. Research and discussion on machine learning algorithm based on big data technology [J]. Science and Technology Information, 2020, 18(15):2.