Information Systems and Economics (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/infse.2024.050210
ISSN 2523-6407 Vol. 5 Num. 2

Data-driven marketing strategy and consumer behavior prediction model construction

Lirong Wan[1], Jiayong Xu[1,2,*]

[1]The Graduate University of Mongolia (GUM), Ulaanbaatar, 14200, Mongolia
[2]Beijing Polwision Technology Development Co., Ltd., Beijing, 100101, China
**Corresponding author*

Keywords: Data-driven marketing; consumer behavior prediction; machine learning; data mining

Abstract: With the advent of the era of big data, the data-driven marketing strategy has gradually become the key to enterprise competition. This paper first introduces the basic concepts of data-driven marketing and its importance, and then details how to build a consumer behavior prediction model. In the process of model building, we used various machine learning and data analysis techniques, including data mining, feature engineering, model selection and optimization, and more. Through example analysis, we show how to use these models to guide the development of marketing strategies, and explore the challenges and prospects of the model in practical application. This paper aims to provide useful reference and guidance for enterprises in the era of data-driven marketing.

1. Introduction

In the digital economy era, data has become the core element of enterprise decision-making. Data-driven marketing strategy is to make the use of advanced data analysis technology, deep mining consumer behavior data, insight into consumer demand, so as to develop more accurate and personalized marketing strategy. This strategy can not only help enterprises improve marketing efficiency, reduce costs, but also effectively enhance consumer satisfaction and loyalty, for enterprises to win an advantage in the fierce market competition.

Consumer behavior prediction model is the key support of data-driven marketing strategy. By integrating all kinds of consumer data and using advanced algorithms such as machine learning and deep learning, the model can achieve accurate prediction of consumer behavior. This prediction can not only help enterprises grasp the market trend and layout in advance, but also provide a strong basis for enterprises to make decisions and ensure the pertinacity and effectiveness of marketing strategy.

With the continuous development and popularization of big data technology, data-driven marketing strategy and consumer behavior prediction model construction have become an important direction of enterprise marketing innovation. This paper aims to discuss how to use big data technology to build a consumer behavior prediction model, and how to apply the model to the formulation and optimization of marketing strategies, so as to provide strong support for enterprises to achieve their marketing goals.

In the following sections, we will detail the methods of data collection and processing, the process of consumer behavior analysis, the technical details of the prediction model construction, and the strategies for marketing strategy development and optimization. At the same time, we will also show the effects and challenges of data-driven marketing strategy and consumer behavior prediction model in practical application through case analysis. We believe that through the elaboration of this article, readers will be able to have a deeper understanding of the connotation and value of data-driven marketing, and provide a useful reference for enterprise marketing innovation.

2. Data collection and processing

2.1 Selection and integration of data sources

In the process of constructing the consumer behavior prediction model, the selection and integration of data sources is a crucial step. The quality, completeness, and diversity of the data directly affect the accuracy and reliability of the model. Therefore, when selecting data sources, we need to fully consider the source, quality, and availability of the data.

The choice of data sources should be broad and comprehensive. This includes internal sales data, customer data, product information, as well as external market research data, social media data, competitor data, etc. Internal business data provides us with critical information about consumer buying behavior, preferences, and loyalty, while external data helps us understand market trends, changing consumer demand, and competitor dynamics.

The quality of the data is also an important factor to consider when selecting the data sources. The data should be accurate, complete, and representative. For inaccurate or missing data, we need to conduct data cleaning and preprocessing to ensure the quality of the data. At the same time, we also need to pay attention to the timeliness of the data, and choose the latest and most relevant data as the model input.

When integrating different data sources, we need to consider the format, structure, and standards of the data. As the data from different sources may have problems such as inconsistent format and inconsistent structure, we need to transform the data format and unify the structure for subsequent data analysis and model building. In addition, we also need to focus on how the data is integrated. One common way is to fuse data from different sources to form a unified data set; another way is to associate and integrate data from different sources to leverage the data in subsequent analysis and modeling.

We also need to pay attention to the security and privacy protection of the data. When selecting and integrating data sources, we need to comply with relevant laws, regulations and ethics to ensure the legitimacy and security of the data. At the same time, we also need to take appropriate encryption and desensitization measures to protect consumers' privacy and data security. The selection and integration of data sources is a key step in constructing the prediction models of consumer behavior. We need to extensively and comprehensively select data sources to ensure the quality, integrity and diversity of data. Meanwhile, we also need to pay attention to the data integration and security issues to provide strong support for subsequent model building and analysis.

2.2 Data cleaning and preprocessing

Data cleaning and preprocessing are an indispensable part of data-driven marketing strategy. Their purpose is to ensure the quality of data and provide an accurate and reliable foundation for subsequent data analysis and model building. The process of data cleaning and pre-processing

involves many steps, including data weight removal, missing value processing, outlier detection and processing, data conversion and so on.

First, data weight removal is an important step in data cleaning. During data collection, duplicate data records may occur due to various reasons (such as data source duplication, data acquisition errors, etc.). These duplicates can interfere with subsequent data analysis and model building, thus requiring de-reprocessing. By comparing the key fields in the data records (such as ID, timestamp, etc.), the duplicate data records can be identified and deleted to ensure the uniqueness of the data set.

Secondly, missing value processing is also a key link of data cleaning. In actual data sets, missing values may occur due to various reasons (such as incomplete data collection, incorrect data entry, etc.). The presence of missing values can affect the quality of the data and the accuracy of the model and therefore require processing. Common methods of handling missing values include deleting records containing missing values, filling in with statistics such as mean, median or crowd, and estimating with interpolation methods. The choice of method depends on the characteristics of the data and the requirements of the model.

In addition, outlier detection and processing is also an important step in data cleaning. Outliers are data points that are clearly inconsistent with the overall data distribution, and they may be caused by data acquisition errors, data entry errors, or other reasons. The presence of outliers can interfere with the prediction results of the model and therefore need to be detected and processed. Common outlier detection methods include statistics-based methods (such as Z-score, IQR, etc.), graph-based methods (such as boxplots, scatter plots, etc.), and machine learning-based methods (such as isolated forest algorithms, etc.).

Finally, data transformation is an important link in data preprocessing. Since the data format, units and dimensions may be different for different data sources, data transformation is required in order to eliminate the impact of these differences on model construction. Common data conversion methods include normalization, standardization, discretization, encoding conversion, etc. Through data transformation, the data from different sources can be unified into the same dimension and format, so as to improve the generalization ability and prediction accuracy of the model.

In conclusion, data cleaning and preprocessing are an indispensable part of data-driven marketing strategies. Through removal, missing value processing, outlier detection and processing, and data transformation, the quality and reliability of data can be ensured, providing strong support for subsequent data analysis and model construction(Figure 1).

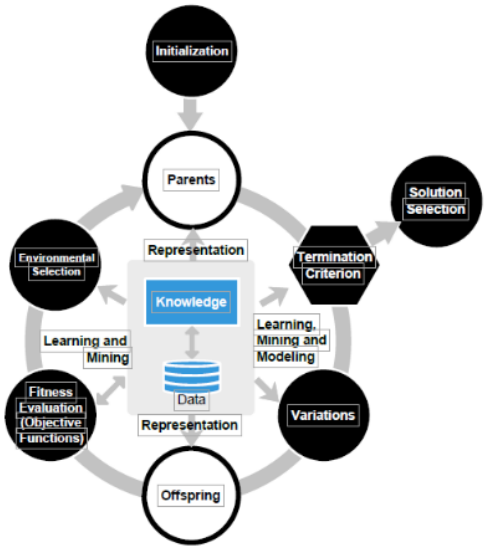

Figure 1: Data-driven optimization

2.3 Data exploration and visualization

Data exploration and visualization is a crucial part of a data-driven marketing strategy. It helps analysts and marketers to deeply understand and discover patterns and trends in the data, and then provide strong support for subsequent model building and strategy development[1].

Data exploration is the process of preliminary analysis of the data with the aim of understanding the overall structure, distribution and characteristics of the data. During the data exploration phase, analysts often use statistical methods, charts, and graphics to initially understand the data. For example, by calculating the mean, median, crowd and other statistics of the data, we can understand the concentration and distribution of the data; By drawing the histogram, boxplot and scatter plot, the charts of the distribution of the data and the relationship between variables can be visually displayed.

In the process of data exploration, analysts also need to pay attention to the data outliers, missing values and extreme values, etc., which may have an impact on the subsequent data analysis and model building. For outliers and missing values, the analyst needs to handle them according to the actual situation, such as deletion, filling or correction. At the same time, analysts need to be careful with extreme values to avoid excessive impact on the overall data analysis results.

Data visualization is about presenting the data in graphic or chart form to understand and analyze the data more intuitively. Data visualization can help analysts quickly discover patterns and trends in the data to better understand the data. Common data visualization tools include Excel, Tableau, Power BI, etc. These tools provide a rich range of chart types and visualization options that can help analysts choose the appropriate chart type and style to present the data as needed.

During data visualization, analysts need to be careful to choose the appropriate chart type and color matching to ensure the clarity and legibility of the chart. At the same time, analysts also need to pay attention to the details of the chart title, axis labels and legends to ensure that the chart can accurately convey the information of the data[2].

Through data exploration and visualization, analysts and marketers can have a deeper understanding of the data, discover the rules and trends in the data, and provide strong support for the subsequent data analysis and model building. At the same time, data visualization can also help team members to better communicate and cooperate, and jointly promote the development and implementation of data-driven marketing strategies (Figure 2).

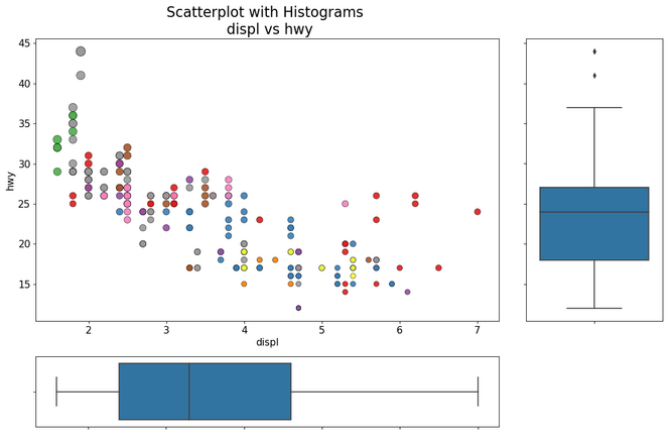

Figure 2: A python plot of the data visualization

3. Prediction model construction

3.1 Machine learning algorithm selection

The choice of machine learning algorithms is a crucial step in constructing prediction models of consumer behavior. Different machine learning algorithms are suitable for different types of data and problems, therefore, choosing the appropriate algorithm can significantly improve the prediction accuracy and effect of the model.

We need to specify the type of prediction task. Common prediction tasks include classification, regression, and clustering. Classification tasks are used to predict discrete variables, such as whether a consumer will buy a product; a regression task is used to predict continuous variables, such as the quantity or amount of consumer product purchases; and a clustering task is used to group similar consumers[3]. Based on the specific prediction task, we can choose a suitable machine learning algorithm.

We need to consider the properties of the data. The characteristics of data include the type, size, distribution and relationship between characteristics. For example, for large-scale data sets, we may need to choose algorithms that can handle large data, such as random forest or deep learning model; for data with nonlinear relationships, we can choose algorithms such as support vector machine or neural network.

We also need to consider the interpretability and robustness of the algorithm. Interpretability refers to the model can explain the reason and basis of the prediction results, which is very important for the development of marketing strategies. Some simple linear models, such as logistic regression and linear discriminant analysis, have good interpretability. Robustness refers to the model to handle noisy data and outliers. Some robust algorithms, such as the decision tree and the random forest, are able to resist the influence of noisy data and outliers to some extent.

When selecting machine learning algorithms, we can also perform model selection and hyperparameter tuning through methods such as cross-validation and grid search. Cross-validation divided the dataset into training and testing sets and repeated the training and testing process several times to assess the stability and generalization ability of the model. The grid search uses different combinations of hyperparameters to find hyperparameter configurations that optimize model performance.

In conclusion, the choice of machine learning algorithms is a process that comprehensively considers factors including the prediction task, data characteristics, interpretability, and robustness. Through reasonable algorithm selection, we can build an accurate, stable and interpretable consumer behavior prediction model, to provide strong support for the marketing strategy formulation of enterprises (Figure 3).

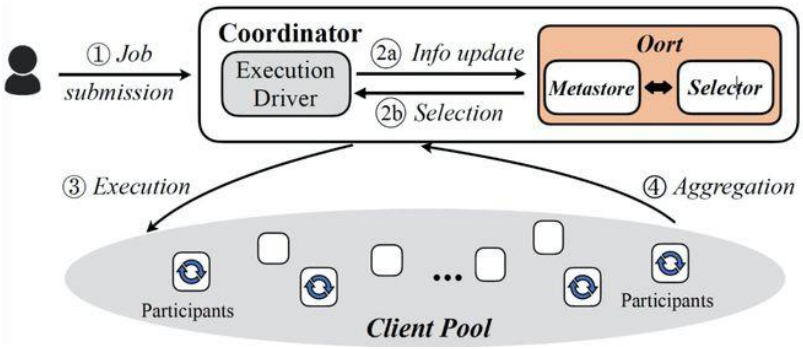

Figure 3: A- -The Markov top algorithm

3.2 Model training and optimization

After selecting the appropriate machine learning algorithm, the next key step is the training and optimization of the model. Model training is to apply the algorithm to real data to learn the rules and patterns in the data and to generate a model that can predict consumer behavior. The model optimization adjusts the parameters and structure of the model to improve the prediction accuracy and generalization ability of the model[4].

Model training usually involves the following steps:

Data division: divide the data set into training set, validation set and test set. The training set is used to train the model, the validation set to adjust the model parameters and hyperparameters, and the test set to evaluate the final performance of the model.

Parameter initialization: setting the initial value for the parameters of the model. These parameters are progressively adjusted by the optimization algorithm during training.

Forward propagation: the training data is input into the model, and the prediction output of the model is obtained through calculation. Loss calculation: calculate the difference between the predicted output and the actual label to obtain the loss function value. Backpropagation and optimization: According to the loss function value, the gradient of the parameters is calculated by the backpropagation algorithm, and the parameters are updated using the optimization algorithm (such as gradient descent, stochastic gradient descent, Adam, etc.). Iterative training: Repeat the above steps until the convergence condition is met or the preset number of iterations is reached. During model training, attention attention to overfitting and underfitting. Overfitting means that the model performs well on the training set but poorly on the test set, which is usually due to excessive complexity of the model or insufficient training data. Underfitting refers to the poor performance on both the training and test sets, which is usually due to the model or insufficient training data.

To solve this problem, we can use techniques such as regularization, integration learning, early stop method to prevent overfitting, or solve the underfitting problem by increasing the data and adjusting the model complexity.

Model optimization mainly focuses on how to improve the prediction accuracy and generalization ability of the model. Common model optimization methods include adjusting hyperparameters, using more complex model structures, and integrating multiple models. Hyperparameter optimization can be performed by methods such as grid search, random search, or Bayesian optimization.

In the process of model training and optimization, the model also needs to be evaluated and compared. Common evaluation indicators include accuracy, recall rate, F1 value, AUC, etc. By comparing the evaluation metrics of the different models, the model with the best performance can be selected.

In conclusion, model training and optimization are a key step in constructing a prediction model of consumer behavior[5]. Through reasonable training and optimization strategy, we can get an accurate and accurate model with strong generalization ability, which provides strong support for the formulation of marketing strategy of enterprises.

4. Conclusion

The data-driven marketing strategy combines modern data technology and marketing concept, and provides an accurate insight into consumer behavior to achieve a personalized and efficient marketing strategy. As the core tool, the consumer behavior prediction model uses machine learning algorithms to effectively integrate and analyze consumer data, which provides a scientific basis for enterprise decision-making. Through the continuous training and optimization of the model, enterprises can more accurately predict the market trend and formulate targeted marketing strategies,

so as to occupy an advantage in the fierce market competition. Data-driven marketing strategy not only improves marketing efficiency, but also enhances consumer satisfaction, and has become an indispensable marketing strategy for modern enterprises.

References

[1] Zhou Xu, Xu Yali, Guo Dong. Review of data-driven marketing strategies [J]. Modern Marketing, 2020 (2): 98-99.

[2] Li Linlin, Yang Lin, Zheng Yaru. Research on consumer behavior prediction model construction based on big data [J]. Business Research, 2019 (14): 121-122.

[3] Zhang Xiaofei, Wang Ming, Zhou Yan. Research on the application of data-driven marketing strategies in the e-commerce industry [J]. E-commerce Research, 2018 (3): 53-55.

[4] Zhu Yan, Zhang Ming. Construction and application of Consumer behavior prediction model based on big data [J]. Science and technology wind, 2017 (12): 94-96.

[5] Huang Ming, Chen Ting, Zhu Zhihong. Research on the impact of data-driven marketing strategies on consumer purchasing decisions [J]. Economic longitude, 2019 (17): 100-102.