DOI: 10.53469/wjimt.2025.08(03).04

Research on Consumer Behavior Prediction Model Based on Big Data Analysis

Shuai Qi

Sichuan Kerui New Laser Technology Co., Ltd. Chengdu 610299, Sichuan Province

Abstract: ***As the core issue in the business field, consumer behavior has always attracted the attention of researchers and practitioners. With the rapid development of information technology, especially the rapid development of Internet technology, the data scale has grown at an alarming rate, and big data has become the cornerstone of modern society. In various industries, the penetration of big data is ubiquitous, profoundly changing our way of life and business environment. For enterprises, accurately capturing and analyzing consumer behavior and needs is the key to developing efficient marketing strategies and improving sales performance. Therefore, in-depth research on consumer behavior prediction models based on big data is of great significance for enterprises to stand out in fierce market competition.***

Keywords: Big data analysis; Consumer behavior prediction model; Case analysis.

1. INTRODUCTION

With the rapid development of Internet technology, big data has become an important tool for enterprises to understand consumer behavior and needs. This paper first analyzes consumer behavior in the context of big data, and then discusses the principle and process of building consumer behavior prediction models based on machine learning and data mining technology. Through practical case analysis, the application effect of predictive models in formulating marketing strategies, product optimization, and improving consumer experience has been demonstrated. Finally, this article summarizes current research and looks forward to the future development trends of consumer behavior prediction models based on big data. Yao [1] conducted research on the local head loss coefficient in short-tube hydraulic testing, providing insights into fluid dynamics within hydraulic systems. Zhao et al. [2] evaluated labor market efficiency using machine learning and the DMP model, highlighting the impact of media news. Chen et al. [3] explored the green innovation effect of the digital economy, emphasizing its economic and environmental benefits. In the field of database management, Ukey et al. [4] proposed an efficient k-nearest neighbor (k-NN) join method for dynamic high-dimensional data. Peng et al. [5] introduced a dual-augmentor framework for domain generalization in 3D human pose estimation, enhancing robustness across diverse scenarios. Their subsequent work [6] presented a 3D vision-language Gaussian splatting technique, further advancing multimodal integration. Deng et al. [7] developed a transformer-based approach for real-time financial fraud detection, leveraging cloud computing. Zhou et al. [8] optimized an automated garbage recognition model using ResNet-50 and weakly supervised CNNs for sustainable urban development. Lyu et al. [9] optimized CNNs for rapid 3D point cloud object recognition, improving computational efficiency. Wang and Liang [10] applied reinforcement learning methods combined with graph neural networks and self-attention mechanisms in supply chain route optimization. Xu et al. [11] discussed AI-enhanced tools for cross-cultural game design, supporting online character conceptualization and collaborative sketching. Yang et al. [12] researched large scene adaptive feature extraction based on deep learning, enhancing scene understanding capabilities. Zheng et al. [13] conducted a comparative study of advanced pre-trained models for named entity recognition. Shen et al. [14] explored the clinical application of an AI system incorporating LSTM for anesthetic dose management in cancer surgery. Lastly, Xu et al. [15] analyzed adversarial machine learning in cybersecurity, focusing on attacks and defenses.

2. CONSUMER BEHAVIOR ANALYSIS UNDER THE BACKGROUND OF BIG DATA

2.1 Deeply explore consumers' emotional tendencies

In the wave of big data, deep mining of consumer emotional tendencies has become a challenging task and a key link in understanding consumer behavior. Traditional consumer behavior analysis often focuses on explicit data such as purchase records and browsing history, while big data gives us the opportunity to touch on the emotional tendencies hidden in consumer language, social media interactions, and even facial expressions. Through natural language processing (NLP) and sentiment analysis techniques, businesses can capture subtle emotional changes in consumers towards products, services, and even brands. These emotional data are like the "voice" of consumers, providing valuable market insights for businesses. For example, comments and feedback on social media are no longer just simple words, they contain consumers' loyalty to the brand, satisfaction with product features, and expectations for the future. Enterprises can use tools such as sentiment dictionaries and sentiment classification algorithms to quantitatively analyze the emotional tendencies of massive comments, in order to identify positive, negative, or neutral attitudes towards consumer emotions. Furthermore, by combining time series analysis, companies can track the changing trends of emotional tendencies, predict fluctuations in consumer emotions, and provide opportunities for crisis public relations, product iteration, and market strategy adjustments.

2.2 Cross platform integration analysis of consumer behavior patterns

In the digital age, consumer behavior is no longer limited to a single platform or channel, but spans multiple touchpoints such as e-commerce platforms, social media, and mobile applications. This cross platform consumption behavior pattern brings unprecedented challenges to enterprises, but also opens up new analytical perspectives. Through the integration capability of big data technology, enterprises can analyze data from different channels and reveal the consistency of consumer behavior and preferences between different platforms. Cross platform fusion analysis is not limited to simple data aggregation, but more importantly, it can identify consumer behavior patterns and motivations in different scenarios through algorithm models [1]. For example, by analyzing consumers' purchase history on e-commerce platforms and their interaction behavior on social media, companies can discover which social media content or activities can effectively promote purchase conversion, thereby optimizing advertising and content marketing strategies. Cross platform analysis can also help companies identify potential high-value customer groups and improve customer retention and repeat purchase rates through precision marketing.

2.3 Micro analysis of consumer decision-making process

The application of big data technology makes micro analysis of consumer decision-making processes possible. Traditionally, consumer decision-making has been considered a linear process, from demand identification to information collection, evaluation, and selection, to purchasing behavior and subsequent evaluation. Big data reveals a more complex and dynamic decision-making process, which includes a large number of factors such as immediate feedback, social influence, and emotional drive. Through big data analysis, enterprises can track every subtle action of consumers in the decision-making process, such as changes in search keywords, differences in page dwell time, and changes in shopping cart content. These micro behavioral data can reflect the process of consumers' inner hesitation, comparison, and final decision. Furthermore, by combining machine learning algorithms, enterprises can construct predictive models for consumer decision-making paths, identify key factors that affect decisions, such as price sensitivity, brand loyalty, social influence, etc. This micro analysis not only helps enterprises better understand consumers, but also provides scientific basis for optimizing product pricing, promotion strategies, and customer service.

3. CONSTRUCTION OF CONSUMER BEHAVIOR PREDICTION MODEL BASED ON BIG DATA

3.1 Design of a predictive model framework integrating multidimensional data

A core challenge in building consumer behavior prediction models based on big data is how to effectively integrate multidimensional data from different sources and formats, which may include basic consumer information such as age, gender, and geographic location, historical transaction records, social media activity, online browsing behavior, and even lifestyle data collected by IoT devices. In order to establish a comprehensive and accurate prediction model, it is necessary to design a framework that can handle such complex data structures. This framework first emphasizes the data preprocessing and cleaning stages, ensuring the data quality of the input model through techniques such as duplicate data deletion, missing value filling, and outlier detection. Then, valuable features are extracted from the raw data using feature engineering methods, which may involve dimensions such as consumer purchasing preferences, purchasing power, and social influence. In the feature selection stage, the most influential feature subset is selected using techniques such as correlation analysis, mutual information, and recursive feature elimination. Next, the framework integrates various machine learning algorithms such as random forests, gradient boosting trees, neural networks, etc., which can capture nonlinear relationships and complex patterns in data. Through optimization strategies such as cross validation and grid search, the best model parameters are found to ensure the generalization ability of the model. The framework also supports ensemble learning methods such as bagging and boosting, and further improves prediction accuracy by combining multiple weak learners.

3.2 Dynamic Learning and Adaptive Adjustment Mechanism

Consumer behavior is constantly changing over time, so effective predictive models require the ability to dynamically learn and adaptively adjust. This means that the model not only needs to be able to process historical data, but also needs to be able to absorb new data in real time and continuously optimize itself to reflect the latest trends in consumer behavior. To achieve this goal, an online learning mechanism has been introduced, allowing the model to update immediately upon receiving new data without the need to retrain the entire model, greatly improving the efficiency and response speed of the model's updates. At the same time, combined with time series analysis techniques, the model can identify the patterns and trends of consumer behavior over time, such as seasonal fluctuations, periodic changes, etc., providing temporal information for prediction. In addition, an adaptive adjustment strategy has been designed. By monitoring changes in the predictive performance of the model, when the prediction error exceeds a preset threshold, the model will automatically trigger a retraining process. This strategy ensures that the model can quickly adjust in the face of sudden changes in consumer behavior and maintain the accuracy of predictions.

3.3 Explanation and Transparency of Consumer Behavior Prediction Models

While pursuing prediction accuracy, the interpretability and transparency of consumer behavior prediction models are equally important. An excellent model not only needs to accurately predict consumer behavior, but also needs to clearly explain the logic and basis behind the prediction results to decision-makers. To achieve this goal, interpretable AI (XAI) techniques are adopted, which include rule-based models (such as decision trees), feature importance scores, local interpretation methods (such as LIME), and global interpretation methods (such as SHAP). Through these techniques, it is possible to intuitively see which features have a significant impact on the prediction results and how these impacts work. In addition, attention is also paid to the transparent construction of the model. Through documentation and visualization, the model construction process, assumptions, data sources, predictive logic, and other information are clearly presented to decision-makers. This not only enhances the credibility of the model, but also promotes cross departmental communication and cooperation, making the model better serve the strategic planning and daily operations of the enterprise.

4. CASE ANALYSIS

4.1 Big Data Driven Consumption Upgrade

In the field of e-commerce, Amazon's personalized recommendation system is a model of big data application. The system constructs a highly personalized product recommendation network by deeply mining multi-dimensional data such as consumers' shopping history, browsing behavior, search records, and evaluation feedback. Amazon's recommendation algorithm not only considers consumers' direct purchasing behavior, but also deeply analyzes consumers' indirect interests, such as browsing but not purchasing products, adding products to the shopping cart, and purchasing patterns similar to other consumers. Taking a consumer named Zhang Hua as an example, he bought many science fiction novels and a high-end camera on Amazon. Based on these purchase records, Amazon's recommendation system not only pushed him more new science fiction novels and camera accessories, but also recommended a series of e-books and video courses related to photography skills based on the photography tutorial books he browsed. This deeply personalized recommendation not only meets Zhang Hua's direct needs, but also stimulates his potential interests and promotes consumption upgrading[2].Amazon's personalized recommendation system can achieve such precision thanks to its powerful data processing capabilities and advanced machine learning algorithms. The system can process massive consumer data in real time, identify similarities between consumers, correlations between products, and trends in consumer preferences through complex algorithm models. In addition, Amazon continuously optimizes its recommendation algorithms through A/B testing, user feedback, and other means to ensure the accuracy and relevance of recommendations.

4.2 Consumer Insights and Experience Optimization Empowered by Big Data

As a globally renowned coffee chain brand, Starbucks is also at the forefront of digital transformation. By collecting and analyzing multi-source data such as consumer transaction data, membership information, and social media interactions, Starbucks has built a comprehensive consumer portrait system and achieved precise insights into consumer needs. Taking Ms. Li, a loyal member of Starbucks, as an example, she often visits Starbucks and prefers drinks such as latte and matcha latte. She likes to spend money in the store on weekends and afternoons. Based on these consumption habits, Starbucks has pushed customized coupons and membership benefits to Ms. Li through its membership system, such as buy one get one free and point redemption, effectively enhancing her consumption experience and brand loyalty. Meanwhile, by analyzing Ms. Li's interactive behavior on social media, Starbucks found that she has a strong interest in coffee culture and baking techniques. Therefore, they recommended Starbucks' coffee tasting and baking courses to her, further deepening her brand identity. Starbucks' digital transformation not only enhances consumers' shopping experience, but also brings significant business value to the company. Through big data analysis, Starbucks can more accurately predict market demand, optimize inventory management, develop personalized marketing strategies, and achieve dual growth in sales and profits. Starbucks also uses big data to insight into changing trends in consumer preferences, continuously launching new products and innovative services, and maintaining the brand's market competitiveness and vitality.

5. CONCLUSION

This article conducts in-depth research on consumer behavior prediction models based on big data analysis, and successfully constructs a prediction model with practical application value. This model can not only help enterprises better understand consumer demand and market trends, but also provide strong support for enterprise market strategy formulation and product optimization. In the future, with the continuous development and improvement of big data technology, the research and application of consumer behavior prediction models will have broader prospects and potential.

REFERENCES

[1] Yao, T. (2024, August). Research on the Local Head Loss Coefficient in Short-Tube Hydraulic Testing. In 2024 3rd International Conference on Applied Mechanics and Engineering Structures (AMES 2024) (pp. 89-97). Atlantis Press.
[2] Zhao, S., Lu, Y., Gong, C., & Xu, Q. (2025). Research on Labour Market Efficiency Evaluation Under Impact of Media News Based on Machine Learning and DMP Model.
[3] Chen, K., Zhao, S., Jiang, G., He, Y., & Li, H. (2025). The Green Innovation Effect of the Digital Economy. International Review of Economics & Finance, 103970.
[4] Ukey, N., Yang, Z., Zhang, G., Liu, B., Li, B., & Zhang, W. (2022, August). Efficient knn join over dynamic high-dimensional data. In Australasian Database Conference (pp. 63-75). Cham: Springer International Publishing.
[5] Peng, Q., Zheng, C., & Chen, C. (2024). A Dual-Augmentor Framework for Domain Generalization in 3D Human Pose Estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2240-2249).
[6] Peng, Q., Planche, B., Gao, Z., Zheng, M., Choudhuri, A., Chen, T., ... & Wu, Z. (2024). 3d vision-language gaussian splatting. arXiv preprint arXiv:2410.07577.
[7] Deng, T., Bi, S., & Xiao, J. (2025). Transformer-Based Financial Fraud Detection with Cloud-Optimized Real-Time Streaming. arXiv preprint arXiv:2501.19267.
[8] Zhou, Y., Wang, Z., Zheng, S., Zhou, L., Dai, L., Luo, H., ... & Sui, M. (2024). Optimization of automated garbage recognition model based on resnet-50 and weakly supervised cnn for sustainable urban development. Alexandria Engineering Journal, 108, 415-427.
[9] Lyu, T., Gu, D., Chen, P., Jiang, Y., Zhang, Z., Pang, H., ... & Dong, Y. (2024). Optimized CNNs for Rapid 3D Point Cloud Object Recognition. arXiv preprint arXiv:2412.02855.
[10] Wang, Y., & Liang, X. (2025). Application of Reinforcement Learning Methods Combining Graph Neural Networks and Self-Attention Mechanisms in Supply Chain Route Optimization. Sensors, 25(3), 955.
[11] Xu, Y., Shan, X., Lin, Y. S., & Wang, J. (2025). AI-Enhanced Tools for Cross-Cultural Game Design: Supporting Online Character Conceptualization and Collaborative Sketching. In International Conference on Human-Computer Interaction (pp. 429-446). Springer, Cham.
[12] Yang, Y., Li, I., Sang, N., Liu, L., Tang, X., & Tian, Q. (2024, September). Research on Large Scene Adaptive Feature Extraction Based on Deep Learning. In Proceedings of the 2024 7th International Conference on Computer Information Science and Artificial Intelligence (pp. 678-683).
[13] Zheng, Z., Cang, Y., Yang, W., Tian, Q., & Sun, D. (2024). Named entity recognition: A comparative study of advanced pre-trained model. Journal of Computer Technology and Software, 3(5).

[14] Shen, Z., Wang, Y., Hu, K., Wang, Z., & Lin, S. (2025). Exploration of Clinical Application of AI System Incorporating LSTM Algorithm for Management of Anesthetic Dose in Cancer Surgery. Journal of Theory and Practice in Clinical Sciences, 2, 17-28.

[15] Xu, J., Wang, Y., Chen, H., & Shen, Z. (2025). Adversarial Machine Learning in Cybersecurity: Attacks and Defenses. International Journal of Management Science Research, 8(2), 26-33.

Author Profile

Shuai Qi (July 19, 1987-), male, Han ethnicity, from Suining, Sichuan, holds a master's degree and specializes in big data and artificial intelligence research.