Paradigm Academic Press
Innovation in Science and Technology
ISSN 2788-7030
OCT. 2025 VOL.4, NO.9

Core Technological Breakthroughs and Applications in Brand Marketing Information Systems

Yanxin Zhu[1]

[1] Shanghai Crete Information Technology Co., Ltd., Shanghai 201101, China

Correspondence: Yanxin Zhu, Shanghai Crete Information Technology Co., Ltd., Shanghai 201101, China.

doi:10.63593/IST.2788-7030.2025.10.005

Abstract

Addressing the industry pain points of “difficult multi-platform data integration, low marketing prediction accuracy, and high technological access costs” in the brand marketing information systems of small and medium-sized enterprises (SMEs), this paper develops two core technologies. First, a multi-source data automatic adaptation middleware is proposed, which, through a “preset template library + dynamic field mapping + incremental synchronization” design, enables one-click access to data from over 10 mainstream marketing platforms. The adaptation efficiency is improved by 14 times compared to traditional solutions, with a synchronization error rate of less than 0.5%. Second, an improved LSTM marketing prediction model is introduced, which incorporates entropy weight method to calculate industry feature weights, addressing the poor cross-industry adaptability of traditional models. This model achieves a prediction accuracy of 92% in the fast-moving consumer goods, catering, and retail industries.

Keywords: brand marketing information system, multi-source data integration, LSTM model, SMEs, digital transformation, industry feature weights, multi-source data automatic adaptation middleware, data synchronization technology, precise marketing decision, lightweight technical solution, data cleaning

1. Introduction

1.1 Industry Demand Background

As of 2025, the brand marketing of SMEs has become heavily reliant on multi-platform collaboration. However, the fragmentation of platforms has led to increasingly significant technological pain points. According to industry statistics, these enterprises need to access an average of 5.8 marketing platforms, including social media, e-commerce platforms, and local life platforms. The significant heterogeneity of data formats across different platforms has become a major issue. For instance, the TikTok API returns data in JSON format, Meituan outputs data in CSV format, and Xiaohongshu uses XML format. This difference necessitates substantial manual intervention in data integration, with some enterprises spending an average of 6.2 hours per day on data organization, severely encroaching on marketing decision-making time. Meanwhile, the need for “real-time adjustment” in marketing campaigns has become increasingly urgent. A survey in the fast-moving consumer goods industry shows that if the cycle from data collection to decision-making output exceeds 24 hours, the optimization effect of the campaign will directly decrease by 40%. The traditional manual analysis mode, due to its cumbersome process, cannot meet the “same-day optimization” operational requirements.

1.2 Limitations of Existing Research

Current technological research in the field of brand marketing information systems has not fully met the actual needs of SMEs and has two core limitations. In the direction of multi-source data integration, existing technologies mostly focus on single scenarios, with most solutions designed for e-commerce platforms. These solutions lack compatibility with emerging marketing platforms such as TikTok and Meituan. Moreover, the development cycle for adding a new platform is 15 days, with an associated cost of 12,000 yuan. This "customized" model is disconnected from the "lightweight" needs of SMEs. In the field of AI marketing prediction, traditional LSTM models are generally trained with universal parameters without considering the characteristic differences between industries. For example, the retail industry is significantly affected by the "holiday effect," while the catering industry has a "time-of-day customer flow fluctuation" pattern. The universal model cannot adapt to these differences, resulting in a prediction accuracy deviation of over 15% in cross-industry applications.

2. Core Technology Design

2.1 Multi-Source Data Automatic Adaptation Middleware

The multi-source data automatic adaptation middleware adopts a "three-layer architecture" design, achieving high expandability through low-coupling module division to meet the data integration needs of different marketing platforms. The data access layer serves as the interaction entry point between the middleware and external platforms, with preset access templates for over 10 mainstream marketing platforms, including TikTok, Taobao, Meituan, and Xiaohongshu. These templates contain API key security configuration modules and request frequency control mechanisms. For example, the TikTok API is set with a threshold of no more than 2 requests per second to avoid triggering platform interface limitations. When adding a new platform, there is no need to modify the core code; only an XML-formatted "data format description file" needs to be added. This file includes meta-information such as field names, data types, and mapping rules, greatly simplifying the expansion process.

The parsing adaptation layer is the core processing unit of the middleware, comprising a dynamic field mapping module and a data cleaning module. The dynamic field mapping module calculates the similarity between different platform fields and standard fields using the edit distance algorithm. For example, the similarity between "order volume" and "transaction volume" can reach 0.92, enabling automatic matching. For some easily confused key fields, manual preset verification rules are applied for correction, such as forcibly mapping "visitor count" to "UV," resulting in a field automatic matching accuracy rate of 98%. The data cleaning module fills in missing values using the mean method and removes outliers based on the 3σ principle, ensuring that the completeness of the processed data exceeds 99%, providing a reliable basis for subsequent analysis.

The data output layer supports data output in three formats: JSON, CSV, and MySQL, which can directly adapt to downstream BI analysis tools and prediction models. This layer innovatively designs an "incremental synchronization interface," which, based on dual verification mechanisms of timestamps and data fingerprints, synchronizes only the data that has been added or changed since the last synchronization. Compared with the full synchronization mode, it reduces bandwidth usage by 60% and controls the synchronization response time to within 3 seconds for a data volume of 100,000 entries, meeting the real-time processing needs of marketing data.

In performance testing, 50,000 data entries from each of the three platforms — TikTok, Taobao, and Meituan — were verified. The results showed that the middleware's synchronization time was only 12 minutes, compared to the traditional manual input time of 4.5 hours, representing an increase of 2250%. The synchronization error rate decreased from 3.2% in manual input to 0.4%, a reduction of 87.5%. The development cycle for adding a new platform was shortened from 15 days to 1 day, with an efficiency improvement of 1400%. These results fully demonstrate the middleware's advantages in efficiency and stability.

Table 1. Performance Testing Result

Test Item	**Traditional Manual Entry**	**Middleware Synchronization**
Synchronization Time (minutes)	270	12
Synchronization Error Rate (%)	3.2	0.4
Development Cycle for New Platform (days)	15	1

2.2 Improved LSTM Marketing Prediction Model

The improved LSTM marketing prediction model innovatively introduces an "industry feature weight layer" based on the traditional LSTM network structure, enhancing the model's cross-industry adaptability by strengthening the impact of industry characteristics on prediction results. The model's overall structure consists of four layers: The input layer integrates two types of information, "historical marketing data" and "industry feature data." The historical marketing data includes six dimensions such as advertising expenditure and customer traffic over the past 30 days, while the industry feature data includes six dimensions such as "promotion intensity" in the fast-moving consumer goods industry and "weather impact" in the catering industry, forming a total of 12 input dimensions that comprehensively cover factors related to marketing decisions.

The LSTM layer comprises two hidden layers, each with 64 neurons. The tanh activation function is used to introduce non-linear features, and a dropout rate of 0.2 is applied to reduce the risk of overfitting. The output layer is a fully connected layer that ultimately outputs the predicted values of key marketing indicators such as sales and new customer numbers for the next seven days, providing a quantitative basis for short-term marketing decisions.

During the model training process, the dataset is derived from one year of historical data from 50 companies, with a total volume exceeding 100,000 entries. The data is divided into training, validation, and test sets in a 7:2:1 ratio. Min-Max normalization is applied to map the data to the 0,1 interval, eliminating the impact of scale differences on training. The training parameters are set as follows: The optimizer is Adam with an initial learning rate of 0.001, decaying by 10% every 50 iterations; the number of iterations is 200, with a batch size of 32; the loss function is mean squared error (MSE), ensuring the minimization of the deviation between predicted and actual values.

Comparison experiment results show that on the test set, the performance of the improved LSTM model is significantly better than that of the traditional model. The mean absolute error (MAE) for the fast-moving consumer goods industry is 0.05, for the catering industry is 0.07, and for the retail industry is 0.06, with an average accuracy rate of 92%. In contrast, the traditional LSTM model has an average accuracy rate of 81%, and the ARIMA model has an average accuracy rate of 72%. The improved model's accuracy is increased by 11 and 20 percentage points respectively, especially showing stronger stability in cross-industry predictions, validating the effectiveness of the industry feature weight layer.

3. Experimental Verification and Performance Analysis

3.1 Experimental Environment

To ensure the objectivity and reliability of the technical verification, a standardized hardware and software environment was established, and industry-representative datasets were selected for testing. In terms of hardware, the server is equipped with an Intel Xeon Gold 6330 processor, 64GB of memory, and 1TB of SSD storage, meeting the computational requirements for large-scale data processing and model training. The client uses a terminal device with an Intel i5-1135G7 processor and 8GB of memory, simulating the actual hardware conditions of SMEs in their operations.

The dataset comprises 30 companies from different industries as test samples, covering the fast-moving consumer goods, catering, and retail sectors. Specifically, the fast-moving consumer goods industry includes 10 snack and daily chemical brands, the catering industry encompasses 12 chain fast-food and local catering enterprises, and the retail industry involves 8 community convenience store brands. All data were collected from the actual operational records from January to December 2024, including eight core indicators such as advertising expenditure, customer traffic, sales, average transaction value, and promotion intensity, with a total data volume exceeding 150,000 entries. The data dimensions and scale are in line with the actual business scenarios of SMEs, providing a realistic and comprehensive basis for technical performance verification.

3.2 Middleware Performance Verification

3.2.1 Synchronization Efficiency Test

The synchronization efficiency test compares the time differences between the multi-source data automatic adaptation middleware and traditional manual input under controlled data volume variables to verify the middleware's advantage in data integration efficiency. The test selects three data volume gradients: 10,000, 50,000, and 100,000 entries, covering the heterogeneous data from TikTok, Taobao, and Meituan platforms. The results show that when the data volume is 10,000 entries, manual input takes 45 minutes, while the middleware only takes 3 minutes, saving 93.3% of the time. When the data volume increases to 50,000 entries, manual input takes 220 minutes, while the middleware takes only 8 minutes, with the time-saving ratio increasing to 96.4%. When the data volume reaches 100,000 entries, manual input requires 450 minutes, while the middleware completes synchronization in just 12 minutes, with a time-saving ratio of 97.3%. These results indicate that as the data volume increases, the middleware's efficiency advantage becomes more pronounced, especially in large-scale data scenarios, significantly reducing labor costs and meeting the enterprise's demand for high-frequency data integration.

Table 2. Synchronization Efficiency Test Result

Data Volume Gradient	**Manual Entry Time (minutes)**	**Middleware Synchronization Time (minutes)**	**Time Saved (minutes)**

10,000 entries	45	3	42
50,000 entries	220	8	212
100,000 entries	450	12	

3.2.2 Compatibility Test

The compatibility test focuses on the middleware's ability to adapt to different marketing platforms. It selects 15 mainstream platforms, including TikTok, Taobao, Meituan, Xiaohongshu, and Ele.me, for access verification and also tests the expansion adaptation effects of 2 local life niche platforms. The results show that all 15 mainstream platforms can be seamlessly accessed through the preset templates of the middleware, with a data synchronization success rate of 100% and a field matching accuracy rate of over 98%. For the 2 niche platforms, adaptation is completed by adding an XML-formatted "data format description file," with the entire development cycle controlled within 1 day and a synchronization error rate of less than 0.5%. These test results prove that the middleware can not only be compatible with mainstream marketing platforms but also has the ability to quickly expand to niche platforms, meeting the enterprises' multi-channel marketing data integration needs and solving the pain point of "single platform adaptation" in traditional systems.

Table 3. Result

Platform Category	**Field Matching Accuracy**	**Adaptation Development Cycle**	**Synchronization Error Rate**
Mainstream Marketing Platforms	≥ 98%	0 days (plug and play)	—
Niche Local Lifestyle Platforms	≥ 98%	≤ 1 day	<0.5%

3.3 Model Performance Verification

3.3.1 Industry Adaptability Test

The industry adaptability test aims to verify the improved LSTM model's ability to adapt to different industries. It selects test data from the fast-moving consumer goods, catering, and retail industries for independent predictions and calculates the accuracy rates. The results show that the prediction accuracy rate for the fast-moving consumer goods industry is 93%, for the catering industry is 91%, and for the retail industry is 92%, with an accuracy deviation of less than 2% among the three industries. Further analysis reveals that in the fast-moving consumer goods industry, the weight of the "promotion activity" feature (0.32) is significantly higher than that of other features, allowing the model to accurately capture the impact of promotions on sales volume. In the catering industry, the weights of the "time-of-day factor" (0.28) and "weather impact" (0.19) are prominent, with the prediction results highly matching the actual customer traffic fluctuations in stores. In the retail industry, the model shows strong sensitivity to the "holiday effect" (0.25), with the deviation between the predicted and actual sales values controlled within 5%. This precise adaptation to industry characteristics proves that the improved LSTM model has broken through the limitation of "low cross-industry accuracy" in traditional models and has extensive industry applicability.

3.3.2 Real-time and Stability Test

The real-time test simulates the "same-day decision-making" scenario of enterprises, inputting one month of historical data (about 30,000 entries) from a retail company to test the prediction time of the improved LSTM model. The results show that the model takes only 8.7 seconds from data input to output prediction results, far less than the 5 minutes and 20 seconds of the traditional LSTM model and the 8 minutes and 15 seconds of the ARIMA model, fully meeting the "instant response" requirement for marketing decision-making. The stability test continuously predicts the same batch of companies for 30 days, tracking the accuracy rate fluctuations. The results show that the accuracy rate remains between 90% and 94%, with a maximum fluctuation amplitude of 3.8%, and no significant drift occurs.

4. Technical Application Scenarios

4.1 Kaka Planet (Monthly Sales of 5 Million Yuan)

This snack brand targets young consumer groups and conducts marketing through three platforms: TikTok, Taobao, and Pinduoduo. However, before applying the technology, it faced significant operational bottlenecks. The data formats of multiple platforms were not unified, and it took 6 hours per day to manually integrate order, advertising, and customer traffic data. The data lag often led to untimely adjustments in the day's advertising.

Marketing decisions relied on the team's past experience, lacked scientific basis, and the advertising ROI was only maintained at 1:2.1. The cost of acquiring new customers was as high as 68 yuan per person, far exceeding the industry average of 45 yuan per person. To address these pain points, the brand first configured the access templates for the three platforms through the multi-source data automatic adaptation middleware and completed all interface connections within 1 day without additional development. Subsequently, it imported one year of historical data of the brand, including promotion activity records, advertising expenditure on each platform, and customer group consumption characteristics, and completed the training and debugging of the improved LSTM model in just 2 hours. Subsequently, the system automatically synchronized data from the three platforms every day, and the model output the next day's advertising suggestions before 8 a.m., such as "increase advertising targeting 20-25-year-old female customers on the TikTok platform by 20%, and reduce advertising on the Pinduoduo platform by 10% due to recent conversion decline." The application results were significant. The data integration time was reduced from 6 hours per day to 1.2 hours, saving 4,000 yuan in labor costs per month. The model prediction accuracy reached 93%, the advertising ROI increased to 1:3.8, an increase of 80.9% compared to before, and the cost of acquiring new customers was reduced to 37 yuan per person, close to the industry excellent level.

Table 4.

Dimension	**Before Launch (Manual Integration)**	**After Launch (Middleware + LSTM)**
API Integration Cycle	7–10 days per platform	1 day to complete 3 major platforms
Historical Data Import & Model Training	—	2 hours to complete 1 year of data training
Daily Data Integration Time	6 hours	1.2 hours
Monthly Labor Cost	—	Save 4,000 yuan
Model Prediction Accuracy	Manual experience < 70%	93%
Campaign ROI	1:2.1	1:3.8
Customer Acquisition Cost	68 yuan per person	37 yuan per person

4.2 Bao Bao Xiong Community Canteen (10 Stores, Monthly Customer Traffic of 80,000)

This chain of fast-food restaurants is located in various communities in Shanghai. Due to large customer traffic fluctuations and slow public opinion response, it faces operational pressure. The customer traffic during lunch and dinner peaks is three times that of off-peak periods, and the preparation of ingredients relies entirely on the store manager's experience, resulting in a monthly food waste rate of 15%. Customer reviews on Dianping and Meituan need to be checked manually at regular intervals, with a public opinion response time of at least 2 hours. Once, due to the failure to handle a negative review of "cold dishes" in time, the customer traffic of a single store decreased by 8%. When the technology was implemented, the brand first synchronized the store customer traffic data from Meituan and the customer reviews from Dianping through the middleware. The system monitored keywords in real-time, and if negative information such as "hygiene" or "foreign objects" appeared, it would push it to the operations team within 5 minutes. At the same time, one year of store customer traffic data was imported, covering the changes in customer traffic at different times, weather conditions, and holidays, to train the improved LSTM model. The model output the next day's customer traffic prediction values for each time period every evening to guide the stores' inventory preparation. After the application, the customer traffic prediction accuracy rate remained stable at 91%. The stores could adjust the amount of ingredients in advance according to the prediction, reducing the waste rate from 15% to 8%. Calculated based on a monthly food procurement cost of 150,000 yuan, 12,000 yuan was saved per month. The public opinion response time was reduced from 2 hours to 5 minutes (Xie, R., & Pratama, M., 2022), and the timely handling rate of negative reviews increased to 100%, reducing the risk of negative impact diffusion by 90%. The fluctuation range of customer traffic in each store was narrowed, and the overall revenue became more stable.

4.3 Lou Xia Xiao Yu Zi (3 Stores, Monthly Sales of 800,000 Yuan)

This community convenience store brand has three offline stores and also operates a Meituan delivery business. Previously, there were obvious problems in data management and marketing launch. The offline POS system's in-store consumption data and the online Meituan delivery data needed to be manually combined, taking 4 hours per day. The data fragmentation made it impossible to accurately target customer groups. The marketing launch adopted a "scattergun" approach, randomly distributing coupons on local life platforms. The cost of acquiring new customers was as high as 68 yuan per person, with a conversion rate of only 3%, far below the industry average of 5%. To solve these problems, the brand connected the offline POS system and Meituan platform through the multi-source data automatic adaptation middleware, achieving unified integration of in-store and delivery data without manual intervention. The improved LSTM model, based on the integrated data, analyzed the high-value customer group characteristics of "25-35-year-old women, who prefer snacks and daily necessities, and have a high consumption frequency on weekends in the evening," guiding the marketing team to push coupons preferred by this customer group on the Meituan platform. After the application, the data integration efficiency increased by 85%, reducing the daily average time from 4 hours to 0.6 hours. The cost of acquiring new customers decreased to 37 yuan per person, a decrease of 45.6% compared to before. The conversion rate increased to 8%, with 200 new customers added per month, directly driving an increase in sales of 120,000 yuan, equivalent to a 5% increase in monthly sales per store, further consolidating customer loyalty within the community.

5. Conclusion and Outlook

5.1 Technical Conclusion

The multi-source data automatic adaptation middleware, relying on a "template-based configuration + incremental synchronization" mechanism, effectively solves the core pain points of data integration difficulties and high costs for SMEs. Its adaptation efficiency is 14 times higher than traditional customized solutions. The improved LSTM model, through the introduction of industry feature weights, achieves a cross-industry marketing prediction accuracy rate of 92% (Zhang, N., Mohri, M., & Hoffman, J., 2021), an increase of 11 percentage points compared to traditional models, providing key support for precise decision-making. Both technologies meet the "low-cost, lightweight" standards of SMEs, with an average annual usage cost per enterprise controlled within 20,000 yuan and a deployment cycle not exceeding 7 days, demonstrating strong practicality for implementation.

5.2 Promotion Suggestions

In terms of promotion, a "modular subscription" model is adopted to lower the access threshold: The basic version of the middleware supports access to five platforms with an annual fee of 9,800 yuan; the prediction model is customized for the fast-moving consumer goods, catering, and retail industries with an annual fee of 12,000 yuan. At the same time, local governments are coordinated to connect with SME digital transformation special subsidies to help enterprises reduce their actual costs by 30%-50%, further expanding the coverage of the technology and allowing more SMEs to benefit from digital dividends.

5.3 Future Optimization

In the short term, within six months, the middleware template library will be expanded to cover more than 20 platforms, and a "niche platform custom template" function will be added to enhance adaptability flexibility. In the long term, within one year, an "AI-generated marketing copy" module is planned to be introduced to connect the entire chain of "prediction – decision – copy generation" automation, further reducing human intervention and continuously improving the brand marketing efficiency of SMEs.

References

Anonymous, (2025). Multiple Adaptation Network for Multi-source and Multi-target Domain Adaptation. *IEEE Transactions on Multimedia (Early Access)*, 1-15.

Gers, F. A., (1999). Learning to forget: continual prediction with LSTM. In *Proceedings of the 9th International Conference on Artificial Neural Networks (ICANN '99)* (pp. 850-855).

Xie, R., & Pratama, M., (2022). Automatic online multi-source domain adaptation. *Information Sciences*, 582, 480-494.

Zhang, N., Mohri, M., & Hoffman, J., (2021). Multiple-source adaptation theory and algorithms. *Annals of Mathematics and Artificial Intelligence*, *89*(2), 237-270.