Design of Marketing Digital Control System Based on the Integration of Big Data Analysis and Machine Learning

Qiming Li[1], Songling Du[2], Shiyuan Zhang[3]*
Marketing Business Department, State Grid Gansu Power Company, Lanzhou 730000, Gansu, China[1, 2]
Information Construction Division, Gansu Tongxing Intelligent Technology Development Co. Lanzhou, Lanzhou 730050, Gansu, China[3]

***Abstract*—The advent of the digital age has made it difficult for traditional marketing methods to meet the rapidly changing needs of the market. To improve the efficiency and effectiveness of enterprise marketing activities, it is particularly important to develop intelligent and precise marketing management systems. Therefore, the study proposes a marketing digital control system based on the integration of big data analysis and machine learning, which utilizes the Apache Flink distributed big data processing framework to design a marketing control system that includes system content and functional requirements. At the same time, machine learning design ideas are introduced into marketing recommendation algorithms, using reinforcement learning to enrich the business logic of the Rete network, dynamically generate and update rules, and calculate user interest while ensuring the fit between marketing recommendation content and user interest information. The study constructed a self-designed dataset (consisting of over 30000 pieces of data) through data simulation and crawling, and compared the research method with other machine learning algorithms on the same dataset. The results show that the maximum matching accuracy of the improved recommendation algorithm reaches 90.12%, and the prediction accuracy of user consumption behavior exceeds 88%, which is better than other comparative algorithms. The mean absolute error value on the product is less than 0.10, and the F1 value is greater than 0.65, indicating significant recommendation effectiveness. The research-designed marketing digital control system effectively integrates big data analysis and machine learning technology, providing support for the digital transformation of enterprises and the intelligent upgrading of related fields.**

Keywords*—Big data; flink; reinforcement learning; rete network; interest level; term frequency-inverse document frequency; rule

I. Introduction

With the continuous maturity and popularization of technologies such as cloud computing, big data, and artificial intelligence, enterprises can use multiple channels and methods to collect, process, and analyze user data, thereby providing reference for marketing decisions. The customer-oriented business philosophy requires enterprises to actively develop marketing activities and occupy favorable market competition shares in order to meet customer needs and respond to market dynamic changes [1]. Modern consumers are increasingly focusing on personalized and convenient shopping experiences, and their shopping needs and preferences are also complex and diverse. Strengthening digital marketing control is an important direction to achieve precise targeting of target customers and provide personalized products and services [2]. In recent years, with the emergence of diversified consumer markets, product consumption patterns have also shown a trend of diversification. Merchants with effective marketing costs often adopt a consumption rebate model to attract consumers, but the mechanism of this model is relatively single and has not fully utilized its advantages in attracting repeat customers and customer sources [3]. Data-driven decision-making has become an important part of marketing practice, and grasping target markets and user information is a key aspect of marketing management [4]. How to achieve coordinated planning of various resources and targeted product and service marketing is an important challenge facing enterprises, and how to mine information for reference and utilization in the complex and massive information data has also become a blind spot faced by enterprises. Previous research has only focused on data mining and prediction, lacking a comprehensive and detailed control design for marketing activity systems. Some of the content focuses on marketing design and often uses association algorithms, clustering algorithms, etc., for data mining, with insufficient consideration for the temporal and differential nature of marketing content data [5-6]. Therefore, the study proposes a digital marketing control system that integrates Flink real-time computing and reinforcement learning to optimize the Rete network, filling the gap mentioned above. The core goal is to achieve highly responsive and interpretable intelligent marketing decisions through dynamic rule generation and interest evolution modeling, and verify their effectiveness in complex business scenarios, thereby solving the performance bottleneck of traditional methods in complex rule sets and ensuring that marketing recommendation results are both accurate and compound with business logic. Specifically, the study proposes to combine big data analysis with machine learning algorithms to design a detailed marketing digital control system from the overall to the local level. Firstly, a processing framework is built using Apache Flink distributed big data technology. Then, reinforcement learning (RL) is used to improve the Rete network to enhance accuracy. To ensure the fit between marketing recommendation content and user interest information, user interest calculation is proposed. The innovation of the research lies in two aspects. One is to leverage Flink's advantages in

*Corresponding Author

block management and distributed parallelization, and use big data technology to achieve intelligent processing of user data. The second is to improve the Rete network by considering the performance of marketing recommendations and the increase in the number of rules, including dynamic matching, interest calculation, etc. The research aims to design a new digital marketing control scheme to help enterprises improve marketing efficiency and effectiveness, and provide support for the implementation of intelligent marketing systems.

The research mainly analyzes the relevant content of the design of the marketing digital control system from six aspects. Section II is a literature review and discussion of the data analysis and related algorithms of current marketing digital control. Section III is to design methods from two aspects: analysis of marketing digital control and construction of the Flink distributed system framework. The digital control analysis includes using a reinforcement learning network rule matching algorithm to generate rules, and considering user preferences, using a word frequency inverse document frequency algorithm to ensure the fit between marketing recommendation content and user interest information. Section IV is to test and analyze the performance comparison and application effects under the research methods. Section V and Section VI are a discussion and analysis of the results, as well as an overview summary of the entire text.

II. Related Work

Precision marketing requires a large amount of data support, including sales basic data, customer target group profile information, and product development information. With the help of big data analysis methods, enterprises can understand user needs, reduce marketing costs, and achieve low-cost, long-term, stable development, achieving precision marketing effects [7, 8]. Scholars Liu and Guo proposed using big data and fuzzy control algorithms to predict digital market preferences in the publishing industry, providing a reference for market analysis and enterprise transformation. The results indicated that this information technology could assist enterprise decision-making and organizational analysis by analyzing consumer information, and promote the development of digital publishing [9]. Rivas and Zhao scholars believed that ChatGPT has great potential for transforming marketing in the future, including understanding customers and achieving service automation [9]. Rithani et al. also believed that big data analysis can be used to collect massive amounts of data, where deep learning algorithms can achieve automatic feature extraction [10].

Liao et al. attempted to use clustering analysis and association rules to design a personalized marketing system for social networks, and believed that rule recommendations could be used to achieve personalized recommendations. The results indicate that this method can effectively achieve marketing and sales goals based on social media content [11]. Cherkaoui et al. proposed using data cleaning, variable statistics, and Apriori algorithm to achieve correlation analysis of marketing data, in order to better understand the relationship between marketing data and consumer behavior. The results indicate that this recommendation method can be combined with artificial intelligence to improve marketing efficiency [12]. Iyelolu et al. found that artificial intelligence significantly improves the accuracy of customer positioning and personalization, dynamically adjusts marketing strategies, and thus increases engagement and conversion rates [13].

Currently, the mainstream approach to precision marketing is to use big data-driven thinking, predictive models, and automated marketing tools for research. For example, reference [12] uses clustering to mine group preferences. However, most studies only focus on a single data source (such as social networks [11] or transaction logs [12]), lacking the fusion analysis of multimodal data (such as text+behavior+spatiotemporal). And traditional clustering/association rules rely on historical data, making it difficult to respond in real-time to interest drift (such as the impact of sudden heat search for products). To address these issues, innovative solutions are proposed: firstly, a multi-source real-time data processing framework based on Flink is constructed, integrating user profiles, behavior data, and external events; Secondly, design the RLHF Rete algorithm to dynamically optimize the rule network through reinforcement learning, achieving millisecond level matching; The research method can provide an interpretable and implementable technical path for intelligent marketing systems.

III. Methods

A. Design of Marketing Digital Recommendation Ideas

The workspace in most rule matching management is based on progressive rule logic, but the matching efficiency is low, and it is difficult to handle the problem of increased computational load during the periodic inspection process [14]. Therefore, the Rete algorithm is proposed for rule generation matching, which utilizes the characteristics of node sharing and state preservation to improve matching efficiency, reduce data transmission time between nodes, and display rule conditions through a data flow network, which is not affected by the number of rules and can effectively reduce the computational complexity of the product. Fig. 1 shows a schematic diagram of the Rete network.

In Fig. 1, the Rete network includes Alpha and Beta networks related to storage event testing, which respectively implement single and multiple test contents. Essentially, they can be seen as directed acyclic graphs of rule sets. When the Rete network runs, it matches rule conditions and facts through an object network in memory. Each working memory element enters the Rete network from the root node and may be stored in intermediate storage through propagation until it reaches the terminal [15]. Although the two major characteristics of Rete networks can improve operational efficiency, it should be noted that they do not consider the interrelationships between inputs, making it difficult to achieve complex connections between rule-based conditional models and factual structures. Therefore, the study utilizes Reinforcement Learning from Human Feedback (RLHF) to enrich the business logic of the Rete network, and dynamically generates and updates rules through continuous learning, namely the Reinforcement Learning from Human Feedback-Rete (RLHF-Rete) algorithm for rule matching in RL networks. The RLHF-Rete algorithm first uses RL to provide a classification model for events during rule formulation processing, allowing the system to explore in the state space. Then, with the continuous improvement of dataset learning strategies, the rule patterns are strengthened. Fig. 2 is a schematic diagram of rule formulation processing.

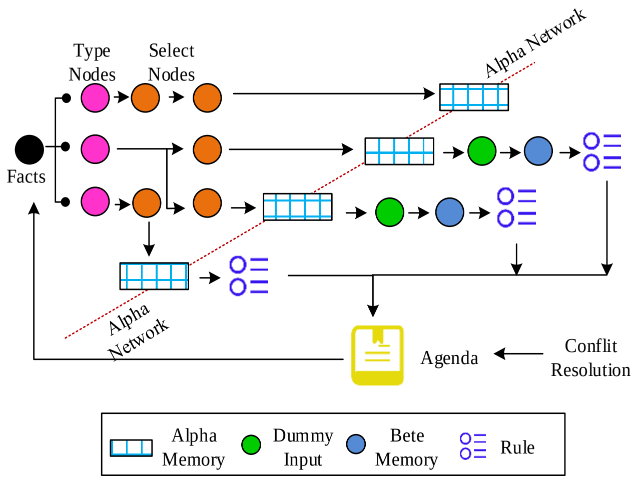

Fig. 1. Schematic diagram of Rete network.

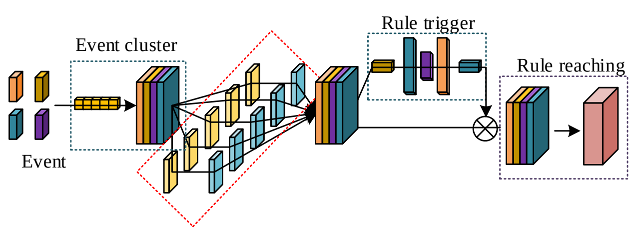

Fig. 2. Schematic diagram of rule formulation processing.

In Fig. 2, the time clusters are sorted according to a rule sequence and processed into clustering patterns. Rule triggers extract and train the event set, identify effective event patterns, and train the classification model to obtain the final rules. When RL is used for feedback control, different reward signal forms are set for different time periods based on different time point deviations, so that the controller and state will execute the new action according to the strategy. If the multi-mode rule sequence of the RLHF-Rete algorithm satisfies the Markov decision process property, it can describe the state transition probability $P(s'|s,a)$ between states and actions, as shown in Eq. (1):

$$P(s'|s,a) = Pr\{St = s'|St-1 = s, At-1 = a\} \quad (1)$$

In Eq. (1), S represents a continuous state space, A represents a finite behavior space, s represents the initial state, a represents the action, s' represents the subsequent state, r represents the reward function, and t represents time [16]. Eq. (1) can be represented as a three parameter function, and given the discount factor and strategy, the value function $V^c(s)$ can be represented as Eq. (2):

$$V^c(s) = E\left[\sum_{t=0}^{\infty} \lambda^t r(s_t, c_t(s_t))\right] \quad (2)$$

In Eq. (2), λ represents the discount factor, c is the given strategy, s_t is the behavior space, and E represents the expected value. The value range of λ generally does not exceed 1.0 to ensure that the optimal strategy is within a bounded range. At the same time, by using the likelihood ratio method to enhance the policy gradient, calculating unbiased gradient estimation, and selecting an appropriate value baseline to simplify the complexity of estimation, the updated policy gradient $\nabla_\gamma M(\gamma)$ can be expressed as Eq. (3):

$$\nabla_\gamma M(\gamma) = E_c\left[\sum_{t=0}^{T-1} \nabla_\gamma \log c(a_t|s_t,\gamma)(\sum_{t'=t}^{T-1} r_{t'} - v^c(s_t))\right] \quad (3)$$

In Eq. (3), $M(\gamma)$ is the objective function, $r_{t'}$ is the reward, $v^c(s_t)$ is the state value function, and γ represents the parameters of $M(\gamma)$. The choice of reward function can affect the effectiveness of rule generation. Therefore, research can provide feedback on the time window, balance rule allocation by controlling the weight of reward values, and consider the impact of rule delay on decision-making, in order to achieve pattern adjustment [17]. After completing the multi rule learning reinforcement mentioned above, considering the uncertainty of the recommendation process, research is conducted on user preference design under semantic pattern extension, including user evaluation information, consumption records, user profiles, etc. This "explicit feedback" can be intuitively fed back to the backend through the RLHF-Rete algorithm. From the perspective of implicit feedback, personalized products are mainly recommended based on user ratings and preferences. Fig. 3 illustrates the preference structure of implicit feedback.

In Fig. 3, user and order ratings are clustered, and preference results are analyzed based on product review results. A personalized recommendation list is generated using the RLHF-Rete algorithm, and the rules are dynamically configured and the recommendation list is updated technically, taking into account information such as time and environment. To ensure the alignment between marketing recommendations and user interest information, the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm is proposed to calculate user interest by constructing IF-IDF vectors of sentence pairs to calculate the cosine distance between vectors. Eq. (4) is the mathematical expression of the IF-IDF vector.

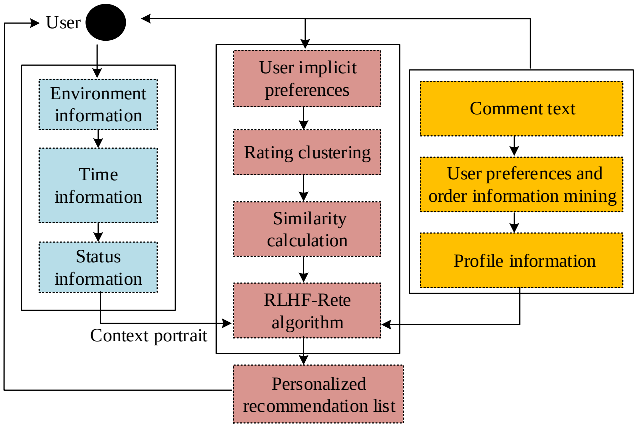

Fig. 3. Schematic diagram of preference structure for implicit feedback.

$$TF-IDF(w_i)=\frac{[Dt(w_i)]\times[\log|N|]}{[Num(SA\cup SB)]\times[\log(1+w_i:w_i\subset N)]} \quad (4)$$

In Eq. (4), $Dt(w_i)$ represents the frequency of occurrence of a word in a sentence, $Num(SA\cup SB)$ is the number of words shared between sentence A and sentence B, $|N|$ is the total number of sentences in the dataset, and w_i is the word [18]. With the help of this equation, the word can be changed to 'product' and the sentence can be changed to 'user', thus calculating the IF-IDF vector between the product and the user. Considering the subjective differences in user ratings, this study introduces vector mean to modify the cosine distance calculation, which calculates the correlation between users and products separately to obtain the mathematical expression of user interest under products, as shown in Eq. (5):

$$Pir_{x,g}=\sum_{y=1}^{q}\tilde{r}_{y,g}*sim(x,y)/\sum_{y=1}^{q}sim(x,y) \quad (5)$$

In Eq. (5), g represents the user, x, y are the products, q is the quantity of similar products, sim is the similarity between products, and $\tilde{r}$ is the user's rating of the products. The rating of different products can be achieved by adjusting the similarity weight, and Eq. (5) can be used to predict the level of interest of users in each product. Afterwards, based on Eq. (5), considering user similarity, the comprehensive formula for user interest is obtained, as shown in Eq. (6):

$$\begin{aligned}Pir_{x,g}'=\overline{r}_g&+[\sum_{p=1}^{j1}(\tilde{r}_{x,p}-\overline{r}_p)*sim(g,p)/\sum_{p=1}^{j1}sim(g,p)]\\&+[\sum_{y=1}^{j2}(\tilde{r}_{y,g}-\overline{r}_g)*sim(x,y)/\sum_{y=1}^{j2}sim(x,y)]\end{aligned} \quad (6)$$

In Eq. (6), $\overline{r}_g$ and $\overline{r}_p$ are the average ratings of all products under user g and p, while $j1$ and $j2$ are the sets of similar users and similar products. Considering that the training dataset designed for research is mostly static analysis content, to ensure the accuracy of predicting user preferences, the study proposes using a time sensitive function to solve the problem of user interest drift. The time function is defined as Eq. (7):

$$f(t)=e^{-\tau\cdot t} \quad (7)$$

In Eq. (7), t is time, e is the base of the natural logarithm, and τ is the time weight. The interest equation that introduces time weight can be expressed as Eq. (8):

$$Pir_{x,g}''=\frac{\sum_{t=1}^{j'}(\tilde{r}_{x,t}-\overline{r}_t)\times\partial(sim(x,y)\times\beta e^{-|t_x-t_y|})\times\tilde{r}_{x,s}\times\beta e^{\frac{t_r-t_{\min}}{t_{\max}-t_{\min}}}}{\sum_{t=1}^{j'}sim(x,y)} \quad (8)$$

In Eq. (8), $\partial()$ represents the logistic activation function, t_x, t_y represent the product rating time in the record set, $t_{\max}$, $t_{\min}$ are the maximum and minimum rating time, and t_r is the user's rating time. By using event sets, users' purchase records and ratings can be recorded, and similarity and similarity matrices can be calculated based on product feature data. Afterwards, the similarity matrix can be used to select the closest content that best matches the product and user. The remaining backup content can be calculated for interest and sorted in descending order, and merged with the previous real-time recommendation result to remove duplicates until the final result that meets the user's interest is obtained.

B. Control System Based on the Integration of Big Data Analysis and Machine Learning

Research uses the Apache Flink distributed streaming data processing framework to process and analyze various types of data. By constructing intelligent marketing systems at various dimensions and recommendation algorithms, the interactive experience between the system and customers can be achieved, including providing differentiated products and services, maintaining customer sources, and improving customer loyalty. The core of the Flink framework is a distributed streaming data engine written in Java and Scala, which can perform stateful computation on unbounded and bounded data streams, including simultaneous processing of batch and stream processing tasks, providing event time and processing time semantics, handling out of order event streams, as well as high throughput, low latency, state consistency assurance, and support for multiple running modes. It can adapt to different application scenarios and requirements. The Flink framework can adapt to different levels of Application Programming Interface (API) layers, and its structure includes a deployment layer, a Runtime core layer, an API layer, and an application layer [19]. Flink supports three temporal semantic expressions for real-time data processing: Event Time, Processing Time, and Ingestion Time. The study added the Watermark mechanism in Flink to solve the problem of temporal semantic disorder in Event Time. The combination of Watermark and Window can trigger a delay mechanism to ensure that the system can automatically verify data and meet time conditions. Fig. 4 shows the system operation process of Flink.

In Fig. 4, the client moves the data on the application to the cluster manager for resource balancing and scheduling, and hands over the creation of jobs to the application manager for resource loading and job management. Flink generates multiple task managers, which dynamically allocate tasks based on task slots and execute them through job graphs. Using Flink as the streaming processing engine for marketing systems, program code conversion of syntax structures can be achieved, and combined with improved matching algorithms, the overall system performance can be improved, ensuring data transmission and output efficiency while enhancing intelligent marketing effectiveness [20]. When designing an intelligent marketing system, research is conducted on e-commerce platforms, analyzing business requirements, functional requirements, and non-functional requirements. In the part of business requirements, it is necessary to find the users who meet the conditions based on the image tag library and time span, and put the query messages into the cache system; in the part of functional requirements, it starts from the aspects of cache management, rule management, rule monitoring, and rule analysis; in the part of non-functional requirements, it starts from the consideration of the system's scalability, operation reliability, and data confidentiality. Fig. 5 is a schematic diagram of the core architecture design of the marketing system.

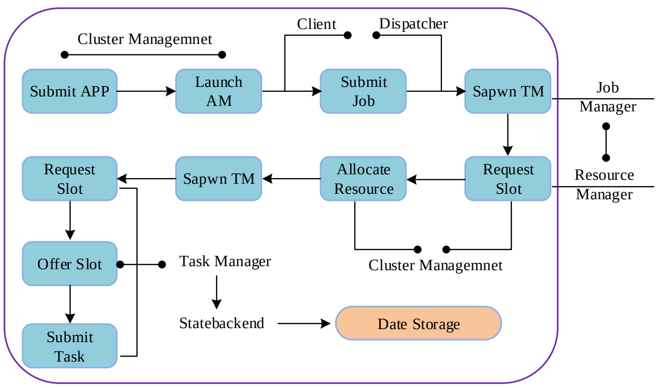

Fig. 4. Flink system operation process.

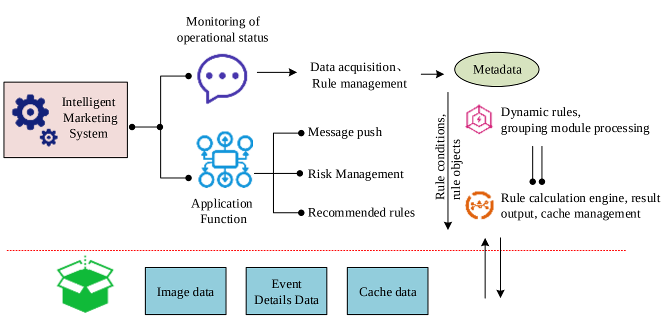

Fig. 5. Schematic diagram of the core architecture design of the marketing system.

In Fig. 5, the core architecture can be divided from top to bottom into rule management module, rule analysis module, application layer, data storage layer, and data caching layer. The data of the marketing system is mainly stored on the big data development platform, including collected user behavior data, business data, and other content. The rule analysis module utilizes collected user data to perform tasks such as analysis, calculation, and matching. In the database design part of the marketing system, it is necessary to design a data tag system related to users, including basic user attributes, activity status, interaction behavior, consumption ability, etc., and store user profile tag data in Hbase, using a two-level indexing method to achieve fast data retrieval [21]. The study was carried out with the help of IntelliJ IDEA 2021.2.3 development and VMware Workstation Pro for the configuration of the system cluster, and set up the auxiliary tools for each part, time tools, manager, event memory for the detailed design of each module.

IV. Results

A marketing digital control system was proposed based on big data analysis and machine learning algorithms, and refined from the perspectives of system function design and requirement analysis. The system control technology parameters used for the experiment were set up as follows: Intel (R) Core (TM) i7-1255H system (16GB memory, 512GB solid-state drive, Intel Iris Xe Graphics integrated graphics card, CentOS 6.3 operating system), data simulator Zookeeper 3.4.6.tar.gz, rule engine Flink, Drools (7.23.0 version), Kafka (2.11-20.0 version), and service cluster tool Hadoop. Considering that there is no standard publicly available dataset for analysis, the study aims to obtain data from two aspects: data policy and real data crawling to construct the dataset (consisting of over 30000 pieces of data). Among them, in the simulation data, a new application logic structure is added to the requirement, and a rule tree is established to achieve rule query and matching, using probability distribution to generate structured data. At the same time, using Zookeeper 3.4.6 to coordinate the state synchronization of distributed data generation nodes, ensuring data consistency; Simulate user behavior event streams (such as clicks, purchases, etc.) using Kafka, while Hadoop clusters store large-scale generated user profiles and behavior logs, supporting subsequent batch analysis. The triggering conditions include user profile attributes and behavior attribute conditions, where the user profile includes gender, age, income, occupation, region, etc., and the behavior attribute conditions include consumption behavior, product browsing status, comment status, etc. Analyze the facts, conditions, thresholds, and time elements based on rule conditions. In the real data acquisition section, the study utilized web crawling tools to obtain legitimate marketing data on platforms such as Taobao and JD.com, including consumption patterns, after-sales ratings, etc. A total of 13256 comments were collected. Firstly, the matching results of the improved algorithm proposed by the research under the rule set were analyzed, and the matching accuracy results are shown in Fig. 6.

Fig. 6(a) shows the accuracy and time consumption of traditional algorithms. The traditional algorithm reached its maximum matching accuracy of 62.15% when the dataset was 6000. However, as the number of rules increased, its matching accuracy showed a downward trend, with a maximum time consumption of 5.4 seconds. Fig. 6(b) shows the accuracy and time consumption of the improved algorithm, where the matching accuracy curve of the improved algorithm shows an upward trend, with a maximum matching accuracy of 90.12%, and is relatively less affected by the number of rule sets, with a time consumption of no more than 3 seconds. Fig. 6(c) shows the product matching results under the influence of time weight. Changes in the time weight coefficient will correspondingly affect the recommendation effect. When the coefficient value was 0.3, the similarity calculation results performed better. The improved algorithm exhibited higher recommendation accuracy compared to traditional algorithms, with an improvement of over 5% in accuracy. Afterwards, the improved algorithm was compared with Recursive Flow Classification (RFC), Spatial Structure Matching (SSM), and TextRank-Hierarchical Clustering (TextRank-HC) algorithms to analyze the rule matching results under different rule numbers and matching times. The comparison results are shown in Fig. 7.

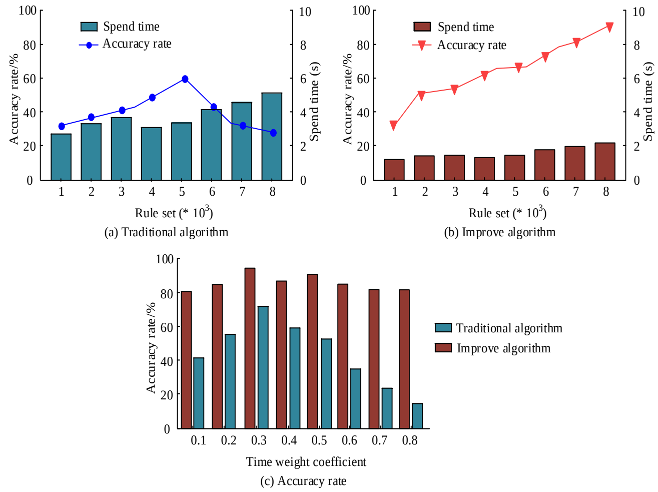

Fig. 6. Improved algorithm for matching results under rule sets.

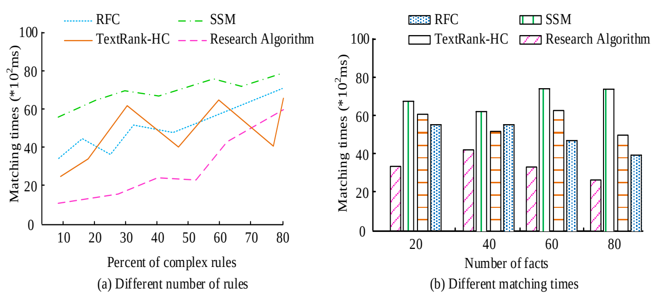

Fig. 7. Rule matching results under different numbers of rules and matching times.

In Fig. 7, under different proportions of complex facts, the improved algorithm had the shortest matching time, with a maximum value not exceeding 6000ms, and the overall spending time showed a relatively slow growth trend as the number of facts increased. Secondly, the RFC algorithm and TextRank-HC algorithm, which require less matching time, had a maximum processing time of no more than 7500ms. The SSM algorithm performed the worst and had a significant growth rate. Under different complex rules, the ranking of algorithms with the highest matching time was: SSM algorithm>TextRank-HC algorithm>RFC algorithm>research algorithm, and the research algorithm had the lowest growth rate of matching duration, with good real-time application and obvious response effect. Based on the application of the algorithm proposed by the research in marketing systems, this study analyzed the prediction of user consumption behavior and the effectiveness of product recommendations, and compared it with the K-Prototype Clustering Algorithm (K-prototype), the Clustering by Fast Search and Find of Density Peaks (CFSFDP), and the Comment perception Graph Neural Networks (RGRS). Fig. 8 to Fig. 9 shows the confusion matrices of different algorithms for predicting user consumption behavior.

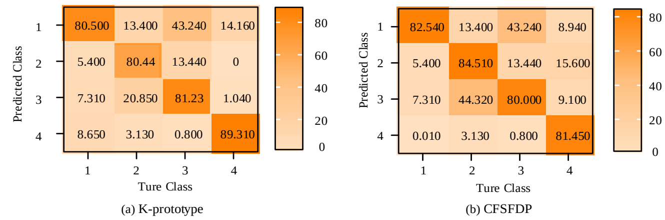

Fig. 8. The confusion matrix of K-prototype algorithm and CFSFDP algorithm for predicting user consumption behavior.

Fig. 8 shows the consumption behavior prediction results of the K-prototype algorithm and CFSFDP algorithm for four types of consumer goods (essential goods category-1, electronic products category-2, clothing category-3, and food category-4). The results showed that the prediction accuracy of the K-prototype algorithm exceeded 80%, and the prediction accuracy in food consumption behavior reached 89.31%. The prediction results of CFSFDP were slightly worse than those of the K-prototype algorithm, but it exceeded 80%. Afterwards, the prediction results of the RLHF-Rete algorithm and the proposed algorithm were analyzed, as shown in Fig. 9.

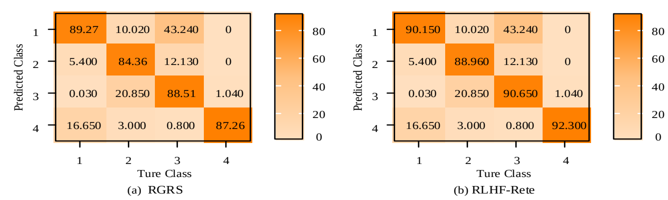

Fig. 9. The confusion matrix of the RLHF-Rete algorithm and the research algorithm for predicting user consumption behavior.

In Fig. 9, the RLHF-Rete algorithm had a prediction accuracy of over 84% for user consumption behavior, while the proposed algorithm had a prediction accuracy of over 88%. The recommendation performance was significant, and the deviation of its recommendation results was relatively small for the four types of products. The above algorithm was subjected to statistical significance testing, and the results are shown in Table I.

In Table I, the RFC algorithm suffers from accuracy loss due to recursive decomposition in terms of prediction accuracy metrics ($p<0.001$, effect size 1.87); The spatial structure calculation of SSM brings significant time overhead (34.8% more matching time); The text feature extraction and clustering of TextRank HC suffer from information loss (accuracy difference of 21.9%); The RLHF Rete algorithm can effectively compensate for information loss and has better prediction accuracy than other algorithms. Afterwards, the application effect of the above algorithm in marketing intelligent control was analyzed, and the results are shown in Fig. 10.

TABLE I. STATISTICAL TEST RESULTS OF DIFFERENT ALGORITHM PERFORMANCE

Evaluation indicators	RLHF-Rete	Comparison algorithm	*p* value	Effect size(Cohen's d)
Accuracy of consumer behavior prediction/ %	88.00±1.2	RFC:76.35±3.2	<0.001	1.87
		SSM:68.42±4.1	<0.001	2.45
		TextRank-HC:72.18±2.9	<0.001	2.01
MAE	0.0712±0.005	RFC:0.1521±0.011	<0.001	2.78
		SSM:0.1843±0.015	<0.001	3.12
		TextRank-HC: 0.1368±0.010	<0.001	2.35
F1 value	0.6923±0.015	RFC:0.5824±0.022	<0.001	1.98
		SSM:0.5217±0.025	<0.001	2.67
		TextRank-HC:0.6035±0.020	<0.001	1.45
Response delay (ms)	5820±210	RFC:7260±315	<0.001	1.89
		SSM:8920±405	<0.001	2.98
		TextRank-HC:8140±380	<0.001	2.45

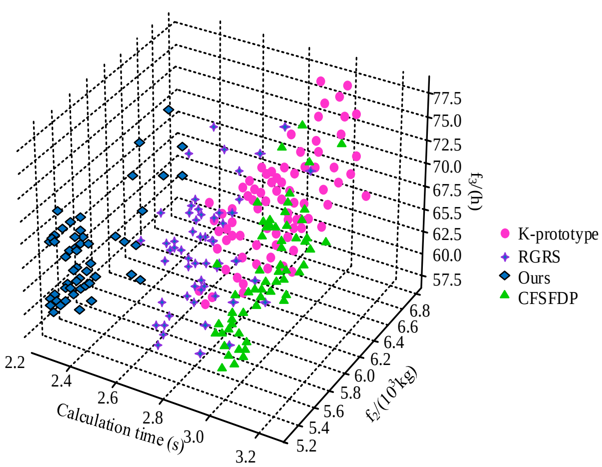

Fig. 10. System testing time results.

In Fig. 10, the research algorithm had a lower computational time compared to other comparative algorithms, with an overall time of no more than 2.2 seconds, and it also performed well in the other two indicators, with a more concentrated overall distribution. However, the performance of the K-prototype algorithm was poor, and the system application effect was not significant. Further analysis was conducted on the application effect of the above algorithms in marketing digitization, and the results are shown in Table II.

TABLE II. PERFORMANCE OF MARKETING DIGITALIZATION APPLICATIONS WITH DIFFERENT ALGORITHMS

Evaluation indicators	RLHF-Rete	K-prototype	CFSFDP	RGRS
F1 value	0.8923±0.015	0.7412±0.021	0.7849±0.018	0.8348±0.017
Response delay (ms)	5820±210	8240±355	7910±320	6850±280
Cold start adaptability (%)	83.5±2.1	72.3±4.2	68.7±3.8	78.6±2.9
Rule interpretability (rating)	4.8/5.0	3.2/5.0	2.9/5.0	2.1/5.0
Coverage rate of long tail products (%)	78.2±3.1	62.5±5.4	58.3±4.8	71.6±4.2
Compute Resource Consumption (GPU-h)	12.5±1.2	8.7±0.9	7.9±0.8	18.3±1.5

In Table II, the recommended F1 values for RLHF Rete algorithm, K-prototype algorithm, CFSFDP algorithm, and RGRS algorithm are 0.8923, 0.7412, 0.7849, and 0.8348, respectively. In the cold start scenario, the RLHF Rete algorithm can achieve a higher click through rate for product recommendations than other algorithms through rule templates and small sample learning. Additionally, its Gini coefficient for long tail product coverage is 0.21, which is much lower than that of the CFSFDP algorithm (0.38), reflecting a more equitable distribution of traffic. The GPU consumption of RLHF Rete algorithm is higher than that of clustering algorithm, but its unit request processing cost is 31.7% lower than that of RGRS algorithm. Afterwards, the marketing digital control system proposed by the research was tested and analyzed, and the results are shown in Table III.

TABLE III. MODULE TESTING AND PERFORMANCE TESTING RESULTS OF THE CONTROL SYSTEM

Project			Before application	After application
Module testing	Cache management	Data connection success rate	87%	94%
		Data query efficiency error	7%	2%
		Query time consumption	2.5s	0.8s
	Rule management	Sensitivity triggered by conditions	Medium	High
		Dynamic rule matching	80%	95%
		Dynamic rule modification	10min	3min
		Display sensitivity	Commonly	Excellent
	Rule monitoring	Data access	Manual	Automatic
		Time triggered event	Instable	Stable
		Execution status	75%	92%

	Rule analysis	Data collection and generation efficiency	82%	91%
		Server response	3.0s	0.7s
		Code running status	Occasionally make mistakes	Stable
Performance testing	Static environment	Data processing time	50s	10s
		Load balancing rate	79%	88%
		CPU usage rate	80%	45%
		Memory utilization rate	70%	50%
		Request (1000 times) query event takes time	30s	5s
		Overall stability	Generally	High
	Dynamic environment	Data processing time	95s	15s
		load balancing	72%	86%
		CPU usage rate	90%	50%
		Memory utilization rate	85%	60%
		Request (1000 times) query event takes time	40s	8s
		Overall stability	Generally	High

In Table III, the control system designed had good data analysis capabilities, with a data query time of only 0.8 seconds and a data collection and generation efficiency of over 90%. It could effectively improve resource allocation efficiency and adapt to system applications in both static and dynamic environments, with relatively stable performance. Afterwards, the application effect of the marketing system was analyzed, and the results are shown in Table IV.

TABLE IV. APPLICATION EFFECT OF MARKETING DIGITAL CONTROL SYSTEM

Project		Before application	After application
Application effect	Marketing conversion rate	2%	6%
	Clue follow-up efficiency	50%	90%
	Customer retention rate (6 months)	30%	50%
	Order conversion rate	10%	15%
	User satisfaction rating	7.5/10	9.0/10

In Table IV, the control system designed could predict consumer behavior, provide personalized recommendation services, timely follow up on 90% of leads, achieve a marketing conversion rate of over 5%, and improve order conversion rates and user satisfaction ratings to varying degrees.

V. DISCUSSION

The emergence of the Internet platform and the development of science and technology information technology make it easy for users to obtain massive data and realize online shopping [22]. Traditional enterprise marketing uses data warehouses to achieve information protection for user groups, and uses a single recommendation algorithm to predict user preferences. However, it ignores the real-time and interactive nature of the marketing system, which makes its application effect limited and difficult to meet the needs of enterprises [23]. The marketing system covers various aspects such as education, social interaction, consulting, life, and finance. Its core can be explained through three modules: input, recommendation, and output, which essentially involve the exchange of information services or products [24]. Accelerating the digitalization process of enterprises and the digital transformation of industries is an important aspect of promoting the development of the digital economy [25]. The existing marketing recommendation system, when using product user similarity to promote sales, fails to fully consider the preferences and needs of each user, with few feedback methods and mostly staying at the level of one-sided product recommendations [26]. To enhance the dynamic interaction between the system and users, research proposes rule-based algorithms and machine RL recommendation algorithms to achieve personalized recommendations.

The results showed that the improved algorithm proposed in the study exhibited a continuously improving matching accuracy as the number of rule sets increased, with a maximum matching accuracy of 90.12%, while the maximum matching accuracy of traditional algorithms was only 56.24%. When the time weight coefficient value was 0.3, the similarity calculation results performed well, and the improved algorithm has improved recall and accuracy by more than 5% compared to traditional algorithms. Under different proportions of complex facts, the rule matching time of the improved algorithm was the shortest, with a maximum value not exceeding 6000ms, and the growth trend was relatively slow. Next were the RFC algorithm and the TextRank HC algorithm, with a maximum time consumption of no more than 7500ms. SSM algorithm performed the worst and had a significant growth rate. The ranking of algorithms with the highest matching time under different complex rules as: SSM algorithm>TextRank-HC algorithm>RFC algorithm>research algorithm. The RLHF Rete algorithm dynamically optimizes the rule network structure through reinforcement learning, reducing redundant rule matching paths and keeping the matching time within 6000ms. The SSM algorithm needs to handle spatial structural similarity calculations, while TextRank HC has a high time complexity due to simultaneous text analysis and hierarchical clustering. Although the RFC algorithm adopts recursive optimization, it does not solve the problem of rule conflicts, making its matching performance second only to research methods. In terms of predicting user consumption behavior, the accuracy values of the three compared models all exceeded 80%, but they were slightly inferior to the RLHF-Rete algorithm proposed by the research. The RLHF-Rete algorithm had a prediction accuracy of over 88% for user consumption behavior, and the overall deviation was relatively small. 88% of the accuracy of consumer behavior prediction comes from the effective integration of structured behavior data and unstructured comments in the research method. Its RLHF mechanism can automatically adjust the threshold, which is more flexible than the CFSFDP fixed density threshold. Compared to the fuzzy control in reference [6], the proposed RLHF Rete improves real-time performance from a minute level to a second level while maintaining comparable interpretability of rules. And compared with the association rules and Apriori algorithm in reference [11], the research method can maintain linear growth through rule pruning. The product recommendation results showed that the MAE value (0.0712) exhibited by the RLHF-Rete algorithm was smaller than the other three comparison models. Among them, the K-prototype algorithm had the largest MAE value, reaching 0.1742, followed by the CFSFDP algorithm. The F1 value of the RLHF-Rete algorithm (0.6923) was greater than that of the K-prototype algorithm (0.6212), CFSFDP algorithm (0.5849), and RGRS algorithm (0.6348). The application effect of marketing intelligent control showed that the research algorithm had less computation time than other comparative algorithms, with an overall time of no more than 2.2 seconds. The data query time of this control system was only 0.8 seconds, and its data collection and generation efficiency exceeded 90%. It could effectively improve resource allocation efficiency and adapt to system applications in static and dynamic environments. Moreover, it could predict consumer behavior through marketing, with 90% of leads being promptly followed up, a marketing conversion rate exceeding 5%, and an order conversion rate exceeding 10%.

VI. Conclusion

The marketing digital control system based on the integration of big data analysis and machine learning proposed by the research has good application effects, which can effectively predict user consumption behavior and achieve high accuracy in product matching. Although the marketing digital control system designed in this study performs well in matching accuracy, recommendation effectiveness, and real-time performance, the system still heavily relies on large-scale user behavior data. However, in reality, there are data sparsity or noise interference that may affect the stability of reinforcement learning rule generation. The Apache Flink framework design may increase rule dynamic latency in high concurrency scenarios, resulting in lagging recommendation results and insufficient balance between personalization and generalization. Therefore, in future improvement work, research attempts to increase multimodal learning and deploy incremental learning frameworks to better model user product social relationships and adapt to data distribution drift. At the same time, by combining knowledge graphs and considering factors such as group behavior patterns and social network dissemination effects on marketing recommendations, more rules are extended to generate content, and a hierarchical rule engine is designed to deploy lightweight recommendation models to enhance the application effectiveness of the model system.

References

[1] Kopalle P K, Gangwar M, Kaplan A, Ramachandran D, Reinartz W, Rindfleosch A. Examining artificial intelligence (AI) technologies in marketing via a global lens: Current trends and future research opportunities. International Journal of Research in Marketing, 2022, 39(2): 522-540.

[2] Plangger K, Grewal D, de Ruyter K, Tucker C. The future of digital technologies in marketing: A conceptual framework and an overview. Journal of the Academy of Marketing Science, 2022, 50(6): 1125-1134.

[3] Ye Z, Zhang D J, Zhang H, Zhang R, Chen X, Xu Z. Cold start to improve market thickness on online advertising platforms: Data-driven algorithms and field experiments. Management Science, 2023, 69(7): 3838-3860.

[4] Abrardi L, Cambini C, Rondi L. Artificial intelligence, firms and consumer behavior: A survey. Journal of Economic Surveys, 2022, 36(4): 969-991.

[5] Nguyen D K, Sermpinis G, Stasinakis C. Big data, artificial intelligence and machine learning: A transformative symbiosis in favour of financial technology. European Financial Management, 2023, 29(2): 517-548.

[6] Ikegwu A C, Nweke H F, Anikwe C V, Alo U R, Okonkwo O R. Big data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions. Cluster Computing, 2022, 25(5): 3343-3387.

[7] Franke M, Pulignano V. Connecting at the edge: Cycles of commodification and labour control within food delivery platform work in Belgium. New Technology, Work and Employment, 2023, 38(2): 371-390.

[8] Hicham N, Nassera H, Karim S. Strategic framework for leveraging artificial intelligence in future marketing decision-making. Journal of Intelligent Management Decision, 2023, 2(3): 139-150.

[9] Liu Q, Guo L. Transformation and development strategy of digital publishing marketing based on big data and fuzzy control algorithm. Journal of Computational Methods in Sciences and Engineering, 2024, 24(1): 489-499.

[10] Rivas P, Zhao L. Marketing with chatgpt: Navigating the ethical terrain of gpt-based chatbot technology. AI, 2023, 4(2): 375-384.

[11] Liao S H, Yang C A. Big data analytics of social network marketing and personalized recommendations. Social Network Analysis and Mining, 2021, 11(1): 21.

[12] Cherkaoui N, Handri K E, Medard D Y T, El Hassani Y, Errafyg A. Consumer behaviour: analysing marketing campaigns through recommender systems and statistical techniques. Marketing i menedžment innovacij, 2024, 15(3): 1-12.

[13] Iyelolu T V, Agu E E, Idemudia C, Ijomah T I. Leveraging artificial intelligence for personalized marketing campaigns to improve conversion rates. International Journal of Engineering Research and Development, 2024, 20(8): 253-270.

[14] Rithani M, Kumar R P, Doss S. A review on big data based on deep neural network approaches. Artificial Intelligence Review, 2023, 56(12): 14765-14801.

[15] Nalbant K G, Aydın S. Development and transformation in digital marketing and branding with artificial intelligence and digital technologies dynamics in the Metaverse universe. Journal of Metaverse, 2023, 3(1): 9-18.

[16] Lopez S, Arjunan G. Optimizing marketing ROI with predictive analytics: Harnessing big data and AI for data-driven decision making. Journal of Artificial Intelligence Research, 2023, 3(2): 9-36.

[17] Paramesha M, Rane N L, Rane J. Big data analytics, artificial intelligence, machine learning, internet of things, and blockchain for enhanced business intelligence. Partners Universal Multidisciplinary Research Journal, 2024, 1(2): 110-133.

[18] Andronie M, Iatagan M, Uță C, Hurloiu I, Dijmărescu A, Dijmărescu I. Big data management algorithms in artificial Internet of Things-based fintech. Oeconomia Copernicana, 2023, 14(3): 769-793.

[19] Lăzăroiu G, Bogdan M, Geamănu M, Hurloiu L, Luminita L, Ștefănescu R. Artificial intelligence algorithms and cloud computing technologies in blockchain-based fintech management. Oeconomia Copernicana, 2023, 14(3): 707-730.

[20] Mokayed H, Quan T Z, Alkhaled L, Sivakumar, V. Real-time human detection and counting system using deep learning computer vision techniques[C]//Artificial Intelligence and Applications. 2023, 1(4): 221-229.

[21] Barzizza E, Biasetton N, Ceccato R, Salmaso L. Big Data Analytics and Machine Learning in Supply Chain 4.0: A Literature Review. Stats, 2023, 6(2): 596-616.

[22] Shah H M, Gardas B B, Narwane V S, Mehta H S. The contemporary state of big data analytics and artificial intelligence towards intelligent supply chain risk management: a comprehensive review. Kybernetes, 2023, 52(5): 1643-1697.

[23] Gallego V, Lingan J, Freixes A, Juan A A, Osorio C. Applying Machine Learning in Marketing: An Analysis Using the NMF and k-Means Algorithms. Information, 2024, 15(7): 368.

[24] Li L, Yuan L, Tian J. Influence of online E-commerce interaction on consumer satisfaction based on big data algorithm. Heliyon, 2023, 9(8): e18322.

[25] Adeniran I A, Efunniyi C P, Osundare O S, Abhulimen A O. Transforming marketing strategies with data analytics: A study on customer behavior and personalization. International Journal of Management & Entrepreneurship Research, 2024, 6(8): 41-51.

[26] Amudharasan A. The impact of recommendation algorithms: Analyzing the influence of data on marketing strategies in the media sector. Open Journal of Business and Management, 2023, 11(6): 3373-3384.