Research on e-commerce user stratification and coupon intelligent placement optimisation based on multi-behavioural RFM and deep reinforcement learning

Xiaokai Tang [1, *], Xin Jiang [2]

[1] School of Mathematics and Statistics, Sichuan University of Science & Engineering, Zigong, China, 643000

[2] School of Information and Control Engineering, Southwest University of Science and Technology, Mianyang, China, 621010

* Corresponding Author Email: suse_xiaokai7856@163.com

Abstract. With the ongoing accumulation of user behavioral data on e-commerce platforms, the precision of coupon allocation based on individual characteristics has emerged as a pivotal challenge for enhancing both resource allocation efficiency and marketing performance. Addressing the limitations of the traditional RFM model—particularly its reliance on single-dimensional user stratification and inability to capture complex interactive behaviors—this study introduces a novel customer segmentation approach that integrates a multi-behavioral RFM model with an enhanced self-organizing map (SOM) algorithm. This integrated method enables a more comprehensive identification of user value and facilitates fine-grained segmentation. Building on this stratification, an intelligent coupon delivery optimization framework powered by deep reinforcement learning (DRL) is further developed. In this framework, user stratification outcomes are embedded into the state space alongside real-time behavioral data, allowing dynamic learning of individual response patterns through a state-action-reward mechanism. Consequently, the system autonomously generates personalized and optimized coupon distribution strategies. Experimental results demonstrate that, under budgetary constraints, the proposed approach significantly increases the activity levels and repurchase frequencies of key users. Furthermore, it substantially outperforms static and random delivery strategies across critical metrics, including coupon conversion rate, delivery efficiency, and overall marketing effectiveness. These findings not only confirm the method's flexibility and adaptability in practical e-commerce settings but also offer a viable pathway and robust theoretical support for advancing precision marketing and optimizing resource allocation.

Keywords: E-commerce platform, Customer segmentation, Multi-behavior RFM, Deep reinforcement learning, Intelligent coupon delivery.

1. Introduction

With the accelerating pace of digital transformation, e-commerce has become deeply embedded within the global economic landscape, evolving into an essential component of consumers' daily lives. Driven by the rapid advancement of Internet and mobile communication technologies, the industry has experienced profound structural shifts, broadening and diversifying user demands to encompass not only everyday commodities but also high-end products and sophisticated services[1,2]. Concurrently, platform competition has intensified, transitioning from basic traffic acquisition to a more nuanced battle centered on refined user management and highly personalized marketing strategies[3~5]. Within this increasingly competitive environment, the imperative to craft differentiated, data-driven strategies capable of enhancing platform stickiness and driving user conversion has become central to sustaining market leadership.

At the core of platform operations lies the intelligent mining and analysis of transaction behavior data, which underpin critical marketing decisions. Key indicators—such as purchase frequency, transaction value, and temporal preferences—serve not only as foundational elements for generating actionable user insights but also as determinants of campaign precision and overall strategic effectiveness. Through systematic data mining, platforms can develop rich, multidimensional user

profiles that enable targeted campaign design, tiered customer management, and precise audience targeting[6~8]. Despite the widespread adoption of the classic Recency-Frequency-Monetary (RFM) model in user stratification, its inherent reliance on static rules has increasingly exposed limitations. In the face of dynamically shifting user behaviors and escalating market competition, this conventional framework struggles to capture the nuanced evolution of individual preferences and interaction patterns [9].

In response to these shortcomings, recent years have witnessed growing interest in the application of Deep Reinforcement Learning (DRL) to e-commerce marketing. Owing to its capacity for dynamic policy optimization within interactive and time-sensitive environments, DRL has demonstrated exceptional potential in modeling user behavioral sequences and refining marketing strategies through iterative learning and adaptive feedback loops. By doing so, it enhances both the personalization and responsiveness of recommendation systems and marketing campaigns[10,11]. Empirical evidence has consistently shown that DRL-based approaches surpass traditional static models across various scenarios, including personalized recommendations, dynamic pricing, and coupon distribution, ultimately delivering superior user experiences and driving platform revenue growth[12,13].

Building on this foundation, the present study introduces an integrated user stratification and coupon placement optimization framework that synthesizes a multi-behavioral RFM (MB-RFM) model, an improved self-organizing map (SOM) algorithm, and deep reinforcement learning. The MB-RFM model enriches user profiling and improves segmentation accuracy by incorporating diverse behavioral data, such as clicks and collections, while the enhanced SOM algorithm refines clustering performance, thereby enabling the precise identification of high-value users and effective exclusion of low-value segments[14~16]. Leveraging these advancements, DRL is subsequently employed to develop a dynamic coupon allocation mechanism capable of autonomously generating and optimizing differentiated incentive strategies tailored to individual users. Experimental findings affirm the superiority of this integrated framework, which significantly outperforms static and random strategies across key metrics, including coupon usage rates, conversion rates, and overall marketing effectiveness. Moreover, the proposed method substantially boosts core user engagement and strengthens platform profitability[17,18].

Against the backdrop of intensifying competition in the e-commerce sector, the design of scientifically grounded and operationally efficient personalized coupon strategies has become indispensable for enhancing user satisfaction and sustaining commercial success. The unified optimization framework proposed herein—driven by the synergistic application of MB-RFM, enhanced SOM, and DRL—not only offers a pragmatic pathway for achieving intelligent user segmentation and adaptive incentive distribution but also contributes valuable theoretical and empirical insights to the academic discourse on user behavior analytics and personalized recommendation systems.

2. Method and Modeling

2.1. Data Source and Description

The dataset used in this study was collected from a major e-commerce platform in China, covering the period from January 1, 2024 to December 1, 2024. The data include user behavioral records such as product browsing and purchasing, and were anonymized prior to analysis. All data were used strictly for academic research purposes and comply with data privacy regulations.

2.2. Multi-behavior RFM Model

In the evolving landscape of digital marketing, user interactions have grown markedly more diverse, rendering the traditional RFM model—reliant solely on purchase behavior—insufficient for capturing the full spectrum of user value. To address this limitation, the present study introduces a multi-behavioral RFM (MB-RFM) model, which systematically incorporates browsing, collection,

and purchase behaviors. By integrating these dimensions, the proposed model significantly enhances the precision and practical applicability of customer segmentation, thereby enabling a more holistic and nuanced understanding of user engagement.

2.2.1. Calculation of R Indicator

R-value measures the degree of a user's recent activity, serving as an indicator of their most up-to-date interaction with the platform. To capture the comprehensive contribution of various user behaviors to overall activity, this study integrates browsing, collection, and purchase actions into the assessment. The calculation formula is defined as follows:

$$R = \sum_{i=1}^{n} \omega_i R_i \tag{1}$$

where R_i denotes the most recent occurrence time of the i-th type of behavior, and ω_i represents the corresponding weight, indicating the importance of the behavior. The weight is determined by a combination of the Analytic Hierarchy Process (AHP) and the Entropy Weight Method, calculated as follows:

$$\omega_i = \frac{\theta_i \eta_i}{\sum_{j=1}^{n} \theta_j \eta_j} \tag{2}$$

where θ_i denotes the subjective weight and η_i denotes the objective weight. By combining both, the approach balances data-driven objectivity and domain knowledge, thereby ensuring a reasonable distribution of weights.

2.2.2. Calculation of F Indicator

The F-value quantifies the frequency of diverse user behaviors within a specified time cycle, serving as a key metric to capture the intensity and consistency of user-platform interactions. It reflects how closely users engage with the platform over time. The calculation formula is expressed as follows:

$$F = \sum_{i=1}^{n} \theta_i F_i \tag{3}$$

where F_i denotes the total number of occurrences of the i-th type of behavior, and θ_i represents the corresponding weight. The weighting method is consistent with that of the R indicator, ensuring the uniformity of classification indicators within the model.

2.2.3. Calculation of M Indicator

The M-value denotes the total monetary expenditure of a user within the observation period, directly indicating their economic contribution to the platform. It is calculated using the following formula:

$$M = \sum_{k=1}^{m} P_k \tag{4}$$

where P_k denotes the amount spent on the k-th purchase by the user. The design of the M value follows the traditional RFM model, emphasizing the user's purchasing power.

2.3. User Segmentation Model Based on Improved SOM Algorithm

Once the MB-RFM feature vector is constructed, customer stratification must be accomplished through an appropriate clustering technique. To this end, this study employs an improved self-organizing map (SOM) neural network algorithm, leveraging its superior nonlinear mapping capabilities and unsupervised learning properties. This approach enables the effective classification of users within a two-dimensional space, ensuring both precision and robustness in user segmentation.

2.3.1. Network Structure and Input

The SOM network is composed of an input layer, which receives (R, F, M) user features, and a competitive layer, consisting of multiple nodes, each corresponding to a potential customer category. Throughout the training process, input data trigger competitive interactions and adaptive adjustments among the nodes. This iterative mechanism ultimately leads to the automatic clustering of users, effectively mapping them into distinct and meaningful customer segments.

2.3.2. Optimization of Dynamic Learning Rate and Neighborhood Function

Building upon the standard SOM algorithm, this study incorporates a dynamic learning rate mechanism designed to facilitate rapid learning during the initial training phase while promoting stable convergence in later stages. The learning rate is adaptively adjusted throughout the training process, ensuring efficient pattern acquisition early on and mitigating oscillations or overfitting as the model approaches optimal clustering. The formulation is as follows:

$$\lambda(t) = \lambda_0 e^{-t/nT} \tag{5}$$

where λ_0 represents the initial learning rate, t denotes the number of iterations, and nT is the decay control parameter. Meanwhile, the neighborhood function was also optimized accordingly.

$$U_c(t) = f(t) \cdot \exp(-\frac{d_c^2}{2\sigma(t)^2}) \tag{6}$$

where d_c denotes the distance between the current node and the winning node, and $\sigma(t)$ represents the neighborhood radius, which is dynamically adjusted according to the following formula.

$$\sigma(t) = \sigma_0 e^{-t/T} \tag{7}$$

Dynamic neighborhood functions enhance clustering performance by enabling a global search during the early training stages, which promotes broad pattern discovery, and gradually shifting toward local optimization in later stages to refine cluster boundaries and improve classification accuracy.

2.3.3. Weight Update Mechanism

The SOM network weights are updated according to the following formula:

$$W_j(t+1) = W_j(t) + \sigma(t) U_c(t)[p_k(t) - W_j(t)] \tag{8}$$

where $W_j(t)$ denotes the current weight of the node, and $p_k(t)$ represents the input vector. This update strategy ensures that the network gradually converges during iterative training, thus improving the stability and accuracy of user segmentation.

2.4. Deep Reinforcement Learning-Based Delivery Decision Module Integrated with User Segmentation and Behavioral Features

Following the generation of user stratification results, developing dynamic and precise incentive programs tailored to distinct user segments becomes critical for enhancing marketing effectiveness. To address this challenge, this study designs a deep reinforcement learning (DRL)-based placement decision module, which integrates user stratification labels with real-time behavioral data. This integrated approach enables the optimization of placement strategies, seamlessly combining personalization with intelligent decision-making to maximize marketing impact.

2.4.1. State Space Modeling

The state design integrates user stratification labels (e.g., high-value users, potential users) with behavioral attributes (e.g., recent activity levels and interaction patterns), thereby offering a holistic representation of each user's current receptiveness. This comprehensive state depiction serves as the foundation for generating context-aware and adaptive marketing decisions.

2.4.2. Action and Reward Mechanism

The action space encompasses a diverse set of coupon issuance strategies executable by the system, including variables such as coupon amount, type, validity period, and delivery method. Meanwhile, the reward is dynamically computed based on user response behaviors—such as coupon collection, redemption, and subsequent repurchase activities—thereby providing a direct and quantifiable measure of campaign effectiveness.

2.4.3. Policy Learning and Updating

The decision module is trained based on Deep Q Networks (DQN) and the core policy update formula is:

$$Q(S,A) \leftarrow Q(S,A) + \alpha[R + \gamma \max_{a'} Q(S',a') - Q(S,A)] \tag{9}$$

where α denotes the learning rate, α' is the discount factor, and S' represents the new state after the user performs an action. This update mechanism enables the model to continuously explore and refine the delivery strategy based on trial-and-error learning and feedback.

3. Results

3.1. Analysis of Customer Segmentation Results

To rigorously evaluate the effectiveness of the MB-RFM model combined with the improved SOM algorithm in user segmentation, users were classified into seven distinct categories based on the model outputs. The number and proportion of users in each category are shown in Figure 1. Overall, the classification results not only demonstrate the model's robust capability in identifying users across multiple levels but also establish a solid empirical foundation for the development of precision marketing strategies. The results are detailed as follows.

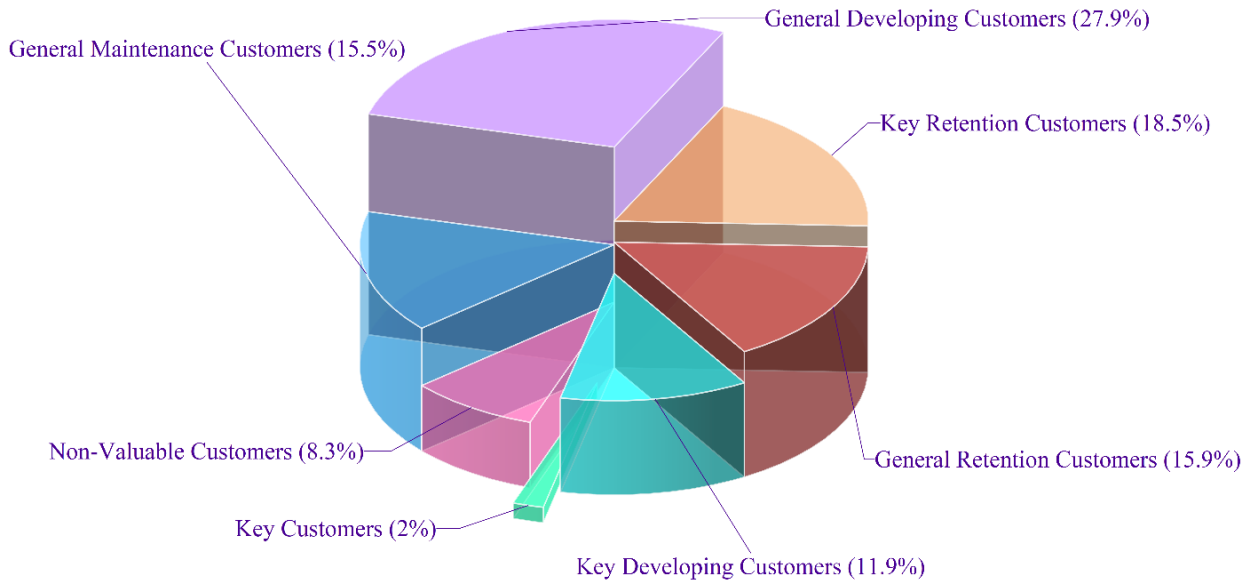

Figure 1. Customer Segmentation Results and Proportions Based on MB-RFM and Improved SOM

(1) User Distribution Characteristics

Analysis of the distribution ratios reveals that general development users constitute the largest proportion, accounting for 27.9% of the total. Although their immediate contribution to platform revenue is modest, their high activity levels and latent purchasing power suggest considerable potential for conversion into high-value customers through targeted incentive programs. In contrast, key retention and general retention customers account for 18.5% and 15.9%, respectively, together representing over 30% of the user base. This sizable proportion highlights the platform's strong core

of stable and active users, which plays a critical role in ensuring operational continuity and resilience. Notably, key customers represent only 2.0% of the total user population, the smallest group identified. Nevertheless, this aligns with established business principles whereby high-value users are both scarce and strategically indispensable. The model's ability to accurately isolate this valuable segment underscores its superior granularity and precision, offering clear focal points for strategic resource allocation and personalized service delivery.

(2) Marketing Value and Strategic Implications

The analysis further reveals pronounced differences in marketing value and corresponding strategic requirements across user groups. Despite comprising the smallest segment, key customers contribute disproportionately to platform performance and exhibit strong loyalty, warranting bespoke, premium service offerings to enhance retention and maximize lifetime value. Key development users (11.9%) and general development users (27.9%) represent critical growth segments. For these groups, deploying aggressive promotional tactics—such as targeted coupons and personalized offers—could accelerate their progression from merely active users to high-spending customers. In contrast, key retention and general retention users should be prioritized for stability-oriented interventions. Strategies emphasizing engagement maintenance and churn prevention, such as adaptive care initiatives and moderate incentives, are particularly appropriate to sustain their purchasing commitment. Finally, although non-value users (8.3%) exhibit limited activity and spending, they remain relevant to the broader strategy. Employing automated, cost-efficient outreach methods ensures continued engagement, while leaving open the possibility of eventual conversion, thus balancing resource allocation efficiency with long-term marketing potential.

3.2. MB-RFM Feature Analysis of Different Customer Types

To comprehensively assess the discriminative power of the MB-RFM model combined with the improved SOM algorithm in user stratification, this study systematically calculated the mean and median values for each customer category across the three dimensions of R (Recency), F (Frequency of behaviors), and M (Monetary value). The results, summarized in Table 1, reveal substantial heterogeneity among user groups across all indicators, thereby reinforcing the validity and robustness of the proposed stratification model. Detailed analyses of each dimension are presented below.

Table 1. R, F, M Statistical Characteristics of Different Customer Types

Customer Type	R		F		M	
	mean	median	mean	median	mean	median
General Retention Customers	64.64	71	1.11	1	135.41	130
General Developing Customers	165.92	169	1.04	1	113.22	84
General Maintenance Customers	12.23	13	1.00	1	115.68	108
Key Retention Customers	134.76	134	1.05	1	129.15	120
Key Developing Customers	102.06	102	1.06	1	139.22	144
Key Customers	12.47	13	2.24	2	242.07	219
Non-Valuable Customers	32.48	34	1.08	1	118.75	113

(1) R-Value (Recency) Characteristics

The R-value, which captures the time elapsed since a user's last interaction with the platform, serves as a proxy for recent activity—lower values indicate higher levels of engagement. Analysis indicates that key customers and general retention customers both exhibit mean and median R-values near 12 days, markedly lower than those observed for other user categories. This underscores the pronounced activeness of these two cohorts, with key customers demonstrating especially frequent platform interactions. By contrast, general development customers and key retention customers display elevated R-values of 165.92 and 134.76, respectively, signaling a notable decline in their interaction frequency. Such trends highlight the necessity of targeted incentives to revitalize these

users and re-engage them with the platform. Non-value customers, positioned within the medium-to-high R-value range, show limited current activity yet remain accessible, suggesting latent potential for reactivation and conversion.

(2) F-Value (Frequency) Characteristics

The F-value reflects the number of user interactions within the observation window, offering insights into behavioral intensity. While the majority of users demonstrate relatively low interaction frequencies, key customers emerge as a clear exception. This group records a notably high mean F-value of 2.24 and a median of 2, far surpassing other categories whose values largely cluster around 1. Such results illustrate the prominent role of key customers in sustaining high-frequency engagement. In contrast, other groups, including general retention and non-value users, exhibit only marginal differences in F-values, indicating that their differentiation is more closely tied to variations in activity levels and spending capacity rather than interaction frequency alone.

(3) M-Value (Monetary) Characteristics

The M-value, representing total user expenditure, serves as a direct indicator of economic contribution. Among all categories, key customers unsurprisingly lead, reflecting their critical commercial significance. For the remaining groups, M-values are relatively concentrated between 100 and 140. Within this range, general development and non-value customers rank toward the lower end, underscoring their currently limited economic impact. These groups thus emerge as priority targets for consumption stimulation through tailored strategic interventions. Particularly noteworthy is the case of key development customers, whose average M-value reaches 139.22—among the highest within non-core user segments. This suggests that, although not yet central contributors, they possess considerable growth potential. With sustained and focused engagement strategies, they are well-positioned to transition into high-value customers, thereby enhancing the platform's long-term profitability.

3.3. Comparative Analysis of Segmentation Performance

To rigorously validate the superior performance of the MB-RFM model combined with the improved SOM algorithm in customer stratification, this study conducted a systematic comparison against the SOM-based classification approach using the traditional RFM model. The evaluation focused on several critical dimensions, including the key customer identification rate, the rejection rate of low-value (worthless) customers, and clustering quality metrics—specifically, the Silhouette Score and Davies–Bouldin Index. The detailed results are summarized in Table 2 and Table 3.are analyzed in detail below.

Table 2. Comparison of Customer Segmentation Performance Between MB-RFM + SOM and Traditional RFM + SOM

Model	Total Customers	Key Customers	Key Customer Identification Rate	Non-Valuable Customers	Non-Valuable Customer Removal Rate
MB-RFM+SOM	18717	6069	0.324	1550	0.083
Traditional RFM + SOM	18717	4141	0.221	1581	0.084

Table 3. Clustering Performance Comparison Between MB-RFM + SOM and Traditional RFM + SOM

Model	Silhouette Score	Davies–Bouldin Index
MB-RFM+SOM	0.553	0.637
Traditional RFM + SOM	0.485	0.774

(1) Key Customer and Low-Value Customer Identification Performance

In terms of identifying key customers, the MB-RFM + SOM model demonstrates clear and substantial advantages. Under identical sample conditions (18,717 users), the enhanced model successfully identified 6,069 key customers, achieving a recognition rate of 32.4%—a remarkable increase of over 10 percentage points compared to the 22.1% rate yielded by the traditional RFM + SOM approach. This pronounced improvement underscores the ability of the MB-RFM model, through its incorporation of multidimensional behavioral features, to significantly deepen user profiling and sharpen value-based differentiation. Consequently, the model exhibits enhanced sensitivity and precision in isolating core user segments. Conversely, both models performed comparably in identifying low-value customers, with rejection rates of 8.3% for MB-RFM + SOM and 8.4% for the conventional model. This near equivalence suggests that, while the proposed model markedly improves high-value customer identification, it simultaneously preserves the robustness of low-value user filtering. Such balanced performance highlights its effectiveness in achieving a desirable trade-off between classification accuracy and comprehensiveness.

(2) Clustering Quality Evaluation

To further quantify the effectiveness of the proposed model in customer classification, two widely recognized clustering performance metrics—Silhouette Score and Davies–Bouldin Index—were employed for in-depth analysis. A higher Silhouette Score reflects better-defined and more discriminative clustering. The MB-RFM + SOM model achieved a Silhouette Score of 0.553, significantly surpassing the 0.485 score of the traditional model, thereby demonstrating superior capability in delineating user groups and establishing clear segment boundaries. Simultaneously, the Davies–Bouldin Index, where lower values indicate more compact intra-cluster cohesion and greater inter-cluster separation, further validated the model's effectiveness. The MB-RFM + SOM algorithm yielded a Davies–Bouldin Index of 0.637, considerably outperforming the traditional model's 0.774. This result reinforces the model’s notable advantage in enhancing inter-segment distinction and ensuring the clarity and interpretability of the classification outcomes.

3.4. Analysis of Delivery Optimization Effects Based on Deep Reinforcement Learning

To rigorously assess the practical effectiveness of the proposed coupon placement optimization framework, which integrates deep reinforcement learning (DRL) techniques, a systematic comparative experiment was conducted against static and random delivery strategies. This evaluation focused on three critical performance indicators: coupon issuance rate, usage rate, and conversion rate, with detailed results presented in Table 4. Furthermore, to examine the convergence dynamics and stability of the DRL strategy throughout the learning process, trends in Reward and Hit Rate were analyzed, as illustrated in Figure 2.The findings are discussed in detail below.

Table 4. Comparison of Coupon Distribution Strategies on Sending, Usage, and Conversion Rates

Strategy	Coupon Sending Rate	Usage Rate	Conversion Rate
DRL	0.802	0.537	0.651
Random	0.751	0.549	0.412
Static	1.000	0.551	0.551

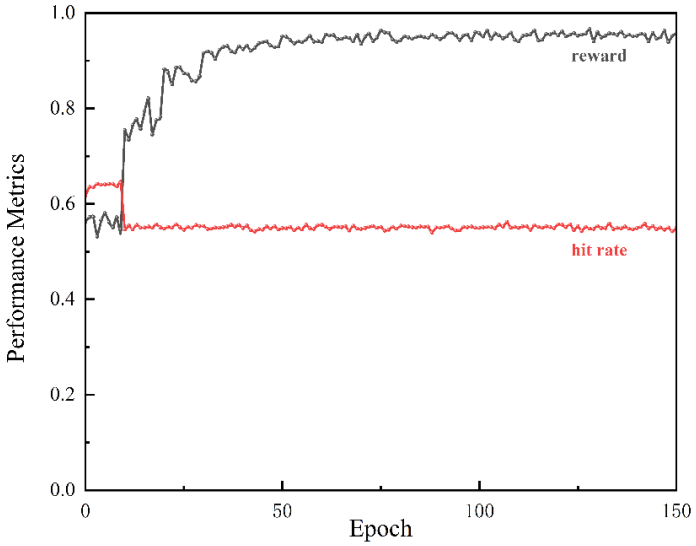

Figure 2. Convergence of Reward and Hit Rate During DRL-Based Coupon Optimization Training

(1) Comparison of Coupon Issuance Rate, Utilization Rate, and Conversion Rate

In terms of coupon issuance, the static strategy uniformly issues coupons to all target users, maintaining a rate of 1.0. By contrast, the random and DRL strategies exhibit issuance rates of 0.751 and 0.802, respectively. Although the DRL strategy marginally exceeds the random approach in issuance frequency, it remains substantially lower than the static rule-based method. This outcome underscores the DRL model's ability to strike an optimal balance between precision targeting and resource conservation. Regarding coupon utilization rates, differences among the three strategies are relatively minor, with values ranging between 0.537 and 0.551. The DRL strategy records a slightly lower utilization rate, suggesting that while refined targeting may not dramatically elevate usage frequency, it does not compromise the effectiveness of coupon application. However, the most striking results emerge in the conversion rate analysis. Here, the DRL strategy achieves a conversion rate of 0.651, significantly outperforming both the static (0.551) and random (0.412) strategies. This substantial improvement strongly demonstrates the DRL model's capability to dynamically generate personalized and motivational coupon distribution strategies by accurately recognizing users' current states and stratification labels. Consequently, it markedly enhances user responsiveness and conversion efficiency.

(2) Analysis of the DRL Strategy Learning Process

An examination of the reward trajectory reveals that the DRL model achieves rapid performance improvement during the initial training phase, with convergence occurring after approximately 30 iterations. Thereafter, the reward stabilizes at a consistently high level near 0.9, indicating the model's proficiency in quickly acquiring effective placement strategies and maintaining stable, high-quality outputs over time. In contrast, the Hit Rate curve exhibits a more intricate pattern of adjustment. Initially, the Hit Rate peaks at around 0.64, but subsequently declines and stabilizes within the range of 0.54 to 0.55. This dynamic reflects the DRL algorithm's adaptive trade-off between exploration and exploitation: while the model deliberately avoids excessive concentration and redundancy in coupon allocation, it ensures balanced user coverage and maintains diversity in delivery strategies. Such behavior illustrates the framework's ability to evolve nuanced and context-sensitive coupon distribution schemes during real-time learning.

3.5. Case Analysis of Personalized Delivery Strategy

To further validate the adaptability of the deep reinforcement learning (DRL) model in formulating personalized and contextually appropriate placement strategies, this study conducted a detailed statistical analysis focusing on three key metrics: the recommended coupon issuance amount, Q-value (action value), and actual coupon usage, based on a subset of sample customers. The corresponding results, presented in Table 5, are discussed in detail below.

Table 5. Case Analysis of Personalized Coupon Recommendation and Customer Response Based on DRL

Customer Type	Recommended Coupon Value	Q-value	Used or Not
Key Customer	10	1.023	1
Key Customer	10	0.966	1
Key Customer	20	0.903	0
Key Customer	20	0.995	1
Key Customer	10	1.065	1
Non-Valuable Customer	0	0.902	0
Non-Valuable Customer	20	0.992	0
Non-Valuable Customer	20	0.961	0
Non-Valuable Customer	0	0.936	0
Non-Valuable Customer	0	0.979	0

(1) Placement Strategy and Response Patterns of Key Customers

For key customers, the DRL model predominantly recommends higher-value coupons ($10 or $20) to encourage sustained engagement and stimulate repeat purchases. Analysis of the relationship between Q-value and user behavior reveals a clear pattern: when Q-values exceed 1 (e.g., 1.023 and 1.065), users almost invariably redeem the coupons. This outcome reflects not only their heightened price sensitivity but also a strong propensity to respond positively to monetary incentives. Conversely, when Q-values decrease to lower levels (e.g., 0.903), coupon redemption is rare, underscoring the model's accuracy in predicting user acceptance and responsiveness. These findings highlight the DRL algorithm's dual capability: it not only forecasts user preferences based on historical behavioral data but also dynamically refines placement decisions in real time. In scenarios where anticipated user response is high, the system proactively deploys incentive-based strategies to reinforce user stickiness and loyalty. However, when expected responsiveness is lower, the model prudently moderates coupon issuance, thereby achieving a delicate balance between maximizing user incentives and optimizing resource allocation. Such dynamic and discriminative decision-making exemplifies the system's sophistication in operationalizing the "high input–high return" paradigm.

(2) Placement Strategy and Response Patterns of Low-Value Customers

In the case of non-valuable users, the DRL model exhibits a distinct placement suppression mechanism and demonstrates clear tendencies toward strategic convergence. In most instances, the model recommends no coupon issuance (amount = 0), coupled with generally low Q-values (approximately 0.9), which aligns with minimal user engagement and negligible coupon redemption. These outcomes indicate that the model effectively identifies user groups with limited activity and purchasing potential, thereby minimizing the risk of inefficient resource allocation. Notably, even in exploratory scenarios—where the model tests the issuance of higher-value coupons ($20) associated with Q-values of 0.992 and 0.961—users largely refrain from redeeming the offers. This behavior reinforces the model's intelligent adjustment capability. Upon detecting unsatisfactory trial outcomes, the DRL model swiftly modifies its placement strategy to avoid redundant or wasteful investments. Such adaptive refinement underscores its practical value in enhancing resource efficiency and in formulating intelligent, personalized marketing strategies that are dynamically responsive to user behavior.

4. Conclusions

To address the inherent limitations of existing customer segmentation approaches—particularly their reliance on singular behavioral representations and their lagging responsiveness to dynamic user interactions—this study introduces a comprehensive user stratification method. By integrating a multi-behavioral RFM (MB-RFM) model with an enhanced self-organizing map (SOM) neural network, and further incorporating deep reinforcement learning (DRL) technology, an intelligent placement optimization framework is established. This framework enables differentiated and precise incentive mechanisms at the individual user level, thereby advancing the granularity and relevance of personalized marketing strategies.

During the customer stratification process, the MB-RFM model substantially improves the descriptive depth of user profiles by incorporating multidimensional behavioral indicators, including clicks, purchases, and collections. Complementing this, the improved SOM algorithm demonstrates outstanding clustering performance and stability, significantly enhancing the hierarchical model's capacity to capture complex user behavior patterns. Empirical findings confirm that this integrated method outperforms the traditional RFM model across several key performance metrics. Specifically, it achieves superior results in key customer identification rates and clustering quality—evaluated through silhouette coefficients and Davies–Bouldin indices—while maintaining a consistently high rejection rate for low-value customers. These results collectively indicate the method's success in balancing classification accuracy with marketing practicality.

With regard to personalized placement decision-making, the DRL-based optimization module introduces user stratification labels and real-time behavioral attributes into the decision-making state space. This integration markedly enhances the model's responsiveness and adaptability in formulating dynamic coupon issuance strategies. Comparative experimental results reveal that, although the DRL model maintains a significantly lower coupon issuance rate compared to static and random strategies, it attains a notably higher conversion rate of 0.651. This figure far surpasses that of traditional approaches, thereby validating the DRL framework's superior capability in optimizing user conversion and resource allocation. Further case analysis underscores the model's nuanced adaptability: while actively incentivizing high-value users to increase repurchase likelihood, it simultaneously exercises strategic restraint in issuing coupons to low-value users, effectively minimizing unnecessary resource expenditure. Such dynamic decision-making exemplifies the model's intelligent adjustment capacity and reinforces its potential in managing user heterogeneity and optimizing commercial outcomes.

Collectively, the proposed joint optimization framework—integrating MB-RFM, improved SOM, and DRL—has demonstrated remarkable effectiveness in enhancing user segmentation accuracy, personalized delivery efficiency, and intelligent decision-making. Moreover, the framework offers significant practical value and broad application potential within the domain of precision marketing. Looking ahead, future research should explore the integration of advanced methodologies, including time series modeling and multi-objective reinforcement learning, to further enhance the dynamic response capability and strategic diversity of the system, ensuring its adaptability to the increasingly complex and multifaceted demands of digital marketing environments.

References

[1] Yin Chunli, Han Jinglong. Dynamic pricing model of e-commerce platforms based on deep reinforcement learning [J]. Computer Modeling in Engineering & Sciences, 2021, 127 (1): 291–307.

[2] Zhang Qi, et al. Optimization of dynamic pricing models for consumer segmentation markets and analysis of big data-driven marketing strategies [J]. Journal of Organizational and End User Computing (JOEUC), 2025, 37 (1): 1–33.

[3] Li Cheng, Chu Min, Zhou Chen, et al. Two-period discount pricing strategies for an e-commerce platform with strategic consumers [J]. Computers & Industrial Engineering, 2020, 147: 106640.

[4] Wang Yihua, Minner Stefan. Deep reinforcement learning for demand fulfillment in online retail[J]. International Journal of Production Economics, 2024, 269: 109133.

[5] Wei Zhehuan, Yan Liang, Zhang Chunxi. Optimization strategies in consumer choice behavior for personalized recommendation systems based on deep reinforcement learning [J]. Journal of Organizational and End User Computing (JOEUC), 2025, 37 (1): 1–35.

[6] Islam Md Rafiqul, Shawon Reza E Rabbi, Sumsuzoha Md. Personalized marketing strategies in the US retail industry: leveraging machine learning for better customer engagement [J]. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence, 2023, 14 (1): 750–774.

[7] Tang Zhipeng, et al. E-commerce platforms offer different free shipping coupon services: surplus or loss? [J]. Journal of Systems Science and Complexity, 2025: 1–28.

[8] Day Min-Yuh, Yang Ching-Ying, Ni Yensen. Portfolio dynamic trading strategies using deep reinforcement learning [J]. Soft Computing, 2024, 28 (15): 8715–8730.

[9] Zhong Xingyu, et al. Deep reinforcement learning for dynamic strategy interchange in financial markets [J]. Applied Intelligence, 2025, 55 (1): 1–19.

[10] Hu Li, Zhang Mengwei, Wen Xin. Optimal distribution strategy of coupons on e-commerce platforms: sufficient or scarce? [J]. International Journal of Production Economics, 2023, 266: 109031.

[11] Zhou Shuang, Hudin Norlaile Salleh. Advancing e-commerce user purchase prediction: integration of time-series attention with event-based timestamp encoding and graph neural network-enhanced user profiling [J]. PLOS ONE, 2024, 19 (4): e0299087.

[12] Gabryel Marcin, et al. Accelerating user profiling in e-commerce using conditional GAN networks for synthetic data generation [J]. Journal of Artificial Intelligence and Soft Computing Research, 2024, 14 (4): 309–319.

[13] Liu Yishu, Hou Jia, Zhao Wei. Deep learning and user consumption trends classification and analysis based on shopping behavior [J]. Journal of Organizational and End User Computing (JOEUC), 2024, 36 (1): 1–23.

[14] Liu Yishu, Chen Chen. Improved RFM model for customer segmentation using hybrid meta-heuristic algorithm in medical IoT applications [J]. International Journal on Artificial Intelligence Tools, 2022, 31 (1): 2250009.

[15] Liao Juan, et al. Multi-behavior RFM model based on improved SOM neural network algorithm for customer segmentation [J]. IEEE Access, 2022, 10: 122501–122512.

[16] Rungruang Chongkolnee, et al. RFM model customer segmentation based on hierarchical approach using FCA [J]. Expert Systems with Applications, 2024, 237: 121449.

[17] Li Liangwei, et al. Spending money wisely: online electronic coupon allocation based on real-time user intent detection [C]// Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020.

[18] Uehara Yuki, et al. Robust portfolio optimization model for electronic coupon allocation [J]. INFOR: Information Systems and Operational Research, 2024, 62 (4): 646–660.