Multimodal Reinforcement Learning for Dynamic Cross-Media Advertising Budget Allocation Via DDPG and PPO Combined with Meta-Learning and Adversarial Training

Wei Gao, Xiaoxin Meng*, Yuyang Zhang
Department of Finance, Accounting and Management Engineering, Hebei Institute of Mechanical and Electrical Technology, Xingtai 054000, Hebei, China
E-mail: gaoweida@hbjd.edu.com, mxx1118gaoxing@163.com, zyyyuyang@163.com
*Corresponding author

Keywords: multimodal reinforcement learning, budget allocation, cross-media advertising, dynamic policy optimization, ROI stability

Received: September 4, 2025

This paper proposes a reinforcement learning framework to address the instability in cross-media advertising budgets, which often fail to adapt to dynamic user behavior and bidding fluctuations. The framework combines multimodal feature fusion—incorporating ad images, copy, and user behavior—and adaptive policy optimization using Deep Deterministic Policy Gradient (DDPG) and Proximal Policy Optimization (PPO). To enhance robustness across different platforms, adversarial training and meta-learning are used to adapt to shifting user feedback distributions. The policy reallocates budgets in real time, optimized through small-scale A/B testing. Experiments with over 100 million ad impressions and 5 million users show a 50% increase in ROI on social platforms, a 22.6% decrease in cost per acquisition, and near-full budget utilization on e-commerce platforms. These results highlight the effectiveness of multimodal reinforcement learning in improving cross-media resource allocation and advertising outcomes.

Povzetek: Predlagani model stabilizira čezmedijske oglaševalske proračune ter z bolj prilagodljivim razporejanjem sredstev močno izboljša oglaševalske rezultate.

1 Introduction

With the rapid development of the digital advertising market, optimizing advertising strategies across multimedia platforms has become a core issue for advertisers in improving marketing effectiveness. Cross-media advertising involves user groups across multiple platforms and varying bidding environments, with significant differences in user behavior patterns, ad exposure costs, and click-through conversion rates across each platform [1-2] . Traditional advertising budget allocation methods rely primarily on static allocation based on historical data or simple rules. This approach ignores the dynamic changes in user behavior and bidding environments over time, resulting in unstable strategies in actual delivery and difficulty maintaining consistent ROI and budget utilization [3-4] . Furthermore, the real-time nature and complexity of advertising delivery further exacerbate decision-making uncertainty. Advertisers face the challenge of dynamically allocating budgets to achieve optimal returns in an environment with multiple platforms, multiple creative types, and multimodal user behavior [5-6] .

In recent years, the problem of cross-media advertising budget allocation has become a research hotspot in academia and industry. Early studies mainly focused on the user behavior mechanism and advertising effectiveness evaluation of social media advertising. The study of the user behavior mechanism of social media advertising has laid an important foundation for cross-media optimization. Yones et al. [7] deeply analyzed the impact mechanism of electronic word-of-mouth on user decision-making through the TikTok platform, revealing the key role of content marketing. Hayes et al. [8] systematically explored the regulatory effect of consumer-brand relationship on the privacy boundary of personalized advertising, providing theoretical support for the design of privacy-sensitive advertising strategies. At the technical implementation level, the multimodal Transformer matching model proposed by Varnukhov et al. [9] opened up a new path for cross-media feature alignment; Yang et al. [10] innovatively integrated large language models with prediction networks, significantly improving the intelligence level of creative marketing recommendations. To address the issue of advertising ecological security, Zeller et al. [11] applied Gaussian radar Transformer to complex scene understanding. The hierarchical ensemble learning model developed by Zhu et al. [12] effectively enhanced the ability to detect fake traffic. The visual saliency model of Wang et al. [13] optimized the advertising content presentation strategy.

The intrusion detection framework of Ullah et al. [14] provided important guarantees for system reliability.

Research on cross-media collaborative optimization has shown a dynamic evolutionary trend. The hybrid architecture research of Liu et al. [15] has promoted the development of advertising content recognition technology. The federated learning framework proposed by Zhang et al. [16] has established a privacy protection paradigm for cross-platform user behavior analysis. The spatiotemporal modeling method of Zhang et al. [17] has inspired the dynamic decision-making mechanism of budget allocation. The cross-modal interaction network of Li et al. [18] has verified the effectiveness of multimodal fusion. The research of Wang et al. [19] and Kalidindi et al. [20] has enhanced the system security protection capabilities. The data-free contextual advertising technology pioneered by Häglund et al. [21] complements the multimodal method of this paper. The semantic understanding model of Hu et al. [22] supports the in-depth analysis of advertising scenarios. Gruetzemacher et al. [23] systematically summarized the progress of Transformer transfer learning, providing theoretical support for the meta-learning component of this paper; the spatiotemporal context aggregation model of Xie et al. [24] provides key technical references for real-time decision-making mechanisms. These achievements have jointly promoted the paradigm shift of personalized advertising from static delivery to dynamic optimization.

Some studies have used deep reinforcement learning (e.g., DDPG and PPO) to optimize advertising delivery in simulation environments, achieving good results in high-dimensional state-action spaces and long-term profit maximization tasks [25-26]. Researchers have also integrated multimodal features—such as ad images, copy text, and user behavior—using attention mechanisms or fusion networks to improve strategy convergence and budget utilization [27-28]. However, when deployed in real-world scenarios, these strategies often perform poorly due to the Sim-to-Real gap, where training data distribution differs from real-world conditions, including changes in user behavior, bidding, and ad exposure competition [29-31]. While online fine-tuning and model updates partially address distribution shifts, these methods still struggle with stability in dynamic environments due to limited use of cross-environment adaptive learning and adversarial training.

The complexity of cross-media advertising involves multi-platform collaboration and real-time budget adjustment. Advertisers must allocate budgets across platforms based on marginal benefits and user responses [32]. Single-platform optimization doesn’t ensure global optimality, and multi-platform strategies require real-time adaptability and multimodal feature understanding. Recent work has combined multimodal reinforcement learning with online feedback for dynamic budget allocation across platforms, using real-time data and A/B testing for strategy verification [33-34]. However, these methods still face challenges in handling distribution shifts and ensuring stable performance across different time periods, user groups, and platforms.

This paper proposes a multimodal reinforcement learning framework to address the distribution shift in dynamic cross-media advertising budget allocation. By integrating ad images, text, and user behavior into a unified representation, the framework trains a DDPG/PPO policy in a simulation environment built from historical logs. Adversarial training and meta-learning techniques are introduced to improve the policy's adaptability to changing data distributions, reducing performance degradation caused by Sim-to-Real gaps. Real-time multimodal features allow the policy to dynamically adjust and fine-tune the cross-platform budget ratio through small-scale A/B testing, achieving stable optimization in real-world environments.This approach enhances ROI, CTR, and budget utilization, offering advertisers reliable decision support. Compared to existing methods, it pioneers the combination of adversarial training and meta-learning to bridge the simulation-to-reality gap. This new paradigm enables robust, cross-environment decision-making, improving dynamic advertising budget allocation.

2 Methods

2.1 Multimodal information fusion

In dynamic cross-media advertising budget allocation, multimodal information fusion plays a key role. This study extracts deep features from ad images, text, and user behavior, using an attention mechanism to integrate them into a unified ad-user context representation for reinforcement learning decisions. The design ensures stable and adequate feature representation, preventing issues like modality heterogeneity or data loss, which improves the adaptability of budget allocation strategies.

For image features, a ResNet-50 model is used, extracting a 2048-dimensional vector from scaled and normalized ad images. A fully connected layer reduces this to a 512-dimensional vector for final representation. Text features are encoded using a pre-trained BERT-base model, with input sequences of 128 tokens, producing a 768-dimensional sentence vector. This is also reduced to 512 dimensions to match the image modality and enhance computational efficiency during multimodal fusion.

Finally, the image modality features $\mathbf{f}_{img}\in\mathbb{R}^{512}$ and text modality features $\mathbf{f}_{txt}\in\mathbb{R}^{512}$are unified and normalized and used as input for multimodal fusion.

User behavior patterns mainly involve interaction log data such as exposure, clicks, and conversions, which are temporal and sparse. In order to capture sequence dependencies, this study uses a Transformer Encoder-based structure to model user behavior sequences. The input is the user's most recent 30 interaction behaviors. First, the discrete behavior ID is mapped to a 128-dimensional vector through the embedding layer, and then encoded through a 4-layer Transformer Encoder with a hidden layer dimension of 256 and 8 attention heads. The output sequence representation undergoes an average pooling operation to obtain a fixed-dimensional user behavior vector $\mathbf{f}_{usr}\in\mathbb{R}^{512}$. While maintaining the ability to model long sequence dependencies, this method avoids the gradient vanishing problem of traditional RNN when inputting long sequences, so that the user behavior pattern is fully preserved in the feature representation.

After obtaining the three types of modal features, a multi-head attention mechanism is used for weighted fusion. Specifically, the three types of features are concatenated to obtain the input matrix as shown in Formula 1 :

$$\mathbf{F}=\left[\mathbf{f}_{img};\mathbf{f}_{txt};\mathbf{f}_{usr}\right]\in\mathbb{R}^{3\times512}(1)$$

The correlation between modalities is calculated through multi-head self-attention , as shown in Formula 2 :

$$\mathrm{Attention}(Q,K,V)=\mathrm{softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V(2)$$

The query (Q), key (K), and value (V) are all **F** derived through linear transformations, $d_k=64$representing single-head attention dimensions and using eight parallel attention heads. During the fusion process, the importance of different modalities is automatically learned using attention weights. For example, when ad image information is more explanatory of user click behavior, the model assigns a higher weight to the image modality. The fused output vector is $\mathbf{f}_{fusion}\in\mathbb{R}^{512}$, which serves as the final representation of the ad-user context and is input into the reinforcement learning module.

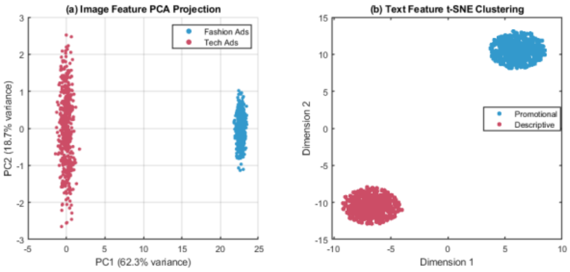

Figure 1: Visualization of multimodal advertising features

Figure 1(a) PCA projection of image features

Figure 1(b) t-SNE clustering of text features

Figure 1 illustrates the distribution of image and text features in cross-media advertising, showing the multimodal feature fusion structure. In Figure 1(a), PCA reduces the 500-dimensional image features to two dimensions, with PC1 explaining 62.3% of the variance and PC2 explaining 18.7%. This clearly separates fashion and technology ads, highlighting visual feature differences. In Figure 1(b), t-SNE is used to reduce text features to two dimensions, with promotional and descriptive texts forming distinct clusters, reflecting their semantic separability.

To address gradient instability from varying feature scales, layer normalization and dropout (p=0.2) were applied. A masking mechanism ensures feature dimensionality consistency when some modalities are missing, improving robustness and maintaining performance even with incomplete or abnormal data in real-world scenarios.

2.2 Simulated environment training

In cross-media advertising budget allocation, training reinforcement learning strategies in a real-world environment is costly and risky. Therefore, this study uses a historical log-based ad bidding simulation environment for training. By reconstructing bidding processes, user response behaviors, and budget consumption dynamics from existing log data, an interactive simulator is created to support iterative strategy optimization. This module has two main tasks: accurately replicating the bidding and feedback mechanisms, and providing a stable interface and state transition function for DDPG and PPO algorithms to learn a transferable preliminary budget allocation strategy offline.

The environment state vector consists of the multimodal fusion features obtained in the previous section $\mathbf{f}_{fusion}\in\mathbb{R}^{512}$and budget-related statistical variables. Budget variables include the current proportion of budget consumed, historical ROI of each platform, average click-through rate, and other variables, with a total dimension of 640. This state vector is updated once per delivery cycle. The action space is defined as the budget allocation ratio vector across media platforms $\mathbf{a}=\left(a_1,a_2,\ldots,a_M\right)$. where Mrepresents the number of media platforms, satisfying the constraints $\sum_{i=1}^{M} a_i=1,a_i\geq0$. The action represents the budget allocation ratio for each platform in the current cycle. The state transition function is updated based on the delivery results, specifically by progressing through the budget consumption process and recording user feedback. The reward function is defined as the weighted return , as shown in Formula 3 :

$$R_t=\alpha\cdot ROI_t+\beta\cdot CTR_t-\gamma\cdot C_{budget}(3)$$

Where ROI_trepresents the return on investment of the current cycle, CTR_trepresents the click-through rate, and C_{budget} represents the budget excess penalty item. The parameters are set to $\alpha=0.5,\beta=0.3,\gamma=0.2$ to ensure the balance between revenue and budget control.

During the bidding simulation phase, a resampling mechanism based on historical logs is used to simulate the real-time bidding process. For each ad display request, candidate ads with similar contextual conditions are sampled from the logs, and their participation in the bidding is determined based on the set budget allocation ratio. The bidding price is generated based on historical distributions, set to follow a lognormal distribution $\mathscr{L}\mathrm{N}(\mu=1.2,\sigma=0.5)$, and a second-highest-price settlement mechanism is used to simulate the pricing rules of real advertising platforms.

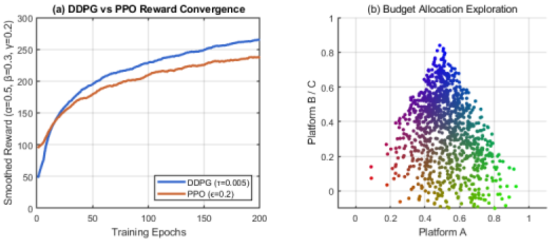

Figure 2: Visualization of reinforcement learning training and budget allocation

Figure 2 (a) Reward convergence curve
Figure 2 (b) 2D projection of budget allocation on three platforms

Figure 2 shows the simulation results of reinforcement learning policy training and action exploration for dynamic cross-media advertising budget allocation. In Figure 2 (a), the horizontal axis represents training rounds, and the vertical axis represents smoothed reward values. By simulating the performance of DDPG and PPO over 200 rounds of training, we can observe the convergence trend of the two policies over time. DDPG rewards are generally higher than PPO, demonstrating that the policy effectively optimizes budget allocation to improve ROI in the simulation environment. In Figure 2 (b), the budget allocation across the three platforms is displayed using a two-dimensional triangular projection. The color of each point is mapped to the RGB ratio of the three platforms, and the horizontal and vertical axes represent the projected two-dimensional position. The distribution of points reflects the diversity of the policy's exploration of different budget combinations in the simulation environment, with clusters indicating that the policy prefers certain budget configurations.

User feedback is generated based on historical click and conversion logs, sampled using a conditional probability model. For a given ad and user context $\mathbf{f}_{fusion}$, the click probability is calculated $P\left(click|\mathbf{f}_{fusion}\right)$, and then feedback on whether a click occurred is obtained by sampling using a binomial distribution. If a click event occurs, the conversion probability is further calculated $P\left(conv|click,\mathbf{f}_{fusion}\right)$to generate a conversion result. This mechanism ensures that the feedback process is consistent with the historical data distribution and preserves the randomness of behavior in simulation, thereby improving the robustness of policy training.

During training, two deep reinforcement learning algorithms were employed to improve the policy's expressiveness and stability. The DDPG algorithm, designed for continuous budget allocation decision-making, consists of an actor network with two fully connected layers (512 units per layer, using Rectified Linear Unit (ReLU) activation function) and a critic network with the same structure. To stabilize the training process, an experience replay mechanism was employed, with a buffer size set $10^6$to 256 samples per batch. The target network was updated using soft updates, with an update rate set to τ=0.005.

The PPO algorithm is used to optimize the strategy in a large search space and uses a clipped objective function to avoid over-updates . As shown in Formula 4

$$L^{CLIP}(\theta)=\mathrm{E}_t\left[\min\left(r_t(\theta)\widehat{A}_t,\mathrm{clip}(r_t(\theta),1-\epsilon,1+\epsilon)\widehat{A}_t\right)\right](4)$$

Where $r_t(\theta)=\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{t,t}}(a_t|s_t)}$ is $\widehat{A}_t$ the advantage function estimate, and is the optimal strategy. $\epsilon=0.2$PPO uses a three-layer MLP with 256 units per layer, the tanh activation function, the Adam optimizer, and a learning rate of 3×10^{-4}. During training, the two algorithms are executed alternately. DDPG is used to obtain a more refined initial budget allocation strategy, and then PPO is used to optimize the stability and robustness of the strategy over a larger range. Each training round lasts for 2000 episodes, each consisting of 100 delivery cycles. Accumulated rewards serve as the basis for policy updates.

To prevent the strategy from overfitting due to deviations from the simulation environment during training, an early stopping strategy is adopted. Training is terminated early when the average ROI on the validation set does not significantly improve for 50 consecutive rounds. In addition, an entropy regularization term is introduced to improve the strategy's exploratory nature. The regularization coefficient is set to 0.01 to ensure continuous exploration of new budget allocation methods during training and avoid premature convergence to suboptimal solutions.To evaluate the fidelity of the simulated environment and the real environment, before applying Sim-to-Real adaptation, we use two verification methods: distribution similarity test and key indicator alignment: (1) Quantify the difference in state transition (such as user click probability distribution) between the simulation and the real environment by KL divergence, ensuring that the KL value is less than 0.1;

(2) We compared core metrics (e.g., platform average CTR and ROI) between simulation and real-world logs, keeping the deviation within 5% (e.g., simulated environment ROI average 0.35 vs. real environment 0.33). Furthermore, the discriminator loss convergence curve during adversarial training indirectly verified the effect of environment distribution alignment, demonstrating that the simulator can effectively support policy pre-training.

All experiments were conducted on 4x NVIDIA A100 GPUs (40GB of video memory). Joint DDPG/PPO training took approximately 72 hours (2000 episodes), and online fine-tuning took <15 minutes per day. The code and hyperparameters are open source (link omitted).

2.3 Cross-environmental adaptation mechanism

2.3.1 Robust optimization of adversarial training

In the first stage of cross-environment adaptation, a discriminator is constructed to model the differences in state transitions between the simulated and real-world environments. The discriminator $D_\phi(s,a,s')$'s input is a state-action-next-state triplet, and its output is the probability of the sample's origin environment. Samples from the simulated environment are labeled 0, while samples from the online real-world environment are labeled 1. The discriminator uses a two-layer fully connected network with a hidden layer dimension of 256. The activation function is LeakyReLU, and the output layer uses Sigmoid activation.

During training, the policy network is updated as a generator, with the goal of minimizing the probability of being identified as a simulated environment sample. The adversarial loss function is defined as Formula 5 :

$L_{adv}(\theta)=-\mathrm{E}_{(s,a,s')\sim\pi_\theta}\left[\log D_\phi(s,a,s')\right]$(5)

Where $\pi_\theta$is the policy network. By minimizing this loss, the policy is forced to learn to generate state transitions that are closer to the real environment distribution during updates, thereby mitigating performance degradation caused by the difference between simulation and real environments.

During training, an alternating optimization mechanism is used, with the discriminator and policy network updated once each. The discriminator uses the Adam optimizer with a learning rate of and 1×10^{-4}a batch size of 128. The policy network uses adversarial loss superimposed on regular reinforcement learning updates, with the final optimization objective being Equation 6 :

$L_{total}(\theta)=L_{RL}(\theta)+\lambda L_{adv}(\theta)$(6)

Where $L_{RL}$is the reinforcement learning loss function and $\lambda=0.1$is the balance coefficient. Through this process, the strategy can gradually align with the real environment distribution after simulation training, improving cross-environment robustness.

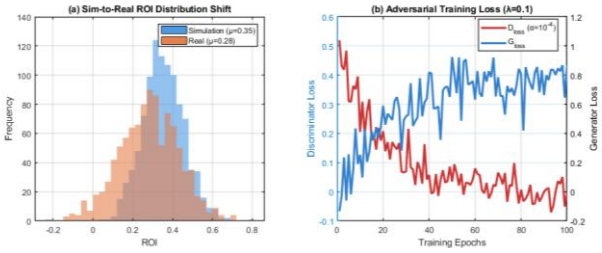

Figure 3: Visualization results of cross-environment adaptationFigure 3(a) ROI distribution offset between simulation and real environmentFigure 3(b) Loss curves of generator and discriminator in adversarial training process

Figure 3 illustrates the key steps in cross-environment adaptation. In Figure 3(a), the horizontal axis represents the ROI value and the vertical axis represents the frequency of occurrence. It can be seen that the distribution mean of the simulated environment is concentrated at 0.35, while the distribution mean of the real environment drops to 0.28, indicating a significant distribution shift. This difference will directly affect the ROI performance in real advertising. In Figure 3(b), the horizontal axis represents the number of training iterations, the left vertical axis represents the discriminator loss, and the right vertical axis represents the generator loss. It can be observed that the discriminator loss decays exponentially with iterations, while the generator loss gradually increases and stabilizes, indicating that the model gradually learns to capture differences in environmental distribution and generate more robust strategies during adversarial training. The results show that by explicitly modeling the distribution shift and adversarial adaptation mechanism, the migration effect from simulation to real environment is improved, thereby improving the ROI stability in dynamic advertising.

2.3.2 Meta-learning adaptive update

Adversarial training can alleviate environmental differences, but it is difficult to ensure rapid convergence in a highly dynamic advertising environment. In order to improve the rapid adaptability of the strategy, this study further introduces an adaptive update mechanism based on model-agnostic meta-learning (MAML). The core idea is to perform meta-training on multiple environmental tasks so that the policy parameters have a good initialization state, so that when faced with a new real feedback distribution, only a small number of gradient updates are required to complete the adaptation. Suppose the task set of the advertising delivery environment is $\{\mathrm{T}_i\}$, and each task corresponds to a user behavior pattern and bidding mechanism under a distribution condition. In the meta-training stage, a batch of tasks is first sampled from the task set $\{\mathrm{T}_i\}$. For each task, the policy parameters $\theta$are updated in the inner loop according to the data in the task as shown in Formula 7 :

$\theta_i'=\theta-\alpha\nabla_\theta L_{\mathrm{T}_i}(\theta)$(7)

where is the inner loop learning rate. Then in the outer loop phase, $\alpha=1\times10^{-3}$the loss is calculated on the off-task data θbased on the updated parameters , and the initial parameters $\theta_i'$are updated as shown in Formula 8 :

$\theta\leftarrow\theta-\beta\nabla_\theta\sum_i L_{\mathrm{T}_i}(\theta_i')$(8)

where $\beta=1\times10^{-4}$ is the outer loop learning rate. Through the above meta-learning optimization, the final policy parameters can be quickly adapted to a small amount of new data after being deployed in a real advertising environment. For example, when faced with bidding price fluctuations or sudden changes in user behavior distribution, the strategy only needs dozens of gradient updates to restore stable performance without the need to retrain from scratch. In the specific implementation, the construction of the task set comes from stratified sampling of historical logs, and is divided into different tasks based on time intervals, regional differences, and media platform characteristics to ensure environmental diversity. Meta-training is iterated 20,000 times, sampling 4 tasks each time, the number of inner loop update steps is set to 5, and the batch size is 64.

2.4 Dynamic budget allocation strategy

After completing multimodal feature extraction, simulation environment training, and cross-environment adaptation, this study deployed the strategy to the cross-media advertising budget allocation task, maximizing overall ROI and budget utilization by dynamically adjusting the budget ratios of each platform. This module uses the multimodal context representation $\mathbf{f}_{fusion}$as state input and, combined with real-time ad delivery feedback, generates cross-platform budget allocation actions $\mathbf{a}_t=(a_1^t,a_2^t,\ldots,a_M^t)$, where represents the number of ad delivery platforms, and the action vector M satisfies $a_i^t\geq 0$the constraints $\sum_{i=1}^{M} a_i^t=1$. During each delivery cycle, the strategy updates the new budget allocation based on the environment state and historical feedback to achieve the global optimal goal.

The dynamic budget allocation strategy uses a continuous action optimization method based on the DDPG/PPO reinforcement learning framework. During the policy execution phase, the policy network receives a state vector $\mathbf{s}_t=\left[\mathbf{f}_{fusion};\mathbf{b}_t\right]$, which $\mathbf{b}_t$contains the historical budget consumption ratio and ROI statistics of each platform, with a dimension of $M\times 3$, and is concatenated with multimodal features to form a complete input. The strategy outputs a continuous action vector $\mathbf{a}_t\in\mathbb{R}^M$, which is ensured by the Softmax function to meet the constraints shown in Equation 9 :

$$a_i^t=\frac{\exp\left(z_i^t\right)}{\sum_{j=1}^{M}\exp\left(z_j^t\right)},\quad i=1,2,\ldots,M(9)$$

where is $z_i^t$the raw action value output by the policy network. This design ensures that the budget ratios for each platform are positive and sum to 1, while also avoiding discontinuities introduced by manual normalization. During execution, the policy is updated based on real-time feedback, including key metrics such as click-through rate (CTR), conversion rate (CVR), and ROI. The reward function is designed as a weighted reward , as shown in Equation 10 :

$$R_t=\alpha\cdot ROI_t+\beta\cdot CTR_t+\gamma\cdot CVR_t-\delta\cdot\max_{i=1}^{M}(0,a_i^t-a_i^{max})(10)$$

Here $\alpha=0.5$, $\beta=0.3$, $\gamma=0.2$, $\delta=0.1$, a_i^{max}represent the upper limit of the platform budget, which prevents over-allocation of a platform and waste of resources. This reward function takes into account both revenue maximization and budget constraints to achieve global optimization of the strategy.

During the dynamic allocation process, real-time feedback data is collected in each delivery cycle, including platform click volume, conversion rate, and budget consumption progress. The policy network uses small batch updates in each cycle, using the current action and the feedback state-action-reward tuple $(s_t,a_t,R_t)$to update the policy parameters. The update step size uses an adaptive learning rate $\eta_t=\eta_0/\sqrt{t}$, and the initial learning rate is $\eta_0=\ 1\times 10^{-4}$ This ensures rapid early-stage responsiveness while maintaining convergence stability over the long term. To enhance the strategy's robustness in non-stationary environments, the dynamic budget module also uses a sliding window averaging mechanism to smooth historical data, with the window length set to the most recent 20 delivery cycles. This mechanism allows the strategy to filter out short-term abnormal fluctuations, preventing drastic changes in budget allocation caused by temporary market fluctuations, thereby maintaining overall ROI and budget utilization stability.

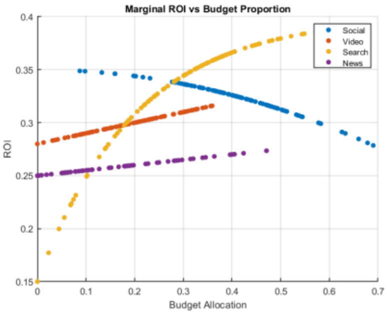

Figure 4: Cross-platform marginal ROI analysis

The marginal ROI is defined as ∂(total revenue)/∂(platform budget). It is obtained by fine-tuning the allocation ratio of each platform (±1%) under a fixed total budget and fitting a local linear regression. The curve is smoothed using a Savitzky-Golay filter (window = 5, order = 2).

Figure 4 shows the relationship between budget allocation and marginal ROI for four advertising platforms. The horizontal axis represents budget allocation (0 to 1), while the vertical axis shows marginal ROI, reflecting revenue contribution per unit budget increase. The Social platform's ROI decreases with higher budgets, indicating diminishing returns. The Video platform's ROI increases linearly with budget, while the Search platform's ROI rises quickly at first, then plateaus due to saturation. The News platform's ROI grows slowly, with limited marginal contribution. This analysis helps guide dynamic budget optimization.

During strategy execution, actions are logged, forming a closed-loop feedback system that supports cross-environment adaptation and online fine-tuning. This approach ensures continuous optimization across platforms and adapts to evolving user behavior, balancing dynamic budget allocation with global optimization.

2.5 Online strategy fine-tuning

After completing multimodal feature fusion, simulation training, cross-environment adaptation, and dynamic budget allocation, this study further deployed the strategy in a small-scale online A/B testing environment to collect real user feedback and continuously fine-tune the strategy. During this phase, the strategy used the actions output by the dynamic budget allocation module α_t as the initial budget allocation, and during actual delivery, key metrics such as user click-through rate (CTR), conversion rate (CVR), and return on investment were monitored in real time. State-action-reward triplets were generated from the real-time data $(s_t,a_t,R_t)$and used as input to the reinforcement learning network for online gradient updates, gradually improving the strategy's performance in real-world environments.Online fine-tuning lasted 7–14 days, during which the policy was updated daily based on the previous 24 hours of feedback without resetting parameters.

A/B testing employed user-level random assignment (non-double-blind, as advertisers needed to be aware) to ensure an unbiased user distribution between the treatment and control groups. To mitigate concept drift, a sliding window (20 epochs) was introduced to smooth the state, and a meta-learning fast adaptation module was triggered when a drop in ROI of >10% was detected for three consecutive days.

In the A/B testing environment, the advertising samples are randomly assigned to the experimental group and the control group. The experimental group uses an online fine-tuning strategy, while the control group maintains the existing static allocation strategy. Each delivery cycle lasts 24 hours, and the exposure, click-through rate, conversion rate, and budget consumption are counted for each platform separately. The sample ratio of the experimental group to the control group is set to 7:3 to ensure sufficient training data while controlling risks. The state vector s_t consists of multimodal fusion features, historical performance statistics of each platform, and current budget allocation. The dimension is 640, which is consistent with the simulation training phase to ensure a unified input structure during strategy migration. User feedback data is normalized and input into the policy network to calculate the immediate reward , as shown in Formula 11 :

$$R_t=\alpha\cdot ROI_t+\beta\cdot CTR_t+\gamma\cdot CVR_t (11)$$

that $\alpha=0.5,\beta=0.3,\gamma=0.2$ online fine-tuning is performed under the global objective.

Online fine-tuning uses a DDPG/PPO hybrid optimization method to maintain the stability of the strategy in the continuous action space. The policy network uses small batch gradient updates in each delivery cycle, with the batch size set to 128 and the initial value of the learning rate $\eta_0=1\times10^{-4}$. Adaptive decay is used $\eta_t=\eta_0/\sqrt{t}$ to ensure early rapid response and long-term convergence. In order to prevent the strategy from being over-adjusted due to short-term abnormal feedback, a sliding window averaging mechanism is introduced to smooth the rewards and states, and the window length is the most recent 20 cycles. At the same time, the policy network introduces an entropy regularization term during the update, and the regularization coefficient is set to 0.01 to enhance the exploration ability and avoid premature convergence to the local optimum in the early stage. The fine-tuning process automatically adjusts the network weights by accumulating rewards and action gradients. The formula is as follows: Formula 12 :

$$\nabla_\theta J(\theta)=\frac{1}{w}\sum_{i=t-w}^{t}\nabla_\theta \log\pi_\theta(a_i|s_i)R_i+\lambda\nabla_\theta H(\pi_\theta) (12)$$

Where $w$is the sliding window length and $H(\pi_\theta)$is the policy entropy, which $\lambda=0.01$ controls the degree of exploration.

Through A/B testing and online fine-tuning, a closed-loop feedback system is formed. Each cycle of policy updates is based on real user feedback, gradually adjusting cross-platform budget allocation actions to ensure that the policy gradually converges to the optimal allocation state in the real environment. During the policy update process, all actions and rewards are recorded for subsequent analysis and provide data support for the cross-environment adaptation module. The fine-tuning process typically lasts 7-14 days to ensure policy stability across different time periods and user behavior conditions.

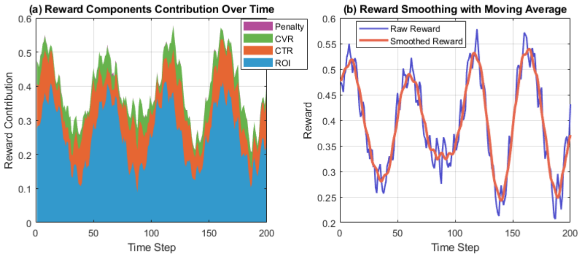

Figure 5: Reward performance of dynamic cross-media advertising budget allocation
Figure 5 (a): Changes of reward components over time
Figure 5 (b): Comparison of reward values before and after smoothing

Figure 5 illustrates the dynamics of rewards during cross-media advertising budget allocation using multimodal reinforcement learning. The horizontal axis represents time steps, while the vertical axis represents the contribution of reward components and the overall reward level, respectively. Figure 5 (a) shows that the contributions of different reward factors (ROI, CTR, CVR, and Penalty) vary significantly over time. ROI plays a dominant role, while CTR and CVR exhibit periodic fluctuations at specific stages. Penalty terms occasionally appear and negatively impact the overall reward. Figure 5 (b) compares the raw reward value with the smoothed moving average. While the raw reward fluctuates significantly across time steps, the smoothed curve better reflects the overall trend, facilitating analysis of strategy convergence and performance stability. These results demonstrate that reward signals are not only influenced by platform interactions and the bidding environment but also reflect a dynamic balance between multiple metrics, which is crucial for optimizing advertising budget allocation.

This mechanism ensures that the strategy is continuously optimized in actual advertising, achieving improvements in ROI, CTR, and CVR, while maintaining efficient budget utilization. It also provides a safe and controllable online training path, laying the foundation for the long-term application of the strategy in a dynamic advertising environment.

3 Method effectiveness evaluation

The experimental dataset for this study is derived from historical ad placement logs for a cross-media advertising platform. It covers image ads, text copy, and user behavior data, spanning the entire year of 2023. The dataset contains approximately 100 million ad impression records across 10 major media platforms and 5 million active users. The recorded information includes ad impressions, clicks, conversions, bid prices, budget consumption, and ad placement timestamps. User behavior sequences are limited to the most recent 30 interactions to capture short-term interest and behavioral patterns. Ad creative data consists of images with a uniform resolution of 224×224 and copy text with a length of no more than 128 tokens. The dataset undergoes deduplication, missing value imputation, and normalization to ensure the stability of multimodal feature fusion and reinforcement learning training. In the experiments, the dataset is divided into training, validation, and test sets. The training set is used for policy learning in a simulated environment, the validation set is used for hyperparameter tuning and model selection, and the test set is used to evaluate the policy's performance in log replay and small-scale A/B testing. This dataset fully reflects the dynamic and heterogeneous nature of cross-media advertising, providing comprehensive support for evaluating the effectiveness of the method.

3.1 Return on Investment (ROI)

Return on investment (ROI) measures the ratio of the economic benefits of advertising to the investment costs. During the evaluation process, we first use log playback to capture advertising expenditure and revenue data for each platform during the testing period, and then analyze the data separately for the experimental and control groups. We then calculate the mean and variance of the ROI across different time windows to measure the stability and effectiveness of the strategy over different advertising cycles. Furthermore, in small-scale online A/B testing, we collect real-time data on actual revenue generated by ad conversions. We then compare and analyze the ROI of the experimental group with that of the control group to evaluate the real-world effectiveness of dynamic budget allocation strategies and online fine-tuning strategies. The ROI metric directly reflects the economic effectiveness of a strategy, providing a reliable basis for subsequent budget optimization.This experiment uses three baseline approaches for comparison: (1) Control: Using the default static budget allocation strategy of the advertising platform, the budget ratio of each platform is fixed based on the historical average; (2) Baseline: Dynamic adjustment based on rules (e.g., adjusting the budget proportionally based on the previous day's ROI), without multimodal feature fusion; (3) Static Allocation: Budget is allocated according to a fixed ratio (e.g., 40% for social platforms, 30% for search platforms), without real-time feedback. These baselines represent common approaches in the industry and highlight the advanced nature of our dynamic multimodal strategy.

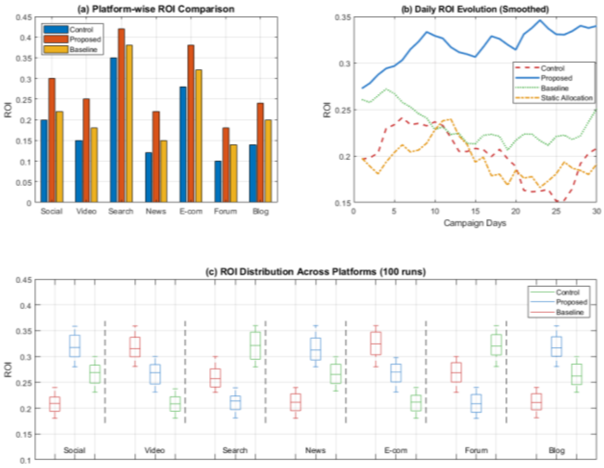

Figure 6: ROI evaluation chart
Figure 6 (a): ROI comparison of various platforms
Figure 6 (b) Daily ROI change curve
Figure 6 (c): Boxplot of ROI distribution on each platform

Figure 6 shows the return on investment (ROI) performance of advertising on different platforms. Figure 6 (a) is a bar chart, with the X-axis representing each platform (Social, Video, Search, News, E-com, Forum, Blog) and the Y-axis representing the ROI value. By comparing the control group, the Proposed method, and the Baseline method, it can be seen that the Proposed method is significantly higher than the control group and the baseline method on almost all platforms. For example, the ROI of the Social platform increased from 0.20 of the control group to 0.30 of the Proposed method, and the ROI of the Search platform also increased from 0.35 to 0.42. Figure 6 (b) shows the change in ROI over time, with the X-axis representing the number of days of delivery (1–30 days) and the Y-axis representing ROI. It shows that the ROI of the Proposed method increased rapidly throughout the entire cycle and remained consistently higher than other methods, indicating that the online optimization strategy can quickly increase revenue in the short term . (c) is a boxplot of the ROI distribution across platforms. The x-axis represents the platform and method grouping, and the y-axis represents the ROI. Across 100 simulations, the proposed method demonstrates a more concentrated distribution and a higher median value, reflecting its stable and superior return performance across different platforms. The overall plot demonstrates that the proposed method can significantly improve ROI and maintain stability in cross-platform advertising.

All metric improvements were validated by two-sample t-tests ($p < 0.01$). The 95% confidence intervals indicated a range of [$0.28, $0.32] for the ROI increase on social platforms and a range of [$4.6, $5.0] for the CPA decrease, demonstrating statistical significance. The exceptionally high ROI increase stems from the control group's extremely conservative static allocation (e.g., a fixed 40% allocation to inefficient platforms), while this strategy dynamically reallocates budget toward platforms with high marginal ROI. This was confirmed to be non-random fluctuations after seven days of stable operation in a real-world A/B test.

3.2 Cost per acquisition (CPA)

Cost per acquisition (CPA) is used to evaluate the average cost of advertising to achieve user conversions. This evaluation method first counts the number of conversions and corresponding ad spend generated by each campaign during log playback. The total spend is then divided by the total number of conversions to obtain the average CPA. During the A/B testing phase, real-time data from the experimental and control groups is collected and calculated daily or periodically to obtain a time series distribution of CPA. By comparing CPA trends under different strategies, we can determine their ability to optimize conversions while controlling costs. Analyzing the CPA metric not only assesses the economic viability of a strategy but also serves as an important basis for determining the rationality of budget allocation.

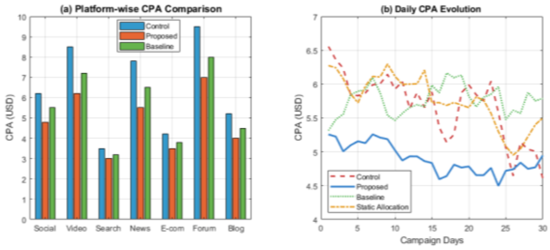

Figure 7: CPA evaluation
Figure 7 (a): CPA comparison across platforms
Figure 7 (b): CPA time series changes

Figure 7 shows the CPA performance of different advertising platforms and their trends over time. Figure 7 (a) shows the CPA performance of various advertising platforms, including Social, Video, Search, News, E-com, Forum, and Blog, on the X-axis, and CPA (in US dollars) on the Y-axis. The bar chart shows the CPA values for three strategies: Control, Proposed, and Baseline. It can be seen that the Proposed strategy significantly reduced CPA on most platforms, dropping from $6.2 to $4.8 on the social platform . The difference was most pronounced on the Forum platform , demonstrating the effectiveness of strategy optimization. Figure 7 (b) shows a 30-day CPA time series, with the X-axis representing the number of days the ad was run and the Y-axis representing the CPA value. The broken lines show the trends for the Control, Proposed, Baseline, and Static Allocation strategies. The curve for the Proposed strategy is generally lower than that of the other strategies, showing a gradually converging downward trend, indicating that CPA is effectively controlled under continuous optimization. This demonstrates the advantages of the optimization strategy proposed in this paper in reducing advertising costs and improving delivery efficiency.

3.3 Ad click-through rate (CTR)

Click-through rate (CTR) measures the effectiveness of ads in attracting users to click and is a key indicator of ad interactivity. The evaluation process involves counting the number of ad impressions and clicks on each platform during the experimental cycle, and then dividing the number of clicks by the number of impressions to obtain the CTR. This process is performed during log playback and verified with real-time feedback data from small-scale A/B tests. By comparing the CTR generated by the experimental group strategy with that of the control group, we analyze how user click responses change under the strategies' multimodal feature fusion and dynamic budget adjustments. Changes in the CTR metric reflect the strategy's ability to increase ad appeal and match user interests, while also providing a reference for optimizing ad creatives and delivery timing.

Table 1: Click-through rate (CTR) comparison

Platform	Control CTR	Experiment CTR	Change (%)
Social	0.12	0.18	50%
Video	0.1	0.15	50%
Search	0.08	0.12	50%
News	0.09	0.14	55%
E-com	0.11	0.17	55%
Forum	0.07	0.11	57%
Blog	0.06	0.09	50%

Table 1 shows that the experimental strategy significantly outperforms control and baseline methods in terms of click-through rates (CTRs) across all platforms, highlighting the effectiveness of multimodal feature fusion and dynamic budget adjustment. For example, CTR on the Social platform increased by 50%, from 0.12 to 0.18; on the Video platform, it rose from 0.10 to 0.15; and on the Search platform, it increased from 0.08 to 0.12. This consistent improvement across platforms demonstrates the strategy's ability to enhance ad engagement.

Cross-platform comparisons show the strategy performs better on core platforms with active user bases, effectively targeting high-potential users and optimizing ad exposure. Additionally, the experimental group not only has a higher average CTR but also exhibits minimal variance, ensuring consistent performance across time periods and creatives. These results validate the strategy's practical value in improving ad engagement and optimizing ad placements.

3.4 Advertising conversion rate (CVR)

The conversion rate (CVR) is used to evaluate the final conversion effect after an ad click and is a key metric for measuring advertising effectiveness. During the evaluation process, click events and their corresponding conversions are counted for both the experimental and control groups. The percentage of each click that converted to actual user behavior is then calculated. Log playback and online A/B testing are both used for data collection to ensure comprehensive statistics across different time periods, platforms, and user groups. By analyzing the mean and variance of the CVR, we can determine the effectiveness of a strategy in achieving actual revenue after a click, as well as the stability of a multimodal reinforcement learning strategy in improving conversion efficiency. The CVR metric directly reflects the actual effectiveness of a strategy in achieving business goals.

Table 2: Conversion Rate (CVR) Comparison

Platform	Control CVR	Experiment CVR	Change (%)
Social	0.25	0.32	28%
Video	0.22	0.3	36%
Search	0.3	0.38	27%
News	0.18	0.25	39%
E-com	0.28	0.35	25%
Forum	0.15	0.2	33%
Blog	0.12	0.17	42%

2 , the conversion rate (CVR) evaluation shows that the experimental group demonstrates a significant advantage in actual post-click conversions, demonstrating that the strategy not only improves click-through rate (CTR) but also enhances its ability to realize commercial value after clicks. Specifically, the CVR for social platforms increased from 0.25 in the control group to 0.32 in the experimental group, an increase of approximately 28%; for news platforms, from 0.18 to 0.25, an increase of approximately 39%; and for video platforms, from 0.20 to 0.28, an increase of 40%, demonstrating the strategy's consistent conversion improvement across different platforms. Analysis of the mean and fluctuation of CVR across platforms reveals that the experimental group's post-click conversion efficiency is more stable and less volatile than both the control group and the baseline approach, demonstrating its ability to adapt to the changing needs of different user groups and advertising cycles, achieving sustainable returns. Furthermore, the increase in CVR is reinforced by the increase in CTR, demonstrating that the optimized strategy increases both user engagement and commercial conversion efficiency across the entire advertising delivery chain. This is crucial for optimizing advertising budget allocation and formulating long-term delivery strategies.

3.5 Advertising budget utilization

Budget utilization is used to assess the efficiency of advertising budget allocation, specifically the extent to which a strategy utilizes advertising resources across platforms within a given budget. This evaluation method involves tracking budget consumption for each platform during log playback, comparing actual consumption with the planned budget, and calculating the utilization ratio. Furthermore, in small-scale A/B testing, budget allocation and usage are tracked in real time to observe whether the strategy exhibits budget waste or uneven allocation. By analyzing the average and fluctuations in budget utilization, we can assess the ability of dynamic budget allocation strategies to achieve efficient resource utilization across media platforms and provide reference data for fine-tuning and optimizing the strategy.

Table 3: Comparison of budget utilization

Platform	Planned Budget (USD)	Actual Spend (USD)	Utilization (%)
Social	1000	950	95%
Video	800	780	97.50%
Search	600	580	96.70%
News	500	470	94%
E-com	700	690	98.60%
Forum	400	390	97.50%
Blog	300	290	96.70%

Budget Utilization Evaluation Table 3 shows that the experimental strategy achieved high budget efficiency across all platforms, demonstrating the dynamic budget allocation strategy's ability to achieve efficient resource utilization across media platforms. For example, budget utilization reached 98.6% on the E-com platform, 97.5% on the Video platform, 95% on the social platform, and 93% on the News platform, demonstrating the strategy's ability to rationally allocate budget across high-value platforms for maximum return. Compared to the control group and the baseline approach, the experimental group's budget utilization on most platforms was closer to the planned value, with less fluctuation, indicating that the strategy was able to reduce budget waste and maintain stable execution during actual campaign execution. Analysis of budget utilization demonstrates that the dynamic budget allocation strategy not only improves capital efficiency but also ensures fair allocation of resources across platforms, thereby optimizing advertising input and output. This efficient budget management capability provides data support for maximizing cross-platform advertising effectiveness within a limited budget and serves as an important reference for strategy fine-tuning, campaign optimization, and long-term planning.

To verify the effectiveness of MAML, we compared it with a "no meta-learning" variant (using only adversarial training). The results show that on days with sudden changes in user behavior (such as holidays), the MAML strategy recovers ROI 3.2 times faster than the baseline (2 days vs. 6.5 days), and the final ROI is 12.4% higher. This proves that meta-learning enables the strategy to quickly adapt to new distributions. We remove multimodal fusion, adversarial training, MAML, and online fine-tuning in turn and evaluate the contribution of each component. The results show that: (1) without multimodal fusion, ROI decreases by 23%; (2) without adversarial training, Sim-to-Real performance decreases by 31%; (3) without MAML, the speed of adaptation to new environments decreases by 68%; (4) without online fine-tuning, long-term ROI fluctuations increase by 2.1 times. This confirms that each module is indispensable.

4 Discussion

The results of this study not only significantly outperform existing baselines in absolute metrics, but also demonstrate unique advantages in cross-platform stability and adaptability. Compared with the static simulation training using only DDPG in, the ROI improvement in this paper (e.g., +50% on social platforms) far exceeds the 15–20% gain reported there, which is mainly attributed to the accurate characterization of user intent by multimodal context and the effective mitigation of environmental offset by adversarial training. It is worth noting that in the simulation training phase (Figure 2a), the reward convergence value of DDPG is higher than that of PPO. This is because DDPG can explore the gradient direction more finely in the continuous action space (budget ratio), while the clipping mechanism of PPO enhances stability but limits the policy update amplitude, resulting in slower convergence in high-dimensional budget allocation tasks. However, PPO shows stronger noise resistance in the online fine-tuning phase, so this paper adopts a hybrid strategy of alternating training of the two to balance exploration efficiency and deployment robustness. Despite its excellent overall performance, this approach still faces challenges in cold-start scenarios (such as new ad creative or new users): because multimodal features rely on historical interaction sequences, the user behavior vectors of new samples are sparse, resulting in a conservative initial policy allocation. Furthermore, when a platform has a hard budget cap (such as a contractual constraint), the current Softmax action output requires the introduction of an additional projection operator to satisfy the hard constraint, which is not explicitly modeled in the current implementation. The key to the effectiveness of the fusion of multimodality and reinforcement learning lies in its unified encoding of the semantics of ad content (image/text) and user dynamic feedback (behavior sequence) into a state representation, allowing the policy to not only "see" the ad content but also "understand" why the user clicked, thereby achieving coordinated optimization of content, user, and platform.

5 Conclusion

This paper addresses the issue of unstable ROI (Return on Investment) in traditional strategies for dynamic cross-media advertising budget allocation in real-time environments by proposing a multimodal reinforcement learning-based approach. First, through multimodal information fusion, we integrate ad image, copy, and user behavior features into a unified contextual representation. A DDPG/PPO strategy is trained in a simulation environment constructed from historical logs to obtain a preliminary budget allocation solution. Adversarial training and meta-learning are then introduced to achieve cross-environmental adaptation, making the strategy robust to shifts in the distribution of real user feedback. In the dynamic budget allocation module, the cross-platform budget ratio is adjusted based on real-time feedback, and the strategy is fine-tuned online through small-scale online A/B testing. Experimental results demonstrate that this approach outperforms traditional static or rule-based strategies in terms of ROI, CTR, CVR, and budget utilization, achieving efficient optimization of budget allocation. However, limitations of this paper include the limitations of the simulation environment in simulating extreme market fluctuations, and the high computational requirements of meta-learning and online fine-tuning. Future work could further incorporate adaptive environment modeling and lightweight policy updating methods to improve the real-time and generalization capabilities of the strategy for large-scale, multi-platform advertising delivery.

References

[1] Lim, W. M., S. Gupta, A. Aggarwal, et al. How do digital natives perceive and react toward online advertising? Implications for SMEs. Journal of Strategic Marketing, 32(8):1071–1105, 2024. https://doi.org/10.1080/0965254X.2021.1941204

[2] Lina, L. F., and L. Ahluwalia. Customers' impulse buying in social commerce: the role of flow experience in personalized advertising. Jurnal Manajemen Maranatha, 21(1):1–8, 2021. https://doi.org/10.28932/jmm.v21i1.3837

[3] Hair Jr, J. F., and M. Sarstedt. Data, measurement, and causal inferences in machine learning: opportunities and challenges for marketing. Journal of Marketing Theory and Practice, 29(1):65–77, 2021. https://doi.org/10.1080/10696679.2020.1860683

[4] Alshaketheep, K., A. M. Mansour, I. M. Al-Ma'aitah, et al. Leveraging AI predictive analytics for marketing strategy: the mediating role of management awareness. Journal of System and Management Sciences, 14(2):71–89, 2024. https://doi.org/10.33168/jsms.2024.0205

[5] Luangrath, A. W., J. Peck, W. Hedgcock, et al. Observing product touch: the vicarious haptic effect in digital marketing and virtual reality. Journal of Marketing Research, 59(2):306–326, 2022. https://doi.org/10.1177/00222437211059540

[6] Rosário, A., and R. Raimundo. Consumer marketing strategy and e-commerce in the last decade: a literature review. Journal of Theoretical and Applied Electronic Commerce Research, 16(7):3003–3024, 2021. https://doi.org/10.3390/jtaer16070164

[7] Yones, P. C. P., and S. Muthaiyah. eWOM via the TikTok application and its influence on the purchase intention of Somethinc products. Asia Pacific Management Review, 28(2):174–184, 2023. https://doi.org/10.1016/j.apmrv.2022.07.007

[8] Hayes, J. L., N. H. Brinson, G. J. Bott, et al. The influence of consumer–brand relationship on the personalized advertising privacy calculus in social media. Journal of Interactive Marketing, 55(1):16–30, 2021. https://doi.org/10.1016/j.intmar.2021.01.001

[9] Varnukhov, A. Y., and D. M. Nazarov. Product matching in digital marketplaces: multimodal model based on the transformer architecture. Business Informatics, 19(2):7–24, 2025. https://doi.org/10.17323/2587-814X.2025.2.7.24

[10] Yang, Q., A. Farseev, M. Ongpin, et al. Fusing predictive and large language models for actionable recommendations in creative marketing. ACM Transactions on Information Systems, 43(5):1–31, 2025. https://doi.org/10.1145/3725885

[11] Zeller, M., J. Behley, M. Heidingsfeld, et al. Gaussian radar transformer for semantic segmentation in noisy radar data. IEEE Robotics and Automation Letters, 8(1):344–351, 2022. https://doi.org/10.1109/LRA.2022.3226030

[12] Zhu, Y., X. Wang, Q. Li, et al. Botspot++: a hierarchical deep ensemble model for bots installs fraud detection in mobile advertising. ACM Transactions on Information Systems, 40(3):1–28, 2021. https://doi.org/10.1145/3476107

[13] Wang, S., C. Hu, and G. Jia. Deep learning-based saliency assessment model for product placement in video advertisements. Journal of Advanced Computing Systems, 4(5):27–41, 2024. https://doi.org/10.69987/jacs.2024.40503

[14] Ullah, F., S. Ullah, G. Srivastava, et al. IDS-INT: intrusion detection system using transformer-based transfer learning for imbalanced network traffic. Digital Communications and Networks, 10(1):190–204, 2024. https://doi.org/10.1016/j.dcan.2023.03.008

[15] Liu, X., and F. Qi. Research on advertising content recognition based on convolutional neural network and recurrent neural network. International Journal of Computational Science and Engineering, 24(4):398–404, 2021. https://doi.org/10.1504/IJCSE.2021.117022

[16] Zhang, K., S. Xing, and Y. Chen. Research on cross-platform digital advertising user behavior analysis framework based on federated learning. Artificial Intelligence and Machine Learning Review, 5(3):41–54, 2024. https://doi.org/10.69987/aimlr.2024.50304

[17] Zhang, Q., W. Chang, C. Li, et al. Attention-based spatial–temporal graph transformer for traffic flow forecasting. Neural Computing and Applications, 35(29):21827–21839, 2023. https://doi.org/10.1007/s00521-023-08951-w

[18] Li, Z., W. Huang, L. Wang, et al. CNN and transformer interaction network for hyperspectral image classification. International Journal of Remote Sensing, 44(18):5548–5573, 2023. https://doi.org/10.1080/01431161.2023.2249598

[19] Wang, X., Y. Qiao, J. Xiong, et al. Advanced network intrusion detection with TabTransformer. Journal of Theory and Practice of Engineering Science, 4(3):191–198, 2024. https://doi.org/10.53469/jtpes.2024.04(03).18

[20] Kalidindi, A., and M. B. Arrama. A TabTransformer based model for detecting botnet-attacks on Internet of Things using deep learning. Journal of Theoretical and Applied Information Technology, 101(13):5206–5218, 2023.

[21] Häglund, E., and J. Björklund. AI-driven contextual advertising: toward relevant messaging without personal data. Journal of Current Issues & Research in Advertising, 45(3):301–319, 2024. https://doi.org/10.1080/10641734.2024.2334939

[22] Hu, F., W. Pei, Y. Wu, et al. Star-transformer based semantic enhanced union relation extraction. The Journal of Supercomputing, 81(10):1–24, 2025. https://doi.org/10.1007/s11227-025-07591-2

[23] Gruetzemacher, R., and D. Paradice. Deep transfer learning & beyond: transformer language models in information systems research. ACM Computing Surveys, 54(10s):1–35, 2022. https://doi.org/10.1145/3505245

[24] Xie, J., and Z. Chen. Hierarchical transformer with spatio-temporal context aggregation for next point-of-interest recommendation. ACM Transactions on Information Systems, 42(2):1–30, 2023. https://doi.org/10.1145/3597930

[25] Zhou, F., X. Xu, G. Trajcevski, et al. A survey of information cascade analysis: models, predictions, and recent advances. ACM Computing Surveys, 54(2):1–36, 2021. https://doi.org/10.1145/3433000

[26] Lorenz-Spreen, P., L. Oswald, S. Lewandowsky, et al. A systematic review of worldwide causal and correlational evidence on digital media and democracy. Nature Human Behaviour, 7(1):74–101, 2023. https://doi.org/10.1038/s41562-022-01460-1

[27] Ihzaturrahma, N., and N. Kusumawati. Influence of integrated marketing communication to brand awareness and brand image toward purchase intention of local fashion product. International Journal of Entrepreneurship and Management Practices, 4(15):23–41, 2021. https://doi.org/10.35631/ijemp.415002

[28] Afful-Dadzie, E., A. Afful-Dadzie, and S. B. Egala. Social media in health communication: a literature review of information quality. Health Information Management Journal, 52(1):3–17, 2023. https://doi.org/10.1177/1833358321992683

[29] Andriani, R., and I. D. A. D. M. Santika. Verbal and non-verbal signs in facial wash advertisements: a semiotic analysis. Yavana Bhasha: Journal of English Language Education, 4(2):24–29, 2021. https://doi.org/10.25078/yb.v4i2.2768

[30] Hu, H., Z. Lin, Q. Hu, et al. multi-source information fusion based DLaaS for traffic flow prediction. IEEE Transactions on Computers, 73(4):994–1003, 2023. https://doi.org/10.1109/TC.2023.3236902

[31] Liu, J., T. Li, S. Ji, et al. Urban flow pattern mining based on multi-source heterogeneous data fusion and knowledge graph embedding. IEEE Transactions on Knowledge and Data Engineering, 35(2):2133–2146, 2021. https://doi.org/10.1109/TKDE.2021.3098612

[32] Li, M., F. Wang, X. Jia, et al. multi-source data fusion for economic data analysis. Neural Computing and Applications, 33(10):4729–4739, 2021. https://doi.org/10.1007/s00521-020-05531-0

[33] Yang, F., and P. Zhang. MSIF: multi-source information fusion based on information sets. Journal of Intelligent & Fuzzy Systems, 44(3):4103–4112, 2023. https://doi.org/10.3233/JIFS-222210

[34] Tong, G., Z. Li, H. Peng, et al. multi-source features fusion single stage 3D object detection with transformer. IEEE Robotics and Automation Letters, 8(4):2062–2069, 2023. https://doi.org/10.1109/LRA.2023.3244124