CDIVR: Cognitive Dissonance-aware Interactive Video Recommendation

Sicong Liu[1], Ke Yan[1,2](✉), Haojie Shi[1], and Ming Jia[1]

[1] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China
{202321081425,202322081428}@std.uestc.edu.cn,
kyan@uestc.edu.cn,
mjia7772@163.com

[2] KASH Institute of Electronics and Information Industry, Kashi, China

Abstract. Interactive video recommendation (IVR) plays a crucial role in enhancing user engagement and satisfaction in short-video platforms. Despite its benefits, IVR suffers from the filter bubble issue, primarily addressed through diversity-based strategies to mitigate overexposure effect. However, such strategies may lead to cognitive dissonance, which can negatively impact overall satisfaction when contents are recommended to the users that deviates from their expectations. Few studies in IVR have adequately considered the potential impact of cognitive dissonance on user experience and offered effective strategies. Therefore, we propose a cognitive dissonance-aware interactive video recommendation (CDIVR) model. Specifically, a cognitive reward model (CRM) is proposed to comprehensively estimate user satisfaction from three aspects: user preference, overexposure effect, and cognitive bias. Additionally, a denoised state tracker (DST) is also employed to the reduce noise interference and accurately track the changes in user states through cognitive state representation and adaptive denoising. Experiments on the KuaiRec and KuaiRand datasets demonstrate that the CDIVR model can effectively alleviate cognitive dissonance issue and outperform the baselines by 17.84% and 11.98% in cumulative user satisfaction, respectively. The code is available at: `https://github.com/sana-mine/CDIVR-codes`.

Keywords: Interactive video recommendation · Cognitive dissonance · Reinforcement learning · Adaptive denoising

1 Introduction

Short-video platforms, such as TikTok and Kuaishou, have rapidly attracted billions of users through personalized recommendations, which has sparked significant research in video recommendation [27]. Unlike traditional static recommendation methods, interactive video recommendation (IVR) can dynamically adjust the recommendation strategies based on real-time user feedback, better aligning with users' actual needs [9]. This interactivity not only improves the accuracy of recommendations but also enhances user satisfaction, making IVR an important area of study [17]. Current research in IVR primarily employs reinforcement learning (RL) methods to optimize the recommendation strategies by maximizing cumulative user satisfaction [1]. However, existing methods in IVR often overlook the impact of cognitive effects. Research in behavioral economics suggests that the cognitive effects can significantly influence user satisfaction and decision-making behavior [4], providing a direction for further improvement in enhancing the user experience during the interaction process.

The “filter bubble” is a common cognitive effect in IVR, where algorithms tend to recommend information highly consistent with users’ past behaviors, leading to information homogeneity [26]. The main approach to address the filter bubble is to offer diverse contents to overexposure effect caused by the repeated recommendation of similar videos [12]. However, the diversity-based approaches may lead to "cognitive dissonance", where users may feel discomfort when the recommended content conflicts with their expectations [23]. For instance, recommending fitness videos to users who are not interested in sports may decrease their satisfaction. Existing methods overlook the impact of cognitive dissonance, resulting in a suboptimal user experience. Furthermore, user historical interaction data may contain noise, such as random browsing behavior and data collection errors [30], which can affect the model’s ability to assess cognitive dissonance and lead to unreasonable recommendations. Therefore, it’s necessary to perceive the user’s cognitive dissonance and filter out the noise during the interaction process.

To address the above issues, we propose a cognitive dissonance-aware interactive video recommendation (CDIVR) model, combining a cognitive reward model (CRM) and a denoised state tracker (DST) within an offline reinforcement learning (Offline RL) framework. Specifically, CRM is proposed to estimate the degree of user preference, overexposure effect, and cognitive bias from interaction data to comprehensively represent user satisfaction while serving as a simulated environment to guide policy training. By incorporating a counterfactual mechanism, the model can learn the recommendation strategy that reduces the effect of cognitive dissonance, thereby enhancing user long-term satisfaction. Moreover, the DST can reduce noise interference and accurately track the changes in user states through cognitive state representation and adaptive denoising. These components enable the CDIVR model to offer effective recommendation strategies that alleviate cognitive dissonance, thereby enhancing user engagement and satisfaction.

Our contributions are summarized as follows:

- To our best knowledge, this is the first work to address the issue of cognitive dissonance in IVR. We propose a CRM to effectively balance between cognitive dissonance and overexposure effect, thereby providing more accurate real-time feedback.
- We propose a DST module, which can track the changes in user states and mitigate the impact of noise through cognitive state representation and adap-

tive denoising, thereby enabling the model to stably learn the strategies for alleviating cognitive dissonance.

- We conduct extensive experiments on two real-world datasets, where the CDIVR model outperforms the baselines in terms of significantly enhancing cumulative user satisfaction. The hyperparameter analysis and case study show that the CDIVR model can effectively alleviate cognitive dissonance.

2 Related Work

2.1 Interactive Recommendation Systems

Interactive Recommendation Systems (IRS) represent a dynamic recommendation mechanism that captures user behavior and feedback in real-time adjusting recommendation strategies to meet the personalized needs of users. Current research in IRS primarily utilizes RL techniques to capture the complex changes in user preferences [9]. Chen et al. [3] proposes a deep RL-based method capable of optimizing recommendation strategies through real-time user feedback. Zou et al. Wang et al. [16] further improves the long-term user engagement in IRS by incorporating exploration and regularization in the latent action space. On the other hand, some studies [17,21] combine causal inference method to enhance the performance of IRS. However, above studies do not provide a comprehensive analysis of the impact of cognitive effects on IRS. Since the short-video platforms contain a lot of interactive behaviors available for analysis, this paper focuses on the IVR task and examines the impact of cognitive effects on user satisfaction.

2.2 Cognitive Effect in Recommendation

In recent years, the impact of cognitive effects on user behavior in recommendation processes has garnered significant attention from the academic community [15]. Filter bubble is a common cognitive effect in recommendation. Existing methods address this issue through exploration [18], opinion injection [7], and fairness-based recommendations [20], which essentially aim to prevent overexposure effect by promoting recommendation diversity [12]. However, these methods in recommendation overlook the potential for excessive diversity to cause cognitive dissonance among users. Some researches [19,24] highlight the necessity for recommendation systems to address cognitive dissonance to maintain user satisfaction. Therefore, this paper aims to enhance user long-term satisfaction by alleviating cognitive dissonance issue.

3 Preliminaries

In IVR, the dynamic interaction strategies can be learned through Offline RL methods. Additionally, the fast fourier transform (FFT) is an efficient frequency domain transformation technique, which can be utilized for deniosing in interaction sequences. In this section, we will introduce the relevant concepts of Offline RL and FFT.

3.1 Offline Reinforcement Learning

The objective of RL is to enable an agent to continuously interact with the environment and learn a policy that maximizes cumulative rewards. RL tasks are typically modeled in the form of a Markov Decision Process, represented as a quintuple $\langle A, S, P, R, \mu \rangle$. This includes the action space A, state space S, state transition probabilities $P : S \times S \times A \to [0, 1]$, reward function $R : S \times A \to \mathbb{R}$, and discount factor $\mu \in [0, 1]$. The core of RL lies in learning a strategy that maximizes cumulative returns. The value function evaluates the expected cumulative reward starting from a certain state s and following a specific policy π. It is defined as $V^{\pi}(s) = \mathbb{E}_{\pi}\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \mid S_t = s\right]$, where R_{t+k+1} is the reward at time step $t + k + 1$.

Offline RL has garnered significant attention within the field of recommendation systems [5]. Offline RL utilizes historical data for policy training without real-time environmental interaction. In contrast to online RL methods, the Offline RL method can minimize risks associated with direct interactions.

In IVR, the Offline RL method employs a trained reward model to evaluate the videos recommended by the agent, thereby simulating user feedback. Based on the simulated feedback, the agent refines the recommendation policy, which is subsequently validated for its effectiveness in real-world settings.

3.2 Fast Fourier Transform

Discrete Fourier Transform (DFT) converts time series data into the frequency domain, revealing a signal's frequency components, essential for time series analysis. For a sequence $\mathbf{x} = [x_0, x_1, \ldots, x_{N-1}]$, the 1D DFT is defined as:

$$X[k] = \sum_{n=0}^{N-1} x_n \cdot e^{-j\frac{2\pi}{N}kn}, \quad k = 0, 1, \ldots, N-1 \tag{1}$$

where $X[k]$ represents the frequency domain, j is the imaginary unit, and k the frequency index. The Inverse DFT (IDFT) reconstructs the original signal:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] \cdot e^{j\frac{2\pi}{N}kn}, \quad n = 0, 1, \ldots, N-1 \tag{2}$$

DFT has a time complexity of $O(N^2)$, making it inefficient for long sequences. The Fast Fourier Transform (FFT) optimizes this to $O(N \log N)$ using a divide-and-conquer approach, also applicable to Inverse FFT (IFFT).

In IVR, FFT can extract frequency features in the time dimension, effectively capturing both short-term and long-term dependencies in user behavior, thereby enhancing the performance of recommendation models.

4 Methodology

To address the issue of cognitive dissonance caused by overly diverse recommendations, we propose a novel model named CDIVR (Cognitive Dissonance-aware

Interactive Video Recommendation). Fig. 1 illustrates the overall framework of CDIVR model. By introducing the CRM and DST, CDIVR model can effectively capture the cognitive state changes during the interactive process and provides accurate recommendation strategies. In addition, CDIVR model trains the interaction strategy based on the offline RL framework and evaluates the effectiveness of the model in the simulated video recommendation environment. In this section, we introduce the three key modules of the CDIVR model.

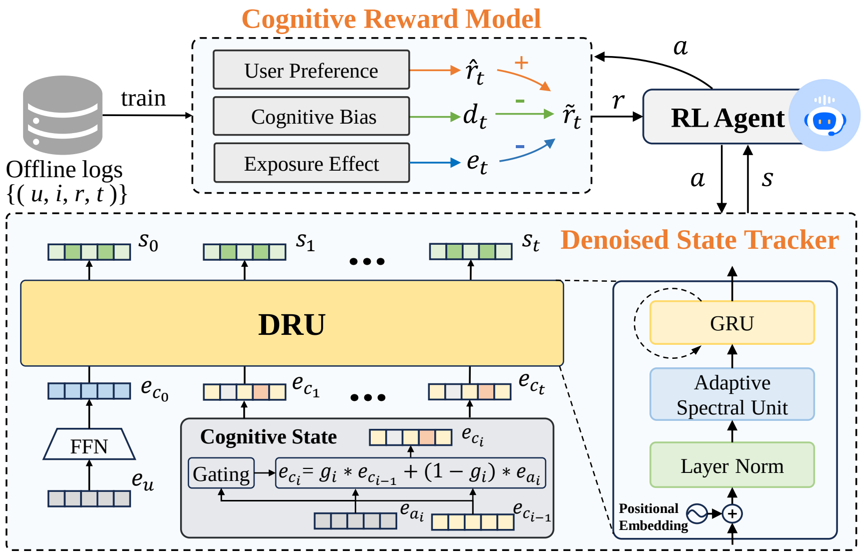


Fig. 1. Overall framework of our proposed CDIVR model.

4.1 Cognitive Reward Model

To provide the comprehensive estimation of user satisfaction, we propose a cognitive reward model (CRM), which is trained on historical data to decouple and analyze the cognitive effects during the interaction process. The CRM can comprehensively estimate user satisfaction from three aspects: user preference, overexposure effect, and cognitive bias. In the RL planning stage, CRM can provide counterfactual reward to balance the influence of overexposure effect and cognitive dissonance, thereby obtaining the effective strategy to alleviate cognitive dissonance.

User Preference Estimation Intrinsic preference refers to the degree of interest of user u in the t-th video i_t when not affected by other factors influence during the interaction process, which can be expressed as $\hat{r}_t = f_\theta(u, i_t)$. In this paper, the DeepFM model [14] is used for preference estimation.

Cognitive Bias Estimation The cognitive dissonance highlights the inconsistency between users' actual experiences and their expectations. To quantify the effect of cognitive dissonance, we estimate cognitive bias d_t before and after watching the t-th video. Specifically, we consider two factors related to cognitive bias: (1) the decline in user preference after interaction and (2) the inconsistency in categories among adjacent recommended videos. Substantial discrepancies in user preference and video categories can reflect the psychological discomfort arising from unmet user expectations, inducing cognitive dissonance. Therefore, cognitive bias can be represented as:

$$d_t = \frac{\alpha_u \times \min(0, (\hat{r}_t - \hat{r}_{t-1}))}{1 + \exp(-\operatorname{dist}(i_t, i_{t-1}))}, \tag{3}$$

where α_u represents the cognitive degree of the user. The term $\min(0, (\hat{r}_t - \hat{r}_{t-1}))$ captures the decline in user preference and $\operatorname{dist}(,)$ represents the distance between two videos, aiming to indicate both the immediate decline in preference and the inconsistency in video categories.

Overexposure Effect Estimation The overexposure effect reflects the fatigue or even aversion that may arise when users are repeatedly exposed to similar videos over a period of time, thereby affecting their overall satisfaction. To accurately measure this effect, we refer to the method described in [12], which employs a weighted calculation based on the similarity of historical recommendation data. The overexposure effect of user u on video i at time t can be expressed as:

$$e_t := e_t(u, i) := \alpha_u \beta_i \sum_{(u, i_l, t_l) \in \mathcal{S}_u, t_l < t} \exp\left(-\frac{t - t_l}{\tau} \times \operatorname{dist}(i, i_l)\right), \tag{4}$$

where β_i represents the unendurableness of the video, $\mathcal{S}_u$ is the complete historical interaction sequence for user u, and τ is a temperature. $t - t_l$ represents the temporal distance from the current time t. Videos that are closer in time should be assigned a higher overexposure weight.

User Satisfaction Estimation To more accurately capture users' actual satisfaction with recommended video, it is essential to consider the combined effects of various factors, including users' intrinsic preferences. However, user preferences alone are insufficient to fully reflect the actual user experience, as cognitive effects also play a crucial role. Therefore, a comprehensive assessment of user satisfaction should incorporate user preference, overexposure effect, and cognitive bias to capture the degree of cognitive effects during the interaction process. Additionally, the counterfactual mechanism is introduced to adjust the degree of cognitive effects during the planning stage, enabling the model training to adopt corresponding RL strategies. Based on the above considerations, the cognitive-aware user satisfaction, denoted as $\tilde{r_t}$, can be expressed as:

$$\tilde{r_t} = \frac{\hat{r}_t + \lambda_1 \cdot d_t}{1 + \lambda_2 \cdot e_t}, \tag{5}$$

where λ_1 and λ_2 are the adjustment coefficients for cognitive dissonance and overexposure effect, respectively.

4.2 Denoised State Tracker

RL-based recommendation strategies rely on the current state to make decisions. To better characterize user state transitions under the influence of cognitive effects, we define the representation of user's cognitive state $\mathbf{e_c} \in \mathbb{R}^C$. Initially, we adopt a feed-forward neural network (FFN) [13] to map the user representation vector $\mathbf{e_u}$ with dimension D into the cognitive state space, i.e., $\mathbf{e_u} \in \mathbb{R}^D \rightarrow \mathbf{e_{c_0}} \in \mathbb{R}^C$. By employing a gating mechanism, we extract the weights of the previous cognitive state representation $\mathbf{e_{c_{i-1}}}$ and action representation $\mathbf{e_{a_i}}$ relevant to the current moment i. Subsequently, the momentum update mechanism is adopted to represent the user's cognitive state, which can be formulated as:

$$\mathbf{e_{c_i}} = g_i \cdot \mathbf{e_{c_{i-1}}} + (1 - g_i) \cdot \mathbf{e_{a_i}}, \tag{6}$$

where $\mathbf{e_{a_i}}$ denotes the representation of the i-th recommended video, and the gating value g_i is defined as follows:

$$g_i = \sigma(\mathrm{Concat}(\mathrm{Proj}_1(\mathbf{e_{c_i}}), \mathrm{Proj}_2(\mathbf{e_{a_i}}))), \tag{7}$$

where σ denotes the sigmoid function, $\mathrm{Concat}(,)$ represents the vector concatenation operation. Both $\mathrm{Proj}_1(.)$ and $\mathrm{Proj}_2(.)$ indicate the FFN operation.

Additionally, high-frequency noise in recommendation data reflects random fluctuations in user behavior and the system, deviating from trends and reducing interpretability. The noise affects the model's ability to perceive transitions in users' cognitive states, impacting recommendation accuracy. Inspired by [8], we propose a denoised recurrent unit (DRU), which can accurately capture the changes in user states during the interaction process through adaptive denoising. After positional encoding and layer normalization, the cognitive state undergoes adaptive spectral unit (ASU) processing for a denoised hidden state, followed by a GRU model [6] to capture temporal dependencies.

Adaptive Spectral Unit To reduce noise impact, we introduce an adaptive spectral unit (ASU) [8] that adjusts the threshold based on spectrum characteristics. This mechanism allows for better distinction between true signals and high-frequency noise.

Fast Fourier Transform Given the sequence feature $S \in \mathbb{C}^{C \times L}$, we apply FFT to obtain the frequency domain representation F:

$$F = \mathcal{F}[S] \in \mathbb{C}^{C \times L'}, \tag{8}$$

where $\mathcal{F}[.]$ denotes the 1D FFT operation, and L' is the sequence length in the frequency domain.

Adaptive Denoising The dominant frequency is identified by the power spectral density $P = |F|^2$. A learnable parameter θ is used to filter out high-frequency noise during training. The adaptive denoising process is expressed as:

$$F' = F \odot (P < \theta), \tag{9}$$

where $\odot$ denotes element-wise multiplication, and the binary mask $(P < \theta)$ is utilized to extract frequency features with density below the threshold.

Frequency Domain Feature Integration Two learnable filters handle the original frequency feature F and the denoised feature F'. The processed features are then integrated to obtain a more comprehensive representation of frequency domain. The integrated feature F_I is expressed as:

$$F_I = W_G \odot F + W_L \odot F', \tag{10}$$

where W_G and W_L represent the globally and locally trainable filters, respectively.

Inverse Fourier Transform The integrated frequency feature is transformed back to the time domain using the inverse FFT:

$$F_T = \mathcal{F}^{-1}(F_I), \tag{11}$$

where F_T is the input to the GRU model.

4.3 Recommendation Strategy Based on Reinforcement Learning

To optimize the policy learning process in IVR, we adopt the proximal policy optimization (PPO) algorithm [22]. PPO is a policy gradient-based algorithm designed to enhance training stability by limiting the extent of policy changes during each update. In this paper, the optimization objective of PPO is to maximize the cumulative user satisfaction. Given the current policy π_θ, PPO updates the policy parameters θ by maximizing the following objective function $J(\theta)$:

$$J(\theta) = E_t \left[\min \left(r_t(\theta)\hat{A}_t, \operatorname{clip}\left(r_t(\theta), 1-\epsilon, 1+\epsilon\right)\hat{A}_t \right) \right], \tag{12}$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$ denotes the ratio of the action probabilities between the new and old policy; $\hat{A}_t$ represents the estimated advantage function for the state-action pair (s_t, a_t); ϵ is a hyperparameter used to control the step size of the policy update. The objective function includes a clipping term $\operatorname{clip}(\cdot)$, which ensures updates are not overly aggressive and thus helps to prevent significant performance fluctuations.

5 Experiments

In this section, we firstly introduce the experimental setup. Then, we conduct experiments on two datasets and compare the results with state-of-the-art baseline models to demonstrate the superiority of the CDIVR model. Finally, through hyperparameter analysis and case study, we validate that the CDIVR model can effectively alleviate cognitive dissonance, ultimately enhancing user satisfaction.

5.1 Experimental Setup

Dataset We adopt two real-world video recommendation datasets, KuaiRec and KuaiRand, to construct recommendation environments for evaluating the effectiveness of CDIVR model. These datasets encompass comprehensive features of users and videos, alongside large-scale interaction logs from a short-video platform, with their statistical details presented in Table 1.

Table 1. Statistics of the two evaluation datasets.

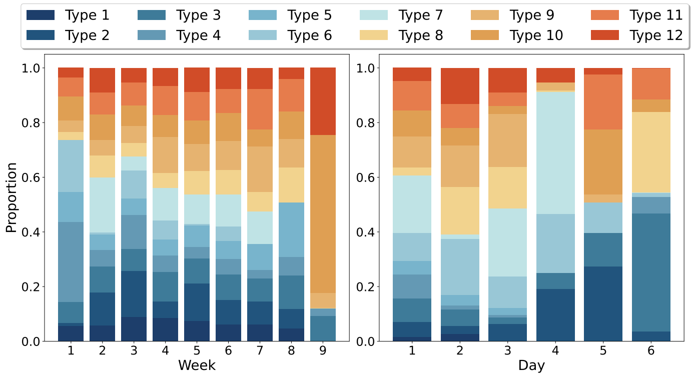
Dataset	Usage	Users	Items	Interactions	Density
KuaiRec	Train	7,176	10,728	12,530,806	16.28%
	Test	1,411	3,327	4,676,570	99.62%
KuaiRand	Train	27,285	7551	1,436,609	6.97%
	Test	27,285	7,583	1,186,059	5.73%

KuaiRec [10] is a high-density recommendation dataset that contains a fully observed user-item interaction matrix, which records the interaction information of users with all videos. In this paper, the normalized watching ratio is used to measure user satisfaction and provide online rewards.

KuaiRand [11] is an unbiased recommendation dataset. It collects user behavior data within two weeks by randomly inserting exposed videos, which can reflect users' unbiased interests. In this paper, "is_click" is used to measure user satisfaction and provide online rewards.

We conduct the visual analysis of the interaction data from the KuaiRec and KuaiRand datasets respectively, examining the proportion of watching ratio for a selected user over a period of time. The variations in watching ratio proportions across different types of videos are depicted in the Fig. 2. It can be observed that user interest fluctuates significantly throughout the interaction process, which can lead to cognitive dissonance.

Baselines To validate the efficiency of CDIVR model, we select two fundamental strategies and five RL methods for comparison. We employ the DeepFM model [14] as the backbone for the model-based methods.



(a) KuaiRec

(b) KuaiRand

Fig. 2. Distribution of watching ratio proportions in two datasets.

- **ϵ-greedy**: This method enhances the diversity of the DeepFM's recommendation outcomes by outputting random results with a probability of ϵ.
- **UCB** [2]: This method calculates an upper bound value for each potential recommendation action, which is based not only on the current estimated mean reward but also incorporates a term related to uncertainty.
- **CRR** [28]: The model-free RL method optimizes policies by critic-regularized regression, addressing the issue of poor performance in offline RL algorithms when learning from fixed datasets.
- **IPS** [25]: This method compensates for selection biases in observed data through weighting, enabling a more accurate assessment and learning of the optimal policy.
- **MOPO** [29]: This method is founded on conservative principles and imposes penalties on actions associated with high uncertainty within the model.
- **DORL** [9]: This method incorporates an entropy penalty to avoid singular and extreme recommendation actions.
- **CIRS** [12]: This method optimizes recommendation strategies under overexposure scenarios by modeling causal user model to provide counterfactual rewards.

Evaluation Metrics We evaluate our model in two video recommendation environments. The objective of IVR is to enhance users' long-term satisfaction. We refer to [9,12] and select three metrics:

- $\mathbf{R_{tra}}$: Cumulative satisfaction, which represents the accumulated reward value over the interaction trajectory, providing an overall reflection of the user's long-term experience with the platform.
- $\mathbf{R_{each}}$: Single-round satisfaction, which indicates the immediate experience effect of users.
- **Length**: Interaction length, which reflects the activity of users and the attractiveness of the platform.

Parameters Settings All model parameters are optimized using the Adam optimizer with an initial learning rate of 0.001. For the model-based RL model, we use a DeepFM model with the same configuration as the backbone network for the reward model. λ_1 is 10, and λ_2 is tuned within the set $\{0.001, 0.1, 0.5, 1, 2, 5\}$. During the experiments, all policies are trained for 200 epochs and then evaluated across 100 non-repeating recommended interaction trajectories in the environment. The average values over the 100 interaction trajectories serve as the final evaluation results, and the maximum round for each trajectory is set to 30.

5.2 Experiment Comparison

Table 2 presents the performance of the CDIVR model and other baselines in two environments. For a more intuitive comparison, Fig. 3 illustrates the training processes of CDIVR and the top-4 baseline models in terms of performance. It can be obviously observed that the ϵ-greedy and UCB methods only yield short interaction sequences and fail to meet users' long-term satisfaction because they are based on simple heuristic strategies.

Table 2. Overall performance in two environments. (Bold: Best; underline: runner-up)

	KuaiRec			KuaiRand		
Methods	$\mathrm{R}_{\mathrm{tra}}$	$\mathrm{R}_{\mathrm{each}}$	Length	$\mathrm{R}_{\mathrm{tra}}$	$\mathrm{R}_{\mathrm{each}}$	Length
ε-greedy	3.61	0.85	4.22	1.65	0.37	4.43
UCB [2]	3.52	0.83	4.22	1.65	0.37	4.43
CRR [28]	4.16	0.90	4.65	1.48	0.23	6.56
IPS [25]	12.83	0.76	16.73	3.63	0.22	16.82
MOPO [29]	11.43	0.89	12.81	10.93	$\underline{0.44}$	25.00
DORL [9]	20.49	0.77	$\underline{26.71}$	$\underline{11.85}$	0.43	$\underline{27.61}$
CIRS [12]	$\underline{22.98}$	**0.96**	23.97	11.05	0.43	25.89
Ours	**27.08**	$\underline{0.94}$	**28.96**	**13.27**	**0.44**	**30.00**
% Improv.	17.84%	N/A	8.42%	11.98%	0.00%	8.66%

Compared to other model-based approaches (IPS, MOPO, DORL, CDIVR), the performance of CRR model is inferior, as it is extremely challenging for model-free RL method to capture cognitive effects and learn the effective recommendation strategies from limited offline data. Therefore, adopting model-based RL methods is more suitable for addressing IRS tasks.

For the model-based baseline models, IPS model suffers from high variance, making it difficult to recommend satisfactory videos to users. Especially in the Kuairand dataset, IPS model fails to achieve a high level of single-round reward after training. MOPO's conservative recommendation strategy limits the acquisition of longer interaction sequences. Moreover, MOPO is subject to the noise

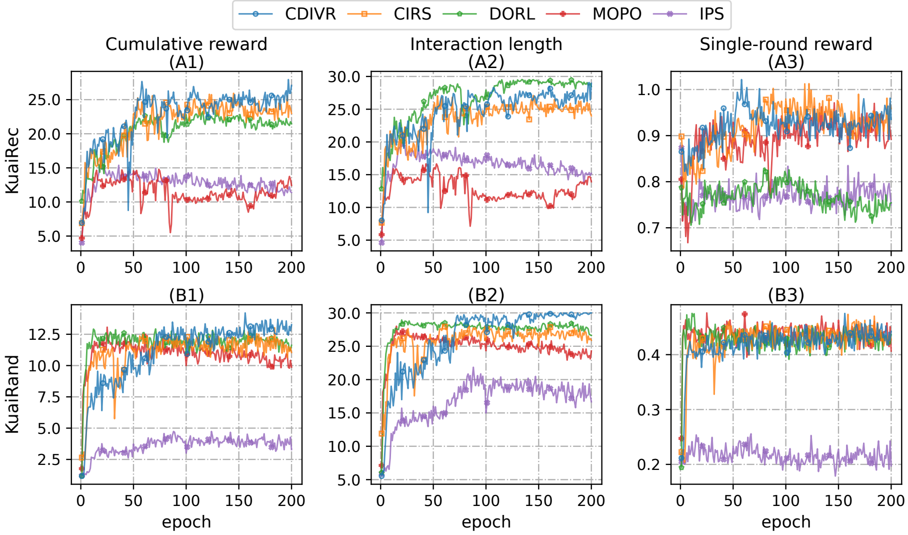


Fig. 3. Results of CDIVR and top-4 baseline models in two environments.

interference in the interaction data, resulting in an unstable training process in the KuaiRec dataset. In contrast, both the DORL and CIRS models, which are based on the idea of diverse recommendations, avoid overexposure to specific categories, thereby achieving better performance. However, these models fail to account for the impact of cognitive dissonance in the interaction process, which prevents them from achieving higher cumulative satisfaction.

As show in Table 2, our proposed CDIVR model outperforms the baseline models by 17.84% and 11.98% in cumulative user satisfaction, respectively. Fig. 3 illustrates the superior performance and stability of the CDIVR model during training. Moreover, CDIVR model achieves a better balance between high single-round satisfaction and long interaction rounds, thereby providing users with greater cumulative satisfaction. This advantage of CDIVR model is attributed to the perception of cognitive dissonance and the provision of counterfactual rewards, enabling the learning of appropriate strategies. Additionally, CDIVR model introduces the DST module to reduce the noise interference and ensure the stability of policy learning.

5.3 Ablation Study

We conduct the ablation study to demonstrate the contributions of key component in CDIVR model to improving cumulative reward $\mathrm{R}_{\mathrm{tra}}$, with the results depicted in Table 3. To investigate the effectiveness of the CRM module, we design the CDIVR model variant by removing the cognitive bias estimation module, denoted as CDIVR w/o C. To examine the validity of the DST module, we replace it with a simple GRU model, denoted as CDIVR w/o D.

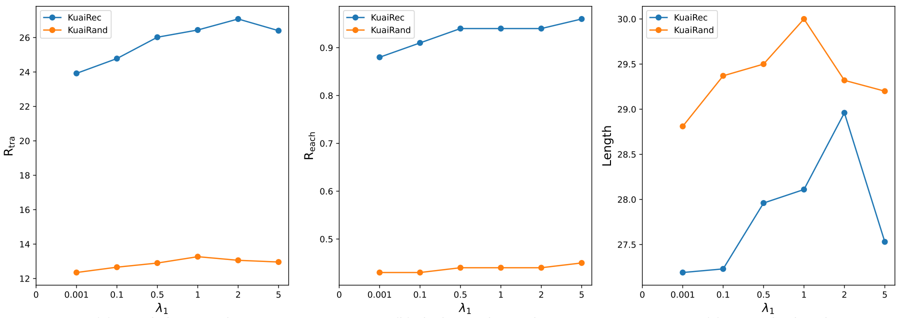
Table 3. Ablation study of the CDIVR model

Dataset-Metric	w/o C	w/o D	CDIVR
KuaiRec-R_{tra}	23.62 (-12.78%)	24.97 (-7.80%)	27.08
KuaiRand-R_{tra}	12.08 (-9.00%)	12.38 (-6.71%)	13.27

As shown in Table 3, the cognitive bias estimation module shows a higher contribution, indicating that accounting for cognitive dissonance can more comprehensively reflect changes in user satisfaction during interactions. Additionally, the CDIVR w/o D model demonstrates a noticeable decline in cumulative satisfaction across both datasets. This suggests that the DST can effectively eliminate the noise impact on the interaction sequence and improve the performance through cognitive state representation and adaptive denoising, thereby enhancing performance.

5.4 Analysis of Alleviating Cognitive Dissonance

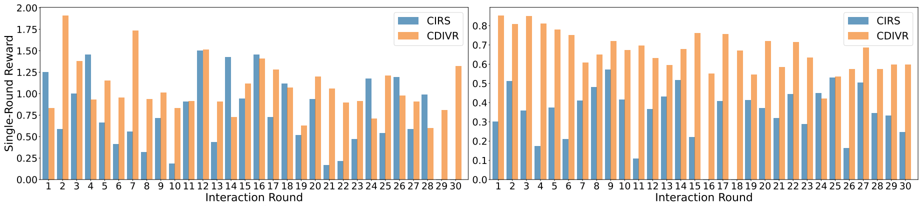
Impacts of Coefficient for Cognitive Dissonance The CDIVR model can adjust the weight of λ_1 to provide counterfactual rewards, thereby learning a recommendation strategy that alleviates cognitive dissonance. Fig. 4 illustrates how this counterfactual mechanism improves user satisfaction: as λ_1 moderately increases, the cumulative reward also rises, demonstrating that reducing cognitive dissonance can significantly enhance user experience. However, an excessive increase in λ_1 will lead to homogenized recommendations and exacerbate the overexposure effect, thereby degrading the long-term user experience. As shown in Fig. 4(a), the CDIVR model achieves the highest cumulative satisfaction when $\lambda_1 = 2$ in KuaiRec and $\lambda_1 = 1$ in KuaiRand, effectively balancing the trade-off between cognitive dissonance and the overexposure effect.



(a) Cumulative reward (b) Single-round reward (c) Interaction length

Fig. 4. Performance with different λ_1.

Case Study To verify the effectiveness of the CDIVR model in alleviating cognitive dissonance, we conduct a case analysis and compare it with the CIRS model, which demonstrates superior performance. The procedure is as follows: Initially, we select a user from each of the two datasets, and the CDIVR and CIRS models are employed to iteratively generate 30 rounds of recommendation results for each selected user. Meanwhile, user satisfaction with the model's recommendations in each round is tracked based on real feedback from the datasets. Fig. 5 records the changes in user satisfaction per interaction round. The CIRS model, failing to account for the potential impact of cognitive dissonance, exhibits significant fluctuations in user satisfaction during interaction process, leading to a lower cumulative satisfaction. On the other hand, the CDIVR model can stably recommend videos that align with user interests by introducing CRM and DST modules. The stable recommendation strategies can effectively alleviate cognitive dissonance and provide users with a better long-term experience.



(a) KuaiRec

(b) KuaiRand

Fig. 5. Results of case study.

6 Conclusion

In this paper, we propose a cognitive dissonance-aware interactive video recommendation (CDIVR) model to address the cognitive dissonance issue in IVR. To comprehensively estimate user satisfaction, we propose a CRM that depicts the satisfaction from three aspects: user preference, overexposure effect and cognitive bias. In addition, the CDIVR model employs a DST module, which can track the changes in user states and mitigate the impact of noise during the interaction process, enabling the model to stably learn the strategies for alleviating cognitive dissonance. Extensive experiments on two real-world datasets demonstrate that the CDIVR model can effectively alleviate cognitive dissonance and significantly outperform baseline models in cumulative user satisfaction. In future work, we will explore the impacts of other cognitive effects (such as anchoring effect) to enhance the user experience in IVR.

Acknowledgments. This research work was partly supported by Xinjiang Science and Technology Program (No.2022D01B185).

References

1. Afsar, M.M., Crump, T., Far, B.: Reinforcement learning based recommender systems: A survey. ACM Computing Surveys **55**(7), 1–38 (2022)
2. Auer, P.: Finite-time analysis of the multiarmed bandit problem (2002)
3. Chen, M., Xu, C., Gatto, V., Jain, D., Kumar, A., Chi, E.: Off-policy actor-critic for recommender systems. In: Proceedings of the 16th ACM Conference on Recommender Systems. pp. 338–349 (2022)
4. Chen, N., Zhang, F., Sakai, T.: Constructing better evaluation metrics by incorporating the anchoring effect into the user model. In: Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval. pp. 2709–2714 (2022)
5. Chen, X., Wang, S., McAuley, J., Jannach, D., Yao, L.: On the opportunities and challenges of offline reinforcement learning for recommender systems. ACM Transactions on Information Systems **42**(6), 1–26 (2024)
6. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. In: NIPS 2014 Workshop on Deep Learning, December 2014 (2014)
7. Donkers, T., Ziegler, J.: The dual echo chamber: Modeling social media polarization for interventional recommending. In: Proceedings of the 15th ACM conference on recommender systems. pp. 12–22 (2021)
8. Eldele, E., Ragab, M., Chen, Z., Wu, M., Li, X.: Tslanet: Rethinking transformers for time series representation learning. In: International Conference on Machine Learning. pp. 1–20. PMLR (2024)
9. Gao, C., Huang, K., Chen, J., Zhang, Y., Li, B., Jiang, P., Wang, S., Zhang, Z., He, X.: Alleviating matthew effect of offline reinforcement learning in interactive recommendation. In: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 238–248 (2023)
10. Gao, C., Li, S., Lei, W., Chen, J., Li, B., Jiang, P., He, X., Mao, J., Chua, T.S.: Kuairec: A fully-observed dataset and insights for evaluating recommender systems. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management. pp. 540–550 (2022)
11. Gao, C., Li, S., Zhang, Y., Chen, J., Li, B., Lei, W., Jiang, P., He, X.: Kuairand: an unbiased sequential recommendation dataset with randomly exposed videos. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management. pp. 3953–3957 (2022)
12. Gao, C., Wang, S., Li, S., Chen, J., He, X., Lei, W., Li, B., Zhang, Y., Jiang, P.: Cirs: Bursting filter bubbles by counterfactual interactive recommender system. ACM Transactions on Information Systems **42**(1), 1–27 (2023)
13. George, B., Michael, G.: Feed-forward neural networks. Ieee Potentials **13**(4), 27–31 (1994)
14. Guo, H., TANG, R., Ye, Y., Li, Z., He, X.: Deepfm: A factorization-machine based neural network for ctr prediction. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17. pp. 1725–1731 (2017)
15. Lex, E., Kowald, D., Seitlinger, P., Tran, T.N.T., Felfernig, A., Schedl, M., et al.: Psychology-informed recommender systems. Foundations and trends® in information retrieval **15**(2), 134–242 (2021)
16. Liu, S., Cai, Q., Sun, B., Wang, Y., Jiang, J., Zheng, D., Jiang, P., Gai, K., Zhao, X., Zhang, Y.: Exploration and regularization of the latent action space in recommendation. In: Proceedings of the ACM Web Conference 2023. pp. 833–844 (2023)

17. Liu, X., Yu, T., Xie, K., Wu, J., Li, S.: Interact with the explanations: Causal debiased explainable recommendation system. In: Proceedings of the 17th ACM International Conference on Web Search and Data Mining. pp. 472–481 (2024)
18. Liu, Y., Xiao, Y., Wu, Q., Miao, C., Zhang, J., Zhao, B., Tang, H.: Diversified interactive recommendation with implicit feedback. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 4932–4939 (2020)
19. Lv, X., Li, J., Wang, Q.: The dark side of recommendation algorithms in chinese mass short video apps: Effect of perceived over-recommendation on users' cognitive dissonance and discontinuance intention. International Journal of Human–Computer Interaction pp. 1–15 (2024)
20. Masrour, F., Wilson, T., Yan, H., Tan, P.N., Esfahanian, A.: Bursting the filter bubble: Fairness-aware network link prediction. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 841–848 (2020)
21. Nie, W., Wen, X., Liu, J., Chen, J., Wu, J., Jin, G., Lu, J., Liu, A.A.: Knowledge-enhanced causal reinforcement learning model for interactive recommendation. IEEE Transactions on Multimedia **26**, 1129–1142 (2023)
22. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347 (2017)
23. Schwind, C., Buder, J., Cress, U., Hesse, F.W.: Preference-inconsistent recommendations: An effective approach for reducing confirmation bias and stimulating divergent thinking? Computers & Education **58**(2), 787–796 (2012)
24. Surendren, D., Bhuvaneswari, V.: A framework for analysis of purchase dissonance in recommender system using association rule mining. In: 2014 International Conference on Intelligent Computing Applications. pp. 153–157. IEEE (2014)
25. Swaminathan, A., Joachims, T.: Counterfactual risk minimization: Learning from logged bandit feedback. In: International Conference on Machine Learning. pp. 814–823. PMLR (2015)
26. Wang, W., Feng, F., Nie, L., Chua, T.S.: User-controllable recommendation against filter bubbles. In: Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval. pp. 1251–1261 (2022)
27. Wang, Y., Ma, W., Zhu, Y., Wang, C., Wang, Z., Tang, F., Yu, J.: Guiding graph learning with denoised modality for multi-modal recommendation. In: International Conference on Database Systems for Advanced Applications. pp. 220–235. Springer (2024)
28. Wang, Z., Novikov, A., Zolna, K., Merel, J.S., Springenberg, J.T., Reed, S.E., Shahriari, B., Siegel, N., Gulcehre, C., Heess, N., et al.: Critic regularized regression. Advances in Neural Information Processing Systems **33**, 7768–7778 (2020)
29. Yuan, F., Karatzoglou, A., Arapakis, I., Jose, J.M., He, X.: A simple convolutional generative network for next item recommendation. In: Proceedings of the twelfth ACM international conference on web search and data mining. pp. 582–590 (2019)
30. Zhao, H., Zhang, L., Xu, J., Cai, G., Dong, Z., Wen, J.R.: Uncovering user interest from biased and noised watch time in video recommendation. In: Proceedings of the 17th ACM Conference on Recommender Systems. pp. 528–539 (2023)