Clicking, Fast and Slow: Towards Intuitive and Analytical Behaviors Modeling for Recommender Systems

Youlin Wu, Haoxi Zhan, Yuanyuan Sun, Haohao Zhu, Bo Xu, Liang Yang, Hongfei Lin✉
School of Computer Science and Technology,
Dalian University of Technology, Dalian, China
{wuyoulin, zhanhaoxi, zhuhh}@mail.dlut.edu.cn
{syuan, xubo, liang, hflin}@dlut.edu.cn

Abstract

Recommender systems personalize content delivery based on users' interaction history. However, not all clicks result from deliberate decisions—many arise from intuitive reactions. Inspired by the dual process theory, we argue that intuitive clicks are primarily driven by System 1, reacting to superficial cues, while analytical clicks involve deeper processing by System 2, considering the semantic meaning and long-term preference. However, existing models overlook these cognitive mechanisms. To address this, we propose DualRec, a novel recommendation method that models both intuitive and analytical behaviors. DualRec encodes items using language models, leveraging shallow layers for superficial understanding (System 1) and deep layers for semantic comprehension (System 2). It employs Transformer-based encoders with two attention mechanisms to capture intuitive "fast" and analytical "slow" click patterns. A learnable fusion layer balances these behaviors. Extensive experiments demonstrate that DualRec outperforms existing methods and highlights the importance of integrating both cognitive processes in recommendations.

Keywords: dual system theory; recommender systems; user preference modeling; language models

Introduction

Recommender systems (RS), as an effective tool to mitigate information explosion, have been widely adopted in many online applications, such as news sites (F. Wu et al., 2020), e-commerce (McAuley, Targett, Shi, & Van Den Hengel, 2015) and social media platforms (Guy, Zwerdling, Ronen, Carmel, & Uziel, 2010) to improve user experience. In general, RS mainly focuses on predicting the next item a user is likely to interact with based on their historical interaction sequence (Boka, Niu, & Neupane, 2024). Traditional RS models mainly rely on item IDs to represent items and users, leveraging techniques like Markov Chains (R. He & McAuley, 2016), matrix factorization (Rendle, 2012; Guo, Tang, Ye, Li, & He, 2017), and more recently, powerful self-attention models (Sun et al., 2019; Kang & McAuley, 2018) to learn collaborative patterns. The advent of large-scale pre-trained language models (PLMs) has propelled growing interest in leveraging text information to learn transferable item representations, aiming to facilitate the user preference modeling, forming the content-based RS paradigm (Hou et al., 2022; Li et al., 2023; Yuan et al., 2023).

However, current content-based RS models mainly focus on capturing user interaction patterns, they often neglect the cognitive processes underlying user behaviors, such as *reading* and *clicking*. Studies of online reading behavior have shown that people spend more time browsing and scanning, rather than in-depth reading (Z. Liu, 2005). This trend results in a shallower and more fragmented understanding of overly abundant online contents, characterized by rapid shifts of attention and reduced deliberation (Loh & Kanai, 2016; Carr, 2020). This cognitive behavior significantly influences how users interact with recommender systems in online applications. For instance, consider a news website where recommended news items are presented after each article (Figure 1, left part). One user might quickly scan the titles, noticing catchy keywords like "sodas" and "toxic", and impulsively click the first recommendation entry based on this superficial impression. Another user, however, might carefully read and analyze each title, discerning that the first article is sensationalized and opting instead for the second one, which aligns with their existing interest in politics (*i.e.* Trump → Biden). This illustrates how varying levels of cognitive engagement can lead to different user behaviors when interacting with an RS. *But why do these differences arise?* In this work, we believe that this observation aligns with Dual Process Theory (DPT)[1] (J. S. B. T. Evans, 2008), which posits that human cognition operates through two distinct systems: System 1, which is fast, automatic, and relies on heuristics and intuitions, and System 2, which is slow, deliberate, and analytical (Stanovich & West, 2002; J. S. B. Evans & Frankish, 2009;

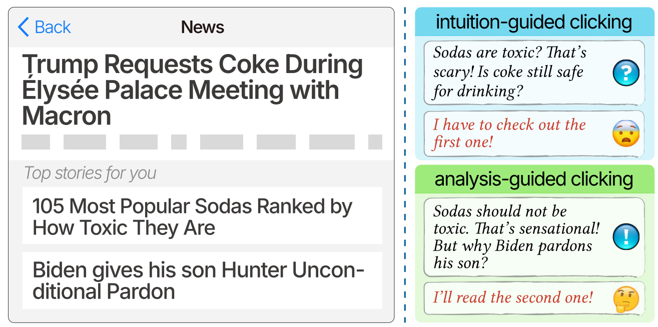

Figure 1: An example of intuition-guided versus analysis-guided click decisions in response to different recommended headlines during online browsing.

[1]The DPT is a family of theories originated from multiple fields of research. In this work, we mainly adopt the System 1–System 2 version from behavioral economics proposed by Kahneman (2011).

J. S. B. Evans, 2011). The first user's rapid, keyword-driven click exemplifies System 1 thinking, while the second user's careful analysis and considered click decision reflects System 2 processing. Unfortunately, current RS models fail to notice these dual cognitive processes, leading to suboptimal performance in personalized recommendations.

Based on these observations, we propose a novel recommendation method, called DualRec, that mimics these cognitive processes to model both intuitive and analytical behaviors. More specifically, we break down users' online behavior into two parts: *content read* and *click decision make*. DualRec takes account of the cognitive duality from these two aspects: item comprehension (read) and user behavior (click) modeling. Considering that different layers in PLMs capture different levels of semantic information (N. F. Liu, 2019; Jawahar, Sagot, & Seddah, 2019), DualRec utilizes both shallow and deep layer outputs for superficial (System 1) and thorough (System 2) semantic comprehensions, respectively. As for click behavior modeling, DualRec adopts a Transformer-based dual-channel architecture to capture both intuitive and analytical behavior patterns. In the intuition channel, DualRec further adopts additive attention to probe for "fast" heuristic click patterns; as for the analysis channel, DualRec utilizes average attention to capture "slow" long-term preferences. Furthermore, a learnable fusion layer is introduced to dynamically adjust the weight of these two channels, allowing the model to flexibly adapt to different users.

In summary, our contributions are as follows:
(1) We scrutinize the cognitive processes underlying users' interactions with an RS and find out they align well with the dual process theory. A novel dual channel method, DualRec, is proposed to model both intuitive and analytical behaviors.
(2) DualRec employs both shallow and deep layer outputs of the PLM to encode items at different comprehension levels; then two Transformer-based preference channels are adopted to capture intuitive and analytical behaviors, respectively.
(3) Extensive experiments on two real-world datasets show that our DualRec has achieved state-of-the-art performance, and further confirm the crucial role of both intuitive and analytical behaviors in personalized recommendations.

Related Works

Dual Process and Online Behavior

Dual Process Theory, a prominent framework in cognitive psychology, has been increasingly utilized in human-computer interaction (HCI) research to understand user behavior and habit formation (Pinder, Vermeulen, Cowan, & Beale, 2018). An ideal situation is that online behavior is driven by the pursuit of information that fulfills needs and interests (Savolainen, 1995). However, users may not always act inline with this ideal in real-life online experience (Lyngs et al., 2019; Lyngs, Binns, Van Kleek, & Shadbolt, 2018; Milli, Belli, & Hardt, 2021). Kleinberg, Mullainathan, and Raghavan (2024) leverage the DPT to explain this discrepancy, suggesting that users may not always conduct deliberate, analytical thinking (System 2) when interacting with online systems. Instead, they may perform quick, intuitive clicks (System 1) driven by immediate gratification, even if these choices don't align with their long-term needs or interests. This is particularly relevant for recommender systems, where Agarwal, Usunier, Lazaric, and Nickel (2024) propose disentangling clicks into impulsive-driven (System 1) and utility-driven (System 2) categories. However, these works only focus on the two behavioral patterns of clicking and fail to pay enough attention of the duality in item comprehension.

Content-based Recommender Systems

Recommender systems, serving as crucial human-computer interfaces, are extensively utilized across diverse online platforms to mitigate information overload (Åberg & Shahmehri, 2000). Traditional RS primarily rely on numerical IDs to represent items and users, leveraging techniques like collaborative filtering (X. He et al., 2017), matrix factorization (Guo et al., 2017), recurrent neural networks (RNN) (Hidasi, Karatzoglou, Baltrunas, & Tikk, 2016), convolutional neural networks (CNN) (Tang & Wang, 2018), and attention mechanisms (Sun et al., 2019) to identify user-item interaction patterns. However, ID lacks semantic information, hindering a deeper understanding of the cognitive processes underlying user behaviors and potentially limiting recommendation accuracy. Thus, recent research has focused on utilizing item content for representation learning (Hou et al., 2022). Two main approaches have emerged: (1) utilizing word embeddings like GloVe (Pennington, Socher, & Manning, 2014) or PLMs like BERT (Kenton & Toutanova, 2019) to encode item descriptions (C. Wu et al., 2019; Hou et al., 2022); and (2) incorporating external knowledge (Qiu, Hu, & Wu, 2022) for enhanced item representations (Yuan et al., 2023). However, they only focus on generating one single "best" representation for each item, which may not adequately capture the multifaceted nature of item semantics, encompassing both superficial cues and deeper meaning.

Problem Formulation

Let $\mathcal{U} = \{u_1, u_2, \ldots, u_{|\mathcal{U}|}\}$ and $I = \{i_1, i_2, \ldots, i_{|I|}\}$ denote the sets of users and items, respectively. The recommender system can be generalized as a click prediction task, where the goal is to predict the next item i_t a user u is likely to interact with based on their historical clicking sequence $s_u = [i_1, i_2, \ldots, i_{t-1}]$. Each item $i \in I$ is associated with textual content (e.g., title, description), represented as $i = [w_1, w_2, \ldots, w_c]$, where w_k represents a word (token) and c is the maximum content length. We employ an item encoder to derive two item embeddings, $\boldsymbol{i}^1$ and $\boldsymbol{i}^2$, capturing superficial and deep semantic information, respectively. Similarly, a user encoder aggregates clicked items into two user embeddings, $\boldsymbol{s}^1$ and $\boldsymbol{s}^2$, representing intuitive and analytical user preferences. Our target can be formulated into modeling the probability $p(i_t|\boldsymbol{s}^1, \boldsymbol{s}^2)$ of the next item i_t being clicked by user u given their historical interactions s_u.

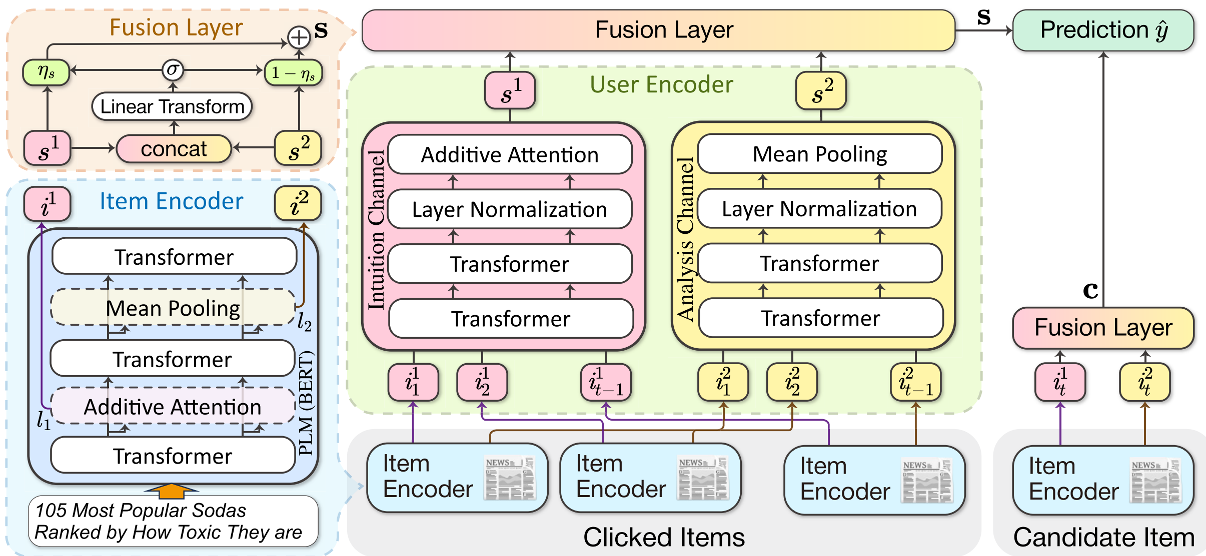

Figure 2: The architecture of the proposed DualRec model.

Methodology

In this section, we provide detailed descriptions of the proposed DualRec. The overall model architecture is shown in Figure 2.

Item Encoder

The item encoder is designed to capture the semantic information of items. As aforementioned, we argue that users may react to catchy words and superficial cues before engaging in deeper comprehension. While traditional methods like static word vectors (Pennington et al., 2014) are trained to focus on deep semantics, they are less effective at capturing the surface-level features. Inspired by Jawahar et al. (2019), we leverage the multi-layered nature of PLMs to represent both levels of information.

Superficial Cues Encoding We model superficial cues–the immediate surface-level features that trigger initial impressions–using the lower (shallow) layers of a PLM, which are known to capture phrase-level information, effectively representing the fragmented understanding of words and phrases that contribute to heuristic judgments (Jawahar et al., 2019).

More specifically, given an item $i = [w_1, w_2, \ldots, w_c]$, we feed the tokens into the PLM and extract the output of the l_1-th layer, denoted as $\boldsymbol{h}_i^1 = [\boldsymbol{h}_{i,1}^1, \boldsymbol{h}_{i,2}^1, \ldots, \boldsymbol{h}_{i,c}^1]$, where $\boldsymbol{h}_{i,k}^1 \in \mathbb{R}^{emb}$ is the representation of the k-th token. To emphasize the salient words in forming the first impressions, we further adopt the additive attention (Bahdanau, 2014) to aggregate these token representations into an item embedding $\boldsymbol{i}^1$. The attention weight α_k for the k-th token is calculated as:

$$a_k = \mathbf{W}^{(2)} \tanh(\mathbf{W}^{(1)} \boldsymbol{h}_{i,k}^1 + \mathbf{b}^{(1)}) + b^{(2)},$$
$$\alpha_k = \frac{\exp(a_k)}{\sum_{j=1}^{c} \exp(a_j)}, \tag{1}$$

where $\mathbf{W}^{(1)}, \mathbf{b}^{(1)}, \mathbf{W}^{(2)}$ and $b^{(2)}$ are learnable parameters. We then build the item's superficial cues embedding $\boldsymbol{i}^1$ as:

$$\boldsymbol{i}^1 = \text{add. attention}(\boldsymbol{h}_i^1) = \sum_{j=1}^{c} \alpha_j \boldsymbol{h}_{i,j}^1. \tag{2}$$

Deep Semantic Encoding Analytical click is primarily driven by deep semantic understanding, which requires a more profound comprehension of the item's content. Jawahar et al. (2019) has also revealed that PLMs encode rich semantic features and long-distance word dependencies at the top (deep) layers. Similarly, we extract the l_2-th layer outputs of the PLM, denoted as $\boldsymbol{h}_i^2 = [\boldsymbol{h}_{i,1}^2, \boldsymbol{h}_{i,2}^2, \ldots, \boldsymbol{h}_{i,c}^2]$. Unlike the superficial cues encoding, to ensure a consistent understanding of all words, we utilize the mean pooling to aggregate these token representations into the deep semantic embedding $\boldsymbol{i}^2$:

$$\boldsymbol{i}^2 = \text{mean pooling}(\boldsymbol{h}_i^2) = \frac{1}{c} \sum_{j=1}^{c} \boldsymbol{h}_{i,j}^2. \tag{3}$$

Importantly, our item encoder produces two embeddings, $\boldsymbol{i}^1$ and $\boldsymbol{i}^2$, to represent one item. Unlike typical PLM usage in item encoding, we do not rely on the `[CLS]` token output, as it is designed for sentence classification and may not be suitable for superficial cues and deep semantics capturing.

User Behavior Encoder

The user behavior encoder is designed to discover clicking patterns. In DualRec, we propose a dual-channel user behavior encoder to model both intuitive and analytical behaviors.

Dual Channel User Encoder To this point, DualRec is able to disentangle users' cognition towards items in I into two levels of interpretations and represent them using two embeddings, $\boldsymbol{i}^1$ (superficial features) and $\boldsymbol{i}^2$ (deep semantic features). Mirroring this dual representation, we model user clicking behavior as arising from two distinct cognitive processes: intuitive (System 1) and analytical (System 2). We therefore employ a dual-channel architecture in our user encoder to capture these distinct aspects.

Given a sequence of clicked items s_u, we can encode each clicked item using aforementioned item encoder and obtain two item embedding sequences $\mathbf{S}^1 = [\boldsymbol{i}_1^1, \boldsymbol{i}_2^1, \ldots, \boldsymbol{i}_{t-1}^1]$ and $\mathbf{S}^2 = [\boldsymbol{i}_1^2, \boldsymbol{i}_2^2, \ldots, \boldsymbol{i}_{t-1}^2]$. For each channel, we first stack L Transformer layers $\mathrm{Trm}(\cdot)$, which are capable of capturing both long and short-term interaction patterns (Sun et al., 2019), on top of each other to encode the behavior sequence:

$$\begin{aligned} \mathbf{S}^{1\star} &= \underbrace{\mathrm{Trm}(\ldots \mathrm{Trm}}_{L\times}(\mathbf{S}^1)), \\ \mathbf{S}^{2\star} &= \underbrace{\mathrm{Trm}(\ldots \mathrm{Trm}}_{L\times}(\mathbf{S}^2)), \end{aligned} \tag{4}$$

furthermore, we employ residual connection and layer normalization to mitigate information loss during the encoding process,

$$\begin{aligned} \widetilde{\mathbf{S}^1} &= \mathrm{Layernorm}(\mathbf{S}^{1\star} + \mathbf{S}^1), \\ \widetilde{\mathbf{S}^2} &= \mathrm{Layernorm}(\mathbf{S}^{2\star} + \mathbf{S}^2). \end{aligned} \tag{5}$$

Finally, considering that System 1 behavior is contextualized (Stanovich & West, 2002; Margolis, 1987), in other words, click decisions are triggered by certain immediate cues during browsing (context), we adopt the additive attention to highlight these cues by aggregating the encoded sequences into an *intuitive* user embedding $\boldsymbol{s}^1$:

$$\boldsymbol{s}^1 = \text{add. attention}(\widetilde{\mathbf{S}^1}), \tag{6}$$

while System 2 behavior is less context-dependent and more reflective of long-term preferences, we directly utilize average attention (*aka.* mean pooling) layer to form an *analytical* user embedding $\boldsymbol{s}^2$ that reflects item-by-item slow consideration:

$$\boldsymbol{s}^2 = \text{mean pooling}(\widetilde{\mathbf{S}^2}). \tag{7}$$

Notably, additive attention, by highlighting specific items, reflects the impulsive nature of System 1-driven clicks, while average attention, by considering all items equally, mimics the deliberate evaluation characteristic of System 2. These distinct attention mechanisms, coupled with the underlying Transformer layers, effectively capture the diverse patterns of user click decisions.

Behavior Pattern Fusion In most cases, users do not conduct online behaviors solely under one cognitive process. In other words, these two cognitive processes may interplay and jointly influence user behavior (Frankish, 2010). To accommodate this, DualRec incorporates a learnable fusion mechanism that dynamically balances the influence of these two cognitive processes, allowing DualRec to flexibly adapt to real-world cases. More specifically, we first fuse $\boldsymbol{s}^1$ and $\boldsymbol{s}^2$ into one user preference embedding $\mathbf{s}$,

$$\begin{aligned} \mathbf{s} &= \eta_s \boldsymbol{s}^1 + (1 - \eta_s)\boldsymbol{s}^2, \\ \eta_s &= \sigma([\boldsymbol{s}^1; \boldsymbol{s}^2]\mathbf{W_a} + b_a), \end{aligned} \tag{8}$$

where $\mathbf{W_a}$ and b_a are parameters to learn; $[\,\cdot\,;\,\cdot\,]$ means concatenation while σ represents the Sigmoid function. Similarly, we then utilize the similar fusion technique to adjust the weight of the two item embeddings $\boldsymbol{i}_t^1$ and $\boldsymbol{i}_t^2$ to get the final candidate item embedding $\mathbf{c}$:

$$\begin{aligned} \mathbf{c} &= \eta_c \boldsymbol{i}_t^1 + (1 - \eta_c)\boldsymbol{i}_t^2, \\ \eta_c &= \sigma([\boldsymbol{i}_t^1; \boldsymbol{i}_t^2]\mathbf{W_a} + b_a). \end{aligned} \tag{9}$$

Model Training

To address the inherent data sparsity issue, where users interact with only a small subset of items in I, and accelerate the training process, we employ negative sampling during training. More specifically, for each impression log j, we first randomly choose k negative items (not clicked) $\{i_j^{1-}, i_j^{2-}, \ldots, i_j^{k-}\}$ and one positive item (clicked) i_j^+. We then utilize the widely adopted (Rendle, Krichene, Zhang, & Anderson, 2020; Li et al., 2022) dot product to calculate the click probability score $\hat{\mathbf{y}}_j = [\hat{y}_j^+, \hat{y}_j^{1-}, \hat{y}_j^{2-}, \ldots, \hat{y}_j^{k-}]$ using the dot product between the user embedding $\mathbf{s}$ and the candidate item embedding $\mathbf{c}$:

$$\hat{\mathbf{y}}_j = \mathrm{softmax}(\mathbf{s} \cdot \mathbf{c}). \tag{10}$$

We then optimize the NCE loss $\mathcal{L}_{\mathrm{NCE}}$ across all positive samples in the training set training set $\mathcal{S}$:

$$\mathcal{L}_{\mathrm{NCE}} = -\sum_{j=1}^{|\mathcal{S}|} \log \frac{\exp(\hat{y}_j^+)}{\exp(\hat{y}_j^+) + \sum_1^k \exp(\hat{y}_j^-)}. \tag{11}$$

Compared to standard cross-entropy, the NCE loss allows our DualRec to leverage additional information derived from negative feedback.

Experiment

In this section, we conduct a series of experiments to answer the following key Research Questions:

RQ1: Does the proposed DualRec model achieve state-of-the-art performance?

RQ2: What are the roles of intuitive and analytical (fast *vs.* slow) behavior in user preference modeling?

RQ3: To what extent does the depth of item content semantics (superficial *vs.* thorough) affect user click behavior?

RQ4: How does the proposed DualRec perform in real-world instances?

Experimental Setup

Datasets and Preprocessing We evaluate our model on two publicly available text-based[2] news recommendation

[2] We choose text-based datasets to avoid interference from other factors such as *price*, which can greatly change user behavior patterns (X. Zhang et al., 2022; J. Zhang et al., 2024).

Table 1: Dataset Statistics.

	MIND-small	MIND-large
#users	94,057	1,000,000
#news	65,238	161,013
#clicks	347,727	24,155,470
#impressions	230,117	15,777,377

datasets: MIND-small and MIND-large (F. Wu et al., 2020) that include one week and six weeks of anonymized interactions on the MSN News website, respectively (see Table 1 for statistics). MIND provides both reading logs (click history) and impression logs (item exposure and click feedback). We filter short (<5) and long (>50) reading logs and truncate titles to 20 words. Following the official splits (5 weeks/6 days *vs.* 1 week/1 day), we first train the model and then evaluate the performance using AUC for click-through rate prediction and MRR, nDCG@5, and nDCG@10 for ranking.

Implementation Details We build the item encoder based on the 12-layer PLM BERT-Base (Kenton & Toutanova, 2019). More specifically, we take the $l_1 = 3$ and $l_2 = 10$ as shallow and deep layers for item encoding, respectively. The user encoder contains two 2-layer Transformer ($L = 2$) channels, where each Transformer layer consists of $H = 20$ attention heads and an attention dimensionality of $A = 400$. For model training, we set the batch size as 32 and employ Adam to optimize equation (11) using a learning rate of $7e^{-6}$ for the BERT parameters and $7e^{-5}$ for the remaining parameters, with a 10% learning rate warmup. All experiments are conducted on one single NVIDIA GeForce RTX 3090 GPU.

Compared Methods We take the following baseline methods for comparison: **NRMS** (C. Wu et al., 2019) uses multi-head self-attention to capture sequential patterns in user interactions; **PLM-NR** adapts NRMS by incorporating a PLM for richer news representations; **HieRec** (Qi et al., 2021) extends this with a hierarchical structure to model multi-grained user interests; **UNBERT** (Q. Zhang et al., 2021) leverages a PLM for multi-grained user-item matching; **TDNR-C^2** (Shu et al., 2024) uses contrastive learning to overcome the item content authenticity bias; **PUNR** (Ma et al., 2023) utilizes PLM-inspired pre-training tasks for enhanced user interest modeling; **GREP** (Qiu et al., 2022) incorporates knowledge graph information to model user interests; **FUM** (Qi, Wu, Wu, & Huang, 2022) uses entities as interest clues within a multi-document framework; **GLORY** (Yang, Liu, Suzumura, Dong, & Li, 2023) combines global and local entity graphs for context-aware modeling.

Main Results (RQ1)

The overall performance of our DualRec and other compared methods are listed in Table 2. Notably, all results are percent numbers with "%" omitted. We run each experiment with different random seeds five times and report the average results to ensure robustness. We first observe that methods incorporating PLMs (*e.g.* PUNR and UNBERT) generally perform well. This is likely due to PLMs being able to capture the deeper semantic information relevant to analytical (System 2) click decisions. We then discover that utilizing entities associated with the content (*e.g.* GREP) also demonstrates strong performance. This may be because entities serve as cues for intuitive (System 1) clicks. However, these existing methods lack an explicit mechanism for differentiating between intuitive and analytical user behaviors and finally lead to suboptimal results. Finally, it became evident that our DualRec, which explicitly models both System 1 and System 2 processes, achieves state-of-the-art performance in all cases.

Table 2: Overall performance on MIND-small and MIND-large. The best and second-best results are highlighted using **boldface** and underline, respectively. "†": also known as "NRMS-BERT"; "*": improvements are significant at the level of 0.05 with paired *t*-test.

MIND-small				
Method	AUC	MRR	nDCG@5	nDCG@10
NRMS	65.63	30.96	34.13	40.52
PLM-NR†	68.60	32.97	36.55	42.78
HieRec	67.95	32.87	36.36	42.53
UNBERT	67.92	31.72	34.75	41.02
PUNR	68.89	33.33	36.94	43.10
TDNR-C^2	68.89	33.57	37.23	43.39
GREP	68.12	33.75	37.25	43.37
FUM	67.11	31.31	35.08	41.42
GLORY	68.15	32.97	36.47	42.78
DualRec	**69.52***	**34.52***	**38.19***	**44.29***
MIND-large				
NRMS	68.24	33.49	36.56	42.24
PLM-NR†	69.50	34.75	37.99	43.72
HieRec	69.03	33.89	37.08	43.01
UNBERT	70.68	35.68	39.13	44.78
PUNR	71.03	35.17	39.04	45.41
TDNR-C^2	70.38	34.62	38.12	44.30
GREP	69.44	34.40	37.54	43.22
FUM	70.01	34.51	37.68	43.38
GLORY	69.45	34.03	37.92	44.19
DualRec	**71.83***	**36.02***	**40.05***	**46.36***

Further Analysis

Ablation Study (RQ2) To analyze the impact of System 1 and System 2 behaviors, we ablate DualRec as follows: (1) *w/o* fusion: Replaces the proposed fusion with a simple average of $\mathbf{s}^1$ and $\mathbf{s}^2$; (2) *w/o* Sys1: Removes superficial cues embedding $\boldsymbol{i}^1$ and intuitive channel in user encoder; (3) *w/o* Sys2: Removes deep semantic embedding $\boldsymbol{i}^2$ and analytical channel. As shown in Figure 3, all variants lead to immediate performance drops, confirming the contribution of both System 1 and System 2 in users' online interactions

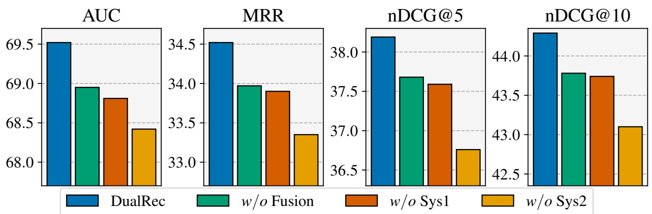

Figure 3: Ablation results on MIND-small. "*w/o*" statnds for "without".

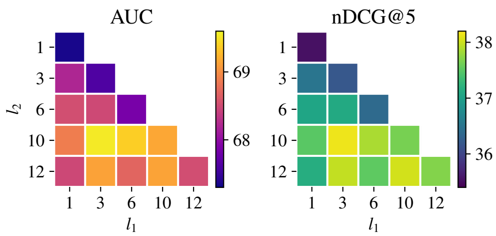

Figure 4: DualRec performance with varied PLM layer settings. l_1=3 and l_2=10 is our default setup.

with the RS. Notably, *w/o* Sys2 performs the worst, indicating that slow and analytical behavior is widely conducted. *w/o* Fusion also performs worse, suggesting that user preferences are jointly shaped by the intuitive and analytical thinking.

The Impact of Item Comprehending (RQ3) To investigate the relationship between shallow and deep semantic comprehension and intuitive/analytical clicks, we vary the PLM layers (l_1 and l_2) used to generate item embeddings, exploring combinations of $l_1, l_2 \in \{1, 3, 6, 10, 12\}$. The results are visualized as heatmaps in Figure 4. We observe that adopting identical item representations (l_1=l_2) in both channels leads to the worst performance, supporting our hypothesis that distinct cognitive processes (System 1 and System 2) are at play throughout the "read then click" process. Furthermore, utilizing shallowest (l_1=1) or deepest (l_2=12) layer outputs does not necessarily guarantee a better performance. Extremely shallow layers may lack sufficient information to represent even superficial features, while overly deep layers may introduce excessive complexity that leads to overfitting.

Case Study (RQ4) We further conduct a case study to illustrate the real-world performance of DualRec. Table 3 shows recommendations for user U19812, whose click history reveals a broad interest and a tendency towards heuristic keywords (*e.g.* "woman dead", "Chick-fil-A apologizes" and "fatal crash"). This confirms the prevalence of intuitive, System 1-driven clicks in real-world interactions. DualRec, designed to capture superficial cues and model intuitive clicks, correctly ranks N23699 ("judges drink & brawl") as the top recommendation. In contrast, PUNR and GREP, which primarily focus on semantic matching and knowledge-based reasoning (System 2), rank N23699 lower. Similarly, DualRec's

Table 3: Case study based on user U19812's log. "D", "P", and "G" represent DualRec, PUNR, and GREP, respectively. Models will rank items more likely to be clicked higher.

User U19812's reading history					
ID	News Title				
N25457	Woman, suspect dead at 'Tarzan' actor Ron Ely's California residence				
N6808	Payton declines to name Saints starting quarterback for Week 8				
N8030	The Warriors' fall from glory, explained				
N15814	Army Officer Who Heard Trump's Ukraine Call Reported Concerns				
N2219	Chick-fil-A Apologizes to Customers for Promoting National Sandwich Day Which Is on a Sunday				
N10972	Cause determined in Jessi Combs' fatal speed record crash				
Recommendation (*rank out of 10 candidate news*)					
ID	Click?	News Title	D	P	G
N23699	✓	3 Indiana judges suspended after a night of drinking turned into a White Castle brawl	1	5	4
N23699	✓	Some believe Mason Rudolph, hit in head with his own helmet, isn't getting enough blame	2	3	1

second recommendation N23699 leverages both System 2 level interest matching (football) and System 1 level keyword intuitions ("not enough blame"). While GREP correctly identifies the entity connection (Payton and Mason are all NFL players) and ranks N23699 at #1, PUNR again ranks it lower, demonstrating the important role of both intuitive and analytical behaviors in providing more accurate recommendations.

Conclusion

In this work, we carefully review how users interact with the recommender systems, and observe that their actions can be attributed to intuitive and analytical behaviors. Inspired by the Dual Process Theory, we present DualRec, which mimics the behavioral duality at both *read* and *click* stages. DualRec utilizes different layers of PLM to capture both superficial cues and deep semantic features of items. Then a Transformers-based dual-channel user encoder, employing dedicated attention mechanisms, models both "fast" intuitive and "slow" deliberate click patterns. Finally, a fusion layer is proposed to dynamically balance these two cognitive processes. Extensive experiments on two real-world datasets demonstrate DualRec's superior performance, confirming the importance of considering both intuitive and analytical behaviors for effective personalized recommendations. In the future, we aim to extend our findings to more complex scenarios, such as multimodal recommendation, where cognitive processes toward images and videos can pose a great challenge to behavioral modeling.

Acknowledgments

We would like to thank our anonymous reviewers for providing insightful comments. This work was supported by the National Natural Science Foundation of China (Grant No. 62376051, 62076046).

References

Åberg, J., & Shahmehri, N. (2000). The role of human web assistants in e-commerce: an analysis and a usability study. *Internet research*, *10*(2), 114–125.

Agarwal, A., Usunier, N., Lazaric, A., & Nickel, M. (2024). System-2 recommenders: Disentangling utility and engagement in recommendation systems via temporal point-processes. In *The 2024 acm conference on fairness, accountability, and transparency* (pp. 1763–1773).

Bahdanau, D. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Boka, T. F., Niu, Z., & Neupane, R. B. (2024). A survey of sequential recommendation systems: Techniques, evaluation, and future directions. *Information Systems*, 102427.

Carr, N. (2020). *The shallows: What the internet is doing to our brains*. WW Norton & Company.

Evans, J. S. B. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental review*, *31*(2-3), 86–102.

Evans, J. S. B., & Frankish, K. E. (2009). *In two minds: Dual processes and beyond.* Oxford University Press.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition [Journal Article]. *Annual Review of Psychology*, *59*(Volume 59, 2008), 255-278.

Frankish, K. (2010). Dual-process and dual-system theories of reasoning. *Philosophy Compass*, *5*(10), 914–926.

Guo, H., Tang, R., Ye, Y., Li, Z., & He, X. (2017). Deepfm: a factorization-machine based neural network for ctr prediction. In *Ijcai* (pp. 1725–1731).

Guy, I., Zwerdling, N., Ronen, I., Carmel, D., & Uziel, E. (2010). Social media recommendation based on people and tags. In *Proceedings of the 33rd international acm sigir conference on research and development in information retrieval* (pp. 194–201).

He, R., & McAuley, J. (2016). Fusing similarity models with markov chains for sparse sequential recommendation. In *2016 ieee 16th international conference on data mining (icdm)* (pp. 191–200).

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T.-S. (2017). Neural collaborative filtering. In *Www* (pp. 173–182).

Hidasi, B., Karatzoglou, A., Baltrunas, L., & Tikk, D. (2016). Session-based recommendations with recurrent neural networks. In Y. Bengio & Y. LeCun (Eds.), *4th international conference on learning representations, ICLR 2016, san juan, puerto rico, may 2-4, 2016, conference track proceedings.*

Hou, Y., Mu, S., Zhao, W. X., Li, Y., Ding, B., & Wen, J.-R. (2022). Towards universal sequence representation learning for recommender systems. In *Proceedings of the 28th acm sigkdd conference on knowledge discovery and data mining* (pp. 585–593).

Jawahar, G., Sagot, B., & Seddah, D. (2019). What does bert learn about the structure of language? In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 3651–3657).

Kahneman, D. (2011). Thinking, fast and slow. *Farrar, Straus and Giroux*.

Kang, W.-C., & McAuley, J. (2018). Self-attentive sequential recommendation. In *2018 ieee international conference on data mining (icdm)* (pp. 197–206).

Kenton, J. D. M.-W. C., & Toutanova, L. K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacl-hlt* (pp. 4171–4186).

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2024). The challenge of understanding what users want: Inconsistent preferences and engagement optimization. *Management science*, *70*(9), 6336–6355.

Li, J., Wang, M., Li, J., Fu, J., Shen, X., Shang, J., & McAuley, J. (2023). Text is all you need: Learning language representations for sequential recommendation. In *Proceedings of the 29th acm sigkdd conference on knowledge discovery and data mining* (pp. 1258–1267).

Li, J., Zhu, J., Bi, Q., Cai, G., Shang, L., Dong, Z., … Liu, Q. (2022). Miner: Multi-interest matching network for news recommendation. In *Findings of the association for computational linguistics: Acl 2022* (pp. 343–352).

Liu, N. F. (2019). Linguistic knowledge and transferability of contextual representations. *arXiv preprint arXiv:1903.08855*.

Liu, Z. (2005). Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of documentation*, *61*(6), 700–712.

Loh, K. K., & Kanai, R. (2016). How has the internet reshaped human cognition? *The Neuroscientist*, *22*(5), 506–520.

Lyngs, U., Binns, R., Van Kleek, M., & Shadbolt, N. (2018). So, tell me what users want, what they really, really want! In *Extended abstracts of the 2018 chi conference on human factors in computing systems* (pp. 1–10).

Lyngs, U., Lukoff, K., Slovak, P., Binns, R., Slack, A., Inzlicht, M., … Shadbolt, N. (2019). Self-control in cyberspace: Applying dual systems theory to a review of digital self-control tools. In *proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1–18).

Ma, G., Liu, H., Xing, W., Qian, W., Lv, Z., Yang, Q., & Hu, S. (2023). Punr: Pre-training with user behavior modeling for news recommendation. In *Findings of the association for computational linguistics: Emnlp 2023* (pp. 8338–8347).

Margolis, H. (1987). *Patterns, thinking, and cognition: A* *theory of judgment*. University of Chicago Press.

McAuley, J., Targett, C., Shi, Q., & Van Den Hengel, A. (2015). Image-based recommendations on styles and substitutes. In *Proceedings of the 38th international acm sigir conference on research and development in information retrieval* (pp. 43–52).

Milli, S., Belli, L., & Hardt, M. (2021). From optimizing engagement to measuring value. In *Proceedings of the 2021 acm conference on fairness, accountability, and transparency* (pp. 714–722).

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 1532–1543).

Pinder, C., Vermeulen, J., Cowan, B. R., & Beale, R. (2018). Digital behaviour change interventions to break and form habits. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *25*(3), 1–66.

Qi, T., Wu, F., Wu, C., & Huang, Y. (2022). Fum: Fine-grained and fast user modeling for news recommendation. *SIGIR*.

Qi, T., Wu, F., Wu, C., Yang, P., Yu, Y., Xie, X., & Huang, Y. (2021). Hierec: Hierarchical user interest modeling for personalized news recommendation. In *Acl-ijcnlp* (pp. 5446–5456).

Qiu, Z., Hu, Y., & Wu, X. (2022). Graph neural news recommendation with user existing and potential interest modeling. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *16*(5), 1–17.

Rendle, S. (2012). Factorization machines with libfm. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *3*(3), 1–22.

Rendle, S., Krichene, W., Zhang, L., & Anderson, J. (2020). Neural collaborative filtering vs. matrix factorization revisited. In *Proceedings of the 14th acm conference on recommender systems* (pp. 240–248).

Savolainen, R. (1995). Everyday life information seeking: Approaching information seeking in the context of "way of life". *Library & information science research*, *17*(3), 259–294.

Shu, Y., Zhang, X., Wu, Y., Xu, B., Yang, L., & Lin, H. (2024). Don't click the bait: Title debiasing news recommendation via cross-field contrastive learning. In *Ccf international conference on natural language processing and chinese computing* (pp. 224–236).

Stanovich, K. E., & West, R. F. (2002). Individual differences in reasoning: Implications for the rationality debate? In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (p. 421–440). Cambridge University Press.

Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W., & Jiang, P. (2019). Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th acm international conference on information and knowledge management* (pp. 1441–1450).

Tang, J., & Wang, K. (2018). Personalized top-n sequential recommendation via convolutional sequence embedding. In *Proceedings of the eleventh acm international conference on web search and data mining* (pp. 565–573).

Wu, C., Wu, F., Ge, S., Qi, T., Huang, Y., & Xie, X. (2019). Neural news recommendation with multi-head self-attention. In *Emnlp-ijcnlp* (pp. 6389–6394).

Wu, F., Qiao, Y., Chen, J.-H., Wu, C., Qi, T., Lian, J., … Zhou, M. (2020, July). MIND: A large-scale dataset for news recommendation. In *Acl* (pp. 3597–3606). Online: Association for Computational Linguistics.

Yang, B., Liu, D., Suzumura, T., Dong, R., & Li, I. (2023). Going beyond local: Global graph-enhanced personalized news recommendations. In *Recsys* (pp. 24–34).

Yuan, Z., Yuan, F., Song, Y., Li, Y., Fu, J., Yang, F., … Ni, Y. (2023). Where to go next for recommender systems? id-vs. modality-based recommender models revisited. In *Proceedings of the 46th international acm sigir conference on research and development in information retrieval* (pp. 2639–2649).

Zhang, J., Cheng, Y., Ni, Y., Pan, Y., Yuan, Z., Fu, J., … Yuan, F. (2024). Ninerec: A benchmark dataset suite for evaluating transferable recommendation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*(01), 1–12.

Zhang, Q., Li, J., Jia, Q., Wang, C., Zhu, J., Wang, Z., & He, X. (2021). Unbert: User-news matching bert for news recommendation. In *Ijcai* (Vol. 21, pp. 3356–3362).

Zhang, X., Xu, B., Yang, L., Li, C., Ma, F., Liu, H., & Lin, H. (2022). Price does matter! modeling price and interest preferences in session-based recommendation. In *Proceedings of the 45th international acm sigir conference on research and development in information retrieval* (pp. 1684–1693).