

DFRec: Dual Fluctuation Modeling of Multi-level Intent Evolution for Next-Item Recommendation

Nengjun Zhu^{1*}, Lingdan Sun¹, Qi Zhang², Jian Cao³, Hang Yu¹

¹School of Computer Engineering and Science, Shanghai University, China

²School of Computer Science, Tongji University, China

³School of Computer Science, Shanghai Jiao Tong University, China

{zhu_nj, sunld1127, yuhang}@shu.edu.cn, zhangqi_cs@tongji.edu.cn, cao-jian@sjtu.edu.cn

Abstract

User sequential behaviors are driven by a variety of complex and evolving intents. Capturing the dynamic change of user intents has become critical yet challenging in the next-item recommendation. Existing studies usually model the transition relationships among multiple intents within a session or integrate temporal information to capture the dynamic evolution of user intents. However, they struggle to identify the precise timing and magnitudes of intent changes, leading to ambiguity in providing consistent or violated recommendations and ultimately yielding subpar performance. To this end, we propose a novel framework called **Dual Fluctuation Modeling of Multi-level Intent Evolution for Next-Item Recommendation (DFRec)** in this paper. DFRec explicitly identifies the user intent changes and further quantifies the magnitude of the changes. Specifically, we assume that a user’s intent fluctuates around an inherent intent, with the magnitude of fluctuations indicating the extent of changes in user intents. Thus, we design an Emerging Intent Generation Module that employs a normal distribution with dynamic variance to capture intent fluctuations at each time step. Furthermore, we introduce a dual-layer dynamic variance update mechanism to capture fluctuation characteristics at different temporal levels, enhancing the representation of possible emergent intents. Extensive experiments on three real-world datasets verify DFRec’s superiority over state-of-the-art baselines.

Code — <https://github.com/vgeek-z/DFRec>

Introduction

Next-item Recommender systems (RS) aim to predict the user’s next action based on historical interactions. User intents are key factors determining users’ interactions with the next items (Zhu et al. 2020; Qin et al. 2024), as they reveal the underlying reasons behind user behaviors. Growing research attention is dedicated to intent modeling for enhanced next-item recommendations, thereby fostering the development of intent-based next-item recommendations. However, user intents are typically intricate and evolve dynamically with time, presenting significant challenges in capturing precise intent dynamics in intent-based recommendations.

*Corresponding author.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

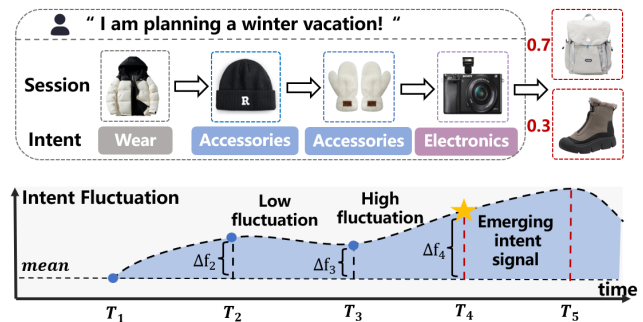


Figure 1: An example illustrates user intent fluctuation and how emerging intent can be triggered in a shopping scenario.

To capture the dynamics of user intent, existing methods (Luo et al. 2020; Liu et al. 2020; Li et al. 2022a) typically model the transition relationships among multiple intents within a session. For example, ISRec (Li et al. 2022a) extracts the intents of users from sequential contexts, and then takes complex intent transitions into account through an intention graph. Alternative approaches (Zhang et al. 2022; Fan et al. 2021; Huang et al. 2023) explicitly or implicitly integrate temporal information to capture the dynamic evolution of user intents. These methods establish intent modeling and reveal its effectiveness to recommendation performance.

However, these methods face a key limitation: failing to identify the precise timing and magnitudes of intent changes, which results in limited performance improvement of current intent-based recommendations. Intuitively, let us consider a familiar shopping scenario where a user plans a winter vacation, as shown in Figure 1. The user’s initial interactions revolve around winter apparel. Herein, a sudden shift to engaging with a digital camera in the subsequent interaction signals a significant fluctuation in user intent, revealing an emerged intent (Zhu et al. 2024) on documenting the user’s vacation experiences. It is essential to promptly and accurately capture such intent fluctuations, as this enables next-item RS to suggest a travel backpack in a more prominent position, rather than shoes within the apparel category.

Accordingly, we contend that it is crucial to identify the intent fluctuations and quantify the fluctuation magnitude

to understand the dynamic change of user intents effectively. Naturally, modeling user intent fluctuations precisely is an extremely challenging task. **First**, the uncertainty of user intent fluctuations makes it difficult to model the randomness of these fluctuations (**CH1**). **Second**, the context-dependency of intent fluctuations presents a persistent challenge in effectively incorporating intricate contextual cues to quantify the magnitude of the fluctuations dynamically (**CH2**).

In this paper, we propose a novel framework called **Dual Fluctuation Modeling of Multi-level Intent Evolution for Next-Item Recommendation (DFRec)** to address the above issues. DFRec generates multiple consecutive item segments and explicitly identifies the magnitude of changes in user intent between different segments. Specifically, we assume that a user’s intent fluctuates around an *inherent intent*, with the magnitude of fluctuations indicating the extent of changes in user intents. Thus, we start with the first item in the session and progressively include more items to generate multiple consecutive item segments. Then **Emerging Intent Generation Module (EIGM)** is proposed to probe users’ intent fluctuations at each time step, implemented by a normal distribution with dynamic variance. Herein, the user’s inherent intent in item categories serves as the base, while the variance from the base represents the magnitude of the user’s intent fluctuation at the current time step. Thus, the emerging intent representation can be captured through sampling at each time step. Compared to existing methods that directly vectorize user intent, our generative approach incorporates uncertainty to model the randomness of user intent fluctuation, effectively tackling the challenge **CH1** and accordingly identifying intent fluctuations. To address the challenge **CH2**, we further design a **Dual Intent Fluctuation Modeling (DIFM)** mechanism to capture fluctuation characteristics across different temporal levels. The short-term fluctuation variance primarily captures the immediate intent fluctuations between consecutive time steps, while the long-term fluctuation variance considers the intent fluctuation from the base over a time segment through a sliding window mechanism. Finally, **Intent Fusion Module (IFM)** introduces a novel attention mechanism to combine multiple consecutive item segments with the corresponding emerging intent representations at each time step to form the final session representation. The representation is then utilized to make recommendations.

Our key contributions are summarized as follows:

- We propose a novel framework, DFRec, which generates multiple consecutive item segments and explicitly identifies the magnitude of user intent changes between different segments.
- We capture the extent of changes in user intents by modeling intent fluctuations at each time step, utilizing a normal distribution with a dynamically adjusted variance.
- We design a dual variance update mechanism to model long- and short-term intent fluctuations simultaneously.
- Extensive experiments on three real-world datasets show the superiority of our DFRec compared with various state-of-the-art solutions.

Related Work

Many recent approaches have turned their attention to studying users’ intentions to improve the performance of recommender systems. Some studies assume that users’ decisions can be influenced by multiple intents. Thus they explore a user’s multiple latent intentions from their interaction sequence in different ways. MCPRN (Wang et al. 2019) has a channel for each implicit intent and uses a purpose routing network to detect the purpose of each item and assign it to the corresponding channel. MSGIFSR (Guo et al. 2022) treats a group of locally consecutive items in a session as user intents and uses a multi-granularity heterogeneous session graph to capture the interaction between different granularity intent units. MiaSRec (Choi et al. 2024) represents various user intents by deriving multiple session representations centered on each item and dynamically selecting the important ones. In addition, some methods (Li et al. 2022b; Gao et al. 2023; Wu et al. 2023) use unentangled representation learning to embed user or item representation into multiple sub-channels, each of which captures a specific user intent. Thus they can explore user intents at different factor levels. However, these methods focus solely on the diversity of user intentions while overlooking that the users’ intents change over time.

To model the dynamic change of user intents, MIT-GNN (Liu et al. 2020) proposes a multi-intent translation graph neural network to mine users’ multiple intents by considering the correlations of the intents. GSRec (Fan et al. 2021) designs a continuous-time bipartite graph, which captures temporal dynamics within the sequential patterns of user-item interactions. DIDN (Zhang et al. 2022) captures the user’s dynamic intentions by learning dynamic item embeddings that incorporate the temporal order of items within the session. As a significant auxiliary information for items, category information has been explored in recommendation areas. Besides, the work (Zhu et al. 2024) proposed decoupling user behaviors to distinguish between the user’s inherent intents and emerging intents. This approach primarily identifies items with emerging intent signals from the item-level attributes. Although the aforementioned approaches have been shown to enhance recommendation performance, they fail to account for the fine-grained modeling of the moments and magnitudes of intent changes. As a result, these methods may fail to model user intent comprehensively.

Unlike previous works, our approach explicitly identifies the intent fluctuations and quantifies fluctuation magnitude to model the extent of changes in user intents. We argue that users’ interaction patterns reflect the sequential fluctuation of their intents. Capturing the fluctuation in user intent beyond their inherent intents at a specific moment can reveal underlying, dynamically shifting intents.

The Proposed DFRec

Problem Definition

Let $\mathcal{I} = \{e_1, e_2, \dots, e_j, \dots, e_N\}$ and $\mathcal{C} = \{c_1, c_2, \dots, c_K\}$ denote a set of items and a set of item categories, where N and K are the numbers of items and categories, respectively. Each item e_j has an initial randomized embedding v_j . Let

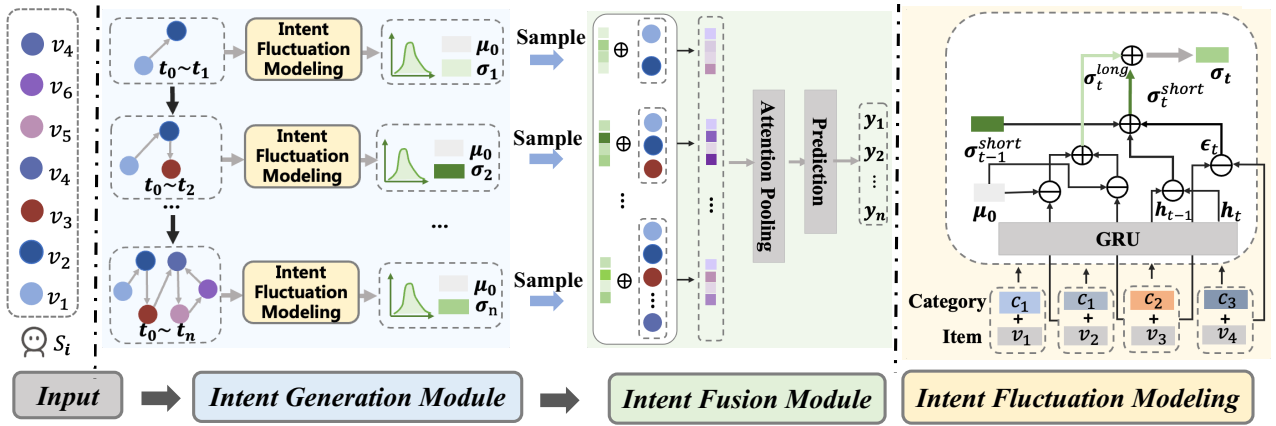


Figure 2: The framework of DFRec.

$\{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_i, \dots\}$ be a set of anonymous sessions, where $\mathcal{S}_i = \{e_1^i, e_2^i, \dots, e_j^i, \dots\}$ represents a session, which is a collection of clicked items following the chronological order, and $e_j^i \in \mathcal{I}$ is an item interacted at time index j . Given a session \mathcal{S}_i and its \mathcal{C}_i , our task is to predict the next item with the highest recommendation scores.

Overall Framework

DFRec has three main components as shown in Figure 2.

1) Emerging Intent Generation Module (EIGM): it utilizes a normal distribution $\mathcal{N}(\mu_t, \sigma_t^2)$ with dynamic variance to capture the user’s intent fluctuation at each time step. First, the inherent intent of users for item categories is obtained through Dirichlet distribution, which is considered as the baseline or mean μ_0 of user intent. Then, the user’s intent fluctuation is captured by the Dual Intent Fluctuation Modeling (DIFM) for each time step, modeled as the variance σ_t^2 at that moment and representing the extent of change in the user’s intent. Therefore, at each time step, the current user’s emerging intent representation is generated by sampling from $\mathcal{N}(\mu_t, \sigma_t^2)$, where $\mu_t = \mu_0$.

2) Dual Intent Fluctuation Modeling (DIFM): it introduces a dual variance update mechanism to consider fluctuation characteristics across different temporal levels, enabling a more comprehensive capture of user intent fluctuations. Short-term variance is updated based on the differences between the current and previous time steps, capturing immediate intent fluctuations within a session. Long-term variance is updated by setting a sliding window over multiple time steps, considering the residuals and the difference between the current step’s intent and the inherent intent.

3) Intent Fusion Module (IFM): it generates multiple consecutive item units with different lengths and fuses them with the emerging intent representations using an attention mechanism, resulting in a comprehensive session representation.

Emerging Intent Generation Module (EIGM)

The uncertainty in user interaction behavior is reflected in the possibility that users may randomly deviate from their

inherent intent and explore new options during a session. To identify the extent of changes in user intents, we capture the user intent fluctuations at a given time step by considering how much it exceeds their inherent intent. This reflects the user’s underlying and dynamically changing exploration and shifting behavior, such as exploring new item categories or a sudden increase in demand for certain specific categories.

To effectively model intent fluctuations, we utilize a normal distribution $\mathcal{N}(\mu_t, \sigma_t^2)$ with dynamic variance to adjust the intent deviation probabilistically at each time step. This generative approach incorporates uncertainty to model the randomness of user intent fluctuation and accordingly identify intent fluctuations. Assuming that the user’s intents fluctuate around the inherent intent μ_0 , the amplitude of these fluctuations is determined by the user’s current behaviors. Thus, the user’s emerging intent can be represented as:

$$\mu_{t+1} \sim \mathcal{N}(\mu_t, \sigma_t^2) \quad (1)$$

The item categories reflect user intents. We first obtain the user’s inherent intent for item categories through a Dirichlet distribution, which is considered as the baseline or mean μ_0 of user intent. The Dirichlet distribution (Blei, Ng, and Jordan 2003) is a common prior distribution used in multinomial distributions, which naturally represents the user’s preference distribution over different item categories. Specifically, for each item category c_j in the current session \mathcal{S}_i , we compute an intent relevance strength $\tilde{\omega}_j$ by incorporating the frequency n_{c_j} of category within the session and the category representation \mathbf{c}_j :

$$\tilde{\omega}_j = \text{ReLU}(\mathbf{W}^T \mathbf{c}_j + \mathbf{b}) + \lambda \cdot n_{c_j} \quad (2)$$

where λ is a scaling factor, $\mathbf{W} \in \mathbb{R}^{d \times 1}$ and \mathbf{b} are trainable parameters. ReLU is an activation function that ensures the $\tilde{\omega}_j$ remains a positive value. Then, we sample the user’s interest distribution in each item category $\mathbf{p}^i = (p_1^i, p_2^i, \dots, p_{|C_i|}^i)$ from the Dirichlet distribution as follows:

$$\mathbf{p}^i \sim \text{Dirichlet}(\tilde{\omega}_1^i, \tilde{\omega}_2^i, \dots, \tilde{\omega}_{|C_i|}^i) \quad (3)$$

where \mathbf{p}^i can be viewed as the coarse-grained intent distribution during the session \mathcal{S}_i , and $|C_i|$ represents the number

of categories in the \mathcal{S}_i . Then, the baseline of user intent μ_0 is determined by:

$$\mu_0^i = \sum_{c_j} p_j^i c_j \quad (4)$$

Dual Intent Fluctuation Modeling (DIFM)

Identifying the extent of changes in user intents can be considered as capturing the user's intent fluctuations beyond the inherent intent at a certain moment, reflecting the user's potential and dynamic exploration and shifting behaviors. Inspired by the GARCH (Bollerslev 1986), which is mainly used to model and predict volatility in financial data, we design a dual intent fluctuation mechanism to adjust users' intent fluctuation dynamically. The core idea is to use the historical fluctuation and the change of user intent to predict future fluctuation, to dynamically update the amplitude of user intent fluctuation σ_t . When σ_t is larger, it indicates that the user's intent fluctuation is stronger, and thus, a higher probability of emerging a new intent. In such a case, we should make a bold recommendation instead of exploiting existing behaviors discreetly.

First, the initial embedding vectors of items and categories in the current session \mathcal{S}_i are combined as the input e_i :

$$e_t^i = v_t^i + c_t^i, t \in \{1, \dots, n\} \quad (5)$$

Then we employ the Gate Recurrent Unit (GRU) to encode the current user's historical interaction and the hidden state for each time step is obtained as follows:

$$\dots, h_t^i, \dots = GRU(\dots, e_t^i, \dots) \quad (6)$$

To consider fluctuation characteristics across different temporal levels, we propose a dual intent fluctuation mechanism, which simultaneously considers both short-term and long-term fluctuation variances. Short-term variance is updated based on the differences between the current and previous time steps, capturing immediate intent fluctuations within a session. We first obtain the user's intent offset for categories between the current and previous time steps:

$$\epsilon_t^i = e_t^i - e_{t-1}^i \quad (7)$$

Then, the fluctuation of short-term user intent, i.e., σ_t^{short} is predicted via historical fluctuation and the change of user intent as follows:

$$\sigma_t^{short} = \alpha \cdot \epsilon_t^2 + \beta \cdot (\sigma_{t-1}^{short})^2 + \gamma \cdot (h_t - h_{t-1})^2 \quad (8)$$

where σ_{t-1}^{short} denotes the short-term variance at $t-1$ step, h_t denotes the hidden states fused with timing series. α , β , and γ are learnable parameters to tune the influence of each part.

Long-term fluctuations refer to changes in user intent over a longer time frame, where the differences between the current step's intent and the inherent intent should be considered. To capture this level of fluctuation, we introduce a sliding window mechanism to track the long-term trend and the degree of fluctuations in user intent:

$$\sigma_t^{long} = \frac{1}{N_w} \sum_{k=t-N_w}^{t-1} (e_k^i - \mu_0)^2 + \delta \cdot (h_k - \mu_0)^2 \quad (9)$$

where N_w denotes the length of the sliding window, and σ_t^{long} is the summation of intent fluctuation at the sliding window.

So far, the fused intent fluctuation variance σ_t^i for session \mathcal{S}_i is obtained as follows:

$$\sigma_t^i = \sigma_t^{short} + \sigma_t^{long} \quad (10)$$

Intent Fusion Module (IFM)

User intents exhibit hierarchy (Zhu et al. 2020), showing different focal points of intent over various time spans. To model the extent of changes in user intents more precisely, we propose an Intent Fusion Module (IFM), which composes the recommendation results from various granularity item segments and the corresponding emerging intents.

We first generate various granularity item segments for the current session \mathcal{S}_i and apply linear transformations to generate multi-level item units. Specifically, we start with the first item, progressively adding more items to form new segments. This allows us to capture the incremental changes in user intent at varying time granularities.

$$q_t^i = W_t \left(\sum_{j=1, \dots, t} \{v_j\} \right) \quad (11)$$

where v_i is the embedding of the i^{th} item in session \mathcal{S}_i , and $W_t \in \mathbb{R}^{d \times d}$, $t = 1, 2, \dots, T$ refers to a learnable matrix. Besides, the emerging intent representation of each time step r_t^i can be sampled from the normal distribution $\mathcal{N}(\mu_t, \sigma_t^2)$ as in Eqn. (1). Thus q_t^i is further updated by incorporating emerging intent information as follows:

$$q_t^i = fc(\text{concat}(q_t^i, r_t^i)) \quad (12)$$

where $\text{concat}()$ is a concatenating operation and $fc()$ represents a fully-connected layer network. Then these units are used to compute multi-head attention weights as follows:

$$\alpha_h = \text{softmax} \left(\frac{QW_Q(KW_K)^T}{\sqrt{d}} \right) \quad (13)$$

where $Q = (q_1, \dots, q_t)^T \in \mathbb{R}^{t \times d}$ is the query matrix, $K = (v_1, \dots, v_n)^T \in \mathbb{R}^{n \times d}$ represents the hidden representation of each item in the sequence and $W_Q, W_K \in \mathbb{R}^{d \times d}$ are trainable parameters. $h = 1, 2, \dots, H$ denotes the different attention head indexes. Then following (Zhang et al. 2023), we apply L_p -pooling to pool the attention map and multiply the hidden representation of the items in the sequence with the corresponding pooled attention weights to get the final session representation s_i^{hybrid} .

$$s_i^{hybrid} = \text{concat}(\{s_h^i\}_{h=1, \dots, H}) \quad (14)$$

$$s_h^i = \sum_{j=1}^n \hat{\alpha}_{j,h} v_j \quad (15)$$

$$\hat{\alpha}_{j,h} = \left[\sum_{m=0}^{l-1} (\alpha_{j,mH+h})^p \right]^{\frac{1}{p}} \quad (16)$$

where $\hat{\alpha}_{j,h}$ is the output of the pooling operator at the location (j, h) , and $\alpha_{j,mH+h}$ is the feature value within the pooling region. s_h^i is the session representation under the h^{th} head.

Model Inference and Recommendation

We utilize s_i^{hybrid} to predict the recommendation scores of an item v_j as the next items for a session \mathcal{S}_i :

$$\hat{y}_{ij} = \sigma(v_j^i \top s_i^{hybrid}) \quad (17)$$

where $\sigma()$ is a sigmoid function. We formulate the learning objective as a cross-entropy loss function, which has been extensively used in recommender systems and is defined as:

$$L = - \sum_i \log(\hat{y}_{ij^+}) + \log(1 - \hat{y}_{ij^-}) \quad (18)$$

where (i, j^+) is a positive pair representing that item v_{j^+} is a ground-truth for session \mathcal{S}_i , and instead, (i, j^-) is a sampled negative pair and $j^- \neq j^+$.

Complexity Analysis

The time complexity of the DFRec is derived from several integral components. For EIGM, the complexity is $\mathcal{O}(TK)$, where T signifies the session length and K denotes the number of categories. In the dual-layer intent fluctuation modeling, the short-term fluctuation calculation, which involves the computation of the GRU hidden layer, has a complexity of $\mathcal{O}(Td^2)$, while the long-term fluctuation has a complexity of $\mathcal{O}(TWd)$, where W represents the window size and d denotes the embedding dimension. The intent fusion module encompasses a multi-head attention mechanism with a time complexity of $\mathcal{O}(T^2d)$ and an L_p -pooling operation with a time complexity of $\mathcal{O}(Td)$. Overall, the total time complexity of the DFRec is $\mathcal{O}(TK + Td^2 + TWd + T^2d + Td)$.

Experiment

Experimental Setting

Datasets. We conduct experiments on three datasets, i.e., (1) **Yoochoose**¹: RecSys’15 Challenge dataset from an e-commerce website. (2) **Jdata**²: a challenge dataset hosted by JD corporation. It is filtered by a 1-hour duration to extract the session data. (3) **Diginetica**³: an e-commerce dataset composed of user purchasing behaviors.

For all datasets, we filter out the sessions whose length is less than 2 and the items that appear less than 5 times. Besides, we apply a data augmentation by treating the i_{th} item as the label and the ones before the i_{th} one as the input sequence. Dataset statistics are recorded in Table 1. The three datasets are chosen to diversify the number of categories, i.e., 12, 79, and 982, which can be used to evaluate their impacts.

Baselines. We compare our DFRec with the following four types of baselines: conventional static approaches, i.e., BPR; typical session-based RSs, i.e., NARM and SASRec; graph-based RSs, i.e., SR-GNN, GCE-GNN, and SHARE; and intent-based RSs, i.e., DIDN, Atten-Mixer, and MiaSRec.

1) **BPR** (Rendle et al. 2012) optimizes the matrix factorization with implicit negative feedback using a pairwise loss

¹<https://www.kaggle.com/chadgostopp/recsys-challenge-2015>

²<https://jdata.jd.com/html/detail.html?id=8>

³<https://competitions.codalab.org/competitions/11161>

Dataset	Yoochoose	Jdata	Diginetica
# of training sessions	189,166	221,879	153,387
# of test sessions	29,780	31,807	24,487
# of items	12,576	31,400	15,660
# of categories	12	79	982
average length	7.35	7.97	7.26

Table 1: Statistical information of evaluated datasets.

function. 2) **NARM** (Li et al. 2017) improves GRU4REC by incorporating attention into RNNs for session-based RSs. 3) **SASRec** (Kang and McAuley 2018) uses a transformer model with a single-head attention mechanism to recommend the next item. 4) **SR-GNN** (Wu et al. 2019) is the first to apply a gated graph convolutional network to establish transition relationships between items within a session. 5) **GCE-GNN** (Wang et al. 2020) exploits item transitions over the current and all-session graphs to learn session-level and global-level embeddings. 6) **SHARE** (Wang et al. 2021) utilizes hypergraph attention networks to exploit the contextual windows to model session-wise item representations. 7) **DIDN** (Zhang et al. 2022) designs a dynamic intent-aware module incorporating item-, user-, and temporal-aware information to learn dynamic item embeddings. 8) **Atten-Mixer** (Zhang et al. 2023) leverages both concept-view and instance-view readouts to achieve multi-level reasoning over 3 item transitions to model multi-level user intent. 9) **MiaSRec** (Choi et al. 2024) represents various user intents by deriving multiple session representations centered on each item and selecting the important ones.

Metrics. We employ commonly used **Recall@K** and **NDCG@K** (Normalized Discounted Cumulative Gain) to evaluate the recommendation results. Recall@K indicates the proportion of test cases in which the desired item appears in the Top-K items. NDCG@K takes the exact position of the correctly recommended items into account.

Parameter Settings. In our experiments, models are tuned for their best performance using Adam to optimize the parameters. For example, the initial learning rate is 0.001 for all datasets. The batch size is set to 100 and the embedding dimension size is set to 256. In addition, the window size W in Eqn. (9) is searched in the ranges of $\{1, 2, 3, 4, 5, 6\}$.

Overall Performance

Table 2 illustrates the overall performance of our DFRec and the baselines on three datasets regarding Recall and NDCG metrics. We can have the following main observations:

First, DFRec generally outperforms all the baselines on all datasets, verifying the superiority of our model. For example, DFRec outperforms the best-performing baseline (i.e., MiaSRec) by 1.07% and the worst-performing session-based methods (i.e., NARM) by 12.19% on Jdata at Recall@5.

Second, the intent-based models (i.e., DIDN, Atten-Mixer, and MiaSRec) perform better than most models that fail to consider intent, such as SR-GNN, and SHARE. Among them, Atten-Mixer and MiaSRec consider multiple user intentions, leading to significantly improved per-

Method	Yoochoose				Jdata				Diginetica			
	R@5	N@5	R@10	N@10	R@5	N@5	R@10	N@10	R@5	N@5	R@10	N@10
BPR	20.36	30.40	30.40	16.69	19.15	12.34	28.14	15.24	18.94	12.26	27.89	15.15
NARM	68.11	60.59	72.35	61.97	32.41	25.18	39.73	27.55	57.46	72.04	77.44	72.68
SASRec	56.10	43.84	67.03	47.38	25.64	18.24	34.53	21.10	58.76	45.71	69.23	49.12
SR-GNN	72.93	60.54	74.46	61.03	41.78	27.81	46.13	29.22	77.14	72.22	77.78	72.42
GCE-GNN	69.87	60.04	74.97	61.71	33.87	23.47	43.71	26.67	76.59	71.77	78.49	72.39
SHARE	72.99	61.19	76.91	61.73	37.15	26.91	44.41	27.88	77.14	70.41	79.03	70.67
DIDN	70.54	60.03	75.81	61.73	36.77	26.16	46.64	29.37	76.34	69.47	78.29	70.43
Atten-Mixer	73.51	64.83	<u>77.57</u>	66.23	40.27	32.09	47.87	34.55	77.77	72.54	<u>79.32</u>	73.05
MiaSRec	<u>74.60</u>	<u>67.64</u>	<u>77.27</u>	<u>67.88</u>	<u>43.53</u>	<u>36.40</u>	<u>50.18</u>	<u>38.29</u>	<u>78.09</u>	<u>74.85</u>	79.31	<u>75.01</u>
DFRec	75.16	69.48	77.85	69.55	44.60	37.58	50.71	39.56	78.18	75.26	79.40	75.21

*Bold value indicates the best performance in each column, and the value with an underline is the second-best one.

* It shows that the performance of our work is significantly ($p < 0.05$) better than the second-best one based on a Wilcoxon test.

Table 2: Overall model performance on both datasets, with the metrics of Recall@K and NDCG@K (K=5, 10).

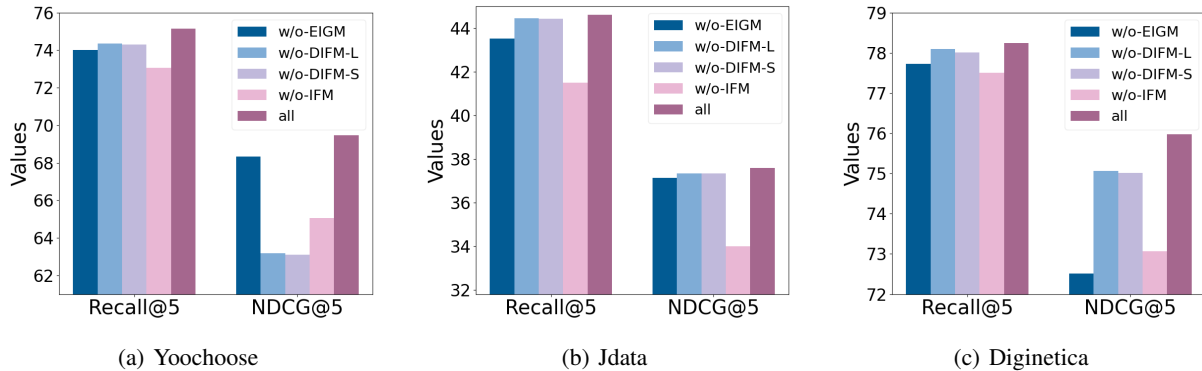


Figure 3: Performance of variants of the proposed DFRec.

formance. However, they focus solely on the diversity of user intentions while overlooking their dynamic evolution. DIDN captures the user’s dynamic intentions by learning dynamic item embeddings that incorporate the temporal order of items within the session. However, this straightforward approach may overlook the complexity of user intention variation.

Compared with them, our DFRec captures multi-level user intentions by generating multiple consecutive item segments of varying lengths. Besides, we introduce intent fluctuation modeling to capture the emerging intention signals between different levels of user intent. Thus, diversity and dynamic variability of intent are simultaneously considered to generate recommendations. Especially by quantifying the amplitude of intent change, it’s more specific for DFRec to make discreet or bold recommendations, thus improving the recommendation performance.

Ablation Study

To evaluate the effectiveness of each module in DFRec, we carry out multiple variants of DFRec. The variants’ performance is shown in Figure 3, where “all” represents DFRec in which all components are considered; the notation “w/o” represents the removal of an element, e.g., w/o-EIGM indicates the emerging intent generation modeling is removed from DFRec. w/o-DIFM-L and w/o-DIFM-S indicate that long-term and short-term fluctuation variances in DIFM are

removed, respectively. Besides, we replace the corresponding continuous item segments with each time step’s item to evaluate the impact of fusing multi-level item segments, denoted as w/o-IFM.

We then have the following observations: 1) “all” outperforms others, showing the effectiveness of its structure. 2) Removing Emerging Intent Generation Module (EIGM) harms the model’s performance, i.e., the performance of “w/o-EIGM” is worse than “all”. It indicates that it is necessary to capture emerging intent signals by modeling user intent fluctuations explicitly. 3) The Recall and NDCG metrics on all datasets degrade regardless of whether short-term or long-term variances are removed. These show that dual variances are valuable signals to model intent fluctuation. 4) Dropping the fused modeling for continuous item segments shows a substantial decline compared to the whole model, reflecting that a set of continuous segments may contain richer information than a single item.

Parameter Analysis

The window size N_w in Eqn. (9) determines how many time-steps of information are considered to capture long-term intent fluctuations of users. To evaluate its influence, we change the size from 1 to 6, and the results are shown in Figure 5. We can observe that the setting $N_w = 2$ for the Yoochoose datasets and dropout $N_w = 3$ for the Jdata datasets are generally proper to DFRec. The Recall and NDCG per-

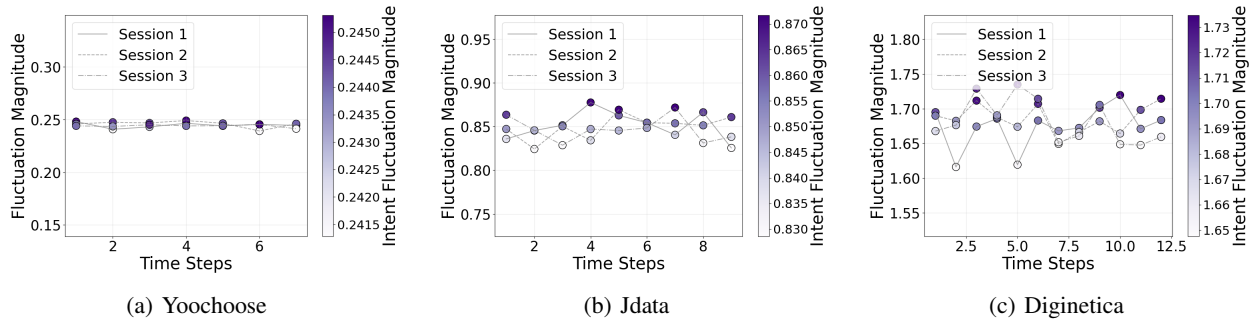


Figure 4: Visualization of user intent fluctuation curves on three datasets. The color scale represents the fluctuation amplitude, with darker shades indicating higher values and lighter shades representing lower amplitudes.

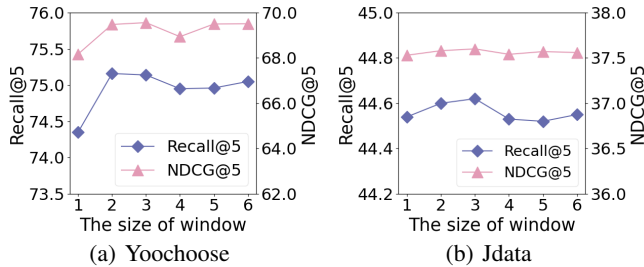


Figure 5: Impact of the size of window size.

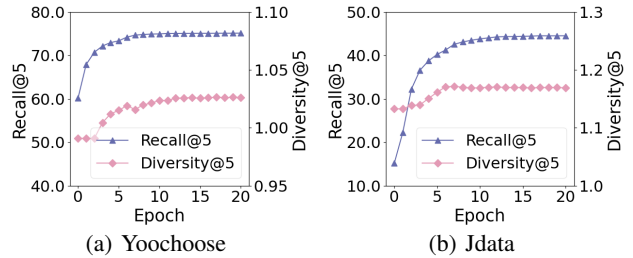


Figure 6: Recommendation Diversity of DFRec.

formance decreases when the window size is set to a small or large value. This is because fewer time steps of interactions are inadequate to capture the user’s long-term information, but overmuch items make it easy to introduce noise.

Recommendation Diversity Evaluation

We evaluate diversity metrics to verify the effectiveness of DFRec in capturing emerging intent signals. Following (Wang et al. 2019), we quantify diversity by utilizing the category distribution of recommendations, measured using information entropy.

$$Diversity@k = - \sum_{i=1}^k Pr_i \log_2 Pr_i \quad (19)$$

while Pr_i is the probability of the category to which item e_i belongs and is estimated by the frequency. Figure 6 illustrates that the Diversity and Recall metrics increase on the Yoochoose and Jdata datasets as training progresses. This indicates the effectiveness of our model in capturing users’ emerging intent signals through intent fluctuation modeling. The results on Diginetica show similar observations, but like Figure 5, the related subfigure is not exhibited to save space.

Case Study

We randomly sample 3 sessions from three datasets to visualize the users’ intent fluctuations learned by DFRec, respectively, as shown in Figure 4. We observe that the intent fluctuations vary between datasets. For Yoochoose, the magnitude of intent fluctuations is minimal, with the curve exhibiting a relatively smooth distribution, indicating stable user intents in this scenario. Conversely, for Diginetica, the

curve shows pronounced fluctuations, suggesting substantial variations in user intents with noticeable differences between consecutive time steps. These observations align with the characteristics of the datasets: Yoochoose has a limited number of categories, 12; Jdata has a moderate number of categories, 79; and Diginetica contains the highest number of categories, 982.

Conclusion

In this paper, we studied intent-modeling methods for the next-item recommendations and claimed that user intent fluctuations embedded in consecutive item segments can serve as the extent of changes in user intents. Accordingly, we proposed a novel framework called DFRec, which generates multiple consecutive item segments and explicitly identifies the magnitude of user intent changes between different segments. Specifically, we hypothesized that a user’s intent fluctuates around an inherent intent, with the magnitude of fluctuations indicating the extent of changes in user intents. Thus, we designed an Emerging Intent Generation Module to capture intent fluctuations at each time step, where larger fluctuations indicate a higher probability of emerging a new intent. Furthermore, we introduced a dual-layer dynamic variance update mechanism to capture fluctuation characteristics at different temporal levels. Finally, extensive experiments on real-world datasets validated the superiority of our model. There are potentially interesting directions for future work. For example, we can incorporate a large language model (LLM) to enhance the modeling of intent fluctuations by considering text attribute changes in sessions.

Acknowledgments

This work is supported by National Natural Science Foundation of China under Grant No. 62202282 and 62406225.

References

- Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan): 993–1022.
- Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3): 307–327.
- Choi, M.; Kim, H.-y.; Cho, H.; and Lee, J. 2024. Multi-intent-aware Session-based Recommendation. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2532–2536.
- Fan, Z.; Liu, Z.; Zhang, J.; Xiong, Y.; Zheng, L.; and Yu, P. S. 2021. Continuous-Time Sequential Recommendation with Temporal Graph Collaborative Transformer. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, 433–442.
- Gao, R.; Tao, Y.; Yu, Y.; Wu, J.; Shao, X.; Li, J.; and Ye, Z. 2023. Self-Supervised Dual Hypergraph Learning with Intent Disentanglement for Session-Based Recommendation. *Knowledge-Based Systems.*, 270(C).
- Guo, J.; Yang, Y.; Song, X.; Zhang, Y.; Wang, Y.; Bai, J.; and Zhang, Y. 2022. Learning Multi-granularity Consecutive User Intent Unit for Session-based Recommendation. In *Proceedings of the International Conference on Web Search and Data Mining*, 343–352.
- Huang, C.; Wang, S.; Wang, X.; and Yao, L. 2023. Modeling Temporal Positive and Negative Excitation for Sequential Recommendation. In *Proceedings of the ACM Web Conference 2023*, 1252–1263.
- Kang, W.-C.; and McAuley, J. 2018. Self-Attentive Sequential Recommendation. In *Proceedings of the IEEE International Conference on Data Mining*, 197–206.
- Li, H.; Wang, X.; Zhang, Z.; Ma, J.; Cui, P.; and Zhu, W. 2022a. Intention-Aware Sequential Recommendation With Structured Intent Transition. *IEEE Transactions on Knowledge and Data Engineering*, 34(11): 5403–5414.
- Li, J.; Ren, P.; Chen, Z.; Ren, Z.; and Ma, J. 2017. Neural Attentive Session-based Recommendation. In *Proceedings of the ACM on Conference on Information and Knowledge Management*, 1419–1428.
- Li, Y.; Gao, C.; Luo, H.; Jin, D.; and Li, Y. 2022b. Enhancing Hypergraph Neural Networks with Intent Disentanglement for Session-based Recommendation. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieve*, 1997–2002.
- Liu, Z.; Li, X.; Fan, Z.; Guo, S.; Achan, K.; and Yu, P. S. 2020. Basket Recommendation with Multi-Intent Translation Graph Neural Network. In *Proceedings of the IEEE International Conference on Big Data*, 728–737.
- Luo, A.; Zhao, P.; Liu, Y.; Zhuang, F.; Wang, D.; Xu, J.; Fang, J.; and Sheng, V. S. 2020. Collaborative Self-Attention Network for Session-based Recommendation. In *Proceedings of the International Joint Conference on Artificial Intelligence*.
- Qin, X.; Yuan, H.; Zhao, P.; Liu, G.; Zhuang, F.; and Sheng, V. S. 2024. Intent Contrastive Learning with Cross Subsequences for Sequential Recommendation. In *Proceedings of the International Conference on Web Search and Data Mining*, 548–556.
- Rendle, S.; Freudenthaler, C.; Gantner, Z.; and Schmidt-Thieme, L. 2012. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the International Conference on Learning Representations*, 452–461.
- Wang, J.; Ding, K.; Zhu, Z.; and Caverlee, J. 2021. Session-based Recommendation with Hypergraph Attention Networks. In *Proceedings of the SIAM International Conference on Data Mining*, 82–90.
- Wang, S.; Hu, L.; Wang, Y.; Sheng, Q. Z.; Orgun, M.; and Cao, L. 2019. Modeling Multi-Purpose Sessions for Next-Item Recommendations via Mixture-Channel Purpose Routing Networks. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 3771–3777.
- Wang, Z.; Wei, W.; Cong, G.; Li, X.-L.; Mao, X.-L.; and Qiu, M. 2020. Global Context Enhanced Graph Neural Networks for Session-based Recommendation. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, 169–178.
- Wu, H.; Zhang, Y.; Ma, C.; Guo, W.; Tang, R.; Liu, X.; and Coates, M. 2023. Intent-aware Multi-source Contrastive Alignment for Tag-enhanced Recommendation. In *Proceedings of the IEEE International Conference on Data Engineering*, 1112–1125.
- Wu, S.; Tang, Y.; Zhu, Y.; Wang, L.; Xie, X.; and Tan, T. 2019. Session-Based Recommendation with Graph Neural Networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 346–353.
- Zhang, P.; Guo, J.; Li, C.; Xie, Y.; Kim, J. B.; Zhang, Y.; Xie, X.; Wang, H.; and Kim, S. 2023. Efficiently Leveraging Multi-level User Intent for Session-based Recommendation via Atten-Mixer Network. In *Proceedings of the International Conference on Web Search and Data Mining*, 168–176.
- Zhang, X.; Lin, H.; Xu, B.; Li, C.; Lin, Y.; Liu, H.; and Ma, F. 2022. Dynamic Intent-Aware Iterative Denoising Network for Session-Based Recommendation. *Information Processing and Management.*, 59(3).
- Zhu, N.; Cao, J.; Liu, Y.; Yang, Y.; Ying, H.; and Xiong, H. 2020. Sequential Modeling of Hierarchical User Intention and Preference for Next-item Recommendation. In *Proceedings of the International Conference on Web Search and Data Mining*, 807–815.
- Zhu, N.; Sun, L.; Luo, X.; Cao, J.; Zhang, Q.; and Lu, X. 2024. Exploitation or Exploration Next? User Behavior Decoupling and Emerging Intent Modeling for Next-Item Recommendation. In *Proceedings of the IEEE International Conference on Data Mining*, 965–970.