ISSN: 1813-4890

Contrastive-Enhanced Hypergraph Neural Network for Session-Based Recommendation via Recency-Intent Alignment

Yifan Huang *

College of Information Technology and Engineering, Tianjin University of Technology and Education, Tianjin 300222, China

Abstract

Existing contrastive learning tasks in session-based recommendation primarily focus on graph structure perturbations, failing to fully leverage the domain prior that "recent interactions better reflect a user's core intent." To address this limitation, this paper proposes a Recency-Enhanced Contrastive Hypergraph Neural Network (RECH-HNN). The model employs a hypergraph convolutional network as its backbone to capture high-order item associations. It further designs a dynamic readout mechanism that fuses content-based attention with a recency prior to generate more precise session intent representations. Crucially, we introduce a novel recency-intent alignment contrastive learning paradigm. By constructing specific augmented views, such as recency-preserving (positive samples) and tail-replacing (hard negative samples), this paradigm guides the model to learn session representations that are robust to early-stage noise yet sensitive to changes in core intent. Experimental results on multiple public datasets demonstrate that the proposed RECH-HNN model consistently outperforms state-of-the-art baselines like S^2-DHCN on key metrics such as P@K and MRR@K, validating the effectiveness of our approach.

Keywords

Session-Based Recommendation; Graph Neural Network; Hypergraph; Contrastive Learning; Recency Preference.

1. Introduction

Session-based recommendation aims to predict a user's next action based on their anonymous and immediate interaction sequence, which is crucial for scenarios involving rapidly shifting user interests or a lack of historical profiles. Research in this field has evolved from early Markov Chain models[1] to Recurrent Neural Network (RNN) models capable of capturing sequential dependencies. However, these models inherently model along a sequential path, making it difficult to effectively characterize the complex non-linear and non-contiguous transition relationships among items within a session.

To overcome these limitations, Graph Neural Networks (GNNs) have been introduced to session-based recommendation[2]. By explicitly constructing session sequences into graph structures, GNNs can capture high-order dependencies between items. Building on this, recent works such as S^2-DHCN[3] have further enhanced model performance by incorporating self-supervised contrastive learning. The core idea of these methods is to construct positive samples by applying perturbations to the graph structure to learn session representations that are robust to topological noise.

Despite significant progress, existing contrastive learning paradigms suffer from a fundamental limitation: their learning objective is set on generic "structural robustness" while overlooking a more critical domain prior in session-based recommendation—the recency of interactions. A large body of research has shown that interactions at the end of a sequence are often the most direct manifestation of a user's current core intent, whereas earlier behaviors may contain considerable exploration or noise. Current contrastive learning methods have not explicitly integrated this prior knowledge into their learning objectives.

Motivated by this insight, this paper aims to deeply integrate the critical domain prior of "recency" into the contrastive learning framework and proposes a new learning objective: recency-intent robustness. An ideal session representation should be insensitive to noise in the early part of a session but highly sensitive to modifications in the tail sequence that alter the core intent. To achieve this, we propose the Recency-Enhanced Contrastive Hypergraph Neural Network (RECH-HNN). This model shifts the goal of contrastive learning from "structural robustness" to "recency-intent robustness." It not only designs a readout mechanism that fuses content and positional priors to accurately locate the core intent but also establishes a novel contrastive learning paradigm. Through specific data augmentation strategies like "recency-preserving" and "tail-replacing," the model is explicitly guided to learn more discriminative session representations, thereby offering a new direction for self-supervised learning research in this domain.

2. Related Work

2.1. Research on Session-Based Recommendation

For general recommendation problems, matrix factorization is a common solution, which decomposes the user-item interaction history into low-dimensional matrices. However, due to the sparsity of the rating matrix, it is not suitable for session-based recommendation. Early research on session-based recommendation was predominantly based on Markov Chain (MC) models. These methods capture short-term item representations within a session but are typically limited in modeling long-term user preference dependencies, as they focus on sequential transitions between adjacent items.

With the rise of deep learning, models based on Recurrent Neural Networks (RNNs) have been widely applied due to their powerful sequence modeling capabilities. GRU4Rec[4] was a pioneering work in this direction, using Gated Recurrent Units (GRUs) to model the sequential behavior of items in a session. Building on this, NARM[5] introduced an attention mechanism, employing a dual-encoder structure to capture both the user's global preferences and current main purpose. STAMP[6] abandoned the recurrent structure of RNNs, relying entirely on self-attention mechanisms and multi-layer perceptrons to capture users' long- and short-term preferences, highlighting the importance of the last click. However, these models still essentially focus on modeling sequential transitions between adjacent items, with limited ability to capture complex, non-contiguous dependencies within a session. Furthermore, Convolutional Neural Network (CNN) models, represented by NextItNet, have also been applied to session-based recommendation, learning short-term sequential patterns by stacking causal convolutional layers, but they also struggle to capture long-range dependencies.

In recent years, to further enhance model performance, contrastive learning has been introduced as a powerful self-supervised learning paradigm. For instance, COTREC designed a graph co-training framework that facilitates information interaction between different views through self-supervised tasks. CL4SRec[7] constructs positive pairs through data augmentation (e.g., item cropping, masking, reordering) to learn more robust sequence representations. More recent work, such as DuoRec[8], improves both generalization and personalization by designing a "general-specific" dual contrastive learning objective. These works have validated the great potential of contrastive learning in alleviating data sparsity and improving representation quality..

2.2. Research on GNN-based Recommendation

To effectively model the complex topological relationships among items within a session, researchers began to construct session sequences into graph structures and use Graph Neural Networks (GNNs) for representation learning, which has become a mainstream direction. SR-GNN is a foundational work in this area, converting session sequences into directed graphs and using Gated Graph Neural Networks (GGNNs) to propagate node information. Subsequently, a series of works have refined and extended the graph construction and information propagation mechanisms. For example, GCE-GNN[9] effectively integrates cross-session global context by constructing a global graph and local session graphs. Disen-GNN proposed a disentangled graph neural network that models diverse user interests by embedding items into multiple independent intent spaces. TAGNN further considered the influence of the target item by introducing a target-aware attention mechanism when learning session representations.

To capture deeper, high-order associations between items, models like S^2-DHCN and DHCN[10] have introduced the hypergraph structure. By treating a session itself as a hyperedge connecting multiple items, hypergraph convolutional networks can more explicitly model the complex co-occurrence relationships of items across different sessions, achieving superior performance. Meanwhile, some of the latest research has started to focus on the dynamic and personalized nature of graph structure learning. For instance, MSG-IF proposed a multi-granularity intent fusion framework that dynamically discovers and fuses users' continuous intent units on a hypergraph. GNN-LSTP explicitly decomposes sessions into long- and short-term preferences and introduces relevance encoding on a global graph.

In the combination of self-supervised learning and GNNs, S^2-DHCN is a landmark work that integrates self-supervised contrastive learning with a hypergraph network. Its core idea is to construct positive samples by applying perturbations to the hypergraph structure (e.g., node/edge dropping) to learn session representations robust to topological noise. Following this line of thought, SG-GNN proposed a multi-task learning framework that simultaneously performs next-item prediction and graph structure reconstruction. However, despite the significant success of these GNN-based and contrastive learning methods, the design of their contrastive tasks still has a common limitation: their objective is set on generic "structural robustness," ignoring a more critical domain prior in session-based recommendation—the recency of interactions. In summary, although existing works have made considerable progress, how to deeply integrate the key domain prior of "recency" into the contrastive learning framework to design more targeted self-supervised tasks remains an issue to be further explored. The RECH-HNN model proposed in this paper is intended to fill this research gap.

3. Model Framework and Algorithm Description

3.1. Problem Definition

Let $V=\{v_1,v_2,\ldots,v_m\}$ be the set of all items, where m is the total number of items. An anonymous user session sequence can be represented as $s=\{v_{s_1},v_{s_2},\ldots,v_{s_n}\}$, where $v_{s_i}\in V$ is the item with which the user interacted at the i -th step in session s ,and n is the session length. The goal of session-based recommendation is to predict the next item $v_{s_{n+1}}$ that the user is most likely to interact with, given the current session s .The model needs to generate a prediction score vector $\hat{y}=y_1,y_2,\ldots,y_m$ for all candidate items V and perform Top-K recommendation based on it.

ISSN: 1813-4890

3.2. Overall Model Framework

The proposed RECH-HNN model aims to learn more discriminative session representations by deeply integrating the "recency" prior into the contrastive learning framework. Its architecture consists of three core components: Global Hypergraph Encoding, Recency-Enhanced Dynamic Readout, and Recency-Intent Contrastive Learning, as shown in Figure 1.

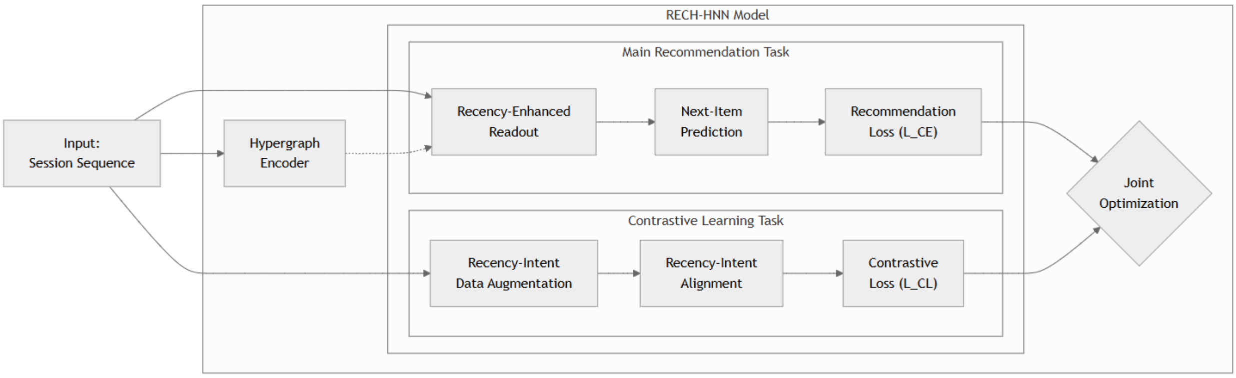

Fig. 1 RECH-HNN model

3.3. Global Hypergraph Construction and Encoding

To capture the complex high-order associations among items, we construct a global hypergraph $G=(V,E)$,where V is the set of nodes (all items) and E is the set of hyperedges (all sessions). Then, we derive item-item adjacency relationships from the hypergraph structure and use a hypergraph convolutional network for information propagation.

Specifically, the structure of the hypergraph can be represented by an incidence matrix $H \in R^{m\times|S|}$,where $|S|$ is the total number of sessions. If item v_i belongs to session s_j,then $H_{ij}=1$;otherwise, $H_{ij}=0$.The propagation process of hypergraph convolution can be formally defined as:

$$\mathbf{X}^{(l+1)} = \sigma\left(\mathbf{D}_v^{-1/2}\mathbf{H}\mathbf{W}_e\mathbf{D}_e^{-1}\mathbf{H}^{\cdot}\ \mathbf{D}_v^{-1/2}\mathbf{X}^{(l)}\right)\mathbf{W}^{(l)} \tag{1}$$

where $\mathbf{X}^{(l)} \in R^{m\times d}$ is the learned item embedding matrix at layer l , d is the embedding dimension, and $\mathbf{X}^{(0)}$ is the randomly initialized item embeddings. $\mathbf{D}_v$ and $\mathbf{D}_e$ are the diagonal matrices of node degrees and hyperedge degrees, respectively, used for normalization. $\mathbf{W}_e$ is a diagonal matrix representing the weights of hyperedges, usually set as an identity matrix. $\mathbf{W}^{(l)} \in R^{d\times d}$ is the trainable weight matrix at the l-th layer, and σ is a non-linear activation function such as LeakyReLU. By stacking L layers of hypergraph convolutions, the model can aggregate information from L-hop neighbors, thereby learning the final item embedding representation $\mathbf{H}_{item} \in R^{m\times d}$ that captures global high-order relationships.

3.4. Recency-Enhanced Dynamic Readout Module

After obtaining item embeddings that capture high-order relationships, the key is to generate an accurate session representation from the corresponding item embedding matrix $\mathbf{H}_s \in R^{m\times d}$ for the current session s.To this end, we design a dynamic readout module that fuses content-based attention with a recency prior.

First, we calculate the content relevance score for each item. This score is determined by the item's own information, its position, and the overall session context. For the i-th item in the session, its information vector n_{s_i} is calculated as follows:

$$\mathbf{n}_{s_i} = \mathrm{sigmoid}(\mathrm{GLU}_1(\tanh(\mathbf{W}_1[\mathbf{p}_i;\mathbf{h}_{s_i}])) + GLU2(\bar{h}_s)) \tag{2}$$

where $\mathbf{h}_{s_i}$ is the i -th row of $\mathbf{H}_S$,i.e.,the embedding of item v_{s_i} , P_i is its corresponding learnable position embedding, and h_s is the session context representation obtained by average pooling all row vectors of $\mathbf{H}_S$. $\left[\mathbf{p}_i;\mathbf{h}_{s_i}\right]$ denotes vector concatenation. $GLU(\cdot)$ represents a Gated Linear Unit, and W_1 is a trainable parameter matrix. Subsequently, the content attention score α'_{s_i} is obtained through a dot product with a trainable vector w_2 :

$$\alpha'_{s_i} = \mathbf{w}_2^{\cdot}\mathbf{n}_{s_i} \tag{3}$$

Next, we introduce the recency prior. Based on the number of steps $\delta_i = n - i$ from the end of the sequence for item i ,we define its recency prior score α''_i as:

$$\alpha''_i = -\gamma \cdot \delta_i \tag{4}$$

where γ is a hyperparameter that controls the strength of recency. The final attention weight α_i is obtained by summing the content attention score and the recency prior score, followed by normalization with a softmax function. This is probabilistically equivalent to a Product of Experts, ensuring that only items important in both content and position receive high weights.

$$\alpha_i = \mathrm{softmax}\left(\alpha'_i + \alpha''_i\right) \tag{5}$$

Finally, the session representation s_{final} is obtained by a weighted sum of all item embeddings within the session according to these attention weights:

$$\mathbf{s}_{\mathrm{final}} = \sum_{i=1}^{n} \alpha_i \mathbf{h}_i \tag{6}$$

3.5. Recency-Intent Contrastive Learning

To guide the model to learn representations with recency-intent robustness, we design a novel contrastive learning task. Unlike traditional methods, our data augmentation strategy does not focus on generic structural perturbations but is closely centered around the core concept of "recency intent." For the anchor representation s_{anchor} generated from each session, we construct specific positive and negative samples through particular transformations to create its positive and negative views.

The augmentation strategies are divided into two categories. The first category aims to construct positive samples that preserve the core intent. We design two methods for this: one is recency-preserving augmentation, which simulates scenarios where users skip exploratory behaviors by randomly dropping some items from the early part of the session; the other is time-warping augmentation, which enhances the model's robustness to minor fluctuations in user interaction rhythm by applying a small amount of Gaussian noise to the step count δ_i when calculating the recency prior. The second category of strategies aims to construct hard negative samples that can confuse the model. For this, we design tail-replacing augmentation, which randomly replaces key items at the end of the session sequence. This operation, while minimizing sequence changes, is highly likely to completely alter the user's core intent, thereby forcing the model to learn more discriminative features.

Based on these augmentation strategies, we use a multi-positive InfoNCE loss to optimize the model. For an anchor s_a ,its contrastive loss L_{CL} is defined as:

$$L_{\mathrm{CL}}\left(\mathbf{s}_a\right) = -log\frac{\sum_{\mathbf{s}_p \in S_{pos}} exp\left(\mathrm{sim}\left(\mathbf{s}_a,\mathbf{s}_p\right)/\tau\right)}{\sum_{\mathbf{s}_p \in S_{pos}} exp\left(\mathrm{sim}\left(\mathbf{s}_a,\mathbf{s}_p\right)/\tau\right) + \sum_{\mathbf{s}_n \in S_{neg}} exp\left(\mathrm{sim}\left(\mathbf{s}_a,\mathbf{s}_n\right)/\tau\right)} \tag{7}$$

where $sim(\cdot,\cdot)$ denotes cosine similarity. S_{pos} is the set of positive samples generated by the two positive augmentation strategies mentioned above. S_{neg} includes hard negative samples generated by tail-replacing, as well as anchor representations from other sessions within the same batch (in-batch negatives), which serve as global negative samples. τ is a temperature hyperparameter that adjusts the difficulty of distinguishing between samples. Through this loss function, the model is driven to pull the anchor closer to positive samples with consistent core intent, while pushing it away from negative samples with conflicting intent, especially those hard negatives that differ only subtly at the tail.

3.6. Joint Optimization and Prediction

After obtaining the final session representation s_{final},we calculate the prediction score for each item by taking the inner product with all item embeddings H_{item}.The result is normalized using a softmax function to obtain the predicted probability distribution y:

$$\mathbf{y} = \mathrm{softmax}\left(\mathbf{s}_{\mathrm{final}} \cdot \mathbf{H}_{\mathrm{item}}^{\cdot}\right) \tag{8}$$

The loss function for the main recommendation task is the standard cross-entropy loss L_{CE}:

$$L_{\mathrm{CE}} = -\sum_{i=1}^{m} y_i log\left(\hat{y}_i\right) \tag{9}$$

where y is the one-hot encoded vector of the target item. The total loss function L_{total} of the model is a weighted sum of the cross-entropy loss L_{CE} and our proposed recency-intent contrastive loss L_{CL}:

$$L_{\mathrm{total}} = L_{\mathrm{CE}} + \beta \cdot L_{\mathrm{CL}} \tag{10}$$

where β is a hyperparameter that balances the importance of the main task and the self-supervised task. By minimizing the total loss L_{total},the model is jointly optimized to simultaneously achieve accurate recommendations and learn session representations with recency-intent robustness.

4. Experiments

4.1. Experimental Datasets

To comprehensively evaluate the effectiveness of the proposed RECH-HNN model, we conducted extensive experiments on three public datasets widely used in session-based recommendation research: Diginetica, Tmall, and Yoochoose 1/64. The Diginetica dataset originates from the 2016 CIKM Cup Personalized E-commerce Search Challenge. The Tmall dataset is extracted from the IJCAI-15 competition and consists of user shopping logs from Alibaba's Tmall e-commerce platform. The Yoochoose dataset is from the 2015 RecSys Challenge and contains a large volume of user clickstream data from e-commerce websites; we used its commonly adopted 1/64 sampled version. To ensure data quality and make the experiments more realistic, we performed necessary preprocessing on the raw data. Specifically, we first filtered out sessions with a length of 1, as such sessions do not provide effective sequential information. Simultaneously, to alleviate data sparsity, we also removed infrequent items that appeared fewer than 5 times in the dataset. After this preprocessing, the detailed statistics of the three datasets are shown in Table 1.

Table 1 Dataset Statistics

Dataset	YOOCHOOSE1/64	Tmall	Diginetica
Clicks	557248	818479	982961
Training Sessions	369859	351268	719470
Test Sessions	55898	25898	60858
Items	16766	40728	43097
Avg. Session Length	6.16	6.69	5.12

4.2. Evaluation Metrics

To quantitatively evaluate the recommendation performance of the models, we adopted two widely used ranking-based evaluation metrics in the field of session-based recommendation: Precision (P) and Mean Reciprocal Rank (MRR). For a given session, the model generates a ranked list of all candidate items. We select the top K items as the final recommendation result and evaluate the model's performance accordingly. Following the common settings in related works, we set K to 20 in our experiments.

P@K measures the proportion of sessions where the true target item is successfully hit within the Top-K recommendation list. Its formula is:

$$\mathrm{P@K} = \frac{n_{\mathrm{hit}}}{|S_{\mathrm{test}}|} \tag{11}$$

where S_{test} is the total number of sessions in the test set, and n_{hit} is the number of sessions where the target item was correctly predicted in the Top-K list.

MRR@K not only considers whether the item was hit but also accounts for the position of the correctly recommended item in the ranked list. The higher the rank, the higher the value of this metric. Its formula is defined as:

$$\mathrm{MRR@K} = \frac{1}{|S_{\mathrm{test}}|} \sum_{s \in S_{\mathrm{test}}} \frac{1}{\mathrm{rank}_s} \tag{12}$$

where $rank_s$ represents the rank position of the true target item in the recommendation list for session s. If the target item does not appear in the Top-K list, its reciprocal rank is counted as 0.

4.3. Overall Performance Comparison

The experimental results in Table 2 show that RECH-HNN achieves the best performance on all three datasets across all evaluation metrics, validating the effectiveness of the proposed method.

Table 2 Compare the experimental results

Models	YOOCHOOSE1/64		Tmall		Diginetica	
	P@20	MRR@20	P@20	MRR@20	P@20	MRR@20
FPMC	45.62	15.01	16.06	7.32	26.53	6.95
GRU4Rec	60.64	22.89	10.93	5.89	29.45	8.33
NARM	68.32	28.63	23.30	10.70	49.70	16.17
STAMP	68.74	29.67	26.47	13.36	45.64	14.32

SR-GNN	70.57	30.94	29.46	13.96	50.73	17.59
DISEN-GNN	71.83	31.25	31.58	15.46	53.70	19.05
GCE-GNN	71.85	32.04	34.35	15.91	54.64	19.20
S^2-DHCN	72.03	32.10	31.41	15.05	53.18	18.43
HyperS^2Rec	72.29	32.30	29.76	16.19	53.29	18.30
Atten-Mixer	72.27	34.55	40.98	16.72	55.12	18.90
RECH-HNN	72.73	34.67	41.70	18.47	55.95	19.53

Compared to the traditional model FPMC, all deep learning-based models show a significant performance advantage, indicating that simple Markov assumptions are insufficient to capture the complex non-linear dependencies in sessions. Among the RNN-based models, NARM and STAMP, by introducing attention mechanisms, perform far better than the basic GRU4Rec, which verifies the necessity of differentiating the importance of items within a session. It is noteworthy that GRU4Rec's performance on the Tmall dataset is even worse than FPMC's, possibly because the user behavior patterns in the Tmall dataset are more complex and volatile, and a single recurrent structure is 反而 constrained when handling such highly dynamic interest drift.

All GNN-based methods outperform RNN-based models in most cases. This result clearly demonstrates that by explicitly constructing session sequences into graph structures, GNNs can capture complex, non-contiguous transition relationships between items that are difficult for RNNs to model. Among these GNN models, GCE-GNN and S^2-DHCN further improve performance by introducing global information or hypergraph structures, illustrating the importance of capturing higher-order, more global item associations.

Most critically, the proposed RECH-HNN model surpasses all strong baseline models, including S^2-DHCN and Atten-Mixer, on all datasets. The core reason for this performance improvement is that RECH-HNN's contrastive learning objective achieves a deeper alignment with the intrinsic characteristics of the session-based recommendation task.

Specifically, on the Yoochoose 1/64 and Diginetica datasets, which have significant clickstream features, user behavior sequences often contain a large amount of exploratory clicks (noise) and a final, clear purchase intent. Models like S^2-DHCN learn "structural robustness" by perturbing the graph structure. While effective, this does not directly address the pain point of "exploratory noise." RECH-HNN, through data augmentation strategies like "recency-preserving" and "tail-replacing," constructs a specialized self-supervised task that forces the model to understand "what recent items determine the essence of a session." This mechanism enables the model to more effectively distill the user's core purchase intent from sequences containing a large amount of noise, thus achieving significant improvements on metrics like MRR@20 that focus more on ranking quality.

On the Tmall dataset, which has longer average session lengths and a more complex commercial setting, the phenomenon of user interest drift may be more frequent. The Atten-Mixer model performs exceptionally well on this dataset, thanks to its powerful sequential feature mixing capabilities. However, RECH-HNN still achieves superior performance. The reason is that RECH-HNN's "recency-intent alignment" not only suppresses early noise, but its "time-warping" augmentation strategy also makes the model more robust to minor fluctuations in user interaction rhythm. This allows the model to stably focus on the final, most critical interaction behaviors even when dealing with multiple intent shifts that may exist in long sessions, thus demonstrating excellent performance on the complex Tmall dataset as well. In summary, the superiority of RECH-HNN is not coincidental but a direct reflection of its core design philosophy effectively playing out on datasets with different characteristics.

4.4. Ablation Study

To verify the effectiveness of each core component in the proposed RECH-HNN model, we conducted a series of ablation studies. Specifically, we sequentially removed the recency prior injection (w/o RecencyPrior), the hard negative samples (w/o HardNeg), and the entire recency-intent contrastive learning framework (w/o CL) from the complete RECH-HNN model and analyzed the performance changes. The experiments were conducted on all three datasets, with results shown in Figure 2 and Figure 3.

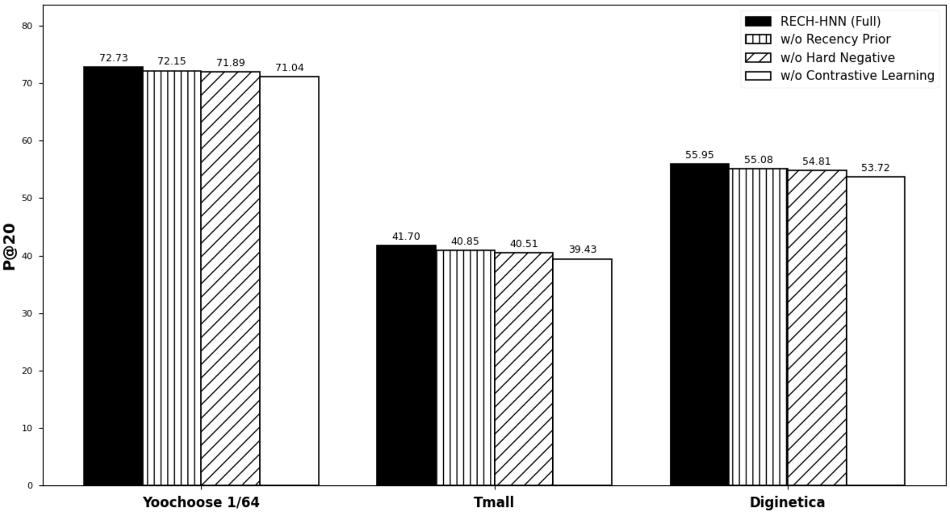

Fig. 2 Recall@20 result when different modules are removed

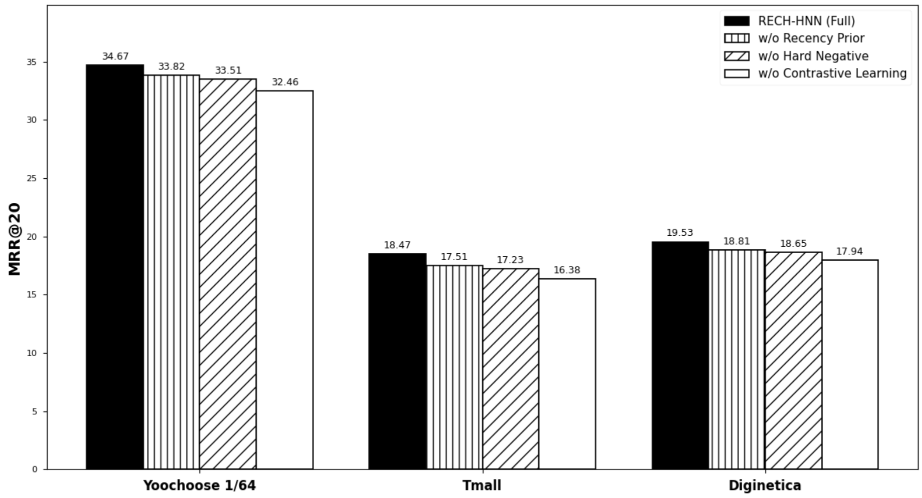

Fig. 3 Recall@20 result when different modules are removed

The results show that removing any core component leads to varying degrees of performance degradation on P@20 and MRR@20 metrics. This clearly demonstrates that each module in the model plays an indispensable role in improving recommendation performance. The specific analysis is as follows:

Impact of the Recency-Intent Contrastive Learning Framework (w/o CL): After removing the entire contrastive learning framework, the model's performance experienced the most significant decline across all datasets. For example, on the Tmall dataset, the MRR@20 metric dropped by about 11.3%. This fully demonstrates that contrastive learning, as a powerful self-supervised learning paradigm, can effectively learn more robust and generalizable session representations by constructing an auxiliary self-supervised task, thereby significantly alleviating the data sparsity problem and providing a crucial regularization signal for the optimization of the main recommendation task.

Impact of the Hard Negative Sample Strategy (w/o HardNeg): When the contrastive learning framework was retained but only the hard negative samples were removed, the model's performance also decreased noticeably. This directly validates the effectiveness and necessity of the "tail-replacing" strategy proposed in this paper. Traditional contrastive learning methods mostly use other sessions within a batch as negative samples. While the model can learn to distinguish between sessions with obvious differences, its ability to discriminate between sessions that are structurally similar but differ subtly in the tail items that determine the core intent is often insufficient. The introduction of hard negative samples, by constructing such "high similarity, low intent relevance" pairs, forces the model to become highly sensitive to tail modifications that are sufficient to change the core intent, thereby obtaining finer and more discriminative session representations.

Impact of the Recency Prior Injection (w/o RecencyPrior): After removing the recency prior injection from the dynamic readout module, the model's performance also showed a consistent decline. This indicates that explicitly incorporating the domain prior that "recent interactions better reflect the user's core intent" into the attention mechanism is highly effective. Without this prior knowledge as "soft guidance," the model has to rely entirely on data-driven content attention to capture the user's main intent, which not only increases the model's learning burden but also makes it more susceptible to interference from noisy items in the early part of the session. The injection of the recency prior can stably guide the model's attention to focus on the tail of the session, which better reflects the user's immediate needs, thus providing a reliable guarantee for generating more accurate session intent representations.

In summary, the results of the ablation study strongly prove the integrity and synergy of the RECH-HNN model design. The recency-enhanced dynamic readout module provides the foundation for accurately locating the core intent, while the novel recency-intent contrastive learning framework, especially the introduction of the hard negative sample strategy, is key to enhancing the discriminative power of the model's representations. The three components are complementary and work together to drive the model to learn session representations that are robust to early noise and sensitive to changes in core intent, thereby achieving excellent recommendation performance.

5. Conclusion

This paper proposes a Recency-Enhanced Contrastive Hypergraph Neural Network (RECH-HNN) for session-based recommendation, designed to address the issue that existing contrastive learning paradigms generally ignore the "recency" domain prior. The model uses a hypergraph convolutional network to capture high-order item associations and designs a dynamic readout mechanism that fuses a recency prior to generate more accurate session intents. Furthermore, by constructing specific augmented views such as "recency-preserving" and "tail-replacing," this paper introduces a novel recency-intent alignment contrastive learning paradigm to learn session representations that are robust to early-stage noise yet sensitive to changes in core intent. Experimental results on multiple public datasets validate the effectiveness of the proposed method.

Future work will explore the following directions: (1)Dynamic Intent Evolution Modeling: Exploring how to model the multiple potential intents a user may have within a long session and their transition processes, rather than just focusing on the final intent. (2)Fusing Multimodal Information: Integrating multimodal attributes of items, such as text and images, into the contrastive learning framework to build richer item representations and explore more challenging hard negative sampling strategies.

References

[1] G. Shani, D. Heckerman and C. Brafman: An MDP-Based Recommender System, Journal of Machine Learning Research, Vol. 6 (2005), p.1265-1295.

[2] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie and T. Tan: Session-Based Recommendation with Graph Neural Networks, Proc. of the AAAI Conference on Artificial Intelligence (Honolulu, HI, USA, January 27 - February 1, 2019), Vol. 33 (2019), p.346-353.

[3] Z. Xia, Z. Li, J. Huang, Z. Yin and G. Li: Self-Supervised Hypergraph Convolutional Networks for Session-Based Recommendation, Proc. of the AAAI Conference on Artificial Intelligence (Virtual, February 22 - March 1, 2022), Vol. 36 (2022) No. 4, p.4871-4879.

[4] B. Hidasi, A. Karatzoglou, L. Baltrunas and D. Tikk: Session-Based Recommendations with Recurrent Neural Networks, Proc. of the International Conference on Learning Representations (ICLR) (San Juan, Puerto Rico, May 2-4, 2016), p.1-10.

[5] J. Li, P. Ren, Z. Chen, Z. Ren, T. Lian and J. Ma: Neural Attentive Session-Based Recommendation, Proc. of the ACM International Conference on Information and Knowledge Management (CIKM) (Singapore, November 6-10, 2017), p.1419-1428.

[6] Q. Liu, Y. Zeng, R. Mokhosi and H. Chen: STAMP: Short-Term Attention/Memory Priority Model for Session-Based Recommendation, Proc. of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (London, United Kingdom, August 19-23, 2018), p.1831-1839.

[7] Z. Xie, T. Wang, J. Zhao, J. Li, F. L. Wang and G. Sun: Contrastive Learning for Sequential Recommendation, Proc. of the IEEE International Conference on Data Engineering (ICDE) (Virtual, April 19-22, 2022), p.125-138.

[8] Y. Qiu, H. Wang, H. Yuan, Z. Li, J. Tang and J. He: Contrastive Learning for Session-Based Recommendation with Universal and Specific Intents, IEEE Transactions on Knowledge and Data Engineering, Vol. 35 (2023) No. 9, p.9419-9431.

[9] C. Wang, Z. Wang, Y. Zhang, Y. Wang, W. Zhang and Y. Wang: Global Context Enhanced Graph Neural Networks for Session-Based Recommendation, Proc. of the ACM SIGIR Conference on Research and Development in Information Retrieval (Virtual, July 25-30, 2020), p.169-178.

[10] Z. Xia, J. Huang, Z. Yin and H. Xu: Session-Based Recommendation with Hypergraph-Infused Knowledge, Proc. of the ACM SIGIR Conference on Research and Development in Information Retrieval (Virtual, July 11-15, 2021), p.1979-1983.