MISRec: Multi-Intention Sequential Recommendation

Rui Chen[1], Dongxue Chen[1], Riwei Lai[1], Hongtao Song[1(✉)], and Yichen Wang[2]

[1] Harbin Engineering University, Harbin, Heilongjiang, China
{ruichen,cdx,lai,songhongtao}@hrbeu.edu.cn
[2] Hunan University, Changsha, Hunan, China
yichenwang.hnu@gmail.com

Abstract. Learning latent user intentions from historical interaction sequences plays a critical role in sequential recommendation. A few recent works have started to recognize that in practice user interaction sequences exhibit multiple user intentions. However, they still suffer from two major limitations: (1) negligence of the dynamic evolution of individual intentions; (2) improper aggregation of multiple intentions. In this paper we propose a novel **M**ulti-**I**ntention **S**equential **Rec**ommender (MISRec) to address these limitations. We first design a multi-intention extraction module to learn multiple intentions from user interaction sequences. Next, we propose a multi-intention evolution module, which consists of an intention-aware remapping layer and an intention-aware evolution layer. The intention-aware remapping layer incorporates position and temporal information to generate multiple intention-aware sequences, and the intention-aware evolution layer is used to learn the dynamic evolution of each intention-aware sequence. Finally, we produce next-item recommendations by identifying the most relevant intention via a multi-intention aggregation module. Extensive experimental results demonstrate that MISRec consistently outperforms a large number of state-of-the-art competitors on three public benchmark datasets.

Keywords: Recommender system · Sequential recommendation · Intention modeling

1 Introduction

In the Internet era, recommender systems have found their way into various business applications, such as e-commerce, online advertising, and social media. Recently, sequential recommendation has emerged as the mainstream approach for next-item recommendation. Learning a user's latent intentions from his/her temporally ordered interactions lies in the core of sequential recommendation.

Supported by the National Key R&D Program of China under Grant No. 2020YFB1710200.

B. Li et al. (Eds.): APWeb-WAIM 2022, LNCS 13423, pp. 191–198, 2023.
https://doi.org/10.1007/978-3-031-25201-3_15

In real-world scenarios, a user normally exhibits multiple intentions in his/her historical interactions. To this end, some very recent studies [4,7,10] have started to explore a user's multiple latent intentions in different ways.

While these studies have confirmed that modeling a user's multiple intentions is a rewarding research direction, we argue that they still suffer from two major limitations. First, they largely neglect the dynamic evolution of individual intentions. While previous studies emphasize the extraction of multiple intentions from user interaction sequences, they overlook the benefits of modeling the dynamic evolution of each individual intention, which is essential for next-item recommendation. Second, modeling a user's intention to interact with an item as a weighted sum of multiple intentions is counter-intuitive. While a user exhibits multiple intentions in his/her historical interaction sequence, the interaction with a particular item is usually driven by a single intention.

In this paper, we propose a novel **M**ulti-**I**ntention **S**equential **Rec**ommender (MISRec) to address these two limitations. We first design a multi-intention extraction module to extract multiple intentions from user interaction sequences. Next, we propose a multi-intention evolution module, consisting of an intention-aware remapping layer and an intention-aware evolution layer. The intention-aware remapping layer incorporates position information and recommendation time intervals to generate multiple intention-aware sequences, where each sequence corresponds to a learned intention. The intention-aware evolution layer is used to learn the dynamic evolution of each intention-aware sequence. Finally, we produce next-item recommendations by explicitly projecting a candidate item into multiple intention subspaces and determining its relevance to each intention. Empowered by Gumbel-softmax, we devise a multi-intention aggregation module to adaptively determine whether each intention is relevant to the target item or not. We perform a comprehensive experimental study on three public benchmark datasets and demonstrate that MISRec consistently outperforms representative state-of-the-art competitors.

2 Related Work

Sequential recommendation has been an emerging paradigm for next-item recommendation. GRU4Rec [1] is the first to employ gated recurrent units (GRUs) to extract sequential patterns from user interaction sequences. Caser [8] considers convolutional neural networks (CNNs) as the backbone network to learn sequential patterns as local features of recent items. NARM [6] uses an attention mechanism to capture more flexible sequential patterns from user interaction sequences. SASRec [2] proposes to leverage self-attention to adaptively consider interacted items. All the above works assume that a user has only a monolithic intention and thus a single embedding representation, which does not reflect the reality well, leaving much room for further improvement. As such, some recent works have started to explore how to better model users using multiple intentions. MCPRN [10] designs a dynamic purpose routing network to capture different user intentions. SINE [7] activates sparse user intentions from a given concept pool and then aggregates the intentions for next-item recommendations.

3 Proposed Method

3.1 Problem Setting

Let $\mathcal{U} = \{u_1, u_2, \cdots, u_{|\mathcal{U}|}\}$ and $\mathcal{I} = \{i_1, i_2, \cdots, i_{|\mathcal{I}|}\}$ be the set of all users and the set of all items, respectively. Given a sequence of user u's historically interacted items $S^u = (s_1^u, s_2^u, \cdots, s_l^u)$ with $s_i^u \in \mathcal{I}$ and the corresponding time sequence $T^u = (t_1^u, t_2^u, \cdots, t_l^u)$ with $t_1^u \leq t_2^u \leq \cdots \leq t_l^u$, the goal of sequential recommendation is to predict the next item with which user u is most likely to interact next. In addition, the recommendation time t is important for recommendation. We transform the interaction time sequence T^u into a new time interval sequence $Tiv^u = (tiv_1^u, tiv_2^u, \cdots, tiv_l^u)$, where $tiv_i^u = \min(t - t_i^u, \tau)$ with τ being a hyperparameter controlling the maximum time interval.

3.2 Embedding Layer

Following previous works, we first transform the user u's interaction sequence $(s_1^u, s_2^u, \cdots, s_l^u)$ into a fixed-length sequence $(s_1^u, s_2^u, \cdots, s_n^u)$, where n denotes the maximum length that our model handles. In the embedding layer, we create an item embedding matrix $\mathbf{E}_i \in \mathbb{R}^{|\mathcal{I}| \times d}$ based on the one-hot encodings of item IDs, where d is the dimension of embedding vectors. Then we retrieve the interaction sequence embedding matrix $\mathbf{E}_{S^u} = \left[e_{s_1^u}, e_{s_2^u}, \cdots, e_{s_n^u}\right] \in \mathbb{R}^{n \times d}$, where $e_{s_i^u}$ is the embedding of item s_i^u in $\mathbf{E}_i$. We also establish two embedding matrices, $\mathbf{E}_P = \left[e_{p_1}, e_{p_2}, \cdots, e_{p_n}\right] \in \mathbb{R}^{n \times d}$ for absolute positions and $\mathbf{E}_{Tiv^u} = \left[e_{tiv_1^u}, e_{tiv_2^u}, \cdots, e_{tiv_n^u}\right] \in \mathbb{R}^{n \times d}$ for relative time intervals.

3.3 Multi-Intention Extraction Module

To capture multiple intentions behind a user's historical interaction sequence, we propose a multi-intention extraction module based on multi-head attention mechanism. Specifically, we map the embedding matrix of a user's interaction sequence $\mathbf{E}_{S^u}$ into different latent subspaces using multiple heads, where each head represents an intention of a user. Let γ be the number of heads and thus the number of intentions. We generate the kth intention m_k^u via

$$m_k^u = head_k^u \mathbf{W}_t, \tag{1}$$

$$head_k^u = Attention(\mathbf{E}_{S^u}\mathbf{W}_k^Q, \mathbf{E}_{S^u}\mathbf{W}_k^K, \mathbf{E}_{S^u}\mathbf{W}_k^V), \tag{2}$$

where $head_k^u \in \mathbb{R}^{1 \times \frac{d}{\gamma}}$ is the output of kth head through a multi-head attention layer. Note that, to match the dimension of item embeddings, a transformation matrix $\mathbf{W}_t \in \mathbb{R}^{\frac{d}{\gamma} \times d}$ is proposed to transform $head_k^u$ from $\mathbb{R}^{1 \times \frac{d}{\gamma}}$ to $\mathbb{R}^{1 \times d}$. $Attention(\cdot)$ is an attention function, and $\mathbf{W}_k^Q$, $\mathbf{W}_k^K$, and $\mathbf{W}_k^V \in \mathbb{R}^{d \times \frac{d}{\gamma}}$ are the trainable transformation matrices of the kth head's query, key and value, respectively. Inspired by previous works [9], we adopt scaled dot-product as the attention function:

$$Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = softmax(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}})\mathbf{V}. \tag{3}$$

After the multi-intention extraction module, we obtain a user u's γ intentions, denoted by $(m_1^u, m_2^u, \cdots, m_\gamma^u)$.

3.4 Multi-Intention Evolution Module

With the extracted multiple intentions from the multi-intention extraction module, we next capture the dynamic evolution of each intention via a multi-intention evolution module, which consists of an intention-aware remapping layer and an intention-aware evolution layer.

Intention-Aware Remapping Layer. Simply capturing the sequential patterns on the learned intentions lacks guarantees to model the dynamic evolution of user intentions precisely [5]. Therefore, we first design an intention-aware remapping layer to explicitly inject sequentiality and temporal information into intention-aware interaction sequences. In particular, we devise an extended scaled dot-product attention mechanism, where the learned intentions play the role of query vectors, and the key and value of the scaled dot-product attention are the interaction sequence injected with positional and temporal information:

$$(\mathbf{Key} : \mathbf{Value}) : (\mathbf{E}_{S^u}\mathbf{W}_S^K + \mathbf{E}_P\mathbf{W}_P^K + \mathbf{E}_{Tiv^u}\mathbf{W}_T^K : \mathbf{E}_{S^u}\mathbf{W}_S^V + \mathbf{E}_P\mathbf{W}_P^V + \mathbf{E}_{Tiv^u}\mathbf{W}_T^V), \tag{4}$$

where $\mathbf{E}_{S^u}$, $\mathbf{E}_P$, $\mathbf{E}_{Tiv^u} \in \mathbb{R}^{n\times d}$ are the embedding matrices of the interaction sequence, position sequence and time interval sequence, respectively. $\mathbf{W}^K$ and $\mathbf{W}^V \in \mathbb{R}^{d\times d}$ are the trainable matrices for keys and values, where the subscripts S, P and T indicate the matrices for the interaction sequence, position sequence and time interval sequence, respectively. Then we compute a new intention-aware interaction sequence $\mathbf{S}_k^u = (s_{k1}^u, s_{k2}^u, \cdots, s_{kn}^u)$ via

$$\mathbf{S}_k^u = softmax\left(\frac{(m_k^u\mathbf{W}_{S_k}^Q)\mathbf{Key}^\top}{\sqrt{d}}\right)\mathbf{Value}, \tag{5}$$

where $\mathbf{W}_{S_k}^Q \in \mathbb{R}^{d\times d}$ is the trainable matrix for intention m_k^u.

Intention-Aware Evolution Layer. To capture the dynamic evolution of each intention, we employ gated recurrent units (GRUs) to model the dependencies between interacted items under each individual intention. Specifically, the input to the GRU for the kth intention is the kth intention-aware interaction sequence $\mathbf{S}_k^u$. We utilize the last hidden state h_k^u of the GRU to represent the user u under the kth intention. We further adopt a point-wise feedforward network (FFN) to endow the model with non-linearity and consider interactions between different latent dimensions:

$$m_k^u = h_k^u + Dropout(FFN(LayerNorm(h_k^u))), \tag{6}$$

$$LayerNorm(x) = \alpha \odot \frac{x-\mu}{\sqrt{\sigma^2+\epsilon}} + \beta, \tag{7}$$

$$FFN(x) = ReLU(x\mathbf{W}_1 + b_1)\mathbf{W}_2 + b_2, \tag{8}$$

where $\mathbf{W}_1, \mathbf{W}_2 \in \mathbb{R}^{d \times d}$ are learnable matrices, and b_1, b_2 are d-dimensional bias vectors. μ and σ are the mean and variance of x, α and β are the learned scaling factor and bias term, respectively. We apply layer normalization to the input h_k^u before feeding it into the FFN, apply dropout to the FFN's output, and add the input h_k^u to the final output.

3.5 Multi-Intention Aggregation Module

Intuitively, a user's interaction with an item is usually driven by a single intention. Directly combining the multiple intention representations as the final intention representation is counter-intuitive and cannot maximize the benefits of extracting multiple intentions. In addressing this issue, we adopt the Gumbel-softmax to adaptively determine whether an intention is relevant to the candidate item or not. Specifically, we first explicitly project the candidate item's embedding e_{n+1} into different intention subspaces (see Eq. 9), and then calculate the relevance between each intention representation and the candidate item's embedding in each intention subspace via the inner product operation (see Eq. 10). Distinct from the previous methods using softmax to aggregate the multiple intention representations, we adopt the Gumbel-softmax to identify the most relevant intention (see Eqs. 11 and 12). Finally, we obtain the final representation m^u of user u at the finer granularity of intentions.

$$e_{n+1}^k = e_{n+1}\mathbf{W}^{\mathbf{k}}, \tag{9}$$

$$r_{n+1}^k = e_{n+1}^k {m_k^u}^\top, \tag{10}$$

$$a_k = \frac{\exp((\log(r_{n+1}^k) + g_i)/\tau)}{\sum_{j=1}^{\gamma} \exp((\log(r_{n+1}^j) + g_j)/\tau)}, \tag{11}$$

$$m^u = \sum_{k=1}^{\gamma} a_k * m_k^u. \tag{12}$$

3.6 Model Training

After we get the final representation m_u of user u, prediction scores are calculated as the inner product of the final user representation m_u and the candidate item's embedding e_i:

$$r_{u,i} = e_i m_u^\top. \tag{13}$$

We use the pairwise Bayesian personalized ranking (BPR) loss to optimize the model parameters. Specifically, it encourages the predicted scores of a user's historical items to be higher than those of unobserved items:

$$\mathcal{L}_{\mathcal{BPR}} = \sum_{(u,i,j) \in \mathcal{O}} -\ln \sigma(r_{u,i} - r_{u,j}) + \lambda \|\boldsymbol{\Theta}\|_2^2, \tag{14}$$

where $\mathcal{O} = \{(u,i,j) | (u,i) \in \mathcal{O}^+, (u,j) \in \mathcal{O}^-\}$ denotes the training dataset consisting of the observed interactions $\mathcal{O}^+$ and sampled unobserved items $\mathcal{O}^-$, $\sigma(\cdot)$ is the sigmoid activation function, $\boldsymbol{\Theta}$ is the set of embedding matrices, and λ is the L_2 regularization parameter.

4 Experiments

4.1 Experimental Setup

Datasets and Evaluation Metrics. We evaluate our framework on three public benchmark datasets that are widely used in the literature. **Amazon-Review** datasets[1] contain product reviews from the online shopping platform Amazon, and we use two representative datasets, Grocery and Gourmet Food (referred to as **Grocery** and **Beauty**). **MovieLens**[2] datasets contain a collection of movie ratings from the website MovieLens. and we use MovieLens-1M (referred to as **ML1M**) in our experiments. Following previous works [2,5], we filter out cold-start users and items with fewer than 5 interactions and sort the interactions of each user by timestamps. Similarly, we use the most recent item for testing, the second most recent item for validation, and the remaining items for training. We evaluate our framework by two widely-adopted ranking metrics, Hit Ratio@N (**HR@N**) and Normalized Discounted Cumulative Gain@N (**NDCG@N**).

Baselines. To demonstrate the effectiveness of our solution, we compare it with a wide range of representative sequential recommenders, including four single-intention-aware methods (**GRU4Rec** [1], **NARM** [6], **Caser** [8], and **TiSASRec** [5]) and a multi-intention-aware method, **SINE** [7].

Implementation Details. Identical to the settings of previous methods, the embedding size is fixed to 64. We optimize our method with Adam [3] and set the learning rate of Grocery, Beauty, and ML1M to 10^{-4}, 10^{-3} and 10^{-4}, respectively, and the mini-batch size to 256 for all three datasets. The maximum length of interaction sequences of Grocery, Beauty, and ML1M is set to 10, 20, and 50, respectively. The maximum time interval is set to 512 sec for all three datasets. The temperature parameter τ in the Gumbel-softmax is set to 0.1. To address overfitting, we use L_2 regularization with the regularization coefficients of 10^{-5} for Grocery and ML1M and 10^{-4} for Beauty.

4.2 Main Results

Overall Comparison. We report the overall comparison in Table 1, where the best results are boldfaced and the second-best and third-best results are underlined. We can draw a few interesting observations: (1) TiSASRec achieves the best performance among single-intention-aware methods, indicating the effectiveness of the self-attention mechanism and temporal information in capturing sequential patterns. However, without considering multiple user intentions, these methods cannot identify a user's true intention accurately, leading to sub-optimal recommendations. (2) As a multi-intention-aware method, SINE performs generally better than most single-intention-aware methods, which shows

[1] http://jmcauley.ucsd.edu/data/amazon/links.html.
[2] https://grouplens.org/datasets/movielens/1m/.

Table 1. Performance of different models on the three datasets. All the numbers in the table are percentages with % omitted.

	Grocery				Beauty				ML1M			
	Metric@10		Metric@20		Metric@10		Metric@20		Metric@10		Metric@20	
	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG	HR	NDCG
GRU4Rec	4.79	2.41	7.89	3.19	3.98	2.09	6.38	2.69	14.17	6.90	23.06	9.13
NARM	6.21	3.21	9.72	4.09	7.28	4.18	10.23	4.92	15.23	7.10	25.66	9.71
Caser	5.65	2.85	9.02	3.69	5.92	3.15	8.91	3.90	15.62	7.48	27.72	11.00
TiSASRec	7.32	3.20	11.00	4.06	8.25	4.23	11.31	5.00	22.37	10.82	33.54	13.64
SINE	6.27	2.96	9.89	3.73	5.84	2.57	8.73	3.30	16.64	7.18	27.91	10.00
MISRec	**7.63**	**3.37**	**11.37**	**4.26**	**8.83**	**4.72**	**12.42**	**5.45**	**22.81**	**11.62**	**34.37**	**14.45**
Improv.	4.23	4.98	3.36	4.16	7.03	11.58	9.81	9.00	1.97	7.39	2.47	5.94

Table 2. Performance of different variants of MISRec. The results of HR@20 and NDCG@20 are omitted due to the space limitation.

	Grocery		Beauty		ML1M	
	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10
w/o PE	7.47	3.30	8.29	4.35	22.50	10.74
w/o TIE	7.46	3.31	8.38	4.49	22.48	10.80
w/o GS	7.59	3.34	8.78	4.68	22.68	11.50
MISRec	**7.63**	**3.37**	**8.83**	**4.72**	**22.81**	**11.62**

that explicitly exploring multiple user intentions is a rewarding direction. However, the performance of SINE is still worse than TiSASRec. We deem that it is caused by the negligence of the dynamic evolution of individual intentions and the improper aggregation of multiple intentions. (3) By addressing the two issues mentioned above, MISRec maximizes the benefits of extracting multiple intentions and consistently yields the best performance on all datasets, which well justifies our motivation.

Ablation Study. To investigate the contributions of different components on the final performance, we conduct an ablation study to compare the performance of different variants of our MISRec model on the three datasets. The variants include: (1) **w/o PE** removes positional embeddings in the multi-intention evolution module. (2) **w/o TIE** removes time interval embeddings in the multi-intention evolution module. (3) **w/o GS** replaces the Gumbel-softmax with the softmax in the multi-intention aggregation module. Table 2 shows the performance of all variants and the full MISRec model on the three datasets. By comparing the performance of different variants, we can derive that both positional embeddings and time interval embeddings lead to performance improvement, which demonstrates the significance of explicitly modeling the dynamic evolution of different intentions. Furthermore, identifying the most relevant intention

rather than aggregating multiple intentions consistently improves the performance by a significant margin, which justifies our motivation.

5 Conclusion

In this paper, we proposed a novel Multi-Intention Sequential Recommender (MISRec) to address the limitations of existing works that leverage users' multiple intentions for better next-item recommendations. We made two major contributions. First, we designed a multi-intention evolution module that effectively models the evolution of each individual intention. Second, we proposed to explicitly identify the most relevant intention rather than aggregate multiple intentions to maximize the benefits of extracting multiple intentions. A comprehensive experimental study on three public benchmark datasets demonstrates the superiority of the MISRec model over a large number of state-of-the-art competitors.

References

1. Hidasi, B., Karatzoglou, A., Baltrunas, L., Tikk, D.: Session-based recommendations with recurrent neural networks. In: Proceedings of the 4th International Conference on Learning Representations (ICLR) (2016)
2. Kang, W., McAuley, J.J.: Self-attentive sequential recommendation. In: Proceedings of the 18th IEEE International Conference on Data Mining (ICDM), pp. 197–206 (2018)
3. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations (ICLR) (2015)
4. Li, C., et al.: Multi-interest network with dynamic routing for recommendation at Tmall. In: Proceedings of the 28th International Conference on Information and Knowledge Management (CIKM), pp. 2615–2623 (2019)
5. Li, J., Wang, Y., McAuley, J.J.: Time interval aware self-attention for sequential recommendation. In: Proceedings of the 13th International Conference on Web Search And Data Mining (WSDM), pp. 322–330 (2020)
6. Li, J., Ren, P., Chen, Z., Ren, Z., Lian, T., Ma, J.: Neural attentive session-based recommendation. In: Proceedings of the 26th International Conference on Information and Knowledge Management (CIKM), pp. 1419–1428 (2017)
7. Tan, Q., et al.: Sparse-interest network for sequential recommendation. In: Proceedings of the 14th International Conference on Web Search And Data Mining (WSDM), pp. 598–606 (2021)
8. Tang, J., Wang, K.: Personalized top-n sequential recommendation via convolutional sequence embedding. In: Proceedings of the 11th International Conference on Web Search And Data Mining (WSDM), pp. 565–573 (2018)
9. Vaswani, A., et al.: Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS), pp. 5998–6008 (2017)
10. Wang, S., Hu, L., Wang, Y., Sheng, Q.Z., Orgun, M.A., Cao, L.: Modeling multi-purpose sessions for next-item recommendations via mixture-channel purpose routing networks. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI), pp. 3771–3777 (2019)