Top-one Recommendation with Anonymous User Behaviors (Student Abstract)

Xiangkui Lu[1], **Jun Wu**[1,2,*]

[1]MoE Key Lab of Big Data & Artificial Intelligence in Transportation, Beijing Jiaotong University, Beijing 100044, China
[2]MoE Engineering Research Center of Integration and Application of Digital Learning Technology, The Open University of China, Beijing 100039, China
{lxkui,wuj}@bjtu.edu.cn

Abstract

Top-one recommendation with anonymous user behaviors, also known as session-based recommendation (SBR), faces challenges of top-one ranking and short anonymous sequences. To this end, we propose a novel objective that combines (i) a reciprocal rank loss to directly optimize the benchmark metric of top-one recommendation, with (ii) a listwise contrastive loss to handle short sequences through listwise augmented consistency regularization. Empirical studies demonstrate that optimizing the proposed objective significantly improves the performance of existing SBR baselines.

Introduction

Due to the prevalence of mobile devices and growing privacy concerns, recommender systems have shifted from top-n recommendation with identifiable user behavior to top-one recommendation with anonymous user behavior, giving rise to Session-Based Recommendation (SBR). SBR faces two key challenges: (i) **Top-one Ranking**. The benchmark metric for top-one task is Mean Reciprocal Rank (MRR), which is notoriously challenging to optimize using gradient descent methods due to vanishing gradients for bottom ranks (e.g., 100, 1000). (ii) **Short Sequences**. Anonymous sequences are typically very short (containing fewer than 5 items), making it difficult to infer user intent accurately.

Since directly optimizing top-one ranking is difficult, existing methods (Li et al. 2017; Wu et al. 2019; Zhang et al. 2023) often use binary cross entropy (BCE) loss as a surrogate to MRR, resulting in sub-optimal recommendations. For short sequences, contrastive learning with sequence augmentation appears feasible. However, augmenting short sequences may lead to large discrepancies, resulting in the assumption of representation consistency in traditional contrastive learning no longer holding.

To address the above challenges, we propose a novel objective, coined **RRLC**, which incorporates (i) a **R**eciprocal **R**ank loss to narrow the gap between the optimization objective and the benchmark metric for top-one ranking, and (ii) a **L**istwise **C**ontrastive loss to promote consistency between two predictive ranked lists generated from the session and its augmentation. Empirical studies show that optimizing RRLC significantly improves the performance of three existing SBR models on two datasets, yielding an average improvement of around 7% in terms of MRR metric.

The Proposed Objective

Let $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$ and $\mathcal{S} = \{s_1, s_2, \dots, s_m\}$ denote the item set and session set, respectively, where n and m are the number of items and sessions. Given a session $s \in \mathcal{S}$, the task of SBR is to predict the next item $v \in \mathcal{V}$ that the user will interact with. To address the challenges in SBR, we propose RRLC objective, which combines a reciprocal rank loss and a listwise contrastive loss.

Reciprocal Rank Loss

MRR metric is calculated by averaging the reciprocal ranks of the ground-truth items across all sessions. The reciprocal rank of the ground truth for session s, $rr(s) = \frac{1}{r_s}$, is inversely proportional to the rank r_s, leading to a flat curve in the bottom ranks (e.g., 100, 1000). Consequently, directly optimizing $rr(s)$ suffers from vanishing gradients, making it challenging for the optimization to converge during the initial training phase when most ground-truth items are ranked at the bottom.

To enable direct optimization of $rr(s)$, we propose to adjust the curve of $rr(s)$ by introducing a flexibility factor σ, which provides an optimizable gradient for bottom ranks while maintaining consistency with the exact reciprocal rank loss (i.e., $1 - rr(s)$) for top ranks. The objective is:

$$\mathcal{L}_{RR} = \frac{1}{(r_s - 1)^{\sigma}}, \sigma \in [-1, 0), \tag{1}$$

where σ controls the shape of the $\mathcal{L}_{RR}$ curve. As illustrated in Figure 1, as σ approaches 0 (e.g., $\sigma = -1/8$), the loss approaches its lower bound, which is also the upper bound of the exact loss (the line in red). As σ decreases, the slope of bottom ranks becomes steeper (e.g., $\sigma = -1/2$), allowing for larger gradients. When $\sigma = -1$, the loss reaches its upper bound, having large gradients but leading to a significant gap with the exact loss. In summary, by adjusting σ, we strike a balance between model optimization and consistency with exact reciprocal rank objective.

[*]Corresponding author.

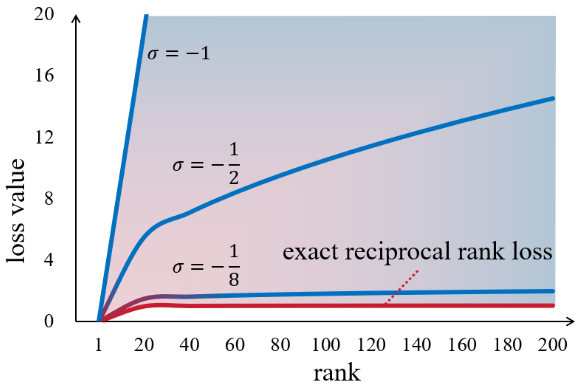

Figure 1: Different approximations to the exact reciprocal rank loss with different σ.

Listwise Contrastive Loss

Listwise contrastive loss involves frequent-pattern-guided augmentation and listwise consistency regularization. First, we mine frequent patterns within $\mathcal{S}$, where each pattern "$v_i \to v_j[c]$" indicates that v_j frequently precedes v_i, with the support value c calculated as the sum of the reciprocal distances between v_i and v_j in each session. All mined patterns form the knowledge base $\mathcal{K}$. Second, each item in s is treated as a query v_q to retrieve its highest-support predecessor v_p from $\mathcal{K}$. If v_q is the first item in s, v_p is directly inserted before v_q. Otherwise, a double check is requires to ensure that v_p precedes v_q and follows the item preceding v_q. This insertion process is iteratively applied to the augmented session s' until a stopping condition (maximum iterations or session length) is reached. Third, by independently feeding s and s' into a recommender (implemented with any existing SBR model), two ranked lists, L_s and $L_{s'}$, are generated. Subsequently, rank-biased overlap (RBO) (Oh et al. 2022) is employed to measure the consistency of the top-K orderings between these lists, defined as the sum of the overlaps between the two lists at each position. Unfortunately, the set union operation used in calculating overlap is non-differentiable and non-decomposable, preventing direct optimization with gradient descent. To this end, we propose an approximation that converts set operation into vector computation. The resulting approximate RBO is:

$$RBO(L_s, L_{s'}, K) \approx \frac{1}{K}\sum_{k=1}^{K}\frac{\mathcal{G}(\hat{\mathbf{y}}_s - \hat{y}_{s,tk};\tau)\cdot\mathcal{G}(\hat{\mathbf{y}}_{s'} - \hat{y}_{s',tk};\tau)}{k}, \tag{2}$$

where $\hat{y}_{s,tk}$ and $\hat{y}_{s',tk}$ are the k-th largest scores in $\hat{\mathbf{y}}_s$ and $\hat{\mathbf{y}}_{s'}$ (the corresponding score vectors of L_s and $L_{s'}$), respectively. $\mathcal{G}(x;\tau)$ is the sigmoid function with temperature τ, and K is the focused K orderings. So far, the listwise contrastive loss in a training batch $\mathcal{B}$ is defined as:

$$\mathcal{L}_{LC} = \sum_{s\in\mathcal{B}} -log\frac{exp(RBO(L_s, L_{s'}, K))}{\sum_{o\in\mathcal{B}} exp(RBO(L_s, L_{o'}, K))}, \tag{3}$$

By combining Eq.(1) and Eq.(3) with hyperparameter α, the total loss is: $\mathcal{L} = \mathcal{L}_{RR} + \alpha\mathcal{L}_{LC}$, which can be optimized using existing optimizers like Adam.

		Diginetica		Tmall	
		HR@5	MRR@5	HR@5	MRR@5
NARM	baseline	26.69	14.97	20.08	13.89
	RRLC	28.14	15.88	22.50	14.97
	Imp.	**5.45%**	**6.08%**	**12.06%**	**7.78%**
SRGNN	baseline	27.20	15.26	18.86	13.12
	RRLC	28.68	16.17	21.40	14.77
	Imp.	**5.46%**	**6.02%**	**13.48%**	**12.56%**
ATTMIX	baseline	27.59	15.52	25.57	17.34
	RRLC	28.17	15.99	27.04	18.23
	Imp.	**2.11%**	**3.03%**	**5.72%**	**5.19%**

Table 1: Performance comparison between each baseline and its RRLC-modified version.

		Diginetica		Tmall	
		HR@5	MRR@5	HR@5	MRR@5
NARM	RRLC	**28.14**	**15.88**	**22.50**	**14.97**
	w/o RR	27.34	15.48	21.02	13.97
	w/o LC	27.79	15.73	20.98	13.99
SRGNN	RRLC	**28.68**	**16.17**	**21.40**	**14.77**
	w/o RR	28.19	16.00	20.09	13.51
	w/o LC	27.94	15.83	21.29	14.77
ATTMIX	RRLC	**28.17**	**15.99**	**27.04**	18.23
	w/o RR	28.10	15.96	26.72	18.14
	w/o LC	28.06	15.90	26.90	**18.26**

Table 2: Performance of RRLC compared to its variants.

Experiments

We conduct empirical studies on the Diginetica and Tmall datasets, select three widely used SBR models (NARM (Li et al. 2017), SRGNN (Wu et al. 2019), and ATTMIX (Zhang et al. 2023)) as baselines, and employ HR (Hit Ratio) and MRR as evaluation metrics.

Overall Performance. We compare the performance of each SBR baseline with its RRLC-modified version. Quantitative results in Table 1 reveal that optimizing RRLC consistently boosts the performance of all baselines across both datasets. Specifically, RRLC achieves an average improvement (Imp.) of 7.38% in HR and 6.78% in MRR, demonstrating the versatility and superiority of RRLC loss.

Ablation Study. To analyze the effectiveness of each part of loss, we compare RRLC to two degenerated variants: one using BCE instead of reciprocal rank loss ('w/o RR') and another omitting listwise contrastive loss ('w/o LC'). As shown in Table 2, RRLC consistently outperforms both variants in most cases, highlighting the effectiveness of reciprocal rank loss for top-one ranking and listwise contrastive loss for addressing short sequences.

Conclusion

We propose a novel objective RRLC to address the challenges of top-one ranking and short sequences in SBR, whose effectiveness has been empirically verified.

Acknowledgements

This work was supported by the Fundamental Research Funds for the Central Universities (2023YJS023), Beijing Natural Science Foundation (L232033), and the Innovation Fund from the Engineering Research Center of Integration and Application of Digital Learning Technology, Ministry of Education (1321004).

References

Li, J.; Ren, P.; Chen, Z.; Ren, Z.; Lian, T.; and Ma, J. 2017. Neural Attentive Session-based Recommendation. In *CIKM*, 1419–1428.

Oh, S.; Ustun, B.; McAuley, J.; and Kumar, S. 2022. Rank List Sensitivity of Recommender Systems to Interaction Perturbations. In *CIKM*, 1584–1594.

Wu, S.; Tang, Y.; Zhu, Y.; Wang, L.; Xie, X.; and Tan, T. 2019. Session-Based Recommendation with Graph Neural Networks. In *AAAI*, 346–353.

Zhang, P.; Guo, J.; Li, C.; Xie, Y.; Kim, J.; 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 *WSDM*, 168–176.