

Adapting Knowledge Graphs for Behavior Denoising in Sequential Recommendation

Zichun Jin^{1,†}, Zihan Zhou^{1,†}, Yinan Liu^{1,*}, Bin Wang¹ and Xiaochun Yang¹

¹Northeastern University, Shenyang, China

Abstract

Sequential recommendation predicts the next item from a user’s interaction history, but not every interaction is equally informative. Real logs combine persistent preferences with temporary needs, exploration, and incidental behavior, so some interactions can distort history representations or provide unreliable supervision. Existing denoising methods judge such interactions mainly from co-occurrence, order, or model predictions, without explicit evidence from relations between items. Knowledge graphs (KGs) offer this evidence, but item popularity, graph degree, uneven coverage, and widely shared entities can inflate connectivity and bias reliability estimates. Here we present ADAPTEKKG, which derives calibrated KG evidence for each training example without adding graph representations to the recommendation model. It first compares the observed context with structurally matched alternatives to identify relational paths that are unusually prominent and uses them to build a local KG view. It then compares each interaction with structurally matched reference items to calibrate its support within that view. The resulting retention coefficients gate historical representations and reweight target losses. All sample-specific scores are computed offline using training interactions and a fixed KG, so the backbone remains unchanged and no KG access is required at inference. Experiments show gains with a standard sequential recommender and multiple behavior-denoising sequential recommenders.

Keywords

Knowledge Graphs, Sequential Recommendation, Behavior Denoising, Structural Matching

1. Introduction

Sequential recommendation relies on a user’s interaction history to predict the next item [1, 2], yet the interactions in this history do not provide equally reliable evidence of user preference [3]. Real-world logs reflect a mixture of long-term preferences, short-term needs, exploration, and incidental behavior, making individual interactions unevenly valuable for sequence modeling [4]. An interaction weakly associated with other interactions in the same sequence may both disrupt historical representations and provide unstable supervision when used as a training target [3]. Existing denoising methods mainly rely on co-occurrence, order, or model predictions to identify such interactions [5, 6, 7], but lack evidence from explicit inter-item relations for making this judgment.

Knowledge graphs encode entities and relations connecting items and therefore offer a potential means of filling this gap [8, 9]. However, richer path connectivity in the graph does not necessarily imply more relevant relations between items [10]. Items with higher popularity, higher KG degree, or more extensive coverage naturally admit more paths, while entities shared by many items can also induce broad connections. Using these connections directly may bias reliability estimation toward richly connected items and common relational patterns. Whether a path provides distinctive support for the current interaction must therefore be assessed against structurally comparable references.

Based on this idea, we propose ADAPTEKKG. Rather than injecting KG representations into the recommendation backbone, it uses two structural-matching stages with distinct roles that are executed sequentially. These stages convert connections in the global KG into KG-consistency evidence for individual behavioral examples and derive conservative retention coefficients to modulate training signals. In the first stage, alternative contexts with similar structural attributes characterize the background distribution of relational paths. The method retains and weights relational patterns that are more



prominent in the current context relative to the matched background, forming an example-specific local KG view. With this view fixed, the second stage uses reference items with similar structural attributes to calibrate the support that these relational patterns provide for the evaluated interaction. The resulting retention coefficients gate historical interaction representations and reweight target losses, respectively. All KG adaptation and sample-specific scoring are performed offline using training interactions and a fixed KG, without changing the backbone architecture or requiring KG access at inference. Experiments show that ADAPTEKKG yields improvements when applied to a standard sequential recommendation model and multiple behavior-denoising models.

2. Related Work

Behavior denoising improves sequential recommendation by correcting unreliable interactions or reducing their influence during training. STEAM [5] learns item-wise keep, delete, and insert operations from synthetically corrupted sequences. BirDRec [6] uses bidirectional recommender predictions to rectify unreliable histories and targets with theoretical error guarantees. SSDRec [7] exploits interaction-derived inter-sequence relations to guide self-augmentation and hierarchical denoising. Despite their different mechanisms, these methods obtain correction evidence from interaction logs, whether through synthetic supervision, recommender predictions, or learned behavioral relations. These log-derived signals do not directly provide external typed relational evidence for assessing an interaction’s contextual support. KG-based recommenders commonly integrate graph-derived representations into prediction models [9]; ADAPTEKKG instead uses structurally matched KG evidence to derive example-conditioned retention coefficients offline, before recommendation optimization.

3. Methodology

Raw KG connectivity is not directly comparable across examples. A context containing popular or high-degree items may contain many common paths, while an easily connected candidate may receive high support for the same reason. ADAPTEKKG handles these effects in sequence: matched-null contexts select paths that stand out from comparable contexts; with those paths fixed, matched references convert support into an interaction-level retention coefficient.

3.1. Example-Conditioned Interaction Reliability

For a training example $\mathcal{D}_s = (H_s, y_s)$, let $H_s = (x_{s,1}, \dots, x_{s,L_s})$ contain only interactions preceding y_s . Let $R(v; C) \in [0, 1]$ be the retention coefficient assigned to v from context C :

$$r_{s,j}^H = R(x_{s,j}; H_s^{-j}), \quad r_s^T = R(y_s; H_s), \quad (1)$$

where H_s^{-j} excludes position j . When scoring $x_{s,j}$, this avoids direct self-inclusion; the target y_s is not in H_s and is scored against the complete prefix. The coefficient $r_{s,j}^H$ scales the historical embedding, while r_s^T scales the target loss.

3.2. Local KG Adaptation with a Matched Null

Let $C = (c_1, \dots, c_n)$ be an ordered context. A directed typed two-hop pattern $p = (r_1, c, r_2)$ links an ordered item pair via connector c ; $\phi_p(u, v)$ indicates whether (u, v) instantiates p . High coverage in C does not by itself make p informative, because similar coverage may also occur in contexts with comparable structural attributes. We therefore form matched-null contexts $\{C^{0,b}\}_{b=1}^{B_A}$ by replacing each position with an item of similar training popularity, KG degree, and linkage status while preserving length and order. For a value t and nonempty multiset $\mathcal{A}$, $Q(t; \mathcal{A})$ records the fraction of values below

t , with ties contributing one half:

$$Q(t; \mathcal{A}) = \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} [\mathbb{I}(a < t) + \frac{1}{2} \mathbb{I}(a = t)]. \quad (2)$$

The observed coverage and its matched-null weight are

$$\begin{aligned} \kappa_p(C) &= \frac{1}{n(n-1)} \sum_{\substack{1 \leq a, b \leq n \\ a \neq b}} \phi_p(c_a, c_b), \\ \alpha_p(C) &= \left[\kappa_p(C) - \text{median}_b \kappa_p(C^{0,b}) \right]_+ Q\left(\kappa_p(C); \{\kappa_p(C^{0,b})\}_{b=1}^{B_A}\right). \end{aligned} \quad (3)$$

The first line measures the fraction of ordered position pairs covered by p . In the second, the bracketed term keeps the path only when its coverage exceeds the matched-null median, while Q gives more weight when that coverage ranks higher among the null values. The weighted paths with $\alpha_p(C) > 0$ form $\mathcal{P}_C^*$ and define the local KG view $\mathcal{G}_C^*$.

3.3. Matched Support Calibration and Training Adaptation

Even after path filtering, an easily connected item may match many retained paths. We therefore match references $\{z_b\}_{b=1}^{B_R}$ to v by training popularity, KG degree, and linkage status. They are scored only after $\mathcal{G}_C^*$ is fixed, so they cannot change the selected paths:

$$\begin{aligned} S_K(v | C) &= \frac{\sum_{a=1}^n \sum_{p \in \mathcal{P}_C^*} \alpha_p(C) \phi_p(c_a, v)}{n \sum_{p \in \mathcal{P}_C^*} \alpha_p(C)}, \\ R(v; C) &= \min \left\{ 1, 2Q\left(S_K(v | C); \{S_K(z_b | C)\}_{b=1}^{B_R}\right) \right\}. \end{aligned} \quad (4)$$

$S_K(v | C)$ is the average fraction of retained path weight connecting v to the context. The second line compares this score with the matched-reference scores: a mid-rank percentile below one half gives $R(v; C) = 2Q < 1$, while one at or above one half gives $R(v; C) = 1$.

Each historical embedding is multiplied by its coefficient, $\tilde{\mathbf{e}}_{s,j} = r_{s,j}^H \mathbf{e}(x_{s,j})$, before sequence encoding. For $\ell_s = \text{CE}(f_\theta(\tilde{H}_s), y_s)$ and batch $\mathcal{B}$, the target-weighted loss is $\sum_{s \in \mathcal{B}} r_s^T \ell_s / \sum_{s \in \mathcal{B}} r_s^T$ when the denominator is positive, and the unweighted mean otherwise.

If $n < 2$, a valid matched background or reference set cannot be constructed, or $\mathcal{P}_C^* = \emptyset$, the KG channel uses $R(v; C) = 1$ and skips undefined quantities. Global matching statistics use the training partition; sample-specific retention coefficients are computed offline using training interactions and a fixed KG, and are detached from recommendation optimization.

4. Experimental Results

Experimental setup. We evaluate KG calibration and sequential recommendation on Steam Games, which contains 25,389 users, 4,089 items, 328,278 interactions, and 462,016 KG triples over six relations. We use SASRec [1], STEAM [5], BirDRec [6], and SSDRec [7], all of which are implemented within the RecBole framework [11]. For each backbone and its ADAPTEDKG-enhanced counterpart, we use identical data splits, candidate sets, and backbone hyperparameters. Retention coefficients and interaction-derived matching statistics are computed using training interactions only.

KG-side calibration. We isolate reference calibration by fixing the target queries, local path views, and number of references while changing only the reference sampler. The substantially weaker correlations under matched sampling (Figure 1) indicate that the estimated retention coefficients are less coupled to item-side structural exposure.

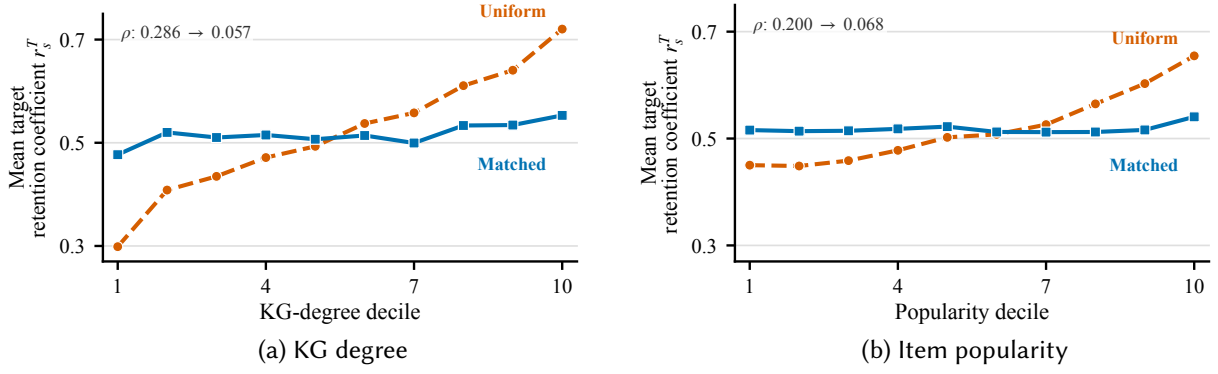


Figure 1: Matched-reference sampling attenuates the structural dependence of interaction reliability on Steam. Mean target retention coefficient r_s^T across deciles of (a) KG degree and (b) item popularity under uniform and matched reference sampling.

Table 1

Steam Games results ($\times 10^3$). Each entry in panel (a) reports the backbone result followed by its ADAPTEdKG-enhanced counterpart; panel (b) reports SASRec-based component ablations. H and N denote HR and NDCG, respectively.

(a) Backbone comparison					(b) Component ablation				
Backbone	H@5	H@10	N@5	N@10	Variant	H@5	H@10	N@5	N@10
SASRec	72.0/ 85.9	121.8/ 141.4	45.3/ 56.5	61.3/ 74.3	w/o Local Adapt.	82.2	137.6	53.4	71.2
STEAM	85.1/ 96.2	148.5/ 154.9	52.7/ 60.3	73.0/ 79.2	w/o Matched Null	84.0	139.5	55.0	72.9
BirDRec	78.3/ 86.0	129.5/ 134.8	50.0/ 55.9	66.5/ 71.6	w/o Matched Ref.	84.8	139.4	55.5	73.0
SSDRec	75.5/ 87.7	127.5/ 145.1	50.8/ 59.2	67.4/ 77.4	ADAPTEdKG	85.9	141.4	56.5	74.3

Recommendation performance. ADAPTEdKG achieves higher values for every reported metric across all four backbones (Table 1(a)). The gains with SASRec show that ADAPTEdKG can benefit a standard sequential recommender without an existing behavior-denoising mechanism. The gains with STEAM, BirDRec, and SSDRec suggest that its KG-derived retention signal is compatible with behavior-denoising mechanisms learned from interaction logs.

Ablation study. In the ablations, *w/o Local Adapt.* removes the entire local adaptation stage, using the global path set with context-independent weights. *w/o Matched Null* retains the context-specific local view but removes matched-null calibration by setting $\alpha_p(C) = \kappa_p(C)$. *w/o Matched Ref.* replaces matched references with uniform references. All three variants lower every reported metric relative to the full model (Table 1(b)), with local adaptation producing the largest drops and either matching stage producing smaller declines. This pattern is consistent with local adaptation forming example-specific path evidence, matched-null calibration refining path selection and weighting, and matched-reference calibration refining interaction support.

5. Conclusion

We presented ADAPTEdKG, which uses knowledge graph relations to assess the contextual support for an interaction. The method compares the observed context with structurally similar alternatives to identify and weight paths that stand out from the background. It then uses similar reference items to calibrate interaction support. The resulting retention coefficients adjust historical representations and target losses during training. All graph processing is completed offline, leaving the recommendation backbone and inference procedure unchanged. The results show that calibrated graph evidence can complement behavior denoising in sequential recommendation.

Declaration of Use of Generative AI

During the preparation of this work, the authors used OpenAI Codex on a limited basis for language polishing and to assist with the implementation and debugging of experimental code. All tool-assisted text and code were critically reviewed, tested, and revised by the authors. The research design, methodological decisions, analysis, and interpretation remained the responsibility of the authors, who take full responsibility for the content of this work.

References

- [1] W. Kang, J. J. McAuley, Self-attentive sequential recommendation, in: IEEE International Conference on Data Mining, ICDM 2018, Singapore, November 17-20, 2018, IEEE Computer Society, 2018, pp. 197–206. doi:10.1109/ICDM.2018.00035.
- [2] M. Quadrana, D. Jannach, P. Cremonesi, Tutorial: Sequence-aware recommender systems, in: Companion of The 2019 World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019, ACM, 2019, p. 1316. doi:10.1145/3308560.3320091.
- [3] Y. Sun, B. Wang, Z. Sun, X. Yang, Does every data instance matter? enhancing sequential recommendation by eliminating unreliable data, in: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021, ijcai.org, 2021, pp. 1579–1585. doi:10.24963/IJCAI.2021/218.
- [4] H. Chen, Y. Lin, M. Pan, L. Wang, C. M. Yeh, X. Li, Y. Zheng, F. Wang, H. Yang, Denoising self-attentive sequential recommendation, in: RecSys '22: Sixteenth ACM Conference on Recommender Systems, Seattle, WA, USA, September 18 - 23, 2022, ACM, 2022, pp. 92–101. doi:10.1145/3523227.3546788.
- [5] Y. Lin, C. Wang, Z. Chen, Z. Ren, X. Xin, Q. Yan, M. de Rijke, X. Cheng, P. Ren, A self-correcting sequential recommender, in: Proceedings of the ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023 - 4 May 2023, ACM, 2023, pp. 1283–1293. doi:10.1145/3543507.3583479.
- [6] Y. Sun, B. Wang, Z. Sun, X. Yang, Y. Wang, Theoretically guaranteed bidirectional data rectification for robust sequential recommendation, in: Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023, 2023.
- [7] C. Zhang, Q. Han, R. Chen, X. Zhao, P. Tang, H. Song, Ssdrec: Self-augmented sequence denoising for sequential recommendation, in: 40th IEEE International Conference on Data Engineering, ICDE 2024, Utrecht, The Netherlands, May 13-16, 2024, IEEE, 2024, pp. 803–815. doi:10.1109/ICDE60146.2024.00067.
- [8] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. de Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, A. N. Ngomo, A. Polleres, S. M. Rashid, A. Rula, L. Schmelzeisen, J. F. Sequeda, S. Staab, A. Zimmermann, Knowledge graphs, ACM Comput. Surv. 54 (2022) 71:1–71:37. doi:10.1145/3447772.
- [9] Q. Guo, F. Zhuang, C. Qin, H. Zhu, X. Xie, H. Xiong, Q. He, A survey on knowledge graph-based recommender systems, IEEE Trans. Knowl. Data Eng. 34 (2022) 3549–3568. doi:10.1109/TKDE.2020.3028705.
- [10] H. Shomer, W. Jin, W. Wang, J. Tang, Toward degree bias in embedding-based knowledge graph completion, in: Y. Ding, J. Tang, J. F. Sequeda, L. Aroyo, C. Castillo, G. Houben (Eds.), Proceedings of the ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023 - 4 May 2023, ACM, 2023, pp. 705–715. doi:10.1145/3543507.3583544.
- [11] W. X. Zhao, S. Mu, Y. Hou, Z. Lin, Y. Chen, X. Pan, K. Li, Y. Lu, H. Wang, C. Tian, Y. Min, Z. Feng, X. Fan, X. Chen, P. Wang, W. Ji, Y. Li, X. Wang, J. Wen, Recbole: Towards a unified, comprehensive and efficient framework for recommendation algorithms, in: CIKM '21: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event, Queensland, Australia, November 1 - 5, 2021, ACM, 2021, pp. 4653–4664. doi:10.1145/3459637.3482016.