Session-Based Recommender Systems Enhanced with Anomaly Detection: A Comparative Study

Bahareh Rahmatikargar

School of Computer Science, University of Windsor, Sunset Avenue, Windsor, N9B 3P4, Ontario, Canada.
{rahmatib}@uwindsor.ca

Abstract. Recommender systems are an application of social network analysis that model user–item interactions. Session-based recommender systems make use of short-term user activities to generate personalized recommendations, but their performance can be affected by anomalous points due to noise, error, or unusual user behavior. This work extends a Graph Neural Network (GNN)-inspired framework by adding four anomaly detection methods: Isolation Forest, Local Outlier Factor (LOF), One-Class Support Vector Machine (SVM), and Density-Based Spatial Clustering of Applications with Noise (DBSCAN). These methods are applied at the session representation level and evaluated on three benchmark datasets—Yoochoose 1/64, Gowalla, and Diginetica. Experimental results indicate that session-level anomaly detection, particularly with moderately sparse datasets, significantly improves hit rate (HR) and mean reciprocal rank (MRR) and offers a reasonable improvement over baseline research.

Keywords: recommender systems, accuracy, graph neural networks (GNNs), anomaly detection

1 Introduction

Recommender systems have risen as a significant research area in artificial intelligence, assisting users in navigating large digital environments to discover items suited to their interests [20]. They predict user-item interactions using collaborative, content-based, and context-based techniques, evolving from content-based filtering to advanced methods like graph-based, session-based, and cross-domain recommendations [15][16][12]. Session-based recommender systems have gained attention for providing real-time recommendations based on a user's current session, ideal for unregistered users with limited historical data [10]. Using Graph Neural Networks (GNNs), these systems model short-term user intent and contextual factors, improving on earlier Recurrent Neural Network approaches [13][7]. GNNs leverage item features and inter-item relationships to enhance recommendation quality [25][26].

However, anomalies from noise, data errors, or rare user behavior can bias embeddings and lower prediction quality [27]. Anomaly detection identifies data points deviating from normal patterns, yet its impact on session-based systems is understudied. We compare four anomaly detection techniques, including Isolation Forest [9], Local Outlier Factor (LOF) [2], One-Class SVM [21][18], and DBSCAN [5], applied at the session representation level in a GNN-based model, using Yoochoose 1/64, Gowalla, and Diginetica datasets.

This extends our prior work [14] to assess multi-technique detection across diverse datasets. To address scalability, we optimize graph sparsity by filtering low-frequency items and use mini-batch training, though large-scale deployment may require graph sampling. Hyperparameter tuning is conducted via grid search, optimizing thresholds (e.g., Isolation Forest's z-score ± 2.2) to maximize HR and MRR across methods. The paper is structured as follows: Section 2 reviews related work, Section 3 details GNN methodology and anomaly detection, Section 4 presents experiments, and Section 5 concludes.

2 Related Work

Session-based recommender systems leverage graph neural networks to capture user behavior. SR-GNN [26] models item transitions with attention mechanisms, while the Dual-channel Graph Transition Network [28] fuses intra- and inter-session dynamics. The Interval-enhanced Graph Transformer [22]incorporates time intervals and user preferences, and RN-GNN [23] addresses repeat consumption with intra-session learning.

Anomaly detection in graph-based systems enhances recommendation quality. TA-Detector [24] uses GNNs and residual networks to identify untrustworthy connections, while GAD-NR [17] employs graph autoencoders for contextual and structural anomaly detection.

Our work builds on a dual GNN framework [8], combining Adaptive GNN (A-GNN) for dynamic item correlations and Single Gate GNN (SG-GNN) for sequential dependencies, embedding items in a low-dimensional space with local and global preferences.

We extend this framework by applying classical anomaly detection methods: LOF [2] for local density-based anomalies, One-Class SVM [18] for boundary-based outliers, and DBSCAN [5] for density-based clustering, to improve recommendation accuracy across Yoochoose, Gowalla, and Diginetica datasets, building on our prior work with Isolation Forest [14].

These classical methods, including LOF, One-Class SVM, and DBSCAN, effectively handle noise and outliers in session-based systems, improving recommendation robustness across diverse datasets [2][18][5]. Their scalability for high-dimensional data is enhanced by optimized implementations, as shown in prior studies [6][19].

3 Methodology

Let $U = \{u_1, u_2, \ldots, u_n\}$ be a set of users and $I = \{i_1, i_2, \ldots, i_m\}$ a set of items. Each user interacts with items through actions such as clicks, purchases, or views, depending on the domain. Modern session-based recommender systems, especially those using graph neural networks, follow a multi-stage pipeline: graph construction, item embedding, session representation, and scoring.

Given a set of sessions $S = \{s_1, s_2, \ldots, s_{|S|}\}$ where $s = [u, i_1, i_2, \ldots, i_k]$, we construct a session graph $G = (V, E)$ with items as nodes and item transitions as edges. We adopt a dual-channel GNN model [8], combining A-GNN (capturing implicit relations via attention) and SG-GNN (modeling sequential dependencies). Their outputs are fused to generate item embeddings, which are aggregated into session representations.

Anomaly detection is applied at the session representation level to remove noisy sessions that may harm recommendation quality. We evaluate four classical unsupervised methods: Isolation Forest, LOF, One-Class SVM, and DBSCAN. Identified anomalies are zeroed out to eliminate their impact on final scoring.

Experiments are conducted on three datasets: sparse, moderately sparse, and dense, to assess how detection performance varies with session length, noise level, and interaction density. The full procedure is outlined in Algorithm 1.

Algorithm 1 Anomaly Detection at Session Representation Layer

Require: Set of sessions S, set of items I, Anomaly detection model $M \in \{\text{Isolation Forest}, \text{LOF}, \text{One-Class SVM}, \text{DBSCAN}\}$
Ensure: Top-k recommended items for each session
Build session graph G from session data S
Generate item embeddings: $E_I = \text{embed}(G)$
Compute two views of item representations:
  $IEmb_1 = \text{A-GNN}(E_I)$
  $IEmb_2 = \text{SG-GNN}(E_I)$
Fuse embeddings: $F_E = \text{Fuse}(IEmb_1, IEmb_2)$
Generate session embeddings: $SessionRep = \text{Encode}(F_E)$
Apply anomaly detector M to $SessionRep$
for each session $s_i \in SessionRep$ **do**
  if anomaly score exceeds threshold **then**
    Zero out the session: $s_i = 0$
  end if
end for
Compute item ranking score from cleaned $SessionRep$
return Top-k recommendations for each session

4 Experiment and Results

We used the Python programming language and its libraries, such as PyTorch [11], to implement the proposed models and conduct experiments. To evaluate the effectiveness of anomaly detection in enhancing session-based recommendation systems, we tested a GNN-based approach, building over a dual GNN technique [8].

As mentioned earlier, our work is an expansion of our previous study [14], where we had worked only with Isolation Forest for anomaly detection over the Yoochoose 1/64 dataset. The results of applying anomaly detection earlier, such as during item embeddings or fusion layer, are presented in our previous work [14], which were less effective or even counterproductive, especially when replacing anomalies with "best" embeddings. We achieved the best results when anomaly detection was applied at the global session representation stage.

In this expanded work, we incorporate three additional anomaly detection techniques: LOF, One-Class SVM, and DBSCAN, and compare their performance on two additional datasets, Gowalla and Diginetica, alongside Yoochoose 1/64. These datasets were selected to capture various session properties, allowing for a solid comparison of anomaly detection techniques.

4.1 Experimental Setup

The initial GNN model was initialized with 100-dimensional embeddings, 2-4 GNN layers, 4-6 A-GNN blocks, and dropout levels of 0.1, 0.5, and 0.9. We trained using the Adam optimizer with a learning rate of 0.0005, decaying by 0.5 every five epochs, and L_2 regularization (10^{-5}) with early stopping to prevent overfitting. Parameters were initialized using a Gaussian distribution (mean = 0, std = 0.1), and a fixed embedding dimension and batch size of 100.

The Yoochoose 1/64 dataset [1] has six months' worth of user sessions with an average of 6 items per session. Its noisy and sparse nature, prompted by diverse user behaviors such as random browsing and suspected bot traffic, provide it with a good test bed for anomaly detection in high-traffic environments. It contains 16,766 unique items and 425,757 total sessions.

Gowalla dataset [3], which was taken from a social network site, contains sparse sessions with an average of 4 items per session, reflecting the sporadic user activity typical of location-based services. Sparsity in this case is a unique type of challenge to anomaly detection since outliers will disproportionately influence recommendation quality. This dataset includes 29,510 unique items and 830,893 sessions, adding to the challenge of learning meaningful patterns.

The Diginetica dataset [4] contains an average of 5 items per session and finds a balance between Yoochoose 1/64 density and Gowalla sparsity with moderate noise because of varied user intent (e.g., browsing vs. purchase). The dataset consists of 42,596 unique items and 777,029 sessions, making it the largest among the three in terms of item variety.

We have dropped sessions of length 1 and items that appear less than 5 times during preprocessing.

Performance was evaluated using two standard metrics: **Hit Rate (HR)** and **Mean Reciprocal Rank (MRR)**, defined as follows:

- **Hit Rate (HR)**:

$$HR = \frac{1}{|U|} \sum_{u \in U} \frac{|R_u \cap G_u|}{|G_u|}$$

where $|U|$ is the number of users, R_u is the set of recommended items for user u, and G_u is the set of ground-truth items for user u.

– **Mean Reciprocal Rank (MRR)**:

$$MRR = \frac{1}{|U|} \sum_{u \in U} \frac{1}{\text{rank}_u}$$

where rank_u is the rank of the first relevant item for user u.

4.2 Anomaly Detection Methods and Parameters

We applied four anomaly detection techniques at the global session representation level, where the final session representation is obtained by concatenating global and local representations followed by a linear transformation. Each method was configured with parameters tailored to the characteristics of the datasets:

– **Isolation Forest** [9]: This method isolates anomalies by constructing a forest of decision trees with random feature selection and split points, determining anomalies as points of shorter path lengths. We used z-score thresholds of $[-2, 2]$ for all datasets, an interval that we found to best balance sensitivity and specificity in our earlier work [14]. We selected this interval from a larger set of intervals ([-1.5, 1.5], [-2, 2], [-2.5, 2.5], [-3, 3]), which was employed in preliminary experiments [14] as it provided the best results.
– **LOF** [2]: This algorithm identifies outliers by comparing the local density of a point with that of its k-nearest neighbors according to reachability distances, in order to compute a density ratio. Parameters were assigned such that Yoochoose 1/64 (k = 20, contamination = 0.1), Gowalla (k = 10, contamination = 0.05), and Diginetica (k = 20, contamination = 0.1). The contamination parameter reflects the expected rate of anomalies with a smaller one for Gowalla since it's sparse and bigger ones for Yoochoose and Diginetica since they are denser.
– **One-Class SVM** [18]: This classifier defines a max-margin hyperplane that starts from the origin to divide normal data and instances on the other side of the boundary as outliers. Parameters were: Yoochoose 1/64 (ν = 0.1, γ = 'scale'), Gowalla (ν = 0.05, γ = 0.001), and Diginetica (ν = 0.1, γ = 'scale'). The ν parameter controls the capacity of outliers, scaling it to support Gowalla with lower noise, while γ = 'scale' scales proportionally with data distribution in dense data.
– **DBSCAN** [5]: This clustering algorithm clusters points in an ϵ-radius and requires a minimum neighbor count (min samples) to cluster, flagging the rest of the points as outliers. Values used were: Yoochoose 1/64 (ϵ = 1.5, min_samples = 10), Gowalla (ϵ = 1.0, min_samples = 3), and Diginetica (ϵ = 1.5, min_samples = 7). These values reflect the density of the datasets where lower ϵ and min samples are used in Gowalla because of its sparsity.

Table 1. Performance of DGNN with Anomaly Detection on Global Session Representation Across Datasets

Model	Gowalla		Diginetica		Yoochoose 1/64	
	HR	MRR	HR	MRR	HR	MRR
Main Model	68.34	49.94	70.34	46.34	79.63	48.74
Isolation Forest	70.04	53.90	73.01	51.82	80.63	49.02
LOF	69.36	52.72	73.04	50.67	64.39	27.39
One-Class SVM	69.77	51.23	72.86	51.04	71.38	35.53
DBSCAN	69.21	49.41	74.12	51.05	80.33	48.88

4.3 Results

Table 1 summarizes the performance of DGNN with and without anomaly detection across three datasets. Anomalous sessions were identified and zeroed out based on the best-performing configurations. Model parameters were also fine-tuned, resulting in improved outcomes compared to those reported in [8].

In **Gowalla**, Isolation Forest and LOF showed the highest gains in both HR and MRR, suggesting their effectiveness in handling sparse and noisy sessions. One-Class SVM and DBSCAN yielded moderate improvements.

For **Diginetica**, all anomaly detection methods outperformed the baseline, with DBSCAN achieving the highest HR (74.12) and Isolation Forest the highest MRR (51.82), indicating that Diginetica's moderate density makes it suitable for anomaly detection.

In **Yoochoose 1/64**, Isolation Forest and DBSCAN again improved over the Main Model. However, LOF and SVM performed worse, likely due to over-pruning in a dense and complex environment.

5 Conclusion

This paper emphasizes the impactful potential of anomaly detection in improving GNN-based session-based recommender systems through the integration of LOF, One-Class SVM, and DBSCAN with Isolation Forest. Our empirical evaluations on Gowalla, Diginetica, and Yoochoose 1/64 datasets show that anomalous session filtering at the global representation level consistently enhances recommendation accuracy, with Isolation Forest and DBSCAN yielding substantial HR and MRR gains, especially on moderately sparse datasets. These findings confirm the effectiveness of anomaly detection in the improvement of session-based recommendations, paving the way to further research into adaptive strategies and deeper comprehension of anomalous session patterns in order to further enhance system performance in diverse data landscapes.

6 Acknowledgment

This work was carried out under the supervision of Dr. Ziad Kobti and Dr. Pooya Moradian Zadeh from the School of Computer Science, University of Windsor.

References

1. Ben-Shimon, D., Tsikinovsky, A., Rokach, L., Meisels, A., Shani, G., Naamani, L.: Recsys challenge 2015 and the yoochoose dataset. In: Proceedings of the 9th ACM Conference on Recommender Systems. ACM (2015)
2. Breunig, M.M., Kriegel, H.P., Ng, R.T., Sander, J.: Lof: Identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. pp. 93–104. ACM (2000)
3. Cho, E., Myers, S.A., Leskovec, J.: Friendship and mobility: User movement in location-based social networks. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1082–1090. ACM (2011)
4. CIKM Cup 2016: Diginetica challenge dataset. Retrieved from https://cikm2016.cs.iupui.edu/cikm-cup/ (2016)
5. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). vol. 96, pp. 226–231. AAAI Press (1996)
6. Goldstein, M., Uchida, S.: A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PloS ONE **11**(4), e0152173 (2016)
7. Jouyandeh, F., Sadeghi, S., Rahmatikargar, B., Zadeh, P.M.: Fake news and covid-19 vaccination: A comparative study. In: Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. pp. 525–531. IEEE (2021)
8. Li, Z., Wang, X., Yang, C., Yao, L., McAuley, J., Xu, G.: Exploiting explicit and implicit item relationships for session-based recommendation. In: Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining. pp. 553–561. ACM (2023)
9. Liu, F.T., Ting, K.M., Zhou, Z.H.: Isolation forest. In: 2008 Eighth IEEE International Conference on Data Mining. pp. 413–422. IEEE (2008)
10. Ludewig, M., Mauro, N., Latifi, S., Jannach, D.: Empirical analysis of session-based recommendation algorithms: A comparison of neural and non-neural approaches. User Modeling and User-Adapted Interaction **31**, 149–181 (2021)
11. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems. pp. 8024–8035 (2019)
12. Rahmatikargar, B., Khan, A.R., Zadeh, P.M., Kobti, Z.: Cross-domain recommendation: Leveraging semantic alignment and user clustering to address data sparsity. Procedia Computer Science **257**, 706–713 (2025)
13. Rahmatikargar, B., Zadeh, P.M., Kobti, Z.: Social isolation detection in palliative care using social network analysis. In: 2022 22nd IEEE International Symposium on Cluster, Cloud and Internet Computing (CCGrid). pp. 905–912. IEEE (2022)

14. Rahmatikargar, B., Zadeh, P.M., Kobti, Z.: Enhancing recommender systems with anomaly detection: A graph neural network approach. In: Complex Networks & Their Applications XIII: Proceedings of The Thirteenth International Conference on Complex Networks and Their Applications: COMPLEX NETWORKS 2024-Volume 3. p. 16. Springer Nature (2024)
15. Rahmatikargar, B., Zadeh, P.M., Kobti, Z.: Two decades of recommender systems: From foundational models to state-of-the-art advancements (2004-2024) (2024)
16. Ricci, F., Rokach, L., Shapira, B.: Recommender systems: Techniques, applications, and challenges. Recommender Systems Handbook pp. 1–35 (2021)
17. Roy, A., Shu, J., Li, J., Yang, C., Elshocht, O., Smeets, J., Li, P.: Gad-nr: Graph anomaly detection via neighborhood reconstruction. In: Proceedings of the 17th ACM International Conference on Web Search and Data Mining. pp. 576–585. ACM (2024)
18. Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C.: Estimating the support of a high-dimensional distribution. Neural Computation **13**(7), 1443–1471 (2001)
19. Schubert, E., Sander, J., Ester, M., Kriegel, H.P., Xu, X.: Dbscan revisited, revisited: Why and how you should (still) use dbscan. ACM Transactions on Database Systems (TODS) **42**(3), 1–21 (2017)
20. Shao, B., Li, X., Bian, G.: A survey of research hotspots and frontier trends of recommendation systems from the perspective of knowledge graph. Expert Systems with Applications **165**, 113764 (2021)
21. Tax, D.M., Duin, R.P.: Support vector domain description. Pattern Recognition Letters **20**(11-13), 1191–1199 (1999)
22. Wang, H., Zeng, Y., Chen, J., Han, N., Chen, H.: Interval-enhanced graph transformer solution for session-based recommendation. Expert Systems with Applications **213**, 118970 (2023)
23. Wang, J., Xie, H., Wang, F.L., Lee, L.K., Wei, M.: Jointly modeling intra- and inter-session dependencies with graph neural networks for session-based recommendations. Information Processing & Management **60**(2), 103209 (2023)
24. Wen, J., Jiang, N., Li, L., Zhou, J., Li, Y., Zhan, H., Kou, G., Gu, W., Zhao, J.: Ta-detector: A gnn-based anomaly detector via trust relationship. ACM Transactions on Multimedia Computing, Communications and Applications (2024)
25. Wu, S., Sun, F., Zhang, W., Xie, X., Cui, B.: Graph neural networks in recommender systems: A survey. ACM Computing Surveys **55**(5), 1–37 (2022)
26. Wu, S., Tang, Y., Zhu, Y., Wang, L., Xie, X., Tan, T.: Session-based recommendation with graph neural networks. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 346–353. AAAI (2019)
27. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P.S.: A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems **32**(1), 4–24 (2020)
28. Zheng, Y., Liu, S., Li, Z., Wu, S.: Dgtn: Dual-channel graph transition network for session-based recommendation. In: 2020 International Conference on Data Mining Workshops (ICDMW). pp. 236–242. IEEE (2020)