Indonesian Journal of Electrical Engineering and Computer Science
Vol. 38, No. 3, June 2025, pp. 2050∼2058
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v38.i3.pp2050-2058

Long-term user engagement in recommender systems: a review

Swathi Edem[1,2], B. V. Ram Naresh Yadav[1]
[1]Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Hyderabad, Telangana, India
[2]Department of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology, Telangana, India

Article Info

Article history:

Received Apr 7, 2024
Revised Jan 12, 2025
Accepted Feb 28, 2025

Keywords:

Long-term metrics
Recommender systems
Reinforcement learning
User engagement
User satisfaction

ABSTRACT

The purpose of recommender systems (RS) is to facilitate user collaboration and communication on the platform. Nevertheless, there is limited knowledge regarding the extent of this relationship and the techniques by which RS could promote persistent user engagement with the platform. In order to fill this void, the present study investigates the role of RS in transforming users' short-term engagement with the RS into long-lasting involvement with the platform. We present a theoretical framework by reviewing relevant literature in the domains of RS and user engagement to probe these issues. We provide open challenges in this field along with metrics in the present study.

Corresponding Author:

Swathi Edem
Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Hyderabad
Telangana, India
Email: swathiedem@gmail.com

1. INTRODUCTION

The use of recommender systems (RS) is on the rise because they make it easier to find and use information in many different areas of our lives, including shopping, food, travel, social platforms, media, and news. In recent years, recommendation algorithms that prioritize instantaneous user responses like likes and clicks have achieved remarkable success. But it's becoming more and more obvious that putting too much weight on short-term engagement might cause clickbait or pigeon-holing effects, which are bad for users' experience in the long run [1]. Algorithm designers on recommendation systems have started optimizing for other objectives that are better aligned with the long-term user experience, after seeing the pitfalls associated with excessive attention on short-term metrics. For instance, according to Zhao *et al.* [2], a RS's ultimate objective is to encourage users to return to the platform more frequently rather than only meeting their needs during the current session.

It is challenging to optimize for long-term satisfaction since the expected long-term outcomes occur less often, are more complicated, and change slowly over time, as opposed to short-term engagement metrics. One can question if there are simpler options to optimize that are more indicative of the eventual outcome. An enhanced correlation between the suggestion and a more straightforward target for optimization is preferred. Optimizing for long-term satisfaction may be tough due to the complexity and slower change of desired outcomes when compared to short-term engagement measures.

In this research, the distinctions between RS that depend on explicit user rating input and those that rely on implicit user action feedback are presented in a clear and demonstrable manner. The ability to predict the activities that users will do in the future is more effective than the ability to predict explicit rankings for the purpose of enhancing engagement. It is not enough to just focus on predicting implicit behaviors in order to improve user engagement. This article will discuss the long-term metrics that should be examined in order to improve user engagement and satisfaction. Higher degrees of negative user feedback, including negative action rates and browsing effort, are shown by the fact that certain research indicate that recommenders that are primarily concerned with predicting implicit actions may not be as accurate as recommenders that are based on ratings. This finding hints to the need of doing more research to investigate different approaches to combining both positive and negative feedback, with a particular emphasis on penalizing products that have received negative feedback from users.

By combining the RS and including online user interaction, the levels of user engagement achieved are comparable to those achieved by using an implicit-action-based recommender alone. However, it did not result in an increase in user browsing, suggesting a compromise between the objectives of user engagement and satisfaction. The observed outcome is expected to be relevant to other algorithms that simulate explicit or implicit feedback data, since it likely reflects the underlying characteristic of both forms of feedback signals rather than the particular methods used. Additional investigation is required to validate this. Nevertheless, it suggests that in order to precisely assess user happiness, it is essential to take into account elements that extend beyond just ratings or behaviors.

We provide the literature review of user behavior on recommendation systems and reinforcement learning (RL) for learning the long-term metrics in RS. Four main categories define the approaches used in the current work: policy learning, point-of- interest recommendation, exploration and inquiry techniques, and reward formulation methods. We then go into considerable detail on the evaluation parameters. Finally, we list the conclusion and contribution of this study.

2. RELATED WORK

2.1. Analysing user behaviour on recommendation systems

Before extensive research has been conducted on analyzing user behaviour in RS across several disciplines such as human-computer interaction, marketing, and information retrieval. The distribution of user interests in long-tail and specialty material was investigated by Goel *et al.* [3]. User behaviour on RS is affected by user preferences, algorithmic suggestions, and other factors such as personality qualities. Anderson *et al.* [4] investigated the influence of suggestions on the variety of content consumed by users. In order to better assess the impact of recommendations, Villermet *et al.* [5] suggested differentiating between algorithmic and human behaviour when listening to music online. Models were proposed to comprehend the evolving user preferences by utilizing a mix of structural and probabilistic methodologies [6]. The influence of individual and contextual elements on user behaviour in recommendation systems was demonstrated by Karumur *et al.* [7]. Another area of study involves creating simulations or doing field research to assess user behaviours while considering possible confounding variables [1], [8]. The feedback loop between recommendation algorithms and user behaviours is studied by Hansen *et al.* [9] in relation to patterns of consumption on video and music streaming services. While many studies have focused on how to measure user engagement with RS, few have sought to understand users' long-term experiences by tracking their sequential and developing behaviours.

2.2. Reinforcement learning for learning the long-term metrics

In the discipline of artificial intelligence and machine learning, reinforcement learning (RL) is a subfield that focuses on teaching computers to make a series of decisions in an unpredictable and sometimes dynamic environment through the process of trial and error. In the context of RS, RL can be used to optimize long-term user engagement by learning from user feedback and adapting to their changing preferences over time. This can involve rewarding recommended items that are well-received by users and penalizing items that are not, in order to incentivize the system to provide recommendations that are more likely to be relevant and engaging to individual users. RS, may be represented as an agent that interacts with users, who function as the environment. After each suggestion request is fulfilled by the agent, we may log the feedback and status changes from users. This data can be used to compute a reward and update the agent's current state. Utilizing RL will result in the development of a recommendation policy that maximizes user engagement over an extended period of time.

RL can also be used to optimize the exploration and exploitation of new recommendations, balancing the need to introduce new items [10]. Some works use RL to look for the best weights in order to make users happy over the long run. Off-policy RL is used by Han *et al.* [11] to find the best weights for an advertiser's expected CTR and bid price. Because interacting with young agents too much will ruin the user experience, they create an environment simulator to get feedback from users while they train their model offline. But the real recommendation world is too complicated for the simulator to fully represent it. The RL model that is built on the simulator will actually be hard to use online, which will make the experience worse for the users. Pei *et al.* [12] suggest using RL to find the best way for a platform to make the most money. To make the model easier to understand, they use an evolutionary strategy to solve the problem. This means that the proposed methods can only improve the profile of the current suggestion. Various simulators like Sim2Rec [13], [14] have been used to test the RL based approaches which mainly focused on long-term user engagement.

RESACT [15] is a sequential RS focuses on residual actor-critic methods, which learn to predict the future rewards or returns. This approach evaluated on return time and session length. RESACT can better capture the complex relationships between user preferences and historical interaction data. This results in more accurate and personalized recommendations, leading to increased user engagement and retention. To minimize the accumulated time interval of multiple user sessions to optimize retention, a dubbed RLUR [16] is proposed. In this work, short video recommendation problem is formulated as an infinite-horizon request-based Markov decision process (MDP).

Scalarized multi-objective RL (SMORL) [17] is proposed with two objectives namely novelty, accuracy and diversity of recommendations. The decision-making strategy known as Sim2Rec [13] is designed to maximize the quality of long-term user engagement by optimizing real-world user involvement. It combines a simulation-based framework with machine learning techniques to simulate user behavior and make data-driven recommendations. This approach has been used in various industry applications, such as e-commerce and personalized medicine, to improve user engagement and retention. ADAREC [14] is an advanced recommendation system that optimizes long-term user involvement through adaptive, sequential decision-making processes, rather than just generating item-to-item recommendations. The goal of ADAREC is to increase user happiness and foster long-term engagement by learning from user feedback and adjusting to shifting preferences.

3. METHODOLOGIES AND ANALYSIS

3.1. Policy learning

For the purpose of enhancing user involvement over the long term, a significant number of research works are focused on policy. The process of finding the ideal approach, or policy, that an agent ought to adopt in order to maximize its cumulative reward in a particular environment is referred to as policy learning in the field of RL. This kind of learning is known as RL, and it involves an agent interacting with its surroundings by performing actions and getting feedback in the form of rewards or penalties. It is the objective of the agent to acquire a policy that will enable them to choose actions that will result in the greatest potential cumulative reward over the course of time period.

Break is suggested as a means of promoting and maintaining the user over a longer amount of time [18]. According to this strategy, encouraging the user to take a break from the RS is a way to boost user satisfaction, which in turn leads to more engagement over a longer period of time. It is possible to use cutting-edge strategies in order to make the policy flexible enough to accommodate shifting preferences among users. A context encoder is used by ADAREC inside the policy network. This encoder makes it possible for RL rules to recognize various patterns of user activity[14].

Both a batch RL framework and an online exploration component are included into the framework, which is referred to as BatchRL-MTF. The former makes use of batch RL to train an optimum recommendation strategy from the fixed batch data offline for long-term user happiness, whilst the latter investigates alternative high-value actions immediately in order to overcome the local optimal dilemma [19]. Optimizing long-term user engagement with the usage of FeedRec [20]. There are two components that make up FeedRec: i) a Q-network, which is constructed in hierarchical long short-term memory (LSTM), is responsible for simulating complicated user behaviours; and ii) an S-network, which has the capacity to mimic the environment, provides assistance to the Q-network, and eliminates the instability of convergence in policy learning.

3.2. Point of interest recommendation

According to the short-term preferences, the next point of interest (POI) that a user will visit is impacted by the objects and venues that the user has recently visited in the trajectory that is now being followed. For instance, a person may go to a pub immediately after eating eaten at a restaurant the previous night. An individual's long-term preferences are a representation of their general interests, which are derived from their historical trajectories. In conclusion, there is a tendency for short-term preferences to fluctuate often throughout the course of time, but long-term choices tend to remain relatively consistent. In the early stages of POI recommendation research, the primary emphasis was placed on assessing the preferences of users via the use of collaborative filtering (CF), particularly algorithms based on matrix factorization (MF). Only the consumers' choices that are static may be modelled using these approaches. These sorts of recommenders are unable to reflect the dynamism of user preferences, thus for instance, when a user who lives in India visits to Landon for a vacation, they may still recommend POIs that are situated in India. approaches that are based on deep learning have lately shown promising results in a variety of recommendation systems. Some examples of these approaches are embedding learning, neural CF, deep latent factor model, and metric learning.

For the purpose of modelling the long-term periodicity, DeepMove [21] makes use of a deep neural network that is equipped with two attention processes. For the purpose of modelling preferences over a short period of time, both deep residual collaboration learning (DRCF) and DeepMove have utilized recurrent neural network (RNN)-based techniques. A gated mechanism that models both long-term and short-term interests is presented by satio-temporal gated network (STGN) [2], which is also an effort to simulate both types of interests. This mechanism falls within the LSTM architecture. RNN-based approaches are becoming more prevalent in the area of next-POI recommendation [21], [13]. Both the LSTM-based and gated LSTM frameworks are used by temporal and multi-level context attention (TMCA) [22] and STGN [2] in order to acquire knowledge about spatial-temporal contexts, respectively. To capture the sequential transition, Deep-Move [21] develops a multi-modal RNN. A geo-dilated RNN is used for short-term preference learning, whereas a nonlocal network is used for long-term preference modeling using long- and short-term preference modeling (LSTPM) [23].

3.3. Explore and exploitation methods

The process of trying out various activities to learn more about the environment and enhance the agent's comprehension of which acts result in desirable outcomes is referred to as exploration. Contrarily, exploitation is choosing courses of action that are anticipated to result in the greatest immediate profit by applying the knowledge or information obtained from prior experiences. By selecting activities that the agent deems to be the best at the moment given its current knowledge, exploitation seeks to maximize the short-term benefit. One of the main challenges in RL is balancing exploration and exploitation. The agent must choose behaviors that are known to be effective based on prior experiences (exploitation) and try out novel actions to learn more about the environment (exploration).

The recommendation is based not just on the predicted number of clicks that the user will make immediately, but also on the anticipated number of clicks that will come from the user's subsequent return. The exploitation for immediate click, the exploitation for projected future clicks, and the exploration of unknowns for model estimate are the three competing variables that are taken into consideration while developing a bandit-based solution for online learning. This method is created on the basis of this idea.

3.4. Reward formulation

Reward formulation in RL involves designing and defining the reward signal that the agent receives from the environment based on its actions. The reward signal serves as feedback to the agent, guiding it to learn a policy that maximizes the cumulative reward over time. Reward formulation is a critical aspect of RL, as it directly influences the behavior of the agent and ultimately determines the success of the learning process.

A surrogate for long-term user experience was offered in a study [24], which suggested that it is possible to develop a surrogate by using data on previous user interactions with the RS and then utilizing that data to model the user's upcoming and current preferences. This makes it possible to represent the long-term user experience by a fictitious object, which can then be utilized to develop the system and measure performance metrics. To increase long-term user involvement, it has been proposed that a reward mechanism be developed using PrefRec. It automatically trains a reward function in an end-to-end manner using the preferences. After that, learning signals are provided by the reward function in order to train the recommendation policy [25].

4. METRICS

After while there isn't a specific set of universally agreed-upon long-term metrics for user engagement in RS. There are several commonly used metrics that can provide insights into long-term user engagement. The metrics in the paper [14] are the users' average return days, the users' return probability on the next day, unlike rate, comment rate, like rate, and the cumulative retention reward. The metrics like returning time and user retention are used to evaluate the performance of the RS [16]. The mean absolute error (MAE), root mean square error (RMSE), and normalized discounted cumulative gain (NDCG) are the metrics used the framework implemented by Ji *et al.* [20]. Table 1 shows what metrics are adopted in the existing methods and in this section, those metrics are described. These metrics are often adapted based on the specific goals and context of the RS [26]. Here are some examples:

– App dwell time: it is the average amount of time that users spend using an app in a given day [19].
– User positive-interaction rate: it represents the proportion of video plays that include favorable user interactions during a 24-hour period [19].
– Cumulative clicks over time: it is the total number of clicks made during the engagement up to that specific moment. A higher total number of clicks over time suggests that an algorithm encourages clients to click and return more frequent [27].
– Click-through rate (CTR): it is calculated as the percentage of clicks received by an algorithm to the total number of suggestions it has generated. Users are more likely to click on the suggested products when the CTR is greater [14], [27].
– Average return time: calculated as the mean amount of time that has passed between a user's successive visits up to a specific point in time [15], [16], [27].
– Return rate: it is the proportion of all the suggestions the algorithm has produced to the number that prompt a user to return within the threshold τ. An algorithm that increases return rates encourages people to use it again [16], [27].
– Improved user ratio: the metric is computed as the proportion of users whose mean return time is decreased compared to their recorded mean return time in the offline data, up to a certain point during contact. A higher proportion of users who have seen improvement implies that the algorithm is more successful in increasing user engagement with the system [27].
– No return count: it refers to the count of users who left the system after a suggestion. A lower no return count shows that an algorithm retains a greater number of users inside the system. The return count is normalized by dividing it by the total number of non-return users in the offline data [27].
– Ratio-based diversity: the fraction of distinct subject clusters in *S*:

$$D_{ratio}(S) = \frac{|unique\_topic\_clusters\_in\_S|}{|S|} \quad (1)$$

$|S|$ S represents the total count of items, including any duplicates, that the user has consumed [24].

– Distribution-based diversity: the consumption history of a consumer may be shown as a distribution across multiple topics. Let N_i represent the quantity of things consumed from subject i, the entropy-based diversity is:

$$D_{entrophy}(S) = -sum_i \widetilde{p_i} log(\widetilde{p_i}) \quad (2)$$

where $\widetilde{p_i} = \frac{N_i}{\sum_i N_i}$ is the proportion of items from topic i in S.

– Hit ration (HR): $HR@k$ is a statistic that quantifies the extent to which an item is ranked inside the $top - k$ slots of a recommendation list [17].

$$HR(click) = \frac{\#hits\_among\_clicks}{\#clicks} \quad (3)$$

– Coverage: it is calculated as the proportion of all things, including less popular items, that are covered by every $top - k$ suggestion from the test or validation sequences [17].
– Repetitiveness: calculates the average number of repeats in the top-*k* places of recommendation lists every session [17].

$$R@K = \frac{1}{N} \sum_{i=1}^{N} \#repetitions\_in\_top - k\_items\_of\_session\_i \quad (4)$$

– Normalized capped importance sampling: y, a policy's score π is determined using:

$$J^{NCIS} = \frac{1}{|U|} \sum_{v \epsilon U} [\frac{\sum_{i=0}^{\top_v} \widetilde{p_i}(\pi, \top_v) L_i^v}{\sum_{i=0}^{\top_v} \widetilde{p_i}(\pi, \top_v)}] \quad (5)$$

where U is the set of test users $\top_{\upsilon}$ is the set of sessions of that user L_i^{υ} is the level of change for the i-th session and $(\upsilon, \widetilde{p_i}(\pi, \top_{\upsilon})$is the probability that the policy π follows the request trajectory of the i-th session in $(\top_{\upsilon})$ [15], [25].

- Retention rate: the proportion of users who continue to utilize the RS over a certain time period is referred to as the retention rate inside the system. The capacity of the system to retain the interest and engagement of users over an extended period of time is shown by this statistic [14], [16].
- NDCG: cumulative gain refers to the entire sum of gains. Discounted cumulative gain (DCG) is an idea that builds upon the idea of cumulative gain (CG) by including the process of discounting the gains based on the rank. Ideal discounted cumulative gain (IDCG) is the computation of the DCG for the ideal ranking, taking into account the gains. NDCG may be seen as the assessment of the similarity between the actual order of relevance and the desired order of relevance [17].

$$NDCG = \frac{DCG}{IDCG} \tag{6}$$

As per the literature study, there are few more metrics could be used for the evaluation of long-term user engagement in RS like session length, frequency of interaction, churn rate, conversion rate, and customer lifetime value. These metrics could be experimented in the future work.

Table 1. Summary of the methods focused on long-term user engagement using RL algorithms

Model	Year	Method	Dataset/domain	Evaluation metric
RESACT [15]	2023	Policy learning	A real-world dataset RecL-25M and synthetic dataset MovieLensL-1M	normalized capped importance sampling (NCIS)
Dubbed RLUR [16]	2023	Policy learning	KuaiRand	Returning time, user retention
Lotka-Volterra dynamical system [18]	2023	Policy learning	MovieLens 1M dataset, Goodreads dataset	Mean long-term engagement rate (LTE)
ADAREC [14]	2023	Policy learning	Real-world E-commerce dataset	The cumulative retention reward, the users' average return days, the users' return probability on the next day (LTE)
SMORL [17]	2022	Reward formulation	RC15 and Retail rocket	Accuracy, diversity, novelty, repetitiveness
PrefRec [25]	2022	Reward formulation	Customized dataset of short-form videos	Session depth, visiting frequency sampling (NCIS)
Multi-task learning model (MTL) + multi-task fusion model (MTF) [19]	2022	Policy learning	A real-world short video dataset	App dwell time (ADTime), user positive-interaction rate (UPIRate)
Reward surrogates in an RL-based RS [24]	2022	Reward formulation	Industrial recommendation platform	Overall user visiting frequency, number of satisfied consumptions, user visiting frequency from low frequency user segment, number of homepage visits
FeedRec [20]	2021	Policy learning	Real-world E-commerce dataset	Depth, return time
LSTPM [22]	2020	POI recommendation	The foursquare check-in dataset and Gowalla dataset	NDCG and recall@K

Session lengthis the average duration of user sessions within the RS. Longer session lengths suggest deeper user engagement and interest in the recommended content [15]. Frequency of interactions refers to the frequency with which consumers engage with the RS over a period of time, such as via clicks, views, or purchases. An increased interaction frequency is indicative of a user's continued continued involvement.

Churn rate is the rate at which users disengage or stop using the RS over time. A lower churn rate indicates higher user retention and long-term engagement. Conversion rate is the proportion of people who buy or subscribe to a service after recommendations. Higher conversion rates show suggestion efficacy in motivating user behaviours.

The detailed study of metrics is presented in this section. These metrics should be adjusted according to the RS's unique aims, the suggested content's type, and user interactions. To get a complete picture of user engagement over time, it's a good idea to incorporate both quantitative metrics and qualitative feedback. Many research works used F-measure, precision, and recall to evaluate the RS though those metrics are not suitable to evaluate a RS. The retention rate, NDCG and NCIS are proved metrics to evaluate the long-term user engagement in RS. As there are no standard metrics to evaluate the long-term engagement of the user in RS, it is the responsibility of the researcher to select the right metrics to fulfil the aim of that research.

5. THE CHALLENGES FOR LONG-TERM USER ENGAGEMENT IN RECOMMENDER SYSTEMS

The use of RS has become more important across a variety of online marketplaces. They want to give consumers with personalized suggestions that are based on the interests and actions of the same users; however, the key challenge lies in maintaining long-term user engagement. Common difficulties in achieving long-term user engagement:

- Limited user feedback: often users provide sparse feedback or ratings on recommended items; this complicates the system's ability to accurately understand their preferences.
- Dynamic user preferences: users' preferences can change over time, leading to discrepancies between their current interests and the system's recommendations; this can result in frustration and disengagement.
- Cold start problem: when new users join a RS platform, there is a deficiency of historical data to personalize recommendations for them; this initial phase can deter user engagement before it even begins.
- Frequent changes in the user preferences: it's natural that a user interests change time to time. Sometimes, due to external factors user interests get changed. But in these cases, it is very difficult to predict the future preferences.
- Lack of randomness in recommendations: sometimes user may get bored if the recommendations are of the same types. So it is important to include variety types of recommendations.

To address these challenges, RS must continually adapt and evolve by incorporating innovative algorithms and techniques; this ongoing development is essential to ensure that users remain actively engaged with the system over extended periods. By studying and resolving the factors that inhibit long-term user involvement, RS may boost customer satisfaction and loyalty, ultimately leading to greater success for online platforms and businesses.

6. CONCLUSION

We provided an exhaustive and current analysis of long-term user engagement with RS. We placed a strong emphasis on the significance of user engagement and satisfaction for RS, as well as the difficulties that are associated with this path. Based on our study we conclude that the RL algorithms are suitable to optimize the long term user engagement. Therefore we presented a literature survey on the role of RL algorithms to increase the user long term engagement in RS. There isn't a specific set of universally agreed-upon long-term metrics for user engagement in RS. We presented the metrics considered in the previous works to evaluate the long-term engagement. We feel the research in this direction is very less and significant progress is needed. In conclusion, we are confident that our survey will aid the researcher in comprehending fundamental concepts and future developments to increase the long-term user engagement.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Swathi Edem	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	
B. V. Ram Naresh Yadav		✓				✓			✓	✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal Analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject Administration
Fu : **Fu**nding Acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] A. J. B. Chaney, B. M. Stewart, and B. E. Engelhardt, "How algorithmic confounding in recommendation systems increases homogeneity and decreases utility," in *Proceedings of the 12th ACM Conference on Recommender Systems*, 2018, pp. 224–232, doi: 10.1145/3240323.3240370.

[2] P. Zhao *et al.*, "Where to go next: a spatio-temporal gated network for next POI recommendation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 5, pp. 2512–2524, Jul. 2020, doi: 10.1109/TKDE.2020.3007194.

[3] S. Goel, A. Broder, E. Gabrilovich, and B. Pang, "Anatomy of the long tail: ordinary people with extraordinary tastes," in *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining*, 2010, pp. 201–210, doi: 10.1145/1718487.1718513.

[4] A. Anderson, L. Maystre, I. Anderson, R. Mehrotra, and M. Lalmas, "Algorithmic effects on the diversity of consumption on Spotify," in *The Web Conference 2020 - Proceedings of the World Wide Web Conference, WWW 2020*, 2020, pp. 2155–2165, doi: 10.1145/3366423.3380281.

[5] Q. Villermet, J. Poiroux, M. Moussallam, T. Louail, and C. Roth, "Follow the guides: disentangling human and algorithmic curation in online music consumption," in *RecSys 2021 - 15th ACM Conference on Recommender Systems*, 2021, pp. 380–389, doi: 10.1145/3460231.3474269.

[6] X. Zhao, Z. Zhu, and J. Caverlee, "Rabbit holes and taste distortion: distribution-aware recommendation with evolving interests," in *The Web Conference 2021 - Proceedings of the World Wide Web Conference, WWW 2021*, 2021, pp. 888–899, doi: 10.1145/3442381.3450099.

[7] R. P. Karumur, T. T. Nguyen, and J. A. Konstan, "Personality, user preferences and behavior in recommender systems," *Information Systems Frontiers*, vol. 20, no. 6, pp. 1241–1265, Dec. 2018, doi: 10.1007/s10796-017-9800-0.

[8] D. Holtz, B. Carterette, P. Chandar, Z. Nazari, H. Cramer, and S. Aral, "The engagement-diversity connection: evidence from a field experiment on Spotify," in *EC 2020 - Proceedings of the 21st ACM Conference on Economics and Computation*, 2020, pp. 75–76, doi: 10.1145/3391403.3399532.

[9] C. Hansen, R. Mehrotra, C. Hansen, B. Brost, L. Maystre, and M. Lalmas, "Shifting consumption towards diverse content on music streaming platforms," in *WSDM 2021 - Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 2021, pp. 238–246, doi: 10.1145/3437963.3441775.

[10] L. Zou, J. Song, L. Xia, W. Liu, Z. Ding, and D. Yin, "Reinforcement learning to optimize long-term user engagement in recommender systems," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019, pp. 2810–2818, doi: 10.1145/3292500.3330668.

[11] J. Han, Y. Yu, F. Liu, R. Tang, and Y. Zhang, "Optimizing ranking algorithm in recommender system via deep reinforcement learning," in *Proceedings - 2019 International Conference on Artificial Intelligence and Advanced Manufacturing, AIAM 2019*, 2019, pp. 22–26, doi: 10.1109/AIAM48774.2019.00011.

[12] C. Pei *et al.*, "Value-aware recommendation based on reinforcement profit maximization," in *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*, 2019, pp. 3123–3129, doi: 10.1145/3308558.3313404.

[13] X. H. Chen *et al.*, "Sim2Rec: a simulator-based decision-making approach to optimize real-world long-term user engagement in sequential recommender systems," in *Proceedings - International Conference on Data Engineering*, 2023, vol. 2023-April, pp. 3389–3402, doi: 10.1109/ICDE55515.2023.00260.

[14] Z. Xue *et al.*, "AdaRec: adaptive sequential recommendation for reinforcing long-term user engagement," *OpenReview*, 2024, [Online]. Available: https://openreview.net/forum?id=waeGeAdZUx.

[15] W. Xue *et al.*, "Resact: reinforcing long-term engagement in sequential recommendation with residual actor," *11th International Conference on Learning Representations, ICLR 2023*, 2023, [Online]. Available: https://openreview.net/forum?id=HmPOzJQhbwg.

[16] Q. Cai *et al.*, "Reinforcing user retention in a billion scale short video recommender system," in *Companion Proceedings of the ACM Web Conference 2023*, Apr. 2023, pp. 421–426, doi: 10.1145/3543873.3584640.

[17] D. Stamenkovic, A. Karatzoglou, I. Arapakis, X. Xin, and K. Katevas, "Choosing the best of both worlds: diverse and novel recommendations through multi-objective reinforcement learning," in *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, Feb. 2022, pp. 957–965, doi: 10.1145/3488560.3498471.
[18] E. Saig and N. Rosenfeld, "Learning to suggest breaks: sustainable optimization of long-term user engagement," in *Proceedings of the 40th International Conference on Machine Learning, PMLR*, 2023, vol. 202, pp. 29671–29696, [Online]. Available: https://proceedings.mlr.press/v202/saig23a.html.
[19] Q. Zhang *et al.*, "Multi-task fusion via reinforcement learning for long-term user satisfaction in recommender systems," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2022, pp. 4510–4520, doi: 10.1145/3534678.3539040.
[20] L. Ji, Q. Qin, B. Han, and H. Yang, "Reinforcement learning to optimize lifetime value in cold-start recommendation," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, Oct. 2021, pp. 782–791, doi: 10.1145/3459637.3482292.
[21] J. Feng *et al.*, "Deepmove: predicting human mobility with attentional recurrent networks," in *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, 2018, pp. 1459–1468, doi: 10.1145/3178876.3186058.
[22] R. Li, Y. Shen, and Y. Zhu, "Next point-of-interest recommendation with temporal and multi-level context attention," in *2018 IEEE International Conference on Data Mining (ICDM)*, Nov. 2018, vol. 2018-Novem, pp. 1110–1115, doi: 10.1109/ICDM.2018.00144.
[23] K. Sun, T. Qian, T. Chen, Y. Liang, Q. V. H. Nguyen, and H. Yin, "Where to go next: modeling long- and short-term user preferences for point-of-interest recommendation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, pp. 214–221, Apr. 2020, doi: 10.1609/aaai.v34i01.5353.
[24] Y. Wang *et al.*, "Surrogate for long-term user experience in recommender systems," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Aug. 2022, pp. 4100–4109, doi: 10.1145/3534678.3539073.
[25] W. Xue *et al.*, "PrefRec: recommender systems with human preferences for reinforcing long-term user engagement," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2023, pp. 2874–2884, doi: 10.1145/3580305.3599473.
[26] P. Pu, L. Chen, and R. Hu, "A user-centric evaluation framework for recommender systems," in *Proceedings of the fifth ACM conference on Recommender systems*, Oct. 2011, pp. 157–164, doi: 10.1145/2043932.2043962.
[27] Q. Wu, H. Wang, L. Hong, and Y. Shi, "Returning is believing: optimizing long-term user engagement in recommender systems," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Nov. 2017, vol. Part F1318, pp. 1927–1936, doi: 10.1145/3132847.3133025.

BIOGRAPHIES OF AUTHORS

Swathi Edem is a research scholar at Jawaharlal Nehru Technological University Hyderabad of Computer Science and Engineering, India and assistant professor at Chaitanya Bharathi Institute of Technology, Hyderabad, India. Her research interests include RS, RL, and machine learning. She can be contacted at email: swathiedem@gmail.com.

B. V. Ram Naresh Yadav is professor at Jawaharlal Nehru Technological University Hyderabad, India. He holds a Ph.D. degree with specialization in security. His research areas are security and information retrieval systems. He can be contacted at email: bvramnaresh@jntuh.ac.in.