Towards an Evaluation Framework for Indoor Recommender Systems

Alessio Ferrato[1]

[1]*Department of Engineering, Roma Tre University, 00146 Rome, Italy*

Abstract

Recommender Systems (RSs) play a crucial role in shaping user experiences on the Web, yet their availability is limited when it comes to indoor environments. Indoor RSs face unique challenges, including user localization, privacy concerns, complex spatial layouts, and user adoption. While several evaluation frameworks exist, they are primarily designed for online domains and may not be suitable for indoor recommendations. This paper introduces an evaluation framework tailored for indoor RSs, addressing the scarcity of publicly available datasets and the complexity of model comparison and metric selection. We also emphasize the absence of a suitable dataset for indoor recommendations and propose the integration of a synthetic data generator to facilitate research in this domain. This paper reviews existing evaluation frameworks and identifies their limitations in the context of indoor recommendations, setting the first step for developing a specialized framework. Our work aims to bridge the gap between traditional RSs and indoor environments, paving the way for more effective recommendations in physical spaces.

Keywords

Recommender Systems, Evaluation Framework, Indoor Environment

1. Introduction

A Recommender System (RS) can filter the content in a given scenario by creating personalized recommendations specific to help users make decisions. To date, these systems are widespread on the Web and, in many cases, shape the success of the platforms we use every day [1]. Despite this, when outside the Web, particularly in indoor environments, we rarely take advantage of RSs [2, 3].

The low proliferation of indoor RS is a combination of several factors, and there are some unique challenges to address in this environment compared to their online counterparts [4, 5]. Some key points to consider are user localization, privacy issues, complex environments, and, most importantly, user adoption [6]. In particular, an indoor RS relies mainly on users' physical movements and interactions in the environment, which can be very tangled (e.g., multiple rooms, floors, and different layouts). Moreover, convincing users to adopt this type of system can be difficult considering the amount of information it needs, but building them with anonymity-

preserving techniques can help to overcome this aspect [6]. It is important to note that they must also face the usual issues impacting any RS [1], such as cold-start, data sparsity, scalability, and fairness.

Indoor environments, such as retail stores, museums, and educational institutions, present distinctive needs, and traditional online RSs may need to adapt more effectively to these physical spaces. Even if there are noteworthy works in indoor environments (e.g., see [5, 7]), to the best of our knowledge, the current research landscape suggests that there is still no established approach in these settings due to several factors: lack of publicly available datasets, difficult model comparison, and suitable evaluation metrics selection. To this end, we introduce and discuss an evaluation framework for indoor RSs with a special synthetic data generation module to simplify research.

The following section will discuss the background by introducing different evaluation frameworks we can employ in this domain. Next, we will illustrate the research goals. Finally, our framework and its main components will be discussed by summarizing the work's contributions and presenting the possible limitations.

2. Background

Considering that "offline evaluations are often the first step in conducting evaluations and there is a logical evolution from offline evaluations, through user studies to online analyses" [1], the development of an evaluation framework plays a key role in helping to reproduce experiments [8]. Over the years, several evaluation frameworks for RSs have emerged, some of them theoretical (e.g., see [1]), others freely distributed under license[1], but to our knowledge, none of them was explicitly built for evaluating indoor RSs.

A review of the literature revealed five frameworks, summarized in Table 1, which could be used and adapted to our scenario: Daisy-Rec [9], Mab2Rec[2], RecBole [10], ReChorus [11], RecPack [12].

DaisyRec, RecPack, and RecBole are created explicitly for Top-K recommendation but support only one type of input, a matrix of user and item interaction. For this reason, these systems are not suitable for our application domain since no information about the placement of items in the environment can be used. In contrast, Mab2Rec accepts a more complex representation as input but is a framework that only implements models based on multi-armed bandits, thus limiting evaluation with different recommendation techniques. RecBole, on the other hand, is a vast framework that allows the evaluation of many recommendation tasks. However, even if more than 64 models are implemented for context-aware/session/sequential recommendation, none of them is made to suggest items specifically in an indoor environment.

In addition, all these frameworks already provide and simplify the loading and preprocessing of the most popular datasets in the literature. However, unfortunately, no one of them belongs to the indoor scenario. Finally, they do not implement any mechanism for generating synthetic data even if it would be useful because "in cases where a natural real-world dataset that would be

---

[1]github.com/ACMRecSys/recsys-evaluation-frameworks (last access: February 25, 2024)
[2]github.com/fidelity/mab2rec (last access: February 25, 2024)

Table 1

A summary table of frameworks we may adapt in the indoor recommendation domain.

System	Recommendation Tasks	Data Input
DaisyRec [9]	Top-K	user-item interaction matrix
Mab2Rec[3]	Content-Aware, Context-aware	user-item interaction matrix, user features, item features, eligibility
RecBole [10]	General, Sequential, Context-aware, Knowledge-based	user-item interaction matrix, user features, item features
ReChorus [11]	Top-K	user-item interaction matrix
RecPack [12]	Top-K	user-item interaction matrix

sufficiently suitable for developing, training, and evaluating a RS is not available, a synthesized dataset may be used"[1].

3. Research Goals

Given the above, the Research Goals can be summarized into three points:

1. Implement a synthetic data generator from indoor environment representation (i.e., items and their location).
2. Implement models used in indoor recommendation (e.g., [5, 7]).
3. Identify metrics to be used for evaluation (e.g., crowdness, coverage, popularity).

4. Framework Overview

In every evaluation framework, we can identify three characteristic components: data module, recommendation module, and evaluation module.

The first element is the module in charge of data loading and processing. In our case, we want to enrich the module with a component in charge of generating synthetic data to overcome their lack in literature. A strategy for generation is presented in [13] where, starting from a limited set of real data, the authors generate synthetic datasets to train context-aware RSs. Different strategies can be used to generate the data at this point, starting from less complex simulations (e.g., random-walking) to elaborate scenarios with different user preference profiles (e.g., visiting style in museum [14]), different user flow (e.g., Google Popular Times [4]), or adding dwell time [15].

Another central element is the recommendation module, where all the models available for training are implemented. This part differs between frameworks in the number of models implemented and the recommendation task. In our framework, we will implement the models

---

[4]blog.google/products/maps/maps101 (last access: February 25, 2024)

found in the literature for offline recommendation(e.g., see [5]). However, we will also add traditional models to test whether it is the most suitable in this domain.

Finally, the evaluation module is used to evaluate the RSs through classical metrics (for a complete list please see [1]) with particular attention to the *so-called* fairness metrics [16] if synthetic data are used since real data (e.g., rating) are not available to compute some results (e.g., prediction accuracy).

The system will be modular enough to write a configuration file to start an experiment, making it very easy to introduce new researchers to this topic.
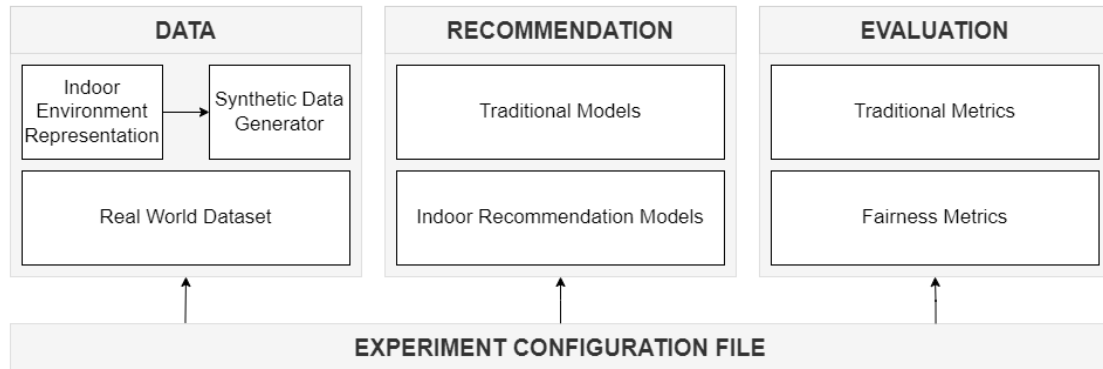
Figure 1: A high level view of the framework.

5. Conclusion

In conclusion, the evaluation of indoor RSs is a challenging task, primarily due to the scarcity of available datasets and models. We recognize this limitation so we are determined to fill this gap, in an underexplored domain, by developing an evaluation framework with new specifics. In this initial contribution, we outlined the research activities planned for this purpose.

References

[1] E. Zangerle, C. Bauer, Evaluating recommender systems: survey and framework, ACM Computing Surveys 55 (2022) 1–38.

[2] F. E. Walter, S. Battiston, M. Yildirim, F. Schweitzer, Moving recommender systems from on-line commerce to retail stores, Information systems and e-business management 10 (2012) 367–393.

[3] A. Ferrato, Challenges for anonymous session-based recommender systems in indoor environments, in: Proceedings of the 17th ACM Conference on Recommender Systems, 2023, pp. 1339–1341.

[4] L. M. F. Otani, V. F. de Santana, Practical challenges in indoor mobile recommendation, arXiv preprint arXiv:2211.15810 (2022).

[5] M. Del Carmen Rodríguez-Hernández, S. Ilarri, R. Hermoso, R. Trillo-Lado, Towards trajectory-based recommendations in museums: evaluation of strategies using mixed synthetic and real data, Procedia computer science 113 (2017) 234–239.

[6] A. Friedman, B. P. Knijnenburg, K. Vanhecke, L. Martens, S. Berkovsky, Privacy aspects of recommender systems, Recommender systems handbook (2015) 649–688.

[7] J. Shin, C. Lee, C. Lim, Y. Shin, J. Lim, Recommendation in offline stores: A gamification approach for learning the spatiotemporal representation of indoor shopping, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 3878–3888. URL: https://doi.org/10.1145/3534678.3539199. doi:10.1145/3534678.3539199.

[8] M. Ferrari Dacrema, S. Boglio, P. Cremonesi, D. Jannach, A troubling analysis of reproducibility and progress in recommender systems research, ACM Transactions on Information Systems (TOIS) 39 (2021) 1–49.

[9] Z. Sun, H. Fang, J. Yang, X. Qu, H. Liu, D. Yu, Y.-S. Ong, J. Zhang, Daisyrec 2.0: Benchmarking recommendation for rigorous evaluation, IEEE Transactions on Pattern Analysis and Machine Intelligence (2022).

[10] W. X. Zhao, Y. Hou, X. Pan, C. Yang, Z. Zhang, Z. Lin, J. Zhang, S. Bian, J. Tang, W. Sun, et al., Recbole 2.0: Towards a more up-to-date recommendation library, in: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, 2022, pp. 4722–4726.

[11] C. Wang, M. Zhang, W. Ma, Y. Liu, S. Ma, Make it a chorus: knowledge-and time-aware item modeling for sequential recommendation, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 109–118.

[12] L. Michiels, R. Verachtert, B. Goethals, Recpack: An(other) experimentation toolkit for top-n recommendation using implicit feedback data, in: Proceedings of the 16th ACM Conference on Recommender Systems, RecSys '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 648–651. URL: https://doi.org/10.1145/3523227.3551472. doi:10.1145/3523227.3551472.

[13] M. Del Carmen Rodríguez-Hernández, S. Ilarri, R. Hermoso, R. Trillo-Lado, Datagencars: A generator of synthetic data for the evaluation of context-aware recommendation systems, Pervasive and Mobile Computing 38 (2017) 516–541.

[14] M. Zancanaro, T. Kuflik, Z. Boger, D. Goren-Bar, D. Goldwasser, Analyzing museum visitors' behavior patterns, in: User Modeling 2007: 11th International Conference, UM 2007, Corfu, Greece, July 25-29, 2007. Proceedings 11, Springer, 2007, pp. 238–246.

[15] A. Ferrato, C. Limongelli, M. Mezzini, G. Sansonetti, A deep learning-based approach to model museum visitors, in: Proceedings of the ACM Intelligent User Interfaces workshops, 2022, pp. 217–221.

[16] Y. Deldjoo, D. Jannach, A. Bellogin, A. Difonzo, D. Zanzonelli, Fairness in recommender systems: research landscape and future directions, 2023.