# Measuring and Mitigating Biases in Location-based Recommender Systems

Pablo Sánchez
*Instituto de Investigación Tecnológica (IIT), Universidad Pontificia Comillas*
Madrid, Spain
psperez@icai.comillas.edu

Alejandro Bellogín
*Universidad Autónoma de Madrid*
Madrid, Spain
alejandro.bellogin@uam.es

Ludovico Boratto
*University of Cagliari*
Cagliari, Italy
ludovico.boratto@acm.org

***Abstract*—This article is a summary of the work published in the journal Data Mining and Knowledge Discovery [1]. It presents an analysis of different types of biases and polarization measurements that affect the area of Point-Of-Interest recommendation. Our results evidence which state-of-the-art recommenders suffer from popularity bias, venue exposure polarization, and geographical distance polarization.**

***Index Terms*—POI recommendation, Bias mitigation, Polarization, Temporal Evaluation**

## I. Introduction

In the tourism sector, consumer needs and service offerings intersect. Travel guides/blogs, traditionally used for trip planning, often highlight popular places. However, to enhance user experience, recommendation algorithms should also suggest novel destinations [2]. The Point-Of-Interest (POI) recommendation, i.e., suggesting new places to visit upon city arrival, is a key problem in tourism [3]. This task is typically based on Location-Based Social Networks (LBSNs) data, where user check-ins enable services like information sharing and venue recommendations [4].

However, LBSN recommendations present challenges, including diverse contextual dimensions and higher user preference sparsity. It is crucial to evaluate LBSNs' impact on tourism. User experiences and venue businesses depend on these recommendations. Nonetheless, Recommender Systems (RSs) may be biased towards certain properties (e.g., popular items), affecting stakeholders differently.

## II. Contributions

In this work, we characterize four forms of polarization in the context of Point-Of-Interest recommendation: towards venue and category popularity, venue exposure, and geographical distance. We achieve this through metrics that have not been used before. Then, we assess if the use of check-ins to capture the interactions of the users with a LBSN to produce recommendations may lead to polarized suggestions from these perspectives. To do this, we consider an evaluation methodology that mimics the real world, by using a temporal split of the user check-ins. We then compare different families of recommenders to inspect these forms of polarization and propose two different, but complementary, approaches to mitigate the observed polarization measurements and biases (further details in [1]).

## III. Results

We performed experiments on the Foursquare global check-in dataset[1]. This dataset is formed by 33M check-ins in different cities around the world. We selected the check-ins from Tokyo, which includes 10,057 users, 24,892 venues, and more than 900K check-ins. Results corresponding to other cities are reported in the full publication [1]. We considered 16 recommendation algorithms grouped in five families:

- non-personalized: Random (Rnd) and Popularity (Pop);
- collaborative filtering: user-based (UB) and item-based (IB) nearest neighbor, matrix factorization using Alternate Least Squares (HKV) and matrix factorization using Bayesian Personalized Ranking (BPR);
- temporal or sequential: user-based neighborhood with a temporal decay function (TD), Factorized Markov Chains (MC), Factorized Personalized Markov Chains (FPMC), Factorized Item Similarity Models with high-order Markov Chains (Fossil);
- purely geographical: Kernel Density Estimation (KDE), based on the closest venues to the user centroid (AvgDis);
- focused on points-of-interest: probabilistic Matrix Factorization (FMFMGM), BPR optimized for POI recommendation (GeoBPR), weighted POI matrix factorization (IRenMF), and a hybrid algorithm that combines UB, Pop, and AvgDis (PGN).

Table I summarizes the results obtained for Tokyo in terms of accuracy (P for precision and nDCG for normalized discounted cumulative gain), novelty (Expected Popularity Complement or EPC), diversity (by measuring how many different venues are being recommended with the Gini coefficient), coverage (user coverage as UC to account for how many users receive at least one recommendation), and polarization measurements considered in our work based on exposure and distance.

We observe that one of the best performing recommenders (in terms of accuracy) is the Pop recommender – this also happens in the other cities, as shown in [1]–, even though the TD model also obtains very competitive results. The accuracy of the POI algorithms is very similar to other classical approaches, like the UB or the BPR.

[1] https://sites.google.com/site/yangdingqi/home/foursquare-dataset

TABLE I
PERFORMANCE RESULTS ON TOKYO CITY, WHERE THE METRICS ARE COMPUTED AT CUTOFF 5. IN ITALICS, THE BEST VALUE FOR EACH FAMILY.

| | **Accuracy** | | **Novelty** | **Diversity** | **Popularity** | | **Exposure** | | **Distance** | | **Coverage** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Recommender** | **P** | **nDCG** | **EPC** | **Gini** | **PopI** | **PopC** | **ExpP** | **ExpR** | **DistT** | **DistU** | **UC** |
| Rnd | 0.000 | 0.000 | **0.999** | **0.551** | **0.303** | *0.760* | **0.000** | **0.001** | 37.2 | 34.7 | **7,253** |
| Pop | **0.071** | *0.087* | 0.746 | 0.000 | 0.000 | 0.960 | 0.131 | 0.121 | *24.9* | *26.4* | **7,253** |
| UB | *0.070* | *0.087* | 0.769 | 0.001 | 0.002 | 0.968 | 0.103 | 0.093 | 26.0 | 25.8 | *6,931* |
| IB | 0.063 | 0.080 | 0.819 | *0.025* | *0.026* | *0.911* | 0.064 | 0.057 | 23.2 | 25.0 | *6,931* |
| HKV | 0.064 | 0.078 | *0.845* | 0.002 | 0.003 | 0.921 | *0.038* | *0.031* | *22.0* | *21.7* | *6,931* |
| BPR | 0.066 | 0.081 | 0.754 | 0.000 | 0.003 | 0.955 | 0.123 | 0.112 | 25.6 | 27.7 | *6,931* |
| TD | **0.071** | **0.088** | 0.776 | 0.001 | 0.003 | 0.965 | 0.097 | 0.087 | 25.9 | 25.4 | *6,931* |
| MC | 0.051 | 0.062 | 0.804 | 0.001 | 0.003 | *0.939* | 0.107 | 0.098 | 26.5 | 30.9 | 6,879 |
| FPMC | 0.053 | 0.064 | 0.807 | 0.001 | 0.001 | 0.943 | 0.103 | 0.096 | 31.0 | 30.1 | 6,884 |
| Fossil | 0.058 | 0.074 | *0.851* | *0.003* | *0.006* | 0.878 | *0.046* | *0.040* | *22.0* | *21.7* | 6,879 |
| KDE | *0.004* | *0.005* | **0.999** | *0.318* | *0.212* | 0.753 | **0.000** | **0.001** | **0.4** | 15.5 | 6,879 |
| AvgDis | 0.001 | 0.001 | **0.999** | 0.202 | 0.187 | **0.719** | **0.000** | **0.001** | 0.6 | **4.2** | *6,931* |
| FMFMGM | 0.063 | 0.079 | 0.772 | 0.001 | 0.002 | 0.979 | 0.105 | 0.095 | 23.7 | 22.7 | 6,931 |
| GeoBPR | 0.065 | 0.081 | 0.756 | 0.000 | 0.001 | 0.957 | 0.120 | 0.110 | 23.7 | 24.2 | 6,931 |
| IRenMF | *0.069* | 0.083 | *0.799* | 0.003 | 0.008 | 0.951 | *0.072* | *0.063* | 23.9 | 23.8 | 6,931 |
| PGN | 0.068 | *0.086* | 0.777 | *0.014* | *0.023* | *0.932* | 0.110 | 0.100 | *23.6* | *20.9* | **7,253** |
| Skyline | 0.784 | 0.996 | 0.982 | 0.231 | 0.087 | 0.796 | 0.000 | 0.000 | 17.5 | 18.8 | 7,241 |

Simultaneously, when measuring the distance (DistT and DistU), we observe that both Rnd and Pop algorithms obtain high values, showing us that the recommended venues of these models are far from each other. However, the geographical influence alone is not enough to obtain high values in terms of relevance, as evidenced by the poor performance of the pure geographical algorithms (AvgDis, KDE).

When analyzing the exposure metrics (ExpP, ExpR), the random recommender obtains lower values in terms of ExpP than most of the algorithms due to the fact that it recommends items in an arbitrary manner, without overrepresenting any subset of items. Similarly, this recommender obtains good results in the ExpR metric because it is recommending almost all the venues in the system, so it is probable that within those recommendations there are relevant venues. However, what the Rnd recommender fails is in recommending the relevant venues to the correct users, as discussed before regarding the accuracy metrics.

Hence, we conclude that most of the recommenders suffer from a great popularity bias, evidencing the difficulty of finding good representatives for all metrics. Therefore, among all the experimented recommenders, we consider IB and PGN to be of particular interest, since even though they do not perform as well in terms of accuracy as Pop, they obtain competitive results in terms of other metrics like novelty, diversity, and item exposure; this is a direct consequence of suffering less from the popularity bias.

## IV. CONCLUSIONS

This article examines various types of polarization affecting Point-Of-Interest recommendations, including popularity bias (for venues and categories), venue exposure, and geographical distance biases. Our results reveal that many state-of-the-art recommenders, including classical and POI-based, exhibit significant popularity bias, especially in sparse datasets. However, some classical algorithms, like the item-based approach, demonstrate decent relevance performance with less bias. We have utilized these "polarization free" algorithms to reduce recommendation bias while maintaining relevance. Despite our findings, we believe further improvements can be made in algorithm testing and polarization mitigation techniques.


## ACKNOWLEDGMENT

This work has been funded by the Ministerio de Ciencia e Innovación (reference PID2019-108965GB-I00), by the European Social Fund (ESF), within the 2017 call for predoctoral contracts, and by the project "AmerMad-CM: America in Madrid. Interconnected heritage and impact on tourism in the Community of Madrid" (ref: H2019/HUM-5694).


## REFERENCES


[1] P. Sánchez, A. Bellogín, and L. Boratto, "Bias characterization, assessment, and mitigation in location-based recommender systems," *Data Min. Knowl. Discov.*, vol. 37, no. 5, pp. 1885–1929, 2023. [Online]. Available: https://doi.org/10.1007/s10618-022-00913-5

[2] D. Massimo and F. Ricci, "Building effective recommender systems for tourists," *AI Magazine*, vol. 43, no. 2, pp. 209–224, 2022. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/aaai.12057

[3] J. Zhang and C. Chow, "Geosoca: Exploiting geographical, social and categorical correlations for point-of-interest recommendations," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval.* ACM, 2015, pp. 443–452.

[4] P. Sánchez and A. Bellogín, "Point-of-interest recommender systems based on location-based social networks: A survey from an experimental perspective," *ACM Comput. Surv.*, vol. 54, no. 11s, pp. 223:1–223:37, 2022. [Online]. Available: https://doi.org/10.1145/3510409