ECAI 2024
U. Endriss et al. (Eds.)

doi:10.3233/FAIA240852

Does Serendipity Enhance Recommendation Quality? Measuring Accuracy and Beyond-Accuracy Objectives of Serendipitous POI Suggestions

Imen Ben Sassi [a,*], Priit Järv[b] and Sadok Ben Yahia[b,c]

[a]LIRMM, Université de Montpellier, CNRS, Montpellier, France
[b]Tallinn University of Technology, Tallinn, Estonia
[c]University of Southern Denmark, Sønderborg, Denmark

Abstract. Point of Interest (POI) recommender systems (RSs) play a primary role in improving Location-based Social Networks' user experience. This paper studies the potential usefulness of serendipity in POI recommendations. We first introduce a new POI RS, called DISCOVERY, that attempts to improve the accuracy-serendipity trade-off. The proposed RS aims to recommend POIs that provide a pleasant surprise, allowing users to discover new venues known as serendipitous POIs. We then look closely at how serendipity affects the quality of POI suggestions by contrasting the outcomes of DISCOVERY with those of three cutting-edge non-serendipitous POI RSs. We use two real-world datasets—Foursquare and Flickr—along with a variety of metrics to test our ideas. These include (i) accuracy, which checks the precision, recall, and f-measure of Top-N recommendations; and (ii) beyond-accuracy, which checks the categorical and geographical diversity, explainability, and coverage in terms of POIs. The reported experimental observations show that serendipity boosts POI recommendation accuracy and favors geographically proximate and explainable POIs. However, standard POI baselines outperform DISCOVERY in terms of categorical diversity and coverage.

1 Introduction

With the emergence of Location-based Social Networks (LBSNs) such as Yelp and Foursquare, users can share their experiences via check-ins to Points of Interest (POIs) about locations they have visited, such as restaurants and tourist spots. The main task of POI Recommender Systems (RSs) is to propose interesting POIs to users to improve their experiences. This task is essential and helpful for residents and tourists to explore interesting locations in a city. It also offers the opportunity for POI owners to attract more potential visitors and increase their revenues. Traditionally, the effectiveness of RSs is judged by accuracy [15]. However, restricting the evaluation to accuracy metrics may cause the "filter bubble" phenomenon raised when users are trapped in limited options that are too similar to their profile [26, 6, 34]. Consequently, they can overlook any opportunity to explore more options, and the RS falls into the "overspecialization" or "lack of serendipity" problem. The original serendipity definition is "[...] *making discoveries, by accidents and sagacity, of things which they were not in the quest for* [...]" [33]. Serendipity defines the user's surprise when receiving a relevant recommendation, which means that the latter should be at the same time relevant to his interests and unexpected. Hence, serendipity guarantees accuracy and positive emotional reactions raised by the RS suggestion [5]. However, the benefits of serendipity for RSs are still unclear since it is tricky to measure its sought-after "surprising aspect". Thus, deciding whether the serendipity would lead us to higher users' satisfaction with the recommendations [18]. For that reason, the authors in [27] have studied how serendipitous suggestions enhance users' engagement and social perceptions of RSs. Their findings show that serendipitous recommendations enhance the rewarding experience of using a RS and promote users to attribute more sociable qualities to the RS. Unlike some scenarios where RSs need only to recommend relevant and familiar items, for example, social connection recommendations, serendipity is a primary element for recommending POIs. Motivated by the lack of a recommendation approach that considers the serendipity concept to suggest POIs, this paper introduces a new recommendation method that recommends POIs that provide a happy accident or pleasant surprise. In doing so, we allow the user to discover new venues called serendipitous POIs. Furthermore, we also pay heed to measuring the effect of serendipity on the quality of POI recommendations. Indeed, we compare the results of our serendipitous RS versus three state-of-the-art non-serendipitous POI RSs. The overarching research questions are as follows: (1) How do we measure the perception of serendipity in the context of POI RSs? (2) Can the suggestion of a non-obvious recommendation affect the accuracy of the RS? and (3) What is the impact of serendipity on beyond-accuracy objectives?

The rest of the paper is organized as follows. We start by reviewing POI RSs and the existing works in the field of serendipitous recommendations in Section 2. Then, we briefly summarize the preliminaries and the problem definition in Section 3. In Section 4, we thoroughly describe our serendipity-driven POI RS. Section 5 shows the experimental evaluation aiming to study the effect of serendipity on the quality of POI recommendations. Finally, Section 6 summarizes our work with a discussion of plans for future research directions.

2 Related Work

In this section, we scrutinize the recent research related to POI recommendation and serendipity-oriented algorithms.

* Corresponding Author. Email: imen.ben-sassi@lirmm.fr.

2.1 POI Recommendation Systems

Recently, POI RSs have played a primary role in improving LBSN users' experiences [30]. The authors in [20] have proposed a ranking based on the geographical factorization method, called Rank-GeoFM, which considers the geographical influence in recommending POIs. The critical contribution of this method is the use of both visited and unvisited POIs in learning the ranking function, which has contributed to alleviating the sparsity problem. The temporal and geographical contexts have been captured by the spatial-temporal prediction method using a recurrent neural network model [21]. The proposed model has demonstrated an exceptional ability to manage continuous data. The sequential data representation is based on the geographical distances between POIs and the time intervals between nearby behaviors. A geographical model, called LGLMF, was proposed in [28], which considers the users' central activity region and the popularity of locations in this region. Later, the authors came up with a new spatio-temporal activity-centers model, called STACP [29]. It takes into account geographical and temporal factors to model users' behavior and solve the problem of not having enough data. The authors of [23] suggested a recommendation model that uses a self-attentive encoder to figure out the nonlinear user-POI relationships and the level of user preference in a number of different areas. A new approach based on word2vec architecture was introduced in [8]. The proposed model, named POI2Vec, leverages the geographical influence of POIs to predict the potential visitors to a location in the next few hours by aggregating user preference and POI sequential influence. Recently, the authors in [10] introduced the DAN-SNR approach. The latter has incorporated the social influence of each user's friends to compute the following POI recommendation based on a deep attentive network for social-aware following POI recommendation. The authors in [9] have proposed RELINE, which embeds eight relational graphs into one shared latent space. The latter includes the geographical proximity, social and temporal, and user preference dynamics for making recommendations. The proposed system has demonstrated remarkable performance in both cold-start POIs and user issues. The authors in [19] have proposed a new tour RS that considers the personalization of the travel destination's characteristics. Based on his tour history, the proposed RS quantifies the user's personal preference for each travel-related aspect, including diversity, popularity, and distance. Finally, after summing the personalized scores for each aspect, it recommends the top-N POIs.

2.2 Serendipity in Recommendation Systems

The concept of serendipity is one of the primary vital aspects that has attracted researchers in the field of RSs [4, 38]. In the context of RSs, serendipity refers to the ability of a RS to suggest items that enable users to come across relevant but still pleasantly surprising items that they would not have discovered by themselves [5, 26]. Accordingly, a serendipitous recommendation may be defined by novel, unexpected, but still valuable and relevant items [17]. The relevance of an item is usually related to its closeness to the user profile [6]. Novelty occurs when a RS suggests an unknown item that the user could not discover autonomously [32]. However, the unexpectedness assessment is not apparent. The content-based approach proposed in [11] was the first to introduce serendipity in RSs. The authors applied a supervised learning method based on items' textual descriptions to predict the probability of relevance of an unseen item to a given user. We consider the items that the RS is uncertain about, meaning they are neither relevant nor irrelevant, as potentially serendipitous and suggest them as recommendations. In [2], a new approach to recommending unexpected items was introduced. The latter identifies expected items by the user and then derives unexpected ones. The expected items define items rated by the user and those similar to these latter. In [22], the authors proposed a new method for a serendipitous item recommendation, which makes the ranking sensitive to the popularity of negative examples. This method extends traditional personalized ranking methods by considering item popularity in AUC optimization to improve accuracy and serendipity.

It is of utmost importance to differentiate serendipity from diversity [7, 17]. Diversity applies primarily to sets of items and pertains to how dissimilar items are based on their properties, such as genres or locations, while serendipity is an item-level property involving the comparison of the recommended items with the user profile. Unlike diversity, serendipity requires relevance. Novelty, often confused with serendipity, includes unpopular and dissimilar items but lacks the necessity of relevance. Serendipitous items are both relevant and unexpected, distinguishing them significantly from items a user rates in a RS, a distinction that novel items may not necessarily share.

To sum up, mainly accuracy metrics have been considered to generate or evaluate RSs by assessing how well the output of a RS matches a proportion of known withheld items. Since accuracy alone cannot guarantee satisfactory recommendations, beyond-accuracy concepts were employed to evaluate the recommendation results. It is worth noting that only the authors in [22] introduced a serendipitous-based ranking method. The latter investigates the relationship between accuracy and serendipity by taking into account item popularity. However, to the best of the authors' knowledge, no previously published research in the POI recommendation domain takes advantage of the serendipity concept to recommend top-N venues.

3 Preliminaries and Problem Definition

We define the POI recommendation problem and its key components.

Definition 1. ***(User)*** *is a unique user $u \in U$ registered on the LSBN and described by the collection of all his check-ins.*

Definition 2. ***(POI)*** *is a unique location $l \in L$ in which users checked in and represented as: $< l_{id}, long, lat, categ >$.*

Definition 3. ***(Check-in)*** *is a record that represents a user u checks-in at a location l at the timestamp t, represented as the triplet: $c =< u, l, t >$, where $u \in U$, $l \in L$, and t is a timestamp. Each check-in can be performed only by one user, but the same user may have multiple check-ins.*

Definition 4. ***(Neighbor$_u$)*** *is the set of the most similar (nearest) users to the user u, defined by users who mostly visited the POIs visited by u. The threshold of the number of commonly visited locations is fixed by the parameter α.*

Definition 5. ***(Serendipity-based POI recommendation problem)*** *given a user u with a check-in location history L_u, predict the top-N serendipitous POIs—those that are simultaneously relevant, unexpected, and novel—that the user would likely want to visit, excluding those already in L_u.*

4 DISCOVERY: a serendipity-based POI recommender system

This section introduces DISCOVERY, a serendipity-based POI RS, which consists of recommending locations that provide a pleasant surprise, allowing the user to discover new venues.

4.1 Computing Relevance

Relevance is a user-specific notion that defines users' interest in items [32]. Relevance can be measured using multiple judgments [17]. In the remainder, we define a POI as relevant to a user u if a user $v \in Neighbor_u$ has visited this location. Given a list of visited POIs by u denoted L_u and a neighborhood of u denoted $Neighbor_u$, the relevant POIs to u are computed as shown in Eq. 1.

$$Relevant\_POI(u) = \bigcup_{v \in Neighbor_u} L_v \ \setminus \ L_u \quad (1)$$

4.2 Computing Unexpectedness

Unexpectedness reflects how dissimilar a suggested item is to a user profile [14]. As we look at POI RS (c.f. Eq. 2), we define unexpectedness by the point-wise mutual information (PMI) that shows how similar two places are based on the number of users who have been to both places and each place separately.

$$PMI(i,j) = log_2 \frac{p(i,j)}{p(i)p(j)} \ / \ -log_2 \ p(i,j) \quad (2)$$

where $p(i)$ is the probability that any user has visited the location i and $p(i,j)$ represent the likelihood of a user visiting locations i and j simultaneously. Thus, PMI ranges from -1 to 1, where -1 indicates that two locations are never visited together, while 1 indicates that two locations are always visited together. To assess the level of unexpectedness of a location l to a user u, we compute its PMI with each visited location in L_u. Then, we calculate the average of the PMI value across the locations from the user's visits L_u (c.f. Eq. 4). Based on the PMI, the list of unexpected locations for a user u is defined by Eq. 3, where L_u is the user's history of visits (i.e. the set of visited locations). The threshold of 0 is chosen to include all locations that the user is more likely to not visit, than to visit.

$$Unexpected\_POI(u) = \bigcup_{i \in L} Unexp^{avg}_{co-occ}(i,u) \mid Unexp^{avg}_{co-occ}(i,u) < 0 \quad (3)$$

$$Unexp^{avg}_{co-occ}(i,u) = \frac{1}{|L_u|} \sum_{j \in L_u} PMI(i,j) \ ; \ where \ i \in L \quad (4)$$

4.3 Computing Novelty

The novelty is defined in [32] through the distance between an item and a user's consumption. In the remainder, the term novel POI stands for a relevant (or even irrelevant) POI that a user has never visited in his life [16]. Therefore, we compute a POI's novelty based on its semantic distance from all the locations a user has already visited. The novelty score ranges from 0 to 1, where a value near to 0 indicates that the POI lacks novelty. As detailed in Eq. 6, to decide whether a location $i \in L$ is novel for a user u, we compute its distance to all the locations in the user's history of visits $j \in L_u$ and take the average value as the overall novelty value. Finally, i is considered novel whenever the obtained average distance is higher than 0.5 (c.f. Eq. 5).

$$Novel\_POI(u) = \bigcup_{i \in L} nov^{dist}(i,u) \mid nov^{dist}(i,u) >= 0.5 \quad (5)$$

$$nov^{dist}(i,u) = \frac{1}{|L_u|} \sum_{j \in L_u} dist(i,j) \quad (6)$$

where $dist(i,j)$ indicates the distance between locations i and j, and is formalized as detailed in Eq. 7.

$$dist(i,j) = 1 - sim(i,j) \quad (7)$$

$$sim(i,j) = similarity_{spaCy}(category_i, category_j) \quad (8)$$

where $sim(i,j)$ is a semantic similarity between locations i and j ($sim(i,j) \in [0,1]$). Below, we figure out how similar two words are based on their location category using the spaCy library (see Eq. 8). Here, $category_i$ and $category_j$ are the LSBN categories of locations i and j, respectively.

4.4 Serendipity-based Algorithm for POI Recommendations

We refer to serendipity as the ability of a RS to suggest serendipitous POIs and two variations for serendipity are used: *(i)* $Serendipitous\_POI^{v1}$ that considers only the relevance and the unexpectedness aspects to define the serendipity [24, 14] (c.f. Eq. 9); and *(ii)* $Serendipitous\_POI^{v2}$ where serendipitous POIs are relevant, novel, and unexpected [31, 11] (c.f. Eq. 10).

$$Serendipitous\_POI^{v1}(u) = Relevant\_POI(u) \cap Unexpected\_POI(u) \quad (9)$$

$$Serendipitous\_POI^{v2}(u) = Relevant\_POI(u) \cap Unexpected\_POI(u) \cap Novel\_POI(u) \quad (10)$$

We select the top-N serendipitous POIs from the intersection of the computed POI lists, prioritizing those with the highest visit counts, to help the user discover the most serendipitous iconic places in a city.

5 Experimental Study

We carried out several experiments to provide evidence of the potential usefulness of leveraging serendipity in POI RSs. To do so, we compared the top-N recommendation list generated by the DISCOVERY algorithm versus those lists suggested by three other non-serendipitous baselines.

5.1 Datasets Description

We use two real-world datasets from Foursquare[1] [36] and Flickr [12]. Their main characteristics are presented in Table 1.

Table 1: Dataset statistics.

Dataset	LBSN	Users	POIs	Check-ins	Categories
Tallinn	Flickr	1,911	1,054	12,413	13
New York	Foursquare	1,083	38,333	227,428	250

The Foursquare dataset, with a sparsity of 99.45%, includes check-in data in a big city, i.e. New York City (USA). Each check-in is defined by: "User ID", "Venue ID", "Venue category ID", "Latitude", "Longitude", and "Timestamp". The Flickr dataset is collected in a small city, i.e. the city of Tallinn (Estonia) and has a sparsity of 99.38%. The dataset provides user tours (check-in sequences) for different POIs, where each check-in is defined by: "user ID", "POI ID", and "Timestamp" and each POI is described by: "Latitude", "longitude", "view", "category", "image IDs", "tags", and "region". We partition each dataset into training data (the earliest 70% check-ins) and test data (the most recent 30%) for each user.

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

5.2 Evaluation Metrics

To assess the impact of serendipity on the quality of the recommendations, we use metrics to assess the accuracy, including Precision@N, Recall@N, and F-measure@N, as well as beyond-accuracy metrics, including Coverage@N, ILDGeo@N, Diversity@N, and Fidelity@N, with $N \in \{5, 10, 15, 20\}$. It is worth mentioning that none of these evaluation metrics capture serendipity, since our model primarily focuses on this concept to generate recommendations. Thus, it does not make sense to recompute it to compare our model versus the other non-serendipitous baselines.

5.2.1 Accuracy Evaluation

(1) Precision is defined by the number of relevant recommendations divided by the total number of items recommended by the RS. The precision is defined by Eq. 11, where $Rec\_List_u^N$ refers to the list of recommended items $@N$ to the user u and Rel_List_u defines the list of relevant items $@N$ to u.

$$Precision@N = \frac{1}{|U|}\sum_{u \in U}\frac{|Rec\_List_u^N \cap Rel\_List_u|}{|Rec\_List_u|} \tag{11}$$

(2) Recall is defined by the number of relevant items recommended by the RS divided by the number of existing relevant recommendations. The Recall is defined by Eq. 12.

$$Recall@N = \frac{1}{|U|}\sum_{u \in U}\frac{|Rec\_List_u^N \cap Rel\_List_u|}{|Rel\_List_u|} \tag{12}$$

(3) F-measure is defined as the harmonic mean of precision and recall. The F-measure is defined in Eq. 13.

$$F-measure@N = 2 * \frac{Precision@N * Recall@N}{Precision@N + Recall@N} \tag{13}$$

5.2.2 Beyond-Accuracy Evaluation

(1) Coverage refers to item coverage and measures the degree to which the recommendations cover the catalog of available items [15]. In the remainder, as detailed in Eq. 14, the coverage is defined by the fraction of POIs appearing in the users' recommendation lists, where $Rec\_List_u^N$ is the set of top N recommendations generated for a user u, U is the set of all users, and L is the POI catalog.

$$Coverage@N = \frac{|\bigcup_{u \in U} Rec\_List_u^N|}{|L|} \tag{14}$$

(2) ILDGeo defines the Average Intra-List Distance [4] and measures how different two POIs in the recommendation list are on average [13] (c.f. Eq. 15). In this paper, we are interested in evaluating the geographical diversity of the generated recommendations. It is worth mentioning that the lower the ILDGeo value, the closer the recommended POIs are. Therefore, the user's journey is more organized, and his experience is better. As shown in Eq. 16, dissimilarity is defined by the distance in kilometers between two POIs i and j and denoted $kmDistance$, where $kmDistance$ is the kilometer distance between two POIs computed using their longitude and latitude.

$$ILDGeo_u@N = \frac{\sum_{i,j \in Rec\_List_u^N \ \wedge \ i \neq j} dissim(i,j)}{|Rec\_List_u^N| * (|Rec\_List_u^N| - 1)} \tag{15}$$

$$dissim(i,j) = kmDistance(loc_i, loc_j) \tag{16}$$

(3) Diversity measures the categorical diversity of the recommended POIs [13]. The DivCat metric computes the number of unique categories included in the recommendation list (c.f. Eq. 17), where Cat_i is the category to which the POI i belongs.

$$DivCat_u@N = |\bigcup_{i \in Rec\_List_u^N} Cat_i| \tag{17}$$

(4) Fidelity refers to the user-based neighbor style explanation, which shows how a user's neighbors rated the recommended item [1]. Fidelity measures the percentage of explainable POIs in the Top-N recommendations and aims to ensure more transparency, trustworthiness, and persuasiveness. By explainable POIs, we mean those determined based on the visits of similar users to the recommended POI (c.f. Eq. 19), where $Neighbor_u$ refers to the users who mostly visited the POIs visited by u, and L_v refers to the locations visited by each user v in $Neighbor_u$.

$$Fidelity_u@N = \frac{|RecList_u^N \cap Explainable_u|}{|Rec\_List_u^N|} \tag{18}$$

$$Explainable_u = \bigcup_{v \in Neighbor_u} L_v \tag{19}$$

5.3 Baseline Models

We compared the proposed approach with state-of-the-art POI recommendation approaches. As discussed in Section 2, there is a dearth of works that focus on serendipity. Furthermore, our primary goal is to show the effect of serendipity in POI recommendations by comparing our serendipity-driven approach versus other baselines that do not heed this concept. The details of the compared methods are listed below:

(1) STACP[2] [29]: a Spatio-temporal algorithm that extracts the users' mobility patterns according to the historical check-in centers of activity depending on their current temporal state.

(2) LGLMF[3] [28]: uses a logistic matrix factorization to formulate the generated users' main regions of activities by a local geographical model to recommend similar checked-in POIs in the zone of each user's activity.

(3) Rank-GeoFM[4] [20]: a ranking-based matrix factorization model that includes the geographical influence of neighboring POIs while learning user preference rankings for POIs.

(4) Versions of DISCOVERY: to evaluate the impact of the novelty, we use two versions of our model:

(a) DISCOVERY$_{v1}$ is a simplified version of DISCOVERY, which only considers the relevance and unexpectedness of POIs to generate recommendations (c.f. Eq. 9).

(b) DISCOVERY$_{v2}$ enriches the previous model with information on the novelty of POIs (c.f. Eq. 10).

5.4 Comparison with Traditional Techniques

In the first series of experiments, we assess the usefulness of serendipity by comparing DISCOVERY versus the non-serendipitous-driven baseline methods in terms of accuracy. Next, we conduct a comparison from a beyond-accuracy perspective. Finally, we discuss the impact of novelty on serendipitous item recommendation results.

---

[2] https://github.com/rahmanidashti/STACP
[3] https://github.com/rahmanidashti/LGLMF
[4] https://github.com/dbgroup-uestc/cuiyue

5.4.1 Accuracy-based comparison

In terms of accuracy, our serendipity-driven RS works much better than methods that either a) learn users' preferences and then use geography to affect them, like Rank-GeoFM; b) look at users' mobility patterns based on historical check-in centers of activity, like LGLMF; or c) make users' central regions of activity, like STACP. This is because we look at more information.

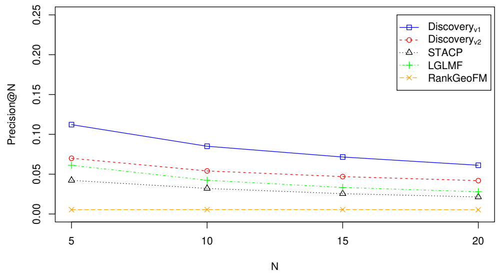

Figure 1: Precision values comparison for $N = 5, 10, 15$, and 20 for Tallinn dataset.

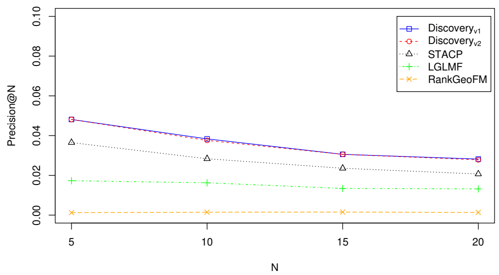

Figure 2: Precision values comparison for $N = 5, 10, 15$, and 20 for New York dataset.

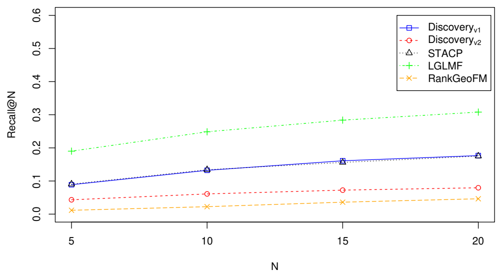

Figure 3: Recall values comparison for $N = 5, 10, 15$, and 20 for Tallinn dataset.

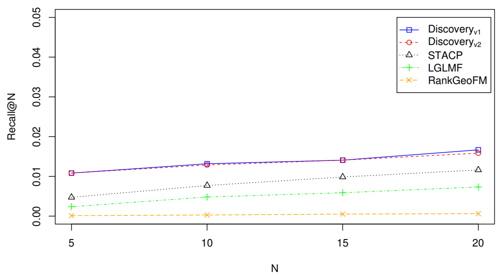

Figure 4: Recall values comparison for $N = 5, 10, 15$, and 20 for New York dataset.

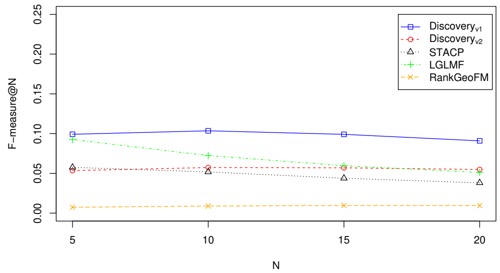

Figure 5: F-measure values comparison for $N = 5, 10, 15$, and 20 for Tallinn dataset.

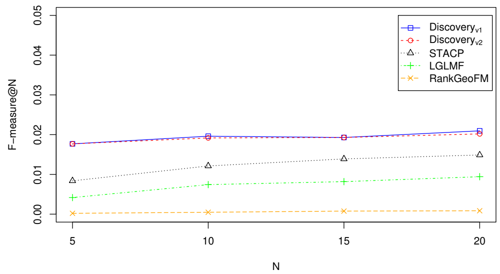

Figure 6: F-measure values comparison for $N = 5, 10, 15$, and 20 for New York dataset.

5.4.2 Beyond-accuracy-based comparison

We look closely at how well serendipity works in POI RSs by comparing and contrasting DISCOVERY with other common methods using the beyond-accuracy metrics. As for the coverage and the diversity, they are shown, respectively, in Figures 7–8 and Figures 9–10. Both RankGeoFM and LGMF achieve high results, which means that their recommendation results cover a high fraction of all POIs and POI categories, especially with the Flickr dataset having a limited number of POIs and categories compared to the Foursquare dataset. At the same time, STACP achieves much lower coverage and diversity than other methods. Our method's results show that the use of serendipity has led to limited diversity and coverage levels. These findings were quite expectable since both versions of the DISCOVERY algorithm deprioritize any POI similar to the previous check-ins.

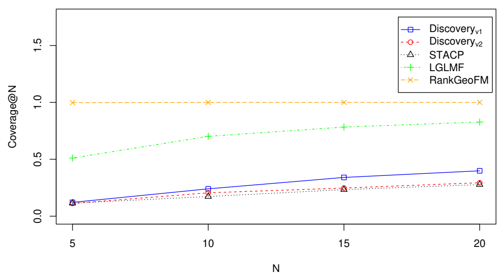

Figure 7: Coverage values comparison for $N = 5, 10, 15$, and 20 for Tallinn dataset.

As for the geographical diversity expressed by the IDLGeo measure, the validation of the obtained results depends on the main goal of the RS. Suppose we deal with an eco-friendly RS that prioritizes the recommendation of geographically close spots to minimize the

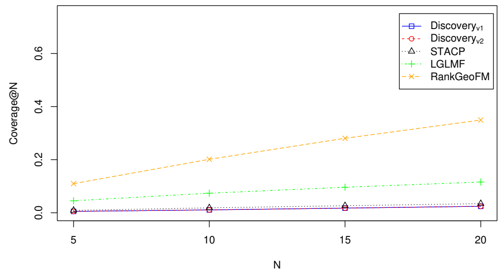

Figure 8: Coverage values comparison for $N = 5, 10, 15$, and 20 for New York dataset.

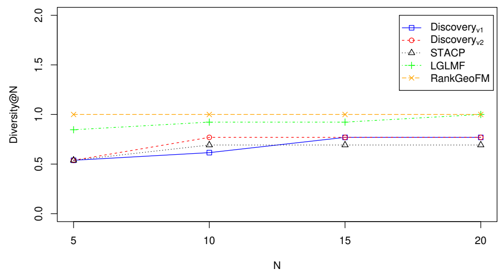

Figure 9: Diversity values comparison for $N = 5, 10, 15$, and 20 for Tallinn dataset.

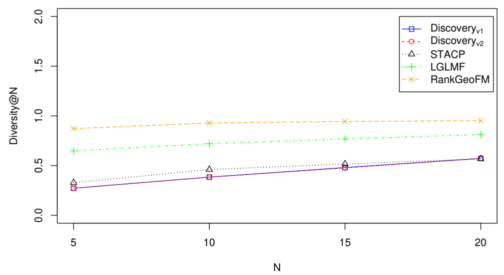

Figure 10: Diversity values comparison for $N = 5, 10, 15$, and 20 for New York dataset.

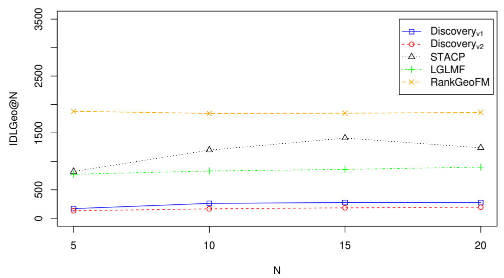

Figure 11: IDLGeo values comparison for $N = 5, 10, 15$, and 20 for Tallinn dataset.

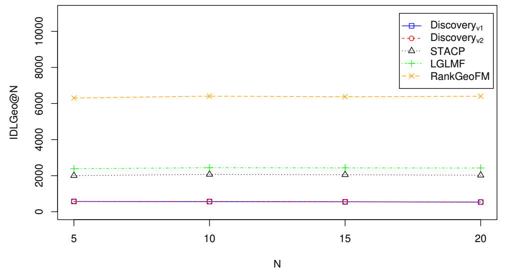

Figure 12: IDLGeo values comparison for $N = 5, 10, 15$, and 20 for New York dataset.

use of transport and promote green transportation. In that case, we can say that our method presents the most promising results, and that our RS would present a pleasing addition to an urban mobility hub for promoting smart city planning and making cities smarter and more sustainable (c.f. Figures 11 and 12). Besides, a tourist may prefer receiving recommendation lists, where each one contains geographically close spots for better organizing his journey, especially when visiting vast cities such as New York. Figure 12 shows that the lowest IDLGeo values have been obtained with DISCOVERY$_{v1}$ and DISCOVERY$_{v2}$. The gap between DISCOVERY$_{v1}$ and the baselines is more remarkable with small-sized recommendation lists IDLGeo@5 was improved by 356.86%, 385.64%, and $1,010.82\%$ compared to LGLMF, STACP, and Rank-GeoFM. Otherwise, if the RS recommends geographically distant POIs, our results are considered the worst compared to traditional baselines, and serendipity negatively affects this metric.

Both versions of DISCOVERY achieve the highest explainability in terms of fidelity. Indeed, we adopt a neighbor-based collaborate filtering approach to computing item relevance (c.f. Figures 13 and 14), which is similar to the definition of explainability of neighbor style proposed in [1]. For instance, in terms of explanability@5, DISCOVERY maintained an explainability of 100%, while other baselines varied between 23.12% and 73.44% for the Tallinn dataset, and 0.82% and 32.78% for the New York dataset. By and large, it comes out that serendipity has improved the transparency of the RS by justifying recommendations to enhance the user's trust in the suggestions.

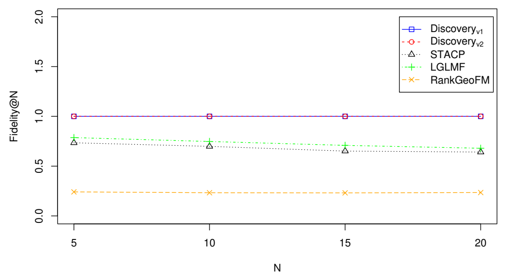

Figure 13: Fidelity values comparison for $N = 5, 10, 15$, and 20 for Tallinn dataset.

5.5 *Impact assessment of Novelty*

To assess the impact of incorporating novelty into the serendipity definition, we compare the results obtained with DISCOVERY$_{v1}$ versus those obtained with DISCOVERY$_{v2}$. In our experiments, we vary the size of the recommendation list N and measure the quality of the two versions of our model.

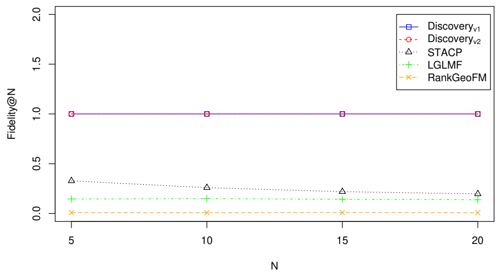

Figure 14: Fidelity values comparison for $N = 5, 10, 15$, and 20 for New York dataset.

The novelty has considerably degraded the accuracy-based results on the Tallinn dataset. We can see that the f-measure@20 decreases by 39.62% when considering the novelty of items to be recommended (c.f. Figure 5). However, for the New York dataset, the difference between the two versions, in terms of f-measure@20, is not as noteworthy and is limited to 3.67% (c.f. Figure 6). Given that the computation of novelty relies on the semantic similarity of POI categories, we can infer that incorporating novelty into the serendipity definition could enhance recommendation accuracy in datasets with a sufficient number of POI categories, such as the New York dataset. The same points are made about coverage and diversity when DISCOVERY$_{v2}$ does not work as well as DISCOVERY$_{v1}$, especially on the Tallinn dataset (see Figures 7 and 9), when the system may exhibit a bias towards certain locations simply because they belong to different categories. A more fine-grained semantic similarity calculation, which allows places in the same category to be semantically distant, could help to mitigate the negative impact of nevolty. However, the integration of the novelty has decreased the IDLGeo results and given rise to lists of elements geographically closer. For instance, for IDLGeo@20 based on the Tallinn dataset, the geographical distance between Top-20 POIs has been reduced by 29.93% (c.f. Figure 11).

5.6 *Parameter Tuning*

In this sub-section, we study the importance of parameter tuning. We examine the impact of the user's neighborhood size, defined by the parameter α, on the performance of our model for both Tallinn and New York datasets. For the Tallinn dataset (c.f. Figure 15), we set the parameter α to a value between 1 and 8. For the New York dataset (c.f. Figure 16), we set it to a value between 4 and 12. In all of our experiments, there is a peak point where our model gets the best performance. Our goal is to investigate to what extent common behaviors can motivate users to visit a new location. Thus, the regularization parameter α defines the importance of social typical behavior influence. The main disadvantage of the precision-and-recall setting is that precision typically decreases as N increases, whereas recall increases as N increases. As a result, we used the f-measure to calculate a trade-off between precision and recall and select the optimal value of α. As underscored in Figures 15 and 16, the optimal value of the parameter α is not the same for both datasets.

Due to space limitations, this paper presents figures only for $N = 20$. Nonetheless, it should be mentioned that the results obtained for other N values, i.e. $N \in \{5, 10, 15\}$, mirror those observed for $N = 20$. For the Tallinn Flickr dataset, the highest f-measure values were obtained with a user's neighborhood equal to 5 for $N \in \{5, 10, 15, 20\}$. For the New York Foursquare dataset, the highest f-measure result was reached with a neighborhood value equal to 10 for $N \in \{5, 10, 15, 20\}$.

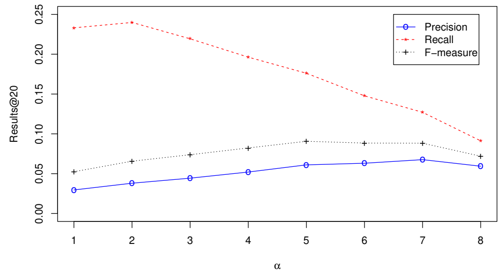

Figure 15: The impact of the tuning parameter α on recommendation@20 precision, recall, and f-measure for Tallinn dataset.

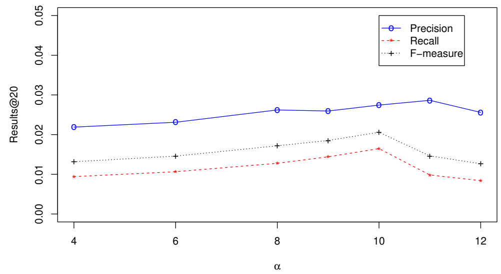

Figure 16: The impact of the tuning parameter α on recommendation@20 precision, recall, and f-measure for New York dataset.

6 Conclusion

This paper studied the usefulness of serendipity in POI recommendations. We first built a new POI RS called DISCOVERY, where the main idea was to shed light on the overlooked spots. We further developed two versions of our method based on two serendipity definitions to illustrate the effect of novelty on the quality of the recommendations. Our findings show that novelty has an adverse impact on almost all experiments. Then, we compared the results of our serendipitous-driven RS versus those obtained with traditional POI RSs. Using the Foursquare and Flickr datasets, the experimental comparison demonstrated a positive impact of serendipity on the accuracy of POI recommendations. It also supports the recommendation of geographically close and explainable POIs. Nevertheless, the standard POI baselines have achieved the best categorical diversity and coverage results.

Despite the encouraging outcomes of DISCOVERY, we are fully aware that there is still room for improvement. A valuable direction is to incorporate rich item features (e.g. payment method and service) into the distance definition used to compute the novelty of the POIs instead of just taking into account item categories. A variety of factors [25], such as temporal and environmental influences like weather, user companions, or the peak season period of the POI, may be considered in the recommendation process. Additionally, we intend to conduct a user-based study and compare our approach with large language models for recommendations [3, 35]. This will enable us to assess the genuine perception of the recommendations as serendipitous. We also plan to validate the generalizability of our results by utilizing the global-scale check-in dataset [37] that encompasses data from 415 cities. This dataset offers a broad spectrum of user behaviors and diverse POI characteristics—including categories, city types, country weather, and more. By leveraging this variety, we aim to gain a deeper understanding of DISCOVERY's strengths and limitations across different scenarios.

References

[1] B. Abdollahi and O. Nasraoui. Using explainability for constrained matrix factorization. In *Proceedings of the Conference on Recommender Systems*, RecSys '17, page 79–83, New York, NY, USA, 2017. ACM. ISBN 9781450346528.

[2] P. Adamopoulos and A. Tuzhilin. On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Trans. Intell. Syst. Technol.*, 5(4), 2014. ISSN 2157-6904. doi: 10.1145/2559952.

[3] K. Bao, J. Zhang, Y. Zhang, W. Wenjie, F. Feng, and X. He. Large language models for recommendation: Progresses and future directions. In *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*, SIGIR-AP '23, page 306–309, New York, NY, USA, 2023. ACM. ISBN 9798400704086. doi: 10.1145/3624918.3629550.

[4] P. Castells, N. J. Hurley, and S. Vargas. *Novelty and Diversity in Recommender Systems*, pages 881–918. Springer US, Boston, MA, 2015. ISBN 978-1-4899-7637-6.

[5] L. Chen, Y. Yang, N. Wang, K. Yang, and Q. Yuan. How serendipity improves user satisfaction with recommendations? a large-scale user evaluation. In *The World Wide Web Conference*, WWW '19, page 240–250, New York, NY, USA, 2019. ACM. ISBN 9781450366748.

[6] M. de Gemmis, L. Pasquale, G. Semeraro, and C. Musto. An investigation on the serendipity problem in recommender systems. *Information Processing & Management*, 51(5):695–717, 2015. ISSN 0306-4573. doi: https://doi.org/10.1016/j.ipm.2015.06.008.

[7] T. Duricic, D. Kowald, E. Lacic, and E. Lex. Beyond-accuracy: a review on diversity, serendipity, and fairness in recommender systems based on graph neural networks. *Frontiers in Big Data*, 6, 2023. ISSN 2624-909X. doi: 10.3389/fdata.2023.1251072.

[8] S. Feng, G. Cong, B. An, and Y. M. Chee. Poi2vec: Geographical latent representation for predicting future visitors. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, page 102–108, San Francisco, California, USA, 2017. AAAI Press.

[9] C. Giannis, P. Kefalas, A. N. Papadopoulos, and Y. Manolopoulos. RELINE: point-of-interest recommendations using multiple network embeddings. *Knowl. Inf. Syst.*, 63(4):791–817, 2021. doi: 10.1007/s10115-020-01541-5.

[10] L. Huang, Y. Ma, Y. Liu, and K. He. Dan-snr: A deep attentive network for social-aware next point-of-interest recommendation. *ACM Trans. Internet Technol.*, 21(1), 2020. ISSN 1533-5399. doi: 10.1145/3430504.

[11] L. Iaquinta, M. d. Gemmis, P. Lops, G. Semeraro, M. Filannino, and P. Molino. Introducing serendipity in a content-based recommender system. In *Proceedings of the 8th International Conference on Hybrid Intelligent Systems*, HIS '08, page 168–173, USA, 2008. IEEE Computer Society. ISBN 9780769533261.

[12] P. Järv. *Place Recommendation with Geo-tagged Photos*. PhD thesis, Tallinn University of Technology, 2020.

[13] H. Jiao, M. Fan, and H. Yamana. Evaluation of poi recommendation system beyond accuracy: Diversity, explainability and computation cost. In *Proceedings of the 12th Forum on Data Engineering and Information Management*, DEIM '20, Tokyo, Japan, 2020. C4-4.

[14] M. Kaminskas and D. Bridge. Measuring surprise in recommender systems. In *Proceedings of the Workshop on recommender systems evaluation: dimensions and design held in conjunction with RecSys 2014*, REDD '14, Silicon Valley, USA, 2014. ACM.

[15] M. Kaminskas and D. Bridge. Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Trans. Interact. Intell. Syst.*, 7(1), Dec. 2016. ISSN 2160-6455. doi: 10.1145/2926720.

[16] K. Kapoor, V. Kumar, L. Terveen, J. A. Konstan, and P. Schrater. "i like to explore sometimes": Adapting to dynamic user novelty preferences. In *Proceedings of the 9th ACM Conference on Recommender Systems*, RecSys '15, page 19–26, New York, NY, USA, 2015. ACM. ISBN 9781450336925.

[17] D. Kotkov, S. Wang, and J. Veijalainen. A survey of serendipity in recommender systems. *Knowl.-Based Syst.*, 111:180–192, 2016. ISSN 0950-7051.

[18] D. Kotkov, J. A. Konstan, Q. Zhao, and J. Veijalainen. Investigating serendipity in recommender systems based on real user feedback. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, SAC '18, page 1341–1350, New York, NY, USA, 2018. ACM. ISBN 9781450351911.

[19] J. Lee, S. J. A., D.-K. Chae, and S.-C. Lee. Personalized tour recommendation via analyzing user tastes for travel distance, diversity and popularity. *Electronics*, 11(7):1120, 2022. doi: 10.3390/electronics11071120.

[20] X. Li, G. Cong, X.-L. Li, T.-A. N. Pham, and S. Krishnaswamy. Rank-geofm: A ranking based geographical factorization method for point of interest recommendation. In *Proceedings of the 38th Conference on Research and Development in Information Retrieval*, SIGIR '15, page 433–442, New York, NY, USA, 2015. ACM. ISBN 9781450336215.

[21] Q. Liu, S. Wu, L. Wang, and T. Tan. Predicting the next location: A recurrent model with spatial and temporal contexts. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, page 194–200, Phoenix, Arizona, 2016. AAAI Press.

[22] Q. Lu, T. Chen, W. Zhang, D. Yang, and Y. Yu. Serendipitous personalized ranking for top-n recommendation. In *Proceedings of the IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, volume 1, pages 258–265, Macau, China, 2012. IEEE.

[23] C. Ma, Y. Zhang, Q. Wang, and X. Liu. Point-of-interest recommendation: Exploiting self-attentive autoencoders with neighbor-aware influence. In *Proceedings of the Conference on Information and Knowledge Management*, page 697–706, New York, NY, USA, 2018. ISBN 9781450360142.

[24] A. Maksai, F. Garcin, and B. Faltings. Predicting online performance of news recommender systems through richer evaluation metrics. In *Proceedings of the Conference on Recommender Systems*, RecSys '15, page 179–186, New York, NY, USA, 2015. ACM. ISBN 9781450336925.

[25] D. Massimo and F. Ricci. Popularity, novelty and relevance in point of interest recommendation: an experimental analysis. *Information Technology & Tourism*, 23:473 – 508, 2021. doi: 10.1007/s40558-021-00214-5.

[26] C. Matt, A. Benlian, T. Hess, and C. Weiß. Escaping from the filter bubble? the effects of novelty and serendipity on users' evaluations of online recommendations. In *Proceedings of the International Conference on Information Systems*, page 1–18, Auckland, New Zealand, 2014.

[27] J. Oh, S. Sudarshan, J. A. Lee, and N. Yu. Serendipity enhances user engagement and sociality perception: the combinatory effect of serendipitous movie suggestions and user motivations. *Behaviour & Information Technology*, 0(0):1–18, 2021.

[28] H. A. Rahmani, M. Aliannejadi, S. Ahmadian, M. Baratchi, M. Afsharchi, and F. Crestani. LGLMF: local geographical based logistic matrix factorization model for POI recommendation. In *Proceedings of the 15th Asia Information Retrieval Societies Conference*, volume 12004 of *AIRS '19*, pages 66–78, Hong Kong, China, 2019. Springer.

[29] H. A. Rahmani, M. Aliannejadi, M. Baratchi, and F. Crestani. Joint geographical and temporal modeling based on matrix factorization for point-of-interest recommendation. In *Proceedings of the European Conference on Information Retrieval*, pages 205–219, Lisbon, Portugal, 2020. ISBN 978-3-030-45439-5.

[30] 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.*, 54(11s), sep 2022. ISSN 0360-0300. doi: 10.1145/3510409.

[31] G. Shani and A. Gunawardana. Evaluating recommendation systems. In F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors, *Recommender Systems Handbook*, pages 257–297. Springer, Boston, MA, 2011. ISBN 978-0-387-85819-7.

[32] S. Vargas and P. Castells. Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys '11, page 109–116, New York, NY, USA, 2011. ACM. ISBN 9781450306836.

[33] H. Walpole. To mann, monday 18 january 1754. *Horace Walpole's Correspondence*, 20:407–411, 1960.

[34] N. Wang and L. Chen. User bias in beyond-accuracy measurement of recommendation algorithms. In *Proceedings of the Fifteenth ACM Conference on Recommender Systems*, RecSys '21, page 133–142, New York, NY, USA, 2021. ACM. ISBN 9781450384582.

[35] T. Wang and C. Wang. Embracing llms for point-of-interest recommendations. *IEEE Intelligent Systems*, 39(01):56–59, jan 2024. ISSN 1941-1294. doi: 10.1109/MIS.2023.3343489.

[36] D. Yang, D. Zhang, V. W. Zheng, and Z. Yu. Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(1): 129–142, 2015. ISSN 2168-2216.

[37] D. Yang, D. Zhang, and B. Qu. Participatory cultural mapping based on collective behavior data in location-based social networks. *ACM Trans. Intell. Syst. Technol.*, 7(3), jan 2016. ISSN 2157-6904. doi: 10.1145/2814575.

[38] Y. C. Zhang, D. O. Séaghdha, D. Quercia, and T. Jambor. Auralist: Introducing serendipity into music recommendation. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, page 13–22, New York, NY, USA, 2012. ACM. ISBN 9781450307475. doi: 10.1145/2124295.2124300.