



Article Indoor Trajectory Prediction for Shopping Mall via **Sequential Similarity**

Peng Wang D, Jing Yang * and Jianpei Zhang

College of Computer Science and Technology, Harbin Engineering University, Harbin 150009, China; peng.wang0110@gmail.com (P.W.); zhangjianpei@hrbeu.edu.cn (J.Z.) * Correspondence: yangjing@hrbeu.edu.cn

Abstract: With the prevalence of smartphones and the maturation of indoor positioning techniques, predicting the movement of a large number of customers in indoor environments has become a promising and challenging line of research in recent years. While most of the current predicting approaches that take advantage of mathematical methods perform well in outdoor settings, they exhibit poor performance in indoor environments. To solve this problem, in this study, a sequential similarity-based prediction approach which combines the spatial and semantic contexts into a unified framework is proposed. We first present a revised Longest Common Sub-Sequence (LCSS) algorithm to compute the spatial similarity of the indoor trajectories, and then a novel algorithm considering the indoor semantic R-tree is proposed to compute the semantic similarities; after this, a unified algorithm is considered to group the trajectories, and then the clustered trajectories are used to train the prediction models. Extensive performance evaluations were carried out on a real-world dataset collected from a large shopping mall to validate the performance of our proposed method. The results show that our approach markedly outperforms the baseline methods and can be used in real-world scenarios.

check for updates

Citation: Wang, P.; Yang, J.; Zhang, J. Indoor Trajectory Prediction for Shopping Mall via Sequential Similarity. Information 2022, 13, 158. https://doi.org/10.3390/info13030158

Academic Editor: Gholamreza Anbarjafari (Shahab)

Received: 4 March 2022 Accepted: 15 March 2022 Published: 19 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

Keywords: indoor trajectory; spatial similarity; semantic similarity; trajectory clustering; indoor location prediction

1. Introduction

While a large number of studies focusing on outdoor trajectory prediction have been carried out [1], research has shown that people tend to spend over 87% of their lifetime in indoor environments such as dwellings, grocery stores, airplane terminals and conference offices [2,3]. However, few studies have performed detailed analyses of the indoor trajectory prediction problem which is crucial to the location-based services for human beings.

One of the main reasons is that the indoor positioning technique is immature [4], which leads to an insufficient amount of appropriate experimental data. The Global Positioning System (GPS) technology which is efficient in outdoor scenarios cannot accurately identify the whereabouts of a user in an indoor environment due to severe signal blocking and the complicated multi-path effects, leading to the localization performance declining greatly [5]. Different from GPS, the indoor positioning products such as Wi-Fi-based devices can only record users' positions in the range of activation, and users' precise locations cannot be captured if users walk away from the activation range. Because it is impractical to deploy positioning devices everywhere, it is hard to track objects at any place in the indoor environment. Another reason for this is that unlike in outdoor environments where people only need to follow a road map, users face multiple features (e.g., elevators, doors, stores, and corridors) in indoor settings, making it hard to precisely predict users' movements in a meaningful way.

Nowadays, with the development of indoor positioning technologies, such as iBeacon [6], radio frequency identification (RFID) [5], Bluetooth [7], Wi-Fi [8] and the prevalence of

Wi-Fi-enabled mobile products, indoor trajectory prediction, which is a promising research field, gradually attracts much more attention. It provides users with flexible services that can be used in location recommendation, movement reconstruction, coupon promotion, and the provision of security services. For example, by being aware of the movement in advance, vendors can quickly target possible shoppers and push advertisements through online advertisement systems, which makes it possible to boost their sales even before customers physically approach the store. Moreover, through this application, pedestrian flow in the shopping mall can be predicted, making it possible to avoid traffic blocks and maximize the effect of shopping coupon promotion. Last but not least, through the indoor prediction system, managers can rearrange workers to achieve a much more efficient operation.

According to the prediction strategy, existing methods can be roughly classified into two groups: individual-based and global-based approaches. The individual-based prediction model assumes that an individual's movement is independent of the others, and we can only use the movement history of the user themselves to predict their future locations [9,10]. Global-based prediction models mainly focus on solving the prediction problem by utilizing the historical movement data of all users to predict a specific user's further location [11]. Data-mining techniques such as recurrent neural network (RNN) [12,13], association rule [14,15] and layered Hidden Markov model (LHMM) [16] have been extensively studied. Though the majority of the aforementioned approaches focus on the spatial aspect, they fall short in describing the unique semantic feature of indoor trajectories, which leads to the poor performance in indoor trajectory prediction.

Therefore, we propose a novel indoor trajectory prediction model, which concentrates on the spatial and semantic aspects simultaneously. A large number of real-world datasets was used to validate the performance of our study. The dataset was collected from about 120,000 anonymous users over a one-year period at a large inner-city shopping mall, which has seven floors, 67 Wi-Fi access points, and more than 200 stores that belong to 34 categories across 90,000 square meters. Other information such as floor plans and stores were provided by the owner of the shopping mall. Customers in the mall need to register for free Wi-Fi service and have to accept the terms and conditions of the service provider. Finally, there are three kinds of logs in the dataset: a Wi-Fi association log, a web browsing log, and a web query log. Compared to our previous work [17], in this study, we focus on the approach of similarity computation rather than model building. The inter-relationship between spatial similarity and semantic similarity in prediction accuracy was investigated, which provides us with a better understanding of human movement in the indoor environment. In conclusion, the contributions of our study are summarized as follows:

- A revised LCSS algorithm is proposed to compute the spatial similarities of the indoor trajectories.
- 2. Semantic features of the indoor trajectory are investigated and a novel algorithm utilizing the semantic R-tree is proposed to compute the semantic similarities.
- 3. Second-order Markov Chain (2-MMC) and k-means algorithms are used to group the trajectories to improve the accuracy of trajectory prediction.
- 4. The performance of our model is evaluated on a large-scale shopping mall dataset. The results imply the advantage of our model against the baseline methods.

Please note that this work is an extended version of our conference paper [18]; compared to the original work, this study has the following improvements: (1) we put forward a new modeling approach that employs 2-MMC and k-means algorithms to construct the mobility model and group the trajectories; (2) in the related work part, indoor positioning and trajectory cleaning techniques are introduced and discussed in detail; (3) in the prediction phase, the tested trajectory is compared to the centroid first instead of all the trajectories; (4) an example of the Wi-Fi access point is illustrated; (5) pseudocodes for the algorithms such as spatial similarity calculation, semantic similarity calculation, model building, and predicting are presented and discussed in detail; (6) the features of the dataset are discussed in the experimental part; and (7) other affecting factors such as the number of clusters and orders of Markov models are investigated in the experiment.

The rest of this study is as follows: In Section 2, we present a literature review on the problem of location prediction. In Section 3, preliminaries about indoor trajectory are introduced. In Section 4, a novel methodological framework for indoor location prediction is proposed. In Section 5, the performance of our approach is evaluated and compared with the baselines. Lastly, in Section 6, we conclude our work, and suggestions for further studies are presented.

2. Related Work

In this part, research on trajectory prediction will be introduced and the differences between these works and our study will be discussed.

Indoor localization: Unlike the outdoor Global Positioning System (GPS), indoor positioning systems have only been mature in recent years and started to emerge in commercial markets [19]. In [20], various positioning technologies used in indoor environments were discussed, and a prototype application for users to navigate through the indoor settings based on the technique of Wi-Fi Received Signal Strength Indicator (RSSI) is proposed. In [21], an evaluation of possible supervised machine learning algorithms is carried out to validate their performances in terms of localization accuracy in the indoor environment; the results show that with the proper selection of algorithms and with a sufficient number of samples for training, we can achieve accurate indoor positioning. While currently the majority of Wi-Fi-based positioning techniques concentrate on increasing localization accuracy, they overlook the diversity of Wi-Fi signal distributions. The authors of [22] proposed a new hybrid model based on the concept of Asymptotic Relative Efficiency (ARE) which exploits signal distributions to strengthen the robustness of the localization systems in complicated indoor environments. To achieve accurate self-positioning and tracking for iPhone users in the indoor scenario, a hybrid method between Wi-Fi and pedestrian dead reckoning (PDR) is provided [23]. In [24], the fingerprint-based positioning algorithms are investigated and a novel criterion is proposed to help better select the reference points. In [25], an automated method is introduced for the calibration of Received Signal Strength (RSS)-fingerprinting-based positioning systems, and a robotic platform is employed to gather fingerprints; then, the gathered fingerprints are used to train machine learning models. To improve the accuracy of range-based localization under the condition of nonline-of-sight in the indoor environment, the authors proposed a localization algorithm by improving the range accuracy [26].

Existing prediction methods can be roughly classified into two groups based on their prediction strategies.

Individual-based prediction considers a users' movement behaviors to be independent of each other, so only the trajectories of the specific object itself are used for the prediction. Regarding the spatial and temporal aspects of trajectory data, time series analysis is first introduced to predict objects' further locations [9], and then Markov model [27] and machine learning techniques [28,29] are investigated. In [28], the authors present a time-ordered vector to model the movement history of customers, while in [29], the authors proposed a classification tree to model the contextual aspects of the trajectory data. Other studies focusing on forecasting the further whereabouts of users in the constrained road networks are also investigated [30,31]. Finally, additional information such as Wi-Fi log [32] or social data [33] are employed to tackle the trajectory prediction problem. The main deficiencies of the individual-based prediction algorithms include: (1) the fact that these methods require long-term trajectory sequences of a certain user which is unrealistic in practice and (2) these approaches need to build an independent prediction model for each user which fails in predicting further whereabouts of non-systematic users.

Global-based prediction models solve the trajectory prediction problem by assuming that users' moving behaviors tend to follow the crowd to a certain extent. Studies mainly focus on mining frequent movement patterns and utilizing this global information to predict

a user's next location. Machine learning techniques such as Markov models, Apriori, and recurrent neural networks (RNN) are extensively investigated. For example, an improved Apriori algorithm is proposed to forecast the whereabouts of a group of shoppers [15]. The authors of [34] employed a Markov model to transfer trajectory points into conversion probabilities for trajectory prediction. Based on RNN, a spatial-temporal RNN model was constructed [12]. However, the aforementioned group-based methods construct prediction models for all users, overlooking the presence of similar subgroups [35]. To solve this problem, a visitor prediction approach is proposed in [36]; the model first mines visitors who are likely to visit the same place, and it then incorporates friends of those visitors, who are influenced by the visitors' activities and are likely to follow them. The authors proposed a novel location prediction method that first considers the trajectories of individuals' familiar strangers [37].

Apart from the model building, the prediction can also be carried out based on the forms of domain knowledge (such as the topology of the map, or constraints on the motility of the people being tracked). For example, (1) with regard to the particle-filtering techniques, the authors of [38] propose a probabilistic model to cleanse RFID data for object tracking; a Bayesian inference based algorithm is utilized and a sequential sampler is proposed to accurately and efficiently clean the RFID data. (2) with regard to proba**bilistic conditioning techniques**, the authors of [39] propose a probabilistic framework for cleaning the data collected by Radio-Frequency IDentification (RFID) tracking systems; the model consists of dumping the inconsistent trajectories and assigning to the others a suitable probability of being the actual one. Probabilistic conditioning is adopted to compute the probabilities. Additionally, a probabilistic cleaning model [40], which treats the trajectories as events with integrity constraints encoding some knowledge about the map and the motility characteristics of the monitored objects, is studied to reduce the inherent uncertainty of trajectory data collected for RFID-monitored objects. (3) With regard to graph-based techniques, the authors of [41] focus on false negatives in raw indoor RFID tracking data, and a probabilistic distance-aware graph is proposed which considers the transition probabilities, the characteristics of indoor topology and RFID readers simultaneously to identify false negatives and recover missing information in indoor RFID tracking data. In [42], the authors propose a sampling technique to interpret RFID data, where a sequence of readings was generated by a set of objects that simultaneously moved for a time interval, their method considers readings, and hard and weak integrity constraints implied by the topology of the floor plan, the capacity of the locations, and the objects' speeds simultaneously.

Due to the limitations of existing indoor trajectory data, the aforementioned methods have only experimented on small datasets: a small number of users and only seven access points were assessed in [43], 2 days' worth of data points as training data, and 1-day worth of data points as testing data were used in [34], the time duration of the data was only 48 h in [16]), and in [44], the dataset was acquired in a limited setting. Different from previous studies, the dataset we used was collected from the general public in a big shopping mall, and over 261,369 indoor trajectories were recorded. The dataset provides us with a unique opportunity to explore the correlation between users' physical movements and semantic movements.

3. Preliminaries and Problem Definitions

In this section, important concepts will be introduced first to help understand our prediction method.

Definition 1 (AP Point). In general, AP point stands for the Wi-Fi access point in the indoor setting where each AP point has a unique ID and its activation range covers multiple stores. In our study, each store belongs to a given category. We denote an AP point as $AP_i = \{i, Sub_{category_i}\}$, where $Sub_{category_i}$ is a subset of all the store categories.

In our study, floor plans of the shopping mall were overlaid with access point locations, and the active ranges of the access points were approximated by Voronoi regions, each centered on a single access point which encompasses all the points that are closest to it. In order to correlate user physical movements captured by the access points, we label each access point with semantics corresponding to its location in the shopping center. As each access point covers a certain area with signal in the shopping mall, the service area is approximated by the so called Voronoi cell, in which any location is closest to its seed location (the access point) than to any other seed location. Once we obtain these Voronoi cells, we know which shop falls under an AP from the shopping center floor plan (on average, there are 3.67 shops in each Voronoi cell). We then assign a list of semantic categories to an access point corresponding to each shop in the region covered by an access point.

Table 1 is an example of an access point; AP_{15} covers six stores belonging to the category of Cafe, Men's Fashion, General Fashion, General Footwear, Men's Footwear, and Underwear. Then, we denote AP_{15} as $AP_{15} = \{15, [Cafe, Men's Fashion, General Fashion, General Footwear, Men's Footwear, Underwear]\}$

Table 1. Illustration of AP_{15} .

AP	Store Name	Store Category
AP_{15}	Starbucks	Cafe
	Armani	Men's Fashion
	Bulgari	General Fashion
	ECCO	General Footwear
	LOTTUSSE	Men's Footwear
	Laperla	Underwear

Definition 2 (Trajectory Point). We denote the indoor trajectory point as $P_i = \{(AP_n, t_i)\}$. When a customer logged into a Wi-Fi access point AP_n at the timestamp t_i , $n \in [1, k]$, where k is the total number of Wi-Fi access points.

Definition 3 (Indoor Trajectory). *The indoor trajectory T is an ordered sequence of trajectory points detected in a user's movement history, where* $T = \{P_1, P_2, \dots, P_m\}$.

Definition 4 (Trajectory Similarity). *Indoor trajectory* T_1 *is similar to* T_2 *if* $D(T_1, T_2) \le \epsilon$ *, where* D(.) *is the distance function and* ϵ *is the distance threshold.*

In the past few years, the trajectory distance function D(.) has been investigated extensively and can be defined in various ways. For instance, the authors of [45] studied examples of trajectory distance functions used for analyzing the movement data from different objectives. During our background study, there was a set of distance functions based on string matching such as dynamic time wrapping distance (DTW) [46], longest common subsequence (LCSS) [47], edit distance for trajectories [48] and Euclidean distance [49]; these kinds of methods provide a straightforward way to depict trajectory as a list of time-ordered spatial points. A survey of various distance metrics can be found in [50]. Although the indoor environment is constrained, it is full of semantic information; however, few studies to date have considered the influence of semantic features in indoor environments. Different from the aforementioned algorithms, in this study, we propose a new distance function that considers the spatial and contextual information simultaneously.

Definition 5 (Indoor Trajectory Prediction:). Given an indoor trajectory $T = \{P_1, P_2, \dots, P_k\}$, the goal is to compute the position of P_{n+1} in the (n + 1)th timestamp based on the previous *n* timestamps.

In the following sections, we use the terms indoor trajectory and trajectory interchangeably unless otherwise specified.

4. The Prediction System

In this part, our indoor trajectory prediction model considering both the spatial and contextual aspects will be introduced. It is capable of solving the following problems: (1) How do we define and compute the contextual similarity between trajectories? (2) How do we group the trajectories and construct the predicting model for the test trajectory without employing the whole dataset?

4.1. Trajectory Sequence Similarity Calculation

The main part of our prediction approach is to cluster the trajectory sequences. We hold the assumption that users tend to follow the same movement patterns when their trajectories are clustered in the same group and vice versa. Our model improves the prediction accuracy by modeling similar trajectories.

Trajectories in indoor settings contain not only spatial information but also contextual information. The spatial information is represented by the sequence of time-ordered Wi-Fi access points the users have logged when moving inside the indoor environment, while the contextual information describes the store categories in the shopping mall and reflects the users' shopping habits to some extent. In our study, we propose a unified framework that considers spatial and semantic information simultaneously.

Spatial similarity mainly computes the similarity of users' physical movement sequences in the geospatial space. We hold the assumption that when customers move around in the indoor space, they will share much more common physical movement patterns as the trajectory sequences become longer. So, we employed the longest common subsequence (LCSS) to calculate the spatial similarity between indoor trajectories.

Definition 6 (Spatial Similarity). *The spatial similarity between two trajectories is defined and computed as follows:*

$$spaRatio(T,R) = \frac{LCSS(T,R)}{min(||T||,||R||)}$$
(1)

||T|| represents the number of Wi-Fi access points in trajectory *T* and the dynamic programming method is used to calculate the LCSS between two trajectories; Algorithm 1 presents the detailed procedure.

Lemma 1. The defined spatial ratio function 1 between two trajectories is symmetric, i.e.,

$$spaRatio(T, R) = spaRatio(R, T)$$
 (2)

Proof. As the LCSS algorithm always returns the sub-sequence in common between two trajectories, and the denominator always returns the minimum length, then the ratio is always symmetric. \Box

Semantic similarity mainly computes the similarity of contextual information between the trajectory sequences; the contextual information will reveal users' shopping behaviors to some extent.

In practice, based on the functions, an indoor building can be classified into various sections. For instance, an airport terminal can be divided into check-in, security checking, waiting room, and boarding sections; a supermarket can be divided into bakery, vegetables, seafood, meat, and cooked food sections. Generally, users' physical movements have strong correlations with the section he/she is currently in [51–53]. For example, in an airport terminal, the majority of passengers will follow the route of check-in, security checking, waiting room, and boarding; and in a shopping mall, people will also show certain navigational hobbies. As shown in Figure 1, a customer first spends some time

buying shoes; after that he wants to eat something, and then he may enjoy the big hit movie in the cinema. Through this, we can infer his sequential patterns and so the customer's behavior is highly predictable.

Algorithm 1 Spatial similarity calculation.

```
Input: trajectory sequences T, R
Output: the spatial similarity ratio between T and R
 1: a \leftarrow ||T||, b \leftarrow ||R||,
 2: Initialize M[a][b] //the matrix that stores the LCSS
 3: for i \leftarrow 1 to a do
 4:
        M[i][0] \leftarrow 0
 5: end for
 6: for j \leftarrow 1 to b do
        M[0][j] \leftarrow 0
 7:
    end for
 8:
 9: for i \leftarrow 1 to a do
10:
        for j \leftarrow 1 to b do
11:
            max\_len \leftarrow = M[i-1][j-1]+1
            if max\_len < M[i-1][j] and M[i-1][j] > M[i][j-1] then
12:
13:
                max\_len \leftarrow M[i-1][j]
            else if max\_len < M[i][j-1] and M[i-1][j] > M[i][j-1] then
14:
                max\_len \leftarrow M[i][j-1]
15:
16:
            end if
17:
            M[i][j] = max\_len
        end for
18:
19: end for
20: len = min(a, b) / / the minimum length of two trajectories
21: Return M[a][b]/len
```



Food Court

Figure 1. An example of a user's movement in a shopping mall. The red squares denote the Wi-Fi APs; the dashed lines show the movement of the customer.

In our study, a specific semantic location can be covered by multiple access points. Figure 2 depicts the detailed information about how many access points cover the same semantic location. For example, the semantic location "P130 Unisex Fashion(MM)" is covered by 17 Wi-Fi access points.



Figure 2. Semantic locations w.r.t number of access points.

Based on the semantic-R-Tree, we calculated the semantic similarities by measuring the correlations between the shop categories. We denote the leaf nodes as semantic labels which represent the actual stores/sections in the indoor settings. As can be seen from Figure 3, none leaf nodes represent the location category of the lower-level nodes. The nodes will represent a larger section when they are at a higher level.



Figure 3. Location Semantic analysis.

Here, a revised DTW algorithm is utilized to compute the contextual similarity of trajectories, as shown in Formula (3), T_i and R_j represent the shop category sequences of trajectories T and R, respectively.

$$dist(T,R) = \begin{cases} 0, & m = n = 0\\ \infty, & m = 0 \text{ or } n = 0\\ dist(T_i, R_j) + min \begin{cases} dist(T_i, R_{j-1}) & m \neq 0, n \neq 0\\ dist(T_{i-1}, R_j) & m \neq 0, n \neq 0\\ dist(T_{i-1}, R_{i-1}) & m \neq 0, n \neq 0 \end{cases}$$
(3)

Definition 7 (Semantic Similarity). *The semantic similarity between two trajectories T and R is defined as:*

$$semRatio(T,R) = \frac{1}{1 + dist(T,R)} \times \frac{mim(m,n)}{max(m,n)}$$
(4)

The first part of the formula uses the dynamic time warping (DTW) method to measure the semantic distance of two trajectories, while it does not take into account the difference in trajectory length. The latter part of the formula introduces the trajectory length ratio to solve the problem that the length of trajectories are not equal. If the trajectories T and R are the same, then distTR(T, R) is zero, and the result of the formula calculation is 1; that is, the trajectories are completely similar.

The pseudo-code of the Semantic Similarity Calculation is shown in Algorithm 2. It first initializes the height of a semantic analysis tree *SRT* and lowest common ancestor (*LCA*) (line 1); it then traverses the semantic analysis tree to obtain its height (line 2). After calculating the length of indoor trajectory T_1 and T_2 (line 3), the semantic distances between the two given trajectories is calculated (lines 4–11).

Algorithm 2 Semantic Similarity Calculation.
Input: two trajectory sequences <i>T</i> , <i>R</i>
semantic analysis R-Tree SRT
Output: the semantic similarity ratio between <i>T</i> and <i>R</i>
1: Initialize height, LCA,
2: $height \leftarrow get_SRT_height(root)$
3: $m \leftarrow T , n \leftarrow R / \text{length of trajectory } T \text{ and } R$
4: //Compute semantic distances
5: while $i \in m, j \in n$ do
6: $LCA \leftarrow getLCA(root, p_i, p_j) / / obtain the LCA$
7: $h \leftarrow getlevel(LCA, root) / / compute the height of LCA$
8: $dist(T_i, R_j) \leftarrow h/height$
9: end while
10: while $i \in m$ do
11: $dist(T, R) \leftarrow dist(T_i, R_j);$
12: end while
13: Return $semRatio(T, R) = \frac{1}{1+dist(T,R)} \times \frac{mim(m,n)}{max(m,n)}$

After calculating the spatial and semantic similarities between trajectories, the final indoor trajectory similarity is superimposed in two parts, which is defined in Formula (5). Min–max normalization is employed here to solve the problem of differences in spatial and semantic similarity.

$$Dist_{total}(T, R) = \alpha * \frac{spaRatio(T, R) - min(spaRatio)}{max(spaRatio) - min(spaRatio)} + (1 - \alpha) * \frac{semRatio(T, R) - min(semRatio)}{max(semRatio) - min(semRatio)}$$
(5)

4.2. Indoor Trajectory Prediction Framework

We define the trajectory prediction problem as follows: Given a user's trajectory sequence in the indoor environment, our goal is to predict the location he is going to visit next. The framework of our approach consists of two parts: (1) Model Building, which finds the most similar trajectory group and builds the prediction model for the given trajectory, and (2) Prediction, which forecasts the next location for a specific user.

Model Building Phrase: During the process of model building, our goal is to find the given user's trajectory similarity group and construct the prediction model. Since each user has to compute similarity with all the others, there will be n(n - 1)/2 times of computations in total. When we have a large number of users, it will be time consuming to conduct the similarity computation. According to social contagion theory, people tend to be affected by others and adjust themselves to crowds [54]. According to this philosophy, series subsets of trajectories are extracted and the tested user's trajectory similarity group C_i is obtained by computing its similarity with the centroids in the subsets. The model building process is shown in Algorithm 3.

Algorithm 3 Model Building Phase. **Input:** Indoor trajectory dataset D, weight coefficient: α **Output:** the location prediction model of every user *model*_i, Cluster centers: $Centroid_i$ 1: $m \leftarrow ||D||$ 2: Initialize Matrix[m][m]for $i \leftarrow 1$ to *m* do */*/compute the spatial and semantic similarity of two trajectories 3: 4: for $j \leftarrow i + 1$ to *m* do $spaRatio(T_i, T_j) \leftarrow Obtain spatial similarity between T_i and T_i$ 5: *semRatio*(T_i , \dot{T}_i) \leftarrow Obtain semantic similarity between T_i and T_i 6: 7: end for 8: end for 9: Compute min(spaRatio), max(spaRatio), min(semRatio), max(semRatio) 10: for $i \leftarrow 1$ to *m* do //compute the trajectory distance matrix for $j \leftarrow i + 1$ to *m* do 11: $dist \leftarrow Dist_{total}(T_i, T_j)$ 12: 13: $Matrix[i][j] \leftarrow dist$ end for 14: 15: end for 16: //obtain the clusters and centroid 17: $\{C_i\}, \{Centroid_i\} = kmeans(Matrix[m][m])$ 18: //construct the model 19: for $T \in C_i$ do 20: model = 2-MMC(T)models.add(model) 21: 22: end for 23: **Return** *Models*, *Centroid*_i

Our proposed approach uses the 2-MMC algorithm for constructing the model. The core of the 2-Markov-based prediction is that the next visiting location of the tested object is only related to its two previous locations without considering the whole movement history.

We did not employ the standard Markov model because its prediction performance is not stable and will be affected by the last state badly, where the next location is only related to the current one. For instance, a customer lives a simple lifestyle of "home \rightarrow subway \rightarrow workplace" and "workplace \rightarrow subway \rightarrow Gym". When his current location is subway, the standard Markov-based approach will return one place between workplace and gym as the prediction result, meaning almost half of the overall predictions are wrong. In contrast, when using the 2-MMC algorithm, we can benefit from his previous location which is workplace, and the current one is subway; obviously, the next place he will attend is gym. As can be seen from the above example, 2-MMC performs better without extremely increasing the state space.

In Markov theory, the probability of moving to a new state depends on the current state and the transition matrix between states. So, predicting the probability of a user u visiting a new location at (n + 1)th timestamp given u's visited location sequence ordered by his/her check-in timestamps can be depicted as Formula (6).

$$P(P_{n+1} = AP_{n+1}|P_n = AP_n, P_{n-1} = AP_{n-1}, \cdots, P_1 = AP_1)$$

= $P(P_{n+1} = AP_{n+1}|P_n = AP_n)$ (6)

Based on Equation (6), the chance of a customer going to a new position AP_{i+1} by considering the first-order Markov Chain can be represented as Formula (7):

$$Pr(AP_{j}|AP_{i}) = \frac{F(AP_{i}->AP_{i+1})}{\sum_{k=1}^{n}F(AP_{i}->AP_{k})}$$
(7)

As for the second-order Markov model, we will build a larger transition matrix because there will be many more combinations between the last two history locations and the probability of moving to a new location considering the last two locations is (Formula (8)):

$$Pr(AP_{i+1}|AP_i, AP_{i-1}) = \frac{F(AP_{i-1}, AP_i - > AP_{i+1})}{\sum_{k=1}^{n} F(AP_{i-1}, AP_i - > AP_k)}$$
(8)

where the numerator is the count of single movement from AP_{i-1} to AP_{i+1} in the training dataset while the denominator represents the sum of the transition trajectories that are originated from AP_{i-1} ; *n* is the total number of Wi-Fi access points.

Predicting phrase: After model building, there is a one-to-one correspondence between centroids and models. That is, *Centroid_i* corresponds to *Model_i*. Given a new customer's indoor trajectory, the goal is to predict his/her next location with the highest possibility. The distances between the given indoor trajectory and centroids are first calculated to determine which group the trajectory belongs to, and then the corresponding prediction model is employed to forecast the most likely further whereabouts of the customer. The prediction process of the proposed approach is shown in Algorithm 4.

Algorithm 4 Prediction Process.

Input: tested trajectory *T*, weight coefficient α Prediction models: {model_i} Cluster centroids: {Centroid_i} Output: Set of predicted locations 1: for Centroid_i in {Centroid_i} do 2: //Compute the similarity between *T* and Centroid_i 3: dist \leftarrow Dist_{total}(*T*, Centroid_i) 4: end for 5: //Obtain the most similar Centroid_i 6: id = argmax(dist) 7: model = models[id] //model building 8: {location_i} = model(*T*) 9: return Set of predicted locations

5. Experimental Results and Analysis

In this section, the experiments we performed will be outlined to validate our proposed algorithm with the dataset collected from a large shopping mall. All algorithms were implemented in Python 3 and ran on a Windows 10 Pro PC with a 2.90 GHz Core i7-7500U CPU. A five-fold cross-validation was carried out in the experiments. We split the dataset into training and testing parts from the perspective of check-in time instead of using a random partition method. This is because it makes no sense to use the values from the future to forecast values in the past, and so we will avoid looking further when we train the model and preserve the relation of temporal dependency between observations.

5.1. Experiment Setup

Data acquisition: The data were collected from about 120,000 anonymized users over 1 year period at a large inner-city shopping center with seven floors and 67 Wi-Fi access points across 90,000 square meters. Customers can register for free Wi-Fi services and have to accept the terms and conditions of the service provider. According to the mall operator, the mall consists of more than 200 stores that belong to 34 shop categories. Details about the dataset are shown in Table 2.

Item	Value
APs	67
associations	907,084
shops	200
shop categories	34
trajectories	261,369

Table 2. Brief overview of the dataset.

An example of the trajectory points is shown in Table 3.

Table 3. An example of the trajectory points.

tra_id	ap_Name	Association_Time	Disassociation_Time
7BTAVNC	wap013	9/5/20:57	9/5/21:17
7BTAVNC	wap023	9/5/20:52	9/5/20:57
7BTAVNC	wap048	9/5/20:47	9/5/20:52

According to the layout of the shopping mall, the proximal areas of the 67 access points (covering over 200 stores) belong to more than 30 categories (e.g., Sports, Books, Entertainment, and Drinks) and can be categorized into three main categories, Retail, Food, and Navigational, which have 46, 11, and 10 Wi-Fi access points, respectively. Detailed analyses of each category with reference to the accessed times are shown in Table 4.

Table 4. Access times w.r.t each store category

Store Category	% Access Times	Average Duration (h)
Food	23%	1.39
Retail	70%	2.29
Navigational	7%	1.00
Total	100%	2.77

Data Processing: Our dataset was collected through Wi-Fi positioning; due to the unstable mobile terminal signal, abnormal, erroneous, and invalid data were easily generated. We consider the following two kinds of trajectory points as noise points:

- 1. Abnormal time points: If the sampling interval of two adjacent trajectory points was 0 s, it was considered as an abnormal time trajectory point.
- 2. Abnormal floor points: If a trajectory point was not in the study area or jumped between different floors in a short time period, it was considered as an abnormal point.

Figure 4 depicts the trend about the number of trajectories w.r.t the length. We omit the trajectories whose lengths are smaller than three, as they are too short to convey meaningful visiting hobbies. As can be seen from the figure, more than 90% of the trajectories tend to access three to nine Wi-Fi access points.

We analyzed the retention time of all users' login data; as can be seen from Figure 5, the majority of customers tend to spend less than thirty minutes in an area which corresponds to their shopping behaviors.



Figure 4. number of trajectories w.r.t the length.



Figure 5. Distribution: Duration of stay.

Parameters: Table 5 summarizes the complete set of the parameters that we used in the experiments. A parameter was set to the default value where any other parameter was being varied.

Table 5. Parameters that we used in our experiments.

Parameter	Values	Default
number of groups (N)	5, 10, 15, 20, 25, 30	20
weight coefficient (α)	0.9, 0.8, 0.7,, 0	0.4
number of correct predicted locations (k)	1, 3, 5, 7, 9	-
order of markov model (r)	1, 2, 3, 4, 5	2

Evaluation Metrics: In this study, our indoor trajectory predicting model ranks each candidate (i.e., Wi-Fi access point) concerning a target trajectory and returns the *top-k* highest locations as prediction results to the tested trajectory. The metric *Accuracy@k* was employed to validate the quality of the proposed method.

 Accuracy@K: This measures how close our predicted location is to the true location of the tested trajectory.

$$Accuracy@k = \frac{number \ of \ locations \ correctly \ predicted}{total \ number \ of \ test \ locations} \tag{9}$$

Note that we did not employ mean absolute error (MAE) as our performance metric in the experiments. The reason is that MAE is a measure of the deviation of recommendations from the true user-specified values. However, in our problem, users have no ratings with regard to the next location and the metric is not suitable for our problem.

Comparative methods: We compare our model with the following baseline methods to verify the effectiveness of our spatial-contextual-based method.

- 1. *HMM*, which is extensively used in human movement prediction [55].
- LHMM is a layered mechanism for modeling the spatial-temporal trajectory sequences [16].
- 3. *MMC*, which is effective in time-ordered sequence prediction [56].

5.2. Experimental Results

In this part, we will evaluate the performances of our indoor trajectory prediction model. The effect of the weight coefficient α is first investigated, and then the influence of group number is studied; after this, the order of Markov model to the prediction accuracy is tested. Finally, the proposed spatial-contextual similarity-based prediction approach is compared with the baseline methods.

5.2.1. The Influence of Weight Coefficient

The weight coefficient α is mainly used to validate the influence of spatial and contextual similarity on the performance of prediction accuracy. During the experiment, we tested and found the optimal value for parameter α from [0, 0.1, 0.2, ..., 1]. When α is set to 0 or 1, it means only the spatial or the contextual similarity is considered to affect the trajectory prediction results.

Figure 6 depicts the influence of coefficient α on the performance metric *Accuracy@k*. When parameter $k \in \{1, 3, 5, 7, 9\}$, the prediction accuracy first shows an increasing and then decreasing trend. Our method exhibits a relatively high performance when $0.3 \le \alpha \le 0.7$. When $\alpha = 0.4$, *Accuracy@*5 achieved 58.2% with improvements of 20.8% and 18.3% compared to $\alpha = 0$ and $\alpha = 1$, respectively, implying that both the spatial similarity and contextual similarity contributed to the prediction accuracy.



Figure 6. Influence of weight coefficient α on prediction accuracy.

5.2.2. Effect Analysis of the Number of Clusters

Because the number of the tested trajectory similarity groups impacts the performance of our method, we employed parameter N to represent the number of clustered groups and observe the changes in performance metric *accuracy@k* under different settings of N. Figure 7 shows the results.



Figure 7. Impact of group number on prediction accuracy.

Initially, with the increment of groups, the prediction accuracy increases gradually. We infer that the reason is that at the early beginning there are too many trajectories in the group and the prediction performance of our model may be interfered by the data noise, so with the decrease in data noises, the prediction accuracy rises. When N = 20, the prediction accuracy achieves the best score, 3%, 3.9%, and 5.7% for P@1, P@3, and P@5, respectively. However, after this, the prediction accuracy shows a trend of decreasing. The reason may be that with the increment of parameter N, the model of the test trajectory has fewer references to make accurate predictions. However, we claim that even though in our experiment the highest prediction accuracy is achieved when N = 20, the optimal value for N varies according to different datasets. Because N represents the balance between the positive information and the data noise, the strength of social contagion and interference varies for datasets [54].

5.2.3. Influence of Markov Order

We validated the influence of Markov order on the prediction accuracy of our algorithm, and the results are shown in Figure 8. As can be seen from the figure, when the order of the Markov model increases from one to two, the prediction accuracy increases greatly, our model achieves prediction accuracies of 29%, 39% and 58% for p@1, p@3 and p@5, respectively. However, after that, the prediction accuracy declines gradually.



Figure 8. Influence of Markov order to prediction accuracy.

We infer that the reason is that, for the first-order Markov model, its prediction ability is affected badly by the last state, where the prediction is only made in the last state, leading to unsatisfactory prediction performances. For the higher-order Markov models, there is a possibility that the tested trajectory cannot find a pattern to match, and that the model will not return prediction results for the given tested trajectory. Since constructing the high-order Markov model is time consuming and the prediction ability has not improved, we just employ the second-order Markov model in our experiments.

5.2.4. Baseline Comparison

To validate the efficiency of our proposed method, it was compared with three baseline methods. We set the coefficient α and cluster number *N* to 0.4 and 20, respectively, as a default. As can be concluded from Figure 9, we ascertain that our method performs the best where the precision of our approach can be improved by 66~98%, 14~72% and 28~70% for P@1, P@3, and P@5, respectively, compared with the baseline methods.

The reason for this may be that the comparing approaches take the trajectory prediction problem as a time-ordered sequential modeling issue and the similarity between trajectories is not taken into consideration. Moreover, the contextual information in the indoor environment is also an important factor for the trajectory prediction problem which none of the baseline methods have considered. In general, our method significantly outperforms the baseline methods in terms of prediction accuracy.



Figure 9. Prediction accuracy comparison.

6. Conclusions and Further Work

Precise and robust trajectory prediction plays a critical role in indoor location-based services, especially in large shopping malls. For instance, predicting the next store to be visited by a customer as well as providing information about the store to the user not only offers a personalized shopping experience but also makes it possible to boost sales even before the customer physically approaches the store.

In this paper, we first consider the spatial and contextual similarities of indoor trajectories, and a revised LCSS and a novel semantic R-tree-based algorithm are proposed to compute the similarity between trajectories. Then, we classify and group the trajectories by employing the K-means algorithm according to the similarity matrix and construct the predicting model based on the 2-MMC algorithm. We evaluated our approach on a real-world indoor trajectory dataset. The experiment results show that our model outperforms the baseline methods remarkably in prediction accuracy.

To this end, we summarize the advantages of our method as follows: (1) We put forward a semantic_RTree-based algorithm to simulate users' semantic movements. (2) Through the similarity-based approach, we have gained a better understanding of the inter-relationship between spatial and semantic aspects on the performance of trajectory prediction. (3) Our method provides us with an insight into the prediction accuracy and the size of the training data. (4) It is efficient to group the trajectories rather than compare the test trajectory with all the other trajectories in the database. However, there are still limitations in this study: (1) Though we used a dataset for the shopping mall, the effectiveness of our method for other indoor environments is unclear. (2) We still do not have a clear understanding of how the temporal aspects such as retention time will affect the prediction accuracy.

Future studies could potentially (1) use more types of indoor trajectory data to verify the model, such as trajectories in the airport or in the hospital or (2) integrate more factors to enhance the robustness of our model, such as temporal factors.

Author Contributions: Conceptualization, P.W., J.Y. and J.Z.; methodology, P.W. and J.Y.; software, P.W.; validation, P.W.; formal analysis, P.W.; investigation, P.W.; writing—original draft preparation, P.W.; writing—review and editing, J.Y.; supervision, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant Nos. 61672179, 61370083, and 61402126; the Natural Science Foundation Heilongjiang Province of China under Grant Nos. F2015030; the Youths Science Foundation of Heilongjiang Province of China under Grant No. QC2016083; the Fundamental Research Funds for the Central Universities under Grant No. HEUCFM180601; the Heilongjiang Postdoctoral Science Foundation No. LBH-Z14071.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are available from the authors upon readers request.

Acknowledgments: The authors acknowledge valuable suggestions of anonymous referees.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Huang, J.; Dahlmeier, D.; Lin, Z.; Ang, B.K.; Seeto, M.L.; Shi, H. Wi-Fi Based Indoor Next Location Prediction Using Mixed State-Weighted Markov-Chain Model. *Int. J. Mach. Learn. Comput.* **2014**, *4*, 505. [CrossRef]
- Klepeis, N.E.; Nelson, W.C.; Ott, W.R.; Robinson, J.P.; Tsang, A.M.; Switzer, P.; Behar, J.V.; Hern, S.C.; Engelmann, W.H. The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants. *J. Expo. Sci. Environ. Epidemiol.* 2001, *11*, 231. [CrossRef] [PubMed]
- 3. Jenkins, P.L.; Phillips, T.J.; Mulberg, E.J.; Hui, S.P. Activity patterns of Californians: Use of and proximity to indoor pollutant sources. *Atmos. Environ. Part A Gen. Top.* **1992**, *26*, 2141–2148. [CrossRef]
- Liu, H.; Darabi, H.; Banerjee, P.P.; Liu, J. Survey of Wireless Indoor Positioning Techniques and Systems. *IEEE Trans. Syst. Man, Cybern. Part C* 2007, 37, 1067–1080. [CrossRef]
- Xu, H.; Ding, Y.; Li, P.; Wang, R.; Li, Y. An RFID Indoor Positioning Algorithm Based on Bayesian Probability and K-Nearest Neighbor. Sensors 2017, 17, 1806. [CrossRef]
- Martin, P.; Ho, B.; Grupen, N.; Muñoz, S.; Srivastava, M.B. An iBeacon primer for indoor localization: Demo abstract. In Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, BuildSys 2014, Memphis, TN, USA, 3–6 November 2014; pp. 190–191. [CrossRef]
- Lie, M.M.K.; Kusuma, G.P. A fingerprint-based coarse-to-fine algorithm for indoor positioning system using Bluetooth Low Energy. *Neural Comput. Appl.* 2021, 33, 2735–2751. [CrossRef]
- Cao, H.; Wang, Y.; Bi, J.; Xu, S.; Si, M.; Qi, H. Indoor Positioning Method Using WiFi RTT Based on LOS Identification and Range Calibration. *ISPRS Int. J. Geo Inf.* 2020, 9, 627. [CrossRef]
- Scellato, S.; Musolesi, M.; Mascolo, C.; Latora, V.; Campbell, A.T. NextPlace: A Spatio-temporal Prediction Framework for Pervasive Systems. In Proceedings of the Pervasive Computing—9th International Conference, Pervasive 2011, San Francisco, CA, USA, 12–15 June 2011; Lyons, K., Hightower, J., Huang, E.M., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2011; Volume 6696, pp. 152–169. [CrossRef]
- Ceci, M.; Appice, A.; Malerba, D. Time-Slice Density Estimation for Semantic-Based Tourist Destination Suggestion. In Proceedings of the ECAI 2010—19th European Conference on Artificial Intelligence, Lisbon, Portugal, 16–20 August 2010; Coelho, H., Studer, R., Wooldridge, M.J., Eds.; Frontiers in Artificial Intelligence and Applications; IOS Press: Amsterdam, The Netherlands, 2010; Volume 215, pp. 1107–1108. [CrossRef]
- Morzy, M. Mining Frequent Trajectories of Moving Objects for Location Prediction. In Proceedings of the Machine Learning and Data Mining in Pattern Recognition, 5th International Conference, MLDM 2007, Leipzig, Germany, 18–20 July 2007; Perner, P., Ed.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4571, pp. 667–680. [CrossRef]
- Liu, Q.; Wu, S.; Wang, L.; Tan, T. Predicting the Next Location: A Recurrent Model with Spatial and Temporal Contexts. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; Schuurmans, D., Wellman, M.P., Eds.; AAAI Press: Palo Alto, CA, USA, 2016; pp. 194–200.
- Yao, D.; Zhang, C.; Huang, J.; Bi, J. SERM: A Recurrent Model for Next Location Prediction in Semantic Trajectories. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, 6–10 November 2017; Lim, E., Winslett, M., Sanderson, M., Fu, A.W., Sun, J., Culpepper, J.S., Lo, E., Ho, J.C., Donato, D., Agrawal, R., et al., Eds.; ACM: New York City, NY, USA, 2017; pp. 2411–2414. [CrossRef]
- Keles, I.; Ozer, M.; Toroslu, I.H.; Karagoz, P. Location Prediction of Mobile Phone Users Using Apriori-Based Sequence Mining with Multiple Support Thresholds. In Proceedings of the New Frontiers in Mining Complex Patterns—Third International Workshop, NFMCP 2014, Held in Conjunction with ECML-PKDD 2014, Nancy, France, 19 September 2014; Appice, A., Ceci, M., Loglisci, C., Manco, G., Masciari, E., Ras, Z.W., Eds.; Revised Selected Papers; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2014; Volume 8983, pp. 179–193. [CrossRef]
- Morzy, M. Prediction of Moving Object Location Based on Frequent Trajectories. In Proceedings of the Computer and Information Sciences—ISCIS 2006, 21th International Symposium, Istanbul, Turkey, 1–3 November 2006; Levi, A., Savas, E., Yenigün, H., Balcisoy, S., Saygin, Y., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2006, Volume 4263, pp. 583–592. [CrossRef]
- Li, Q.; Lau, H.C. A Layered Hidden Markov Model for Predicting Human Trajectories in a Multi-floor Building. In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT 2015, Singapore, 6–9 December 2015; pp. 344–351. [CrossRef]

- 17. Wang, P.; Yang, J.; Zhang, J. A Spatial-Contextual Indoor Trajectory Prediction Approach via Hidden Markov Models. *Wirel. Commun. Mob. Comput.* 2022, 2022, 6719514. [CrossRef]
- Wang, P.; Yang, J.; Zhang, J. Location Prediction for Indoor Spaces based on Trajectory Similarity. In Proceedings of the DSIT 2021: 4th International Conference on Data Science and Information Technology, Shanghai, China, 23–25 July 2021; ACM: New York City, NY, USA, 2021; pp. 402–407. [CrossRef]
- 19. Jensen, C.S.; Lu, H.; Yang, B. Indoor-A New Data Management Frontier. IEEE Data Eng. Bull. 2010, 33, 12–17.
- Golenbiewski, J.; Tewolde, G. Wi-Fi Based Indoor Positioning and Navigation System (IPS/INS). In Proceedings of the 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Vancouver, BA, Canada, 9–12 September 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–7.
- Polak, L.; Rozum, S.; Slanina, M.; Bravenec, T.; Fryza, T.; Pikrakis, A. Received Signal Strength Fingerprinting-Based Indoor Location Estimation Employing Machine Learning. *Sensors* 2021, 21, 4605. [CrossRef]
- 22. Zhou, M.; Li, Y.; Tahir, M.J.; Geng, X.; Wang, Y.; He, W. Integrated statistical test of signal distributions and access point contributions for Wi-Fi indoor localization. *IEEE Trans. Veh. Technol.* **2021**, *70*, 5057–5070. [CrossRef]
- Vy, T.D.; Nguyen, T.L.; Shin, Y. Pedestrian Indoor Localization and Tracking Using Hybrid Wi-Fi/PDR for iPhones. In Proceedings of the 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring), Helsinki, Finland, 25–28 April 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–7.
- 24. Tao, Y.; Zhao, L. Fingerprint localization with adaptive area search. IEEE Commun. Lett. 2020, 24, 1446–1450. [CrossRef]
- Kolakowski, M. Automated Calibration of RSS Fingerprinting Based Systems Using a Mobile Robot and Machine Learning. Sensors 2021, 21, 6270. [CrossRef] [PubMed]
- Lee, B.h.; Park, K.M.; Kim, Y.H.; Kim, S.C. Hybrid Approach for Indoor Localization Using Received Signal Strength of Dual-Band Wi-Fi. Sensors 2021, 21, 5583. [CrossRef] [PubMed]
- Nishino, M.; Nakamura, Y.; Yagi, T.; Muto, S.; Abe, M. A location predictor based on dependencies between multiple lifelog data. In Proceedings of the 2010 International Workshop on Location Based Social Networks, LBSN 2010, San Jose, CA, USA, 2 November 2010; Zhou, X., Lee, W., Peng, W., Xie, X., Eds.; ACM: New York City, NY, USA, 2010; pp. 11–17. [CrossRef]
- Anagnostopoulos, T.; Anagnostopoulos, C.; Hadjiefthymiades, S. Mobility Prediction Based on Machine Learning. In Proceedings of the 12th IEEE International Conference on Mobile Data Management, MDM 2011, Luleå, Sweden, 6–9 June 2011; pp. 27–30. [CrossRef]
- Tran, L.H.; Catasta, M.; McDowell, L.K.; Aberer, K. Next place prediction using mobile data. In Proceedings of the Mobile Data Challenge Workshop (MDC 2012), Newcastle, UK, 18–22 June 2012; Number CONF.
- Kim, S.; Won, J.; Kim, J.; Shin, M.; Lee, J.; Kim, H. Path Prediction of Moving Objects on Road Networks Through Analyzing Past Trajectories. In *Knowledge-Based Intelligent Information and Engineering Systems, Proceedings of the 11th International Conference, KES* 2007, XVII Italian Workshop on Neural Networks, Vietri sul Mare, Italy, 12–14 September 2007; Apolloni, B., Howlett, R.J., Jain, L.C., Eds.; Part I; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4692, pp. 379–389. [CrossRef]
- Jeung, H.; Yiu, M.L.; Zhou, X.; Jensen, C.S. Path prediction and predictive range querying in road network databases. *VLDB J.* 2010, 19, 585–602. [CrossRef]
- Gomes, J.B.; Phua, C.; Krishnaswamy, S. Where Will You Go? Mobile Data Mining for Next Place Prediction. In Proceedings of the Data Warehousing and Knowledge Discovery—15th International Conference, DaWaK 2013, Prague, Czech Republic, 26–29 August 2013; Bellatreche, L., Mohania, M.K., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8057, pp. 146–158. [CrossRef]
- Yang, N.; Kong, X.; Wang, F.; Yu, P.S. When and Where: Predicting Human Movements Based on Social Spatial-Temporal Events. In Proceedings of the 2014 SIAM International Conference on Data Mining, Philadelphia, PA, USA, 24–26 April 2014; pp. 515–523. [CrossRef]
- Ang, B.K.; Dahlmeier, D.; Lin, Z.; Huang, J.; Seeto, M.L.; Shi, H. Indoor next location prediction with Wi-Fi. In Proceedings of the Fourth International Conference on Digital Information Processing and Communications (ICDIPC2014), Kuala Lumpur, Malaysia, 18–20 March 2014; The Society of Digital Information and Wireless Communication: Hong Kong, China, 2014; pp. 107–113.
- 35. Wang, P.; Wu, S.; Zhang, H.; Lu, F. Indoor location prediction method for shopping malls based on location sequence similarity. *ISPRS Int. J.-Geo-Inf.* 2019, *8*, 517. [CrossRef]
- Saleem, M.A.; Costa, F.S.D.; Dolog, P.; Karras, P.; Pedersen, T.B.; Calders, T. Predicting Visitors Using Location-Based Social Networks. In Proceedings of the 19th IEEE International Conference on Mobile Data Management, MDM 2018, Aalborg, Denmark, 25–28 June 2018; pp. 245–250. [CrossRef]
- Zhang, S.; Li, C.; Li, X. Predicting Location Trajectories of Humans by Their Diverse Social Ties. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, SMC 2018, Miyazaki, Japan, 7–10 October 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1221–1226. [CrossRef]
- Zhao, Z.; Ng, W. A model-based approach for RFID data stream cleansing. In Proceedings of the 21st ACM International Conference on Information and Knowledge Management, CIKM'12, Maui, HI, USA, 29 October–2 November 2012; Chen, X., Lebanon, G., Wang, H., Zaki, M.J., Eds.; ACM: New York City, NY, USA, 2012; pp. 862–871. [CrossRef]
- Fazzinga, B.; Flesca, S.; Furfaro, F.; Parisi, F. Exploiting Integrity Constraints for Cleaning Trajectories of RFID-Monitored Objects. ACM Trans. Database Syst. 2016, 41, 24:1–24:52. [CrossRef]

- Fazzinga, B.; Flesca, S.; Furfaro, F.; Parisi, F. Cleaning trajectory data of RFID-monitored objects through conditioning under integrity constraints. In Proceedings of the 17th International Conference on Extending Database Technology, EDBT 2014, Athens, Greece, 24–28 March 2014; pp. 379–390. [CrossRef]
- Baba, A.I.; Lu, H.; Pedersen, T.B.; Xie, X. Handling False Negatives in Indoor RFID Data. In Proceedings of the IEEE 15th International Conference on Mobile Data Management, MDM 2014, Brisbane, Australia, 14–18 July 2014; Zaslavsky, A.B., Chrysanthis, P.K., Becker, C., Indulska, J., Mokbel, M.F., Nicklas, D., Chow, C., Eds.; IEEE Computer Society: 2014; Volume 1, pp. 117–126. [CrossRef]
- 42. Fazzinga, B.; Flesca, S.; Furfaro, F.; Parisi, F. Interpreting RFID tracking data for simultaneously moving objects: An offline sampling-based approach. *Expert Syst. Appl.* **2020**, 152, 113368. [CrossRef]
- Lam, L.D.; Tang, A.; Grundy, J. Predicting indoor spatial movement using data mining and movement patterns. In Proceedings of the 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju Island, Korea, 13–16 February 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 223–230. [CrossRef]
- Gambs, S.; Killijian, M.O.; del Prado Cortez, M.N. Next place prediction using mobility markov chains. In Proceedings of the First Workshop on Measurement, Privacy, and Mobility, Bern, Switzerland, 10–13 April 2012; ACM: New York City, NY, USA, 2012; p. 3.
- 45. Giannotti, F.; Nanni, M.; Pedreschi, D.; Pinelli, F.; Renso, C.; Rinzivillo, S.; Trasarti, R. Unveiling the complexity of human mobility by querying and mining massive trajectory data. *VLDB J.* **2011**, *20*, 695–719. [CrossRef]
- 46. Berndt, D.J.; Clifford, J. Using Dynamic Time Warping to Find Patterns in Time Series. In Proceedings of the Knowledge Discovery in Databases: Papers from the 1994 AAAI Workshop, Seattle, WA, USA, 31 July–4 August 1994; Fayyad, U.M., Uthurusamy, R., Eds.; Technical Report WS-94-03; AAAI Press: Palo Alto, CA, USA, 1994; pp. 359–370.
- Vlachos, M.; Kollios, G.; Gunopulos, D. Elastic Translation Invariant Matching of Trajectories. *Mach. Learn.* 2005, 58, 301–334. [CrossRef]
- Cao, H.; Wolfson, O.; Trajcevski, G. Spatio-temporal data reduction with deterministic error bounds. VLDB J. 2006, 15, 211–228. [CrossRef]
- Cai, Y.; Ng, R.T. Indexing Spatio-Temporal Trajectories with Chebyshev Polynomials. In Proceedings of the ACM SIGMOD International Conference on Management of Data, Paris, France, 13–18 June 2004; Weikum, G., König, A.C., Deßloch, S., Eds.; ACM: New York City, NY, USA, 2004; pp. 599–610. [CrossRef]
- Agrawal, R.; Faloutsos, C.; Swami, A.N. Efficient Similarity Search In Sequence Databases. In Proceedings of the Foundations of Data Organization and Algorithms, 4th International Conference, FODO'93, Chicago, IL, USA, 13–15 October 1993; Lomet, D.B., Ed.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 1993; Volume 730, pp. 69–84. [CrossRef]
- Cho, E.; Myers, S.A.; Leskovec, J. Friendship and mobility: User movement in location-based social networks. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; ACM: New York City, NY, USA, 2011; pp. 1082–1090. [CrossRef]
- 52. Gonzalez, M.C.; Hidalgo, C.A.; Barabasi, A.L. Understanding individual human mobility patterns. *arXiv* 2008, arXiv:0806.1256. [CrossRef]
- Song, C.; Qu, Z.; Blumm, N.; Barabási, A.L. Limits of predictability in human mobility. *Science* 2010, 327, 1018–1021. [CrossRef] [PubMed]
- Christakis, N.A.; Fowler, J.H. Social Contagion Theory: Examining Dynamic Social Networks and Human Behavior. *Stat. Med.* 2013, 32, 556–577. [CrossRef] [PubMed]
- Mathew, W.; Raposo, R.; Martins, B. Predicting future locations with hidden Markov models. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing, Pittsburgh Pennsylvania, 5–8 September, 2012; ACM: Pittsburgh, Pennsylvania, USA, 2012; pp. 911–918. [CrossRef]
- 56. Gambs, S.; Killijian, M.O.; del Prado Cortez, M.N. Show me how you move and I will tell you who you are. In Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS, San Jose, CA, USA, 2 November 2010; ACM: New York City, NY, USA, 2010; pp. 34–41. [CrossRef]