

Article

A New Approach to Measuring the Similarity of Indoor Semantic Trajectories

Jin Zhu ^{1,2}, Dayu Cheng ^{1,3}, Weiwei Zhang ², Ci Song ^{1,4} , Jie Chen ^{1,4} and Tao Pei ^{1,4,5,*} 

- ¹ State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, CAS, Beijing 100101, China; zhujin@usts.edu.cn (J.Z.); chengdy@reis.ac.cn (D.C.); songc@reis.ac.cn (C.S.); chenj@reis.ac.cn (J.C.)
- ² School of Geography Science and Geomatics Engineering, Suzhou University of Science and Technology, Suzhou 215009, China; zhangweiwei@usts.edu.cn
- ³ School of Mining and Geomatics, Hebei University of Engineering, Handan 056038, China
- ⁴ University of Chinese Academy of Sciences, Beijing 100101, China
- ⁵ Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
- * Correspondence: peit@reis.ac.cn; Tel.: +86-10-6488-8960

Abstract: People spend more than 80% of their time in indoor spaces, such as shopping malls and office buildings. Indoor trajectories collected by indoor positioning devices, such as WiFi and Bluetooth devices, can reflect human movement behaviors in indoor spaces. Insightful indoor movement patterns can be discovered from indoor trajectories using various clustering methods. These methods are based on a measure that reflects the degree of similarity between indoor trajectories. Researchers have proposed many trajectory similarity measures. However, existing trajectory similarity measures ignore the indoor movement constraints imposed by the indoor space and the characteristics of indoor positioning sensors, which leads to an inaccurate measure of indoor trajectory similarity. Additionally, most of these works focus on the spatial and temporal dimensions of trajectories and pay less attention to indoor semantic information. Integrating indoor semantic information such as the indoor point of interest into the indoor trajectory similarity measurement is beneficial to discovering pedestrians having similar intentions. In this paper, we propose an accurate and reasonable indoor trajectory similarity measure called the indoor semantic trajectory similarity measure (ISTSM), which considers the features of indoor trajectories and indoor semantic information simultaneously. The ISTSM is modified from the edit distance that is a measure of the distance between string sequences. The key component of the ISTSM is an indoor navigation graph that is transformed from an indoor floor plan representing the indoor space for computing accurate indoor walking distances. The indoor walking distances and indoor semantic information are fused into the edit distance seamlessly. The ISTSM is evaluated using a synthetic dataset and real dataset for a shopping mall. The experiment with the synthetic dataset reveals that the ISTSM is more accurate and reasonable than three other popular trajectory similarities, namely the longest common subsequence (LCSS), edit distance on real sequence (EDR), and the multidimensional similarity measure (MSM). The case study of a shopping mall shows that the ISTSM effectively reveals customer movement patterns of indoor customers.

Keywords: indoor trajectory similarity; semantic similarity; edit distance; indoor positioning data; indoor walking distance



Citation: Zhu, J.; Cheng, D.; Zhang, W.; Song, C.; Chen, J.; Pei, T. A New Approach to Measuring the Similarity of Indoor Semantic Trajectories. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 90. <https://doi.org/10.3390/ijgi10020090>

Academic Editors: Wolfgang Kainz and Bart Kuijpers

Received: 9 December 2020
Accepted: 18 February 2021
Published: 20 February 2021

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



Copyright: © 2021 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/>).

1. Introduction

Indoor positioning devices, such as WiFi, Bluetooth, and RFID (radio frequency identification) devices, generate an extensive number of indoor trajectories for objects moving indoors. With indoor trajectories, insightful indoor movement patterns reflecting complex human spatial behavior can be discovered by adopting various methods, such as clustering analysis [1–4]. These methods are frequently based on an indoor trajectory

similarity that measures the similarity degree of two indoor trajectories. Raw trajectories embodying spatial and temporal information can be combined with semantic data to constitute semantic trajectories [5]. Semantic trajectories can enhance our understanding of the movement process, such as revealing the intention of movement. Therefore, semantic trajectory similarity measures focus on the semantic similarity and have attracted the attention of many researchers during the past few years [6–9]. However, little attention has been paid to measures of the similarity of indoor semantic trajectories.

Indoor trajectories have two characteristics distinct from those of outdoor trajectories. One characteristic is the mechanism of the positioning devices. Outdoor trajectories are usually supposed to be generated by global positioning system (GPS)-type devices. GPS devices track the positions of objects continuously. Unlike GPS devices, indoor positioning devices, such as RFID and WiFi devices, report the positions of objects in the activation range of the devices [10]. If an object leaves the activation range, its position is unknown. It is costly and almost impossible to deploy indoor positioning devices everywhere in an indoor space. Unlike outdoor objects, indoor objects cannot be tracked in some places. The other characteristic is the space model. The space model accommodating outdoor moving objects is free space, namely, unconstrained Euclidean space [11]. Migrating birds equipped with GPS receivers move in free space. Indoor space (or constrained Euclidean space [11]) comprises indoor entities (e.g., rooms, doors, and corridors), and these entities constrain indoor movement.

Figure 1 shows an indoor trajectory. The green circles are the activation ranges of indoor positioning devices. Suppose a person walks from point *a* to point *b* along an indoor walking path. The path is a winding path owing to the movement constraints imposed by the indoor space; however, only two points *a* and *b* are recorded by the positioning device along the indoor walking path. Therefore, indoor trajectories may miss more trajectory points than outdoor trajectories. Because of the indoor movement constraints and the missing trajectory points, the path of two consecutive indoor trajectory points cannot be seen as a straight line segment. The distance between the two points is not the Euclidean distance but rather the indoor walking distance.

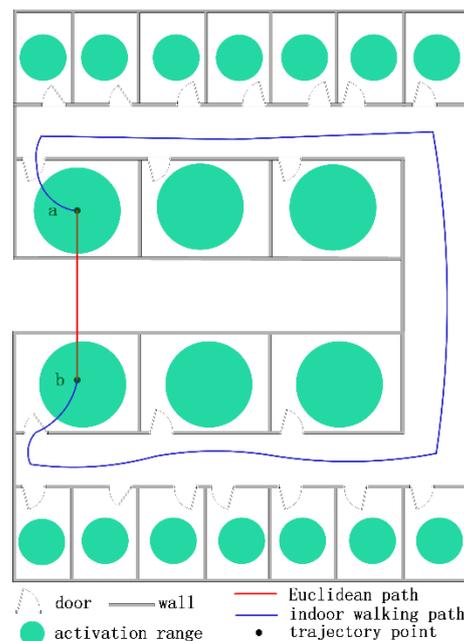


Figure 1. Indoor trajectory in an indoor space.

There is a wealth of trajectory similarity measures [12]. Representative measures include dynamic time warping (DTW) [13], longest common subsequence (LCSS) [14], and edit distance on a real sequence (EDR) measures [15]. DTW was originally used

for time series and then adapted to two-dimensional trajectories. The LCSS and EDR extend similarity for string sequences to similarity for trajectories. Although these methods measure the similarity of trajectories, their basis depends on the distance between two trajectory points. The cited works neglect the movement constraints imposed by the indoor space and the specialty of the indoor positioning device. They have an underlying assumption that the path of two consecutive trajectory points is the Euclidean path and the distance between the two points is the Euclidean distance. However, in a complicated indoor environment, the distance between two points on a walking path is typically appreciably longer than the Euclidean distance. This results in an inaccurate similarity measure when these methods are applied to indoor trajectories.

Most of the above existing trajectory similarities are defined according to the spatial aspect of trajectories. However, the spatial closeness of two trajectories does not necessarily mean that the trajectories are similar, because the semantic similarity may be small; that is, the categories of points of interest (POIs) that they pass may be different.

Previous studies were not accurate for indoor trajectories regarding the characteristics of indoor trajectories and paid less attention to the semantic aspect of indoor trajectories. The objective of the present study is to measure the similarity of indoor semantic trajectories accurately. This paper presents a novel indoor trajectory similarity measure called the indoor semantic trajectory similarity measure (ISTSM), to measure indoor trajectories accurately. The ISTSM considers the characteristics and the semantic dimension of indoor trajectories simultaneously. The ISTSM is modified from the edit distance for a string sequence and each point of the indoor semantic trajectory is viewed as a character of a string using indoor semantic information. Accurate indoor walking distances between trajectory points are obtained by incorporating an indoor navigation graph representing the complex indoor space. Indoor semantic information and indoor walking distances are fused to compute the ISTSM. The advantage of the ISTSM is that indoor movement constraints are considered, and it is a more accurate measure of the similarity of indoor semantic trajectories. Experiments conducted with a synthetic dataset and a real indoor trajectory dataset verify that the ISTSM is a more accurate and reasonable similarity measure for indoor trajectories, and it facilitates the understanding of indoor movement. The ISTSM can be applied in different indoor environments (shopping malls, railway stations and airports) and is potentially useful for applications such as indoor evacuation [16].

The remainder of this paper is organized as follows. Section 2 reviews related work on trajectory similarity. Section 3 presents our method in detail. Section 4 gives experimental evaluation results. Section 5 concludes the paper and provides an outlook for future work.

2. Literature Review

A variety of trajectory similarity measures have been proposed over a period of decades. These approaches can be classified into two categories according to whether the semantic dimension of the trajectory is considered.

The first category focuses on either the spatial dimension or the temporal dimension, or both. The Euclidean distance is the sum of Euclidean distances between corresponding points of two trajectories. The discrete segment Hausdorff distance and discrete segment Fréchet distance respectively extend the Hausdorff distance and the Fréchet distance for the point distance to segment distance [17]. DTW is modified from one-dimensional time series, but is sensitive to noise [13]. The LCSS is robust against noise, but ignores possible non-matching gaps in the two trajectories [14]. Kang et al. extended the LCSS to compute the indoor trajectory similarity [18]. Their method considers the common visit time interval in which two indoor moving objects stay in the same indoor entity. However, as inherited from the LCSS, the similarity ignores the spatial distance between mismatched points. The EDR proposed by Chen et al. also reduces the effect of noise and avoids the gap disadvantage of the LCSS [15]. Wang et al. proposed a distance called the edit distance combined with Euclidean distance (EDEU) to measure the similarity of RFID

trajectories [19]. However, the EDEU does not consider indoor movement constraints and the Euclidean distance is not accurate for indoor moving objects.

In recent years, the study of trajectory has shifted from raw trajectories to semantic trajectories embodying meaningful semantic information. Semantic trajectories are raw trajectories combined with related contextual information, such as POIs, land use, and weather. The second category of trajectory similarity takes into account the semantic aspect of the trajectory and can help us understand trajectories well. Ying et al. proposed the maximal semantic trajectory pattern (MSTP) similarity [6]. Frequent semantic trajectories are first mined from semantic trajectories, and a modified LCSS is then applied to calculate the MSTP similarity. Wan et al. put forward a semantic–geographic similarity that considers semantic similarity and geographic similarity simultaneously [20]. The semantic similarity and geographic similarity are computed using the cosine and Hausdorff distances, respectively. Furtado et al. introduced the multidimensional similarity measure (MSM), which considers the spatial, temporal, and semantic dimensions altogether [7]. The cited studies considered stops, where moving objects stay still for a certain amount of time. Lehmann et al. recently proposed the stops and moves similarity measure, which considers both stops and moves [8]. Petry et al. presented a multi-aspect trajectory similarity measure, which considers the relationship between semantic attributes [9]. Jin et al. introduced an indoor trajectory similarity based on spatial and hierarchical semantic similarity [2]. The spatial similarity is measured using a distance in three-dimensional space and the hierarchical semantic similarity is computed using a semantic classification tree. Neither outdoor nor indoor trajectory similarities consider indoor movement constraints and indoor semantic information simultaneously, and they are thus not accurate for indoor applications.

Raw indoor positioning data contain noise, and are inherently uncertain due to the limitations of indoor positioning devices and the complicated indoor environment. Plenty of methods have been proposed to clean indoor raw data. Most of these studies consider the indoor constraints through various approaches, such as the graph model-based approach [21], the hidden Markov model (HMM)-based approach [22], the particle filtering approach [23], the probabilistic conditioning approach [24] and the Metropolis–Hastings approach [25]. Data cleansing is the data preprocessing phrase, but our method differs from these approaches in that ISTSM integrates indoor walking distance directly into the indoor trajectory similarity measure.

The indoor walking distance considers the constraints imposed by the indoor space and is more accurate than the Euclidean distance for an indoor space. An indoor navigation graph is needed to compute the indoor walking distance. The indoor navigation graph is a graph model supporting indoor navigation for indoor objects. A node of the graph represents a location in the indoor space, while an edge of the graph represents a path between nodes. Researchers have proposed a variety of indoor navigation graph models, such as the generalized Voronoi diagram [26], Delaunay triangulation (DT) [27], the visibility graph model (VG) [28], straight medial axis transformation (S-MAT) [29], and the grid graph [30], to name a few. These methods use various techniques to subdivide an indoor space into a set of entities and then construct navigation graph models. Each model has its own pros and cons. Hahmann et al. compared these models using four criteria: the number of graph edges, graph creation time, route computation time, and route quality [31]. They identified that the VG delivers the shortest possible routes and is a promising method; however, the VG creates many redundant edges and requires optimization. DT produces satisfying route results most times, but sometimes gives modest results. Nevertheless, the divergence between DT and the VG is not large. DT is therefore employed to construct the indoor navigation graph in the present study.

3. Methodology

The overall flow of the ISTSM has three steps and is shown in Figure 2. Indoor raw trajectories are converted into indoor semantic trajectories that comprise semantic labels. The ISTSM is based on the edit distance between string sequences. We treat semantic

labels as characters of a string sequence. The edit distance seeks the minimum cost of transforming one sequence into another sequence, with insertion, deletion, and substitution operations on the characters. To fuse accurate indoor distances into the ISTSM, the indoor floor plan representing the indoor space is transformed into an indoor navigation graph. With the navigation graph, indoor walking distances between semantic trajectory points are calculated. The substitution costs of semantic trajectory points are calculated using the indoor walking distances and semantic information of indoor entities. Finally, the ISTSM of two trajectories is computed with a classic dynamic programming algorithm.

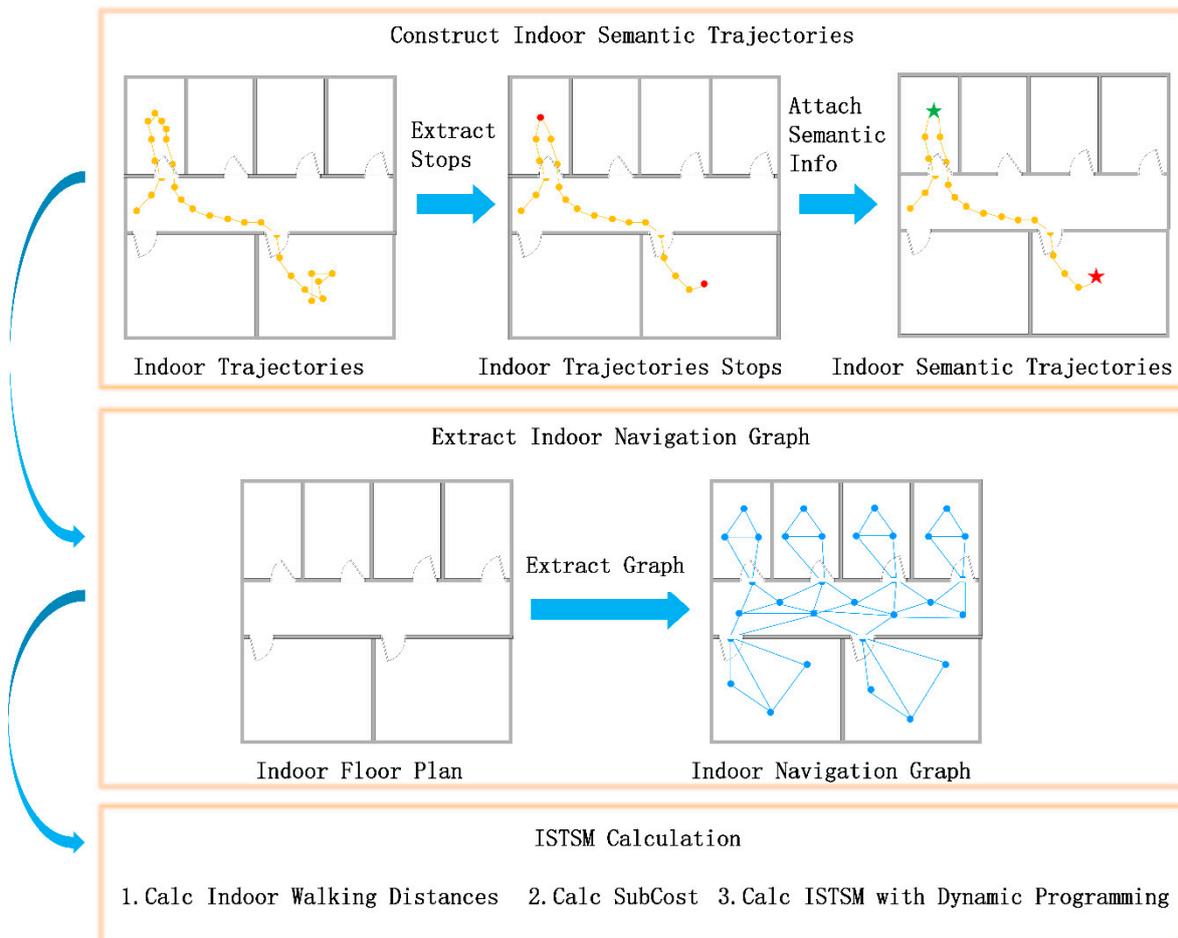


Figure 2. Overall flow of the indoor semantic trajectory similarity measure (ISTSM).

3.1. Constructing Indoor Semantic Trajectories

Several related definitions of an indoor trajectory are first stated.

Definition 1. Indoor Trajectory. An indoor trajectory (ITr) is a sequence of space–time points with time stamps when the indoor moving object moves in an indoor space.

$$ITr = \{(p_1, t_1), \dots, (p_i, t_i), \dots, (p_n, t_n)\} \quad (1)$$

Point $p_i = (x_i, y_i)$ is the indoor position that is recorded by indoor positioning sensors at time stamp $t_i (i = 1, \dots, n)$, x_i and y_i are the coordinates for the point position p_i .

Definition 2. Indoor Semantic Trajectory. The indoor semantic trajectory ($ISTr$) of an indoor moving object is a sequence of space–time points with semantic labels.

$$ISTr = \{(p_1, s_1, t_1), \dots, (p_i, s_i, t_i), \dots, (p_n, s_n, t_n)\} \quad (2)$$

Here, s_i is the indoor semantic label associated with the corresponding indoor entity in which p_i is located at time stamp $t_i (i = 1, \dots, n)$.

Constructing indoor semantic trajectories from raw indoor trajectories involves two steps, as in Figure 3. The first step involves extracting stops from indoor trajectories. A trajectory can be viewed as a sequence of alternating stops and moves [32]. Stops are trajectory episodes where the moving object stays in a region for a while, and can be represented by the centroid of the points of the trajectory episode. Moves are trajectory episodes in which the moving object keeps moving. Stops are usually important or interesting places and have semantic information. The second step is attaching indoor semantic labels to stops.

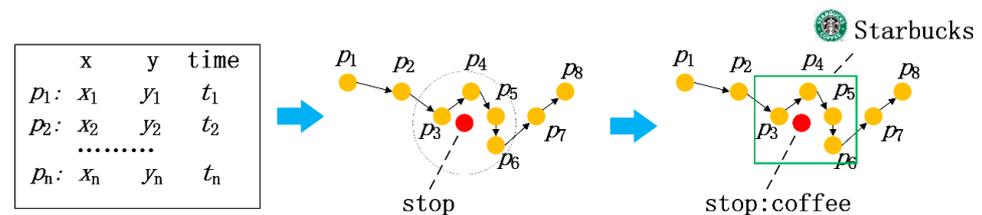


Figure 3. Extracting indoor semantic trajectories.

Stops can be extracted from raw indoor trajectories with the stop detection algorithm [33]. The stop detection algorithm relies on two parameters. One is the distance threshold θ_d that restricts the region size of the stop. In Figure 3, the radius of the circle is θ_d . The other is the time threshold θ_t representing the minimal amount of time for which the moving object must stay. After all stops are extracted, a trajectory can be seen as a sequence of stops.

Stops have semantic meanings and can reflect the goal and status of the moving object. For an indoor space, each indoor entity can be represented by a polygon. If a stop is within the polygon of an entity, the semantic label of the entity is attached to the stop. In Figure 3, the green rectangle is a Starbucks coffee shop. As the stop is within the shop, the stop is attached to the coffee shop semantic label. After all stops are attached to semantic information, indoor trajectories are transformed into indoor semantic trajectories.

3.2. Extracting an Indoor Navigation Graph

The accurate calculation of indoor walking distances relies on the indoor navigation graph. An indoor navigation graph represents an indoor space as a graph used for navigation. The indoor space can be partitioned into various types of indoor entity, such as rooms, corridors, and stairs. Each indoor entity can be represented by a graph node and each graph edge represents connections between indoor entities. The indoor navigation graph extracted from an indoor floor plan can be used to calculate the indoor walking distances between entities.

In this work, we employ DT to build the navigation graph. Figure 4 depicts the detailed steps. Given a floor plan (a), the indoor space of the floor plan is extracted as a whole and converted into a polygon. The vertices of the polygon are used to generate the first triangulated irregular network (TIN). Afterward, the centroids of each triangle in the first TIN are extracted (b) and employed to generate the second TIN (c). Finally, the nodes and edges of the second TIN that are completely within the indoor space are extracted to generate the indoor navigation graph (d).

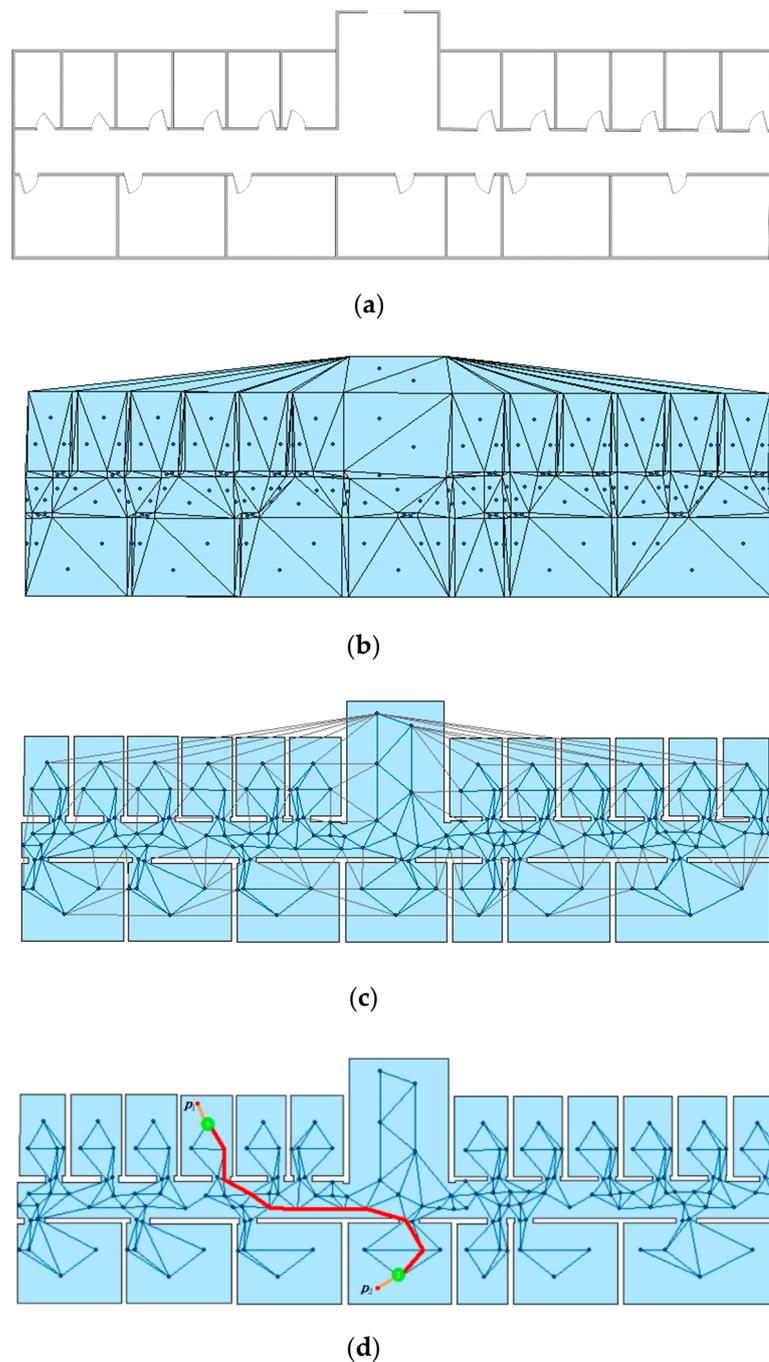


Figure 4. Construction of the indoor navigation graph and computation of indoor walking paths: (a) the floor plan; (b) generation of the first triangulated irregular network (TIN) and the extraction of the centroid of each triangle; (c) generation of the second TIN; and (d) final indoor navigation graph and the computation of the indoor walking path.

Figure 4d shows the computation of the indoor walking distance between any two points. To calculate the indoor walking distance between two points p_1 and p_2 , $IndoorDist(p_1, p_2)$, we first search the nearest graph nodes n_1 (denoted as ①) and n_2 (denoted as ②) for p_1 and p_2 separately. After obtaining n_1 and n_2 , the shortest path between n_1 and n_2 is computed using the Dijkstra shortest path algorithm [34], and the path distance is denoted as $dist(n_1, n_2)$. The Euclidean distance between p_1 and n_1 is denoted as $dist(p_1, n_1)$, and

the distance between p_2 and n_2 is denoted as $dist(p_2, n_2)$. Finally, $IndoorDist(p_1, p_2)$ is calculated as the sum of $dist(p_1, n_1)$, $dist(n_1, n_2)$, and $dist(p_2, n_2)$, as in Equation (3).

$$IndoorDist(p_1, p_2) = dist(p_1, n_1) + dist(n_1, n_2) + dist(p_2, n_2) \quad (3)$$

3.3. ISTSM Computation

The sequence of semantic labels representing indoor semantic trajectories is exploited for determining similarity. The edit distance is extended as the ISTSM to compute the similarity of indoor semantic trajectories. The edit distance was originally used for comparing string sequences and has applications in information retrieval and computational biology. As a classic edit distance, the Levenshtein distance is defined as the minimum number of operations that transfer one string to another, where the operations include insertion, deletion, and substitution [35]. Here, the costs of insertion, deletion, and substitution are all 1. The ISTSM treats the semantic labels of an indoor semantic trajectory as a string characters. Similar to the case for the Levenshtein distance, the insertion and deletion costs of the ISTSM are defined as 1. However, the substitution cost is defined in terms of semantic labels and indoor walking distances between stops. The contribution of the ISTSM is that it integrates semantic information and accurate indoor walking distances into the edit distance such that it provides an accurate measure of the similarity of semantic trajectories.

The substitution cost of semantic labels fuses indoor walking distances and indoor semantic information. For two semantic labels of different semantic types, we consider that the difference between the two semantic labels is large, while the substitution cost is defined as 1, the same as the insertion and deletion costs. For two semantic labels of the same semantic type, the difference between the two semantic labels relates to the indoor walking distance between the indoor entities with which the two semantic labels are associated. The substitution cost is defined as the ratio of the indoor walking distance between the centroids of the two entities and the maximum indoor walking distance between indoor entities both having the same semantic type. The substitution cost is thus between 0 and 1. In this way, the substitution cost not only differentiates different semantic types, but also reflects the relative distances of stops of the same semantic type.

An indoor entity can be represented by the centroid of the polygon of the indoor entity. The indoor walking distance between any two indoor entities can be computed with the indoor navigation graph built previously.

Formally, the substitution cost $subCost$ is defined by Equation (4).

$$subCost(s_i, s_j) = \begin{cases} 1 & \text{if } s_i.SType \neq s_j.SType \\ \frac{IndoorDist(s_i, s_j)}{IndoorDist_{max}^{s_i.SType}} & \text{otherwise} \end{cases} \quad (4)$$

Here, s_i and s_j are two semantic labels (i.e., indoor entities), $SType$ is their semantic type information, $IndoorDist(s_i, s_j)$ is the indoor walking distance between the indoor entities s_i and s_j , and $IndoorDist_{max}^{s_i.SType}$ is the maximum indoor walking distance between two indoor entities that are both of the same semantic type as $s_i.SType$.

When s_i and s_j are of different semantic types ($s_i.SType \neq s_j.SType$), the substitution cost is directly assigned the maximum value of 1. When they are of the same semantic type, the substitution cost is between 0 and 1. The substitution cost is zero only when s_i and s_j refer to an identical indoor entity ($i = j$). The substitution cost is 1 when s_i and s_j refer to two indoor entities that have a maximum indoor walking distance of the same semantic type. After the substitution cost of all indoor entity pairs is computed, the result is stored in the substitution cost matrix $subCostMat$.

Dynamic programming is a common and efficient method for computing the edit distance. Using $subCostMat$, the edit distance matrix $D_{0...n,0...m}$ for two indoor semantic

trajectories $P[1 \dots n]$ ($|P| = n$) and $Q[1 \dots n]$ ($|Q| = m$) is computed element by element, adopting the dynamic programming method expressed as Equation (5).

$$D_{i,j} = \begin{cases} j & \text{if } i = 0 \\ i & \text{if } j = 0 \\ D_{i-1,j-1} & \text{if } i, j > 0, \text{ and } P_i = Q_j \\ \min \begin{cases} D_{i,j-1} + 1 \\ D_{i-1,j} + 1 \\ D_{i-1,j-1} + \text{subCostMat}(P_i, Q_j) \end{cases} & \text{otherwise} \end{cases} \quad (5)$$

When $D_{0\dots n,0\dots m}$ has been computed, $D_{n,m}$ is the distance for the two indoor semantic trajectories. $D_{n,m}$ is related to the lengths of the two trajectories. $D_{n,m}$ is normalized using the trajectory lengths n and m in Equation (6) to eliminate the effect of the trajectory length. Liu et al. showed that the result of normalization is a metric distance that can be used with indexing technology to accelerate the trajectory distance query [36].

$$D_{norm} = \frac{2 \times D_{n,m}}{D_{n,m} + n + m} \quad (6)$$

The value of distance $D_{n,m}$ ranges from 0 to 1. The ISTSM is therefore the opposite of $D_{n,m}$ and is defined by Equation (7).

$$ISTSM = 1 - D_{norm} = 1 - \frac{2 \times D_{n,m}}{D_{n,m} + n + m} \quad (7)$$

4. Experimental Evaluation

The ISTSM was evaluated with two datasets: a synthetic dataset and a real dataset. The effectiveness of the ISTSM was evaluated in comparison with popular trajectory similarities in the synthetic dataset. Customer movement patterns can reflect similar movement characteristics for a group of shopping customers in shopping malls. The study of customer movement patterns is valuable in understanding customer shopping behaviors, thus supporting applications such as sales promotion and friend recommendation. Customer movement patterns can be discovered from indoor customer trajectories. A real dataset from a shopping mall was used to perform clustering analysis with the ISTSM and thus determine customer movement patterns in a case study.

4.1. Experimental Analysis of the Synthetic Dataset

4.1.1. Data Transformation Methodology

The synthetic dataset was generated using a data transformation method to evaluate the effectiveness of the ISTSM. The idea is borrowed from Refs. [7,37]. The advantage of the method is that the ISTSM can be evaluated and compared for different conditions in a controlled manner.

First, assuming a floor plan was given as an indoor space, an indoor semantic seed trajectory $ISTr = \{S_1, \dots, S_n\}$ (where point locations and time stamps are omitted) was generated within the indoor space. Two different types of trajectory transformation were then applied to the seed trajectory to generate new trajectories. The trajectory transformation method had a varying rate that controlled the amount of transformation. For trajectory transformation type K with varying rate r , a set of η transformed indoor semantic trajectories $W_r^K = \{ISTr_1, \dots, ISTR_\eta\}$ was generated. Afterwards, we computed the average ISTSM between $ISTr$ and each transformed trajectory set W_r^K for a different varying rate r , and we then compared the result with representative trajectory similarities, namely the EDR, LCSS, and MSM. While DTW is also a classic trajectory similarity, it is sensitive to noise, and previous work has revealed that the EDR and LCSS are superior to DTW [15]. Therefore, the ISTSM was not compared with DTW. The two trajectory transformation methods are as follows.

1. Change in order of stops. A number of stops controlled by varying rate r were randomly selected and the positions of these stops were changed. Figure 5 presents the order change of one stop labeled Lab 1. Note that the varying rate r may affect more stops. In Figure 5, one stop order change ($r = 20\%$) affects two stops for the seed trajectory.

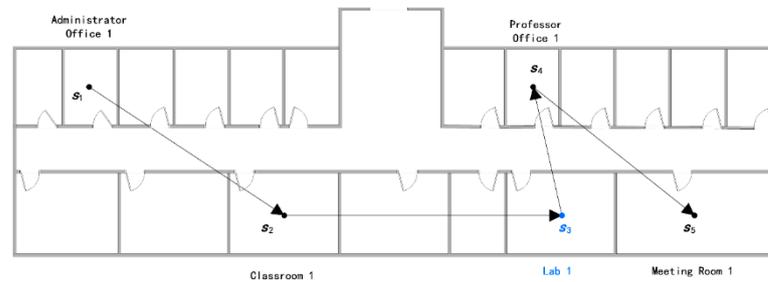


Figure 5. Order change of stops for $r = 20\%$.

2. Replacement of stops. This transformation randomly generated a number of stops controlled by varying the rate r , and replaced the stops to generate transformed trajectories. Figure 6 presents the replacement of one stop labeled Lab 1 with another stop labeled Professor Office 2.

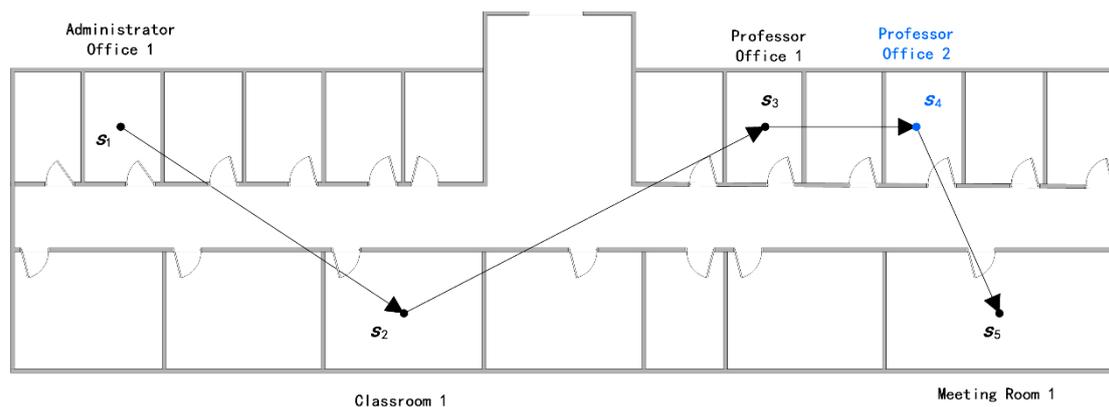


Figure 6. Replacement of stops for $r = 20\%$.

The floor plan of Figure 4 was used as the indoor space in this experiment. The seed indoor semantic trajectory had 10 stops. Although a raw trajectory may have hundreds or thousands of points, there are far fewer stops on the transformed semantic trajectory. The length of 10 for a seed indoor semantic trajectory is thus appropriate. The number of transformed trajectories η was set to 1000 for each type of transformation. The varying rate r was the proportion of transformed stops and ranged from 0.1 to 1 in steps of 0.1. The experiments were performed on a laptop personal computer with an i5-5200U 2.20-Hz CPU and 16 GB memory.

The ISTSM was compared with the EDR, LCSS, and MSM. The ISTSM has no input parameter, whereas the EDR, LCSS, and MSM had at least one input parameter. The EDR had a parameter ϵ as the space-matching threshold. The value of ϵ was set to 3 (where the unit is meters in our indoor space), because the room width was 6 for the input indoor space and half of the width is a proper value for ϵ . The LCSS method has two parameters: ϵ is the same as for the EDR and δ is the translation threshold in the time dimension. ϵ was set to 3, the same as for the EDR. δ was set to 2 to match nearby points in the time dimension. The MSM is a multidimensional similarity where each dimension has its own weight, distance function, and threshold. To compare with the ISTSM that focuses on space and semantic dimensions, the weights for space and semantic dimensions were both set at 0.5. The spatial distance function is the Euclidean distance. For the semantic distance function, the distance is zero when the semantics match, and is 1 otherwise. The spatial

threshold ε_1 was set at 3, the same as for the EDR and LCSS, and the semantic threshold ε_2 was set at 0.5.

4.1.2. Experimental Results

1. Change in order of stops. The average similarity scores computed for the seed trajectory and the transformed trajectories with a varying rate of change in the order of stops are shown in Figure 7. The value of the MSM was always 1 for different rates because it did not depend on the stop order. The ISTSM was between the LCSS and EDR. We see that the ISTSM was a little larger than the EDR. This was because the ISTSM considered the semantic dimension and indoor walking distances between stops, whereas the EDR did not. That is to say, two stops of the same semantic type were more similar than two stops having different semantic types.

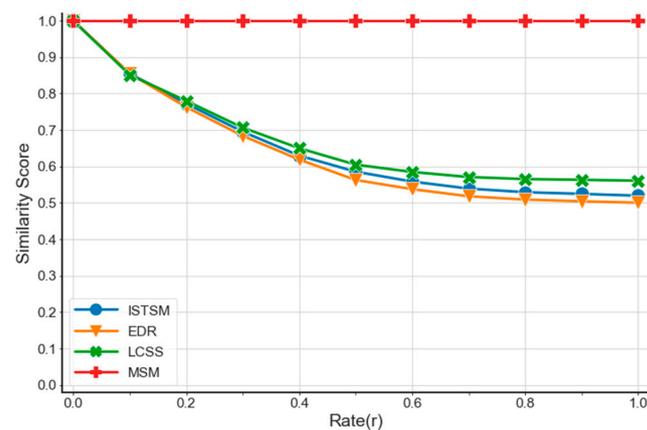


Figure 7. Result of changing the stop order, where the x axis is the trajectory transformation rate r and the y axis is the similarity score of the original trajectory and transformed trajectory.

2. Replacement of stops. The average similarity scores computed for the seed trajectory and transformed trajectories with a varying rate of replaced stops are illustrated in Figure 8. All four similarities decreased as the varying rate increased. The (ascending) order of the four similarities was LCSS, EDR, ISTSM, and MSM, at a given varying rate. The LCSS was the lowest among them, as it ignored the similarity between the original stops and the replaced stops when they were of the same semantic type. The MSM was the highest because it did not consider the stop order and searched for the best matching stops. The LCSS thus tends to underestimate similarity, while the MSM tends to overestimate similarity. The ISTSM and EDR were between the LCSS and MSM. The reason for the ISTSM being slightly higher than the EDR is the same as described above.

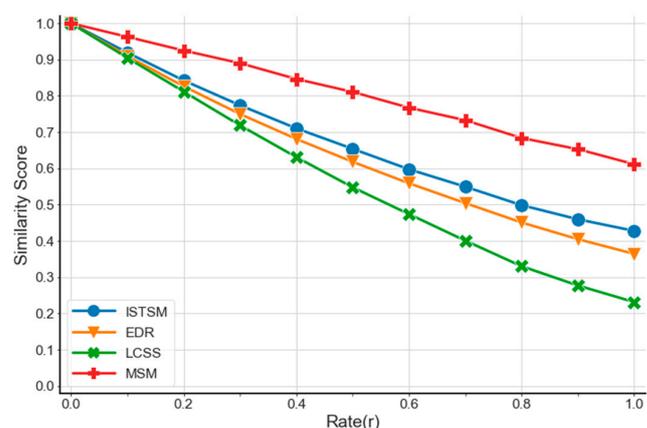


Figure 8. Result of replacing stops, where the x axis is the trajectory transformation rate r , while the y axis is the similarity score of the original trajectory and transformed trajectory.

The above two figures reveal that the ISTSM considers the semantic information and is able to detect the small difference in stops with the same semantic type. We will later show that the difference relates to corresponding indoor walking distances between stops. With the consideration of the indoor semantic information and the complex indoor environment, high-level knowledge may be discovered from trajectories using the ISTSM.

4.2. Case Study—Determining Customer Movement Patterns

A real dataset of a shopping mall was used to perform clustering analysis with the ISTSM in determining customer movement patterns. The real dataset contains customer indoor trajectories for 1 day in a shopping mall. The shopping mall is a six-floor building in Chongqing, China. For brief and clear presentation, we selected the fourth floor as the indoor space and extracted customer indoor trajectories passing through the fourth floor, as shown in Figure 9. There are 102 indoor entities on the fourth floor that can be classified into nine semantic types: elevator, clothing, restaurant, general merchandise, coffee, shoe/bag, toy, home furnishing, and massage entities. The ISTSM values of six sample customer trajectories were first computed for trajectory comparison to demonstrate the advantages of the ISTSM. Clustering analysis was then performed for all customer trajectories to determine customer movement patterns.

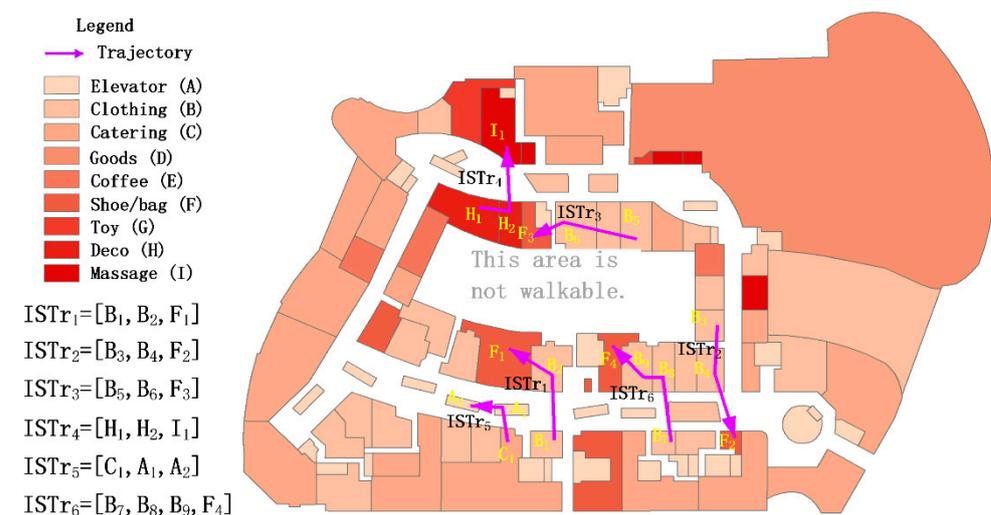


Figure 9. Floor plan of a shopping mall and six indoor semantic sub-trajectories.

4.2.1. Trajectory Comparison

Raw indoor trajectories were first preprocessed to extract indoor semantic trajectories. The number of raw indoor trajectories was 799. The distance threshold parameter θ_d and the time threshold parameter θ_t of the stop detection algorithm were respectively set at 5 m and 180 s. After the stops were detected, all stops were attached to semantic labels of corresponding indoor entities. We then filtered those indoor semantic trajectories that had only one stop. Finally, 301 indoor semantic trajectories remained.

To simplify the trajectory comparison, from the 301 indoor semantic trajectories, we selected six sample trajectories and extracted a trajectory episode as a sub-trajectory from each of the six trajectories. The six sub-trajectories are shown in Figure 9. In the figure, the nine semantic types of different indoor entities are shown according to the legend and represented by the nine capital letters A–I. The subscript of semantic labels refers to different indoor entities. As examples, A_1 and A_2 are two different semantic labels (indoor entities), both of semantic type A.

With the six indoor semantic sub-trajectories $ISTr_1, ISTR_2, \dots, ISTR_6$, we calculated ISTSM values for $ISTr_1$ and $ISTr_2, \dots, ISTR_6$. The result is shown in Figure 10. We see that $ISTSM(ISTr_1, ISTR_2)$ is the largest, having a value of 0.61, while $ISTSM(ISTr_1, ISTR_6)$ is the second largest at 0.57, and $ISTSM(ISTr_1, ISTR_3)$ is the third largest at 0.36.

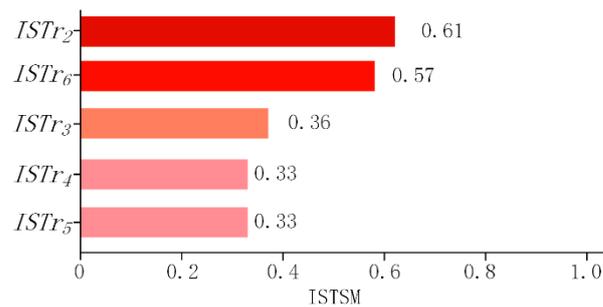


Figure 10. ISTSM of $ISTr_1$ and $ISTr_2$ – $ISTr_6$.

The semantic types ($[H,H,I]$ and $[C,A,A]$) of the semantic labels of the trajectories $ISTr_4$ and $ISTr_5$ are different from that of $ISTr_1$ ($[B,B,F]$), and $ISTSM(ISTr_1, ISTR_4)$ and $ISTSM(ISTr_1, ISTR_5)$ are thus both equal to 0.33 without considering the indoor walking distance according to Equation (5). The result illustrates the effect of the semantic dimension on the ISTSM.

$ISTSM(ISTr_1, ISTR_2)$ is larger than $ISTSM(ISTr_1, ISTR_3)$. The reason for this is that all three indoor semantic trajectories ($ISTr_1$, $ISTr_2$, and $ISTr_3$) had the same semantic type ($[B,B,F]$) of semantic labels, and these ISTSM values thus depended on the indoor walking distances of these semantic labels according to Equation (4). Note that in Figure 9, the central area is empty and not walkable. Thus, if a customer would like to walk from B_1 to B_5 , he or she has to walk around the central area. As a result, $IndoorDist(B_1, B_5)$ is larger than $IndoorDist(B_1, B_3)$. By the same token, $IndoorDist(B_2, B_6)$ is larger than $IndoorDist(B_2, B_4)$, while $IndoorDist(F_1, F_3)$ is larger than $IndoorDist(F_1, F_2)$. Therefore, for $ISTr_1$ and $ISTr_3$, the indoor walking distances between all the corresponding semantic labels are larger than those of $ISTr_1$ and $ISTr_2$. Consequently, $ISTSM(ISTr_1, ISTR_2)$ is larger than $ISTSM(ISTr_1, ISTR_3)$. The result shows the effect of the indoor walking distance on the value of the ISTSM.

Interestingly, $ISTSM(ISTr_1, ISTR_6)$ is larger than $ISTSM(ISTr_1, ISTR_3)$ but is smaller than $ISTSM(ISTr_1, ISTR_2)$. For $ISTr_6$, the semantic type ($[B,B,B,F]$) of semantic labels is similar to that of $ISTr_2$ ($[B,B,F]$). The indoor walking distances between all the corresponding semantic labels of $ISTr_1$ and $ISTr_2$ are larger than those between $ISTr_1$ and $ISTr_6$, but $ISTr_6$ (4) is longer than $ISTr_2$ (3). As such, $ISTSM(ISTr_1, ISTR_6)$ (0.57) is a little smaller than $ISTSM(ISTr_1, ISTR_2)$ (0.61). For $ISTr_6$ and $ISTr_3$, the indoor walking distances between all the corresponding semantic labels of $ISTr_1$ and $ISTr_3$ are much larger than those of $ISTr_1$ and $ISTr_6$. $ISTr_6$ (4) is also longer than $ISTr_3$ (3), and $ISTSM(ISTr_1, ISTR_6)$ (0.57) is thus larger than $ISTSM(ISTr_1, ISTR_3)$ (0.36). This demonstrates the effect of the trajectory length on the ISTSM.

The ISTSM can be integrated with a clustering algorithm to discover different groups of trajectories. Figure 11 illustrates the ISTSM matrix of the six indoor semantic trajectories and the dendrogram resulting from the agglomerative hierarchical clustering with the ISTSM. In Figure 11, the cell value of the grid is the ISTSM value, and the numbers (3, 6, 1, 2, 4, 5) on the left and top of the grid are the IDs of the six trajectories. The dendrogram is on top of the grid. The agglomerative hierarchical clustering algorithm begins with each trajectory as a cluster, and then, at each step, combines two clusters with the largest ISTSM (least distance) as a new cluster until only one cluster remains. The clustering algorithm first clustered $ISTr_1$ and $ISTr_2$ to form the combined cluster p , as these have the largest ISTSM value (0.61), and cluster p is then clustered with $ISTr_6$ to form cluster q . The process is iterated until the root cluster t is formed. We see from the dendrogram that the clustering result is reasonable. $ISTr_4$ and $ISTr_5$ in the clusters are different in the semantic dimension from the trajectories in cluster r . The trajectory clustering order of cluster r reflects the effect of indoor walking distances on the ISTSM.

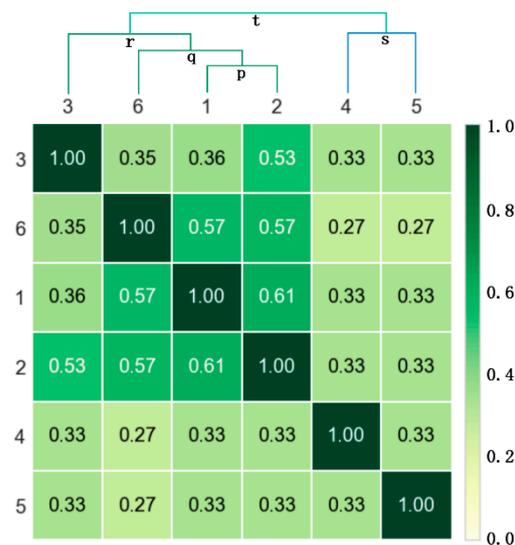


Figure 11. ISTSM matrix and dendrogram of hierarchical clustering.

In summary, the above results show that the ISTSM is effective, accurate, and reasonable for indoor semantic trajectories, with the consideration of indoor walking distances and the semantic information of stops.

4.2.2. Precision Evaluation of ISTSM

In this section, the precision of ISTSM is evaluated with the information retrieval approach [38,39]. Indoor trajectories were grouped into different groups and were used as the ground truth trajectories. All trajectories in the same group are considered to be relevant and trajectories in different groups are considered to be non-relevant. The relevance of trajectories was measured with different trajectory similarity measures. The most similar trajectory is considered to be the most relevant one. Thus, the precision evaluation task of trajectory similarity measurement is turned into an information retrieval task. The precision of ISTSM can be evaluated with mean average precision (MAP) and precision at recall. ISTSM was compared with EDR, LCSS and MSM.

From 301 indoor semantic trajectories, two trajectory groups were selected as the ground truth trajectories based on the different semantic types of stops. One group is composed of trajectories having at least two stops of *clothes* semantic type, and the other group is composed of trajectories having at least two stops of *canteen* semantic type. There are 76 trajectories in the first group and 43 trajectories in the second group. The trajectories in one group were assumed to be more similar than the trajectories in the other group.

The parameters of ISTSM, EDR, LCSS and MSM were set as in Section 4.1.1. Table 1 shows the MAP for the four trajectory similarities. ISTSM (0.80) has the largest MAP. This is because other trajectory similarities do not consider the semantic information and the indoor constraints at the same time. MSM considers the semantic aspect but not the indoor walking distances, and has the second largest MAP (0.75). LCSS takes into account the stops with the same semantic types, so its MAP can reach 0.71. EDR has the lowest MAP (0.68), as it cannot deal with semantic information. Figure 12 demonstrates the precision at recall. ISTSM outperformed the other trajectory similarities at all recall levels. When the recall level is between 0 and 0.5, the precision of ISTSM shows a significant improvement compared with the other three other trajectory similarities.

Table 1. MAP evaluation result.

	ISTSM	MSM	LCSS	EDR
MAP	0.80	0.75	0.71	0.68

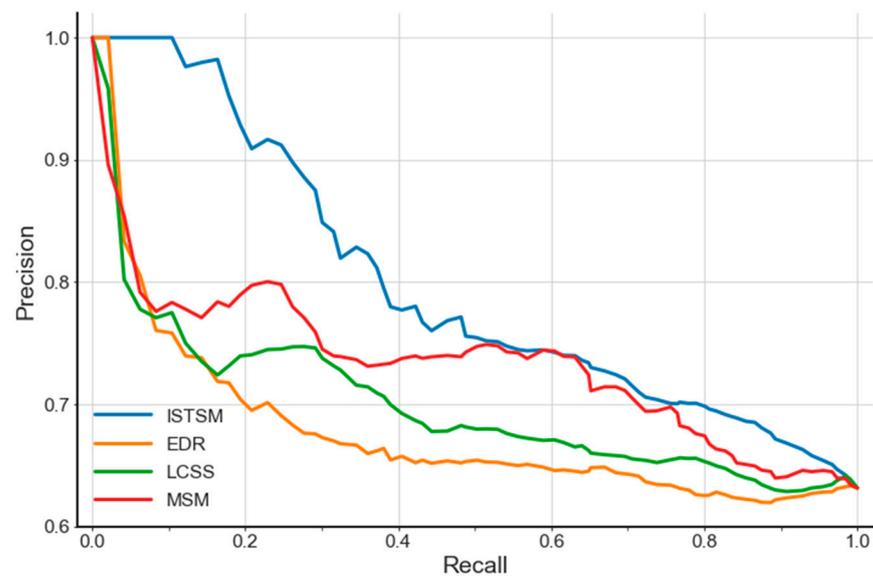


Figure 12. Precision at recall evaluation result.

4.2.3. Determining Customer Movement Patterns

The agglomerative hierarchical clustering algorithm was applied to the overall shopping mall dataset to determine customer movement patterns. Determining the number of clusters is a nontrivial task. To find the number of clusters, the knee-point detection method was adopted, and a graph that has the number of clusters on the x axis and the merge distance on the y axis is shown in Figure 13. The number of clusters that we tried ranged from 1 to 20. The merge distance is the sum of squared distances from each trajectory to its assigned cluster center. The knee-point detection method is a popular method that can identify the correct number of clusters with a mathematical approach [38]. First, a straight line is drawn from the first point to the last point. Euclidean distances between each point to the straight line are then computed. The point of the largest distance is the knee-point. In Figure 13, the point of three clusters is the knee-point, and the dataset is thus clustered into three clusters. Each cluster reveals the spatial and semantic characteristics of the customer movement patterns. The spatial and semantic patterns of the clusters are thus described in the following.

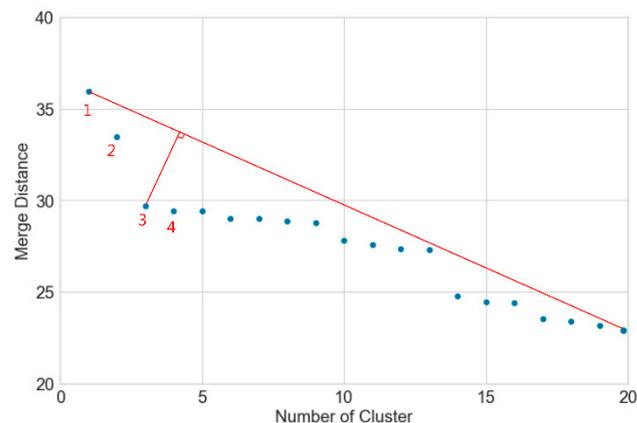


Figure 13. Determining the number of clusters for hierarchical clustering.

The spatial patterns of the three trajectory clusters are presented in Figure 14. For an indoor semantic trajectory, we split the trajectory into consecutive semantic origin–destination (OD) flows. As an example, an indoor semantic trajectory $\{s_1, s_2, s_3\}$ is split into two OD flows, $\{s_1, s_2\}$ and $\{s_2, s_3\}$. The centroid of an indoor entity is used to represent

the corresponding semantic label of the indoor entity. For all OD flows of one trajectory cluster, we count the number of distinct semantic label pairs and show them with graduated symbols. The three clusters have obvious differences in their spatial patterns. For cluster 1, the OD flows of $\{A_1, B_1\}$, $\{A_1, B_2\}$, $\{B_1, B_3\}$, $\{C_1, B_3\}$, and $\{C_2, B_5\}$ are strong compared with the other flows. All these OD flows except $\{C_2, B_5\}$ are located in the central right part of the shopping mall. Owing to the small trajectory number of cluster 2, $\{B_1, C_1\}$ is the only strong OD flow. For cluster 3, $\{A_1, B_1\}$, $\{A_1, B_2\}$, and $\{A_2, B_4\}$ are strong OD flows, and $\{A_2, B_4\}$ is the OD flow that cluster 1 does not have. The next largest flows with a flow count of 6–10 also have clear differences from cluster 1.



Figure 14. Cont.

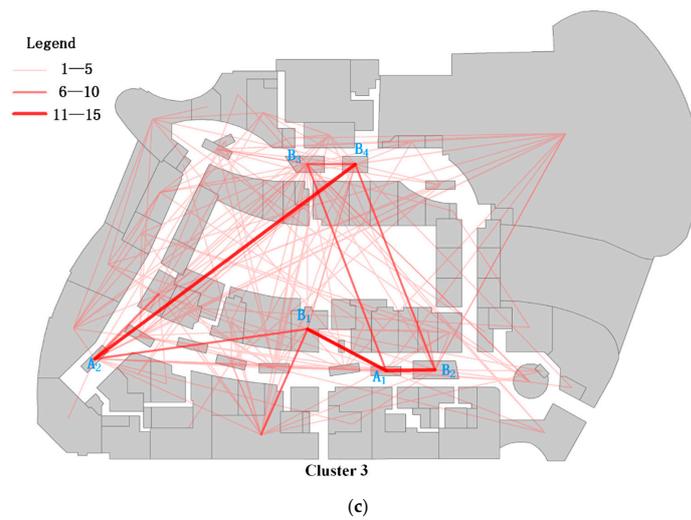


Figure 14. Spatial patterns of indoor semantic trajectory clusters: (a) cluster 1; (b) cluster 2; and (c) cluster 3.

A chord diagram is adopted to show the relationship of the semantic labels, and thus reveal the semantic pattern of one cluster, in Figure 15. The nine semantic types of semantic label are represented by nine fragments of the outer circle of the diagram, and arcs are drawn between them. The size of the arc is proportional to the number of semantic label pairs. We see that $\{catering, elevator\}$ and $\{clothing, elevator\}$ are strong semantic flows for cluster 1, while $\{catering, clothing\}$ and $\{clothing, clothing\}$ are strong semantic flows for cluster 2. Cluster 3 is similar to cluster 1, but the flows of $\{catering, catering\}$, $\{catering, massage\}$, and $\{catering, shoe/bag\}$ are stronger than those of cluster 1. Interestingly, $\{catering, catering\}$ is a relatively strong flow for cluster 3. A possible reason for this is that wandering customers do not have an appointed restaurant beforehand. When they encounter a restaurant, they spend some time browsing the menu and if they do not like the menu, they move on to the next restaurant.

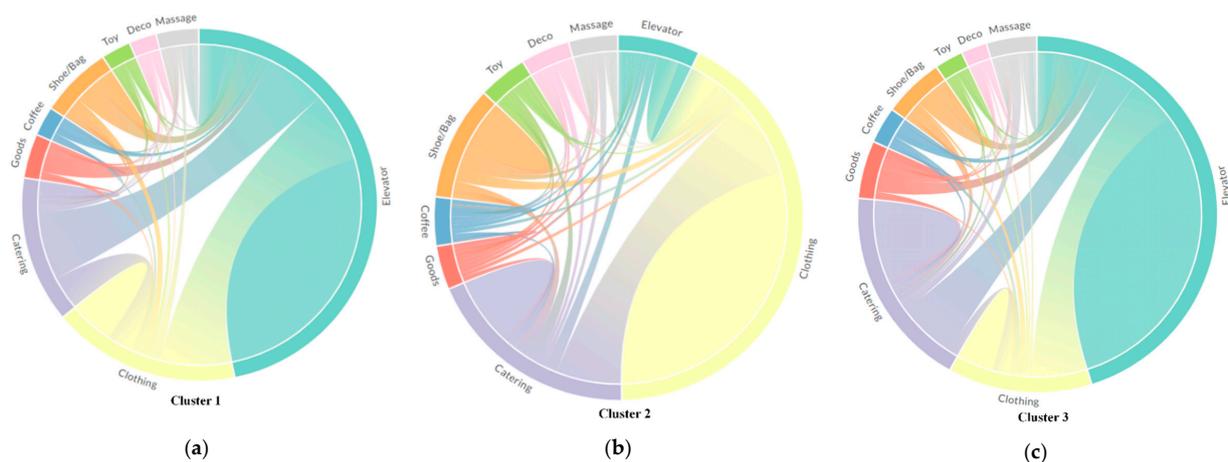


Figure 15. Semantic patterns of indoor semantic trajectory clusters: (a) cluster 1; (b) cluster 2; and (c) cluster 3.

5. Conclusions and Future Work

The indoor trajectory similarity measure is important to indoor trajectory analysis. However, most existing trajectory similarities do not account for indoor movement constraints imposed by the indoor space and the characteristics of indoor positioning sensors. Furthermore, indoor semantic information is often neglected. In this paper, we propose a new indoor trajectory similarity measure called the ISTSM, which considers the constraints

and semantic information of the indoor space simultaneously. The ISTSM extends the classic edit distance. When computing the substitution cost of the edit distance, the ISTSM integrates accurate indoor walking distances with an indoor navigation graph and the semantic information of stops. The ISTSM is more accurate and reasonable because it fuses accurate indoor walking distances between trajectory stops. Experiments revealed that the ISTSM is more accurate and reasonable than the state of the art, and it can unveil customer movement patterns for customers with trajectory clustering analysis. Other trajectory similarities usually need two weights, namely, the spatial weight and the semantic weight, whereas the ISTSM has no weights and is thus convenient for use.

One limitation of our approach is that the construction of the indoor navigation graph model is a non-trivial and time-consuming task. Additionally, the ISTSM does not consider indoor positioning error. We plan to integrate the ISTSM with various indoor data-cleansing approaches, and assess it in more challenging scenarios where there are noises and more ambiguities in the readings of indoor positioning devices. We also plan to evaluate ISTSM with user survey data for other indoor scenarios, such as railway stations and airports, in future work.

Author Contributions: Conceptualization, Jin Zhu; methodology, Jin Zhu; software, Jin Zhu and Dayu Cheng; validation, Weiwei Zhang and Ci Song; formal analysis, Tao Pei; investigation, Jie Chen; resources, Tao Pei; data curation, Tao Pei; writing—original draft preparation, Jin Zhu; writing—review and editing, Weiwei Zhang, Ci Song, Jie Chen, Tao Pei; visualization, Jin Zhu and Dayu Cheng; supervision, Tao Pei; project administration, Jin Zhu and Tao Pei; funding acquisition, Jin Zhu and Tao Pei. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China (No. 41525004, 42071436 and 41701477), Grant of State Key Laboratory of Resources and Environmental Information System (No.201816), and Research Foundation for Talent Introduction of Suzhou University of Science and Technology (No. 331511203).

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from RTMAP Science and Technology Ltd. (<http://www.rtm.com>).

Acknowledgments: The authors thank RTMAP Science and Technology Ltd. for providing the shopping mall dataset.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Kim, J.; Hwangbo, H.; Kim, S.J.; Kim, S. Location-Based Tracking Data and Customer Movement Pattern Analysis Using for Sustainable Fashion Business. *Sustainability* **2019**, *11*, 6209. [[CrossRef](#)]
2. Jin, P.; Cui, T.; Wang, Q.; Jensen, C.S. Effective Similarity Search on Indoor Moving-Object Trajectories. In Proceedings of the International Conference on Database Systems for Advanced Applications, Dallas, TX, USA, 16–19 April 2016; Springer: Cham, Switzerland, 2016; pp. 181–197.
3. 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](#)]
4. Yoshimura, Y.; Sobolevsky, S.; Ratti, C.; Girardin, F.; Carrascal, J.P.; Blat, J.; Sinatra, R. An Analysis of Visitors' Behavior in the Louvre Museum: A Study Using Bluetooth Data. *Environ. Plan. B Plan. Des.* **2014**, *41*, 1113–1131. [[CrossRef](#)]
5. Parent, C.; Spaccapietra, S.; Renso, C.; Andrienko, G.; Andrienko, N.; Bogorny, V.; Damiani, M.L.; Gkoulalas-Divanis, A.; Macedo, J.A.; Pelekis, N.; et al. Semantic trajectories modeling and analysis. *ACM Comput. Surv.* **2013**, *45*, 1–32. [[CrossRef](#)]
6. Ying, J.; Lu, E.; Lee, W.; Weng, T.; Tseng, V. Mining user similarity from semantic trajectories. In Proceedings of the Workshop on Location-Based Social Networks, San Jose, CA, USA, 2 November 2010; ACM: New York, NY, USA, 2010; pp. 19–26.
7. Furtado, A.S.; Kopanaki, D.; Alvares, L.O.; Bogorny, V. Multidimensional Similarity Measuring for Semantic Trajectories. *Trans. GIS* **2016**, *20*, 280–298. [[CrossRef](#)]
8. Lehmann, A.L.; Alvares, L.O.; Bogorny, V. SMSM: A similarity measure for trajectory stops and moves. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 1847–1872. [[CrossRef](#)]
9. Petry, L.M.; Ferrero, C.A.; Alvares, L.O.; Renso, C.; Bogorny, V. Towards semantic-aware multiple-aspect trajectory similarity measuring. *Trans. GIS* **2019**, *23*, 960–975. [[CrossRef](#)]
10. Liu, H.; Darabi, H.; Banerjee, P.; Liu, J. Survey of Wireless Indoor Positioning Techniques and Systems. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2007**, *37*, 1067–1080. [[CrossRef](#)]

11. Laube, P. Computational Movement Analysis. *Geogr. Inf. Sci. Technol. Body Knowl.* **2017**, *2017*, 12–13. [[CrossRef](#)]
12. Ranacher, P.; Tzavella, K. How to compare movement? A review of physical movement similarity measures in geographic information science and beyond. *Cartogr. Geogr. Inf. Sci.* **2014**, *41*, 286–307. [[CrossRef](#)]
13. Vlachos, M.; Gunopulos, D.; Das, G. Rotation invariant distance measures for trajectories. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '04), Seattle, DC, USA, 22–25 August 2004; ACM: New York, NY, USA, 2004; pp. 707–712.
14. Vlachos, M.; Kollios, G.; Gunopulos, D. Discovering similar multidimensional trajectories. In Proceedings of the 18th International Conference on Data Engineering (ICDE '02), San Jose, CA, USA, 26 February–1 March 2002; pp. 673–684. [[CrossRef](#)]
15. Chen, L.; Özsu, M.; Oria, V. Robust and fast similarity search for moving object trajectories. In Proceedings of the 2005 ACM SIGMOD international conference on Management of data (SIGMOD '05), Baltimore, MD, USA, 14–16 June 2005; ACM: New York, NY, USA, 2005; pp. 491–502.
16. Zhao, H.; Winter, S. A Time-Aware Routing Map for Indoor Evacuation. *Sensors* **2016**, *16*, 112. [[CrossRef](#)]
17. Xie, D.; Li, F.; Phillips, J.M. Distributed trajectory similarity search. *Proc. VLDB Endow.* **2017**, *10*, 1478–1489. [[CrossRef](#)]
18. Kang, H.; Kim, J.; Li, K. Similarity measures for trajectory of moving objects in cellular space. In Proceedings of the 2009 ACM symposium on Applied Computing, Honolulu, HI, USA, 8–12 March 2009; ACM: New York, NY, USA, 2009; pp. 1325–1330.
19. Wang, Y.; Yu, G.; Gu, Y.; Yue, D.; Zhang, T. Efficient similarity query in RFID trajectory databases. In *International Conference on Web-Age Information Management*; Springer: Cham, Switzerland, 2010; pp. 620–631.
20. Wan, Y.; Zhou, C.; Pei, T. Semantic-Geographic Trajectory Pattern Mining Based on a New Similarity Measurement. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 212. [[CrossRef](#)]
21. Baba, A.I.; Lu, H.; Pedersen, T.B.; Xie, X. Handling False Negatives in Indoor RFID Data. In Proceedings of the 2014 IEEE 15th International Conference on Mobile Data Management, Brisbane, Australia, 14–18 July 2014; pp. 117–126.
22. Baba, A.I.; Jaeger, M.; Lu, H.; Pedersen, T.B.; Ku, W.; Xie, X. Learning-Based Cleansing for Indoor RFID Data. In Proceedings of the 2016 International Conference on Management of Data, San Francisco, CA, USA, 26 June–1 July 2016; Association for Computing Machinery: San Francisco, CA, USA, 2016; pp. 925–936.
23. 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, Maui, HI, USA, 29 October–2 November 2012; ACM: New York, NY, USA, 2012; pp. 862–871.
24. Fazzinga, B.; Flesca, S.; Furfaro, F.; Parisi, F. Exploiting Integrity Constraints for Cleaning Trajectories of RFID-Monitored Objects. *ACM Trans. Database Syst.* **2016**, *41*, 1–52. [[CrossRef](#)]
25. 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](#)]
26. Choset, H.; Burdick, J. Sensor-Based Exploration: The Hierarchical Generalized Voronoi Graph. *Int. J. Robot. Res.* **2000**, *19*, 96–125. [[CrossRef](#)]
27. Xu, M.; Wei, S.; Zlatanova, S. An indoor navigation approach considering obstacles and space subdivision of 2D plan. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, 339–346. [[CrossRef](#)]
28. Giblin, P.; De Berg, M.; Van Kreveld, M.; Overmars, M.; Schwarzkopf, O. Computational Geometry: Algorithms and Applications. *Math. Gaz.* **2001**, *85*, 175. [[CrossRef](#)]
29. Lee, J. A Spatial Access-Oriented Implementation of a 3-D GIS Topological Data Model for Urban Entities. *GeoInformatica* **2004**, *8*, 237–264. [[CrossRef](#)]
30. Li, X.; Claramunt, C.; Ray, C. A grid graph-based model for the analysis of 2D indoor spaces. *Comput. Environ. Urban Syst.* **2010**, *34*, 532–540. [[CrossRef](#)]
31. Hahmann, S.; Miksch, J.; Resch, B.; Lauer, J.; Zipf, A. Routing through open spaces—A performance comparison of algorithms. *Geo-Spat. Inf. Sci.* **2017**, *21*, 247–256. [[CrossRef](#)]
32. Spaccapietra, S.; Parent, C.; Damiani, M.L.; De Macedo, J.A.; Porto, F.; Vangenot, C. A conceptual view on trajectories. *Data Knowl. Eng.* **2008**, *65*, 126–146. [[CrossRef](#)]
33. Li, Q.; Zheng, Y.; Xie, X.; Chen, Y.; Liu, W.; Ma, W. Mining user similarity based on location history. In Proceedings of the 16th ACM Sigspatial International Conference on Advances in Geographic Information Systems, Irvine, CA, USA, 5–7 November 2008. [[CrossRef](#)]
34. Dijkstra, E.W. A note on two problems in connexion with graphs. *Numer. Math.* **1959**, *1*, 269–271. [[CrossRef](#)]
35. Levenshtein, V.I. Binary codes capable of correcting deletions, insertions, and reversals. *Dokl. Akad. Nauk SSSR* **1966**, *163*, 845–848.
36. Yujian, L.; Bo, L. A Normalized Levenshtein Distance Metric. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 1091–1095. [[CrossRef](#)]
37. Wang, H.; Su, H.; Zheng, K.; Sadiq, S.; Zhou, X. An Effectiveness Study on Trajectory Similarity Measures. In Proceedings of the Twenty-Fourth Australasian Database Conference (ADC 2013), Adelaide, Australia, 29 January–1 February 2013; pp. 13–22.
38. Satopaa, V.; Albrecht, J.; Irwin, D.; Raghavan, B. Finding a “Kneedle” in a Haystack: Detecting Knee Points in System Behavior. In Proceedings of the 31st International Conference on Distributed Computing Systems Workshops, Minneapolis, MN, USA, 20–24 June 2011; pp. 166–171. [[CrossRef](#)]
39. Manning, C.D.; Raghavan, P.; Schütze, H. *Introduction to Information Retrieval*; Cambridge University Press: Cambridge, UK, 2008.