International Journal of
Geo-Information

Article

Semantic-Geographic Trajectory Pattern Mining Based on a New Similarity Measurement

You Wan [1], Chenghu Zhou [2] and Tao Pei [2,3,4,*]

[1] School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China; wanyou9@gmail.com
[2] State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, CAS, Beijing 100101, China; zhouch@lreis.ac.cn
[3] University of Chinese Academy of Sciences, Beijing 100049, China
[4] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
* Correspondence: peit@lreis.ac.cn; Tel.: +86-10-6488-8960

Received: 4 May 2017; Accepted: 5 July 2017; Published: 14 July 2017

Abstract: Trajectory pattern mining is becoming increasingly popular because of the development of ubiquitous computing technology. Trajectory data contain abundant semantic and geographic information that reflects people's movement patterns, i.e., who is performing a certain type of activity when and where. However, the variety and complexity of people's movement activity and the large size of trajectory datasets make it difficult to mine valuable trajectory patterns. Moreover, most existing trajectory similarity measurements only consider a portion of the information contained in trajectory data. The patterns obtained cannot be interpreted well in terms of both semantic meaning and geographic distributions. As a result, these patterns cannot be used accurately for recommendation systems or other applications. This paper introduces a novel concept of the semantic-geographic pattern that considers both semantic and geographic meaning simultaneously. A flexible density-based clustering algorithm with a new trajectory similarity measurement called semantic intensity is used to mine these semantic-geographic patterns. Comparative experiments on check-in data from the Sina Weibo service demonstrate that semantic intensity can effectively measure both semantic and geographic similarities among trajectories. The resulting patterns are more accurate and easy to interpret.

Keywords: trajectory pattern; semantic similarity; geographic similarity; pattern mining; clustering

1. Introduction and Motivations

Owing to advanced positioning technologies, a large amount of movement data can now be conveniently collected from daily activities. Some activities share the same movement pattern, which reflects similar lifestyles, habits, or behaviors. The study of movement data can reveal individual movement patterns, facilitate the understanding of the characteristics of human dynamics, and thus support recommendation, activity prediction, urban planning, and traffic monitoring alike. Therefore, movement pattern mining has become a hot topic when considering human movement activities [1,2].

Trajectories can be derived from movement data sampled from daily activities. They typically consist of a sequence of spatiotemporal points represented as (latitude, longitude) tagged with timestamps. Trajectory data can be divided into three main types based on sampling mode: time-frequency sampling data (e.g., animal migration, hurricane data), location-based sampling data (e.g., population migration data) and event-triggered sampling data (e.g., mobile phone call, check-in data). Whereas previous studies have mainly focused on processing the geometric and/or temporal properties of trajectory data, recent studies have gone one step further. They enrich a movement track with more semantic, application-oriented information by adding geographical context. Figure 1 shows an example in which trajectories 1 and 2 are geometrically similar according to their shapes and distance. In addition, trajectories 1 and 3 have the same sequence of actions and share the same semantic pattern (i.e., school, park, and restaurant). In fact, these are two different trajectory similarity measurements that can generate two types of movement patterns: a geographic pattern and a semantic pattern.

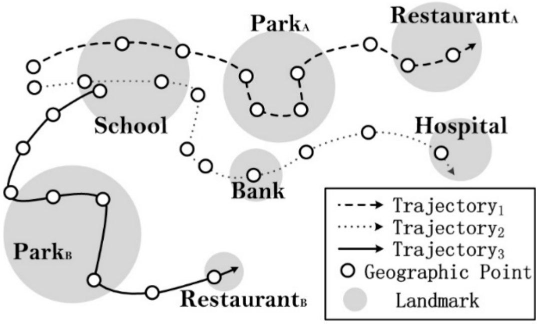


Figure 1. Example of geographic and semantic trajectory patterns.

However, these two types of patterns are not sufficiently accurate to incorporate into recommendation systems or other applications. In Figure 1, two trajectories in the geographic pattern do not share similar semantic meanings because they pertain to different types of activities in their second and third steps. On the other hand, two trajectories in the semantic pattern are not geometrically similar; they pertain to the same type of activity in different places. Therefore, users who define these patterns are not actually similar. More accurate and valuable trajectory patterns that contain both semantic and geographic information are required. People who perform similar activities in nearby places share the same movement activity pattern, and they may have more common interests. Here, "similar activities" are determined by semantic similarity, and "nearby places" are determined by geographic similarity. It is thus necessary to combine these two similarities. However, it is not a trivial task to calculate geographic and semantic similarity simultaneously. Challenges are posed by the complexity of people's movement activity, the large size of trajectory datasets, different sampling strategies, and innate differences between these two types of dimensions. To the best of our knowledge, an appropriate method is still lacking.

To obtain more valuable trajectory patterns that share the same movement activities according to both sematic meaning and geographic distribution, we first define a novel trajectory pattern called a semantic-geographic pattern. Then, a flexible density-based clustering algorithm is used to mine this new type of pattern. The clustering algorithm involves a new similarity measurement called semantic intensity to calculate similarities between trajectories by combining the geographic and semantic features. We prove the effectiveness of our method through comparative trajectory pattern mining experiments on 4340 users over 14,729 check-in traces using four different similarity measurements.

The remainder of this paper is organized as follows. Section 2 provides a review of existing trajectory similarity measurements. Section 3 outlines the definitions of the proposed semantic-geographic pattern. Section 4 describes the framework and major steps of the new pattern mining method. We present the experimental results obtained for a real social media dataset in Section 5 and discuss our conclusions and future work in Section 6.

2. Related Work

There are three major steps to convert raw trajectory data to interesting movement patterns: which are trajectory representation, similarity calculation, and pattern mining [3,4]. Among them, the similarity calculation can be the most important step [5]. Different similarity measurements will generate quite different trajectory patterns. Early research on similarity analysis was based on spatial and/or temporal information in trajectories. For the spatial dimension, classical methods like average distance [6], Hausdorff distance [7], Fréchet distance [8], and Minkowski distance can be used for the similarity calculation. For the temporal dimension, several sequence matching approaches can be directly applied to trajectory similarity analysis, such as Edit Distance [9], Longest Common Subsequence [10], and Dynamic Time Warping [11]. More recently, a number of extensions were developed to deal with trajectory data more effectively. These extension methods can be divided into two categories: spatiotemporal-based and semantic-based.

(1) Spatiotemporal-based methods focus on both spatial and temporal features of trajectories. It's not an easy task to measure two moving objects' spatiotemporal similarity, because their activities are updating asynchronously in both spatial and temporal dimensions. Giannotti et al. [12] defined the T-pattern in a collection of GPS trajectories. A T-pattern is a region-of-interest (ROI) sequence with temporal annotations, where each ROI is a rectangle corresponding to a trajectory density greater than a threshold. Lu et al. [13] proposed a transaction similarity measurement named LBS-Alignment to calculate the similarity of mobile users. By using the longest common sequence, the ratio of the common parts of the sequential motion patterns was taken as the similarity. Etienne et al. [14] defined an ordered sequence of spatial zones (called a "zone graph") to extract and filter trajectories following a similar itinerary, and then qualified spatiotemporal patterns such as the main routes and spatiotemporal channels through statistical computations. Buchin and Purves [8] used space-time prisms to model trajectories, and then computed the similarity of these prisms based on equal time and Fréchet distance. Lv et al. [15] addressed the problem of mining mobile users' long-term activity similarity. After transforming users' GPS trajectories into reference places using a three-layered hierarchical clustering algorithm, a bottom-up agglomerative clustering algorithm based on cosine coefficient similarity was used to group users' one-day activities. Each cluster contained a set of users' one-day activities, which represented a routine activity. Then, the similarity of users' routine activities was calculated based on the optimal matching sequence similarity of their reference place sets. Finally, users' similarities were calculated based on all of their routine activities multiplied by the number of times the user follows each routine activity as weights. Dodge et al. [4] introduced a novel technique for spatiotemporal trajectory similarity that relied on trajectory segmentation based on the movement parameters (e.g., speed, acceleration, or direction). Each segmentation was assigned to a movement parameter class, which can transform a trajectory into a sequence of class labels. Then, a modified version of edit distance called normalized weighted edit distance was developed to measure the similarity between different sequences. Yuan and Raubal [16] also extended the traditional edit distance algorithm by incorporating the spatial distribution of cell towers, and then applied a newly developed spatiotemporal edit distance to compare the trajectories extracted from call detailed records and conduct a hierarchical clustering analysis. Etienne et al. [17] defined a new spatiotemporal pattern called Trajectory Box Plot (TBP) to describe trajectories by a median trajectory, a 3D box and a 3D fence. The median trajectory depicts the typical movement of mobile objects. The box and the fences describe the spatial and temporal spreading around the central tendency. Etienne et al. then used visual analysis to highlight the density of trajectory cluster that changes over time. The above methods measure the trajectories' similarity based on the geometric and temporal information without considering the semantics information. As a consequence, they tend to discover spatiotemporal patterns such as sequences of locations, which for some applications may not help the user extract more meaningful information. Moreover, similar spatiotemporal trajectories may not necessarily be semantically similar, because the activities implied by the nearby locations they pass through may be different.

(2) Semantic-based methods aim to obtain more meaning from trajectory information. A semantic trajectory fundamentally consists of a sequence of locations with semantic tags describing the corresponding landmarks [18]. According to the encoding and organization of semantic knowledge, there are mainly three kinds of semantic similarity approaches [19–21]: semantic or taxonomic relations (also called path distance measures) based, information content-based, and feature models (also called

classic models) based. They all can be used for carrying the semantic meaning of trajectories beyond the low-level pure geographic positions. Alvares et al. [18] first identified the stops in the GPS trajectories of mobile users, and mapped these stops to semantic landmarks by using a background map. They then applied a sequential pattern-mining algorithm to extract the frequent place sequences (i.e., the semantic trajectory pattern) to represent the frequent semantic behaviors. Unfortunately, because of spatial discontinuities and the randomness of users' movement activity, such place-level sequential patterns can appear only when the support threshold is very low. Bogorny et al. [22] took both hierarchical geographic and semantic properties into consideration, and provided two different methods (IB-SMoT and CB-SMoT) to automatically integrate trajectory samples and geographic information in a higher abstraction level (i.e., stops and moves), where the user can define the important parts of trajectories. Ying et al. [5] proposed a novel similarity measurement called Semantic Trajectory Pattern Similarity to evaluate the similarity between two trajectories. First, a frequent sequence pattern mining algorithm was used to get users' Maximal Semantic Trajectory Patterns (MSTPs). Then, the similarity between two MSTPs was calculated by the MSTP-Similarity measurement, which was based on the longest common sequence. Finally, a sequential pattern in the form of a sequence of semantic labels (e.g., school to park) was obtained. Xiao et al. [23] proposed a method to estimate the similarity between user trajectories. Their method first modeled a user's GPS trajectories with a semantic location history (SLH) based on the semantic location hierarchy and users' stay points. Then, the similarities between different users' SLHs were calculated by using a maximal travel match algorithm that summarized the weighted similarity of semantic location sequences detected at each layer of the hierarchy. Although these approaches overcome the geographic constraint on user similarity measurements, the patterns obtained lack detailed geographic distribution information, and are difficult to display using geo-visualization tools. Moreover, the patterns are usually represented with a high-level abstraction of semantic landmarks (e.g., <school, park>, <school, hospital, restaurant>) [5,24], which may be difficult to interpret.

To obtain more valuable trajectory patterns that share the same movement activities according to both semantic meaning and geographic distribution, it is necessary to combine these two measurements and create a unified framework to calculate the similarity between trajectories. Recently, Ying et al. [25] provided a new definition of geographic-temporal-semantic (GTS) pattern tree to model users' historical trajectories. They measured GTS similarity between two users' trajectories by the sum of three weighted dimensional similarity scores to predict a mobile user's next location. However, the efficiency for GTS similarity to mining interesting trajectory patterns is uncertain. Buchin et al. [26] defined a context-aware similarity measure to integrate the various surrounding contexts of trajectories. The final similarity score was obtained by summing up all the weighted context distances. However, suitable prior knowledge is needed to set proper weights for each factor, and the calculation of the trajectory's context distance requires refined and classified land cover and land use data, which is difficult to fetch.

3. Problem Statement

This section presents and defines a new trajectory pattern called a semantic-geographic pattern to describe people's movement activity. Several related preliminaries are provided to help explain this pattern.

3.1. Semantic-Geographic Trajectory Pattern

Generally, people who conduct similar activities in nearby places may share common interests and the same movement activity pattern. This type of pattern involves both semantic and geographic similarity; therefore, we refer to it as a semantic-geographic pattern. In contrast to a purely geographic pattern, people with similar semantic-geographic patterns must conduct semantically similar activities. In contrast to a purely semantic pattern, people must be geographically close within a certain geographic scale when they are conducting these similar activities.

3.2. Preliminaries

Several strategies have been proposed to transfer the original GPS trajectory into a semantic trajectory [23,25]. This paper does not focus on this preprocessing step. Instead, a special type of event-triggered trajectory called "check-in" data is used. Check-in data always contains geographic locations, and the semantic meaning in most check-in records can be easily obtained from a POI (Point Of Interest) database. We provide definitions of terms relevant to the use of this type of data. In addition, Figure 2 describes three users' check-in traces to provide examples of all definitions.

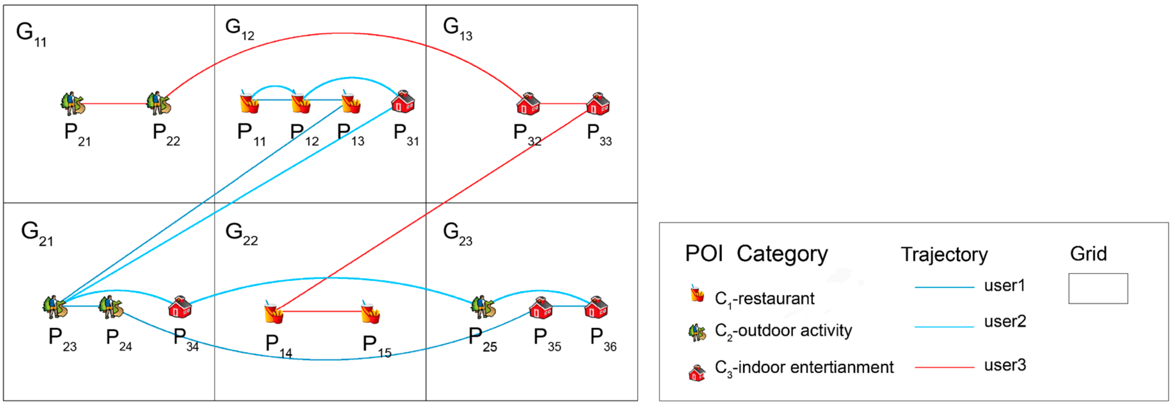


Figure 2. A sample map of three users' check-in activities.

Definition 1. ***Check-in trace.*** *A user's check-in trace (CTrace) is a set of POIs with the check-in time of each POI. The POIs contain both geographic locations and semantic categories, which are usually based on land use and land cover classification standards or certain domain knowledge.*

Example 1. *The three users' check-in traces in Figure 2 are as follows (where P_{ij} is a serial of POIs for one POI category C_i, and t_i is the time sequence in the corresponding POI serial):*

CTrace(user$_1$) = {(P_{11}, t_1), (P_{12}, t_2), (P_{13}, t_3), (P_{23}, t_4), (P_{24}, t_5), (P_{35}, t_6), (P_{36}, t_7)}.
CTrace(user$_2$) = {(P_{11}, t_1), (P_{12}, t_2), (P_{31}, t_3), (P_{23}, t_4), (P_{34}, t_5), (P_{25}, t_6), (P_{36}, t_7)};
CTrace(user$_3$) = {(P_{21}, t_1), (P_{22}, t_2), (P_{32}, t_3), (P_{33}, t_4), (P_{14}, t_5), (P_{15}, t_6)}.

Definition 2. ***Geographic pattern.*** *Users who have small distance values among their check-in traces are grouped together, and their common location pairs and check-in times are considered a geographic pattern (GPattern). Note that the check-in times for a user at a certain location are not always the same; therefore, a time duration is set for each check-in location.*

Example 2. *One geographic pattern between user$_1$ and user$_2$ in Figure 2 is as follows (where G_{ij} is the id of a grid, and POIs are considered nearby and grouped together when they locate in the same grid): GPattern(user$_1$,user$_2$) = {(G_{12}, t_1~t_3) → (G_{21}, t_4~t_5), (G_{21}, t_4~t_5) → (G_{23}, t_6~t_7)}.*

Definition 3. ***Semantic trace.*** *A user's semantic trace (STrace) is a set of POI category pairs with the corresponding check-in times and frequencies.*

Example 3. *STrace(user$_1$) = {((C_1, t_1~t_3) → (C_2, t_4~t_5), count$_1$), ((C_2, t_4~t_5) → (C_3, t_6~t_7), count$_2$)}, where the counts represent the check-in frequency of the corresponding semantic category pairs.*

Definition 4. ***Semantic pattern.*** *Users with high semantic similarities in their semantic traces are grouped together. Their common POI category pairs and check-in times are considered a semantic pattern (SPattern).*

Example 4. *One semantic pattern between $user_1$ and $user_2$ in Figure 2 is as follows: SPattern($user_1$, $user_2$) = {(C_1, t_1~t_2) → (C_2, t_4), (C_2, t_4) → (C_3, t_7)}.*

Definition 5. ***Semantic-geographic trace.*** *A user's semantic-geographic trace (SGTrace) consists of several sets of timestamped POI locations that are grouped by the semantic category of his/her check-in place.*

Example 5. *SGTrace($user_1$) = {(C_1, G_{12}, t_1~t_3), (C_2, G_{21}, t_4~t_5), (C_3, G_{23}, t_6~t_7)}.*

Definition 6. ***Semantic-geographic pattern.*** *Users who have high semantic similarity in their semantic trace and small distances in their semantic-geographic trace are grouped together. Their user set and common SGTraces comprise a semantic-geographic pattern (SGPattern).*

Example 6. *One semantic-geographic pattern between $user_1$ and $user_2$ in Figure 2 is as follows, and its geographic distribution is the same as the geographic pattern: SGPatten($user_1$,$user_2$) = {(C_1, G_{12}, t_1~t_2) → (C_2, G_{21}, t_4), (C_2, G_{21}, t_4) → (C_3, G_{23}, t_7)}.*

4. Semantic-Geographic Pattern Mining Method

From the data mining perspective, clustering is one of the most powerful methods for trajectory pattern mining [27–29]. In this study, the semantic-geographic pattern mining process consists of two parts and four main steps, as shown in Figure 3. First, in the data preprocessing step, users' original check-in traces are transformed into semantic-geographic traces and semantic traces according to the definitions presented in Section 3. Then, in the pattern mining step, geographic similarity and semantic similarity are calculated for each type of trace. The semantic-geographic pattern combined semantic intensity similarity is then calculated to obtain the final similarities. Finally, a density-based clustering algorithm is used to detect the semantic-geographic combined movement patterns.

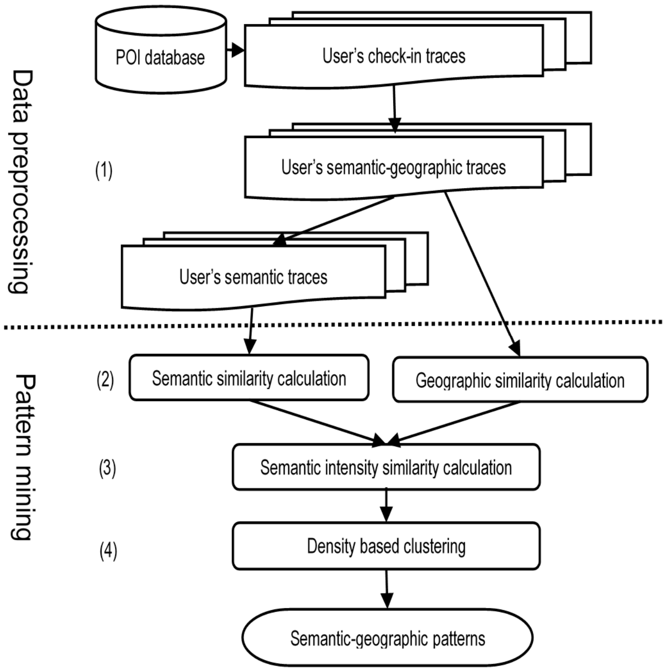

Data preprocessing
POI database
User's check-in traces
(1)
User's semantic-geographic traces
User's semantic traces
Pattern mining
(2)
Semantic similarity calculation
Geographic similarity calculation
(3)
Semantic intensity similarity calculation
(4)
Density based clustering
Semantic-geographic patterns

Figure 3. Steps of semantic-geographic pattern mining method.

4.1. Semantic-Geographic Combined Similarity Measurement

Similarity measurement is the key step of semantic-geographic pattern mining. In this context, a new concept called *semantic intensity* is introduced to calculate trajectory similarities based on both semantic and geographic features. Semantic intensity is a combination of semantic similarity and geographic similarity. It can be considered the average semantic similarity in a geographic unit distance. The calculation of this new semantic intensity measurement is described in the following subsection.

4.1.1. Semantic Similarity Calculation

Because trajectory data are transformed into semantic vectors according to Definition 3, the feature model-based method is suitable for measuring their similarity. Cosine similarity is a feature model-based method that is widely applied in the field of information retrieval [30] and text mining [31], and for this reason is chosen to calculate the semantic similarity between two users' semantic traces.

The semantic similarity of the users' semantic traces $STrace_1$ and $STrace_2$ is as follows:

$$\text{Semantic similarity}(STrace_1, STrace_2) = \frac{\sum_{i=1}^{n} W_{1i} * W_{2i}}{\sqrt{\sum_{i=1}^{n} (W_{1i})^2} * \sqrt{\sum_{i=1}^{n} (W_{2i})^2}} \tag{1}$$

where n is the total number of POI category pairs (e.g., school → park) and W_{ji} is the corresponding weight for each type of POI category pair. In addition, the start time and end time in each POI category pair must overlap in order to calculate the two traces' inner product.

Different weighting schemes might be used in this context; one that is effective and widely used is term frequency and inverse document frequency (TF-IDF):

$$W_{ji} = tf_{i,j} \times log_2(\frac{N}{df_i}) \tag{2}$$

where $tf_{i,j}$ is the number of times term i appears in a document dj, and df_i is the number of documents in which the certain term appears. In this paper, a term corresponds to one POI category pair, and a document corresponds to a semantic trace. The TF factor indicates the importance of a POI category pair in the trace. The IDF factor is a global statistic that measures how widely a POI category pair is distributed over a collection. POI category pairs that appear often in a trace and do not appear in many traces therefore carry significant weight. Unlike in the weighting strategies of Buchin, et al. [26], the TF-IDF weight is directly obtained from original data. This weight carries more accurate semantic meaning, and does not require any prior knowledge. For detailed explanations of cosine similarity and TF-IDF weight please refer to [30–32].

4.1.2. Geographic Similarity Calculation

Calculating the geographic similarity calculation between two semantic-geographic traces is not a trivial task. Because trajectory data are collected over a long period, the places visited may not be the same for one person at a certain time. Thus, unlike in traditional trajectory distance measurements (such as DTW (Dynamic Time Warping), Fréchet, and edit distance), two polyline sets must be calculated at each time interval.

The Hausdorff distance is the maximum distance of a set to the nearest point in another set [33]. Because the computation of the Hausdorff distance can be extended to polygons and line segment sets, curves and surfaces, curve sets, etc., it is widely used for pattern matching and recognition [34,35]. The Hausdorff distance from point set A to point set B is formally defined as follows:

$$h(A,\ B) = \max_{a \in A}\{\min_{b \in B}\{dist(a,\ b)\}\} \tag{3}$$

where a and b are points in sets A and B, respectively, and $dist(a, b)$ is usually the Euclidean distance between these two points.

This Hausdorff distance is not symmetric, which means that most of the time *h(A, B)* is not equal to $h(B, A)$. Therefore, the Hausdorff distance between A and B can be more generally defined as follows:

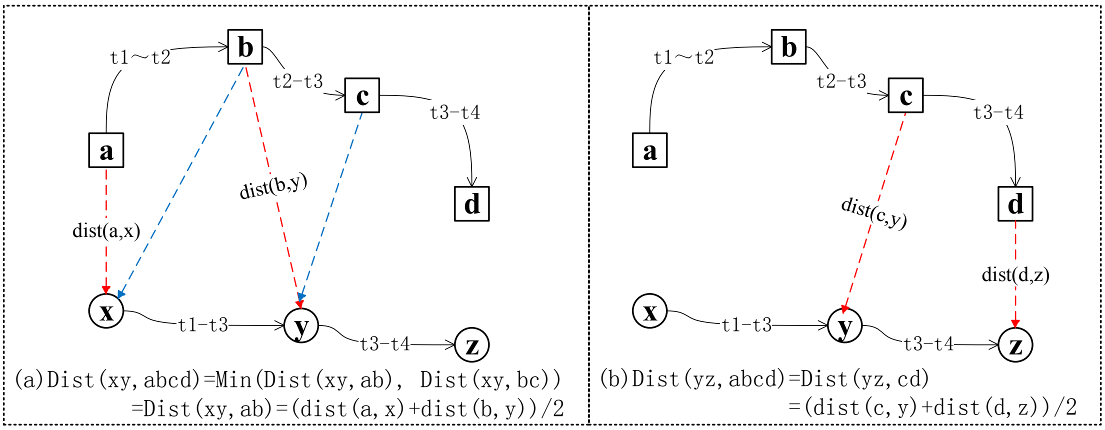
$$H(A,\ B) = \max\{h(A,B),\ h(B,\ A)\} \tag{4}$$

Two users' geographic similarity is calculated by the root mean square value of the Hausdorff distance at each POI category pair that is shared in their corresponding semantic- geographic traces. Therefore, only semantically related POI category pairs are used to calculate the geographic similarity, and the semantically unrelated POI category pairs are omitted. As shown in Figure 4, given Trace$_1$ and Trace$_2$ as one type of a POI category pair's polylines over the same period (t$_1$~t$_4$) for user$_1$ and user$_2$, respectively, the Hausdorff distance is calculated as follows:

$$HausDistAtPoiCate_1(Trace_1, Trace_2) = \mathrm{Max}(H(Trace_1, Trace_2), H(Trace_2, Trace_1)) \tag{5}$$

where:

$$\begin{gathered} H(Trace_1, Trace_2) = \mathrm{Max}(Dist(xy,\ abcd),\ Dist(yz,\ abcd)), \\ H(Trace_2, Trace_1) = \mathrm{Max}(Dist(ab, xyz),\ Dist(bc,\ xyz),\ Dist(cd, xyz)). \end{gathered} \tag{6}$$



(a) Dist (xy, abcd) =Min (Dist (xy, ab), Dist (xy, bc)) =Dist (xy, ab) = (dist (a, x) +dist (b, y)) /2

(b) Dist (yz, abcd) =Dist (yz, cd) = (dist (c, y) +dist (d, z)) /2

Trace1={(x,t1), (y,t3), (z,t4)}; Trace2={(a,t1), (b,t2), (c,t3), (d,t4)}

Figure 4. The Hausdorff distance from trace$_1$ to trace$_2$.

As shown in Figure 4a, the distance between one segment pair from two traces (e.g., *Dist(yz, cd)*) is the arithmetic mean of the corresponding distances of their start points and end points. Moreover, as shown in Figure 4a, the distance from one trace's segment to the other trace (e.g., *Dist(xy, abcd)*) is the minimum distance between xy and trace2's segments (supposing *Dist(xy, ab)* < *Dist(xy, bc)*, then *Dist(xy, ab)* remains). Note that the duration of the segments in trace$_2$ must overlap the duration of xy's segments.

As the Hausdorff distance increases, geographic similarity decreases. The final geographic similarity between two check-in traces is:

$$\begin{aligned} &\text{Geographic similarity}(CTrace_1,\ CTrace_2) \\ &= 1 - \left(\frac{\sqrt{\sum_{ciCategory}(HausDistAtPoiCate_i)^2/n}}{r}\right) \end{aligned} \tag{7}$$

where n is the number of related POI category pairs, and r is a distance threshold value such that a Hausdorff distance greater than r results in a geographic similarity equal to the minimum value of 0. The upper limit of geographic similarity is 1, which occurs when the Hausdorff distance between two traces is 0.

4.1.3. Semantic Intensity Similarity Calculation

Two users' semantic intensity is the semantic similarity within a geographical unit distance. Semantic intensity is calculated as a mixture of both semantic similarity and geographic similarity:

$$\text{Semantic intensity}(\text{user}_1,\ \text{user}_2) = \frac{\text{Semantic similarity}(STrace_1,\ STrace_2)}{2 - \text{Geographic similarity}(GTrace_1,\ GTrace_2)} \quad (8)$$

The divisor of Formula (8) consists of the geographic distance plus 1 (according to Formula (7)), which avoids obtaining an infinite semantic intensity when the geographic similarity equals 0. Additionally, this modification forces the semantic intensity to vary between 0 and 1. As semantic similarity and geographic similarity increase, semantic intensity increases as well. Semantic intensity is assigned the maximum value of 1 when semantic similarity is equal to 1 and geographic similarity is equal to 1. Because semantic intensity scales with both semantic similarity and geographic similarity, a suitable similarity threshold is needed to identify trajectories with high semantic intensity similarity. This issue is addressed in the next subsection.

4.2. Density-Based Clustering

After transforming original trajectory data into feature vectors, generic clustering algorithms are then used to group them based on the similarity measurements mentioned above. However, as argued by Salvador and Chan [36], one important problem associated with these clustering algorithms entails determining the number of clusters. Moreover, a large amount of data corresponding to noise or outliers may adversely affect the efficiency and accuracy of these algorithms. Therefore, in this paper, a density-based spatial clustering of applications with noise (DBSCAN) algorithm [37] is used to overcome these drawbacks and effectively mine valuable trajectory patterns.

In DBSCAN, objects with many nearby neighbors are considered a high-density region. DBSCAN can then group these objects together to form a cluster, and make outliers objects that lie alone in low-density regions outliers. DBSCAN requires only two parameters that are insensitive to the order of the objects in a dataset. One is the neighborhood region's radius, epsilon (ε). The other is the minimum number of nearest objects (*MinPts*, or k) required to form a dense region. The performance of DBSCAN depends on the distance/similarity measure, because it defines the neighborhood region for objects and can strongly affect the region of final clusters. To obtain meaningful trajectory patterns by the similarity measures mentioned above, the two parameters (ε and k) are set based on the three following factors:

(1) The geographic distance between two trajectories calculated by Formula (5) should be no greater than 1 km. A large distance is indicative of a non-similar relation, regardless of how semantically similar the trajectories are. Moreover, the neighborhood region's radius of geographic similarity is also set to 1 as explained in Formula (7).

(2) The neighborhood region's radius of semantic intensity similarity between two trajectories should be no less than 0.5. According to Formulas (7) and (8), trajectories that have minimum geographic similarity (a geographic distance equals to 1) are considered neighbors only if they have a maximum semantic similarity equal to 1. Furthermore, trajectories that have a maximum geographic similarity equal to 1 are considered neighbors only if they have a minimum semantic similarity equal to 0.5. Thus, no matter how large the geographic similarity is, a semantic similarity less than 0.5 is not able to generate high-density regions in a semantic intensity similarity measurement.

(3) A user is considered a dense object when the number of neighboring users in its neighborhood region is no less than a given parameter $k \in [2,9]$. A small k value will generate a large number of clusters that have very small cluster sizes. The cluster number will decrease as k increases. However a large k value tends to generate one special cluster that is much larger than other clusters. Therefore, the final value of k has to be determined experimentally.

5. Experimental Analysis

This section validates the accuracy and efficiency of the proposed similarity measurement. We conducted experiments comparing four similarity measurements: semantic similarity (Sem), geographic similarity (Geo), semantic intensity similarity (SG), and GTS similarity (Ying, et al. [25]). We then analyzed the mining results to demonstrate the advantages and disadvantages of each.

5.1. Data Collection and Preprocessing

The experimental data were extracted from the Chinese microblogging site Sina Weibo. We collected over 238,000 users' check-in records in Beijing during the year 2013. The total data comprised three million records, with about 54,000 POIs. The data were preprocessed based on three rules:

(1) The 54,000 POIs were classified into five simple categories (educational institution (EI), hotel or restaurant (HR), indoor entertainment (IE), outdoor activity (OA) transportation facility (TF)), and one compound category (which is social institution or building (SB)). The compound category includes industrial and commercial buildings, banks, hospitals, and government organizations. These six categories cover a user's daily movement activities, comprising working, studying, eating, etc.

(2) One check-in trace record for a user had to be generated in one day, from 0:00~23:59, and contained at least two POIs.

(3) Users who had checked in at the same trace more than twice and in at least two different months were selected as experimental users. Their traces were considered regular movement activities.

After these three criteria were applied, 14,729 check-in traces from 4340 users remained. Sample records for one user are listed in Table 1. XA, YA, XB, YB are the coordinates of POI A and POI B; TA and TB are the check-in duration of POI A and POI B.

Table 1. A sample user's check-in traces.

ID	POI A	POI B	Cate AB	Weight	XA	YA	XB	YB	TA	TB	Count
1	McDonald's	Ceramic base	HR-IE	0.5	977	2565	5358	3871	7–8	9–15	2
1	Ceramic base	UBC Coffee	IE-HR	0.7	5358	3871	943	2621	8–10	19–22	3
1	Ceramic base	Xinzhao Apartments	IE-SB	0.3	5358	3871	908	2443	9–18	19–20	2

5.2. Experiment Design

The aim of the experiment was to compare the accuracy and efficiency of the trajectory pattern mining method based on four different trajectory similarity measurements.

Semantic similarity and semantic intensity similarity were measured using Formulas (1) and (8). The calculation of geographic similarity differed slightly from that described by Formula (7). The Hausdorff distance was calculated for two users' trajectories, without considering the POI categories. The GTS similarity between two user's trajectories was based on the method of Ying et al. [25], and is expressed as follows:

$$\text{GTS similarity}(\text{U},\ \text{V}) = \frac{\sum_{\text{i}=1}^{\text{m}} \sum_{j=1}^{n} GTSSimilarity\left(M_i, M_j\right)}{m \times n} \tag{9}$$

In Formula (9), M_i and M_j are two segments of the trajectories of two users (U and V); m and n are the size of segmentations in each trajectory. When two segments' geographic distance is less than 1 km, their GTS similarity can be obtained as the sum of semantic similarity and temporal similarity. Their semantic similarity is based on the cosine similarity measurement, and their temporal similarity is the proportion of the intersection of their stay time intervals to the union of their stay time intervals at a certain place. If two segments have a geographic distance greater than 1 km, they have no GTS similarity.

The setting of the similarity threshold for pattern mining relies on the three factors described in Section 4.2. The similarity thresholds for geographic and semantic intensity measurements are 1 and 0.5, respectively. The threshold for GTS similarity is also set to 0.5 for comparison with other measurements. Because the POI categories for each user's check-in records are highly similar, the semantic similarity threshold is set to 0.9.

Two factors are used to verify the efficiency of the four measurements. One is the total user number of the five largest clusters (N5). The greater N5 is, the better measurement it becomes. The other factor is the ratio of the largest cluster user number in N5. An excessively large ratio is considered an inefficient measurement. On the other hand, accuracy is determined by analyzing the user set and POI set in each pattern's contents.

5.3. Results Analysis

By using four similarity measurements and running the DBSCAN algorithm with the parameters mentioned above, a large number of clusters are obtained. The clustered user number is considered an important indicator for describing the cluster results. Usually, the measurement and cluster results improve with the number of clustered users. However, as a by-product of the DBSCAN algorithm, many clusters are very small. Ultimately, only the five largest clusters of each similarity measurement are chosen as the most meaningful patterns to be analyzed. In addition, the largest cluster may cover a very large portion of the whole clustered user base as the parameter k increases, which will greatly impair the representation of other clusters. Therefore, the proportion of the largest cluster's user number to that of the five largest clusters is also used to determine the efficiency of each similarity measurement. The two indicators for four similarity measurements' cluster results are shown in Figure 5.

Figure 5 shows that as the parameter k increases, the top cluster user ratio for semantic similarity jumps to a very high level when k is greater than 7. The clustered user number for geographic similarity reaches a maximum when k equals 6. On the other hand, semantic intensity similarity and GTS similarity generate much smaller clusters, and their total values change only slightly when k ranges from 2 to 6. Finally, the clustering parameter k is set to the same value of 6 to make better comparisons among the four similarity measurements. Table 2 shows the parameter settings and cluster results for the four similarity measurements.

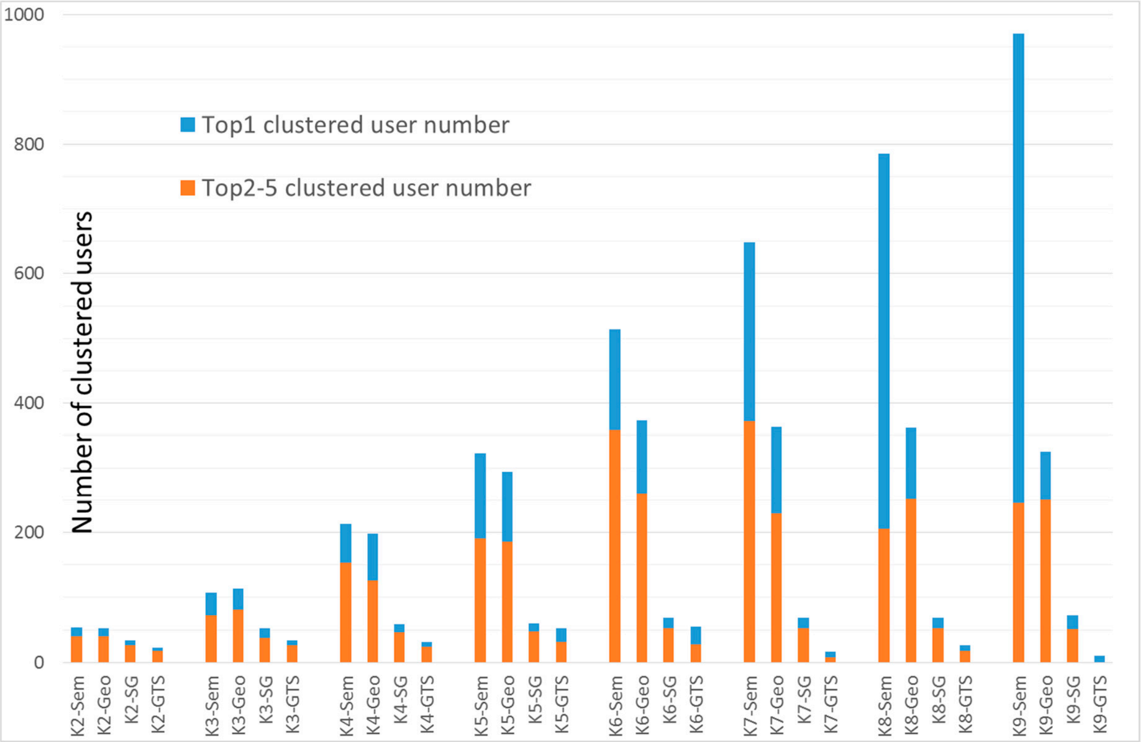


Figure 5. User numbers of the Top-5 clusters and the proportion of the largest cluster for four similarity measurements.

5.3.1. Cluster Result Analysis by User Set

Based on the cluster results presented in Table 2, we perform an overlap analysis of the clustered user sets to demonstrate the differences among the four similarity measurements.

Table 2. DBSCAN parameter settings and mining results for the four similarity measurements.

Measurements	Parameter Setting		Total Clusters			Top 5 Clusters		
	Similarity Threshold	MinPts (k)	Cluster Number	Clustered User	Clustered User Ratio	Top 1 User Size	Top 5 User Size	Top 1 User Ratio
Semantic	0.9	6	2084	3539	81.5%	156	358	30.4%
Geographic	1 km	6	164	1405	32.4%	113	260	30.3%
Semantic intensity	0.5	6	32	242	5.6%	16	69	23.2%
GTS	0.5	6	8	72	1.7%	27	55	49.1%
Semantic	0.9	6	2084	3539	81.5%	156	358	30.4%

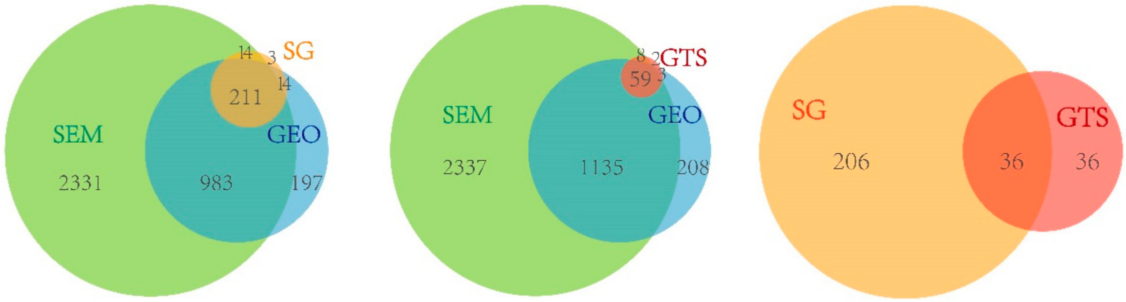
(1) As shown in Table 2 and Figure 6, semantic similarity generated thousands of clusters, most of which have very small user sizes. Furthermore, its clustered users covered 81.5% users of the whole data set, which means that most of the users have high semantic similarities with at least 6 other users. Among those users, 33.7% are also geographically similar.

(2) Only approximately 32.4% of users' activities were conducted within 1 km of other activities. The ratio of semantically similar users to geographically similar users was 85%, which is only a little greater than the ratio of semantically similar users to the whole dataset. These two ratios indicate that there is no correlation between the two types of similar users.

(3) As shown in Figure 6a, the semantic intensity similarity was practically a small subset of the union of the two similarities described above. Over 82.3% users in that union were both semantically and geographically similar. They were not, however, similar to each other when the two similarities were combined. In fact, they were often conducting different types of activities in nearby locations;

thus, they were not truly similar. On the other hand, approximately 12.7% of semantic intensity similar users were identified among non-semantic similar and/or non-geographic similar users. These users were selected because their check-in POI sets were geographically similar for some of the semantic POI categories. In other words, they were not always conducting similar activities in nearby areas, but they did some particular type of activities in nearby areas.

(4) As shown in Figure 6b, GTS similarity exhibited the smallest cluster size. Figure 6c shows that half of the GTS-clustered users overlapped with semantic intensity-clustered users. A detailed comparison between these two similarities is performed in the next subsection.



(**a**) Overlapping of SEM, GEO and SG (**b**) Overlapping of SEM, GEO and GTS (**c**) Overlapping of SG and GTS

Figure 6. Overlapping analysis of the clusters performed using four similarity measurements.

The overlap analysis described above demonstrates that the semantic intensity measurement performs quite differently from the semantic similarity and geographic similarity measurements. Its clustering results contain both semantically and geographically similar users, indicating more valuable patterns. In addition, semantic intensity identified more clustered users than GTS similarity.

5.3.2. Pattern Result Analysis by POI Set

In this section, the POI sets in each cluster are used to analyze the semantic and geographic features of each pattern obtained by using different similarity measurements.

(1) The semantically similar clusters pertain mainly to three types of POIs: outdoor activity (OA), social institution and building (SB), and indoor entertainment (IE). The semantic patterns are as follows:

① SPattern$_1$ (149 users) = {(OA, 9~18) → (SB, 16~22), (SB, 9~18) → (OA, 16~22), (SB, 9~18) → (SB, 16~22)}.

② SPattern$_2$ (156 users) = {(IE, 11~17) → (OA, 18~22), (OA, 11~17) → (IE, 17~22)}.

Figure 7 shows the geographic distributions of two semantic patterns based on their POI traces. Because the semantic similarity measurement does not consider the geographic information in the traces, the two semantic patterns overlap over a large portion of Beijing, and their geographic distributions are difficult to describe and distinguish.

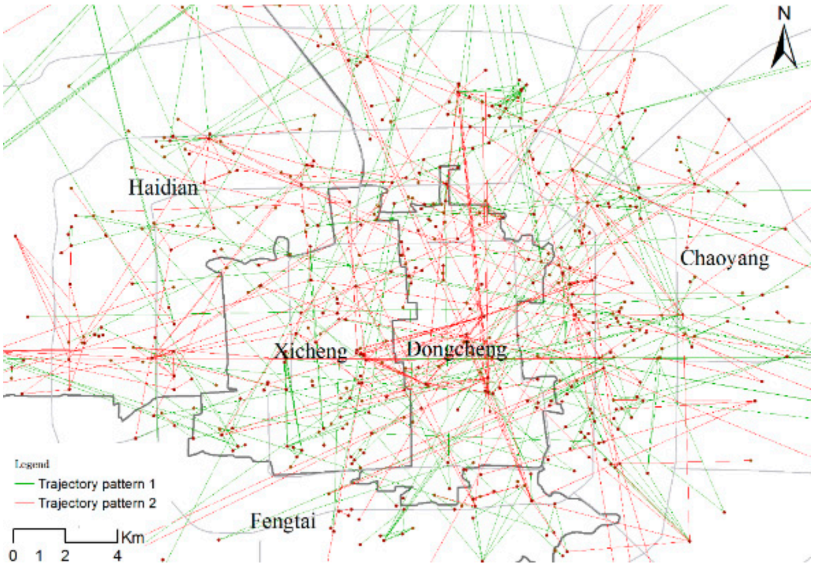


Figure 7. Two largest semantic patterns in user check-in data.

(2) The five largest geographic patterns are illustrated in Figure 8. They are located in different places in Beijing and have different area sizes. However, no detailed semantic rules can be determined from these geographic patterns. It is thus unclear what types of activities are conducted by each pattern's users. In fact, the overlapping analysis in Figure 6 shows that most of the users are not performing similar activities in nearby places.

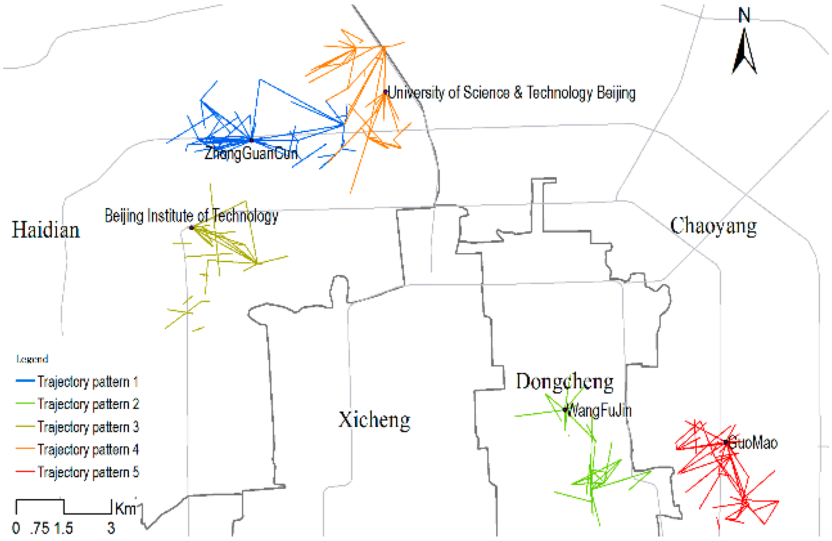


Figure 8. Five largest geographic patterns in user check-in data.

(3) The semantic intensity similarity measurement generated 32 small semantic-geographic patterns. The geographic distributions of the top 5 pattern are shown in Figure 9. In SG pattern 1, for example, people from northern and western areas frequently check in at one place: Zhongguancun. In fact, Zhongguancun is the most popular information technology center in Beijing. Many active Weibo users are working or consuming at Zhongguancun; they may share common interests that can be used for accurate user or business recommendations. In addition, the geographic distribution and time duration of pattern 1 can provide detailed information that city planners can use to make better decisions in this area.

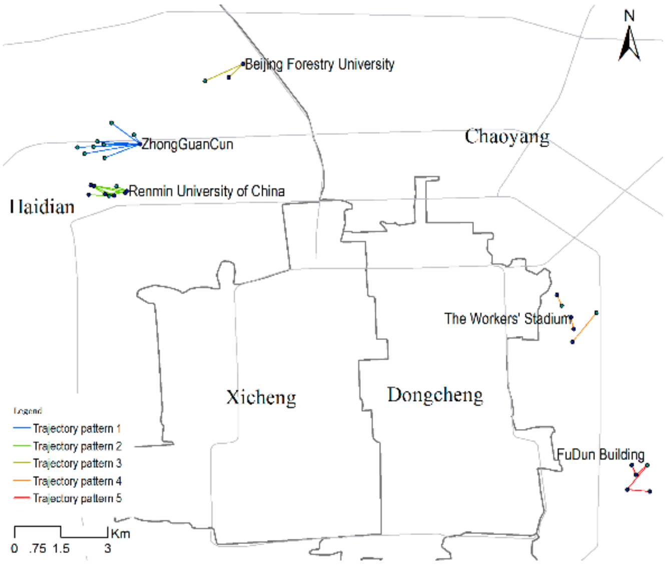


Figure 9. Five largest semantic-geographic patterns in user check-in data.

SGPattern$_1$(14 users) = {(SB, 11~14) → (OA, 16~20)}.

(4) The GTS similarity measure only generated six GTS patterns as can be seen in Figure 10. Except pattern 1, all the other patterns contain very few users. According to Formula (9), the calculation of GTS similarity doesn't consider the value of geographic similarity when users' trajectories have a smaller distance than 1 km. They were all treated the same as those who have high geographic similarities. In the end, users who have low geographic similarity (e.g., geographic distance equals to 1 km) can still be clustered in GTS patterns. That's why pattern 1 and pattern 4 in Figure 10 both have larger areas than corresponding clusters in semantic-geographic patterns.

The cluster results suggest that, compared with the three other similarity measurements, semantic intensity is more effective in identifying various and valuable trajectory patterns. The pattern results reveal more accurate information pertaining to geographic distribution and semantic meaning. This information is important for interpreting patterns, and the patterns obtained can reveal more detailed movement activities among users.

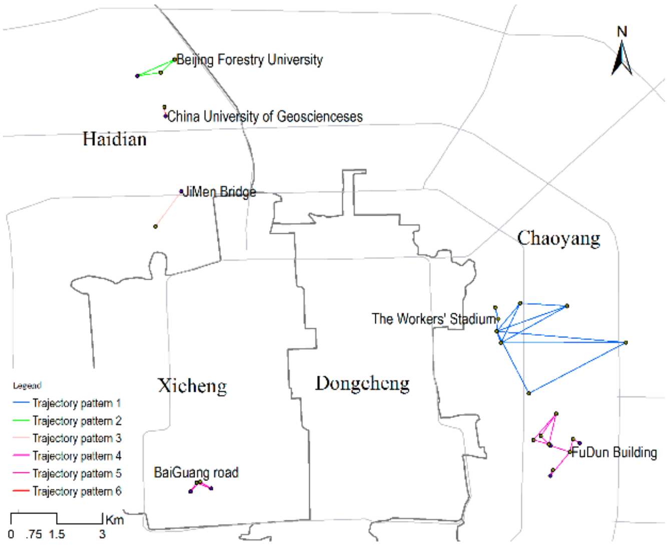


Figure 10. Six GTS patterns in user check-in data.

6. Discussion and Conclusions

In this paper, we discuss several similarity measurements for trajectory data. We formally define the new concept of a semantic-geographic pattern and propose a novel similarity measurement called semantic intensity to calculate semantic and geographic similarity within a unified framework. We then use a flexible density-based clustering algorithm to mine these semantic-geographic patterns. Comparative trajectory pattern mining experiments conducted using four different similarity measurements show that semantic intensity can effectively measure both semantic and geographic similarity among trajectory data. The experimental results also show that the patterns obtained using semantic intensity contain more interesting information than the other measurements. This information can be used to interpret patterns, clarify people's movement activities more carefully, and supply more accurate location and user recommendations.

Two factors enable semantic intensity to perform better than the other measurements. One is the proper combination of two different dimensions: semantic similarity and geographic similarity. The deviation strategy not only limits semantic intensity to values between 0 and 1, but it also avoids the task of setting weights for each dimension. Therefore, compared with the pure semantic similarity and geographic similarity measurements, semantic intensity can measure the two similarities simultaneously and is suitable for various trajectory data. The other factor is that semantic intensity only considers the segments of trajectories that have similar semantic meanings and shorter geographic distances. Because users' movement activities vary widely in the semantic, geographic and temporal dimensions, it is difficult to obtain high similarity values between two users' whole trajectories. Thus, compared with the calculation of GTS similarity, that of semantic intensity makes it easier to obtain high similarity values and to identify more common trajectory patterns.

Data quality problems are an inevitable issue that can affect the efficiency of semantic-geographic pattern mining task for all kinds of similarity measurements. In fact, the experimental data are limited by two main factors. First, check-in data can be considered a small and biased sample dataset corresponding to the entire population's activities. The check-in frequency for one person is quite low, and the check-in locations only cover a small portion of the places he/she has visited. Several data pre-processing strategies are used to fetch the representative places for each person. However, those strategies greatly reduce the number of experimental users and lead to a very small number of semantic-geographic pattern users. A recent study on mining human activity patterns (not trajectory pattern) from Twitter data also showed that only 2.72% users had very similar activity patterns based on space-time and semantics, whereas the majority (87.14%) showed different activity patterns (i.e., similar spatiotemporal patterns and different semantic patterns, similar semantic patterns and different spatiotemporal patterns, or different in both) [38]. Therefore, it is difficult to select a suitable sample dataset for the accuracy verification. Second, experimental users cover only a small portion of population, and their representativeness is hardly investigated because of privacy limitations. This limitation makes it difficult to identify the types of people in one trajectory pattern. For example, we can draw conclusions about characteristics shared by users in the semantic-geographic patterns described above. However, it is very hard to verify whether the users are actually IT employees. Thus, the conclusions can only be drawn directly from the dataset. Other trajectory data also have shortcomings that must be addressed when conducting semantic-geographic pattern mining.

In addition to the similarity measurement and data quality problem, several challenges still remain with respect to this new pattern mining task. Given the diversity of people's movement behavior, it is questionable to set fixed parameters in data clustering processing. A more flexible parameter-setting strategy and more stable clustering algorithms that can identify more accurate and valuable patterns are needed.

Acknowledgments: This work was funded through supported from the National Key Research & Development Plan of China (Project Number: 2017YFB0503601), and the National Natural Science Foundation of China (Project Number: 41471327, 41525004 and 41231171).

Author Contributions: Chenghu Zhou and Tao Pei conceived the main idea of the study and designed the experiments. You Wan performed the experiments and analyzed the results. You Wan wrote the first version of the paper and all the authors improved the following versions.

Conflicts of Interest: The authors declare that there are no conflicts of interest.

References

1. Long, J.A.; Nelson, T.A. A review of quantitative methods for movement data. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 292–318. [CrossRef]
2. Dodge, S.; Weibel, R.; Ahearn, S.C.; Buchin, M.; Miller, J.A. Analysis of movement data. *Int. J. Geogr. Inf. Sci.* **2016**. [CrossRef]
3. Claramunt, M.C. LIDU: Location-based approach to identify similar interests between users in social networks. Ph.D. Thesis, Federal University of Ceará, Fortaleza, Brazil, 2013.
4. Dodge, S.; Laube, P.; Weibel, R. Movement similarity assessment using symbolic representation of trajectories. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1563–1588. [CrossRef]
5. Ying, J.-C.; Chen, H.-S.; Lin, K.W.; Lu, E.H.-C.; Tseng, V.S.; Tsai, H.-W.; Cheng, K.H.; Lin, S.-C. Semantic trajectory-based high utility item recommendation system. *Expert Syst. Appl.* **2014**, *41*, 4762–4776. [CrossRef]
6. McMaster, R.B. A statistical analysis of mathematical measures for linear simplification. *Am. Cartogr.* **1986**, *13*, 103–116. [CrossRef]
7. Rucklidge, W.J. Efficiently locating objects using the hausdorff distance. *Int. J. Comput. Vis.* **1997**, *24*, 251–270. [CrossRef]
8. Buchin, M.; Purves, R.S. Computing Similarity of Coarse and Irregular Trajectories Using Space-Time Prisms. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*; ACM: Orlando, FL, USA, 2013; pp. 456–459.
9. Chen, L.; Ng, R. On the marriage of lp-norms and edit distance. In *Proceedings of the Thirtieth International Conference on Very Large Data Bases-Volume 30*; VLDB Endowment: Toronto, ON, Canada, 2004; pp. 792–803.
10. Vlachos, M.; Kollios, G.; Gunopulos, D. Discovering similar multidimensional trajectories. In *Proceedings. 18th International Conference on Data Engineering*; IEEE: San Jose, CA, USA, 2002; pp. 673–684.
11. Kim, S.-W.; Park, S.; Chu, W.W. Efficient processing of similarity search under time warping in sequence databases: An index-based approach. *Inf. Syst.* **2004**, *29*, 405–420. [CrossRef]
12. Giannotti, F.; Nanni, M.; Pinelli, F.; Pedreschi, D. Trajectory pattern mining. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: San Jose, CA, USA, 2007; pp. 330–339.
13. Lu, E.H.-C.; Tseng, V.S.; Yu, P.S. Mining cluster-based temporal mobile sequential patterns in location-based service environments. *IEEE Trans. Knowl. Data Eng.* **2011**, *23*, 914–927. [CrossRef]
14. Etienne, L.; Devogele, T.; Bouju, A. Spatio-temporal trajectory analysis of mobile objects following the same itinerary. *Adv. Geo-Spat. Inf. Sci.* **2012**, *10*, 47–57.
15. Lv, M.; Chen, L.; Chen, G. Mining user similarity based on routine activities. *Inf. Sci.* **2013**, *236*, 17–32. [CrossRef]
16. Yuan, Y.; Raubal, M. Measuring similarity of mobile phone user trajectories–a spatio-temporal edit distance method. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 496–520. [CrossRef]
17. Etienne, L.; Devogele, T.; Buchin, M.; McArdle, G. Trajectory box plot: A new pattern to summarize movements. *Int. J. Geogr. Inf. Sci.* **2016**, *30*. [CrossRef]
18. Alvares, L.O.; Bogorny, V.; Kuijpers, B.; de Macelo, J.; Moelans, B.; Palma, A.T. *Towards Semantic Trajectory Knowledge Discovery*; Technical Report; Hasselt University: Hasselt, Belgium, 2007.
19. Li, W.; Raskin, R.; Goodchild, M.F. Semantic similarity measurement based on knowledge mining: An artificial neural net approach. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1415–1435. [CrossRef]
20. Rettinger, A.; Lösch, U.; Tresp, V.; d'Amato, C.; Fanizzi, N. Mining the semantic web. *Data Min. Knowl. Discov.* **2012**, *24*, 613–662. [CrossRef]
21. Schwering, A. Approaches to semantic similarity measurement for geo-spatial data: A survey. *Trans. GIS* **2008**, *12*, 5–29. [CrossRef]
22. Bogorny, V.; Kuijpers, B.; Alvares, L.O. St-dmql: A semantic trajectory data mining query language. *Int. J. Geogr. Inf. Sci.* **2009**, *23*, 1245–1276. [CrossRef]

23. Xiao, X.; Zheng, Y.; Luo, Q.; Xie, X. Inferring social ties between users with human location history. *J. Ambient Intell. Humaniz. Comput.* **2014**, *5*, 3–19. [CrossRef]
24. Zhang, C.; Han, J.; Shou, L.; Lu, J.; La Porta, T. Splitter: Mining fine-grained sequential patterns in semantic trajectories. In *Porceedings of the 40th International Conference on Very Large Data Base-Volumn 7, No.9*; VLDB Endowment: Hangzhou, China, 2014; pp. 769–780.
25. Ying, J.J.-C.; Lee, W.-C.; Tseng, V.S. Mining geographic-temporal-semantic patterns in trajectories for location prediction. *ACM Trans. Intell. Syst. Technol.* **2013**, *5*, 2. [CrossRef]
26. Buchin, M.; Dodge, S.; Speckmann, B. Similarity of trajectories taking into account geographic context. *J. Spat. Inf. Sci.* **2014**, *2014*, 101–124. [CrossRef]
27. Li, Y.; Han, J.; Yang, J. Clustering moving objects. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: Seattle, WA, USA, 2004; pp. 617–622.
28. Nanni, M.; Pedreschi, D. Time-focused clustering of trajectories of moving objects. *J. Intell. Inf. Syst.* **2006**, *27*, 267–289. [CrossRef]
29. Parent, C.; Spaccapietra, S.; Renso, C.; Andrienko, G.; Andrienko, N.; Bogorny, V.; Damiani, M.L.; Gkoulalas-Divanis, A.; Macedo, J.; Pelekis, N. Semantic trajectories modeling and analysis. *ACM Comput. Surv. (CSUR)* **2013**, *45*. [CrossRef]
30. Berry, M.W.; Castellanos, M. *Survey of Text Mining II*; Springer: London, UK, 2008.
31. Aggarwal, C.C.; Zhai, C. *Mining Text Data*; Springer Science & Business Media: New York, NY, USA, 2012.
32. Zheng, Y.; Zhou, X. *Computing with Spatial Trajectories*; Springer Science & Business Media: New York, NY, USA, 2011.
33. Shonkwiler, R. Computing the Hausdorff set distance in linear time for any L_P point distance. *Inf. Process. Lett.* **1991**, *38*, 201–207. [CrossRef]
34. Alt, H.; Guibas, L.J. Discrete geometric shapes: Matching, interpolation, and approximation. *Handb. Comput. Geom.* **1999**, *1*, 121–153.
35. Li, L.; Goodchild, M.F. An optimisation model for linear feature matching in geographical data conflation. *Inf. J. Image Data Fusion* **2011**, *2*, 309–328. [CrossRef]
36. Salvador, S.; Chan, P. Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. In *ICTAI '04 Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence*; IEEE: Boca Raton, FL, USA, 2004; pp. 576–584.
37. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*; AAAI Press: Portland, OR, USA, 1996; pp. 226–231.
38. Huang, W.; Li, S. Understanding human activity patterns based on space-time-semantics. *ISPRS J. Photogramm. Remote Sens.* **2016**, *121*, 1–10. [CrossRef]