Towards Semantic-Aware Multiple-Aspect Trajectory Similarity Measuring

Lucas May Petry[1], **Carlos Andres Ferrero**[1,2], **Luis Otavio Alvares**[1], **Chiara Renso**[3], **Vania Bogorny**[1]

[1]Programa de Pós-Graduação em Ciência da Computação (PPGCC)
Universidade Federal de Santa Catarina (UFSC)
Florianópolis – SC – Brazil

[2]Instituto Federal de Santa Catarina (IFSC)
Lages – SC – Brazil

[3]Istituto di Scienza e Tecnologie dell'Informazione (ISTI)
Consiglio Nazionale delle Ricerche (CNR)
Pisa – Italy

lucas.petry@posgrad.ufsc.br, andres.ferrero@ifsc.edu.br,

luis.alvares@ufsc.br, chiara.renso@isti.cnr.it, vania.bogorny@ufsc.br

NOTICE: This is a pre-peer reviewed version of the article entitled **Towards Semantic-Aware Multiple-Aspect Trajectory Similarity Measuring**, accepted for publication in Transactions in GIS (TGIS).

Abstract. *The large amount of semantically rich mobility data becoming available in the era of Big Data, has led to the need for new trajectory similarity measures. In the context of multiple-aspect trajectories, where mobility data are enriched with several semantic dimensions, current state-of-the-art approaches present some limitations concerning the relationships between attributes and their semantics. Existing works are too strict requiring a match on all attributes, or too flexible, considering all attributes as independent. In this paper we propose MUITAS, a novel similarity measure for a new type of trajectory data with heterogeneous semantic dimensions, which takes into account the semantic relationship between attributes, thus filling the gap of the current trajectory similarity methods. We evaluate MUITAS over two real datasets of multiple aspect social media and GPS trajectories. With precision at recall and clustering techniques, we show that MUITAS is the most robust measure for multiple-aspect trajectories.*

1. Introduction

An important research topic in mobility data mining that is significantly growing in recent years is *similarity analysis*. Similarity measures are the basis for several knowledge discovery methods such as clustering, classification, sequential pattern mining, outlier detection, location prediction, etc. Most state-of-the-art works for similarity measuring have focused on the so called raw trajectories, basically considering the properties of space or space-time. This is the case for the measures LCSS (Longest Common Subsequence) [Vlachos et al. 2002], EDR (Edit Distance on Real sequence) [Chen et al. 2005], and UMS (Uncertain Movement Similarity) [Furtado et al. 2018]. These measures are very effective for answering questions about the physical movement of objects such as *which trajectories follow similar routes?* or *which trajectories visit a similar sequence of places?*

With the explosion of Big Data generated from the Internet as weather information, social network interactions (e.g. Facebook, Foursquare, Twitter), and geolocations (e.g. OpenStreetMap), mobility data can be enriched with several layers of semantic information. Examples are the visited places or Points of Interest (POIs) [Alvares et al. 2007], the means of transportation and the goal of the trip [Bogorny et al. 2014], the weather conditions during the movement, the mood of the person, his/her posts on social media. This new type of enriched trajectory is what we call big *multiple-aspect trajectory* [Ferrero et al. 2016]. An *aspect* is a point of view from which an enriched trajectory may be observed [Noël et al. 2015]. The great challenge that we want to address in this paper is how to compute the similarity of such multiple-aspect trajectories considering several aspects together. So the question that we want to answer in this paper is: *how similar are two multiple-aspect trajectories?* How can we compare two multiple-aspect trajectories with potentially many aspects and where each of them has a number of heterogeneous attributes?

A multiple-aspect trajectory is not a simple semantic trajectory represented as a sequence of stops and moves [Spaccapietra et al. 2008], since the aspects need a more complex representation. In Figure 1 we show an example of a multiple-aspect trajectory of a tourist visiting Paris that is enriched with five *aspects*: *visited places, weather conditions, transportation mode, social media posts*, and *health*. Each aspect is described by its own attributes, as for instance: (i) the *visited places* have a spatial position, a category, a rating (the stars in the figure), and a price (the dollar symbols); (ii) the *weather condition* has a spatial position, a description (e.g. sunny, cloudy, etc) and a temperature; and the (iii) aspect *health* has the heart rate. Among these aspects and their attributes, we observe that the attributes rating and price have a relationship with POI category, since they specifically refer to the POI category. The attributes temperature and description refer to the weather condition, and not to the POI category. Similarly, the heart rate of the object is related to the moving object and not to the POI category or the weather condition. Existing works for trajectory similarity fail in catching the relationships between attributes, because the semantics that relies behind the trajectory attributes has not been considered so far.

The well-known similarity measures LCSS and EDR consider two points of a trajectory as similar when all their attributes match, independently of the aspect being considered, thus implying a strong dependency relationship among all attributes. This is a problem in multiple-aspect trajectories where the number of attributes is very high, and requiring a match in all attributes of all aspects significantly reduces the number of matchings. MSM, on the other hand, gives some degree of similarity if two points of a trajectory match in *at least* one attribute. However, MSM does not consider any relationships that may exist between attributes thus considering all attributes as independent. Both assumptions that attributes are either all related or independent are too limiting for multiple-aspect trajectories.

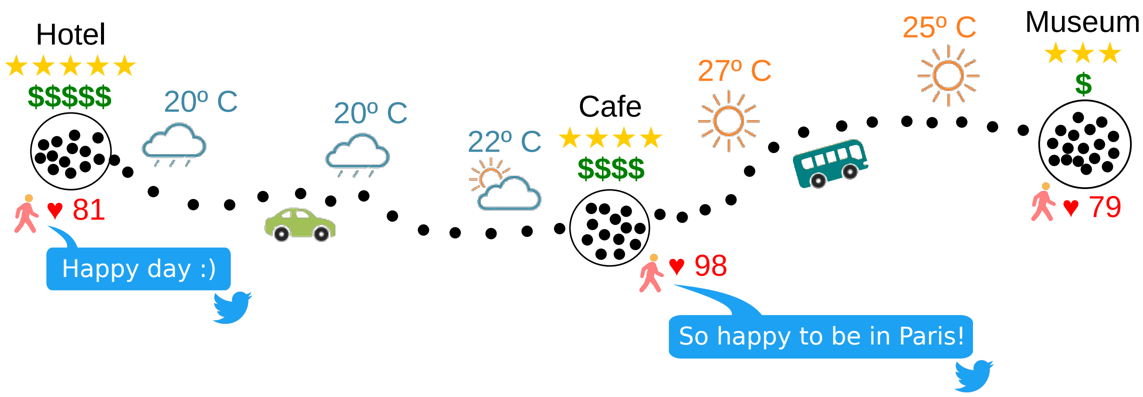

Figure 1. A multiple-aspect trajectory.

1.1. Problem Definition

Let us consider the example shown in Figure 2, with trajectories P, Q, and R. For the sake of simplicity we consider only three attributes in the example: the category of the visited place and its rating, representing the POI aspect; and the temperature representing the aspect weather. Trajectories P and Q visit the same categories of places (Hotel, Cafe, and Museum) and with the same rating. The main difference between P and Q is that P occurs where the weather temperature is low, while for trajectory Q the temperature is always high. Trajectory R, on the other hand, goes to different POIs (Barbershop, Park, and Restaurant), but their ratings are the same of trajectories P and Q, and the weather temperature is always low.

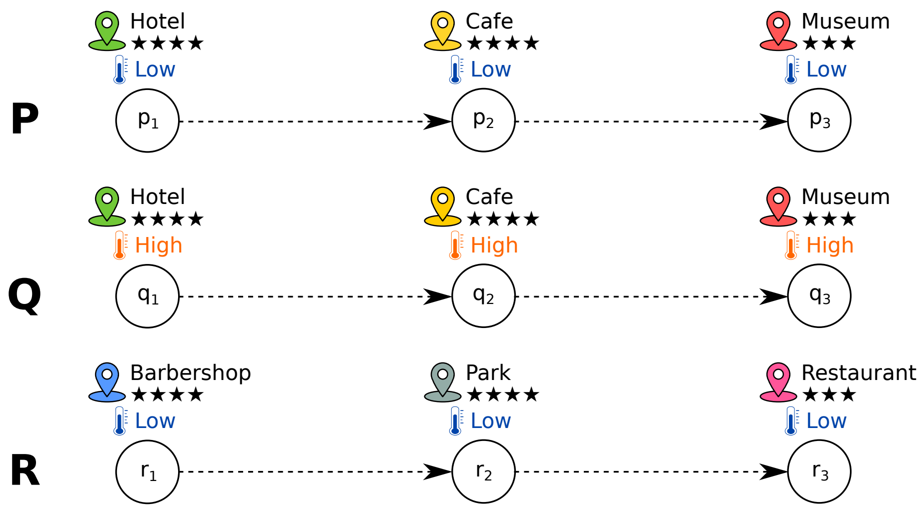

Figure 2. Example of trajectories P, Q, and R.

Now, let us suppose that we want to find the trajectory that is the most similar to P. In our example, trajectory Q is the most similar to P, because both trajectories visit places of the same category (Museum, Cafe, and Hotel) and with the same rating, only differing in the weather temperature. However, without considering the semantics of the attributes and their relationships, trajectories P and Q, and trajectories P and R would have the same similarity score by state-of-the-art methods, because they all share two common attributes: P and Q share the POI category and rating, while P and R share the rating and the temperature.

We claim that when analyzing the similarity of multiple-aspect trajectories, the semantics of the attributes and their relationships is more important than simply counting the number of attribute values that match or do not match. We believe that even though trajectory P shares two attribute values with R (rating and temperature), the trajectories are semantically different, because they visited completely different categories of places. In our example the attribute *rating* is associated to the *visited place*, so its semantics relies on the aspect POI, and its meaning is loss without the POI category, so these attributes should not be disassociated. Existing measures fail to distinguish the similarity between P and Q, and between P and R because they consider all attributes as dependent or independent (POI category disassociated to rating). MSM considers all attributes as independent, and so it gives the same similarity score of 0.66 for both P and Q, and P and R, given that in both comparisons two attributes of the trajectories match. LCSS and EDR consider all attributes as dependent, requiring a match for all three attributes, so the similarity score is zero between all trajectories, because they do not match in all attributes.

For multiple-aspect trajectories, the number of attributes increases significantly, so existing measures tend to give misleading trajectory similarity scores, because they cannot threat attributes of different aspects and do not allow the definition of attribute semantic relationships. A good similarity measure for multiple-aspect trajectories should be flexible to consider both independent and semantically related attributes.

In this paper we propose a new similarity measure called MUITAS (MUltIple aspect Trajectory Similarity), which is robust to consider the semantics behind trajectory attributes, considering both attributes with relationships and independent attributes. In summary, we make the following contributions: we present a new flexible similarity measure for *multiple-aspect trajectories* that - (i) supports both independent and dependent attributes, allowing the definition of attributes that have a semantic relationship, (ii) supports the use a different distance function for each attribute, and (iii) allows the definition of a weight that represents the importance degree of each attribute. We evaluate the proposed measure using an information retrieval and a clustering approach over two real datasets with completely different characteristics, having different and heterogeneous attributes. We use the Mean Reciprocal Rank [Craswell 2009], Mean Average Precision [Manning et al. 2008], and Hierarchical Clustering [Manning et al. 2008] to measure the quality of our work.

1.2. Scope and outline

The scope of this paper is limited to the proposal of a new similarity measure for big trajectory data that involve multiple semantic dimensions. How to integrate different sources of information in order to generate multiple-aspect trajectories is a whole new world of research, and this process is out of the scope of this paper. In this paper we assume that the trajectories are enriched with multiple aspects.

The rest of the paper is organized as follows: Section 2 presents related works, their limitations, and the main differences to our approach. Section 3 introduces the proposed similarity measure and its properties. Section 4 presents the experimental evaluation, validating the accuracy and improvements made by our approach. Lastly, Section 5 concludes the paper describing advantages and limitations of this work, in addition to potential future work.

2. Related work

The similarity of sequences and time series was the primary problem discussed in the literature, long before first works started analyzing actual trajectories. A well-known method for measuring the distance between time series was designed by Berndt [Berndt and Clifford 1994], called *Dynamic Time Warping (DTW)*. DTW aligns two sequences in order to minimize the distance between their elements. A matrix with the distances between elements of both series is created, which is then used to find the contiguous path with the minimum total distance between the series. Given DTW limitation to uni-dimensional data, [ten Holt et al. 2007] extended DTW to create *Multidimensional Dynamic Time Warping (MD-DTW)*. MD-DTW normalizes the distance of elements for all attributes and then builds the distance matrix, whose elements are the sum of the distances in all attributes for every two elements in the sequences. DTW and MD-DTW tend to be sensitive to noise because all elements of the sequences being compared are taken into consideration. Both DTW and MD-DTW consider a single distance function for all dimensions, and deal with numerical attributes only, so not being applicable to multiple aspect trajectories.

The *Longest Common Subsequence (LCSS)*[1] was introduced as a robust similarity measure for raw trajectories [Vlachos et al. 2002]. It is based on the longest common subsequence concept, in which two sequences are considered as similar if they have similar behavior for a large part of their length. Differently from DTW and MD-DTW, LCSS reduces the impact of noisy data by defining distance and matching thresholds. Two points match and are assigned a similarity value of 1 if their distance lies below the matching threshold; otherwise, they do not match and have a similarity of 0. Although it works well with noisy data, LCSS has some disadvantages. LCSS ignores possible gaps of points in trajectories, which, for certain problems, would mean giving the same similarity value for different pairs of trajectories. A gap refers to the existence of a subtrajectory in between two similar components of two trajectories. Additionally, LCSS considers all attributes to be dependent, so two points are similar only when all their attributes match. With this limitation, the more trajectory attributes we consider in the similarity assessment, which is needed for multiple-aspect trajectories, the less similar trajectories tend to be.

Chen [Chen et al. 2005] proposed *Edit Distance on Real sequence (EDR)*, a distance measure for trajectories based on Edit Distance (ED) that is widely used for measuring similarity between strings. The underlying idea in EDR is that, being A and B two trajectories, $EDR(A, B)$ is given by the minimum number of insert, delete and replacement of points needed to transform A into B. EDR assigns 0 when two points are similar and 1 otherwise. Besides reducing the effects of noise, EDR overcomes a major drawback present in LCSS: it assigns penalties according to the length of the gaps between two matched sub-trajectories, which results in more accurate similarity scores. However, EDR also computes a match for two points only if all attributes match, which may be too restrictive for analyzing multiple-aspect trajectories.

[1] Even though LCSS was first designed for time series in [Bollobás et al. 1997], we only consider the most recent approach proposed by Vlachos [Vlachos et al. 2002] for trajectory data, since it is more robust than the first one.

Table 1. Features and limitations of main related works.

	LCSS [Vlachos et al. 2002]	EDR [Chen et al. 2005]	MSM [Furtado et al. 2015]	DTW$_A$ [Shokoohi-Yekta et al. 2017]	UMS [Furtado et al. 2018]	MUITAS
Robust to noise (outliers)	✓	✓	✓		✓	✓
Trajectory gaps		✓	✓		✓	✓
Different distance functions			✓			✓
Attribute weighing			✓			✓
Non-rigid sequence			✓			✓
Multiple-aspect trajectories						✓
No attribute relationship			✓	✓		✓
Partial attribute relationship						✓
Full attribute relationship	✓	✓		✓	✓	✓

An important remark about LCSS and EDR is that both measures were proposed when trajectory data were still limited to the space and time dimensions. Therefore, it was appropriate to consider all attributes as interdependent. However, with multiple-aspect trajectory data and many different attributes, these measures are not robust in the similarity assessment.

More recently, Furtado presented MSM [Furtado et al. 2015], a new similarity measure that overcame several limitations of previous works, because it explicitly adds the semantic dimension in addition to the space and time. MSM also defines weights for every attribute, given that an attribute might be more or less important for different problems. Essentially, given two trajectories A and B, for every point of A, MSM looks for the best match in B. Subsequently, the weighed scores of the matches are added to compose the parity of A with B. Since the parity is not symmetric, $MSM(A, B)$ is computed by the average of $parity(A, B)$ and $parity(B, A)$. Rather than considering pairs of points only if they match for all attributes, MSM treats all attributes separately, and assigns partial similarity according to the number of attributes in which the points match. This flexibility tends to increase the general similarity score. MSM disregards any relationships that might exist between aspects or attributes, so being less robust for multiple-aspect trajectories.

Shokoohi-Yekta in [Shokoohi-Yekta et al. 2017] proposed an adaptive DTW-based approach for multidimensional time series classification, namely DTW$_A$. DTW$_A$ runs both an independent and a dependent version of DTW, DTW$_I$ and DTW$_D$, respectively, and then chooses the best approach according to a scoring function and a threshold. Despite being an adaptive approach, DTW$_A$ only considers all attributes either dependent or independent, not allowing specific relationships between attributes. In addition, DTW$_A$ carries limitations present in DTW, such as rigidity to the sequence of points, sensitivity to noise, and supports numerical attributes only.

A recent work proposed by Furtado [Furtado et al. 2018] is Uncertain Movement Similarity (UMS), which is more robust than previous works regarding different sampling rates and the heterogeneity of raw trajectory data. Despite of its robustness, UMS is limited to spatial attributes, so focusing on spatial similarity, not being appropriate for multiple aspect trajectories.

Table 1 summarizes the main related works. So far, to the best of our knowledge, there is no similarity measure in the literature for multiple-aspect trajectories. Indeed, previously mentioned works address trajectory similarity regarding trajectory attributes, either in a too restrictive or too flexible manner. MUITAS is more flexible than existing measures because (i) it supports full attribute relationship as previous works, (ii) supports partial attribute relationships, and (iii) no attribute relationships as well. Table 1 compares the characteristics of the main discussed approaches and our similarity measure, such as robustness to noise, use of different distance functions for different attributes, the capability of considering attribute relationships, among others. As shown in Table 1, MUITAS has the challenge to group together the main characteristics of other works, thus supporting multiple-aspect trajectories. It is worth mentioning that only MSM and MUITAS were developed for trajectories with semantic attributes.

3. MUITAS: Multiple-Aspect Trajectory Similarity Measure

In this section we introduce the fundamental concepts of our work and we define MUITAS (Multiple-Aspect Trajectory Similarity Measure), a similarity measure for multiple-aspect trajectories. Afterwards, we show a running example of MUITAS. We begin by defining an aspect and multiple-aspect trajectory in Definition 1 and Definition 2, respectively, which, to the best of our knowledge, have not been introduced before.

Definition 1. *Aspect.* An aspect is a set $A = \{a_1, a_2, \ldots, a_l\}$ of l characterizing attributes that semantically represent A.

An aspect is essentially any sort of information that can be annotated to a trajectory. For instance, we may define aspects such as the weather, the POI, and the means of transportation. The weather may have as attributes the condition description, the temperature, and humidity; the POI could be described by the attributes type, rating, and price tier; and the means of transportation could be characterized by its type and average speed. These different aspects and their attributes are associated to the trajectory points, as stated in Definition 2.

Definition 2. *Multiple-aspect trajectory.* A multiple-aspect trajectory is a sequence of points $T = \langle p_1, p_2, \ldots, p_n \rangle$, with $p_i = (x, y, t, A)$ being the i-th point of the trajectory at location (x, y) at timestamp t, described by the set $A = \{A_1, A_2, \ldots, A_r\}$ of r aspects.

Definition 2 states that a multiple-aspect trajectory is annotated with any sort of information, which we call aspects. A point of a multiple-aspect trajectory can be as simple as a point of a raw trajectory ($\mathcal{A} = \emptyset$), or a more complex element with other aspects besides space and time. In order to measure the similarity between two multiple-aspect trajectories it is necessary to quantify the distance between points. Notice that attributes may refer to different types of data, and so for each point we must quantify the distance for each attribute. Having distinct natures they require different distance functions. To measure the attribute similarity we introduce the concept of attribute matching.

Definition 3. *Attribute matching.* Let P and Q be two multiple-aspect trajectories $P = \langle p_1, p_2, \ldots, p_m \rangle$ and $Q = \langle q_1, q_2, \ldots, q_n \rangle$. For any two points $p \in P$ and $q \in Q$, the distance between p and q on an attribute a_i of an aspect A_j is given by the function $dist_i : p \times q \rightarrow \mathbb{Q}$. Two points $p \in P$ and $q \in Q$ will *match* on attribute a_i if $dist_i(p, q) \leq \delta_i$, where δ_i is a distance threshold for attribute a_i.

For each attribute a different distance function can be used, as for instance the Euclidean distance for a spatial attribute, a hierarchy-based distance for the category of a POI, a simple discrete distance for the weather condition, etc. As a different distance function can be used for each attribute, the measure becomes feasible for a variety of applications. Having defined the way we measure the attribute distance, we now must define how to aggregate attributes that belong to the same aspect. For this we introduce the concept of *Feature*.

Definition 4. *Feature*. A feature $f = \{a_1, a_2, \ldots, a_z\}$ is a nonempty set of attributes that describe a unit of analysis of a multiple-aspect trajectory.

To avoid misunderstanding and conflict of concepts, we hereafter refer to *attribute* as an atomic view of a point, and to *feature* as a unit of analysis of a trajectory. In other words, attributes that are independent are defined as features with a single attribute, while attributes with relationships are defined together in the same feature.

As the important features for similarity analysis are application dependent, we give the formal definition of *Application* in Definition 5. For instance, in a tourism application, the important features can be {*place category and price tier*}, {*place category and duration of the visit*}, {*weather condition*}, etc. We define an application according to the attributes, distance functions, distance thresholds, and features used in the analysis.

Definition 5. *Application*. An application $\mathbb{A}$ is defined by a tuple $\mathbb{A} = (\mathcal{A}, \mathcal{D}, \Delta, \mathcal{F}, \mathcal{W})$, where $\mathcal{A} = \{a_1, a_2, \ldots, a_l\}$ is a nonempty set of attributes, $\mathcal{D} = \{dist_1, dist_2, \ldots, dist_l\}$ is a nonempty set of distance functions, $\Delta = \{\delta_1, \delta_2, \ldots, \delta_l\}$ is a nonempty set of distance thresholds, $\mathcal{F} = \{f_1, f_2, \ldots, f_k\}$ is a nonempty set of features, and $\mathcal{W} = \{w_1, w_2, \ldots, w_k\}$ is a nonempty set of weights. $dist_i$ and δ_i are the distance function and threshold of attribute a_i. For each feature $f_i \in \mathcal{F}$ we define a corresponding weight $w_i \in \mathcal{W}$, and $\sum_{i=1}^{|\mathcal{F}|} w_i = 1$.

An application essentially defines the context of the problem, i.e., how trajectories will be analyzed. Different applications may imply different features, different distance functions and/or different thresholds. The weight w_i of a feature f_i represents the importance of that feature for computing the similarity between trajectories for a specific application. Given an application $\mathbb{A}$, we must now define how to measure the similarity between trajectory points and the trajectories themselves. Definition 6 presents the *score* function used to compute the similarity score between trajectory points.

Definition 6. *Score*. Given two trajectory points $p \in P$ and $q \in Q$, and an application $\mathbb{A} = (\mathcal{A}, \mathcal{D}, \Delta, \mathcal{F}, \mathcal{W})$, the matching score between p and q is given by the function $score : P \times Q \rightarrow [0, 1]$, defined as follows

$$score(p, q) = \sum_{i=1}^{|\mathcal{F}|} (match_{f_i}(p, q) * w_i),$$

$$\text{where } match_{f_i}(p, q) = \begin{cases} 1, & \text{if } \forall a_j \in f_i,\ dist_j(p, q) \leq \delta_j \\ 0, & \text{otherwise.} \end{cases}$$

At this point, we have the basic definitions necessary to propose the multiple-aspect trajectory similarity measure. Furtado in [Furtado et al. 2015] defines a parity function which is the basis of the similarity measure MSM. The parity function adds the scores of the best matches of the points of one trajectory with points of another trajectory. We use the same function in our similarity measure, which is given in Definition 7.

Definition 7. *Parity.* Given the set $\mathcal{S}$ of multiple-aspect trajectories, and P and Q two multiple-aspect trajectories in $\mathcal{S}$, the parity of P with Q is given by the function $parity : \mathcal{S}^2 \rightarrow [0, |P|]$, defined as follows

$$parity(P, Q) = \sum_{p \in P} max(\{score(p, q) \mid q \in Q\})$$

It is worth highlighting that, differently to existing similarity measures, MUITAS allows the definition of relationships between attributes for assessing trajectory similarity. The similarity of two multiple-aspect trajectories P and Q, computed by MUITAS, is given by the average parity of P and Q, which is given in Definition 8.

Definition 8. *MUITAS.* Given the set $\mathcal{S}$ of multiple-aspect trajectories, and P and Q two multiple-aspect trajectories in $\mathcal{S}$, the similarity score of P and Q is calculated by the function $MUITAS : \mathcal{S}^2 \rightarrow [0, 1]$, defined as

$$MUITAS(P, Q) = \begin{cases} 0, & \text{if } |P| = 0 \text{ or } |Q| = 0 \\ \dfrac{parity(P, Q) + parity(Q, P)}{|P| + |Q|}, & \text{otherwise.} \end{cases}$$

Similarly to MSM, MUITAS holds the properties of *non-negativity* (Lemma 1), *relaxed identity of indiscernibles* (Lemma 2) and *symmetry* (Lemma 3).

Lemma 1. *Non-negativity.* Given two multiple-aspect trajectories P and Q, $MUITAS(P, Q) \geq 0$.

Proof: Direct from Definitions 7 and 8.

Lemma 2. *Relaxed identity of indiscernibles.* Given two multiple-aspect trajectories P and Q under an application $\mathbb{A} = (\mathcal{A}, \mathcal{D}, \Delta, \mathcal{F}, \mathcal{W})$, then $MUITAS(P, Q) = 1$ if and only if $P = Q$ or $(\forall p \in P\ \exists q \in Q \mid \forall dist_i \in \mathcal{D}, \delta_i \in \Delta\ dist_i(p, q) \leq \delta_i)$ and $(\forall q \in Q\ \exists p \in P \mid \forall dist_i \in \mathcal{D}, \delta_i \in \Delta\ dist_i(q, p) \leq \delta_i)$.

Proof: By Definition 6, if $dist_i(p, q) \leq \delta_i$ for all attributes a_i in a feature f_k, then $match_{f_k}(p, q) = 1$. Hence, $score(p, q) = 1$, because $match_{f_k}(p, q) = 1$ for all features $f_k \in \mathcal{F}$. Therefore, $parity(P, Q) = |P|$, because for any $p \in P$ there is a $q \in Q$ where $score(p, q) = 1$. Similarly, $parity(Q, P) = |Q|$. By Definition 8, $MUITAS(P, Q) = \frac{|P|+|Q|}{|P|+|Q|} = 1$. If $P = Q$, by Definition 8 then $MUITAS(P, Q) = 1$. On the other hand, if for one attribute a_i and at least one point $p \in P$ there is no $q \in Q$ such that $dist_i(p, q) \leq \delta_i$, then $score(p, q) < 1$, $parity(P, Q) < |P|$ and, therefore, $MUITAS(P, Q) < 1$.

Lemma 3. *Symmetry.* Given two multiple-aspect trajectories P and Q, $MUITAS(P, Q) = MUITAS(Q, P)$.

Proof: Direct from Definition 8.

To better understand the proposed measure and how it differs from state-of-the-art works, in the following section we compare the similarity scores for the introductory example in Section 1.

3.1. Running example

In this section we present a running example using trajectories P, Q, and R introduced in Section 1 in Figure 2, for which existing measures give undesired results. As previously mentioned, we want to find the trajectory most similar to P. We instantiate an application $\mathbb{A} = (\mathcal{A}, \mathcal{D}, \Delta, \mathcal{F}, \mathcal{W})$, for which Table 2 describes the set of features $\mathcal{F}$, the weights $\mathcal{W}$, the attributes $\mathcal{A}$, the distance functions $\mathcal{D}$, and thresholds Δ. For the sake of simplicity, all distance functions are binary, i.e., any two attributes match only if they are equal. Also, we defined the feature weights according to the number of attributes the features contain, in order to make a fair comparison with MSM.

Let us compute the similarity of P and Q. The first step is to compute the similarity scores between all points of both trajectories. Starting from p_1 and q_1, the only attribute they have in common is the rating. Therefore, the score of p_1 and q_1 is zero because the rating is in the feature f_1, and so the categories should also be equal for a match on the feature f_1 to occur. Similarly, p_1 and q_2 have only the attribute rating in common, so their score is zero as well. However, p_1 and q_3 have both the same category and rating. Because the feature category and rating attributes are equal, there is a match on the feature f_1, and the score of p_1 and q_3 is $2/3$.

Table 2. Features, weights, attributes, distance functions, and thresholds.

$\mathcal{F}$	$\mathcal{W}$	$\mathcal{A}$	$\mathcal{D}$	Δ
f_1	$2/3$	a_1 = Category	$dist_1(p,q) = 0$ if p.a_1 = q.a_1, 1 otherwise	0
		a_2 = Rating	$dist_2(p,q) = 0$ if p.a_2 = q.a_2, 1 otherwise	0
f_2	$1/3$	a_3 = Temperature	$dist_3(p,q) = 0$ if p.a_3 = q.a_3, 1 otherwise	0

Table 3. Scores of the points of P and Q.

P × Q	q_1	q_2	q_3
p_1	0	0	$2/3$
p_2	0	$2/3$	0
p_3	$2/3$	0	0

Table 4. Scores of the points of P and R.

P × R	r_1	r_2	r_3
p_1	$1/3$	$1/3$	$1/3$
p_2	$1/3$	$1/3$	$1/3$
p_3	$1/3$	$1/3$	$1/3$

Table 3 presents the computed scores between the points of trajectories P and Q. For computing $parity(P, Q)$, for each point in P, we need to sum its best scoring point in Q, i.e., $parity(P, Q)$ is the sum of the highest scores on each line of Table 3. Thus, $parity(P, Q) = 3 \times 2/3 = 2$. Correspondingly, $parity(Q, P)$ is the sum of the highest scores on each column of the table, which gives us $parity(Q, P) = 3 \times 2/3 = 2$. Finally, we compute $MUITAS(P, Q)$ as follows:

$$MUITAS(P, Q) = \frac{parity(P, Q) + parity(Q, P)}{|P| + |Q|} = \frac{2 + 2}{3 + 3} = \frac{2}{3}$$

For computing the similarity of trajectories P and R we also need to compute the scores between their points, which is shown in Table 4. The score of p_1 and r_1 is $1/3$, because the only feature in which they entirely match is f_2 with the temperature attribute. Although they have the same rating, their categories are different. The same occurs for all point comparisons. We have $parity(P, R) = parity(R, P) = 3 \times 1/3 = 1$, and so $MUITAS(P, R) = 1/3$. Table 5 shows the similarity scores given by existing works for P and Q, and P and R. MUITAS is the only measure that is able to distinguish between Q and R, assigning a lower similarity score for P and R, thus retrieving only Q as the most similar to P.

Table 5. Similarity scores given by different measures.

	LCSS	EDR	MSM	MUITAS
$sim(P, Q)$	0	0	2/3	2/3
$sim(P, R)$	0	0	2/3	1/3

In the next section we present the experimental evaluation over real-world datasets.

4. Experimental evaluation

In this section we evaluate the accuracy of the proposed similarity measure over two real trajectory datasets with different characteristics, to show the robustness of MUITAS considering different application domains: (i) a dataset of Foursquare check-ins in the city of New York collected between April 2012 and February 2013 [Yang et al. 2015], and (ii) a dataset of semantic trajectories collected in Pisa, Italy between May 20, 2014 and September 30, 2014[2]. We evaluate the Mean Average Precision (MAP), the Mean Reciprocal Rank (MRR), and we perform Hierarchical Clustering Analysis (HCA), similarly to the evaluations reported in [Chen et al. 2005], [Furtado et al. 2018]. The similarity measures were implemented in Java and R[3] and the experiments were conducted on a PC running Linux Ubuntu 18.04 LTS, equipped with an Intel Core i7-3630QM CPU @ 2.4GHz x 8 and 6GB RAM. The next sections describe the datasets (Section 4.1), the ground truth definition (Section 4.2), the experimental setup (Section 4.3), and the achieved results (Section 4.4).

[2]We performed this experiment on data collected in the context of the TagMyDay experiment under a non-disclosure agreement, during a visit funded by the SOBIGDATA Project, so we cannot redistribute it. More information about the dataset can be found at `http://kdd.isti.cnr.it/project/tagmyday`.

[3]Link to source code will be disclosed upon acceptance.

Table 6. Foursquare dataset attributes description.

Attribute	Type	Range/example	N.	Distance function
Price tier	Numeric	$\{-1, 1, 2, 3, 4\}$	5	Euclidean
Rating	Numeric	$\{-1\} \cup [4.0, 10.0]$	62	Euclidean
Time	Temporal	[00:00,23:59]	1440	Euclidean[6]
Venue category	Nominal	{Arts & Entertainment, College & University, ...}	10	Binary
Weather	Nominal	{Clear, Clouds, Rain, ...}	6	Binary
Weekday	Nominal	{Weekday, Weekend}	2	Binary

4.1. Datasets

The Foursquare dataset contains 227,428 check-ins of 1,083 different users, and each check-in is composed of a timestamp and the corresponding Foursquare venue ID. We then collected venue information, including the spatial position, the rating, and the price tier, from the Foursquare API[4]. Subsequently, historical weather data were collected via the Weather Wunderground API[5] and combined with each Foursquare check-in. Table 6 describes the attributes and the distance functions used for each attribute.

The Pisa dataset was collected by 157 volunteers in Pisa, via an app installed in the user mobile phone. The trajectories used in this experiment are composed by movement segments that represent the user daily routine. Each segment was annotated with the means of transportation, the purpose of the trip, the weather conditions, the traveled distance and time. In total the dataset has 10,880 segments, each described by the attributes shown in Table 7. Both the Foursquare and the Pisa dataset are important for our evaluation because they contain multiple aspect information.

Having described the datasets, in the next section we describe the ground truth defined for evaluating the similarity measure.

4.2. Ground truth definition

The check-ins of the Foursquare dataset and the segments of the Pisa dataset are not labeled with a class and, for that reason, we use a similar approach to the Trajectory-User Linking problem introduced by [Gao et al. 2017] to evaluate our method. We applied a few transformations to the datasets in order to ensure variability and consistency, as described in the following.

We first removed 26 check-ins with missing information about their category on Foursquare. Next, we removed 21,332 noisy check-ins that belong to broad categories such as roads, rivers, neighborhoods, etc, because the geographic location is unique for each venue. Subsequently, we removed 1,230 check-ins that were duplicated, considering a 10-minute threshold.

[4]https://developer.foursquare.com/

[5]https://www.wunderground.com/weather/api/

[6]Difference in minutes.

[7]Difference in minutes.

Table 7. Pisa dataset attributes description.

Attribute	Type	Range/example	N.	Distance function
Activity	Nominal	{Going home, Refueling, ...}	14	Binary
Start time	Temporal	[00:00,23:59]	1440	Euclidean[7]
Distance	Ordinal	{Up to 1km, 1 to 2km, ..., Above 10km}	5	Euclidean
End time	Temporal	[00:00,23:59]	1440	Euclidean[7]
Time duration	Temporal	(0, 86400]	86400	Euclidean
Transportation	Nominal	{Bike, Car, Taxi, ...}	9	Binary
Weather	Nominal	{Sunny, Clouds, ...}	5	Binary
Weekday	Nominal	{Weekday, Weekend}	2	Binary

We then created weekly trajectories of check-ins for each user, given the whole set of check-ins. We claim that a weekly trajectory of a user is more similar to trajectories of the same user and less similar to other user trajectories. Hence, we labeled each weekly trajectory with the corresponding user, which defines our ground truth. We filtered the weekly trajectories in order to ensure variability in the evaluation: (i) we removed short trajectories with less than 10 check-ins and (ii) removed all trajectories of users with less than 10 trajectories. The final dataset contains a total of 66,962 check-ins distributed in 3,079 weekly trajectories of 193 different users, with an average length of 22 points (check-ins) per trajectory and an average of 16 trajectories per user.

For the Pisa dataset we created daily trajectories, because differently from the Foursquare dataset, the trajectory points are less sparse and they represent the detailed user movement and daily routine. In order to ensure variability, we removed small trajectories with less than 3 segments and then removed users with less than 5 trajectories. The final dataset contains a total of 8,800 segments in 1,535 daily trajectories of 67 different users. The trajectories have an average length of 6 segments and an average of 23 trajectories per user. In the next section we detail the metrics used to evaluate the results.

4.3. Experimental setup

Similarity measures are commonly used in clustering analysis, recommendation and information retrieval systems. We evaluate the proposed method with three different analyses also performed in previous works, as in [Chen et al. 2005, Furtado et al. 2018, Esuli et al. 2018]. We measure the Mean Average Precision (MAP) and the Mean Reciprocal Rank (MRR) in an information retrieval task, and we perform Hierarchical Clustering Analysis (HCA) using EDR, LCSS, MSM, and MUITAS for computing similarity between trajectories. We do not compare our work to MD-DTW, UMS, and DTW_A, because multiple-aspect trajectories have categorical attributes, and these measures were designed for numerical attributes of time-series and trajectories.

MAP and MRR are rank-based measures commonly used for evaluating information retrieval systems. MAP summarizes in a single value the precision at different levels of recall, measuring how well similarity measures can retrieve all relevant trajectories for

Table 8. Threshold values employed for attributes in the Foursquare dataset.

Attribute	Unit	Δ
Price tier	Price score	$\{0, 1, 2\}$
Rating	Rating score	$\{0.5, 1.0, 1.5\}$
Time	Minutes	$\{15, 30, 60\}$

Table 9. Threshold values employed for attributes in the Pisa dataset.

Attribute	Unit	Δ
Begin time	Minutes	$\{30, 60, 120\}$
Distance	Distance unit	$\{0, 1, 2\}$
End time	Minutes	$\{30, 60, 120\}$
Time duration	Seconds	$\{1800, 3600, 7200, 10800\}$

each class. MRR, on the other hand, measures how well similarity measures can retrieve one relevant trajectory for each class. Given a trajectory T in the dataset, we rank all other trajectories according to their similarity scores with T. The closer to the top a trajectory of the same class of T is, the higher will be the MRR score. Both MAP and MRR range from 0 to 1, being 1 the best score.

We run complete-linkage hierarchical clustering and evaluate the generated clusters using the F-score, as described in [Manning et al. 2008]. F-score weighs individual cluster quality and the number of generated clusters. For instance, as the number of classes within a cluster increases, precision decreases and the score is penalized. Similarly, as the number of clusters overpasses the number of classes, recall increases and the F-score falls. In other words, F-score is equal to 1 if and only if all clusters are pure and the number of clusters is equal to the number of classes.

We instantiate an application $\mathbb{A} = (\mathcal{A}, \mathcal{D}, \Delta, \mathcal{F}, \mathcal{W})$, where the attributes $\mathcal{A}$ and distance functions $\mathcal{D}$ are presented in Tables 6 and 7 for each dataset. Table 8 and Table 9 show the thresholds Δ tested for each attribute in the datasets. Only attributes with a threshold value greater than zero are displayed. We chose different reasonable threshold values for each attribute in order to evaluate the similarity measures under different threshold configurations. We analyze MAP, MRR and HCA for all configurations of thresholds, totaling 27 results for the Foursquare dataset and 108 results for the Pisa dataset.

For the Foursquare dataset we defined the features $\mathcal{F}$ as: (i) the venue category and the rating; (ii) the venue category and the time; and other attributes remained as separate features. For the Pisa dataset the features $\mathcal{F}$ are: (i) the activity and the traveled distance; (ii) the activity and the time duration; (iii) the activity and the transportation means; and the remaining features are considered with single attributes. In order to have a fair comparison, we defined the set of weights $\mathcal{W}$ to be equal for all attributes for MSM, and proportional weights according to the number of attributes for the features of MUITAS. We report and discuss the achieved results in these datasets in the following section.

Table 10. Average scores and standard deviations for MRR, MAP and HCA for all methods on the Foursquare dataset (reported in the format AVG $\pm$ SD).

	MRR	MAP	HCA
EDR [Chen et al. 2005]	0.447 ± 0.012	0.207 ± 0.005	0.086 ± 0.008
LCSS [Vlachos et al. 2002]	0.412 ± 0.024	0.213 ± 0.009	0.112 ± 0.005
MSM [Furtado et al. 2015]	0.570 ± 0.014	0.300 ± 0.012	0.265 ± 0.024
MUITAS	$\mathbf{0.657} \pm 0.011$	$\mathbf{0.383} \pm 0.012$	$\mathbf{0.382} \pm 0.022$

Table 11. Average scores and standard deviations for MRR, MAP and HCA for all methods on the Pisa dataset (reported in the format AVG $\pm$ SD).

	MRR	MAP	HCA
EDR [Chen et al. 2005]	0.480 ± 0.008	0.173 ± 0.004	0.054 ± 0.006
LCSS [Vlachos et al. 2002]	0.451 ± 0.009	0.172 ± 0.003	0.055 ± 0.007
MSM [Furtado et al. 2015]	0.551 ± 0.016	0.248 ± 0.009	0.176 ± 0.014
MUITAS	$\mathbf{0.604} \pm 0.020$	$\mathbf{0.280} \pm 0.009$	$\mathbf{0.206} \pm 0.011$

4.4. Results and discussion on Pisa and Foursquare datasets

Table 10 and Table 11 report the results in the Foursquare and Pisa datasets, respectively, for EDR, LCSS, MSM, and MUITAS, under each of the three evaluated measures, MRR, MAP, and HCA. The results show that the average scores achieved by our method are higher than state-of-art approaches for both datasets.

EDR and LCSS achieved the lowest average scores for MRR, MAP, and HCA in both datasets, because of the strict way through which they analyze trajectory points. This means, for instance, that some attributes in the datasets are independent from each other, and when considered all together they do not represent a discriminating pattern for most of the users. MSM had better results than EDR and LCSS, especially for hierarchical clustering analysis. This confirms our claim that some attributes may be independent, since MSM does not consider any relationships between attributes.

MUITAS achieved the best averages regardless of dataset and evaluation technique, since it allows partial relationship definition, considering both dependent and independent attributes. MUITAS is neither too strict as LCSS and EDR, nor too flexible as MSM. It is also important to highlight that MUITAS achieved better results independent of dataset, thresholds and evaluation technique.

In order to evaluate the significance of the results we also perform a statistical analysis using the ANOVA (Analysis of Variance) test [Montgomery 2017], with level of significance $\alpha = 0.05$, which results in a *p-value* < 0.05. Subsequently, we perform the *Dunnett's post hoc test (control vs. all)* with MUITAS as the control, obtaining all *p-values* < 0.05. We achieved the same test results for MRR, MAP, and HCA. Therefore, MUITAS significantly outperforms existing similarity measures for both datasets under all evaluated measures.

One important remark on these experiments is that the best results for both datasets were obtained with the measures MSM and MUITAS, specifically developed for semantic trajectories, and which according to Table 1 do not consider the element sequence. This means that for the Foursquare and Pisa datasets, the sequence of the users behavior in trajectories may not be relevant to discriminate them from each other.

In summary, the results in these datasets show that using MUITAS for measuring similarity gives more precise results (as in MRR and MAP measures) and more concise clusters (as in HCA), in comparison to state-of-art trajectory similarity measures.

5. Conclusions and future work

The enrichment of movement data with different contexts and several data attributes has led to a new type of trajectory, that we call *multiple-aspect trajectory*. We claim that for a better understanding of movement patterns of human mobility, these attributes should be considered in the similarity assessment. However, these heterogeneous data attributes, including space, time, and several layers of semantics, makes the trajectory similarity problem more complex than traditional spatio-temporal data. To the best of our knowledge, there are no similarity measures in the literature that consider multiple-aspect trajectories. In this paper we propose MUITAS, a similarity measure that supports both independent and dependent attributes, a different distance function for each attribute, as well as a weight that represents the importance of each attribute. The state-of-the-art methods consider all attributes as independent or dependent. MUITAS overcame this limitation, by allowing partial attribute dependency. Indeed, a distinctive characteristic of MUITAS is the definition of the features of an aspect as part of the application definition that drive the similarity measurement. It is important to point out that MUITAS does not depend on the specific application domain and it can be easily applied to different scenarios.

In order to evaluate the relevance and effectiveness of MUITAS we performed a robust experimental evaluation over two real-world data sets with complementary characteristics with three different evaluation techniques. The results showed that MUITAS is more accurate than existing trajectory similarity measures.

Even though we focused on multiple-aspect trajectories, the proposed similarity measure can be applied to any type of trajectories or sequenced data of a variety of applications. As future work we will analyze the similarity of heterogeneous points of trajectories, and propose an extension of MUITAS to support trajectories that have points with different aspects and/or different attributes.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo a Pesquisa e Inovação do Estado de Santa Catarina (FAPESC) - Project Match - co-financing of H2020 Projects - Grant 2018TR 1266, and the European Union's Horizon 2020 research and innovation programme under GA N. 777695 (MASTER[8]). A special acknowledgement to the SOBIGDATA Project, for the transnational visiting support that allowed access to the TagMyDay dataset[9]. The views and opinions expressed in this paper are the sole responsibility of the author and do not necessarily reflect the views of the European Commission.

[8]http://www.master-project-h2020.eu/

[9]http://sobigdata.eu

References

[Alvares et al. 2007] Alvares, L. O., Bogorny, V., Kuijpers, B., de Macedo, J. A. F., Moelans, B., and Vaisman, A. (2007). A model for enriching trajectories with semantic geographical information. In *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*, page 22. ACM.

[Berndt and Clifford 1994] Berndt, D. J. and Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In *AAAIWS'94 Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, pages 359–370, Seattle, WA. AAAI Press.

[Bogorny et al. 2014] Bogorny, V., Renso, C., Aquino, A. R., Lucca Siqueira, F., and Alvares, L. O. (2014). Constant–a conceptual data model for semantic trajectories of moving objects. *Transactions in GIS*, 18(1):66–88.

[Bollobás et al. 1997] Bollobás, B., Das, G., Gunopulos, D., and Mannila, H. (1997). Time-series similarity problems and well-separated geometric sets. In *SCG '97 Proceedings of the thirteenth annual symposium on Computational geometry*, pages 454–456, New York, NY. ACM.

[Chen et al. 2005] Chen, L., Özsu, M. T., and Oria, V. (2005). Robust and fast similarity search for moving object trajectories. In *SIGMOD '05 Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pages 491–502, New York, NY. ACM.

[Craswell 2009] Craswell, N. (2009). Mean reciprocal rank. In *Encyclopedia of Database Systems*, pages 1703–1703. Springer.

[Esuli et al. 2018] Esuli, A., Petry, L. M., Renso, C., and Bogorny, V. (2018). Traj2user: exploiting embeddings for computing similarity of users mobile behavior. *arXiv preprint arXiv:1808.00554*.

[Ferrero et al. 2016] Ferrero, C. A., Alvares, L. O., and Bogorny, V. (2016). Multiple aspect trajectory data analysis: research challenges and opportunities. In *GeoInfo*, pages 56–67.

[Furtado et al. 2018] Furtado, A. S., Alvares, L. O. C., Pelekis, N., Theodoridis, Y., and Bogorny, V. (2018). Unveiling movement uncertainty for robust trajectory similarity analysis. *International Journal of Geographical Information Science*, 32(1):140–168.

[Furtado et al. 2015] Furtado, A. S., Kopanaki, D., Alvares, L. O., and Bogorny, V. (2015). Multidimensional similarity measuring for semantic trajectories. *Transactions in GIS*, 20:280–298.

[Gao et al. 2017] Gao, Q., Zhou, F., Zhang, K., Trajcevski, G., Luo, X., and Zhang, F. (2017). Identifying human mobility via trajectory embeddings. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 1689–1695. AAAI Press.

[Manning et al. 2008] Manning, C. D., Raghavan, P., Schütze, H., et al. (2008). *Introduction to information retrieval*, volume 1. Cambridge university press Cambridge.

[Montgomery 2017] Montgomery, D. C. (2017). *Design and analysis of experiments*. John wiley & sons.

[Noël et al. 2015] Noël, D., Villanova-Oliver, M., Gensel, J., and Quéau, P. L. (2015). Modeling semantic trajectories including multiple viewpoints and explanatory factors: application to life trajectories. In *Proceedings of the 1st International ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics, UrbanGIS@SIGSPATIAL 2015, Bellevue, WA, USA, November 3-6, 2015*, pages 107–113. ACM.

[Shokoohi-Yekta et al. 2017] Shokoohi-Yekta, M., Hu, B., Jin, H., Wang, J., and Keogh, E. (2017). Generalizing dtw to the multi-dimensional case requires an adaptive approach. *Data mining and knowledge discovery*, 31(1):1–31.

[Spaccapietra et al. 2008] Spaccapietra, S., Parent, C., Damiani, M. L., de Macedo, J. A., Porto, F., and Vangenot, C. (2008). A conceptual view on trajectories. *Data & knowledge engineering*, 65(1):126–146.

[ten Holt et al. 2007] ten Holt, G. A., Reinders, M. J. T., and Hendriks, E. A. (2007). Multi-dimensional dynamic time warping for gesture recognition. In *Thirteenth annual conference of the Advanced School for Computing and Imaging*, Heijen.

[Vlachos et al. 2002] Vlachos, M., Kollios, G., and Gunopulos, D. (2002). Discovering similar multidimensional trajectories. In *ICDE '02 Proceedings of the 18th International Conference on Data Engineering*, pages 673–684, Washington, DC. IEEE.

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