International Journal of Geographical Information Science
Vol. 00, No. 00, Month 200x, 1–22

THIS IS A COPY OF THE POST-PRINT MANUSCRIPT.

THE PUBLISHED VERSION:

Prelipcean, A.C., Gidófalvi, G. and Susilo, Y.O. (2016) Measures of transport mode segmentation of trajectories. *International Journal of Geographical Information Science*, ahead-of-print, 1-22 (10.1080/13658816.2015.1137297), found at <http://www.tandfonline.com/doi/full/10.1080/13658816.2015.1137297>

Preferred citation:

Prelipcean, A.C., Gidófalvi, G. and Susilo, Y.O. (2016) Measures of transport mode segmentation of trajectories. *International Journal of Geographical Information Science*, ahead-of-print, 1-22 (DOI: 10.1080/13658816.2015.1137297)

RESEARCH ARTICLE

Measures of transport mode segmentation of trajectories

Adrian C. Prelipcean^{a*}, Győző Gidófalvi^a, and Yusak O. Susilo^b

^a*Division of Geoinformatics, Department of Urban Planning and Environment, KTH, Sweden;* ^b*Division of Transport and Location Analysis, Department of Transport Science, Sweden*

(Received 00 Month 200x; final version received 00 Month 200x)

Rooted in the philosophy of point- and segment-based approaches for transportation mode segmentation of trajectories, the measures that researchers have adopted to evaluate the quality of the results (1) are incomparable across approaches, hence slowing the progress in the field, and (2) do not provide insight about the quality of the *continuous transportation mode segmentation*. To address these problems, this paper proposes new error measures that can be applied to measure how well a continuous transportation mode segmentation model performs. The error measures introduced are based on aligning multiple inferred continuous intervals to ground truth intervals, and measure the cardinality of the alignment and the spatial and temporal discrepancy between the corresponding aligned segments. The utility of this new way of computing errors is shown by evaluating the segmentation of three generic transportation mode segmentation approaches (implicit, explicit–holistic and explicit–consensus-based transport mode segmentation), which can be implemented in a thick client architecture. Empirical evaluations on a large real-world dataset reveal the superiority of explicit–consensus-based transport mode segmentation, which can be attributed to the explicit modeling of segments and transitions that allows for a meaningful decomposition of the complex learning task.

Keywords: Continuous Model Evaluation; Transportation Mode Segmentation and Detection; Trajectory Data Mining; Error Analysis; Interval Algebra;

1. Introduction

Communication and computing technologies together with location-, orientation-, and motion sensors in smartphones enable the large-scale collection of movements of individuals. Motivated by the steadily decreasing response rate of travelers to traditional declarative

*Corresponding author. Email: acpr@kth.se

pen-and-paper activity-travel surveys (Axhausen et al. 2003, Ogle et al. 2005, Stopher et al. 2008), one promising use of this information lies in the automatic generation of activity-travel diaries (Wolf 2000). An activity-travel diary is a verbose description of what the user did during a time frame, namely where did the user go (destination), why did the user go there (purpose) and how did the user travel there (travel modes). This paper focuses on the far from trivial task of assessing the quality of a travel mode detection method.

The travel mode detection literature so far can be divided into two main approaches: (1) a point-based approach, where an algorithm analyzes different dimensions of a location and proposes a transportation mode close to real time, and (2) a segment-based approach where an algorithm first detects transition points (i.e., the points where a user changes from one transportation mode to another) and then proposes a transportation mode for the entire segment in post-processing stages (see Section 2 for references). Point-based approaches cater for Location-Based Services and answer the question “How is the user travelling now?” and compute *precision and recall at a point level*. Segment-based approaches cater for transport sciences and answer the question “Given a set of locations grouped into segments, which segments are accurately detected and what transportation mode is associated with the accurately detected segments?” by first matching proposed segments to ground truth, and then computing different error *measures on the matched segments*. This paper argues and demonstrates that neither of the evaluations of the two approaches can guarantee that they provide an adequate answer to the transportation mode segmentation question in the context of activity-travel diaries, which is “How was the user traveling during a *whole* time frame?”. The performance measurements associated with the two approaches offer indicators based on *discrete* entity comparisons instead of evaluating the *continuous* output, as a whole, which makes the task of choosing between any two or more transportation mode segmentation methods next to impossible.

To address the aforementioned problem, this paper proposes new error measures that can be applied to any continuous dataset and segmentation method to measure how well a model performs. The proposed error measures are instantiated in the context of automatic activity-travel survey generation to assess how well different *continuous* transportation mode segmentation models perform. The introduced error measures are based on aligning multiple inferred continuous intervals to ground truth intervals, and measure the cardinality of the alignment and the spatial and temporal discrepancy between the corresponding aligned trajectory segments. The utility of this new way of computing errors is shown by evaluating the segmentation of three generic transportation mode segmentation approaches: *implicit*, *explicit-holistic* and *explicit-consensus-based* transport mode segmentation. In the first approach, a point-based transportation mode classification followed by a low-pass filtering that removes noisy classifications produces *implicit* transportation mode segments. In the second approach, an *explicit* detection of segments of GPS points that have similar movement characteristics is performed and the detected segments are classified based on *holistic* segment aggregates. In the third approach, a *consensus-based* transportation mode classification is performed on each point that falls within the *explicitly* detected segments. The generic segmentation approaches can be implemented in a scalable, distributed, stream processing architecture where the mobile devices with powerful computing capabilities process the high volume measurement streams in an online fashion. The proposed alignment method allows the definition of novel error measures and their evaluations for the three generic transportation mode segmentation approaches, which constitute the main contributions of this paper.

The remainder of the paper is organized as follows. Section 2 reviews related work. Section 3 provides necessary preliminaries and formalism. Section 4 presents the issues

of traditional performance measures, introduces the notion of segment alignment, the errors associated with the alignment, and introduces performance measures suitable for transportation mode segmentation evaluation. Section 5 presents the empirical evaluations of the proposed methods. Finally, Section 6 discusses the findings and Section 7 concludes and presents future work.

2. Literature Review

This section reviews the contributions of previous research in the field of trajectory segmentation and of transportation mode detection, with a focus on precision and recall.

2.1. *Semantic segmentation of trajectories*

As Parent et al. (2013) noted, trajectory segmentation depends on the application of the services for which it is performed. For many applications trajectories can be split into periods when the object is stationary, i.e., stops, and periods where the object is moving, i.e., moves. This type of trajectory segmentation is commonly known as stops and moves.

While useful for trajectory semantics, the use of this type of semantic trajectory segmentation in transportation science is limited to identifying where trips start and end (Ashbrook and Starner 2003), and the lack of ground truth data (Alvares et al. 2007, Palma et al. 2008, Rocha et al. 2010, Krum and Horvitz 2006), the dependency on external POI datasets (Alvares et al. 2007, Xie et al. 2009), the reliability on scientists' subjective judgment and the lack of automated methods for the assessment of detection accuracy (Andrienko et al. 2007, Yan et al. 2011, Xie et al. 2009) makes these concepts difficult to use in other fields. Further information that lies in the extraction of segments of same transportation modes cannot be accurately extracted from such a segmentation because a traveler can switch transportation modes without stopping. This paper investigates methods that segment trajectories into transportation modes, as well as trips, and presents a methodology to meaningfully compare the performance of the different methods.

2.2. *Transportation mode segmentation of trajectories*

Two different branches of science have focused on proposing methods that automate the segmentation of trajectories into same mode segments, namely, location based services (LBSes), which investigated point-based transportation segmentation (i.e., close to real-time, for each location) for enriching applications and services with information derived from knowing a user's transportation mode, and transportation science, which investigated segment-based transportation segmentation (i.e., post processing, for a batch of locations) as an effort of automating activity-travel diary generation (Ogle et al. 2005).

2.2.1. *Point-based transportation mode segmentation*

As point-based methods focus on providing close to real-time mode detection, most of them rely on data provided by the GPS receiver (Stenneth et al. 2011), the accelerometer sensor (Hemminki et al. 2013, Yu et al. 2014) or a combination of the two (Reddy et al. 2010, Manzoni et al. 2010, Shah et al. 2014, Prelipcean et al. 2014). The authors that use point-based mode detection methods report the algorithms' performance via traditional precision and recall measurements, which is sufficient if no further information is to be derived from the measurements, such as the time spent by a user while traveling via a

certain mode, or the traveled distance via a mode. The reported precision values vary based on the number of classified transportation modes and the method employed: 80.1% using AdaBoost together with decision tree classifiers for seven modes (Hemminki et al. 2013), 92.5% using support vector machine classifiers for five modes (Yu et al. 2014), 93.7% using random forest classifiers for six modes (Stenneth et al. 2011), 93.6% using a discrete hidden Markov classifier for five modes (Reddy et al. 2010), 90.8% using random forest classifiers for seven modes (Prelipcean et al. 2014), 82.1% using decision tree classifiers for eight modes (Manzoni et al. 2010), and 90% using multi-hierarchy decision tree classifiers for four modes (Shah et al. 2014).

While the point-based mode detection approach is suitable for real time applications, the collected data are usually not suitable for further analysis because the penalty assigned to an error is independent of the position of the error within the sequence as well as the spatial and temporal relations between the measurements in the sequence. Consequently, since all errors are equal when computing precision and recall, there is no trivial way of knowing how any model would perform if the task would be to segment a trajectory into same-mode segments. This research proposes a new methodology that penalizes errors based on the spatio-temporal proximity to ground truth and investigates how a generic point-based mode detection approach performs the transport mode segmentation task.

2.2.2. Segment-based transportation mode segmentation

The segment-based mode detection approach is usually attributed to the automation of activity-travel diary collection. Previous research investigated the steps needed to automate the collection (Axhausen et al. 2003), the suitable thresholds for stop detection (Shen and Stopher 2013) and identifying transitions (Das et al. 2014). For the segmentation of trajectories into same transportation mode segments, the research has been linear and built mostly on top of the approach proposed by Chung and Shalaby (2005), which identifies mode changing points by using a heuristic rule (the first record with a speed faster than 10km/h and a distance larger than 150 meters from the previous point). The reported precision values vary based on the number of classified transportation modes and the method employed: 91.7% using decision trees for four modes (Chung and Shalaby 2005), 94% using fuzzy logic for four modes (Tsui and Shalaby 2006), 95% using multi-hierarchy decision trees for five modes (Stopher et al. 2008), 83% using fuzzy logic for five modes (Schüssler et al. 2011), 85.8% using random forests for five modes, 70% using rule-based classifiers for five modes, 91.6% using membership functions for ten modes (Biljecki et al. 2013), and 71.5% using decision trees for four modes (Zheng et al. 2010).

Discussions regarding the segmentation performance are seldom and brief: Schüssler and Axhausen (2009) do not have labeled data and compare aggregates derived from the studied classification with data available from a microcensus, Schüssler et al. (2011) report a 68% segmentation accuracy by performing a temporal match with a 45 seconds buffer and continue performing the transportation mode segmentation on the original segments instead of the inferred ones, and Biljecki et al. (2013) remap the transportation modes present in two datasets to a common transportation mode set and do not discuss the implications of the computed precision and recall. Finally, Zheng et al. (2010) define the following measures for the performance of transportation mode detection: *accuracy by segment* computed as the number of segments correctly predicted divided by the total number of segments, and *accuracy by distance* computed as the distance covered by the correctly inferred segments divided by the total traveled distance. The authors use a 150 meters buffer to match the inferred segments with the ground truth segments. The authors define precision and recall values for change point detection, and emphasize that even

though the recall has higher priority in this case, a balance between precision and recall should be kept. Even though this approach is thorough in terms of accuracy assessment, its output is discontinuous, i.e., the statistics are presented only for the inferred segments that are matched to ground truth segments, and all errors are treated as equal. The approach presented in this paper analyzes the errors of a model that produces a continuous output and penalizes miss-match errors according to spatio-temporal proximity to ground truth.

Since scientists focused on different methods for the automatic trajectory segmentation and mode detection, they reported the performance of the proposed algorithms using traditional metrics that do not cover all dimensions of error associated with continuous intervals and with the error propagation due to multi-steps approaches. This paper proposes a new way to measure the impact of errors on continuous intervals that can be used to understand the differences, at various levels of analysis, between classification methods. This analysis is proven by comparing the performance of three generic approaches: implicit, explicit-holistic and explicit-consensus-based transport mode segmentation.

3. Preliminaries

This section provides the necessary preliminaries, formalism and terminology used in the remainder of this paper. This paper uses the definitions and notations that the authors previously proposed in Prelicean et al. (2014) slightly modified so that the moving objects are *users* that are traveling with *transportation means*.

Let $M = \{m_1, m_2, \dots, m_k\}$ be an exhaustive set of possible *transportation modes*, and $\{m_t, m_s\}$ be two *mobility states* transition and stop. Furthermore, let a *segmented trajectory* be defined by a sequence of labeled locations $L = \langle l_1, l_2, \dots, l_n \rangle$, where $l_i = (x, y, t, m)$. Let l_i denote the i -th location in L , $l_i.x$ and $l_i.y$ represent its coordinates, $l_i.t$ denote the time at which it was recorded, and $l_i.m$ represent its transportation mode, which belongs to $M \cup \{m_t, m_s\}$.

Let a *tripleg* be defined by the maximal length time period $[s; e]$, where s and e represent the start and stop time of the triplex, respectively, that is traveled with a single mode and overlaps a sequence of locations, as shown in Equation 1.

$$TL^m = \langle l_i^{x,y,t,m} : l_{i+1}.t > l_i.t \wedge l_i.t \in [s; e] \wedge l_i.m \in M \wedge l_i.m = l_{i+1}.m \rangle \quad (1)$$

Let a *trip* be defined as the maximal length sequence of triplets en route to a destination, where between every two triplets TL_i and TL_{i+1} whose modes belong to M , there is a period W_i^{i+1} in a transition state, m_t , as shown in Equation 2. W_i^{i+1} represents the period associated to the waiting time between the two modes and can be of 0 length.

$$T = \langle TL_i^{m_i} : T.s = TL_0.s \wedge T.e = TL_n.e \wedge W_i^{i+1}.s = TL_i.e \wedge W_i^{i+1}.e = TL_{i+1}.s \rangle \quad (2)$$

Considering the aforementioned equations, a *mode-segmented trajectory* can be defined as a sequence of trips, where between every two trips T_i and T_{i+1} (formed of triplets and transition state periods W_i^{i+1} between triplets), there is a period WT_i^{i+1} whose in a stop state, m_s , as shown in Equation 3. WT_i^{i+1} represents the period associated to the time between two consecutive trips and can be of 0 length.

$$L = \langle T_i : T_{i+1}.s \geq T_i.e \wedge T_{i+1}.m = m_s \wedge WT_i^{i+1}.s = T_i.e \wedge WT_i^{i+1}.e = T_{i+1}.s \rangle \quad (3)$$

This section first presents the difficulties of using current error measures used for mode detection, then provides a brief discussion on the difficulties of associating errors to segments and intervals, and proposes new error measures to overcome these difficulties.

As discussed in Section 2, there are two main approaches for transportation mode detection: point-based and segment-based. The quality of the results of the respective approaches are evaluated based on the entities over which predictions are performed: points and segments. For point-based approaches the particular error measures used are: precision, which is defined as the percentage of correct positive predictions, and recall, which is defined as the percentage of the positive cases identified. The example on the left side of Figure 1, illustrates the limitation of these error measures of point-based approaches in the context of transportation mode segmentation. In particular, the ground truth contains two periods associated with two transportation modes, W - walking, and S - subway, by considering the results of two different point-based classification methods PB_1 and PB_2 . While the precision and recall values for both methods are equal, 60%, when the output is translated into segments, PB_1 generates 9 segments and PB_2 generates 3 segments. The critical drawback of points-based error measures is that they cannot explain how well the methods describe the travels of the user during the studied time frame, a continuous phenomenon, because they are impervious to modeling the error propagation when generating segments from the sequences of consecutive points of the same transportation mode.

In comparison, segment-based approach contains two steps: first, the trajectories are partitioned into segments (by detecting trips and triplegs), and, second, a classifier takes as input certain dimensions that are computed from the segments and infers the transportation mode. In general the quality of results of segment-based approaches are also evaluated according to the precision-recall framework, but there is a large variation on

which predicted segments are matched to which ground truth segment for the purposes of error definition. Most commonly, research uses a spatial or temporal buffer around the edges of the ground truth intervals and considers any inferred interval whose **both** edges fall within the buffer area as a match and report mode precision and recall on the matched segments. In an attempt to measure the quality of the match, Zheng et al. (2010) propose two additional measures: accuracy by segment, which represents the percentage of segments that were identified, and accuracy by distance, which represents the percentage of distance covered by correctly identified segments. In order to confine the error measure, Schüssler et al. (2011) consider the simplified task of detecting the transportation modes of segments that are *given a priori*. Regardless of which variant, as the error measures are calculated based on a subset of predictions or based on questionable a priori knowledge they provide little understanding about the real-life performance of the methods. The example on the right side of Figure 1, illustrates the limitation of the error measures of segment-based approaches. In particular, the outputs of three different segment based algorithms, SB_1 , SB_2 and SB_3 , and the outputs translated to segments of the two previously presented point-based methods, PB_1 and PB_2 are visualized in relation to the ground truth. Because a segment is only considered matched when both of its edges fall within the buffer, denoted in Figure 1 by the hatched area, none of the methods could be matched. As such, the five seemingly different methods are equal in terms of the precision and recall metrics that can be computed only on matched segments, which are 0% in this case. The critical drawback of segment-based error measures is that whenever the matched segments do not coincide with **all** ground truth segments, the continuity of the phenomenon is not respected by and reflected in the error measures.

4.2. Interval alignment method

As seen in the previous section, there is a problem with explaining error when comparing a set of ground truth segments with a set of inferred segments without breaking the continuity of the studied phenomenon. To overcome this limitation, this paper proposes the replacement of a buffer-based one-to-one matching with a more permissive way to find corresponding segments between two datasets, namely *interval alignment*¹.

The proposed method finds a correspondence between a set of inferred intervals L^{inf} and a set of ground truth intervals L^{GT} such that after the alignment every interval L_i^{inf} has a correspondence in the ground truth set so that $L_i^{inf}.s = L_j^{GT}.s$ and $L_i^{inf}.e = L_k^{GT}.e$, where $k \geq j$. The relationship between j and k is discussed in Section 4.2.1.

The details of the alignment process are explained in relation to the pseudocode in Algorithm 1. The `align(.)`-function is implemented recursively taking four arguments: the *ordered* list of inferred intervals L^{inf} , the *ordered* list of ground truth intervals L^{GT} , the index of the inferred interval that has to be matched idx_inf , and the index from which the alignment search starts in the ground truth intervals idx_gt . Each call to `align(.)` aligns the start and end of $L^{inf}[idx_inf]$ to the start of a ground truth interval from L^{gt} and the end of a *possibly different* ground truth interval from L^{gt} , advances the indexes to reflect the alignment, and makes a recursive call to align the remaining inferred intervals to ground truth intervals. $L^{inf}[idx_inf]$ is either aligned 1) to the start and end of the first ground truth interval that $L^{inf}[idx_inf]$ is within (lines 3-4) or 2) $L^{inf}[idx_inf]$'s start

¹The alignment method is presented for *intervals* of an *arbitrary dimension* but is subsequently applied on the time dimension where the intervals represent time *periods*.

(lines 8-13) and end (lines 15-20) are aligned to the respectively closest start and end of the ground truth intervals that succeed $L^{inf}[idx_inf].s$ and $L^{inf}[idx_inf].e$.

Since the recursive calls to `align(.)` collectively iterate through the lists at least once and in each iteration one inferred interval is aligned in a constant number of operations, `align(.)` has a complexity of $\Theta(M + N)$, where M and N are the lengths of L^{inf} and L^{GT} .

Algorithm 1 `align( $L^{inf}, L^{GT}, idx\_inf, idx\_gt$ )`

Require: ordered L^{inf}, L^{GT}

```

1: aligned_s  $\leftarrow$  aligned_e  $\leftarrow$  false;
2: for  $i = idx\_gt$  to  $L^{GT}.len$  do
3:   if  $L^{inf}[idx\_inf]$  is within  $L^{GT}[i]$  then
4:     emit_alignment( $L^{inf}[idx\_inf]$ ,  $L^{GT}[i]$ )
5:     align( $L^{inf}, L^{GT}, idx\_inf + 1, i$ )
6:   else
7:     if not aligned_s then
8:       if  $L^{GT}[i].s \geq L^{inf}[idx\_inf].s$  then
9:         aligned_s  $\leftarrow$  true
10:        if  $|L^{GT}[i].s - L^{inf}[idx\_inf].s| \geq |L^{GT}[i + 1].s - L^{inf}[idx\_inf].s|$  then
11:          idx_s  $\leftarrow$   $i$ 
12:        else
13:          idx_s  $\leftarrow$   $i + 1$ 
14:      else if not aligned_e then
15:        if  $L^{GT}[i].e \geq L^{inf}[idx\_inf].e$  then
16:          aligned_e  $\leftarrow$  true
17:          if  $|L^{GT}[i].e - L^{inf}[idx\_inf].e| \geq |L^{GT}[i + 1].e - L^{inf}[idx\_inf].e|$  then
18:            idx_e  $\leftarrow$   $i$ 
19:          else
20:            idx_e  $\leftarrow$   $i + 1$ 
21:      if aligned_s and aligned_e then
22:        emit_alignment( $L^{inf}[idx\_inf]$ , [ $L^{GT}[idx\_s].s, L^{GT}[idx\_e].e$ ])
23:        align( $L^{inf}, L^{GT}, idx\_inf + 1, idx\_s$ )

```

After the alignment is performed, there are two issues that, when considered, can be used for proposing error measures: the cardinality of alignment is not always one-to-one (e.g., multiple inferred intervals can be aligned to one ground truth interval, i.e., oversegmentation, or one ground truth interval is not aligned to any of the inferred intervals, i.e., undersegmentation) and there is a difference between the inferred and the ground truth interval edges (e.g., $L^{inf}[idx_inf].s$ is aligned to $L^{GT}[i].s$, but $L^{inf}[idx_inf].s \neq L^{GT}[i].s$). These discrepancies represent penalties, which are discussed in the following subsections.

4.2.1. Cardinality penalty

This section studies the error by analyzing the cardinality of the alignment of inferred periods to ground truth periods. To do so, based on the visionary work of Allen (1983) who proposes 13 relationships between two intervals, the present work defines four relationships that are illustrated in Figure 2 and are suitable for the current setting as follows:

- (1) A direct match occurs when one inferred period is aligned to one ground truth period and vice versa (one-to-one join) – Figure 2 (a).
- (2) An oversegmentation occurs when multiple inferred period are aligned to one ground truth period (many-to-one join), which is more important for some applications than for others – Figure 2 (b). For example, if transportation modes for

Figure 2. The four cardinality relationships derived from Allen (1983).

periods are correctly inferred then in some cases consecutive same-mode periods can automatically be merged, whereas if the correctness of mode inferences are uncertain then tedious manual merging is required.

- (3) A miss occurs when there is no information regarding an inferred period or its edges (one-to-many join) – Figure 2 (c). Any miss is accompanied by two one-handed joins since instead of splitting one period into “n” segments, the algorithm detected only one segment, which contains edge information on the first ground truth period and the last ground truth period within the period.
- (4) A one-handed join is a special case of a miss in which the start of an inferred period is aligned to the start of a ground truth period and the end of the inferred period is aligned to the end of any proceeding ground truth period (one-to-many join) – Figure 2 (d). This is a special case because the inferred period merges the information between two or more consecutive ground truth periods. The severity of merging ground truth periods is application specific and the one-handed join cardinality relationship can be used to identify these cases as misses or treated them separately to identify which types of periods are hard to distinguish.

4.2.2. Space and time penalties

Any tripleg TL_i has a temporal (T_i) and a spatial (S_i) dimension associated with it, which are used when computing the temporal and spatial penalties that are a consequence of the approximate alignment. Whenever a tripleg TL'_i is aligned to a tripleg TL_i and

the edges corresponding to the periods are different, e.g., $TL'_i.s \neq TL_i.s$, which is further referred to as *alignment disagreement window*, an alignment error is made. These errors have spatial and temporal components which are further referred to as *spatial* and *temporal penalties*, respectively. By analyzing the locations that were wrongfully moved in between the triplets within the alignment disagreement window, one can define the temporal penalty as the time that the edge of the triplet would have to be shifted in order to correct the mistake, and the spatial penalty as the distance that the edge of the triplet would have to be moved *across the consecutive points* within the alignment disagreement window in order to correct the mistake.

The temporal penalty is defined as the duration between all consecutive locations that were wrongfully moved to / from a ground truth triplet TL_i from / to the triplets whose periods overlap the start of the alignment disagreement window – between $TL_i.s$ and $TL_{i-j}.s$ – and the end of the alignment disagreement window – between $TL_i.e$ and $TL_{i+k}.e$ – when an inferred triplet TL'_i is aligned to TL_i . The temporal penalty for the start of the triplet (the analogous formula is valid for the end of the triplet) is calculated with the formula presented in Equation 4 and it is sensitive to negative values.

$$t_p_{TL'_i.s}^{TL_i.s} = TL'_i.s - TL_i.s \quad (4)$$

Similar to the temporal penalty, the spatial penalty is defined as the distance between all the consecutive locations that are wrongfully moved to / from a ground truth triplet TL_i from / to the triplets whose periods overlap the start of the alignment disagreement window – between $TL_i.s$ and $TL_{i-j}.s$ – and the end of the alignment disagreement window – between $TL_i.e$ and $TL_{i+k}.e$ – when an inferred triplet TL'_i is aligned to TL_i . The spatial penalty of the start of the triplet (the analogous formula is valid for the end of the triplet) is calculated with the formula presented in Equation 5, where k is the index of the locations within the given interval, and its values should be analyzed with regards to the temporal penalty values to understand whether the segment has a greater or smaller length.

$$s_p_{TL'_i.s}^{TL_i.s} = \sum_{l_k.t \in (TL'_i.s, TL_i.s]} dist(l_k.(x, y), l_{k+1}.(x, y)) \quad (5)$$

Analyzing the formulas in Equations 4 and 5, the following observations can be made about $t_p_{TL'_i.s}^{TL_i.s}$, also referred to as $t_p.s$, and about $s_p_{TL'_i.s}^{TL_i.s}$ also referred to as $s_p.s$:

- (1) when $t_p.s = 0$, the edges $TL_i.s$ and $TL'_i.s$ were perfectly matched,
- (2) when $t_p.s > 0$, all the locations between $TL_i.s$ and $TL'_i.s$ were wrongfully *moved* from TL_i to the preceding triplets within the alignment disagreement window TL_{i-j} during the alignment and the segment's length *decreased* by $s_p.s$, and
- (3) when $t_p.s < 0$, all the locations between $TL_i.s$ and $TL'_i.s$ were wrongfully *moved* from the preceding triplets within the alignment disagreement window TL_{i-j} to TL_i during the alignment and the segment's length *increased* by $s_p.s$.

One important aspect of this alignment is that the spatial and temporal penalties for the start of a given segment TL_i are the opposites of the spatial and temporal penalties for the end of its preceding segment, i.e., $t_p_{TL'_i.s}^{TL_i.s} = -1 * t_p_{TL_{i-1}.e}^{TL'_{i-1}.e}$ and $s_p_{TL'_i.s}^{TL_i.s} = -1 * s_p_{TL_{i-1}.e}^{TL'_{i-1}.e}$.

Based on the description of the proposed interval alignment method and the space and time penalties, one can easily see a resemblance to string sequence alignment and similarity measures (Levenshtein 1966, Wagner and Fischer 1974) and their extensions to trajectory similarity measures (Chen and Ng 2004, Chen et al. 2005). However, a quick

examination reveals that this similarity is only on the surface and the proposed methods and penalties are fundamentally different from these Edit Distance (ED) based alignment methods and similarity measures. Most importantly, the proposed penalties 1) measure the spatial and temporal dimensions of segments that are within the alignment disagreement window (not the number of edit operations) and 2) are a relatively complex spatial and temporal function of the locations that are within the alignment disagreement window (not a constant function of edit operations).

4.3. Generic transportation mode segmentation approaches as proof of concept

This section describes three different approaches to transportation mode segmentation and proposes new performance measures for assessing such algorithms.

4.3.1. Tried generic transportation mode segmentation approaches

As previously discussed in Section 2, there are two main approaches of deriving transportation mode: a point-based approach, and a segment-based approach, either of which has its own performance measures. This paper uses different prototypical approaches associated with each of the two approaches and proposes a novel, hybrid approach, which uses information specific to both point- and segment-based approaches for transportation mode segmentation. While, not contrary to the “no free lunch” theorems for machine learning (Wolpert 1996), different machine learning methods for different derived features for different transportation modes on different datasets yield different quality results (see Section 2), the choice of evaluation of generic approaches is motivated by the conjecture that the differences that result from how these generic approaches decompose the task dominate the aforementioned differences and are therefore of key interest. The proposed generic approaches, together with a brief description of the mechanisms that underlie the approaches are presented in the following sections.

4.3.1.1. Implicit point-based transportation mode segmentation. The first approach is a point-based one, where the points are classified one-by-one and the transportation segment are the implicit results of the classification, namely all the consecutive, same transportation mode points are grouped into segments after the classification is performed (see Equation 1). As in any classification task, the chosen dimensions along which the classification is performed affect the outcome and should be carefully chosen. In the context of this paper, the authors use the same methodology proposed in previous research (Prelicean et al. 2014), with two main differences: 1) the user id is *not* a dimension of the classification, and 2) the transportation modes used in this research are different from the previous one. Similar to previous research, the current approach also uses a random forest classification method, accompanied by 10-fold cross-validation for model validation. Since the output of such an approach is sensitive to local, non-sequential errors, such as spuriously inserting an incorrect mode within a sequence of correct modes, this paper compensates for this drawback by smoothing the raw classifications using a five element long majority filter.

4.3.1.2. Explicit holistic segment-based transportation mode segmentation. The second approach focuses on first detecting the segments that have the same movement characteristics, and then classifies each segment into a transportation mode based on

holistically computed dimensions at the segment-level. Since there are multiple ways to define what a segment is or how a segment can be detected, this paper proposes three distinct segment detection methods:

- (1) **stop-oriented segment detection:** a method that is finely tuned for detecting stops, makes use of a traditional heuristic “if a user’s speed is less than 3.6km/h for more than 2 minutes, then that represents a dwell period”. Furthermore, the method merges the segments that are wrongfully detected due to noise into the neighboring triplets. A segment is labeled as “deleted due to noise” when the cumulative distance between the centers of consecutive lines formed between consecutive locations is significantly different from the cumulative distance between consecutive locations, and when the polyline formed between all the consecutive points self-intersects more than three times.
- (2) **transition-oriented segment detection:** a method that is sensitive to the fluctuations in speed and accelerometer values that are associated with transitions, and detects a point as a *period boundary* when the accelerometer and speed differences between consecutive locations are greater than a threshold value, and when the projecting distance (defined next) is greater than a threshold value. The *projecting distance* of a given location l_i is defined by two entities: a point l_i^{past} is generated by projecting the distance that could be traveled from l_{i-1} along bearing l_{i-1}^{i-1} using the average speed between l_{i-2} and l_{i-1} in the time frame between l_{i-1} and l_i , and a point l_i^{future} is generated by projecting the distance that could be traveled from l_{i+1} along bearing l_{i+1}^i using the average speed between l_{i+2} and l_{i+1} in the time frame between l_{i+2} and l_{i+1} . The projecting distance is computed as the maximum distance between l_i^{past} , l_i^{future} and l_i .
- (3) **a hybrid segment detection:** a method that first uses the stop-oriented segment detection algorithm, extracts the locations that are within segments that have high standard deviation values for speed and accelerometer, and applies the transition-oriented segment detection algorithm on the extracted consecutive locations.

A traditional 10-fold cross-validation method should not be used to train a classifier, since the inferred segments are different than the ground truth segments, and, as such, the approach considered in the current research uses a leave-one-user-out approach. The dimensions used for the classification of segments into modes are the following: number of locations within the segment, the duration that spans the segment, the average speed, the average duration between consecutive points, the length of the segment, the average, minimum, standard deviation, and median of 1) the number of steps between two consecutive locations and 2) the accelerometer values (Montini et al. 2014, Prelipcean et al. 2014).

4.3.1.3. Explicit consensual point-based transportation mode segmentation.

The last approach combines the insights of both point- and segment-based approaches by using the sliding windows that are in the point-based mode detection and the segments present in the segment-based mode detection. In particular, in the segment detection phase, the window-size is fixed to a predefined value (of five locations), in the transportation segment mode classification phase the window coincides with the segment that contains a point that will be filtered, thus allowing every point inside a segment to vote for the transportation mode of the segment.

4.3.2. Transportation mode segmentation performance metrics

After implementing all the methods, it is important to have accurate metrics that indeed reflect on the performance of each method. As such, the traditional precision and recall metrics are not suitable because they do not capture the essence of what error is in the context of transportation mode segmentation.

Since the proposed methodology makes use of segment alignment to find the correspondence between inferred segments and ground truth segments, the following five evaluation metrics that make use of the segment alignment penalties and segment alignment cardinality are proposed:

- (1) *precision*: the percentage of *correctly* inferred segments that are aligned to a ground truth segment.
- (2) *recall*: the *minimum* percentage of length / duration explained by correctly inferred segments divided by the total length / duration of the ground truth segments.
- (3) *shift-in penalty*: the penalty of moving every missed ground truth segment inside the inferred segment so that all the ground truth segments occupy the same position in the inferred sequence as in the ground truth. The shift-in penalty has a penalty in the dimension of time and one in the dimension of space.
- (4) *shift-out penalty*: the penalty of moving *each* of the incorrect segments – i.e., the segments that do not have the same value for the inferred transportation mode as the ground truth transportation mode – outside of the inferred segment when *more than one* inferred segments are aligned to one ground truth segment.
- (5) *oversegmentation*: the number of inferred segments a ground truth segment is split by the tested method / algorithm.

5. Empirical evaluations

This section presents the analysis of the performance of different transportation segmentation methods when applied to a real-world GPS-based activity-travel survey dataset. The data set was collected during a trial between 29th of September 2014 (Monday) to 5th of October 2014 (Sunday) where 26 users installed the MEILI Mobility Collector (Prelipcean et al. 2014) and collected 66107 GPS points fused with accelerometer summaries in the region of the Stockholm County, Sweden. The users then annotated the collected data via a web application by specifying trip ends, trip purposes, triplets and transportation modes, which resulted in 1307 triplets (38% stop, 28% walk, 12% bicycle, 0.4% moped, 7.6% car, 5% bus, 9% subway, 0.2% tram, 2% train). An overview of the employed segmentation methods is presented in Table 1. In this section, the performance metrics proposed in this paper are compared with the traditional metrics for generic implicit and explicit-holistic transportation mode segmentation methods in Section 5.1, and the performance boost of using an explicit-consensus method is shown in Section 5.2.

5.1. Implicit and explicit-holistic methods

5.1.1. Performance metrics

First, it is important to note what traditional precision and recall values are associated with P_B and with S_B . The performance measures reported for P_B by using traditional metrics are: **64.5%** precision and **65.1%** recall. The performance measures for the S_B

Table 1. The empirical evaluations are performed on the three main segmentation methodologies: *implicit*, *explicit*, and *consensus-based*, out of which the explicit and consensus-based segmentation methods are subsequently split into based on their focus into: stop (*stop*), transition (*tr*), and hybrid (*hybrid*).

P_B	implicit segmentation method based on point-based classifiers
S_B^{stop}	explicit segmentation method that focuses on stop detection
S_B^{tr}	explicit segmentation method that focuses on transition detection
S_B^{hybrid}	explicit segmentation method that combines S_B^{stop} and S_B^{tr}
PS^{stop}	consensus-based S_B^{stop} segmentation superimposed to the P_B classification
PS^{tr}	consensus-based S_B^{tr} segmentation superimposed to the P_B classification
PS^{hybrid}	consensus-based S_B^{hybrid} segmentation superimposed to the P_B classification

Table 2. Precision and recall values for the point-based and segment-based methods. The instances in bold represent the highest recorded values for the attribute.

	Precision				Recall			
	P_B	S_B^{stop}	S_B^{tr}	S_B^{hybrid}	P_B	S_B^{stop}	S_B^{tr}	S_B^{hybrid}
Stop	45.7%	85.0%	47.6%	69.1%	6.0%	65.6%	50.6%	65.4%
Walking	44.0%	51.7%	42.9%	61.6%	10.9%	45.2%	37.3%	64.8%
Bicycle	41.8%	58.1%	41.1%	75.6%	25.6%	39.4%	33.5%	71.0%
Moped	38.9%	0.0%	30.0%	25.0%	20.7%	0.0%	7.1%	6.9%
Car	35.8%	30.4%	22.5%	39.0%	26.5%	25.9%	16.6%	27.4%
Bus	13.3%	45.8%	30.4%	46.2%	20.6%	31.1%	22.9%	44.6%
Subway	9.6%	38.9%	19.0%	47.5%	7.5%	21.4%	15.0%	37.8%
Tram	0.0%	0.0%	0.0%	0.0%	4.9%	0.0%	0.0%	0.0%
Train	6.2%	27.6%	20.3%	36.0%	0.0%	13.3%	12.9%	27.5%
Avg.	26.1%	37.5%	28.2%	44.4%	13.4%	26.9%	21.8%	38.4%

methods report a precision of **80.1%** and a recall of **82.4%** (computed by detecting the transportation modes of *ground truth segments*).

Using the evaluation measures proposed in Section 4.3.2, the new precision and recall values are *up to 40% lower* than both P_B and S_B specific precision and recall values, as shown in Table 2. From the three proposed methods, S_B^{stop} suggests the most stop segments that are correctly aligned to “stop” ground truth segments, followed by S_B^{hybrid} , which even though has problems with the over-segmentation of “stop” segments, it correctly captures a significant percentage of duration / length of ground truth stop segments. For most other modes, S_B^{hybrid} records the highest values for precision and recall, except for moped, where P_B outperforms it. It is noticeable that P_B records motorized transportation modes such as car or moped well, which can be explained by the discrepancy between the number of segments traveled by an infrequent mode and the number of points within those segments (P_B has more samples to learn from the characteristics of infrequent modes than the any of the S_B methods).

5.1.2. A detailed analysis of S_B^{hybrid}

The method with the highest overall precision and recall values, S_B^{hybrid} , was chosen to be analyzed with regards to the proposed performance measures (Table 3). As expected, the spatial shift-in penalty for the “stop” and “walk” modes is small because of the high recall values and of the travel speed of these modes –low speed (walking) or no speed (stop). The transportation modes with low precision and recall values (train, moped or tram) have a large value for the shift-in spatial and temporal penalty, which implies that S_B^{hybrid} either over-segments or fails to segments the triplegs traveled by these modes. Analyzing

Table 3. The proposed performance measures computed for S_B^{hybrid} . The shift-in penalties indicate that the motorized modes are more often confused with one another and that the non-motorized modes have a higher temporal shift-in penalty, which suggest that stop triplegs overlap with walk triplegs. The shift-out penalty shows that oversegmentation is most critical for bus, car and train.

	Prec.	Rec.	Shift-in penalty		Shift-out penalty		Over(#)
			Space(m)	Time(min)	Space(m)	Time(min)	
Stop	69.1	65.4	60±750	3.5±40	2250±15500	51±380	3
Walk	61.6	64.8	50±200	2.5±30	8500±32000	112±750	3
Bicycle	75.6	71.0	150±750	1±4	2000±3800	51±175	4
Moped	25.0	6.9	2150±3150	3.5±5	7500±4750	323±552	3
Car	39.0	27.4	1000±3000	8.5±40	14000±29000	142±380	4
Bus	46.2	44.6	670±1850	2±7	15000±38000	54±115	4
Subway	47.5	37.8	750±2000	3±7	10000±17000	132±411	3
Tram	0.0	0.0	4150±6000	25±37	21000±30000	85±115	0
Train	36.0	27.5	15000±48750	11±32	109000±185000	380±1464	3

the equivalent shift-out penalty, it is noticeable that the time penalty takes precedence over the space penalty for low speed modes, and the opposite is true for high speed modes. Walking segments are over-segmented in the middle of the segment since the spatial and temporal shift-out penalties have a *high* value. Bike segments are over-segmented but mostly at the beginning or at the end of a tripleg since the spatial and temporal shift-out penalties have a *low* value.

Table 4. Precision and recall values for the consensus-based methods. The instances in bold represent the highest recorded value for the attribute.

	PS^{stop}	Precision			Recall	
		PS^{tr}	PS^{hybrid}	PS^{stop}	PS^{tr}	PS^{hybrid}
Stop	91.3%	64.3%	81.2%	69.4%	58.1%	71.1%
Walk	44.0%	57.6%	57.3%	35.0%	43.2%	53.3%
Bicycle	61.0%	71.2%	78.0%	39.2%	47.2%	73.4%
Moped	25.0%	43.3%	45.0%	38.1%	40.5%	54.9%
Car	60.3%	54.9%	66.4%	54.5%	40.1%	62.5%
Bus	33.8%	47.9%	49.6%	17.5%	27.6%	36.2%
Subway	10.4%	33.5%	31.9%	3.8%	18.4%	15.9%
Tram	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Train	12.8%	15.9%	37.8%	5.8%	13.1%	20.6%
Avg.	37.6%	43.2%	49.7%	29.3%	32.0%	43.1%

5.2. Explicit-consensus methods

5.2.1. Performance metrics

Next, the intention was to observe the effectiveness of using the segments as a sliding window, allowing every point within it to vote for the transportation mode, which is presented in Table 4. First, both precision and recall values have increased for all modes. Second, even though the precision of PS^{stop} for “stop” segments is higher than that of PS^{hybrid} , the associated recall value of PS^{hybrid} is higher than that of PS^{stop} , which implies that PS^{hybrid} suggests fewer but more meaningful (with a longer duration) “stop” segments. Finally, PS^{tr} has higher overall precision and recall values than PS^{stop} , which illustrates the information gain from using explicit-consensus methods compared to holistic methods, in which case S_B^{stop} had higher overall accuracy and recall values than S_B^{tr} .

Table 5. Space and time shift-out penalties for S_B^{hybrid} , PS^{hybrid} and PS_{perf}^{hybrid} . The results indicates that part of the oversegmentation problems diminished for most high speed transportation modes.

	Space(m)			Time(min)		
	S_B^{hybrid}	PS^{hybrid}	PS_{perf}^{hybrid}	S_B^{hybrid}	PS^{hybrid}	PS_{perf}^{hybrid}
Stop	2250±15500	2300±16500	2400±15500	51±380	28±228	51±380
Walk	8500±32000	9300±30000	8500±32000	112±750	136±800	112±750
Bicycle	2000±3800	2500±6250	620±1500	51±175	60±200	60±200
Moped	7500±4750	4000±6000	4500±6500	323±552	316±555	380±600
Car	14000±29000	14500±72000	14000±76000	142±380	150±380	150±400
Bus	15000±38000	14000±33000	910±2000	54±115	56±113	20±90
Subway	10000±17000	17000±26000	10000±20000	132±411	138±400	150±450
Tram	21000±30000	21000±30000	0	85±115	85±115	0
Train	109000±185000	73000±93000	30000±60000	380±1464	380±1523	400±1600
Avg.	19000±36000	15000±31000	9400±30000	138±470	140±465	144±550

Table 6. Space and time shift-in penalties for S_B^{hybrid} , PS^{hybrid} and PS_{perf}^{hybrid} . The results clearly indicate the superiority of PS_{perf}^{hybrid} , where there has been a substantial decrease in the shift-in spatial and temporal penalties, which indicates that fewer transportation modes have been misclassified by the new method.

	Space(m)			Time(min)		
	S_B^{hybrid}	PS^{hybrid}	PS_{perf}^{hybrid}	S_B^{hybrid}	PS^{hybrid}	PS_{perf}^{hybrid}
Stop	60±750	55±750	60±750	3.5±40	3.5±43	3.5±40
Walk	50±200	75±250	50±200	2.5±30	3±30	2.5±30
Bicycle	150±750	140±680	30±200	1±4	1±3.5	0±2
Moped	2150±3150	1140±2500	1000±2500	3.5±5	2±4	1.5±4
Car	1000±3000	700±2600	360±1800	8.5±40	8±40.5	7.5±40
Bus	670±1850	650±1800	200±800	2±7	2±7	1±3
Subway	750±2000	1092±2500	500±1800	3±7	3.5±8	2±5.5
Tram	4150±6000	4150±6000	0	25±37	25±37	0
Train	15000±48750	14500±48500	670±2300	11±32	11±32	2.5±6.5
Avg.	2500±7400	2500±7300	350±1200	7±22	7±23	2±14.5

5.2.2. Information gain from using a user-specific point-based method

In order to understand the information gain of each method, this paper makes use of another point-based method, P_B^{perf} , that has been successfully used in previous research (Prelicean et al. 2014). This method takes into account the user id when inferring the transportation mode, which makes it more precise but less general, and has one major drawback: it does not consider stop points. As such, the stop periods proposed by each of the segmentation methods are used, which means that there will be no difference in the performance of stop detection. This point-based approach is a strong candidate for because of its traditional performance metrics: 94.4% precision and 94.5% recall.

The accuracy gain is discussed with relation to S_B^{hybrid} , its initial performance and the performance modification after using PS and PS_{mob} for consensus-based classification. As seen in Table 7, using the more precise point-based classifier gave all transportation modes except “stop” and “walk” a substantial precision and recall boost. The logic behind this performance boost can be observed in Table 6, where one can observe a substantial decrease in penalties for the following modes: bicycle, moped, care, bus, subway and train. This can be explained by the higher precision of the point-based classifier used for the consensus, which results in fewer misclassifications. When analyzing the shift-out penalties, it is noticeable that the new consensus performs better when it comes to space penalty, but with no visible difference for the temporal penalty, which implies that the classification corrected wrongfully inferred segments close to the middle of ground truth segments trav-

Table 7. Precision and recall value for S_B^{hybrid} , PS^{hybrid} and PS_{perf}^{hybrid} . The consensus obtained while using the points classified by the superior point-based method, P_B^{perf} , yield a high increase in overall precision and recall.

	S_B^{hybrid}	Precision			Recall	
		PS^{hybrid}	PS_{perf}^{hybrid}	S_B^{hybrid}	PS^{hybrid}	PS_{perf}^{hybrid}
Stop	69.1%	81.2%	69.1%	65.4%	71.1%	65.4%
Walk	61.6%	57.3%	60.6%	64.8%	53.3%	64.1%
Bicycle	75.6%	78.0%	89.6%	71.0%	73.4%	86.7%
Moped	25.0%	45.0%	70.0%	6.9%	54.9%	71.6%
Car	39.0%	66.4%	75.2%	27.4%	62.5%	77.2%
Bus	46.2%	49.6%	77.8%	44.6%	36.2%	75.6%
Subway	47.5%	31.9%	63.4%	37.8%	15.9%	58.8%
Tram	0.0%	0.0%	100.0%	0.0%	0.0%	100.0%
Train	36.0%	37.8%	71.9%	27.5%	20.6%	59.7%
Avg.	44.4%	49.7%	75.3%	38.4%	43.1%	73.2%

eled with high speed modes (e.g., train, subway, bus, car). However, as the shift-in penalty is also explained by misclassifications, one can expect that the spurious misclassification in the middle of segments are eliminated by a low-pass filter, but those that are closer to the edges are not as easy to eliminate. Investigating and explaining this property is planned for future research.

6. Discussion

The paper tackles a problem that is imperative to understand before getting deeper insights into human behavior, in general, and traveler behavior, in particular, namely understanding how to model transportation mode from sequences of spatio-temporal measurements about travelers. Common practice relies on using statistical packages that, based on a given statistical inference method, takes as input a trajectory dataset and, using predefined dimensions, infers the transportation mode with well established performance measures (e.g., precision, recall, etc). This paper challenges that the traditional and widely used performance measures do not adequately capture the essence of modeling continuous phenomena such as how users travel. As a way to overcome these issues, a new method that aligns *any* entity within a continuous dataset to an entity within a ground truth continuous dataset (while maintaining its continuity) is proposed together with novel performance measures finely tuned for continuous models. These concepts have been tested and evaluated on a real-world dataset and the results suggest that the traditional performance metrics do not adequately capture information that can be used to describe the performance of a model.

In particular, two distinct traditional performance measures (point-based and segment-based) computed for the dataset suggest that the classical precision and recall values do not capture the information needed to analyze models for continuous datasets. On one hand, discrete point-based performance measures are not propagated in the continuous space. On the other hand, the segment-based performance measures are applicable only in the second stage of a segmentation process, i.e., putting labels on segments, and do not embed the errors that are due to the first stage, i.e., the segmentation stage. The proposed methods investigate how the output of the tested methods translated into a continuous model reflects the reality. This shows that any error made in the incipient phases is propagated to the later phases and output at a greater magnitude, but this paper does not identify the dimensions that affect the magnitude.

One of the interesting results was that some segmentation methods, when used together

with a point-based method in a consensus-type of scenario, have a higher accuracy and / or recall gain than others. While some of the proposed performance metrics seem to be correlated with this accuracy gain, this paper does not go into more detail.

Finally, the authors do not intend to compare the results obtained for transportation mode segmentation in the case study of the present paper with results obtained by papers focused on the transportation mode detection tasks, but to draw attention towards the fact that widely used traditional errors do not offer a realistic description of the performance of the algorithms employed for this task.

7. Conclusions and Future Work

This paper proposes a novel approach to modeling and evaluating methods designed to study continuous phenomenon such as transportation mode segmentation. To circumvent the drawbacks of the classical methodology used to find the correspondence between a ground truth entity and an inferred entity, i.e., matching, this paper uses a variation of the time algebra alignment proposed by Allen (1983). The advantages of using alignment instead of matching are stated, discussed and explained on real data obtained from a case study where different generic transportation segmentation approaches were compared. Since the traditional measures for performance rely on the concept of matching inferred segments with ground truth segments and not on alignment, new performance measures are proposed. These measures capture how an error that is produced at any step preceding the transportation mode segmentation propagates down to the modeled output, which contrasts the traditional measures, which only count the number of “hits” and “misses”.

The definition of more robust error measures opens up new research opportunities such as understanding which classification dimensions are sensitive to error, identifying the critical error propagation channels, understanding why certain methods produce a drastically different output when a different dataset is considered, and, most importantly, allowing scientists to answer the question “How feasible is the automation of the transportation mode segmentation task?” while keeping in mind what it is that researchers / government authorities need when asking the question. Future studies aim at validating the proposed methodology over larger datasets and studying the critical dimensions that affect the magnitude of error while it is propagating through the classification stages.

Finally, while this paper presents the concept of error propagation with regards to space and time, this approach is not limited to the two dimensions. Complementary research objectives might involve performing similar analysis for other continuous phenomenon, e.g., studying how an individual’s perception of well being varies over time, or, how an individual’s mental map varies over time. To conclude, while models are built using discrete data, in the case of continuous phenomenon, it is important to understand the error in relation to the modeled phenomenon and not only in relation to the collected discrete data.

References

- Allen, J. F. 1983. Maintaining Knowledge About Temporal Intervals. In *Communications of the ACM*, vol. 26, no. 11, pp. 832–843.
- 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*, no. 22, pp. 22:1–22:8.
- Andrienko, G., Andrienko, N., and Wrobel, S. 2007. Visual Analytics Tools for Analysis of Movement Data. In *SIGKDD Exploration Newsletter*, vol. 9, no. 2, pp. 38–46.
- Ashbrook, D., and Starner, T. 2003. Using GPS to Learn Significant Locations and Predict Movement Across Multiple Users. In *Personal Ubiquitous Computing*, vol. 7, no. 5, pp. 275–286.
- Axhausen, K. W., Schönfelder, S., Wolf, J., Oliveira, M. and Samaga, U. 2003. 80 weeks of GPS-traces: Approaches to enriching the trip information. In *Arbeitsbericht Verkehrs- und Raumplanung*, no. 178.
- Biljecki, F., Ledoux, H., and van Oosterom, P. 2013. Transportation mode-based segmentation and classification of movement trajectories. In *International Journal of Geographical Information Science*, vol. 27, no. 2, pp. 385–407.
- Bohte, W., and Maat, K. 2009. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. In *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 3, pp. 285–297.
- Chen, L., Ng, R. 2004. On the Marriage of lp-norms and Edit Distance. In *Proceedings of VLDB*, pp. 792–803.
- Chen, L., Ozsü, M.T., Oria, V. 2005. Robust and Fast Similarity Search for Moving Object Trajectories. In *Proceedings of SIGMOD*, pp. 491–502.
- Chung E., and Shalaby A. 2009. A Trip Reconstruction Tool for GPS-based Personal Travel Surveys. In *Transportation Planning and Technology*, vol. 28, no. 5, pp. 381–401.
- Das, R. D., Ronald, N., and Winter, S. 2014. Clustering based transfer detection with fuzzy activity recognition from smart-phone GPS trajectories. In *17th International Conference on Intelligent Transportation Systems, IEEE*, no. 17, pp. 3138–3143.
- Hemminki, S., Nurmi, P., and Tarkoma, S. 2013. Accelerometer-based transportation mode detection on smartphones. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, no. 13, pp. 13:1–13:14.
- Krumm, J., and Horvitz, E. 2006. Predestination: Inferring Destinations from Partial Trajectories. In *Proceedings of the 8th International Conference on Ubiquitous Computing*, pp. 243–260.
- Levenshtein, V. I. 1966. Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. In *Soviet Physics-Doklady*, vol. 10, no. 8, pp. 707–710.
- Manzoni, V., Maniloff, D., Kloeckl, K., and Ratti C. 2010. Transportation mode identification and real-time CO2 emission estimation using smartphones. Technical Report at *Massachusetts Institute of Technology*.
- Montini, L., Rieser-Schüssler, N., and Axhausen, K. W. 2008. Personalisation in multi-day GPS and accelerometer data processing. In *14th Swiss Transport Research Conference (STRC)*.
- Ogle, J., Guensler, R., and Vetri, E. 2005. Georgia’s commute Atlanta value pricing program: Recruitment methods and travel diary response rates. In *Journal of the Transportation Research Board*, vol. 1931, no. 1, pp. 28–37.
- Palma, A. T., Bogorny, V., Kuijpers, B., and Alvares, L. O. 2008. A Clustering-based Approach for Discovering Interesting Places in Trajectories. In *Proceedings of the 2008 ACM Symposium on Applied Computing*, pp. 863–868.
- Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Damiani, M. L., Gkoulalas-Divanis, A., Macedo, J., Pelekis, N., Theodoridis, Y. and Yan, Z. 2013. Semantic Trajectories Modeling and Analysis. In *ACM Computing Surveys*, vol. 45, no. 4, pp. 42:1–42:32.

- Prelipcean, A. C., Gidófalvi, G., and Susilo, Y. O. 2014. Mobility Collector. In *Journal of Location Based Services*, vol. 8, no. 4, pp. 229–255.
- Prelipcean, A. C., Gidófalvi, G., and Susilo, Y. O. 2015. Comparative Framework for Activity-Travel Diary Collection Systems. In *Proceedings of Models and Technologies for Intelligent Transportation Systems (MT-ITS) 2015*, pp. 251–258.
- Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., Srivastava, M. 2010. Using Mobile Phones to Determine Transportation Modes. In *ACM Transactions on Sensor Networks*, vol. 6, no. 2, pp. 13:1–13:27.
- Rocha, J.A.M.R. and Oliveira, G. and Alvares, L.O. and Bogorny, V. and Times, V.C. 2010. DB-SMoT: A direction-based spatio-temporal clustering method. In *Intelligent Systems, 2010 5th IEEE International Conference*, no. 5, pp. 114–119.
- Schüssler, N., and Axhausen, K. W. 2009. Processing Raw Data from Global Positioning Systems Without Additional Information. In *Journal of the Transportation Research Board*, no. 2105, pp. 28–36.
- Schüssler, N., Montini, L., and Dobler, C. 2011. Improving post-processing routines for GPS observations using prompted-recall data. Presented at *9TH International Conference on Survey Methods in Transport*, Termas de Puyehue, November.
- Shah, R. C., Wan, C., Lu, H., and Nachman, L. 2014. Classifying the Mode of Transportation on Mobile Phones using GIS Information. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 225–229.
- Shen, L. and Stopher, P.R. 2013. Should we change the rules for trip identification for GPS travel records? In *Proceedings of the Australasian Transport Research Forum (ATRF)*, vol. 36, pp. 261–267.
- Stenneth, L., Wolfson O., Yu P. S., and Xu B. 2011. Transportation mode detection using mobile phones and GIS information. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 54–63.
- Stopher, P., Fitzgerald, C., and Zhang, J. 2008. Search for a global positioning system device to measure person travel. In *Transportation Research Part C: Emerging Technologies*, vol. 16, no. 3, pp. 350–369.
- Tsui, S. Y., and Shalaby A. 2006. Enhanced System for Link and Mode Identification for Personal Travel Surveys Based on Global Positioning Systems. In *Journal of the Transportation Research Board*, no. 1972, pp. 38–45.
- Wagner, R. A., Fischer, M. J. 1974. The String-to-String Correction Problem. In *Journal of the ACM*, vol 21. no. 1, pp. 168–173.
- Wolf, D. 2000. Using GPS Data Loggers To Replace Travel Diaries In the Collection of Travel Data. PhD Thesis at *Georgia Institute of Technology, School of Civil and Environmental Engineering*.
- Wolpert, D. 1996. The Lack of A Priori Distinctions between Learning Algorithms. In *Neural Computation*, vol. 8, issue 7, pp. 1341–1390.
- Xie, K., Deng, K., and Zhou, X. 2009. From Trajectories to Activities: A Spatio-temporal Join Approach. In *Proceedings of the 2009 International Workshop on Location Based Social Networks*, ACM, pp. 25–32.
- Yan, Z., Parent, C., Spaccapietra, S., and Chakraborty, D. 2010. A Hybrid Model and Computing Platform for Spatio-semantic Trajectories. In *Proceedings of the 7th International Conference on The Semantic Web: Research and Applications - Volume Part I*, pp. 60–75.
- Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S., and Aberer, K. 2011. SeMiTri: A Framework for Semantic Annotation of Heterogeneous Trajectories. In *Proceedings of the 14th International Conference on Extending Database Technology*, pp. 259–270.

- Yu M. C., Lin C. J., Yu T., Chang E. Y., and Wang S. 2014. Big data small footprint: The design of a low-power classifier for detecting transportation modes. In *The Proceedings of the VLDB Endowment*, vol. 7, pp. 1429–1440.
- Zheng, Y., Chen, Y., Li, Q., Xie, X., and Ma, W.Y. 2010. Understanding Transportation Modes Based on GPS Data for Web Applications. In *ACM Transactions on the Web*, vol. 4, no. 1, pp. 1:1–1:36.