Trip Identification for Smartphone-based Cycling GPS Data[1]

Elmira Berjisian, University of British Columbia

Alexander Bigazzi, University of British Columbia

Introduction

Due to numerous advantages of non-motorized modes of transportation such as gained physical activity, meager travel costs, and lack of deleterious emissions policymakers have strived to promote them among citizens. Following the popularity of non-motorized modes of transportation, they began to be considered in transportation demand models. Individuals tend to undertake shorter trips via non-motorized modes of transportation which require a microscopic travel behavior analysis compared to motorized modes of transportation.

Complementary to traditional travel surveys, Global Positioning System (GPS) was employed in the 1990s and since then GPS studies began to emerge in travel behavior analysis (Li Shen & Stopher, 2014). With GPS household travel surveys, the potential for recording detailed travel behavior was realized in travel surveys. GPS studies offer many advantages over traditional travel surveys. Under-reporting trip frequency and over-reporting trip duration are the most common issues in traditional travel surveys which are resolved by GPS due to accurate recording of time and location (Houston, Luong, & Boarnet, 2014). The abundance of GPS data from different avenues such as fitness and travel smartphone apps, bike-share systems with location logging, and smartphone-based travel surveys enables investigation of non-motorized travel behavior on a microscopic scale.

Although GPS studies address many issues of traditional surveys and they ease the burden on respondents, they have some issues such as signal loss that generates interruption in the data stream or signal noise that is signified by jumps around the respondent's real location. These issues should be considered in post-processing of GPS data. In addition to those shortcomings, GPS studies lack key information such as mode and purpose of trips and whether a person stops to do an activity or is moving as a part of their trip. Therefore, there are sequential post-processing steps to 1) filter data 2) identify trips 3) detect mode for each trip 4) infer trip purpose 5) match GPS records to the underlying street network (Schüssler, Montini, & Dobler, 2011).

While most of studies focus on the third step in processing GPS data (i.e., mode detection), there is less attention paid to trip identification (the second step in processing GPS data) particularly in case of non-motorized travel which is the overarching aim of this paper. Most studies employ rule-based algorithms (RB-algorithms) to distinguish trips from activities in a GPS trajectory (Biljecki, 2010; Bohte & Maat, 2009; Dalumpines & Scott, 2017; Schuessler & Axhausen, 2009; Stopher, FitzGerald, & Zhang, 2008; Tsui & Shalaby, 2006; Usyukov, 2017; Wolf, Guensler, & Bachman, 2001). These algorithms use a set of rules designed by the analyst to identify trips in GPS trajectories. While these methods are advantageous for their simple implementation, they might be less accurate due to arbitrary thresholds selected by analysts. For example, most of RB algorithms use dwell time to decide if a person is stationary and doing an activity (Dalumpines & Scott, 2017; Stopher et al., 2008; Usyukov, 2017; Wolf et al., 2001). These studies often adopt 120 seconds as dwell time because it is the maximum cycle length according to Highway Capacity Manual and stopping at a traffic light should not be identified as an activity (L Shen & Stopher, 2013). A

[1] Presented at the 54th Annual Meetings of the *Canadian Transportation Research Forum*, May 26-29, 2019 at Vancouver , BC

recent literature review on processing GPS data underscored the necessity to develop more robust methods to identify trips (Gong, Morikawa, Yamamoto, & Sato, 2014). Trip identification is a critical step in processing GPS data because analyses based on spurious trips can lead to inaccurate inferences of travel behavior and culminate in devising misleading policies.

A number of recent studies undertakes more advanced machine learning algorithms (ML algorithms) to tackle the problem of trip identification (Gong, Yamamoto, & Morikawa, 2018; van Dijk, 2018; Zhou, Jia, Juan, Fu, & Xiao, 2017). Gong et al., applied Density Based Spatial Clustering of Application with Noise (DBSCAN) in conjunction with two constraints of temporal and entropy (DBSCAN-TE) on GPS trajectories to segments them into stopping and moving records (Gong et al., 2018). Further, they used a supervised learning method (support vector machine, SVM) to classify stopping records into activity and non-activity stops. In their study, ground-truth data were employed to adjust parameters of DBSCAN-TE and implement SVM. Another study used GPS trajectories to develop a set of 24 attributes in 6 groups (local attributes, global extreme attributes, speed-related attributes, acceleration-related attributes, tracking points clustering attributes, and heading change attributes) and fed them into a random forest classification algorithm to identify two classes (trip-ends and trip records) (Zhou et al., 2017). To train random forest, the authors used a prompted recall (PR) survey as ground-truth data. After applying random forest, a greedy algorithm corrected misclassified records. The greedy algorithm detected a class with duration less than 190 seconds and assigned records in that class to the other class.

Studies which apply ML algorithms use ground-truth data to train their models which is not available in many large-sample GPS studies. Furthermore, PR surveys as a source of ground-truth data are subject to the same human-related biases from that traditional travel surveys suffer (L Shen & Stopher, 2013). Apart from one study by van Dijk (van Dijk, 2018) that evaluated supervised learning algorithms against RB algorithms, there is no similar study to help discern the more appropriate method to address trip identification. Furthermore, van Dijk conducted his study on labeled simulated data and it has not yet been any study that compares unsupervised learning methods that function without ground-truth data and RB algorithms.

The objective of this study is two folds. First, proposing an unsupervised learning method to segment trips into activity and trip records with no available ground-truth data. Second, comparing different trip identification algorithms (six RB and two ML algorithms) on the basis of concordance and discrimination measures. Concordance measures evaluate agreements between different algorithms while discrimination measures are developed to help identify a more accurate method to distinguish trips from activities.

Method

Filtering Data

To remove erroneous GPS records, a set of rules is applied on GPS trajectories. These rules are displayed in Figure 1. The fourth rule addresses negative time gaps in data. Two cases may occur when there is a negative time gap. One, there is a sudden change in a date for a record compared to its adjacent records, such record is identified as an error and is removed. Second, the record at which negative time gap occurs signals a stop while adjacent records show movement (i.e., a stop for just a second), this is identified as an error and is removed from the trajectory. Otherwise, it may occur a jump, this record is treated as a correct record and moved to its proper place in the trajectory. Fifth rule screens out speed outliers by identifying position jumps in a trajectory according to (Schuessler & Axhausen, 2009). The sixth rule is consistent with (Biljecki, 2010) when there is a signal loss greater than eight hours, the participant is most likely doing an activity, thereby split the trajectories at such time gaps.

RB algorithms

Figure 2 displays sequential rules employed in six RB algorithms (Biljecki, 2010; Bohte & Maat, 2009; Dalumpines & Scott, 2017; Tsui & Shalaby, 2006; Usyukov, 2017; Wolf et al., 2001). Note the fourth rule of the first RB algorithm (written in **bold** in Figure 2) is not applied in our study due to lack of data.

Figure 1 Filtering rules

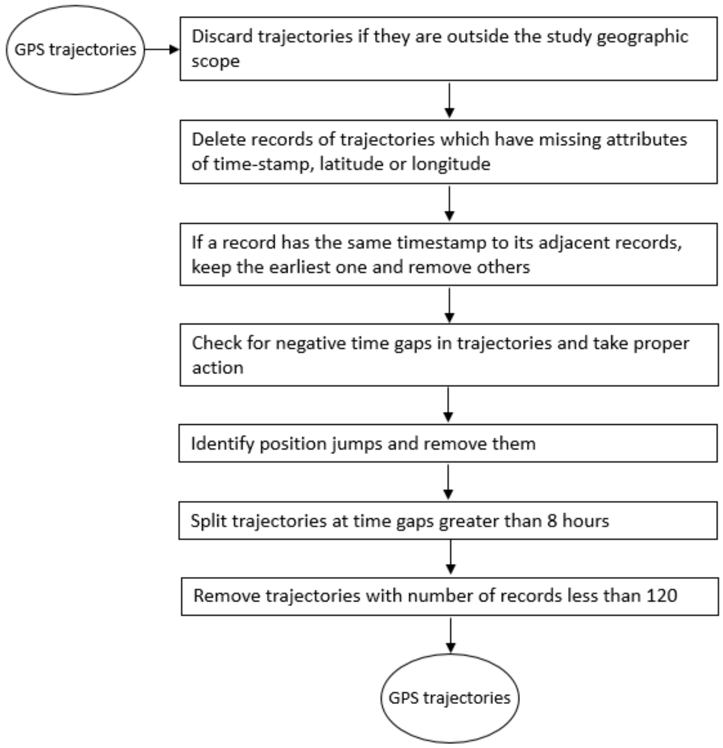


Figure 2 RB algorithms

Rule-based algorithm #1 by (Bohte & Maat, 2009)

- Remove GPS record If distance between 2 consecutive records < 10 m
- Split if time difference between 2 consecutive GPS records ≥ 180 sec
- Remove GPS record if its speed > 200 km/h
- **Merge trips if adjacent trips are within same shopping center polygon**
- Label record as garbage if speed < 5km/h and time difference between a record and its preceding record> 60 sec
- Remove trip if three times of number of garbage records within trip > number of records within trip
- Remove trip if $\sqrt{((\text{trip x max} - \text{trip x min})^2 + (\text{trip y max} - \text{trip y min})^2)}/\text{trip length}$ <0.3
- Remove trips if number of records within trip < 4

Rule-based algorithm #5 by (Biljecki, 2010)

- Split if time gap exceeds ≥ 8 hours
- Split if for time gaps ≥ 600 sec, average speed ≤ 2 km/hr

Rule-based algorithm #2 by (Tsui & Shalaby, 2006)

Signal loss:

- Split if 120 sec ≤ time difference between 2 consecutive GPS records ≤600 sec and distance traveled within this period ≤50 m
- Split if time difference between 2 consecutive GPS records > 120 sec and Euclidean distance traveled is ≥ 500 m
- Split if time difference between 2 consecutive GPS records > 600 sec

Ongoing recording:

- Flag an activity if GPS points have zero speed ≥ 120 sec

Rule-based algorithm #3 by (Wolf et al., 2001)

- Flag a trip end if there is no movement for 120 seconds.

Rule-based algorithm #4 by (Dalumpines & Scott, 2017)

- Flag a trip end if speed ≤ 0.15 m/s for ≥ 120 sec.

Rule-based algorithm #6 by (Usyukov, 2017)

- Flag a trip end if there is no movement detected for>600 sec.

ML algorithms

Filtered trajectories with signal loss more than 10 minutes are further split if average speed in signal loss is less than 0.56 m/s consistent with (Biljecki, 2010). Trajectories then are entered into ML algorithms to identify which portion of the trajectory is trip. Two unsupervised learning algorithms are used in this study. The first algorithm is the one proposed by this study in which random forest (RF) and partition around medoid (PAM) algorithms are combined (RF-PAM). A thorough set of variables to capture the dynamic of GPS trajectory is developed. This set consists of 24 attributes in 6 groups employed in a study by (Zhou et al., 2017). These attributes capture different dimensions of GPS trajectory at record-scale (for example, instantaneous speed or time gap between two consecutive records) and in a time window (for example, the density of points in k-neighborhood point set). According to (Zhou et al., 2017), k-neighborhood point set of a record consists of the record and its k preceding records and k proceeding records which yields a window of 2k+1 records. In this study, k is considered as 10 seconds. 24 attributes are then fed into a RF clustering method which yields a proximity matrix that represents how every two records in GPS trajectory are alike with regard to their 24 attributes. RF proximity between two records is yielded by counting number of times that they are located in a same terminal node (supplementary notes from (Seligson et al., 2005)).

RF proximity measure is employed in PAM clustering with two clusters (corresponding to trip and activity). PAM was developed by Kaufman and Rousseeuw (Kaufman & Rousseeuw, 2009) and is similar to k-means clustering but unlike k-means that minimizes Euclidean distance, PAM can use any arbitrary distance matrix (here, RF proximity measure). The result of RF-PAM method is two clusters of records (trip and activity). Before labeling these clusters to either activity or trip, the greedy algorithm developed by (Zhou et al., 2017) is used to correct misclassified records. The greedy algorithm results in some trajectories to have one cluster which is labeled as "trip". To label records within trajectories with two clusters to either trip or activity, two statistical attributes of clusters are compared (average of average speed in k-neighborhood point set and average density). According to (Zhou et al., 2017), speed and density contribute more than other statistical attributes to identify trip and activity records. Cluster with lower average speed and higher density is labeled as activity while the other cluster as trip.

Second ML algorithm, DBSCAN-TE, was used in a study by (Gong et al., 2018) to identify trip records from trip ends. DBSCAN-TE partitions records into noise (trip records) and clusters (stop records). DBSCAN identifies clusters by their higher density. Gong et al, added two constraints to the original DBSCAN to improve its application for identifying trip and trip end records in a GPS trajectory. First constraint, temporal constraint, ensures the temporal order of records within clusters and second constraint, entropy constraint, further excludes clusters that contain slowly moving records. Additionally, Gong et al, used SVM algorithm that needs ground-truth data to exclude short stops (non-activity stops) from activity records (Gong et al., 2018). Due to lack of ground-truth data, DBSCAN-TE with same parameters as (Gong et al., 2018) is employed in this study. Furthermore, to exclude short stops from activity records, instead of using SVM, only clusters with minimum duration of 190 seconds are maintained (consistent with the greedy algorithm in (Zhou et al., 2017)) and others are moved into trip records.

Concordance measures

To measure agreement between eight trip identification algorithms, concordance measures are used. The statistical tests to measure concordance determine if an algorithm can substitute another algorithm (Ranganathan, Pramesh, & Aggarwal, 2017). Table 1 demonstrates concordance measures for various comparison basis.

Table 1 concordance measures

#	Comparison variable	Sample	Statistical test
1	Dichotomous label of each record (activity or trip) for all GPS records across eight algorithms	All records	Fleiss kappa
2	Number of trips for all GPS trajectories across eight algorithms	All trajectories	intraclass correlation coefficient for 2-way mixed effect, absolute agreement, single measurement
3	Distribution of trip duration for all trips across eight algorithms	All trips	Kruskal-Wallis test and post-hoc Dunn test

To measure agreement between algorithms in terms of labeling a GPS record as trip or activity, Fleiss kappa test is implemented (Fleiss, Levin, & Paik, 2013). First and fifth RB algorithms are excluded from this test as they only specify if a GPS record is a trip or not (as opposed to dichotomous labeling of trip and activity). To capture agreement with respect to addressing signal loss, number of trips within each trajectory should be compared across different algorithms. For this concordance measure, intraclass correlation coefficient (ICC) is computed (Koo & Li, 2016). Third concordance measure explores if different algorithms agree in terms of trips duration extracted from GPS trajectories. Distribution of trip duration for each algorithm is obtained and compared with other algorithms. Due to unequal sample size (number of trips resulted from each algorithm is different), appropriate test to determine if there is a significant difference in the average of trip duration between every two algorithms is Dunn test. Additionally, Kruskal-Wallis test compares distributions to test if they come from the same population.

Discrimination measures

Discrimination measures indicate to what extent a trip identification algorithm differentiates between trips and activities. Different data attributes (speed, density, change in heading, and acceleration) are computed for trips and activities across eight algorithms and their distributions are compared using one-sided Welch's t-test (except first and fifth RB algorithms that have no record labeled as activity). Furthermore, to meticulously assess each algorithm transition between activities and trips, a 20-second transition window is defined where in a trajectory a trip ends in an activity (transition window). Different data attributes for trip records and activity records within the transition window in all trajectories are extracted and two resultant distributions are compared across algorithms. An additional discrimination measure is defined to compare how algorithms address the signal loss. The distribution of average speed in a signal loss at which an algorithm splits trajectories (computed by dividing the net distance moved in signal loss by time gap) for each algorithm is obtained and the average of the distribution is compared to 0.56 m/s (according to (Biljecki, 2010), this is the low threshold for walking and consequently if an activity is performed during signal loss, average speed in signal loss should be lower than this threshold). Table 2 demonstrates different dimensions of discrimination measures.

Table 2 discrimination measures

#	Comparison basis
1	Distribution of speed in the transition window for trip and activity records across algorithms.
2	Distribution of density in the transition window for trip and activity records across algorithms.
3	Distribution of change in heading in the transition window for trip and activity records across algorithms.
4	Distribution of acceleration in the transition window for trip and activity records across algorithms.
5	Distribution of speed for all trip records and activity records across algorithms.
6	Distribution of density for all trip records and activity records across algorithms.
7	Distribution of change in heading for all trip records and activity records across algorithms.
8	Distribution of acceleration for all trip records and activity records across algorithms.
9	Distribution of average speed for all trips and all activities across algorithms.
10	Distribution of net moved distance (distance between first and last record) for all trips and activities across algorithms.
11	Distribution of average speed in signal loss across algorithms

Data

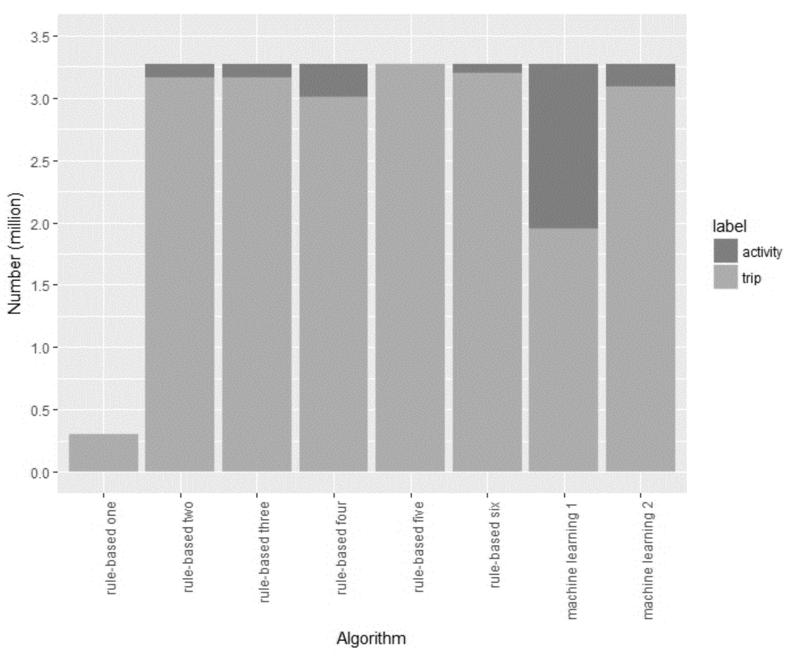
GPS trajectories for this study come from a GPS-based non-motorized travel survey conducted in 2017, Vancouver. Participants were persons of at least 14 years of age who "typically cycle at least once a week". After obtaining consent, participants completed an online questionnaire which contained demographic information, travel habits and preferences. Trajectories were recorded by a smartphone application which also allowed participants to indicate their missed trips as well as trip purposes. Trajectories contained second-by-second time stamp and location of participants. Recruitment and data collection occurred June through October, 2017, in which 146 persons recorded 1729 trajectories (70% on conventional bicycles and the rest on electric bicycles, by foot or unknown). After applying filtering rules, 94% of GPS records (1686 trajectories) were maintained which then entered in trip identification algorithms.

Results

Concordance measures

GPS trajectories with labeled records are concatenated to compare the agreement among different algorithms in terms of labeling a record as trip or activity. Figure 3 displays number of labeled GPS records among different algorithms.

Figure 3 labeled GPS records across algorithms



To measure whether different algorithms agree with each other in labeling a GPS record, Fleiss Kappa test is implemented which results in Kappa of 0.17 (99% confidence level) which indicates slight agreement (Landis & Koch, 1977).

The second measure of concordance (ICC) explores agreement among different algorithms with respect to number of trips. Despite the first concordance measure, this measure can reveal how an algorithm addresses the signal loss. 95% confidence interval of ICC ranges between -0.002 and 0.002 which indicates poor agreement (Koo & Li, 2016).

Last concordance measure investigates if the duration of trips is significantly different across eight trip identification algorithms. Kruskal-Wallis test indicates a significant difference among trip duration distributions. Dunn test is implemented to compare pairs of algorithms. At 99% confidence level, the average of trip duration across all algorithms are different.

Discrimination measures

To identify algorithms that distinguish activity and trips more clearly, 11 discrimination measures are computed for each algorithm (except first and fifth RB algorithms, for which only last measure is computed due to incapacity of these algorithms to identify activities). Statistical attributes within activities and trip are expected to be different. Activity records should signify higher density due to less movement along with low speed and acceleration. Consequently, in activity episodes direction of records is expected to change more frequently. Moreover, net moved distance in a trip is expected to be longer than an activity episode.

Discrimination measures 1 to 4 and 5 to 8 investigate different data attributes in a transition window and for whole trip and activity records, respectively. Intuitionally, a more distinct transition and a greater difference between activity and trip attributes can represent a more accurate algorithm. Finally, different algorithms are assessed with respect to how they address the signal loss by discrimination measure 11. In other words, to test the hypothesis that an algorithm splits trajectory at a signal loss because an activity was undertaking as opposed to data interruption in the trip. In the latter case, if the algorithm misidentifies signal loss as an activity, it generates spurious trips. For each algorithm, one-sided Weltch test (conforming to expectations explicated above for example, higher density in activity episodes rather than trip episodes) is conducted to compare two distribution of activity and trip records (in transition window or for whole trip/activity records).

Table 3 shows the results of statistical tests with respect to different discrimination measures, 1 indicates that algorithm significantly (at 90% confidence level) differentiates between trips and activities with respect to discrimination measures 1 to 10 while 0 signifies the inability of the algorithm in distinguishing between activities and trips. For discrimination measure 11, 1 shows if algorithms split trajectories at a signal loss where an activity was undertaking (signified by average speed during signal loss lower than 0.56 m/s). The last row in Table 3 demonstrates total score of an algorithm in terms of how many of discrimination measures indicate a significant difference between activities and trips (and in case of discrimination measure 11, correct splitting at signal loss).

Discussion and Conclusion

Only 7% of GPS records have the same label among different trip-identification algorithms. According to different concordance measures, none of the algorithms agree with respect to identified trips. Therefore, trip identification algorithms are not interchangeable.

Among RB algorithms, the first algorithm exhibits a weak performance as it removes many valid GPS records by its first rule (demonstrated in Figure 3). This result indicates the importance of developing a trip-identification algorithm that is appropriate for non-motorized travel where moved distance per second is much less than motorized travel. Similar to the first RB algorithm, the fifth RB algorithm does not provide rules to identify activities within ongoing recording of GPS. Among different RB algorithms, the fourth one distinguishes between activity and trips significantly (Table 3) and label more records as activities compared to other RB algorithms (Figure 3). Due to signal noise, a stationary GPS (during an activity) might show some movement which results in non-zero speed. The fourth RB algorithm is the only RB algorithm that considers such error and assigns low speed (0.15 m/s as opposed to zero) to activities.

Table 3 Discrimination measures for different algorithms

Discrimination measure	Algorithm							
	RB 1	RB 2	RB 3	RB 4	RB 5	RB 6	ML 1	ML 2
1	-	1	1	1	-	1	1	1
2	-	0	0	1	-	0	1	1
3	-	0	0	0	-	0	1	1
4	-	1	1	1	-	1	1	1
5	-	1	1	1	-	1	1	1
6	-	1	1	1	-	1	1	1
7	-	0	0	1	-	0	1	1
8	-	1	1	1	-	1	1	1
9	-	1	1	1	-	1	1	1
10	-	1	1	1	-	1	1	1
11	0	1	-	-	1	-	1	1
SCORE	0	8	7	9	1	7	11	11

According to Table 3, ML algorithms differentiate more distinctly between activities and trips (the highest scores are obtained by ML algorithms in Table 3) particularly in terms of change in heading and density. Further investigation revealed that the difference in the average of 8 out of 11 attributes in discrimination measures is greater for the second ML algorithm. Therefore, it surpasses the first ML algorithm in terms of capturing differences between attributes of activities and trips. Visual inspection reveals that the first ML algorithm identifies spurious activities where participants wait at intersections or where signal loss occurs.

In summary, ML algorithms pursue a data-driven approach to label a record as activity or trip while their performance can be enhanced by ground-truth data. By contrast, most of the existing RB algorithms were developed for car travel. Thus, such heuristic algorithms lead to inaccurate results when they are employed to identify non-motorized trips on a finer scale compared to car travel. ML algorithms function without the need for an arbitrary threshold but they can result in spurious activities where temporary stops during the trip are erroneously labeled as activities. RB algorithms are simple in terms of implementation but sensitive to arbitrary thresholds and might not identify all activities in a trajectory. Considering robustness of ML algorithms, they are recommended to be used for travel behavior studies where accurate GPS data are necessary.

Limitation and Future work

A sensitivity analysis on the degree that parameters of ML algorithms (for example, k in k-neighborhood point set in first ML algorithm) can influence the results needs to be conducted. Due to unavailable ground-truth data, validation of algorithms merely rely on their distinction between statistical attributes of activities and trips. In future research, external validation by trip purposes provided by participants will be conducted to infer clearer insights about performance of trip-identification algorithms. Furthermore, transferability of these algorithms is of question and depends on local characteristics of each city. For example, 120 seconds dwell time that is used in RB algorithms might be improper in congested areas such as downtown to identify activities. Furthermore, preprocessing GPS data (i.e., filtering step) may affect results of each algorithm. In this study, a set of simple rules were used to remove erroneous records while in original papers of RB algorithms, different approaches for filtering data were used. The premise of this study for filtering step is that if rules in filtering step remain homogenous for different trip-identification algorithms, results of each algorithm is only contingent on the performance of the algorithms.

References

References available on request: eberjis@mail.ubc.ca