Extracting networkwide road segment location, direction, and turning movement rules from global positioning system vehicle trajectory data for macrosimulation

Adham Badran^{1,2,\infty}, Ahmed El-Geneidy³, Luis Miranda-Moreno¹

¹Department of Civil Engineering, McGill University, Montreal H3A 0C3, Canada ²National Capital Commission, Ottawa K1P 1C7, Canada ³School of Urban Planning, McGill University, Montreal H3A 0C2, Canada

Received: February 12, 2024; Revised: February 27, 2024; Accepted: May 3, 2024

© The Author(s) 2024. This is an open access article under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).

ABSTRACT: The emergence of road users' global positioning system (GPS) trajectory data is attracting increasing research interest in knowledge discovery to improve transport planning-related methods and tools. In fact, the widespread use of GPS-enabled smartphones and the mobile internet has increased the availability and size of such data. With the increase in GPS data coverage and availability, some research has expanded its use to estimate state-wide vehicle-miles travelled, to classify driving maneuvers for road safety assessment, or to estimate environmental performance indicators, such as vehicular fuel consumption and pollution emissions. In computer science, research has used GPS data to infer road network maps. Although the inferred maps provide a correct topology and connectivity, they lack the essential details to be used for transport modeling. Therefore, this work proposes a method to extract network-wide road direction and turning movement rules. In addition, building a road network model under the widely used macroscopic transport modeling software serves as a proof of concept. A sensitivity analysis was carried out to determine the output quality and recommend future improvements. Road segment geometry and directionality were extracted accurately (case study accuracy of 95%); however, turning movement rules can be extracted more accurately using a larger GPS vehicle trajectory sample (case study accuracy of 68%).

KEYWORDS: global positioning system (GPS), transport model, road network, intersection control, map inference, road direction, turning movement

1 Introduction

Transport network modeling requires large quantities of data, depending on the project size and level of detail. For example, building a microsimulation network model for a neighborhood requires detailed road geometry, road type, transport demand matrices, intersection control type, and traffic light phasing. The model results also require validation, which is usually performed by comparing the model output to traffic counts and observed travel times and delays. These data are collected by different means and for different sample sizes depending mainly on modeling needs and available resources.

One of the data sources that is usually collected and used for transportation studies for validation and calibration is global positioning system (GPS) data. Recently, data from floating vehicles and probe vehicles were collected to estimate travel time, queue length, and traffic volume, as described by Zhao et al. (2019) and Zito and Taylor (1994). This technique estimates trip characteristics for a specific fleet or for predefined corridors, which can introduce bias when the sample is limited spatially (only a few corridors) or in terms of the fleet (only buses, taxis, or commercial vehicles). Tantiyanugulchai and Bertini (2003) used GPS-equipped

⊠ Corresponding author. E-mail: Adham.Badran@ncc-ccn.ca



transit vehicles to determine whether transit vehicle speeds and travel times are good proxies for general traffic conditions to be used in real-time advanced traffic management and traveler information systems. El-Geneidy and Bertini (2004) used probe transit vehicle GPS data to determine the optimal temporal resolution and speed for reporting traffic conditions obtained from loop detector data. Although these methods are useful for answering specific questions, Mennis and Guo (2009) found that an increase in the sample size and coverage of GPS data enables researchers to perform data mining and increase geographic knowledge discovery.

Recently, the widespread use of GPS-enabled smartphones and the mobile internet has made collecting and saving GPS data simple and relatively inexpensive. As these data become more widely available, they are attracting much research interest in the transport field. High-sample GPS databases are being built, and knowledge discovery research from GPS data has already started. For example, Fan et al. (2019) examined the use of GPS data to estimate vehicle miles traveled within the state of Maryland in the USA. In another study, Phondeenana et al. (2013) used vehicular GPS data to classify driving maneuvers to improve road safety. In the environmental field, studies have proposed methods to estimate congestion, vehicle fuel consumption, and pollution emissions using GPS data (Gately et al., 2017; Kan et al., 2018; Lin et al., 2019).

In parallel, computer science and geography researchers have been exploring the use of GPS data to infer road network geometry, topology, and connectivity. Some studies have compared different algorithms for inferring road networks from GPS data. These algorithms can be divided into three categories: point clustering, intersection linking, and incremental track insertion. Through a clustering approach, few studies in China have developed techniques to automatically extract road networks from GPS points or segments (Chen et al., 2021; Guo et al., 2022; Zhang et al., 2020). The main idea was to fit the road centerline according to the vehicular GPS data density distribution. Other researchers explored intersection linking techniques to generate road segments from vehicle trajectory data (Karagiorgou and Pfoser, 2012; Xie et al., 2016; Zhang et al., 2019). This approach divides the network inference process into two main steps: (1) detecting intersections using GPS data, for example, based on turning angles, and (2) using GPS vehicle trajectories to link the intersections together and form a directed road network. Finally, the track alignment approach has been studied by several researchers (Leichter and Werner, 2019; Xie et al., 2015; Ni et al., 2018). This technique incrementally adds global positioning system (GPS) tracks to an initially empty map and can also be seen as an incremental averaging of the GPS tracks. Considering the three methods, it is not possible to identify a single method that is preferred with respect to others; however, algorithms that produce maps with higher accuracy have lower coverage, and the opposite is also true. It can also be said that the most popular approaches are clustering and intersection linking, based on the number of publications under each of these approaches.

These techniques can serve as the basis for future research that aims to build a detailed road network for transport modeling or autonomous driving. In fact, road network building is a very active research area in autonomous driving. Providing detailed network features is essential for autonomous vehicle operation since they require precise knowledge of network topology and geometry. For example, Bender et al. (2014) aimed to develop the first map model usable by autonomous vehicles by representing road lanes and intersections not only in terms of directional lines but also in terms of drivable surfaces by introducing right and left bounds. The map model also needed to integrate driving rules.

Although the past developed work is useful for generating simple road networks based on GPS data with a correct topology and connectivity, there is a need to develop methodologies that help extract detailed network features for transport network modeling. For example, map inference methods lack the ability to extract detailed information on network-wide features such as turning movement permissions at intersections or the number of road segments in lanes, which are essential input data for transport network models. The development of macroscopic network models is labor- and data-intensive. Therefore, the development of automated methods can help significantly reduce the resources required in transport modeling tasks.

The objective of this work is to develop a method for extracting road network features from GPS vehicle trajectory data for use in transport network modeling. In addition, this study aims to provide a proof of concept by building a road network model with the widely used macroscopic transport modeling software EMME. More precisely, GPS vehicle trajectory data are used to extract network-wide road direction and turning movement information. This information is essential in transport modeling and land use studies when the study area is large, and network features cannot be collected as efficiently using other methods.

2 Methods

Four main input datasets are used: (1) GPS vehicle trajectory points, (2) geographic representation of the road network, (3) geographic location of all intersections, and (4) Google Maps and Street View. GPS data were collected during the spring of 2014 in Quebec City, Canada. The data were collected over 21 days by 2,000 voluntary users through the Mon Trajet smartphone app, which is made available by the municipality. Each point is described by the following attributes: map-matched X and Ycoordinates, trip ID, speed, and timestamp (Year-Month-Day-Hour-Minute-Second). Fig. 1 is a map of the raw GPS vehicle



Fig. 1 Study area - GPS vehicle trajectory points.



trajectory points (226,000 points) inside the study zone, which consists of 81 intersections.

A shapefile file of the up-to-date study area road network is available online (Adresses Québec, 2022). The locations of all intersections were obtained from the municipality (Données ouvertes, 2022). Finally, Google Maps and Street View were used to validate the results by serving as the ground truth for road segment direction and intersection movement permissions.

This study was completed using QGIS, ArcGIS, FME, and EMME. QGIS was used to visualize geographic data, perform visual validations, and create maps. The travel paths were constructed using the network analyst extension in ArcGIS. FME was used to manipulate the data and perform geographic operations. Finally, EMME is the macrosimulation software used to construct the road network model.

The extraction of road network features from GPS vehicle trajectories can be divided into three main parts: (1) initialization, (2) link direction extraction, and (3) turning movement permission extraction.

2.1 Initialization

The first part of the process is the initialization. It consists of creating an initial base network using EMME software and a simple road network shapefile. This creates a digital network representation composed of links and nodes (Fig. 2). Each node is uniquely identified and located exactly at the intersections depicted in the simple road network shapefile. On the other hand,



Fig. 2 Initial road network model for EMME - links and nodes.



links are created assuming that all roads are two-way streets, and each link is represented using its origin and destination nodes. It should also be noted that the initial road network model created by EMME allows all movements at intersections except for U-turns.

In parallel, the GPS points are filtered to remove outliers, defined as consecutive points separated by more than 30 m. This threshold was determined following the visual inspection of GPS vehicle trajectory points. Outlier removal created small gaps within the GPS trajectories, which were connected using a shortest path algorithm and a simple road network shapefile using the network analyst extension in ArcGIS. This process produced full trip trajectories that were geographically snapped to the simple road network, which enabled the following geographic processing steps.

2.2 Link direction extraction

The road network model created during the initialization step assumed 2-way links for all road segments. However, this is not always true because some roads are only one way. The link direction extraction process aims to extract the directionality information from the observed GPS data to remove modeled links that do not exist.

Following the initialization phase, the trip trajectories are divided into straight segments for which the segment azimuth is calculated. The azimuth corresponds to the angle between the segment orientation and the North measured clockwise. After examining the study area, a direction dictionary was created to associate different azimuth ranges to cardinal directions (Fig. 3). Each segment was then associated with a cardinal direction depending on its azimuth. The same procedure was applied to the initial EMME link table to determine the link directions. Once the directions are determined for the GPS vehicle trajectory segments and EMME links, a geographic operation is carried out to determine the nearest link for each GPS vehicle trajectory segment. The segment direction was used as a criterion to select only the nearest link in the same direction.

For each link, the number of observed GPS segments associated with it was calculated and compared with the number of segments associated with the reverse link by computing their ratio. For example, a ratio of 0.05 (or 5%) for a given link signifies that the number of GPS segment observations for that link is equal to 5% of the observed number of segments on the reverse link. This indicates a high likelihood of that link (or direction of travel) not existing since it is expected to have a similar count magnitude in both directions for a given road.

Azimuth angle range (°)	Direction
275-4	North
5-94	East
95-184	South
185-274	West

South



To determine the optimal ratio indicating the presence/absence of a link, a sensitivity analysis was carried out by testing different ratio value limits between 1% and 10% and comparing the results with the ground truth obtained from Google Maps. For a given ratio limit, a link with a ratio value smaller than the ratio limit value is considered to not exist, while a ratio value greater than or equal to the ratio limit value is considered to exist. Once the prediction was made for each link, the accuracy was calculated as the number of correct link direction predictions divided by the total number of links. The ratio limit value producing the highest accuracy was selected as the optimal ratio limit value. Once the optimal value was determined, the absolute number of observed segments for each road was analyzed to determine the impact of sample size on the link direction prediction accuracy. A second sensitivity analysis was performed by introducing different segment count cutoff thresholds. In other words, the prediction accuracy was computed on a subset of links that had at least a minimum number of observed segments in one of the two link directions. The tested cutoff threshold values were 0 (or all links), 10, 20, 30, 40, 50, 100, and 200. At each of the cutoff thresholds, the link directions were predicted and compared with the ground truth to calculate the prediction accuracy. A summary of the steps, including initialization, is presented in Fig. 4.

2.3 Turning movement permission extraction

Having created a correct link and node representation of the road network, the following step was to determine the permitted turning movements at intersections. In the initial road network model created using EMME, all intersection turning movements are allowed except for U-turns; however, the objective of this step is to extract and allow only the turning movements that were observed within the GPS trajectories. Fig. 5 presents a summary of the process used to extract turning permissions from observed GPS trajectories.



Fig. 5 Turning permission extraction process.



Since not all nodes are intersections, the intersection locations obtained from the Municipality were used to create 20-m radius buffers, which were selected through inspection of the study area, and the nodes within these buffers were selected as intersection nodes. The selected nodes were then used to create new 3-m radius buffers. These intersection node buffers were used to clip only the parts of the GPS trajectories located within each buffer. Since the modeled road network is geographically based on the simple road network and the GPS trajectories were snapped on the same simple road network during the shortest path operation, a good superposition of both geographic features was ensured. The clipping operation removed GPS vehicle trajectory segments that were considered to not be intersection movements. The remaining trajectory segments within the node buffers were then divided into two segments, inbound (toward the node) and outbound (outward from the node). The next step was to determine the azimuth for the inbound and outbound segments per node per trajectory segment. The azimuth was then used to determine the direction of every segment using the correspondence between the azimuth angle and the direction established in the previous step (Fig. 3). The next geographic operation was to find the nearest link to each inbound and outbound segment while ensuring a matching direction between them. At this point, each trajectory with an inbound and outbound segment within a node buffer is associated with two links (inbound and outbound) and can be expressed in terms of the intersection node, origin node (from the inbound link) and destination node (from the outbound link). A compilation of all observed movements at the different nodes provides the number of times that each movement has been made. A turning movement was predicted to be permitted if there was at least one observation from the GPS trajectories for that specific movement.

To determine the intersection movement prediction accuracy, the extracted turning movements for a subset of nine intersections (90 turning movements) were compared to the ground truth obtained from Google Maps and Street View. A sensitivity analysis was performed to assess the effect of sample size on prediction accuracy by evaluating the prediction results for turning movements for which at least one or two observations were extracted.

3 Results

3.1 Link direction extraction accuracy

The highest link direction extraction accuracy was 95%, which was obtained using a ratio limit of 5%. In other words, a link that has a GPS segment accounting for less than 5% of that of the reverse link can be considered to not exist with 95% accuracy. The sensitivity analysis results for the different ratio limit values are presented in Fig. 6.

Following the selection of the optimal ratio limit value (5%), an attribute was added to the initial modeled road network to indicate whether the directional road segment, represented by a link, existed. The resulting link representation of the road network is presented in Fig. 7. Links presented in blue were determined to be nonexistent since they do not have enough GPS segment observations compared to the reverse link (ratio < 5%). A special case is also presented in caption A of Fig. 7, where the initial road network model was created as four parallel links (compared to two links in regular situations). This is explained by the way the simple road network used as an input, represented that road. Since it has







Fig. 7 Final network model - extracted link results.

a large median, it was represented as two lines in the simple road network and therefore understood as two different roads by EMME. However, the developed method was able to determine which link corresponds to an existing road segment and filter the nonexistent links.

Considering sample size in the prediction accuracy assessment was found to have an impact. The introduction of a threshold on the segment count ensured that only roads with a minimum number of GPS segments were considered. Fig. 8 shows that increasing the minimum threshold is correlated with an increase in the link direction prediction accuracy.

For example, using a minimum threshold of 200 segments for a given road segment results in a prediction accuracy of 98.7%, as opposed to not having a minimum threshold, which results in 95% accuracy. However, for this study area and GPS dataset, setting the highest threshold implies that predictions can only be made for 304 links instead of all 674 links, as shown in Table 1.

3.2 Turning movement extraction accuracy

After comparing the extracted turning movements (n = 90) to the ground truth obtained from Google Maps and Street View, an accuracy of 68% was found. Moreover, 97% of the incorrect predictions correspond to turning movements that are permitted within the ground truth dataset but for which no observation was extracted from the GPS dataset. Furthermore, the prediction accuracy was 98% when only turning movements with at least one observation were examined. However, this restriction reduced the number of turning movements for which a turning movement is predicted by 37 and reduced the probability of detecting prohibited turning movements. Finally, the prediction accuracy for turning movements with at least 2 observed movements was 100%, but prediction could only be performed for 51% of the total number of turning movements. Fig. 9 presents an example of the result for one intersection that was extracted with 100% accuracy. Permitted turning movements are presented in red, while prohibited movements are presented in green. It should also be noted that no U-turns were extracted from the GPS data; therefore, it was not possible to determine the turning permissions for that turn type. from the GPS data sample; therefore, it was not possible to determine the turning permissions for that turn type.







Fig.9 Extracted intersection movement permissions (permitted turning movements in red and prohibited movements in green).

4 Conclusions

In this work, we develop a method that can extract link directions and turning movement rules from GPS vehicle trajectory data with high accuracy. Considering a link corresponding to a directional road segment with an observed GPS segment count less than 5% of that of the reverse link is a good indicator of the absence of that road segment. This resulted in a minimal prediction accuracy of 95%. The performance of a sensitivity analysis on the sample size (GPS segment count per road segment) proved that an increase in sample size will only improve prediction results (up to 99%). This level of accuracy is adequate for macroscopic models that require this type of information for large regions. The contribution of this method is the automatic extraction of directional road segments for very large regions, assuming good coverage of GPS data observations.

At intersections, turning movement permission prediction achieved a lower accuracy (68%) than link direction extraction. This is due to the lower number of observations for each intersection turning movement. In fact, turning movements with at least one observation were predicted with 98% accuracy. However, 37% of the permitted turning movements did not have any observations extracted from the GPS data. Therefore, an increase in sample size will allow better coverage of intersection turning movements.

Using this GPS dataset, it was not possible to extract network features (link directions and turning movements) for different times of the day simply due to the sample size. A larger dataset will allow for better knowledge discovery by providing greater temporal coverage of the different times of day. This is important for road networks with varying numbers of lanes available for traffic at different times of the day. For example, some road networks have restricted lanes for transit use during peak periods or for parking during off-peak periods. Similarly, some intersections have varying turning movement permissions by time of day for traffic optimization and safety purposes.

Overall, the developed method demonstrates the feasibility of automatic road network feature extraction for modeling and macrosimulation purposes. This work presents the required input data and the proposed methodology to achieve this objective. A proof of concept was also made by building a road network model using the EMME software for the study area in Quebec City, Canada, using GPS vehicle trajectory data collected by motorists. Data tables in the EMME format were created, indicating which



links and turns had to be removed from the base network to better represent real road network features within the study zone.

In summary, large datasets of GPS points/trajectories can be used to extract road network features to construct road network models. The extraction accuracy was found to depend mainly on the sample size. Therefore, the main limitation of this work is the sample size of the GPS vehicle trajectory data. The increased use of GPS-enabled devices and the increased availability of larger GPS datasets will only increase the prediction accuracy by providing greater spatial and temporal coverage. Spatial and temporal coverages dictate the area for which network features can be extracted and the possibility of extracting features for different periods of the day. It should be noted that for this case study, the GPS vehicle trajectory data and the Google validation data were not collected during the same period, which should be considered if the study area is undergoing changes in topology or major road redevelopment. For the current study, it was assumed that no significant changes in road topology occurred between the time that the GPS trajectories were recorded and the time that the validation data were observed.

In addition to the use of larger datasets, future works include the use of machine learning techniques, such as classification learners, to determine intersection movement permissions. Additionally, future research can explore the possibility of extracting more network features required for macroscopic modeling from GPS data, such as the number of lanes, road types, and link performance relationships. Autonomous driving can also benefit from the extracted network features by adding them to maps used in autonomous vehicles.

Replication and data sharing

xxxx

Acknowledgements

The authors would like to acknowledge the generous support of McGill University's Faculty of Engineering and the Vadasz Scholars Program through the scholarship to complete the doctorate degree.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

References

Adresses Québec 2022. https://adressesquebec.gouv.qc.ca/aqreseau.asp

- Bender, P., Ziegler, J., Stiller, C., 2014. Lanelets: Efficient map representation for autonomous driving. In: 2014 IEEE Intelligent Vehicles Symposium Proceedings, 420–425.
- Chen, B., Ding, C., Ren, W., Xu, G., 2021. Automatically tracking road centerlines from low-frequency GPS trajectory data. ISPRS Int J Geo Inf, 10, 122.

- Données ouvertes, 2022. https://www.ville.quebec.qc.ca/services/donneesservices-ouverts/index.aspx
- El-Geneidy, A. M., Bertini, R. L., 2004. Toward validation of freeway loop detector speed measurements using transit probe data. In: The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No.04TH8749), 779–784.
- Fan, J., Fu, C., Stewart, K., Zhang, L., 2019. Using big GPS trajectory data analytics for vehicle miles traveled estimation. Transp Res Part C Emerg Technol, 103, 298–307.
- Gately, C. K., Hutyra, L. R., Peterson, S., Wing, I. S. 2017. Urban emissions hotspots: Quantifying vehicle congestion and air pollution using mobile phone GPS data. Environ Pollut, 229, 496–504.
- Guo, Y., Li, B., Lu, Z., Zhou, J., 2022. A novel method for road network mining from floating car data. Geo Spatial Inf Sci, 25, 197–211.
- Kan, Z., Tang, L., Kwan, M. P., Zhang, X., 2018. Estimating vehicle fuel consumption and emissions using GPS big data. Int J Environ Res Public Health, 15, 566.
- Karagiorgou, S., Pfoser, D., 2012. On vehicle tracking data-based road network generation. In: Proceedings of the 20th International Conference on Advances in Geographic Information Systems, 89–98.
- Leichter, A., Werner, M., 2019. Estimating road segments using natural point correspondences of GPS trajectories. Appl Sci, 9, 4255.
- Lin, C., Zhou, X., Wu, D., Gong, B., 2019. Estimation of emissions at signalized intersections using an improved MOVES model with GPS data. Int J Environ Res Public Health, 16, 3647.
- Mennis, J., Guo, D. 2009. Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. Comput Environ and Urban Syst, 33, 403–408.
- Ni, Z., Xie, L., Xie, T., Shi, B., Zheng, Y., 2018. Incremental road network generation based on vehicle trajectories. ISPRS Int J Geo Inf, 7, 382.
- Phondeenana, P., Noomwongs, N., Chantranuwathana, S., Thitipatanapong, R., 2013. Driving Maneuver Detection System based on GPS Data. In: Future Active Safety Technology Toward Zero Traffic Accidents, 1–6.
- Tantiyanugulchai, S., Bertini, R. L., 2003. Arterial performance measurement using transit buses as probe vehicles. In: Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems, 102–107.
- Xie, X., Liao, W., Aghajan, H., Veelaert, P., Philips, W., 2016. A novel approach for detecting intersections from GPS traces. In: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 1816–1819.
- Xie, X., Bing-YungWong, K., Aghajan, H., Veelaert, P., Philips, W., 2015. Inferring directed road networks from GPS traces by track alignment. ISPRS Int J Geo Inf, 4, 2446–2471.
- Zhang, C., Xiang, L., Li, S., Wang, D., 2019. An intersection-first approach for road network generation from crowd-sourced vehicle trajectories. ISPRS Int J Geo-Inf, 8, 473.
- Zhang, Y., Zhang, Z., Huang, J., She, T., Deng, M., Fan, H., et al., 2020. A hybrid method to incrementally extract road networks using spatiotemporal trajectory data. ISPRS Int J Geo-Inf, 9, 186.
- Zhao, Y., Zheng, J., Wong, W., Wang, X., Meng, Y., Liu, H. X., 2019. Various methods for queue length and traffic volume estimation using probe vehicle trajectories. Transp Res Part C Emerg Technol, 107, 70–91.
- Zito, R., Taylor, M. A. P., 1994. The use of GPS in travel-time surveys. Traffic Eng Control, 35, 685.



Adham Badran is a professional engineer and senior transportation planner at the National Capital Commission. He has completed his graduate studies at McGill University in transport data and modelling to improve data-driven decision making. His professional background is in the field of transport planning and project management, mainly for major transport infrastructure projects.



Luis Miranda-Moreno has been an Associate Professor at the Department of Civil Engineering, McGill University since 2008. His scientific and professional contributions over more than 15 years have been mostly in the areas of sustainable urban mobility and road safety. His areas of research and expertise include emission modeling methods, the impact assessment of built environment and road design strategies on climate and road injuries, the strategies to promote non-motorized urban mobility and the integration of emerging technological solutions.



Ahmed El-Geneidy is a Professor in the school of urban planning McGill University. He leads TRAM, a multidisciplinary team generating research in several transportation areas. His research interests include land use and transportation planning, transit operations and planning, travel behavior analysis including both motorized (Auto and Transit) and non-motorized (Bicycle and pedestrian) modes of transportation, travel behavior of disadvantaged populations (seniors and people with disabilities) and measurements of accessibility and mobility in urban contexts.

