Expert System for Urban Multimodal Mobility Estimation Based on Information from Public Mobile Network

Krešimir Vidović[1], Sadko Mandžuka[2(✉)], and Marko Šoštarić[2]

[1] Ericsson Nikola Tesla, 10000 Zagreb, Croatia
kresimir.vidovic@ericsson.com
[2] Faculty of Transport and Traffic Science, University of Zagreb, 10000 Zagreb, Croatia
sadko.mandzuka@gmail.com

Abstract. Paper present new approach of urban multimodal mobility (UMM) estimation using anonymized data from public mobile network (PMN). The data set is derived from Call Data Records database, and urban multimodal mobility indicators were defined and relativized. Usage of indicators relativized values ensures that they can be applied for mobility estimation in all urban environment regardless of their physical differences, with existing public mobile network as single prerequisite. Travel mode classification is based on Adaptive Neuro Fuzzy Inference System (ANFIS) and trained using set of rules that were determined using method of surveying experts in domain of urban mobility. Accurate estimate creates a foundation for improvement of existing end creation of new services in urban mobility. Also, this approach has potential through implementation within advanced applications of Intelligent Transport Systems with the goal to improve travel modal shift, passenger comfort, efficiency of urban transport etc.

Keywords: Urban mobility · Multimodality · ANFIS · Mobile communication network · Call detail record

1 Introduction

Fostering multimodal transport is one of the key enablers for solving mobility issues in cities resulting in easing the movement of people and goods. This paper presents an approach in which the urban multimodal mobility indicators (UMMI) can be derived from non-location services related data being logged in public mobile network (PMN). Authors will elaborate on potential of usage of anonymized data from mobile network operators' databases with the goal of estimating urban multimodal mobility (UMM) by utilizing three base and one derived indicator. Additional data set, information about availability of transport modes will be used as well. This approach has a potential of creation of multimodal mobility estimation methodology applicable for all urban environments, since PMN, as (besides digital map) only required data source, exists in all urban environments. It is possible to apply this methodology through advanced

© Springer Nature Switzerland AG 2019
S. Misra et al. (Eds.): ICCSA 2019, LNCS 11620, pp. 3–11, 2019.
https://doi.org/10.1007/978-3-030-24296-1_1

applications within Intelligent Transport Systems (ITS), where similar applications, like driver and passenger information systems, traffic management, incident management systems, urban public transport management and similar, already exists [1–3].

2 PMN Generated Data in Urban Mobility Research

Recent developments have strengthened capabilities to understand and process telecom related big data sets and produce new data insights and can offer advanced data analytics that exclusively utilizes anonymized telecom network's data sets. The technology has evolved and user habits have changed. Mobile phone penetration in 2008 was 59,4%, and in 2018 it was 105% (4,9 billion users). More than 54% of those users own a smartphone (2,5 billion), and 85% of users are using mobile internet [4]. Therefore, increase in mobile phone penetration, together with increased number of smartphone share, enlarged number of mobile internet users and possibilities of precise user location determination in 3G and 4G networks have created the basis for efficient use of such data types in traffic engineering. Extreme computing performance and telecom knowledge makes it possible to understand telecom data, process millions of data records and produce new data insights. Scientific community has recognized telecom big data sets as quality data source for transport engineering which is evident from numerous scientific and professional bibliographies in this area. Authors have, in general, proven that this data type is representative, reliable and have been using them for various applications, like monitoring traffic flows, urban planning, travel mode detection, origin-destination matrix determination, public transport capacity evaluation and optimization, traffic volume determination, transport planning, traffic anomalies identification, mobility analysis etc., just to name few [5–14].

3 Spatial Decomposition

The telecommunication domain consists of a telecommunication network with the corresponding elements. It is a contextual, information world in which a CDR record is generated. In the context of the telecommunication network architecture in the contextual (information world), the physical locations of individual elements are not important, but how they are interrelated and what are their interrelated relationships. The transport network and the entities in the transport system are primarily related to the physical world, i.e. the traffic infrastructure (roads, tunnels, bridges). Traffic Engineering studies that part, or where the entity on the network is located and where it is and how it goes. The telecommunications system belongs to the contextual, information domain, while the traffic system belongs to the physical, material domain. When data from a CDR record is added to the information on the position of base stations in the space, the first step in transition from the contextual, information world into physics has been made, in which mobility as such takes place. Therefore, it is necessary to make that transition using the appropriate procedure (Fig. 1). Each base station has the appropriate coverage area, and its size and shape depend on several parameters. Base cell coverage limits are not isolated so that overlaps of coverage area

occur, especially in urban, densely populated areas. The surface area covering the base station therefore needs to be approximated to decompose the space in such a way that the exact surface area in the space, using the appropriate methodology, is assigned to the appropriate base station.



Fig. 1. Spatial decomposition (hexagonal, raster, Voronoi)

The space can be commonly divided into hexagonal cells, raster fields or Voronoi cells. Voronoi cells represent a special kind of decomposition on discrete sets of objects in space. Discrete sets of space points are defined by distance, where the position of the base station is taken as the center of the cell and includes all points that are closer to the center of that cell than the center of any other cell. Points that are equally distanced from both centers represent the boundary of the cell. Voronoi cells give a better, more realistic and high-quality image distribution over hexagonal cells and raster fields, and therefore will be used in this paper.

4 Multimodal Mobility Indicators and Calculation Methodology

In this paper, we have identified and defined three base indicators of urban mobility, and one derived from them. Number of trips refers to the number of movements between certain urban areas and presents a sum of all the recorded movements of users between certain urban areas within an appropriate timeframe. Second indicator is travel time indicator. The travel time indicator refers to the duration of each trip in the observed time frame. Value of indicators travel time can be made relative by calculating the contribution of each trip in relation to the longest trip, and the value of travel time for each trip is normalized with respect to a trip that has lasted longest during the typical day. In this case, the longest trip is the one in which the distance travelled is the largest, and that in fact the origin and destination remain within the same coverage area. Indicator “distance” is used as an approximation of the distance travelled. It refers to the distance the user has travelled between individual urban areas, i.e. the distance between the source and the destination of the trip. The length of travel or the distance travelled differs between the different urban agglomerations, as it may depend on the number of inhabitants, the density of housing, the surface of the city and the attractiveness of the attraction. Value of indicator distance can be relativized by calculating the contribution of each travel in relation to the maximum recorded travel, with the value of the distance travelled each trip is being normalized with respect to the longest journey in a typical day. In this case, the longest trip is the one in which the distance

travelled is the largest, and that in fact the origin and destination remain within the same coverage area. Calculation methodology is described in following parts of paper. After filtration of unneeded data, it contains temporary ID of users, the time of start of the telecommunications activities and geographical location of the base station. Prior to indicator calculation process, several assumptions and terms need to be defined and describer. The trip is the movement of the user between the two base station of the PMN (corresponding urban areas) where he has achieved telecommunication activity registered in the CDR record. Authors are aware that real origin and destination of trip might differentiate from those identified by telecommunication activities. This will be the subject of further researches, where the relationship among migrations identified from telecommunication activities and migrations that really occurred will be analyzed. False motion or false travel is a user's trip identified by changing the base station, which did not actually happen, or there was no physical user location change. This occurs due to the operation of PMN and the tendency to optimize the operation of some of base stations by transferring excess users on the other base station coverage area at that time less load (Load balancing). Time threshold is used to filter records on the "false" movements and is 10 min < x < 60 min, which means it eliminates all trips that have lasted for less than 10 or more than 60 min. Spatial threshold is used to filter the records of the "false" movements which eliminate all trips shorter than 1 km. Speed threshold is eliminating all trips that, due to speed indicator values, assume that they are not realistic or are the result of an error. (average speed > 100 km/h). Application of these filters might also affect some real migrations, which might affect estimation in smaller or medium size cities. These issues will be addressed in further research, where the signaling data will be utilized and fused with CDR data. User position is more precisely determined, by using trilateration method, and therefore, previously mentioned uncertainties, like false movements, will be mitigated. Additionally, that methodology also solves the problem of capturing only segment of trip in which the user has performed telecommunication activities, since signaling occurs even without telecommunication activities, since the only prerequisite is that the mobile terminal is switched on.

Therefore, the input data file is transformed into a table. Base stations are then identified and are stored in the corresponding table. A variable ID location is being introduced, containing the unique ID for the same pair of LAT, LONG, it is matched and is added to the table of base stations. Following steps include determination of Euclidean distance between each pair of base stations identified in previous steps. The values are stored in the table for each pair of base stations end are calculated using Haversine formula. Then, the process continues for identifying all location changes for each trip, for each user, in each individual data set, and calculating the value of time difference between the change of state (Δt). In the next step, the travel time indicator is defined as the difference in time between the recorded start of telecommunication activities at the source base station and recorded start of telecommunication activities at the destination base station for each trip. Then, in following step the travel speed is being calculated for each trip as the quotient of Euclidean distance value and time changes. It is presented in km/h. Next step is the calculation of the values of the indicators using analyzed input data. Origin destination matrix contains number of trips between all pairs of base stations among which the trips occurred within a certain time.

Indicator of the number of trips between individual base station pairs is calculated using the relation (1):

$$A(i,j) = \sum Trip_ID(i,j) \tag{1}$$

where, the *A(i, j)* represents the total number of recorded trips between individual pairs of base stations (Origin-Destination Matrix), *Trip_ID(i, j)* represents trip between two base stations in the appropriate time frame, where all assumptions are met. The matrix (or the table) of travel time contains the time values of average travel time between all pairs of base stations and is calculated as the average of all recorded travel times between same pairs of base stations within a predefined time. Indicator travel time (B) is calculated using the expression (2):

$$\bar{B}(i,j) = \frac{\sum B(i,j)}{n} \tag{2}$$

Where $\bar{B}(i,j)$ represents the average value of all travel times between base stations in a predefined time frame, B(i,j) represents the value of the travel time for each trip in the time frame, and n represents the number of registered trips between base station pairs within the time frame. Following indicator is matrix of distances and it contains distances between all pairs of base stations among which the trips occurred within a certain time. Also, it is important to note that the Euclidean distance between base station pairs is the same regardless of the direction of travel.

Indicator travel speed is calculated using expression (3):

$$D1(i,j) = \frac{C(i,j)}{B(i,j)} \tag{3}$$

Where D1(i, j) stands for the travel speed for each trip in the time frame, C(i, j) stands Euclidean distance between base station pairs and B(i, j) stands for the duration of each travel trip in a time frame. The result of the previous step are values stored in four matrices for each time frame; in the matrix of number of trips (Origin Destination Matrix), travel time matrix and distance matrix, and travel speed. To ensure that the model is applicable to all urban areas, regardless of their size, the data obtained should be normalized, i.e. the data should be kept in the interval [0, 1]. Normalization is performed so that the value of the indicator in a given time interval is normalized with the highest value of the individual indicator for all eight-time intervals. The purpose of the normalization conducted in this way is to include in the calculation and the extent to which each indicator participates in the overall mobility of urban agglomeration in each individual frame, i.e. to identify its maximum. This will result in a more credible calculation of the UMM estimation for each time frame.

5 Design of ANFIS Based Expert System

Following the process of calculation of urban mobility indicators, next step includes determination of relationship between the values of the UMMI and the value of the multimodal mobility estimation. This paper utilizes so called expert knowledge gathering method for this purpose. Expert knowledge is what qualified individuals know as a result of their technical practices, training, and experience. It may include recalled facts or evidence, inferences made by the expert based on "hard facts" in response to new or undocumented situations, and integration of disparate sources in conceptual models to address system-level issues. Experts are usually identified based on qualifications, training, experience, professional memberships, and peer recognition. Therefore, selected experts will be presented with a set of questions (survey). The question arises from an expert's opinion ("knowledge gathering") on how and to what extent the combination of values of indicators affects the UMM, in the context of mode identification. Survey questions include all permutations of the value of indicators included in the UMM estimation, enriched with several modal options (walking, car, public transport) The expert answers to the question by giving the appropriate value of the multimodal mobility estimate based on their own knowledge and experience in each issue. The scenarios are formed used following principle:

If <value of number of trips indicator> & <value of travel time indicator> & <value of distance travel indicator> & <value of speed indicator>, travel mode is <travel mode> and multimodal mobility is <mobility estimation> .

Through the questionnaires, experts then point to the value of the offered scale for which the mobility estimate can be ranked in different categories.

Upon completion of the survey process, the responses of all the experts participating were evaluated and processed and were used in further steps in the multimodal mobility estimation process. Analysis of the data will be performed for checking the consistency of the response. In order to establish an expert system for multimodal mobility estimation, a method of fuzzy logic, or an adaptive neuro-fuzzy inference system (ANFIS), will be used. ANFIS process overview is presented in Fig. 2. Training process begins by obtaining training data set and checking data set. Training data consist of two vectors, first is input and second is output vector. Those vectors are used to train the ANFIS system. Training data set is used to determine parameters for membership function and threshold value for the error between the actual and desired

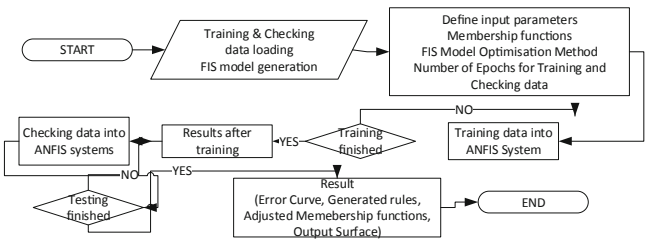

Fig. 2. ANFIS training process, based on [15]

output is determined. Consequent parameters are found using least square method, and when the error becomes less than defined threshold value, process will terminate.

ANFIS training learning rules use hybrid learning, combining the gradient descent and the least squares method. Training begins by creating a set of training data to be able to train the Neuro-Fuzzy system. The data set used as the input must be in a matrix form. First four columns represent input data with values of urban mobility indicators, as presented to the experts, and the last column in the matrix is the output, value of mobility estimation, generated by processing responses from survey data. Initial membership functions for al input variables are created (for four mobility indicators), and the parameter of membership function is determined. Input and output variables for fuzzy interference system are presented at Fig. 3. Following the initial membership functions creation, system training begins. When the training process is finished the final membership functions and training error from the training data set are produced. To identify the most appropriate system, system must be trained with a variety of settings for items such as data set sample, epoch number, membership function type and number, and number of inputs to achieve the best performance. The checking data is then used in conjunction with the training data set for accuracy enhancement.

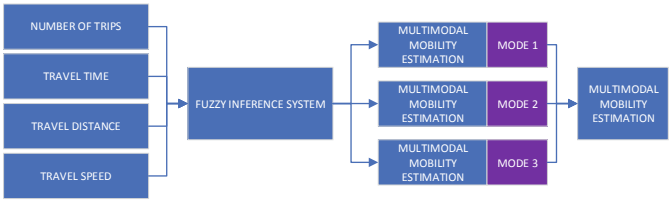

Fig. 3. Input and output variables to fuzzy interference system

Preliminary training result (without multimodality analysis) was network with 78 nodes, 108 linear parameters, 36 nonlinear parameters. In this simulation 654 data pairs used for model training, Fig. 4.

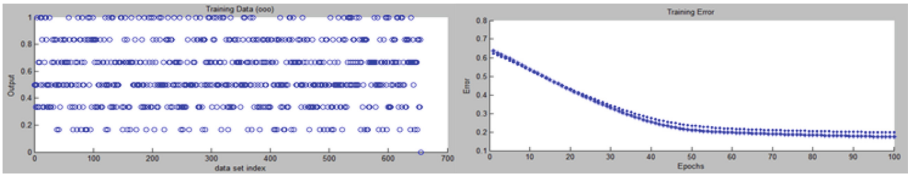

Fig. 4. Characteristics of preliminary training results (distribution of training data and dynamics of training error)

Analysis of model prediction results by comparing the results of models derived from the data set for learning, the data set for validation and the data set for model verification, Fig. 5. The average error check of the model was 0.1907. The process of

validation was carried out for each Fuzzy Inference System (total of 32 proposed candidates). The result of this process is ANFIS based expert system for urban mobility estimation based on user's telecommunication activities in PMN. Therefore, based on values of UMMI, identified and calculated from data from PMN, and rules gathered from surveying experts, using ANFIS method, expert system was formed that is capable of categorizing user migration according to travel mode (based on indicator speed) and evaluate UMM. UMM is separately calculated for each identified transport mode and is stored in multimodal mobility estimation matrix for each defined mode. So, in this case, it will result in three matrixes that will contain multimodal mobility estimation values for each base station pair. In case that user have combined different transportation modes during the trip, this trip will be divided by segments based on identified mode, and those segments will be addressed as independent trips.

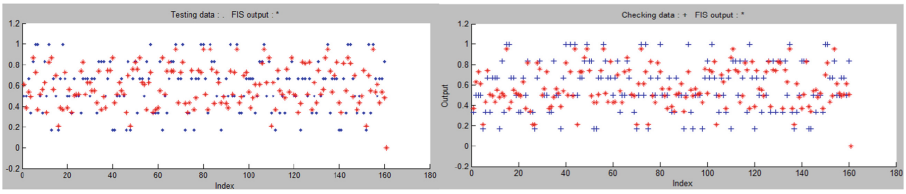

Fig. 5. Testing data and checking data

To determine the extent to which each mode of transport affects total multimodal mobility, it is necessary to determine the share of different modes represented by the number of trips in the total number of trips. Therefore, after the identification of travel mode, it will be possible to identify how much each mod participates in migrations between the appropriate pair of urban areas. The identified value will then be used to relativize the share of each individual travel mode in total multimodal mobility. At the end, multimodal mobility estimation value between the two pairs of base stations (equivalent urban areas) will be calculated as the sum of multiple values of urban mobility estimates for each mode of travel and the relative value of the representation of each mode of travel.

6 Conclusion

This paper demonstrated the potential of the data derived from the public mobile telecommunications networks in the domain of UMM research. It has presented a novel approach in which the data derived from the CDR database from public telecommunication networks can be used in area of smart mobility. Paper has described mobile network topology and addressed issues regarding spatial decomposition and propose solution to get an accurate multimodal mobility estimate. Paper has presented a methodology for modelling four UMMI (number of trips, distance, travel time and speed), and has proposed an approach in which the relativized values of UMMI can be

used in all urban environments, regardless of its size, population or local characteristics. This paper has proven that the urban mobility of inhabitants within an urban environment can in a specific time frame be described with an urban mobility indicator based on information about recorded telecommunication activities of users in a PMN.

References

1. Vidović, K., Mandžuka, S., Brčić, D: Estimation of urban mobility using public mobile network. In: Proceedings of 59th International Symposium, ELMAR-2017. Faculty of Electrical Engineering and Computing, Zadar, Zagreb, p. 21–24 (2017)
2. Škorput, P., Mandžuka, S., Jelušić, N.: Real-time detection of road traffic incidents. Promet - Traffic Transp. **22**(4), 273–283 (2010)
3. Mandžuka, S., Kljaić, Z., Škorput, P.: The use of mobile communication in traffic incident management process. J. Green Eng. **1**(4), 413–429 (2011)
4. GSMA: The Mobile Economy Global (2018). https://bit.ly/2oTFrsl. Accessed 29 Jan 2019
5. Zhang, D., Zhao, J., Zhang, F.: UrbanCPS: a cyber-physical system based on multi-source big infrastructure data for heterogeneous model integration. In: Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems, ICCPS 2015, pp. 238–247 (2015)
6. González, M.C., Hidalgo, C.A., Barabási, C.A.: Understanding individual human mobility patterns. Nature **453**, 779–782 (2008)
7. Gillis, D., Semanjski, I., Lauwers, D.: How to monitor sustainable mobility in cities? Literature review in the frame of creating a set of sustainable mobility indicators. Sustainability **8**(1), 1–30 (2016)
8. Qiu, Y., Tatem, A.J.: Data-mining cellphone-based trajectory data for collective human mobility pattern analysis. World Wide Web Internet Web Inf. Syst. **2009**, 2–3 (2009)
9. Calabrese, F., Diao, M., Lorenzo, G., Ferreira, J., Ratti, C.: Understanding individual mobility patterns from urban sensing data: a mobile phone trace example. Transp. Res. Part C **26**, 301–313 (2013)
10. Diana, M., Pirra, M.: A comparative assessment of synthetic indices to measure multimodality behaviours. Transp. A Transp. Sci. **12**(9), 771–793 (2016)
11. Qu, Y., Gong, H., Wang, P.: Transportation mode split with mobile phone data. In: Proceedings of the IEEE Conference on Intelligent Transportation Systems, ITSC, vol. 2015, October, pp. 285–289 (2015)
12. Bachir, D., Khodabandelou, G., Gauthier, V., El Yacoubi, M., Vachon, E.: Combining Bayesian inference and clustering for transport mode detection from sparse and noisy geolocation data. In: Brefeld, U., et al. (eds.) ECML PKDD 2018. LNCS (LNAI), vol. 11053, pp. 569–584. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-10997-4_35
13. Biljecki, F., Ledoux, H., Van Oosterom, P.: Transportation mode-based segmentation and classification of movement trajectories. Int. J. Geogr. Inf. Sci. **27**(2), 385–407 (2013)
14. Williams, N.E., Thomas, T.A., Dunbar, M., Eagle, N., Dobra, A.: Measures of human mobility using mobile phone records enhanced with GIS data. PLoS ONE **10**(7), 1–16 (2015)
15. Jang, J.S.R.: ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans. Syst. Man Cybern. **23**(3), 665–685 (1993)