A Contextualized Origin-Destination Matrix Derived from Telecommunications Activity Record

Petra Martinjak*, Vjera Turk**, Luka Grubišić*1

*Department of Mathematics, Faculty of Science, University of Zagreb, Bijenička cesta 30, 10000, Zagreb, Croatia

**Department of Computer Engineering, Faculty of Engineering, University of Rijeka, Vukovarska ul. 58, 51000 Rijeka, Croatia

1luka.grubisic@math.hr

Abstract - Traditional origin-destination matrix, a tool often used in traffic science, contains the amount of migrations within certain region and time period. Those migrations are indicators of socio-economic activity in the area, but their motives are unknown (untracked). Matrices can be derived from telecommunications activity records. The goal of this research was to propose a method for contextualization of such origin-destination matrix using information from material world. Spatial objects within the area of interest were categorized by type of socio-economic activity they are involved with. This new information about context of each origin and destination cell allowed development of probability model which separates initial migration flows into contextualized streams. Contextualized origin-destination matrices are visually presented with surface plot and interactive map.

***Keywords* — origin-destination matrix; ODM; socio-economic activities; contextualization; contextualized OD matrix**

INTRODUCTION

An origin-destination matrix is an important tool for describing group mobility and measuring socio-economic activity in a certain region, often used in traffic science for urban migration analysis and strategic planning. Opposed to traditional field-collected data, amount of socio-economic activity and mobility in certain area can be estimated from data related to telecommunication activity [1][2][3]. If socio-economic purpose of objects in the observed region is considered, origin-destination matrix can give valuable insight about context of urban migrations in the area. Such contextualized origin-destination matrix provides better understanding of nature of socio-economic migrations. They can provide valuable information used for optimization of urban planning and help develop better city infrastructure. As a proof of concept, we present one concrete approach to enriching the origin-destination matrix with context (motivation) of migration.

MATERIALS AND METHODS

The origin-destination matrices used in this research were derived from Charging Data Records as described in [4][5] and use space partition based on Voronoi cells. They represent migration data during a single day divided on 8 time-frames (3 hours each).

Our method utilizes spatial data from open and crowdsourced project OpenStreetMap. The specific data structure of OpenStreetMap data influenced modeling of algorithms described in sections A and B, so we begin with a brief summary of this structure. In OpenStreetMap, spatial objects are described following hierarchical order, consisting of *nodes*, *ways* and *relations*. *Nodes* are contained in *ways* and *ways* are contained in *relations*. *Nodes* represent points with geographic position and are used to describe points of interest. *Tags* are key-value pairs assigned to *nodes* which are used to store metadata about map objects (type, name, physical properties etc).

The method consists of acquisition of spatial data, area segmentation, purpose-based object categorization, spatial object filtration and development of probabilistic model for segmentation of migration flows given in original origin-destination matrix.

A. Data acquisition

Algorithm for extraction pulls *node* objects from hierarchical structure. It proceeds to look at the *tags* of the extracted *nodes* and pair *nodes* geolocation with descriptive attributes (e.g. restaurant, school, hospital). This results in data set containing spatial objects along with their geolocations and purpose-oriented attributes.

B. Area segmentation

The region area is then segmented into Voronoi cells in order to follow the space partition inherited from origin-destination matrices [4]. Spatial objects are assigned to respective cell based on their geolocation. This is done by forming *spatial points* through detaching any additional data in node structure from node identification number and geolocation, then defining subsets of *spatial points* lying within each cell, and finally restoring node structure by attaching additional data back to *spatial points* based on the same node identification number.

C. Purpose-based object categorization

Algorithm is deployed to categorize spatial objects based on their socio-economic purpose defined by attributes assigned to them. Five categories are defined to cover basic socio-economic motives: *home, work, education, health* and *leisure*. Specific attributes are given to each category regarding the nature of categories purpose (e.g. attribute "hospital" is assigned to *health* category). Spatial objects are then classified according to their attributes being in a certain category. Objects not classified to either of aforementioned categories were labeled as "*other"*.

D. *Spatial object filtration*

Information gained from area segmentation and purpose-based categorization is combined into algorithm for filtration of objects per purpose and cell. For every Voronoi cell:

1. 6-dimensional vector *attr* is defined and initialized with zeros. Elements of vector *attr* represent number of objects in one of earlier described category
2. For every category *a*:
    2.1. From all the nodes in current cell, those from category *a* are extracted.
    2.2. Appropriate element of vector *attr* is updated by number of nodes extracted in 2.1
    2.3. Extracted nodes are removed from list of all nodes.
3. Vector *attr* is appended to id of current cell

Result of this step is a matrix which rows are Voronoi cell ids and columns are numbers of objects in specific category (per cell).

E. *Probabilistic model development*

Probabilistic model is now created and deployed with intention of merging contextualized data given as a result of previously described steps and migration data given from the origin-destination matrices. Algorithm considers time-frames during which origin matrices are obtained. For every time-frame, the motivators for visiting each object category varies. For example, work category objects are going to be more likely visited during morning hours than in the evening, but in the evening places for leisure are going to be more visited. Considering these relationships of visitation likelihood for each category and during each time-frame, a table of weights for every category in dependence of time-frame is made.

For every time-frame, following steps are taken:

1. Vector *w* of weighted values per category and Origin-destination matrix *O* are taken for the matching time-frame
2. Utilising number of objects per cell for each category and the vector *w*, weighted distribution of migration flow in specific time-frame is deployed
3. Elements of matrix *O* represent migrations from some *i*-th to some *j*-th Voronoi cell. For every migration *m* of matrix *O:*
    3.1. Migration flow *m* is separated using weighted distribution. Result is a 6-dimensional vector whose every element is number of migrations from i-th Voronoi cell to object of specific category in j-th Voronoi cell.
4. By combining these vectors, 3-dimensional contextualized origin-destination matrix is assembled

RESULTS

The results are visually represented in form of bar-plots and surface plots for every category and every time-frame separately, as well as in interactive map. Interactive map allows for choosing origin and destination cells and yields separated migration flow between them.

DISCUSSION

Further research will focus on enriching origin-destination matrix with additional data and optimization of probabilistic model for contextualizing migration flows using GPS data and methods of statistical learning.

REFERENCE

[1] Gonzalez, M C, Hidalgo, C A, and Barabasi, A-L. (2008). Understanding individual human mobility patterns. Nature, 453, 779-782. doi:10.1038/nature06958. Available at: http://barabasi.com/f/250.pdf, accessed on 5 September 2017.

[2] Dong, Y et al. (2015). Inferring Unusual Crowd Events from Mobile Phone Call Detail Records. Proc of Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Part II, 474-492. Porto, Portugal. Preprint available at: https://arxiv.org/abs/1504.03643, accessed on 15 September 2017.

[3] Šćepanović, S et al. (2015). Mobile Phone Call Data as a Regional Socioeconomic Proxy Indicator. PLoS One, 10(4), e0124160. doi: 10.1371/journal.pone.0124160. Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0124160, accessed on: 5. September 2017.

[4] Dešić, S, Filić, M, Filjar, R. (2017). Determination of origins and destinations for an O-D matrix based on telecommunication activity records. Proc of MIPRO 2017, 544-547. Opatija, Croatia. Available at: http://bit.ly/2xIzbq5, accessed on 5 September 2017.

[5] Filjar, R, Filić, M, Lučić, A, Vidović, K, Šarić, D. (2016). Anatomy of Origin- Destination Matrix derived from GNSS alternatives. Coordinates, 12(10), 8-10. Available at: http://mycoordinates.org/anatomy-of- origin-destination-matrix-derived- from-gnss-alternatives/, accessed on 5 September 2017