ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 14 (2016) 1802 – 1808

Transportation Research Procedia

www.elsevier.com/locate/procedia

6th Transport Research Arena April 18-21, 2016

Theoretical method for building OD matrix from AFC data

Viktor Nagy [a,*]

[a]*Egyetem tér 1, Győr 9026, Hungary*

Abstract

The smart cards conquered a serious space in the traffic in the past years. The cards are used like an automatic fare collection system by the different public transport companies. The fare collection is easier, faster and last but not least more efficient with this method. However the generated data are utilizable for other aims apart from the original aim.
The origin-destination matrix plays a key role in a transport company's life. We may adapt to the travel demands efficiently in the knowledge of the matrix. Thanks to the proper service the passengers will be satisfied, and they use the service with pleasure. The additional benefit, that the unnecessary capacities can be forced back in the knowledge of the demand, and thereby the expenses of the service provider's decrease. The origin-destination matrix was able to create only with classical counting and interview methods till now. Their disadvantage, that the production consumes considerable workforce and time with a method like this, and hereby the costs are higher. The counting concern to a given time only and the sampling proportion compared to the number of travels of the year is quite slight. Because of that the distortion effect is considerable, the matrix cannot reflect the effects of an intervention. These effects get into the system only after the next counting.
The penetration of the smart cards increases, the card user passengers' number increases continuously. Thanks to that, not only the cost efficiency of the fare collection is growing, but at the same time the statistical sample size too. If more passengers from young persons to older ones accept the usage of the cards, we may receive more accurate picture about the transportation network.
If the fare collection system was built up somewhere, the origin-destination matrix can be manufactured with lower expenses, and because of the continuous usage the matrix will be current always, the changes become traceable. We may know the demands in all the moments, the origin-destination matrix may change dynamically in function of the weather, time of the day, season and so on. The check-out data can be manufactured from exclusively check in data with different algorithms. The

* Corresponding author. Tel.: +36-96-613-561; fax: +36-96-613-561.
E-mail address: nagy.viktor@sze.hu

Peer-review under responsibility of Road and Bridge Research Institute (IBDiM)
doi:10.1016/j.trpro.2016.05.146

passenger's boarding and alighting places become known on the given route. Linking the sections belonging to a single trip and defining the target point, the origin destination matrix can be developed.

Peer-review under responsibility of Road and Bridge Research Institute (IBDiM)

Keywords: smart card; OD matrix; E-ticketing

1. Introduction

Disregarding the few exceptions, the currently in use e-ticketing systems imply only check-in data, however, there are researches with systems without place data too, like it was examined by Ma et al. (2012), and can be found the check in-check out type systems, when the passenger uses his smart card in the case of boarding and alighting too, like in the research from Wang et al. (2015). The used the travel information to estimate the impact of a fare change and to know the prospective revenue. Zhong et al. (2015) also have both information and examined the network flows and the mobility patterns. In these papers the passengers starting and landing stations are known in the given route. Unfortunately, we can't find this information everywhere, and these systems cannot show a whole picture about the traveling habits. The data don't provide the accurate information from the given trip's beginning and ending points, only the single segments of the trip are known. This study presents a theoretical method, which can help us creating the origin–destination matrix from check in data. The research leans on assumptions which determine or influence the passengers traveling habits.

2. Trip estimation

In the case of a check-in – check-out system the task is easier, because the boarding point and the alighting point for the given route are also known. We can see in Fig. 1. that the trip started at the stop point A and the travel ended in C, and meanwhile the passenger touched stop point B, where he transferred. The problem is, that the card data concern the single segment of the whole trip only, and we can only deduce to the whole trip. The difficulty is that it is necessary to define it, that which stop points count as a destination, and which are only transferring stops. In the knowledge of the travel habits, we may receive a more accurate picture about the whole travel beginning and end point.

To finish the trip it is necessary to determine what points count as a traveling aim, where an activity may have happened. We can define certain conditions in this situation, which can be the followings according to Munizaga et al. (2014):

- If the passenger exceeds a certain time constraint, which assumption also appeared in a research from Guedes et al. (2012)
- If the passenger does not transfer immediately (tolerance period assured)
- If the passenger ride forward on an the identical line after a little time
- If the passenger returns on the identical line after a little time
- If the passenger makes a too big roundabout with his travel

In the above cases reasonable, that the passenger alight with some kind of aim in the given stop, and not transferring only. In this situation a trip ends. Defining the trips of all the passengers and gather them, the origin--destination matrix can be developed. For this, we need to collect all the complete trips, and we can fill up the matrix row's (origin) and column's (destination).

In those cases, where the traveling sections' starting points are known only, the number of the steps grows. It's not enough to solve the trip only, but the single landing points too. To this, we have to apply different base assumptions too, which are based on the travel habits, and they help to define the alighting points. These criteria may be the followings after Barry et al. (2002):

- The starting point of the travel is the endpoint of the previous travel
- The endpoint of the day's last trip is equal to the starting point of the next day's travel
- If the passenger stayed longer time at a stop, it is probable, that the trip ended there

According to starting assumption, the endpoint of the travel is equal to the starting point of the next travel. Refining this a little, it is verifiable, that in the reality the criterion can sound like this: the endpoint of the travel is the stop, from where the next starting point can be reached most comfortably. The establishment of the alighting stops happens as follows: we examine the position of the stops on the given line, where the passenger can transfer, and then we select the closest one. The shortest distance brings up a problem however. According to Munizaga and Palma (2012), in some cases – because of the forming of the line – the achievement of the nearest stop point to the next stop takes more time, as the passenger gets off earlier and walk to the stop point.

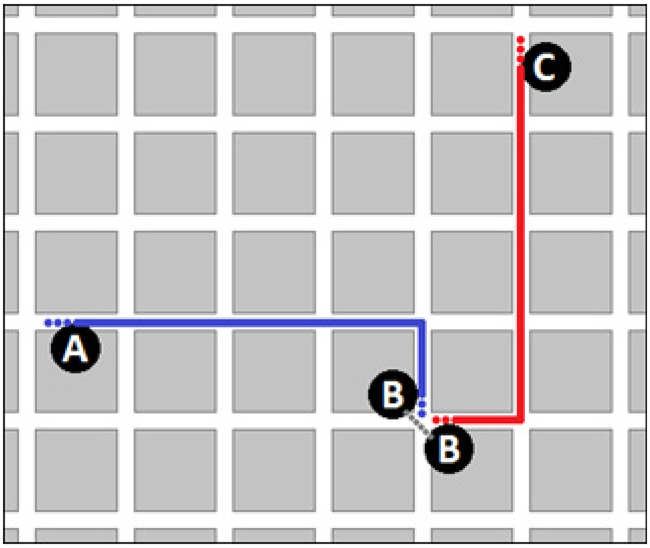

Fig. 1. A whole trip.

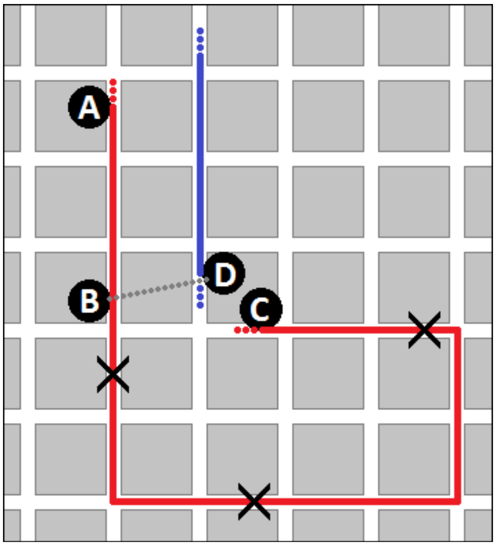

Fig. 2. Problem with loops.

In this case of Fig. 2, the latter stop point on the line is nearer to the next boarding, but because of the loop, it yields considerable plus running-time for the passenger. In the reality, the passengers alight from the vehicle sooner, and rather choose a little bit more walking. Certainly the transferring and the walking yield a discomfort factor, so it is necessary to take into account this discomfort. The elder ones, the pensioner have different requirements with time and walking. They are less set in time, but the walking between two stop points cause difficulty for them in general. The pupils and employees are more bound in time, but their displacement does not cause such a problem. Based on the above ones, we can state, that both of the time and the distance parameters are essential. We may supply the attributes of the single card types with different weight numbers, which define it, what kind of preference prevails for the given cluster.

2.1. Time as a common factor

If we bring the single values onto a common factor, then the values can be compared, just like in cost-benefit analysis. In this case the common factor may be the time. We may measure the travel time against the time profit achieved with the earlier landing. However the walking causes a discomfort, moreover, this discomfort is different for every single cluster. Munizaga and Palma (2012) thing, if the profit is bigger in the case of walking, then the discomfort caused loss, then the passenger chooses the walking.

The maximum walking inclination belongs to the issue of the walking, which one is generally defined in the 500–1500 meters by Zhao et al. (2007) and Munizaga et al. (2014). According to this assumption, the passengers don't walk to larger distances from own strength, but this is in correlation with the city's public transport network and the economics, according to Li et al. (2011). However, a Chilean study from Munizaga et al. (2014) found a context between the walking inclination and the stops physical position. According to the results of the study, the passengers are able to walk for longer distances in the downtown. Presumable, that the thicker inbuilt and the common services are the reasons. Based on these, the walking maxima can be more precise, with the handle of the districts' position.

3. Possible problems

The production of the single travel segments and the trips arise different kind of problems, which ones may cause the rip of the trip, or the trip can be non-generatable. One of the capital problems arises from the walk. If the line's possible landing stops and the distance between the next boarding point exceed the maximum walking distance, then it is impossible to manufacture the landing place, and because of that, the trip neither. This mistake may follow from longer walking actually, or playing the wag, the choice of other traffic method or because of the stop's building up, where it is not necessary to use the travel card. These cause such disturbances, which corrections meet with difficulties. The number of the endpoints which can be manufactured grows, if we take the downtown's bigger walking inclination into consideration.

It's also a problem in the case of trip estimation, if the given passenger travels from an identical line and stop point daily, but travel only once per day. In a case like this, only the initial station is the known one, which means the home presumably, but the destination is indefinable. The passenger makes the return journey with other traffic method and we don't know where he gets off from the given vehicle.

The other projection of the problems is technical. With the usage of the cards, mistakes and deficiencies may come about in the data rows. We do not have to disregard these records, handling the single mistakes, the information can be obtained. The data rows duplication is an often appearing problem. The system reads the card repeatedly for any reason, and the single card data can be found two times after each other. In this case, the system can calculate transferring data only to one of the records, the other one will be valid on the given section only, and using the same card twice is meaningless. Onto the prevention of this, filtering the repeated card touch out, we can leave the first of the data rows only in the given route and wipe the others. In most cases, we can find the data in a form like this:

- Card ID
- Card type
- Date and time
- Line
- Route
- Stop point

Li et al. (2011) state, that we may recover the deficiencies occurring in the records (if some of the other cells contain information). If the system reads at least two data plus the number of the card, in most cases the content of the cells can be manufactured.

- We may deduce the boarding point from the time and the route
- We may deduce the time from the route and the stop point
- We may deduce the route from the line and the time
- The vehicle identifier definable from the route and time

The absent information can be reckoned with proper safety with the help of these values' historical data, so it is not necessary to delete the given record, the trip does not split and therefor the sample going to be larger. Last but not least the origin-destination matrix will show a more realistic picture.

4. Development and other application opportunities

It is necessary to go back to the bases entirely in the interest of correcting the alighting information manufactured from the boarding data. The system has to rest on reality close bases, and it can reflect the real habits better so. This correction can be reached with the extension of the basic assumptions. If we formulate more statements, the received results may be more precise. The statements have to be based on the passengers' behavior experienced in reality, which need to be mapped in a suitable measure. The implement of the data is a fundamental viewpoint. These corrections have an effect on the waiting and the walking inclination mostly, but some of that reinterpret the assumptions. The result of their application gives help in the definition of the dubious alighting places mostly. My assumptions are the followings:

- The travel habits of the same card type owners may be different
- The travel habits of the same age passengers may be different
- The travel habits depend on the weather
- The travel habits depend on the built up of the station
- The travel habits depend on the character of a district
- The travel habits depend on the time of day, and the season
- The travel habits depend on the aim of the travel
- Passengers' travel habits are different with a season ticket and with a normal ticket
- The travel habits depend on the country's culture
- There is a strong connection between the previous assumptions, and they depend on each other

The first two dots point beyond the differences between the elder ones and young persons, which problem was examined by few researchers, like Morency et al. (2007) and Nishiuchi et al. (2013). They can be habits different in the same age group. An elementary school student and a high school student have another habit and preferences. The college student is different from a same aged young employee. The weather and the building of the station are linked to each other. They have a big effect on waiting and walking. The character of the district influences the walking inclination, however the safety level of the neighborhood or the forming have an effect on those too, and it is also necessary to take the time of day, and the season into consideration. The aim of the travel influences the character of the trip. We go to work, go home, and go shopping in a different way. The given country's culture and the wage-level (passengers with normal ticket) influence all of the dots in a smaller measure, so only with these assumptions can we develop the system.

Based on these dots it is can be declared, that the single individuals, or individual types are choosing a route according to different preferences. Because of this, the individual passengers' cognition is inescapable in the interest of receiving a real picture about the network. Based on boarding and alighting points, the traveling aims and the information of travelling times, card types and partner companies, we may order the passengers to clusters, and we supply the single groups with different characteristics.

My additional remark, that the Euclidean distance, as a walking distance, we cannot use in all cases, because more obstacles may impede the pedestrians (railroad, busy road…). The Euclidean distances may be misleading in these cases, so it is necessary to examine the road network and the permeability between the single sections in the interest of the accurate picture.

These variables appear in the calculation like influencing factors. With the help of these measurable and known values, we have not only the classical points of references, but a much more extensive background basis. Important to notice, that from one factor we cannot establish the expected behavior. It is necessary to define the stochastic variable onto the totality of the factors. This variable determines that what kind of behavioral pattern we may expect from the owner of the card, in a given and well determined situation.

4.1. Income supplement

The opportunities for application of smart card data point beyond the matrix production. Compared to the questioning and counting methods, the demands are traceable, in a case of an intervention, therefore the conformity may happen sooner. The appearance place, time and possibly the reason of the single disturbances are definable with these methods. Adapting to the demand increases not only the clientele's satisfaction, but reduces the service provider's expenses. The economicalness is increasable and the measure of the subsidy is reducible with the suppression of unnecessary capacity kilometers and the concurrent demand increase.

The income is increasable if we know the traveling habits and we are sharing it with different firms. These incomes may stem from advertisings, but even from market collaboration, where the companies persuade the card users to traveling and shopping with the help of different discounts or loyalty points, like in the study of Páez et al. (2011).

In a case like this the owner of the card gets loyalty points after a certain amount of travels, which points can be exchanged in the partner businesses. The passengers may buy depreciated products with these points, or they can use their card as a loyalty card, which idea appeared in the research of Bagchia and White (2005). The supporting system of the discounts may work inversely too. It is possible to persuade the card owner into public transport with the points obtained after the bought products. With the obtained extra income the service can be improved and the loss can be decreased.

The information from the partner businesses can be useful because of the single travelers' cognition. The usage of the card at different partners helps the passenger's habits cognition, and helps in the traveling aim definition. We will know the target, which is equal to the end of the trip.

After the mapping of individuals, in the knowledge of the traveling volume and the combination of clientele, we can use direct marketing to allot the transportation company for an additional income supplement.

5. Conclusion

Increasingly more researchers are interested in the opportunities of Smart Card data usage. Their adaptability has a wide circle, but the solution methods are in their infancy. The main assumptions are equal in the researches, but the principles don't cover the whole spectrum of the circumstances affecting the travel behavior. With the usage of my assumptions the system comes closer to the reality. It is necessary to focus on the single individuals and groups first, to be able to examine all of the network's passenger flows.

In my further research, I would like to elaborate a model, what based on the literature and could take the above mentioned rules and the outlined influencing factors into a common framework and system. The effects of some of the factors come to know directly from the data, if the necessary additional information is known, but we cannot clarify the other set of factors with methods like this. For the exploration of the accurate behavior of these factors I plan to prepare a survey, which can brighten the effects of these factors.

References

Bagchia, M., White, P.R., 2005. The potential of public transport smart card data. Transport Policy 12, 464–474.

Barry, J.J., Newhouser, R., Rahbe, A., Sayeda, S., 2002. Origin and destination estimation in New York City with automated fare system data. Transportation Research Record 1817, 183–187.

Guedes, M., Oliveira, N., Santiago, S., Smirnov, G., 2012. On the evaluation of a public transportation network quality: Criteria validation methodology. Research in Transportation Economics 36, 39–44.

Hiroaki Nishiuchi, H., King, J., Todoroki, T., 2013. Spatial-Temporal Daily Frequent Trip Pattern of Public Transport Passengers Using Smart Card Data. International Journal of Intelligent Transportation System Research 11/1, 1–10.

Li, D., Lin, Y., Zhao, X., Song, H., Zou, N., 2011. Estimating a Transit Passenger Trip Origin-Destination Matrix Using Automatic Fare Collection System. DASFAA Workshop, 2011, 502–513.

Ma, X., Wang, Y., Chen, F., Liu, J., 2012. Transit smart card data mining for passenger origin information extraction. Journal of Zhejiang University, 750–760.

Morency, C., Trépanier, M., Agard, B., 2007. Measuring transit use variability with smart-card data. Transport Policy 14, 193–203.

Munizaga, M., Devillaine, F., Navarrete, C., Silva, D., 2014. Validating travel behaviour estimated from smartcard data. Transportation Research Part C 44, 70–79.

Munizaga, M., Palma, C., 2012. Estimation of a disaggregate multimodal public transport Origin-Destination matrix from passive smartcard data from Santiago, Chile. Transportation Research Part C 24, 9–18.

Páez, A., Trépanier, M., Morency, C., 2011. Geodemographic analysis and the identification of potential business partnerships enabled by transit smart cards. Transportation Research Part A 45, 640–652.

Wang, Z., Li, X., Chen F., 2015. Impact evaluation of a mass transit fare change on demand and revenue utilizing smart card data. Transportation Research Part A 77, 213–224.

Zhao, J., Rahbee, A., Wilson, N., 2007. Estimating a rail passenger trip origin–destination matrix using automatic data collection systems. Computer-Aided Civil and Infrastructure Engineering 22, 376–387.

Zhong, C., Manley, E., Arisona, S.M., Batty, M., Schmitt, G., 2015. Measuring variability of mobility patterns from multiday smart-card data. Journal of Computational Science 9, 125–130.