

On railway stations statistics in Smart Cities

Dmitry Namiot, Oleg Pokusaev, Vasily Kupriyanovsky

Abstract— According to studies on the standards of Smart Cities, the transport (or transport component) is one of the defining services in Smart Cities. This explains the large funds and human resources that are invested in the planning, development, and analysis of transport communication in cities all over the world. In this article written on the results of work on the design of the new urban rail system, the authors analyze data on the use of railway stations both within the city and in the urban agglomeration area. One of the key moments and an initial part of any transport design is always the estimation of traffic (the estimation of the use) of the proposed transport system. This requires an understanding of the patterns of the movement of passengers (models of using the transport system). Identification of such patterns and their use in urban analytics are the subjects of this article. Obviously, the use patterns reflect the current state of the transport system and the urban environment. Accordingly, the recorded changes in usage patterns can serve as indicators and metrics for changes in the urban environment.

Keywords— urban analytics, transport data, time series, events proceedings.

I. INTRODUCTION

This article is an extended and updated version of the work presented at the DAMDID 2018 conference.

All cities in the world pay attention to own transport services. In all Smart City definitions, transport (or transport components) is one of the key services in Smart Cities. This means the exceptional importance of projects linked with planning, developing, and analysis of transport communication in cities. This article presents the results of the work on designing a new network of urban railways in Moscow. In our work, we investigated the patterns of the use of railway stations in the city and the suburbs. The first question for any new transport project is always the prediction of traffic (the estimation of use) of the new transport system. Actually, the traffic defines all economic and social aspects of any transport project. In order to provide such predictions, we need to understand the patterns of the movement of passengers (the models of using the transport system). The existing patterns and their possible changes are the base for our prediction. Identification of such patterns and their use in urban analytics are the main subjects of this paper. Obviously, the transport use patterns

reflect the current state of the transport system and the urban environment. Accordingly, the discovered changes in usage patterns can serve as indicators and metrics for changes in the urban environment. It is how the transport behavior patterns could be used in urban analytics.

Mobility (or smart mobility) is one of the key characteristics of the Smart City according to all standards. Accordingly, all cities are constantly involved in the development of transport projects. Cities are no longer designed for cars. Modern cities are oriented, rather, to pedestrians (so-called pedestrian economy) [1]. A convenient opportunity to use several modes of transport is the basic requirement for smart transport in a smart city. Mobility services should be multimodal. Intellectual mobility is defined as the use of technology and data to create links between people, places, and goods in all modes of transport. Mobility as a service is a new concept that offers consumers access to various types of vehicles and experience of travel [2]. Mobility as a service can be perceived as a "better choice" for organizing a move, and this can change the way we are currently treating transport. The key issue for creating applications of the class "mobility as a service" is the availability of digital information from various sources of the urban economy [3].

Recently, all cities in the world show interest in rail transport development (urban railways). This is due to several and quite natural causes that are common to all cities. Railway transport (the urban railway) for today, for example, is the only way to organize traffic without traffic jams.

Note also that the very concept of the city has expanded significantly. Today, in most cases, we should talk about agglomerations, since a large number of people living outside the city constantly go to the city to work, study, etc.

It is obvious that unfortunately, all transport projects are quite expensive. And the first part of any such project always includes an estimate of possible traffic. As we pointed out above, this is based on the understanding of the models of the use of the transport system in the city. For example, passengers, in reality, cannot always choose the shortest route, if it is connected, for example, with extra movement. The choice will be affected by the possible time savings and comfort of a change [4].

In this paper, we present the results of an analysis of the use of railway stations, which were, in fact, aimed at identifying patterns (models) for using the urban railroad. This was a part of the project to build new lines of urban railways in Moscow (Fig. 1).

Received Feb, 6, 2019.

Dmitry Namiot - Lomonosov Moscow State University; RUT (MIIT) (email:dnamiot@gmail.com)

Oleg Pokusaev - Center of digital high-speed transport systems Russian University of Transport (MIIT) (email: o.pokusaev@rut.digital)

Vasily Kupriyanovsky - National Competence Center for Digital Economy Lomonosov Moscow State University; RUT (MIIT) (email: vpkupriyanovsky@gmail.com)

to weekends. The description of the distributions is exactly the pattern for the use of the railway station. For some stations, these templates are almost the same, for some stations they are different. This difference is explained by the location of the station, the density and composition of the population in the vicinity of the station, the presence of centers of attraction.

Technically, the search for templates is reduced to determining the similarity of time series. So, the main tool here is a time series similarity measurement [10]. Given two time series T_1 and T_2 , a similarity function calculates the distance between the two time series. In our case, we will refer to distance measures that compare the i -th point of one time series (T_1) to the i -th point of another (T_2). The typical example is Euclidean distance. There are other methods for the distance measures [11], but the key moment in our case is the equal length for time series. Most other methods were invented just to compensate for the difference in the sizes of the time series being compared. For example, we could mention here Dynamic time warping (DTW) [19]. DTW is one of the often used algorithms for measuring the similarity between two temporal sequences, which may vary in speed. Also, it is seen that it can be used in partial shape matching application. In our case, the speed is always the same and sequences always have the equal size. Most measures (metrics) are dealing with individual data points composing the compared time series. There is so-called derivative DTW [20], which is based on approximated local derivatives instead of data points. It is interesting because derivatives based approaches should be more suitable for dealing with outlines.

In our work, we have successfully used a shape-based similarity measure - Angular Metric for Shape Similarity (AMSS) [12]. This approach treats a time series as a vector sequence and focus on the shape of the data and compares data shapes by employing a variant of cosine similarity. It is illustrated in Fig. 2.

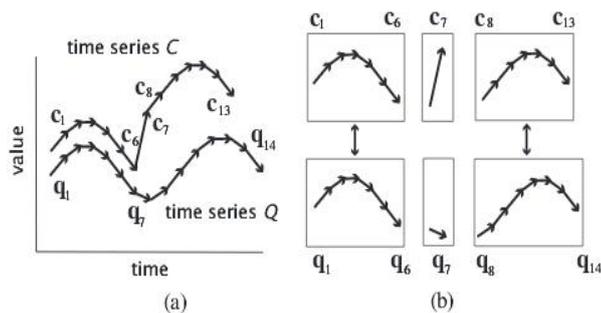


Fig. 2. Angular Metric for Shape Similarity [12]

The cosine similarity metrics minimize the influence of outliers in similarity computation, where outliers are defined as much bigger or smaller data points than their immediate neighbors [12]. Also, our choice was influenced by a table with data from extensive testing of various approaches to the similarity of time series, given in the above-mentioned paper, where AMSS was compared with other metrics: Euclidean distance, DTW, DDTW, etc. As per the provided study, AMSS was found to be the best (dis)similarity measure.

It is possible, of course, that a given similarity measure is not effective for every kind of time series data. It could be

suitable for some types of data, and not so for the others types of data. This could be true for AMSS too. In our experiments, AMSS was applicable. However, the general questions of using the similarity metrics of time series were not the subject of this article. Our goal was to obtain practical conclusions based on the data provided. We could note in the conclusion of this section that the main problem for similarity measures is the outlines.

IV. ON RAILWAY STATIONS USE IN MOSCOW AGGLOMERATION

The idea of data analysis consists of two main points. The patterns of movement that will be extracted from the data (can be found in the data) are a reflection of some existing socio-economic processes in the urban agglomeration. Where do the workers live, where are their jobs in the city, what mode of work, etc. Accordingly, the conclusions that will be made on the basis of data analysis must have some explanation from the point of view of these processes. Let's call this "urban" explanation. And vice versa. Some changes in the observed data can serve as an indicator (or even a metric) of changes in the city (in the agglomeration area).

As far as we know, earlier the analysis of information about the arrivals and departures was not carried out. We can say that the data was processed with only an accounting goal: the number of people who passed and the total accumulated revenue.

The first thing we wanted to investigate is the mode of using the road. According to the general ideas, people go to work in the morning and return back in the evening. Accordingly, we should see a peak at the entrances in the morning at some stations, then (with a delay for the duration of the trip) - the peak in the number of exits at other stations. In the evening, the picture should be reversed.

Here are the hourly figures of the passes for the station outside Moscow. The upper part (Fig. 3) presents entrances (departures from the station) and the lower one (Fig. 4) presents exits (arrivals).

For all images below, the X-axis represents the time, the Y-axis represents the number of passengers entering or leaving a particular station. All the graphs for a particular station represent an average pattern for using this station on a working day or a weekend. Before the presentation of this template, the time series similarity for different days of the week was checked (Section 3).

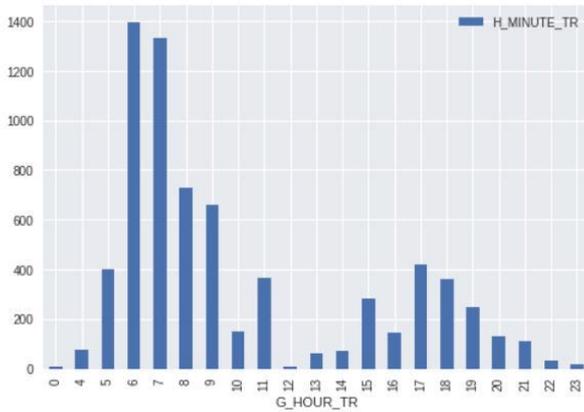


Fig. 3. Outside of Moscow entrances

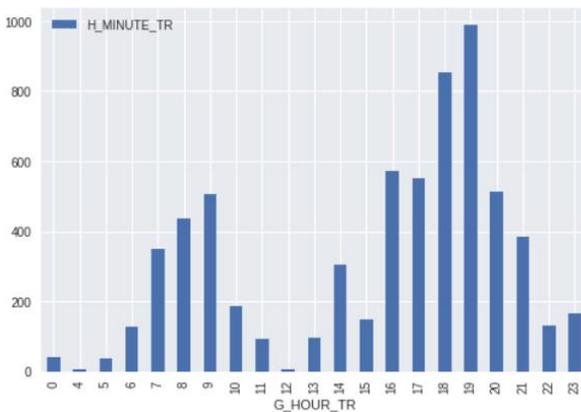


Fig. 4. Outside of Moscow exits

The picture seems to meet expectations. At 6-7 am, passengers are sent to work (peaks on the first graph - entrances to the station), at 6-7pm hours - they are returned (peaks on the second graph are exits). But let us see the same graphs for the station inside Moscow.

At the entrances, we already have two peaks - in the morning and in the evening. In the morning, people go to work (the station is used as a subway station), in the evening - they leave this station on the way home to the suburb. The morning peak is shifted relative to the "regional" for one hour - people are closer to work and can leave home later. The evening peak is at the same time, since the end time for work is the same.

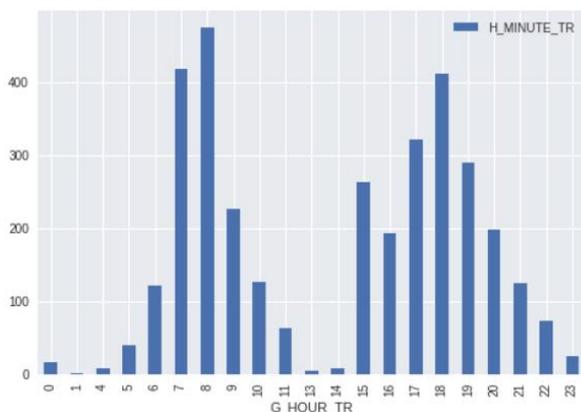


Fig. 5. Entrances inside of Moscow

Note the very low level at the middle of the day (1-2 pm). Partly, it is linked to so-called technological break (lack of trains). Immediately after the break, there is quite a lot of

demand. For stations outside the city, we do not see such demand. This suggests, that the proposed cancellation of a break on new lines will be highly demanded within the city. Technically, without this interruption, the city rail line will function as a metro line with the regular intervals between trains.

The picture on the exits looks similar:

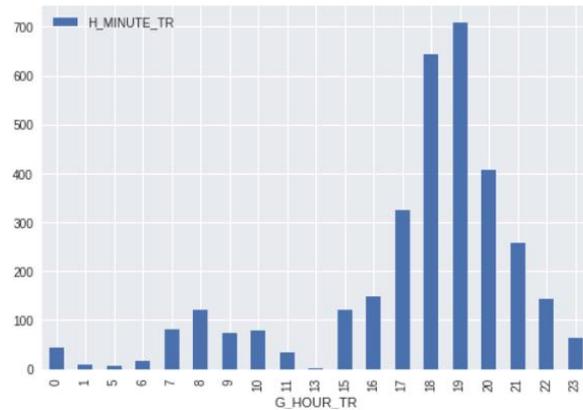


Fig. 6. Exits inside of Moscow.

But what is interesting is that the peak on exits is at the same time as in the suburb. Although the trip from work should take less time comparing with outside stations. Either there is greater mobility (something done after the work and before the boarding), or, taking into account the time on the road, people choose jobs with bigger travel time within the city than people from the suburb. In other words, there is some kind of constant for the time that people are willing to spend on the road to work.

On working days of the week, these distributions remain stable. But on weekends the picture changes.

People from the regional station continue to leave for work in the morning, but also they go to the city during the day, apparently already with some private purposes (Fig. 7).

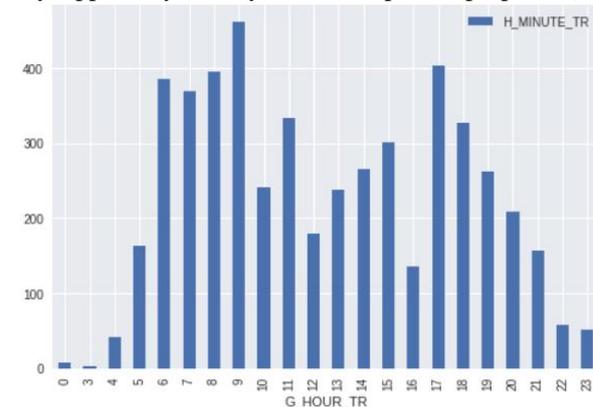


Fig. 7. Weekends entrances outside of Moscow

The departures last till 7-9pm hours. And here is the picture of the exits (arrivals to the station):

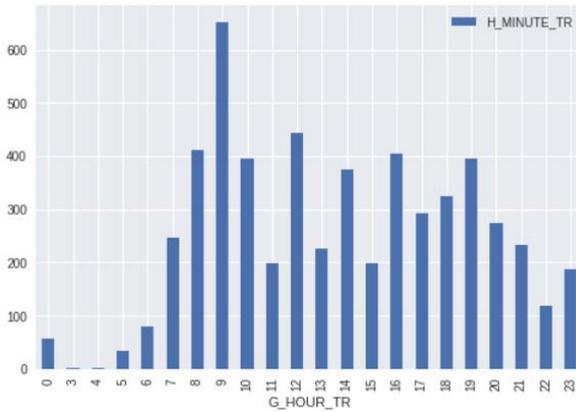


Fig. 8. Weekends exits outside of Moscow

Morning peaks on this chart are, apparently, summer residents or those who live on weekdays in the city and go home for the weekend.

At the station inside of Moscow, we see the morning traffic, which is determined by the people working at the weekend. It is interesting that the number of passengers is practically the same as on working days. Also, there are no daily dips - people continue to drive, unlike working days.

For stations inside of Moscow, there is no evening peak at the weekend. In other words, those working on weekends these days do not return home through this station, since they did this on weekdays (see the Fig. 5 above). One possible explanation is the use of cars. Alternatively, these people (their offices, businesses, for example, in the vicinity of the station) do not work on weekends. So, at the weekends, the stations inside of Moscow work more for passengers from Moscow. The flow of passengers to the suburb area (especially in the evening) decreases noticeably.

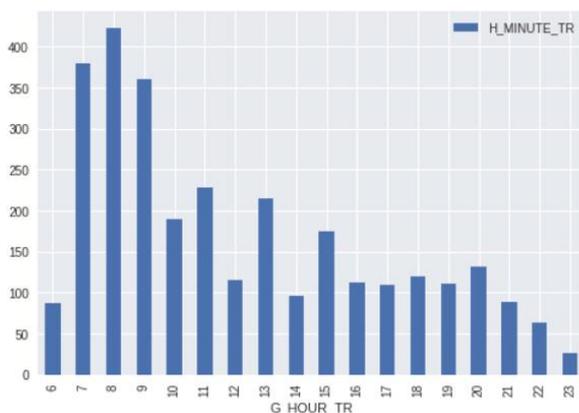


Fig. 9. Weekends entrances (departures) inside of Moscow

The picture on the exit (arrivals to the station) also changed. More than on weekdays it arrives in the morning and starting at 3 pm the departed locals come back.

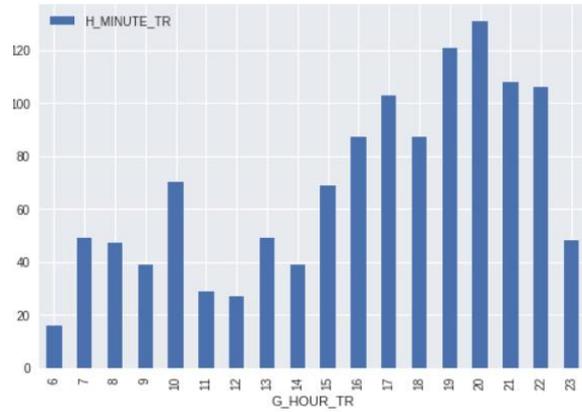


Fig. 10. Weekend arrivals inside of Moscow

Let us describe other elements of the movement that were investigated.

We can investigate the ratio of one-time and permanent tickets. Hypothesis: reusable tickets are cheaper. Consequently, a larger percentage of reusable tickets correspond to more constant traffic (constant flow).

Exactly the same picture can be obtained from the analysis of one-time tickets to one side versus back and forth tickets. The former correspond more to random traffic.

Also, we can investigate the asymmetry in the use of stations on the entrances and exits. For example, those leaving the station return back not from the station to which they originally left. Obviously, there is some kind of mobility of passengers, when they somehow move around the city before returning. This asymmetry can be associated with combined trips - a car and a railway.

Another interesting issue is the absence of peaks (arrivals, departures or even both). There are two possible explanations. For the stations outside of Moscow, it is probably the villa (dacha) zone. For stations within the city, it is associated with the former industrial zone, where the construction is not finished yet).

V. ON RELATED WORKS

The review of algorithms for analysis and recovery of routes based on validation of travel cards is contained, for example, in our paper [6].

In the paper [13], the authors use smart card records resolved in both time and space for getting collective spatial and temporal mobility patterns at large scales and reveal the regularity of these patterns. The main goals declared there are very close to our contribution: demonstrate the potential of using smart card records as data sources to gain insights into city dynamics and aggregated human behavior; explore the relationship between spatiotemporal patterns of smart card usage and underlying city behavior and geography; study patterns in smart card usage, including an analysis of how factors such as the time of the day affect this prediction. However, it should be noted that the work is trying to identify the individual patterns of displacement too. In our case, this is impossible, since there is no information on the

identification of travel documents.

In the paper [14], the authors target the fundamental dilemma: how to make an urban development less dependent upon mobility by car. Examples of locating transit patterns, including extensive literature analysis, are contained in the paper [15]. In the paper [16], authors describe check-in/check-out processing models for traffic prediction. Smart cards (travel cards) present the main source for data mining and predictions. The typical example is paper [17]. Authors provide in-depth temporal and spatial analysis for individual travel patterns, analyze the relationship between temporal and spatial features, and perform abnormal detection. Another good example is the paper [18]. The main idea here is to detect clusters for passengers with the similar transport behavior.

VI. CONCLUSION

The paper discusses the use of data on the use of railway stations in the Moscow agglomeration as a tool for analyzing transport behavior. The results outlined in this article have found practical application in the works on designing a new system of urban railways in the Moscow region. The paper investigated the relationship of the results of processing data on the use of railway stations with socio-economic aspects of the life of the inhabitants of the Moscow agglomeration. The authors propose the method for getting usage patterns for railway stations. This construction (allocation) of usage patterns of stations connected with work traffic is considered. The paper describes several usage patterns for railway stations inside and outside of Moscow. Also, the paper offers estimates of how changes in the city (for example, the construction of former industrial zones) will be reflected (respectively, can be tracked) in the modes of use of railway stations. As a basic analysis tool, similarity methods for time series and distributions were used.

ACKNOWLEDGMENTS

We would like to thank the staff for Center of digital high-speed transport systems Russian University of Transport (MIIT) for providing access to railways data.

REFERENCES

- [1] Namiot, Dmitry, et al. "Pedestrians in the Smart City." *International Journal of Open Information Technologies* 4.10 (2016): 15-21.
- [2] Kupriyanovsky, Vasily, et al. "On intelligent mobility in the digital economy." *International Journal of Open Information Technologies* 5.2 (2017): 46-63.
- [3] Kupriyanovsky, Vasily, et al. "Intellectual mobility and mobility as a service in Smart Cities." *International Journal of Open Information Technologies* 5.12 (2017): 77-122.
- [4] Namiot, Dmitry, Oleg Pokusaev, and Varvara Lazutkina. "On passenger flow data models for urban railways." *International Journal of Open Information Technologies* 6.3 (2018): 9-14.
- [5] Namiot, D. and Sneps-Sneppe, M., 2011. Customized check-in procedures. In *Smart Spaces and Next Generation Wired/Wireless Networking* (pp. 160-164). Springer, Berlin, Heidelberg.
- [6] Namiot, Dmitry, and Manfred Sneps-Sneppe. "A Survey of Smart Cards Data Mining." In *Supplementary Proceedings of the Sixth International Conference on Analysis of Images, Social Networks and Texts (AIST 2017) Moscow, Russia, July 27 - 29, 2017*
- [7] Steenbrugge, John, et al. "Mobile phone data from GSM networks for traffic parameter and urban spatial pattern assessment: a review of applications and opportunities." *GeoJournal* 78.2 (2013): 223-243.
- [8] Ratti, Carlo, et al. "Mobile landscapes: using location data from cell phones for urban analysis." *Environment and Planning B: Planning and Design* 33.5 (2006): 727-748.
- [9] Google Collaboratory <https://research.google.com/colaboratory/unregistered.html> Retrieved: May, 2018
- [10] Gunopulos, Dimitrios, and Gautam Das. "Time series similarity measures and time series indexing." *Acm Sigmod Record*. Vol. 30. No. 2. ACM, 2001.
- [11] Ding, Hui, et al. "Querying and mining of time series data: experimental comparison of representations and distance measures." *Proceedings of the VLDB Endowment* 1.2 (2008): 1542-1552.
- [12] Nakamura, Tetsuya, et al. "A shape-based similarity measure for time series data with ensemble learning." *Pattern Analysis and Applications* 16.4 (2013): 535-548.
- [13] Liu, Liang, et al. "Understanding individual and collective mobility patterns from smart card records: A case study in Shenzhen." *Intelligent Transportation Systems, 2009. ITSC'09. 12th International IEEE Conference On*. IEEE, 2009.
- [14] Bertolini, Luca, and Frank Le Clercq. "Urban development without more mobility by car? Lessons from Amsterdam, a multimodal urban region." *Environment and planning A* 35.4 (2003): 575-589.
- [15] Kusakabe, Takahiko, and Yasuo Asakura. "Behavioural data mining of transit smart card data: A data fusion approach." *Transportation Research Part C: Emerging Technologies* 46 (2014): 179-191.
- [16] Li, Yexin, et al. "Traffic prediction in a bike-sharing system." *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2015.
- [17] Zhao J. et al. Understanding temporal and spatial travel patterns of individual passengers by mining smart card data //Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on. – IEEE, 2014. – C. 2991-2997.
- [18] Agard B., Morency C., Trépanier M. Mining public transport user behaviour from smart card data //IFAC Proceedings Volumes. – 2006. – T. 39. – №. 3. – C. 399-404.
- [19] Müller, Meinard. "Dynamic time warping." *Information retrieval for music and motion* (2007): 69-84.
- [20] Keogh, Eamonn J., and Michael J. Pazzani. "Derivative dynamic time warping." *Proceedings of the 2001 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, 2001