On mobility patterns in Smart City

Dmitry Namiot
Lomonosov Moscow State University
MSU, Russia, 119991, Moscow, GSP-1, 1-52, Leninskiye Gory
dnamiot@gmail.com

Oleg Pokusaev
RUT (MIIT)
127994, Moscow, 9b9 Obrazcova Street
o.pokusaev@rut.digital

Abstract: The article is devoted to the analysis of transport data in the Moscow region, which was used in the design of new urban rail lines. Mobility (smart mobility) is one of the main characteristics of a smart city. Understanding how residents in a city actually move is necessary for the reliable provision of transport services, the construction of new transport lines, etc. Understanding patterns of displacement allows you to identify anomalies in urban displacements, which are evidence of some changes in the urban environment. The basis for the analysis is the data of mobile operators and information on the use of travel documents.

Keywords: transport modeling; geospatial flow; data mining.

1 Introduction

The paper considers the analysis of data on the movement of passengers in the Moscow metro. The basis for the analysis is the information collected from the data of mobile operators - the starting and ending stations of individual trips. Such data (the so-called correspondence matrix) exists in certain time sections and, theoretically, describes all the characteristics of the passenger load of the metro lines. Currently, they are used, for the most part, to calculate simple statistics. For example, such data is used mostly for getting the total number of passengers passing through the station. Naturally, at the same time, all the space-temporal information that characterizes the trip is lost (it is simply not used). And these are very important characteristics. The transport system in Moscow is constantly evolving. This, in particular, is expressed in the fact that the metro (as the main mode of transport) is integrated with other modes of transport (for example, the urban railway). There is a redistribution of traffic flows and passenger traffic. Another reason for the change in passenger traffic is high-rise housing. All this makes it necessary to understand how the passenger traffic in the subway is organized, how they are distributed over time, what can happen if the flow increases, is there any reserve for bandwidth and so on. A simple empirical assessment of the fact that usually, a certain number of passengers use this transport per day or per month becomes insufficient. This article is an attempt to describe data analysis models that may be applicable for more meaningful analysis of such data. First of all, the main goal is to describe the temporal characteristics of the movement of passengers.

The scheme of use of the Moscow metro involves the payment of fare at the entrance to the station. In terms of social networks, this is the so-called check-in [1]. There are no marks on exit. If payment cards are used for payment (smart cards in English-language literature), then, knowing the card number (address), you can try to restore the route. All ideas for trips recovery are based on the sequential use of the card. If some card, for example, was used to pay at station A, and then at station B, then we can assume that there is a route from station A to some transfer point closest to B, making a transfer and continuing the route from station B. If there is interruption in the use of a payment card, it can be assumed that there were two routes: from station A to the station that is closest to B (closest to the next use), and then - from station B to the final destination. Depending on the time of day, it can be assumed that the gap in usage is linked with work / study (in case of daytime) or stay at home (evening and night time). Thus, we can detect the place of work and residence [2].

A station in these examples could be also presented by any point of payment in the transport. E.g. a payment terminal in some bus is also a “station” for this algorithm.

This approach to the restoration of routes in the cities devoted a lot of work [3]. It is quite a working tool, its applicability depends, naturally, on the level of penetration of fare collection tools, which allow collecting statistics [4]. The result of this analysis is a matrix that describes the number of people moving from point A (or a district of a city) to another point B (another area of a city). In the English literature, this is called the origin-destination matrix (OD matrix). In the Russian papers, this is usually called the correspondence matrix. In particular, our own work [2] is dedicated to OD matrices.

Let us consider the transport network as a planar graph:

G* = *(V, E), where ***V*** is the set of vertices, E is the set of edges of the network

A certain transport hub (station) may be associated with each vertex as a place of departure (source) or arrival (drain) of passengers. Let us consider:

$\boldsymbol{O} \in \boldsymbol{V}, \boldsymbol{D} \in \boldsymbol{V}$ – are sets of vertices of the graph, which can be called, respectively, sources and drains of the network.

Then the correspondence matrix in general is:

$$\boldsymbol{p(t)} = \{p_{ij}(t), i \in \boldsymbol{O}, j \in \boldsymbol{D}, t \in \boldsymbol{T}\}$$

It determines the distribution of passenger traffic in the network and can be characterized, for example, by the number of passengers who have moved from area i to area j per unit of time t. The correspondence matrix is considered as an enlarged transport model describing some topology of the city's transport network or agglomeration. This matrix serves as the basis for building a detailed distribution model for traffic flows.

A similar scheme will work for the Moscow metro, given the prevalence of Troika payment cards. However, for Moscow (the Moscow metro), we can use another approach.

The support of mobile communications in the Moscow metro means that a telecom operator can collect data about its subscribers in the same way as it is done in other modern digital urban projects. The events that the operator registers include, naturally, the events of the change of base stations. Knowing their location, you can translate this information into using metro lines. Thus, mobile operators without access to payment card data can determine the entry point to the subway and exit point. Accordingly, assuming that the passenger chooses the shortest route, you can generate data about the trip (from where and where).

To preserve privacy, operators not only remove data on mobile devices from this information but also produce a temporary grouping of data. As a result, for a certain given time interval, we get the total number of passengers who traveled from station A to station B. There is no way to separate individual trips from this information. As a result, for a specific time slice, we have a square matrix in which the columns and rows denote stations, and the values indicate the number of passengers moving from one station to another. Such calculations are repeated with a given frequency (for example, 15 or 60 minutes). The result is a time-dependent sequence of OD-matrices (correspondence matrices)

At the present time, these data (in Moscow area) are used either to verify financial reports (for example, the total number of passengers passing through specific stations), or as baseline data in traffic flow modeling problems (as passengers entering the station get to it). It seems that with this approach the spatial and temporal aspects of traffic (metro usage) disappear. This article attempts to present a flow analysis model (mobility patterns discovery) that takes into account spatio-temporal aspects of traffic.

Most of the available literature (research) is focused on building the OD matrix [5, 6]. In our case, as was shown above, the matrix is already there, and the task is in its analysis.

The remainder of the article is structured as follows. In Section 2, we discuss metrics and approaches to classifying the use of stations. In section 3, we stop at the traffic analysis.

2 On metrics for classifying the use of metro stations

The idea is to convert the source data into a set of metrics that would allow comparing stations usage with each other.
Metrics can be static and dynamic. For example, average path length, clustering coefficients, robustness, efficiency, passenger flow, etc. [7, 8]. To select the metrics we used the results of our previous work [9].

In particular, the use of the notion of centricity was proposed to evaluate the transport system. An example of such an analysis is in [10]. Initially, the concept of centricity was first introduced in [11] to measure the importance of a node in large social networks that were not fully connected. Another example of the use of this concept for the analysis of traffic flows is the work [12]. The idea is that the structural properties of the network determine the flow through the network.

The centricity for a node *v* (it is a station, in our case) from graph *V* is defined as:

$$b_c(v) = \sum_{s,t \in V, s \neq t} \frac{\sigma(s \to t|v)}{\sigma(s \to t)} \tag{1}$$

where *σ (s−> t / v)* is the number of shortest paths from station *s* to station *t* passing through *v*, and *σ (s−> t)* is the total number of shortest paths from *s* to *t*.
It could be normalized by dividing by the total sum of centricity for all stations so that the distribution of values can then be easily compared against the distribution of throughput.

$$\hat{b}_c(v) = \frac{b_c(v)}{\sum_{v \in V} b_c(v)} \tag{2}$$

And according to the study, the traffic passing through the station correlates with its centricity. The more shortest paths through the station exist, the more through it drives passengers. This makes it possible to predict the volumes of passengers, since the characteristics of the nodes can be calculated statically (based only on the network topology).

This concept has a completely intuitive and clear interpretation. It is clear, for example, that the central transfer hub is more important for travel than the end station of some radius. Our centric calculations confirm this. To calculate the duration of trips, Yandex.Metro data on travel time between stations was used.

Here an important practical moment arises. In addition to the expansion of the metro itself, it is expanding due to the appearance of connections (transfers) with the urban railways (Moscow central ring, Moscow diameters – Fig. 1). Accordingly, new transfer hubs and new short routes appear, which will include metro stations and city railway stations.

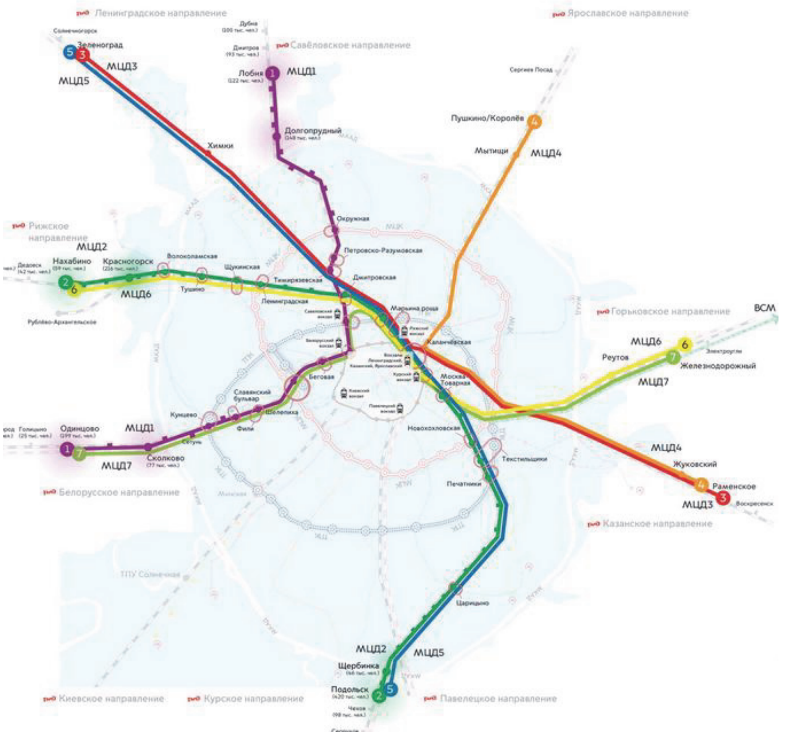

Fig.1. Urban railways [9]

This is a very important point. The commissioning of new transport routes is always associated with an assessment of the traffic flow for them. Urban railways, firstly, will attract new users from suburban areas, and secondly, they will cause the redistribution of metro traffic flows. So this redistribution and can be assessed by changes in the centrality of stations. Accordingly, it is possible to calculate the change in traffic after entering transfers. If we normalize the data of centricity, then we can estimate the weight of the “unit” of centricity in passengers. After that, knowing the changes in centricity, we will be able to estimate changes in traffic.

Using the same data on the duration of trips, it is possible to construct the distribution of trips by duration. This is illustrated for the two lines in the Fig. 2.

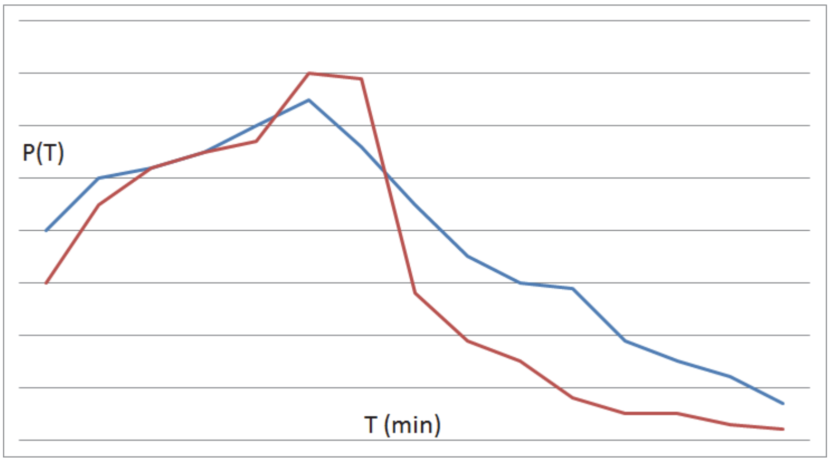

Fig. 2. Duration of trips probability

Here we can note the following. Our calculations show that the duration of trips (distribution of probabilities) differ not only in the whole metro lines but also in the initial stations. For different initial stations, the duration of the trips is different. We concluded that the estimate of the duration of trips depending on the starting station is more correct than the same estimate for the entire metro line. One of the hypotheses - longer trips are typical for stations where there is a transfer from railway lines. Passengers are transferred to the metro system at the earliest opportunity and, accordingly, they travel more by metro.

Also, different lengths of trips are recorded for weekdays and weekends. Here another feature is revealed. In general, passenger traffic is fairly stable along the lines of the metro during weekdays and falls on weekends. Moreover, this fall is absolutely not uniform. On some lines, the flow during the weekend falls less. The explanatory hypothesis is the location on these lines of points of attraction, characteristic for weekends (recreation areas) (Fig. 3). This shows the drop in traffic on weekends as a whole on the metro line. Also, an open issue for further research is the drop in traffic on weekends for individual stations. At least, as will be indicated below, there are stations where traffic is kept during the weekends at about the same level as at work days.

To analyze the stability of flows, we used known approaches to the analysis of the similarity of time series [13]. 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 [14] 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). DTW is one of the often used algorithms for measuring the similarity between two temporal sequences, which may vary in speed.

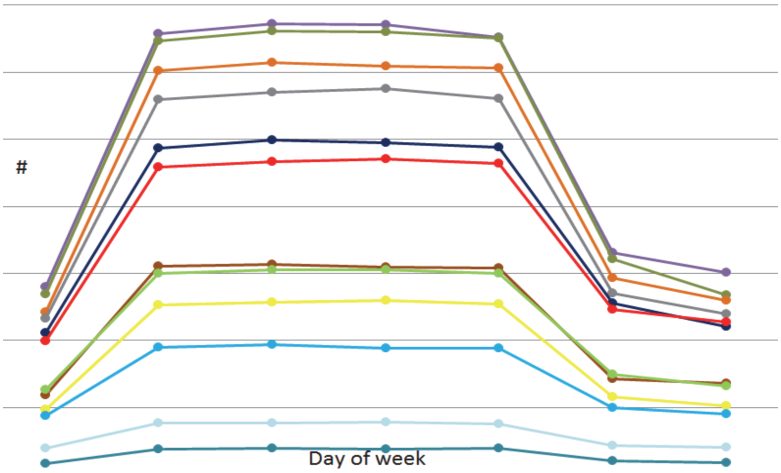

Fig. 3. Traffic for metro lines by days of the week

Also, it is seen that it can be used in a partial shape matching application. In our case, the speed is always the same and sequences always have an equal size. Most measures (metrics) are dealing with individual data points composing the compared time series. There is a so-called derivative DTW, 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 [16].

In our work, we have successfully used a shape-based similarity measure - Angular Metric for Shape Similarity (AMSS) [15]. 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. 4.

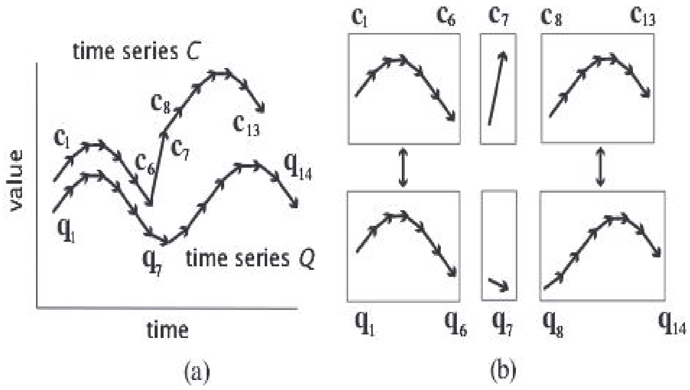

Fig. 4. Angular Metric for Shape Similarity [15]

As it is mentioned in our previous paper [16], where the detailed discussion on time series similarity is provided, the cosine similarity metrics minimize the influence of outliers in similarity computation. And of course, outlines present a big issue for transport data.

As the next step, we investigate the patterns of the inflow and outflow at each station with the average counts in each timeslot. In this connection, we also followed to the ideas, presented in the paper [16].

We have found that all stations could be split into several groups according to their flow patterns. E.g.:

1) The stations where the inflow shows one peak during the morning hours and the outflow shows one peak during the evening hours (time of the day and amount of passengers -Fig. 5). The explanation is obvious. These stations are located in the residential areas (sleeping areas in local language). Most of the travelers are commuters in these areas. Usually, they depart from home to work in the morning and return back to home in the evening.

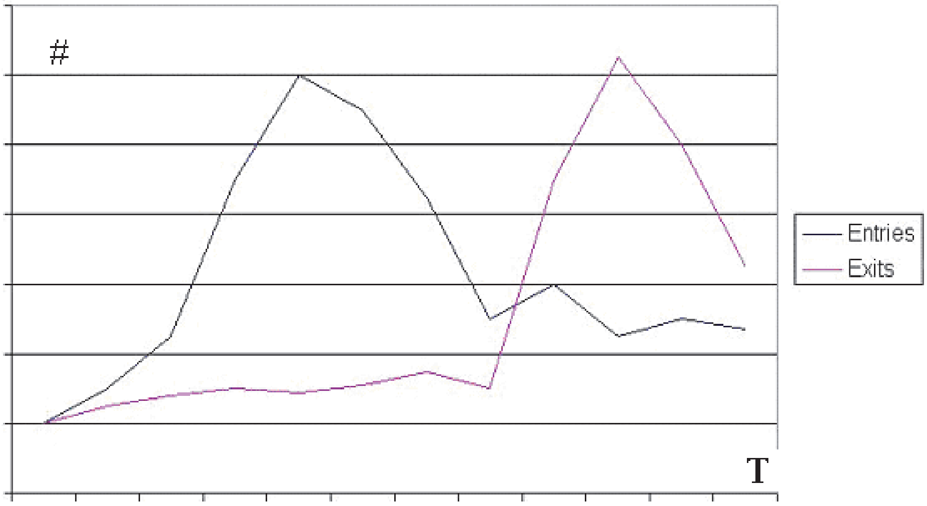

Fig. 5. Home – work commute

Usually, for such stations, there is a significant drop in traffic on weekends. Also, apparently, the previously noted sign of the width of the morning peak works. A wide (longer in time) morning peak is in favor of some kind of external "transportation" of passengers. For example, residents of the area are transplanted here by metro. Buses and electric trains, on which passengers arrive, are naturally stretched in time, the morning rush hour extends from here to time.

2) Stations where morning and evening peaks are observed at the entrance and exit. The explanation is also obvious - these stations are located in areas that are both working and residential. In the morning someone leaves for work but others arrive at the same time.
3) Stations where the morning peak at the exit is shifted in time (later than the peak at the time of arrival at work). A possible explanation is some commercial centers. Buyers arrive later.

4) The absence of pronounced peaks at the entrance and exit. As a rule, for such stations, this pattern is saved and the weekend.

In this paper, we used also the results obtained in [25]. In Fig. 6 we show the intensity of the flow of passengers to each station during each hour of the day (for clarity, the values for the whole month were summed up):

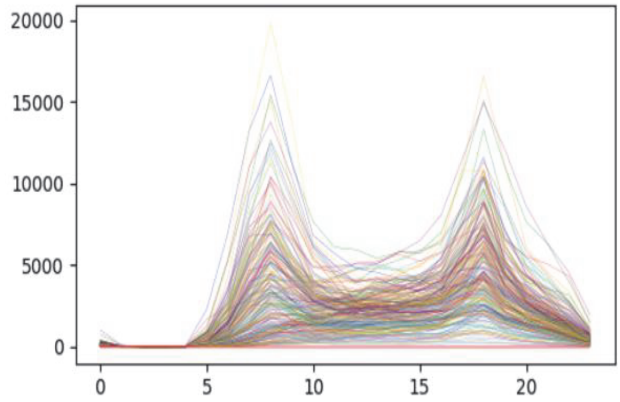

Fig. 6. The intensity of the flow of passengers to different metro stations

It is clearly seen that the largest load on the transport system is in the so-called rush hours. At the same time, the values of these very peaks at different stations are very different. Stations can be divided into three broad categories (see Fig. 7). The first station will be assigned to the station with a strongly prevailing morning peak, the second - with the evening one, and the third - those whose peaks are approximately the same. Such a breakdown clearly shows the reason for the difference in dynamics, namely, the specifics of each separate part of the city:

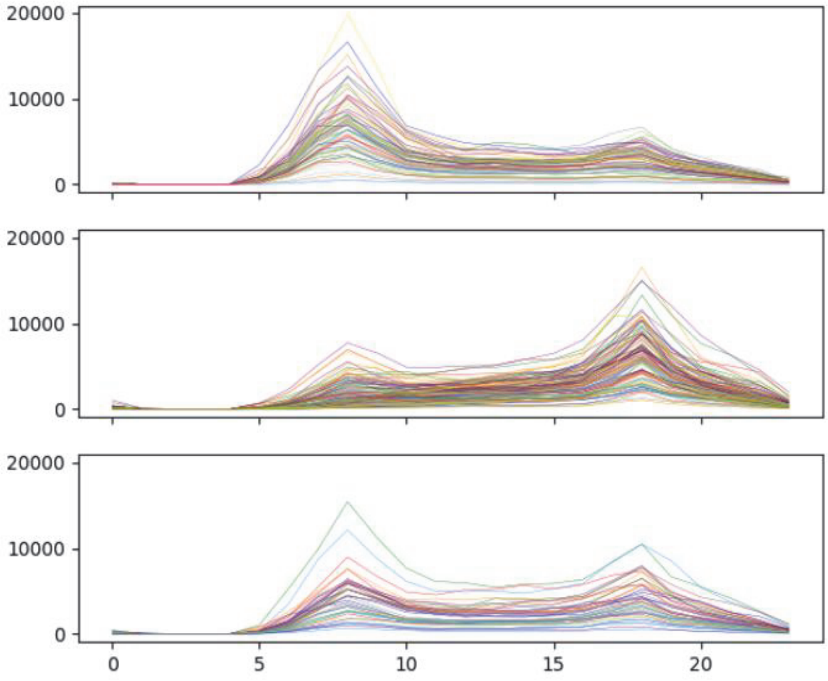

Fig. 7. On stations splitting by peak ratio [25]

it is clear that in the "sleeping" (dormitory) areas the morning peak, when people go to work or study will be more; and in those parts where there are business centers, educational institutions, enterprises the evening peak will prevail.

If, in accordance with this division, the metro scheme is divided into three colors (Fig. 8), then it becomes clear why these categories of stations are distinguished: the morning peak prevails in sleeping areas, and the evening one where business centers, educational institutions, etc. are located. d. Thus, the blue color, for example, is a description of the boundaries of the working area of the city. The prevailing flow here (the pattern of use of stations) is as follows: passengers exit the station more in the morning (go to work), and enter more in the evening (go home). The housing zone (red color) and the mixed type zone (green) are similarly distinguished. However, zoning, of course, depends on the coefficient for splitting in Fig. 7. In this case, the peak is considered to be predominant if the passenger traffic on it exceeds the passenger traffic at another peak by more than 1.5 times.

These studies have confirmed the practical importance of the classification of transport facilities according to the ratio of input and output flow during peak hours and during the day. According to similar ideas, we classified railway stations in the Moscow region [16]. The advantages of this approach are obvious:

- we solve fairly well-known clustering problems. There are many effective algorithms here
- there is a clear algorithm for the formation of features for clustering
- interpretation of the results obtained is completely transparent. The results are easily explained and verifiable
- the groups obtained as a result of clustering are the directly used data (conclusions) for urbanists.

The resulting classifications can be used as basic metrics to track changes in the zoning of a city. For example, our studies show very high flow stability (Fig. 3), naturally, with the available seasonal factor. Changes in the classification will indicate changes in traffic flows (act as a signal of such changes), and, accordingly, are a signal to search for explanatory factors from the point of view of urban policy, transport systems, etc.

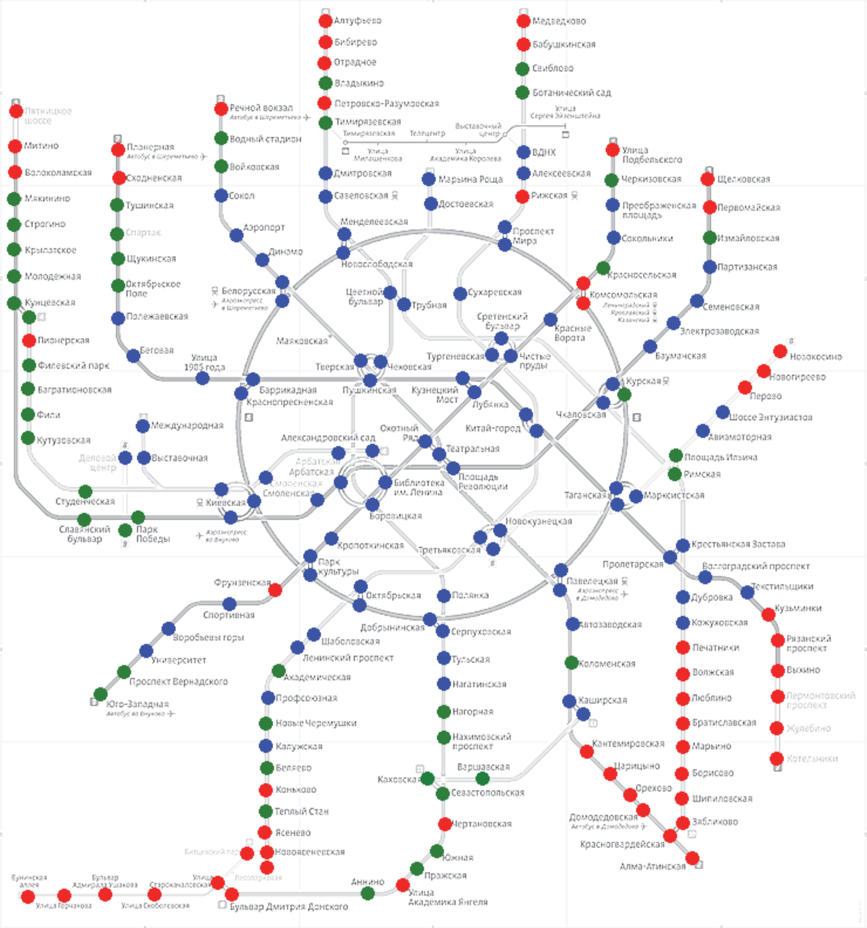

Fig. 8. Moscow’s zones by peak size [25].

3 On traffic analysis

The relationship between every two origin-destination pairs in a complex network, especially the correlation among spatially neighboring OD-pairs, has always been a research interest [17]. There is a possibility that related OD flows could be used as potential features to improve the prediction. One interesting idea is to investigate the correlation of the input traffic at the end station and the exits at subsequent stations for the nearest possible change of lines. Our initial results show, for example, that for some lines there is no dependence between the inputs at the end station and the outputs at subsequent stations until a possible change. This means that for most passengers the endpoints of their routes are already behind the change points along this line. For example, so look trips on the north of Moscow.

Analysis of the literature on the study of traffic shows two main tasks. They are flow prediction and OD forecasting [18]. In the flow prediction tasks, we estimate passenger counts at selected stations regardless of where they come from or where their target points. In our research, we study two papers suggested neural networks for short-term forecasting [19, 20]. Unfortunately, our attempts to reproduce the presented results were unsuccessful. The accuracy of the forecasts was very low.

OD Matrix Forecasting predicts trip counts between the different stations. Here it is necessary to mention the various approaches associated with the analysis of time series. E.g., Kalman filters [21] and ARIMA (an autoregressive integrated moving average) model [22].

In the paper [23], the authors compare the different neural network architectures (the multilayer perceptron, bidirectional recurrent neural networks) for the task of destination prediction. In the paper [24], the authors used Recurrent Neural Networks (RNN) for OD prediction. In the paper [18], authors conclude that while RNNs are powerful tools capable of representing context, they are reported to show theoretical and experimental limitations for long-term dependencies. In our case, this is a deep (time long) dependence of current traffic on previous results, where the temporal contingency between predictors and outputs span over extended periods). By this reason, the authors suggested Long Short-Term Memory (LSTM) networks, which are free from this restriction.

In our research, we followed the model proposed in [17]. It uses a set of sequential flow measurements as independent variables for predicting data for the next time slot (Fig. 9).

Sequential flows have a simple explanation here. Since the neighboring stations, obviously, more often belong to the same area of the city with its passenger traffic dynamics, it can be assumed that the data at the neighboring stations may be an informative feature for traffic forecast.

The main idea in forecasting is that if at one station a surge of passenger activity begins (for example, the outflow of workers of a certain company in the evening or 'stuffing' of passengers of adjacent transport systems to extreme metro stations in the morning), then at neighboring stations, probably activity will also increase. To test this hypothesis, correspondence correlations were considered between the two stations at Moscow’s Nord, as well as all neighboring ones (3 nearest neighbors in each direction). It turned out that the passenger traffic of the first nearest neighbors, indeed, correlate with the target passenger traffic more strongly than all the others. Thus, it makes sense to take this information into account when building a short-term forecast. In addition, obviously, you need to take into account the historical data of passenger traffic.

These data exhaustively describe all passenger traffic, but at the same time, they contain a lot of redundant information, which later will lead to a decrease in the accuracy of the forecast. To avoid this, and also in order to reduce the computational complexity of the algorithm, we reduce the number of attributes using the principal component method — one of the most frequently used methods for reducing the dimension with minimal information loss.

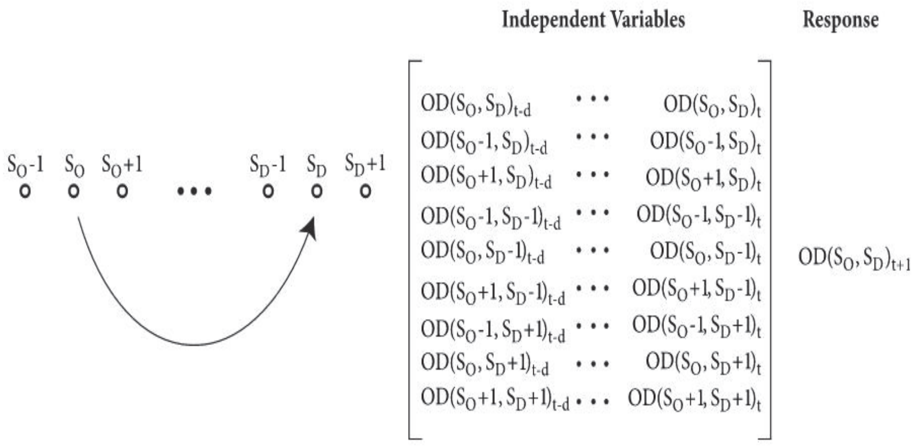
Independent Variables
Response
S_O-1 S_O S_O+1 ••• S_D-1 S_D S_D+1
OD(S_O, S_D)_t-d ••• OD(S_O, S_D)_t
OD(S_O-1, S_D)_t-d ••• OD(S_O-1, S_D)_t
OD(S_O+1, S_D)_t-d ••• OD(S_O+1, S_D)_t
OD(S_O-1, S_D-1)_t-d ••• OD(S_O-1, S_D-1)_t
OD(S_O, S_D-1)_t-d ••• OD(S_O, S_D-1)_t
OD(S_O+1, S_D-1)_t-d ••• OD(S_O+1, S_D-1)_t
OD(S_O-1, S_D+1)_t-d ••• OD(S_O-1, S_D+1)_t
OD(S_O, S_D+1)_t-d ••• OD(S_O, S_D+1)_t
OD(S_O+1, S_D+1)_t-d ••• OD(S_O+1, S_D+1)_t
OD(S_O, S_D)_t+1

Fig. 9 On model features [17].

Then the principal component analysis (PCA) is used to decompose the original OD flow into a number of principal components and scores. In this way, the systematic variations (outlines) in OD flows are captured in lower dimensions. As per the original paper, the weekdays and weekends are treated separately. As the prediction method, we’ve used kNN.

Technically, the proposed system on the window in 5 measurements (5 * 15 = 75 minutes) allows predicting the OD traffic for the next 15 minutes with an error (MAPE - Mean Absolute Percentage Error) of 25% (slightly more than in the above-mentioned paper [17]). The negative point that emerged during testing was that the accuracy of the prediction strongly depends on the time of day. At the beginning of the day, accuracy drops dramatically. Namely, the morning hours are the most unstable traffic. Accordingly, this is a matter of further research.

4 Conclusion

The paper deals with the classification of metro stations in accordance with the identified patterns of their use. The paper also proposes a model for short-term traffic forecasting based on data from the correspondence matrix. In this article, we have proposed several approaches to the analysis of data on movements in the Moscow metro. In our study, we were based on data collected by mobile operators. This data is a correspondence matrix for periodic time slices (every 15 minutes).

We considered two main tasks. First, it is the classification of metro stations by type of use. There are two possible options. This is a static model that splits stations by the number of shortest paths passing through them (the so-called centrality). The dynamic model uses real travel data and classifies stations according to the ratio of inputs and outputs during peak hours and weekends. It is this approach that made it possible to obtain an interesting classification of metro stations by the type of their use. The resulting classification can also be used to monitor changes in traffic flows in the city. The second task considered is the prediction of passenger traffic at short time intervals. The principal possibility of predicting station traffic based on the traffic of neighboring stations is shown.

Acknowledgement

We would like to thank the staff of the Center for Digital High-Speed Transport Systems of the Russian University of Transport (MIIT) for the data provided and extremely useful discussions and criticism in the process. Also, we are grateful to the participants of the scientific seminar of the Laboratory of Open Information Technologies of the Lomonosov Moscow State University for valuable discussions.

References

[1] Namiot, Dmitry, and Manfred Sneps-Sneppe. "Customized check-in procedures." Smart Spaces and Next Generation Wired/Wireless Networking. Springer, Berlin, Heidelberg, 2011. 160-164.

[2] Namiot, Dmitry, and Manfred Sneps-Sneppe. "A Survey of Smart Cards Data Mining." Supplementary Proceedings of the Sixth International Conference on Analysis of Images, Social Networks and Texts (AIST 2017) Moscow, Russia. 2017.

[3] Anda, Cuauhtemoc, Alexander Erath, and Pieter Jacobus Fourie. "Transport modelling in the age of big data." International Journal of Urban Sciences 21.sup1 (2017): 19-42.

[4] Gentile, Guido, and Klaus Noekel. "Modelling public transport passenger flows in the era of intelligent transport systems." Gewerbestrasse: Springer International Publishing (2016).

[5] Sherali, Hanif D., Arvind Narayanan, and R. Sivanandan. "Estimation of origin–destination trip-tables based on a partial set of traffic link volumes." Transportation Research Part B: Methodological 37.9 (2003): 815-836.

[6] Cascetta, Ennio, Domenico Inaudi, and Gerald Marquis. "Dynamic estimators of origin-destination matrices using traffic counts." Transportation science 27.4 (1993): 363-373.

[7] Derrible S, Kennedy C (2010) The complexity and robustness of the metro network. Physica A 389: 3678-3691. View Article Google Scholar

[8] Soh H, Lim S, Zhang TY, Fu XJ, Lee GKK, et al. (2010) Singapore public transportation system. Physica A 389: 5852–5863.

[9] 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.

[10] Ramli, Muhamad Azfar, et al. "A method to ascertain rapid transit systems' throughput distribution using network analysis." Procedia Computer Science 29 (2014): 1621-1630.

[11] Freeman, Linton C. "A set of measures of centrality based on betweenness." Sociometry (1977): 35-41.

[12] Altshuler, Yaniv, et al. "Augmented betweenness centrality for mobility prediction in transportation networks." International Workshop on Finding Patterns of Human Behaviors in NEtworks and MObility Data, NEMO11. 2011.

[13] Gunopulos, Dimitrios, and Gautam Das. "Time series similarity measures and time series indexing." Acm Sigmod Record. Vol. 30. No. 2. ACM, 2001.

[14] 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.

[15] 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.

[16] Namiot, Dmitry, Oleg Pokusaev, and Vasily Kupriyanovsky. "On railway stations statistics in Smart Cities." International Journal of Open Information Technologies 7.4 (2019): 19-24.

[17] Dai, Xiaoqing, Lijun Sun, and Yanyan Xu. "Short-Term Origin-Destination Based Metro Flow Prediction with Probabilistic Model Selection Approach." Journal of Advanced Transportation 2018 (2018).

[18] Toqué, Florian, et al. "Forecasting dynamic public transport origin-destination matrices with long-short term memory recurrent neural networks." 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2016.

[19] Y. Chen, H. S. Mahmassani, and Z. Hong, "Data mining and pattern matching for dynamic origin–destination demand estimation: Improving online network traffic prediction," Transportation Research Record: Journal of the Transportation Research Board, no. 2497, pp. 23–34, 2015.

[20] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach," Transportation Research Part C: Emerging Technologies, vol. 13, no. 3, pp. 211–234, 2005.

[21] X. Chen, S. Guo, L. Yu, and B. Hellinga, “Short-term forecasting of transit route od matrix with smart card data,” in 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), Oct 2011, pp. 1513–1518.

[22] E. Van der Hurk, L. G. Kroon, G. Maroti, and P. Vervest, “Dynamic forecast model of time dependent passenger flows for disruption management,” in 12th conference on advanced systems for public transport in Santiago, Santiago, Chile, 2012, pp. 23–27.

[23] A. de Br´ebisson, ´ E. Simon, A. Auvolat, P. Vincent, and Y. Bengio, “Artificial neural networks applied to taxi destination prediction,” CoRR, vol. abs/1508.00021, 2015.

[24] F. Qian, G. Hu, and J. Xie, “A recurrent neural network approach to traffic matrix tracking using partial measurements,” in 2008 3rd IEEE Conference on Industrial Electronics and Applications, June 2008, pp.1640–164.

[25] Nekraplonna, Mariia, and Dmitry Namiot. "Metro correspondence matrix analysis." International Journal of Open Information Technologies 7.7 (2019): 68-80.