Method of determining user preferences for the personalized recommender systems for public transport passengers

САМАРСКИЙ УНИВЕРСИТЕТ
SAMARA UNIVERSITY

A.A. Borodinov, V.V. Myasnikov

AIST 2019

Main problems of recommender systems

- The “Cold Start”.
- The receiving information method from the user is not formalized.
- Individual characteristics such as personal income, age, gender, numbers of family members, access to public transport influence the choice of the route for the same purpose of the trip.
- User preferences change over time. In addition, context influences user selection.
- Typical existing solutions mainly use the Bayesian approach with a sequential parameter recalculation scheme.
- It is possible to use transfer learning to improve recommendations.
- The problem of determining traffic flow on the vehicle route.

Basic definitions

Data from the mobile application "Pribyvalka-63“ (Figure 1):

- public transport stop information (identifiers and coordinates);
- public transport route information (identifiers and stop identifier list);
- information about the vehicle (identifiers), location coordinates (the vehicle transmits its coordinates two times per minute), route identifier;
- coordinates of users and request parameters are recorded with requests (request results are not saved since they can be restored from vehicle traffic data);
- user response to the request is not saved.

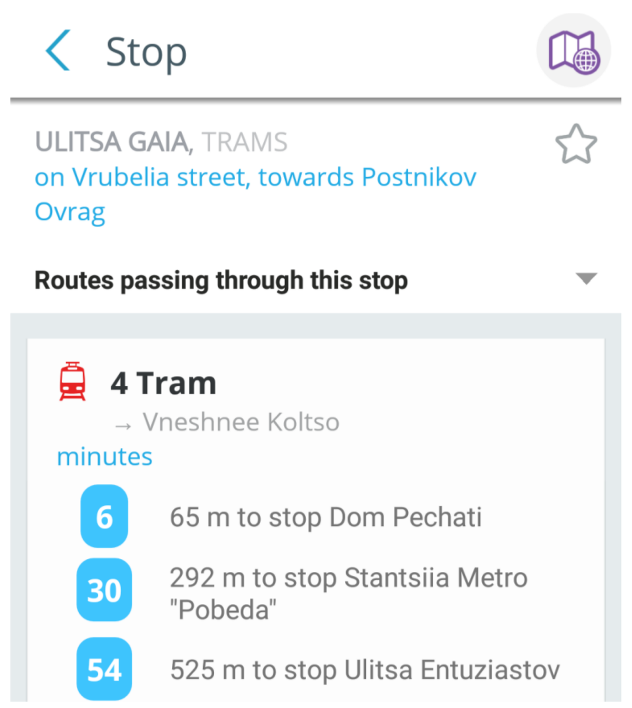

Fig. 1. Mobile application "Pribyvalka-63"

User preferences:

- user-preferred stops at specific space-time coordinates (Figure 2);
- user-preferred "transport correspondence", also considered in the space-time context. "Transport correspondence" refers to the actual movement from one stop to another, the route chosen and the route vehicle type (Figure 3).

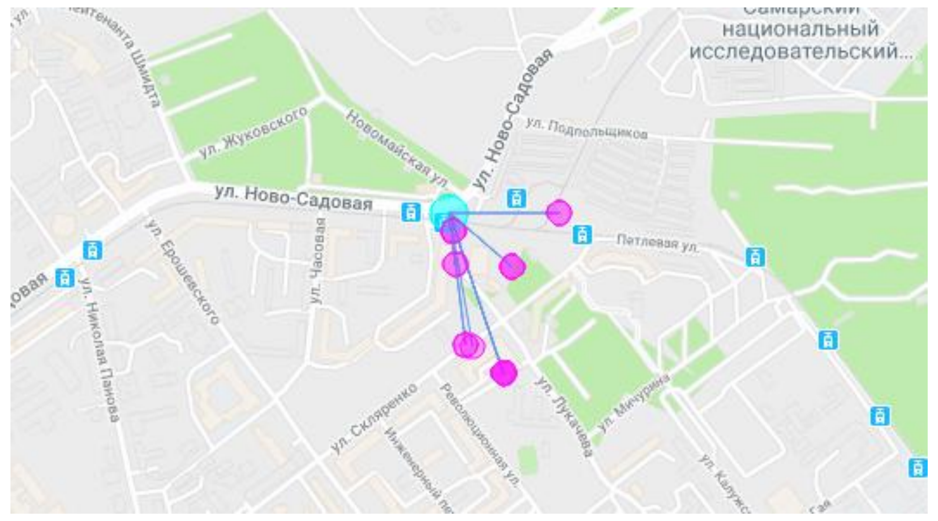

Fig. 2. Blue circle — stop location, purple circles — user location points at request time

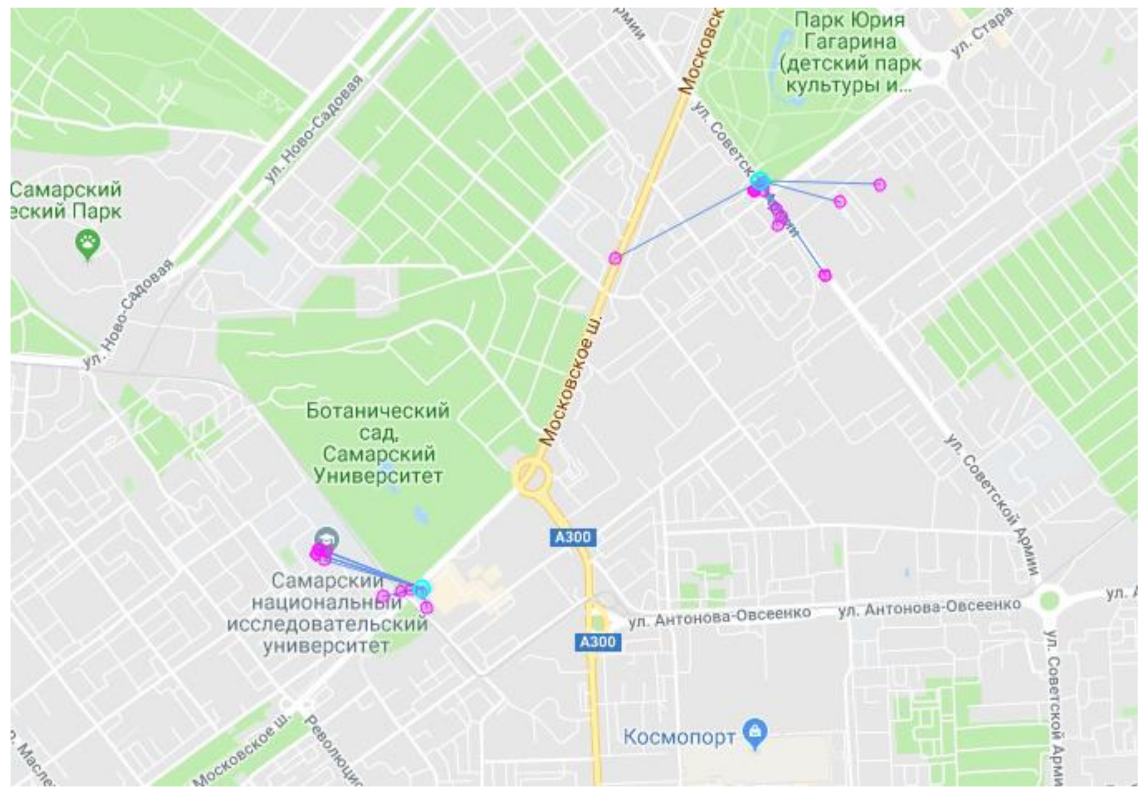

Fig. 3. "Start" and "End" stops of a specific user

User-preferred transport stops

Step 1. Calculate the values for all stops from the set S :

$$\Gamma\left(\mathbf{x},d,t;ID\left(s_i\right)\right),\quad i=\overline{1,|\mathrm{S}|}$$

$$\Gamma\left(\mathbf{x},d,t;z\right)=\sum_{i\in\Im}\mu\left(\mathbf{x},d,t;\mathbf{x}_i,d_i,t_i\right)I\left(z_i=z\right)$$

$$\mu\left(\mathbf{x},d,t;\mathbf{x}_i,d_i,t_i\right)=$$
$$=I\begin{pmatrix}\left(w(d)\in W_0\wedge w(d_i)\in W_0\right)\vee\\ \left(w(d)\in W_1\wedge w(d_i)\in W_1\right)\end{pmatrix}\cdot\exp\left(-\alpha\left|t-t_i\right|\right)\cdot\exp\left(-\beta\left\|\mathbf{x}-\mathbf{x}_i\right\|\right)$$

$$I(a)=\begin{cases}1, & a=true;\\ 0, & a=false.\end{cases}$$

Step 2. Sort by decreasing the set of values and get a permutation

$$\sigma:\mathbb{N}_{|\mathrm{S}|}\to\mathbb{N}_{|\mathrm{S}|}$$

User-preferred "transport correspondences"

- $p_u\left(t|s1,s2,m,\mathrm{W}_a\right)$ $\left(m\in\mathrm{M}(s1,s2),\quad a=\overline{0,1}\right)$ is a correspondence time distribution density *s1→s2* with the route choice *m*:

$$p_u\left(t|s1,s2,m,\mathrm{W}_a\right)=\frac{\sum\limits_{d\in\{d:w(d)=W_a\}}\ \sum\limits_{(i,\ell)\in I_{d,s1,s2,m}}\frac{1}{h_0}K_0\left(\frac{t-t\left(d,m_i^\ell,k_i^\ell,s_i^{start}\right)}{h_0}\right)}{\sum\limits_{d\in\{d:w(d)=W_a\}}\left|I_{d,s1,s2,m}\right|},$$

$$m\in\mathrm{M}(s1,s2),\quad a=\overline{0,1};$$

- $p_u\left(t|s1,s2,\mathrm{W}_a\right)$ is a correspondence time distribution density *s1→s2*:

$$p_u\left(t|s1,s2,\mathrm{W}_a\right)=\frac{\sum\limits_{d\in\{d:w(d)=W_a\}}\ \sum\limits_{(i,\ell)\in I_{d,s1,s2}}\frac{1}{h_0}K_0\left(\frac{t-t\left(d,m_i^\ell,k_i^\ell,s_i^{start}\right)}{h_0}\right)}{\sum\limits_{d\in\{d:w(d)=W_a\}}\left|I_{d,s1,s2}\right|};$$

- $P_u\left(s1,s2|\mathrm{W}_a\right)$ is a correspondence probability *s1→s2*:

$$P_u\left(s1,s2|\mathrm{W}_a\right)=\frac{\sum\limits_{d\in\{d:w(d)=W_a\}}\left|I_{d,s1,s2}\right|}{\sum\limits_{d\in\{d:w(d)=W_a\}}\left|I_{d,}\right|};$$

- $P_u\left(m|s1,s2,\mathrm{W}_a\right)$ is a probability of choosing the route *m* for implementing the correspondence *s1→s2*:

$$P_u\left(m|s1,s2,\mathrm{W}_a\right)=\frac{\sum\limits_{d\in\{d:w(d)=W_a\}}\left|I_{d,s1,s2,m}\right|}{\sum\limits_{d\in\{d:w(d)=W_a\}}\left|I_{d,s1,s2}\right|};$$

- $P_u\left(m|\mathrm{W}_a\right)$ is a probability of choosing the route *m* for implementing the correspondence:

$$P_u\left(m|\mathrm{W}_a\right)=\frac{\sum\limits_{d\in\{d:w(d)=W_a\}}\left|I_{d,m}\right|}{\sum\limits_{d\in\{d:w(d)=W_a\}}\left|I_{d}\right|};$$

- $P_u^*(s|\mathrm{W}_a)$ is a the probability that the stop s is the “end/start”:

$$P_u^*\left(s|\mathrm{W}_a\right)=\frac{\sum\limits_{d\in\{d:w(d)=W_a\}}\eta_u^*\left(s|d\right)}{\sum\limits_{d\in\{d:w(d)=W_a\}}\left|I_{d,s}\right|};$$

- $p_u(\tilde{\rho}|\mathrm{W}_a)$ is a distances distribution that the user is able to overcome without using route vehicles:

$$p_u\left(\tilde{\rho}|\mathrm{W}_a\right)=\frac{\sum\limits_{j\in\left\{j\in J_u:s_j^0\neq s_j^1\right\}}\frac{1}{h_1}K_1\left(\frac{\tilde{\rho}-\rho\left(\mathbf{x}\left(s_j^0\right),\mathbf{x}\left(s_j^1\right)\right)}{h_1}\right)}{\left|\left\{j\in J_u:s_j^0\neq s_j^1\right\}\right|};$$

- $p_u\left(\tilde{\tau}|\tilde{\rho},\mathrm{W}_a\right)$ is a time distribution that the user spends in overcoming the corresponding distance:

$$p_u\left(\tilde{\tau}|\tilde{\rho},\mathrm{W}_a\right)=\frac{\sum\limits_{j\in\left\{j\in J_u:s_j^0\neq s_j^1\right\}}\frac{1}{h_0}K_0\left(\frac{\tilde{\tau}-\left(t_j^1-t_j^0\right)\frac{\tilde{\rho}}{\rho\left(\mathbf{x}\left(s_j^0\right),\mathbf{x}\left(s_j^1\right)\right)}}{h_0}\right)}{\left|\left\{j\in J_u:s_j^0\neq s_j^1\right\}\right|}.$$

The problem is to calculate user-preferred "transport correspondences“ and calculate the statistical values based on data from the mobile application "Pribyvalka-63“. The algorithm for calculating the values is based Kernel density estimation via the Parzen-Rosenblatt window method.

Results

The database obtained during the experiments contains information about requests of 57190 users. Each user is represented by a unique impersonal identifier *ID*(*u*), which is defined by the device ID hash code. The database contains 4103161 user requests for an arrival forecast at a public transport stop. From 1478 stops of the tosamara.ru service, users made requests to 1417 stops.
Maps with different parameters $\alpha,\beta\in\mathbb{R}_+$ and request time were built to visualize the results of the proposed approach. The color of the area on the map corresponds to the first stop from the ordered list (Figure 4).

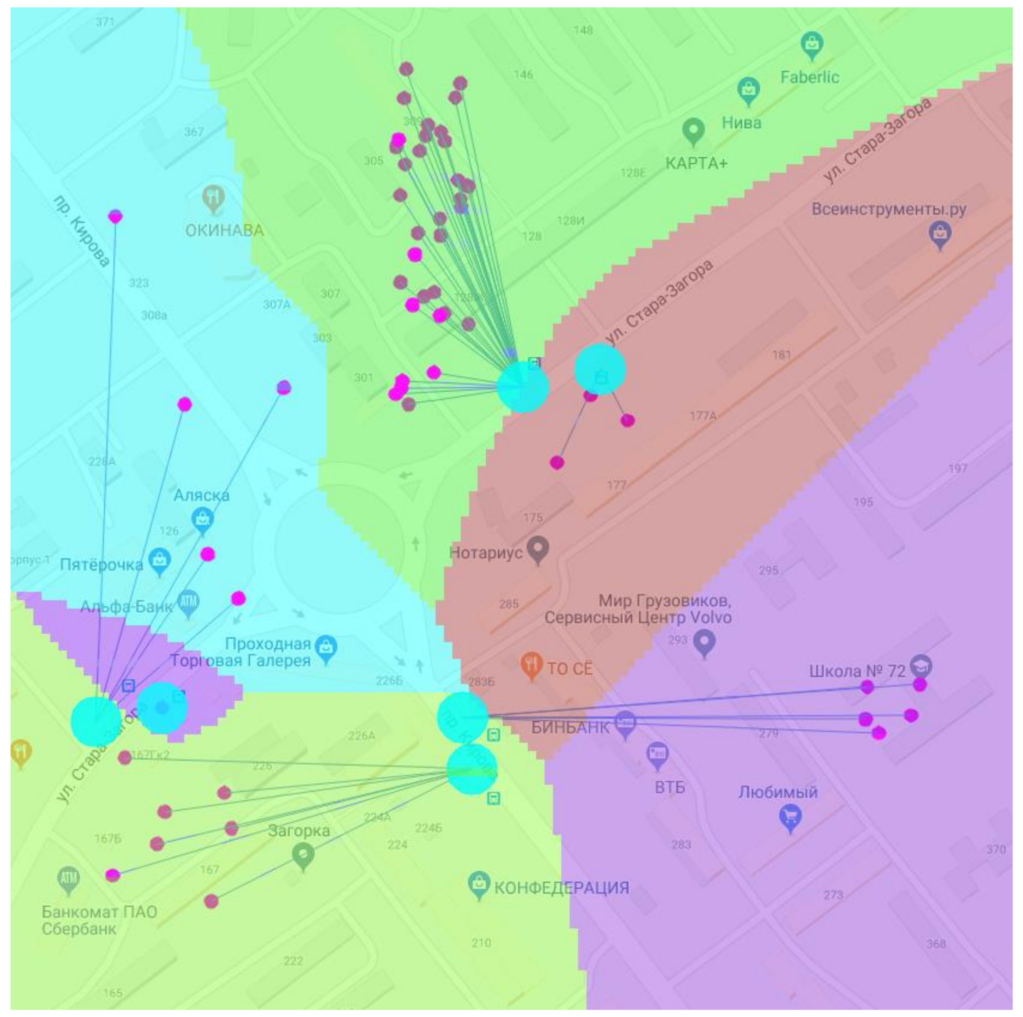

Fig. 4. Preferred stops map depending on user location.

Also, another method was implemented for comparison with the proposed algorithm, in which the user was offered the nearest stop, without taking into account previous requests. The proposed algorithm accuracy was 93%, while the nearest stop prediction resulted in 65% of this quality measure.

Acknowledgments

The work was funded by the Ministry of Science and Higher Education of the Russian Federation (unique project identifier RFMEFI57518X0177).

 ALEKSANDR BORODINOV– aaborodinov@yandex.ru VLADISLAV MYASNIKOV – vmyas@geosamara.ru