Analysis of the preferences of public transport passengers in the task of building a personalized recommender system

САМАРСКИЙ УНИВЕРСИТЕТ
SAMARA UNIVERSITY

A.A. Borodinov, V.V. Myasnikov

Recommender systems major problems

- The cold start.
- The receiving information method from the user is not formalized.
- Individual characteristics such as personal income, age, gender, family size, access to public transport influence the choice of the route even 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.

Definitions and designations

For the mobile service "Pribyvalka-63" data for analysis are presented as follows:

- stop data (identifiers and coordinates);
- route information (identifiers and stop identifier list);
- data on the vehicle (identifiers), location coordinates (with a frequency of 2 times per minute), the destination to routes;
- coordinates of users and request parameters are recorded during requests (request results are not saved, since they can be restored from vehicle traffic data) in the form: $ID(s), d, t, \quad ID(u), \mathbf{x}(u,d,t)$
- user response to the request is not saved.

Two options "user preferences":

- user-preferred stops at specific space-time coordinates (Figure 1);
- 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. Information about the "starting" and "end" stops of a particular user is optional (derived) information (Figure 2).

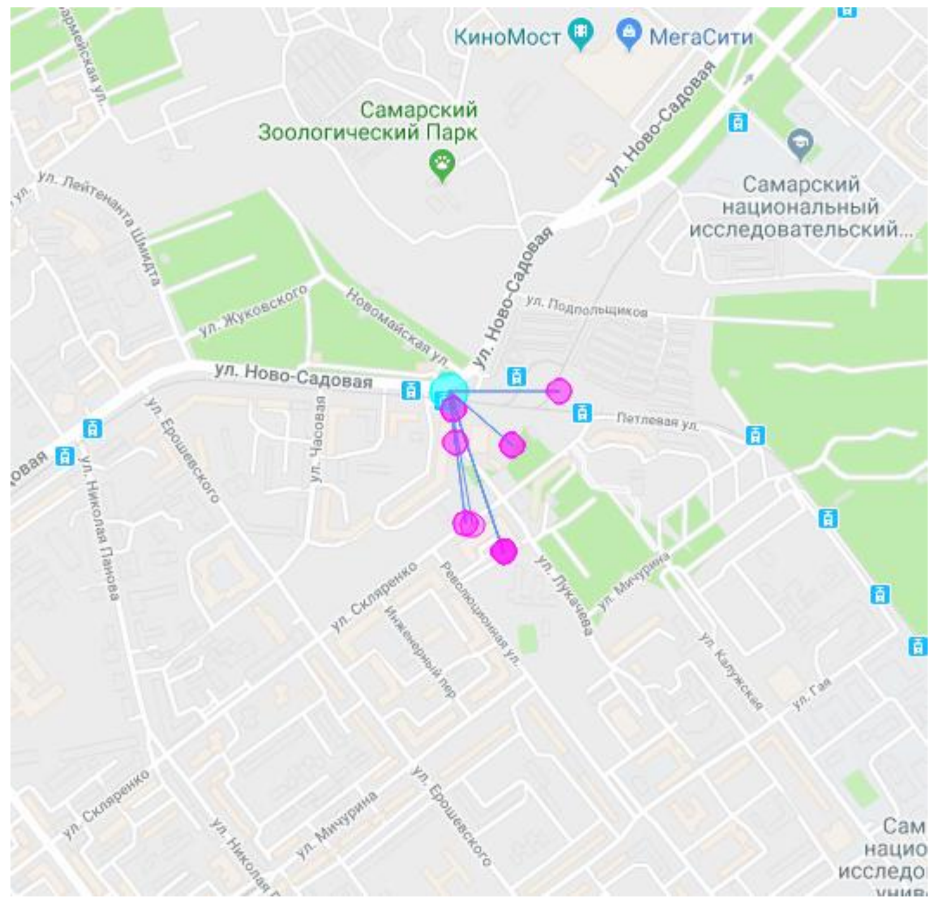

Fig. 1. Blue circle — stop location, purple circles — user location points at request time

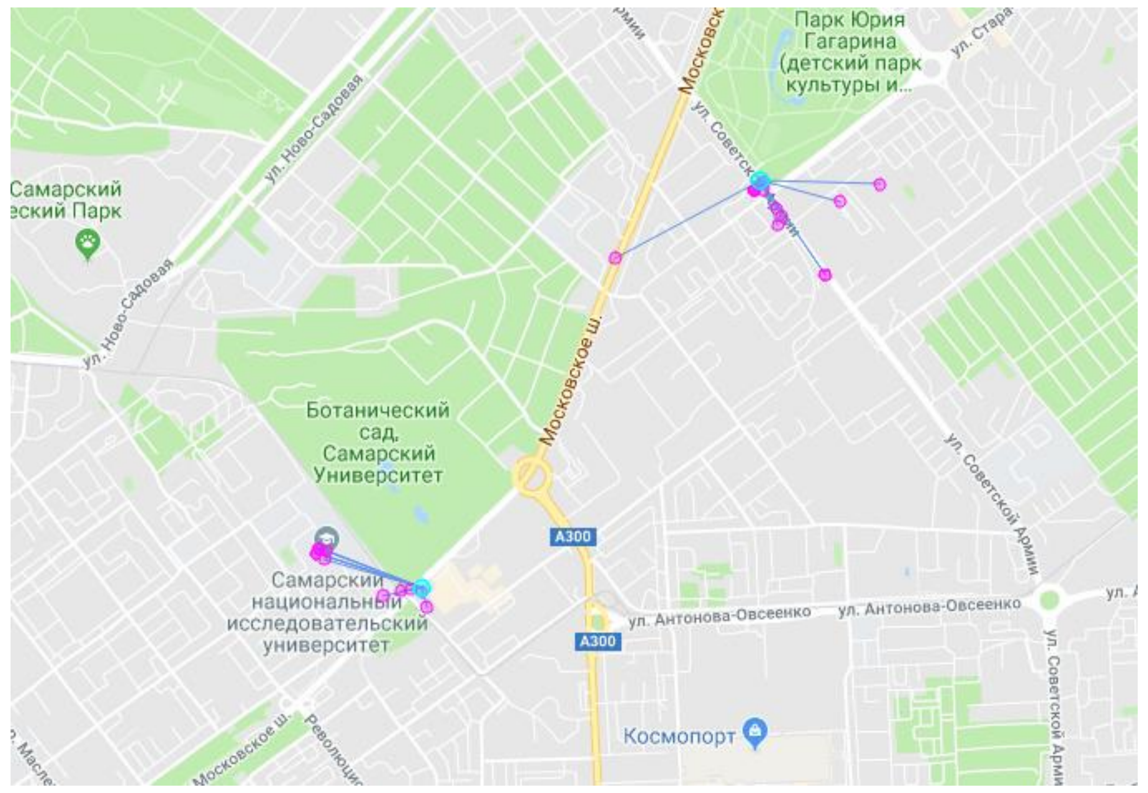

Fig. 2. "Start" and "End" stops of a specific user

User-preferred stops

Step 1. For all stops from the set S calculate the values:

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

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

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

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

Step 2. The values set is ordered in descending order — a permutation is formed

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

User-preferred "transport correspondence"

- $p_u\left(t \middle| s1, s2, m, \mathrm{W}_a\right) \quad \left(m \in \mathrm{M}(s1, s2), \quad a = \overline{0,1}\right)$ correspondence time distribution density *s*1→*s*2 with the route choice *m*
- $p_u\left(t \middle| s1, s2, \mathrm{W}_a\right)$ - correspondence time distribution density *s*1→*s*2
- $P_u\left(s1, s2 \middle| \mathrm{W}_a\right)$ - correspondence probability *s*1→*s*2
- $P_u\left(m \middle| s1, s2, \mathrm{W}_a\right)$ - probability of choosing the route *m* for implementing the correspondence *s*1→*s*2
- $P_u\left(m \middle| \mathrm{W}_a\right)$ - probability of choosing the route *m* for implementing the correspondence
- $P_u^*\left(s \mid \mathrm{W}_a\right)$ - the probability that the stop s is the “end/start”
- $p_u\left(\rho \mid \mathrm{W}_a\right)$ - distances distribution that the user is able to overcome without using route vehicles
- $p_u\left(\tau \middle| \rho, \mathrm{W}_a\right)$ - time distribution that the user spends in overcoming the corresponding distance

Results

The database obtained during the experiments, contains information about requests 57190 users. Each user is represented by a unique identifier *ID*(*u*), which is defined by the device ID hash code and is impersonal. The database contains a total of 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 3).

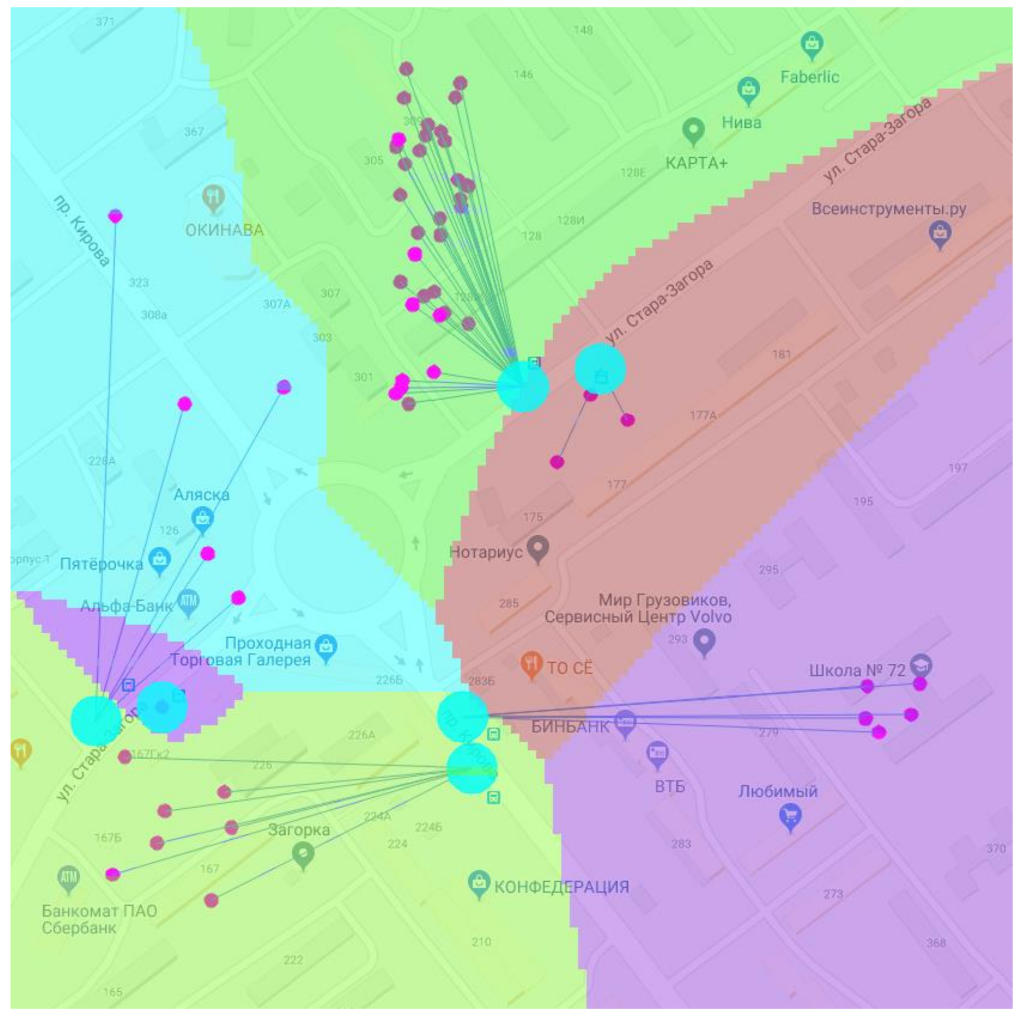

Fig. 3. Preferred stops map depending on user location.

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%, for the nearest stop algorithm - 65%.

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