



А.А. Бородинов

## СПОСОБ ОПРЕДЕЛЕНИЯ ПОЛЬЗОВАТЕЛЬСКИХ ПРЕДПОЧТЕНИЙ В СИСТЕМАХ ПЕРСОНАЛИЗИРОВАННЫХ РЕКОМЕНДАЦИЙ ДЛЯ ПАССАЖИРОВ ОБЩЕСТВЕННОГО ТРАНСПОРТА

(Самарский национальный исследовательский университет имени академика  
С.П. Королева)

Благодаря широкому распространению и активному использованию современных систем электронных коммуникаций, глобальных навигационных систем, активных и пассивных датчиков различного типа и назначения, стало возможным получение большого объема разнородных данных, характеризующих транспортную ситуацию в городе. Подобная информация используется в навигационных или рекомендательных системах (сервисах) достаточно широко [1]. Однако, наряду с развитием сервисов и их популяризацией растет не только количество информации, которую приходится учитывать при планировании перемещений, но также растут ожидания и требования пользователей. Вместо классических задач поиска «кратчайшего пути» [2] или получения «прогноза прибытия на остановку общественного транспорта» [3, 4] пользовательский спрос индивидуализируется, смещая ожидания от сервисов в сторону т.н. персональных ассистентов. Учет всех обозначенных факторов возможен в «самоподстраивающихся» под индивидуальные предпочтения пользователей системах на основе методов машинного обучения [5]. Настоящая работа представляет один из возможных способов постановки и решения задачи определения индивидуальных предпочтений пользователей общественного маршрутного транспорта по данным их взаимодействия с мобильным сервисом в рамках проблемы построения персонализированной рекомендательной системы.

Целесообразными вариантами "пользовательских предпочтений" для проектируемой системы представляются следующие два (считаем пользователя известным):

- предпочитаемые пользователем остановки при его нахождении в определенных пространственно-временных координатах (см. рисунок 1);
- предпочитаемые пользователем "транспортные корреспонденции", рассматриваемые также в срезе пространственно-временного контекста. Под "транспортными корреспонденциями" понимаются фактические перемещения от одной остановки до другой и выбираемый при этом маршрут и/или вид маршрутного ТС. Дополнительной (производной) информацией является информация о "стартовых" и "конечных" остановках конкретного пользователя (см. рисунок 2).

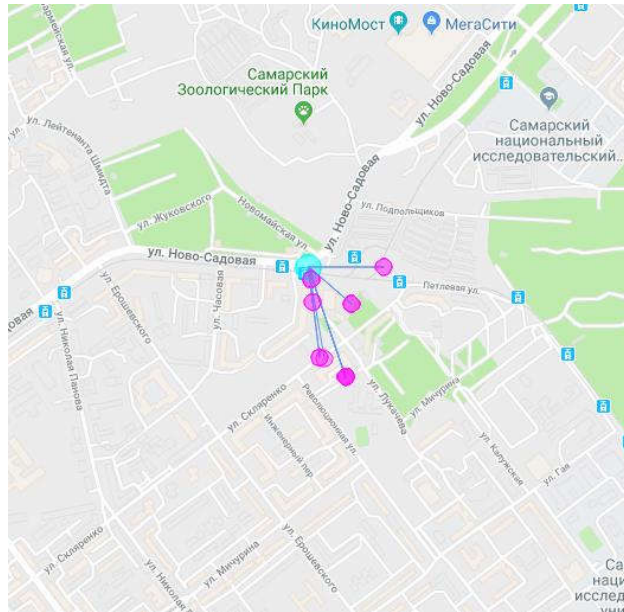


Рисунок 1 - Голубой круг - местоположение остановки, фиолетовые круги - точки положения пользователя при формировании запроса

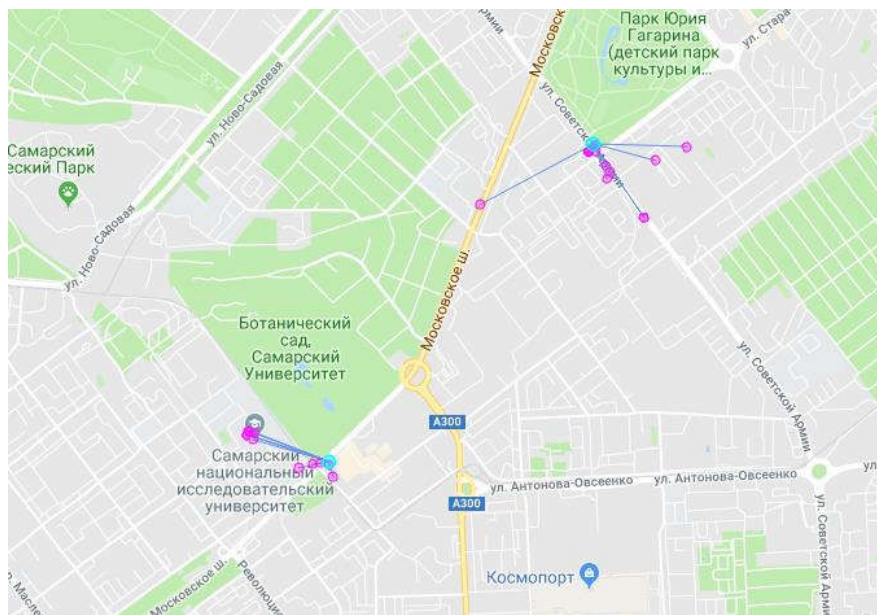


Рисунок 2 - "Стартовые" и "конечные" остановки конкретного пользователя

Пусть  $S$  - множество остановок общественного транспорта. И пусть для каждой остановки  $s \in S$  определены ее пространственные (например, географические) координаты  $\mathbf{x}_s \equiv (x_s, y_s, z_s)$  и некоторый уникальный идентификатор остановки, обозначаемый  $ID(s)$ . Без ограничения общности можем считать множество  $S$  упорядоченным (например, по  $ID(s)$ ):  $S = \{s_1, s_2, \dots, s_{|S|}\}$ .

Пусть величина  $d$  определяет календарную дату, величина  $t$  - время суток, а  $w(d) \in W$  - день недели, принимающий значения из множества:  $W = W_0 \cup W_1$ ,  $W_0 \equiv \{MON, TUE, WEN, THU, FRI\}$ ,  $W_1 \equiv \{SAT, SUN\}$ .



Задача определения "предпочитаемых пользователем остановок" при его нахождении в определенных пространственно-временных координатах может быть формализована следующим образом.

Дано:

а) множество  $\{x_i, d_i, t_i, z_i\}_{i \in S}$  - прецедентов в виде (вектор признаков; ответ)

б) вектор признаков новой ситуации  $x, d, t$ .

Надо: Для указанного вектора  $x, d, t$  определить перестановку  $\sigma : \{1, \dots, |S|\} \rightarrow \{1, \dots, |S|\}$  объектов из упорядоченного множества (остановок). Результатом для пользователя является упорядоченный список остановок:  $s_{\sigma(1)}, s_{\sigma(2)}, \dots, s_{\sigma(|S|)}$ .

Решение: В настоящей работы используется подход, основанный на идеи алгоритма вычисления оценок, предложенного академиком Ю.И. Журавлевым, и метода непараметрического оценивания плотности вероятности Парзена. А именно, зададим величину  $\Gamma(\text{вектор признаков; класс})$ , характеризующую принадлежность вектора признаков к классу, в виде

$$\Gamma(x, d, t; z) = \sum_{i \in S} \mu(x, d, t; x_i, d_i, t_i) I(z_i = z)$$

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

Здесь индикатор события

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

Величины  $\alpha, \beta \in \mathbb{R}_+$  - некоторые коэффициенты,  $|t - t_i|$  - числовая величина (например, число секунд), характеризующая разницу между моментами времени  $t, t_i$ . В результате алгоритм решения задачи определения "предпочитаемых пользователем остановок" будет иметь следующий вид:

Шаг 1. Для всех остановок из множества  $S$  рассчитываются величины:

$$\Gamma(x, d, t; ID(s_i)), \quad i = 1, \dots, |S|$$

Шаг 2. Полученное в (15) множество значений упорядочивается по убыванию, формируется перестановка  $\sigma : \{1, \dots, |S|\} \rightarrow \{1, \dots, |S|\}$ .

Визуализация результатов проводилась на основе сервиса Google Maps. В экспериментальном исследовании работоспособности предложенных методов и алгоритмов использовались данные мобильного приложения «Прибывалка-63» сервиса tosamara.ru.

Для наглядного представления результатов работы предложенного подхода были построены карты с различными параметрами  $\alpha, \beta \in \mathbb{R}_+$  и временем запроса. Цвет области на карте соответствует первой остановке из упорядочен-



ного списка. Пример определения предпочитаемой остановки для исследуемого пользователя представлен на рисунке 3.



Рисунок 3 - Карта предпочитаемых остановок в зависимости от местоположения пользователя во время использования рекомендательной системы

Также для сравнения с предложенным алгоритмом был реализован метод, при котором пользователю предлагалась ближайшая остановка, без учета предыдущих запросов. Точность классификации остановки для предложенного алгоритма составила 93%, для алгоритма ближайшей остановки – 65%, что подтверждает правильность ранжирования остановок представленным методом.

### Литература

1. Chorus, C.G. et al. Use and effects of Advanced Traveller Information Services (ATIS): A review of the literature / C.G. Chorus, E.J.E. Molin, B. Van Wee // *Transport Reviews*. — 2006. — Т. 26(2). — С. 127–149. — DOI: 10.1080/01441640500333677.
2. Agafonov, A.A. et al. Numerical route reservation method in the geoinformatic task of autonomous vehicle routing / A.A. Agafonov, V.V. Myasnikov // *Computer Optics*. — 2018. — Т. 42(5). — С. 912–920. — DOI: 10.18287/2412-6179-2018-42-5-912-920.
3. Agafonov, A.A. et al. Big data analysis in a geoinformatic problem of short-term traffic flow forecasting based on a K nearest neighbors method / A.A. Agafonov, A.S. Yumaganov, V.V. Myasnikov // *Computer Optics*. — 2018. — Т. 42(6). — С. 1101–1111. — DOI: 10.18287/2412-6179-2018-42-6-1101-1111.
4. Agafonov, A. et al. Traffic flow forecasting algorithm based on combination of adaptive elementary predictors / A. Agafonov, V. Myasnikov // *Communications in Computer and Information Science*. — 2015. — Т. 542. — С. 163–174. — DOI: 10.1007/978-3-319-26123-2\_16.
5. Portugal, I. et al. The use of machine learning algorithms in recommender systems: A systematic review / I. Portugal, P. Alencar, D. Cowan // *Expert Systems with Applications*. — 2018. — Т. 97. — С. 205–227. — DOI: 10.1016/j.eswa.2017.12.020.