



одновинтового вертолета с ионно-меточными и аэрометрическими измерительными каналами // Вестник КГТУ им. А.Н. Туполева. 2019. Том 75. №1. С. 82-88.

[4] Кремлевский П.П. *Расходомеры и счетчики количества*. Изд. 3-е перераб. и доп. // Машиностроение, 1975. 776 с.

А.А. Бородинов

МЕТОД ЧИСЛЕННОГО АНАЛИЗА ДАННЫХ ТРАНСПОРТНЫХ КОРРЕСПОНДЕНЦИЙ ПОЛЬЗОВАТЕЛЕЙ ОБЩЕСТВЕННОГО ТРАНСПОРТА

(Самарский университет)

В данной работе рассматривается задача определения множества вероятностных корреспонденций пользователя на общественном маршрутном транспорте или «пешком» и их вероятностные характеристики по данным мобильного приложения «Прибывалка-63» сервиса tosamara.ru, используемого в настоящее время для информирования жителей г. Самара о движении общественного транспорта и прибытия его на остановки.

Получаемое множество вероятностных корреспонденций пользователя в дальнейшем может использоваться в навигационных или рекомендательных системах (сервисах) [1] в качестве предыдущих действий и решений пользователей [2, 3]. Учет всех обозначенных факторов возможен в «самоподстраивающихся» под индивидуальные предпочтения пользователей системах на основе методов машинного обучения [4].

Используемые в работе данные содержат информацию о идентификаторах и координатах остановок, координатах пользователя во время запроса информации о прибытии общественного транспорта к остановке и идентификаторах маршрута и списка идентификаторов остановок на маршруте.

Задача определения осуществляемых пользователем транспортных корреспонденций по косвенным данным может быть определена как задача оценки вероятностных характеристик соответствующих корреспонденций, то есть перемещений от остановки s_1 к остановке s_2 с использованием некоторого маршрута. Множество корреспонденций на маршрутном общественном транспорте представим следующим образом:

$$\left\{ \left(d_i, s_i^{start}, s_i^{end}, m_i^\ell, k_i^\ell \right) \right\}_{i \in I_u; \ell \in \mathcal{S}_i},$$

где I_u – множество вероятных корреспонденций пользователя и от известной (по запросам пользователя) остановки s_i^{start} к неопределенной на момент выбора РТС s_i^{end} для каждого дня d_i ($w(d_i) \in W_a$).



Множество \mathfrak{S}_i определяет подмножество маршрутных ТС $\{m_i^\ell, k_i^\ell\}_{\ell \in \mathfrak{S}_i}$, которые могли реализовать соответствующую корреспонденцию $s_i^{start} \rightarrow s_i^{end}$. Для пары m_i^ℓ, k_i^ℓ время $t(d_i, m_i^\ell, k_i^\ell, s_i^{start})$ определяет время появления этого маршрутного ТС на остановке s_i^{start} . Параметр δ_x определяет допустимую предельную удаленность пользователя от остановки по координатам, а δ_t допустимую максимальную длительность временного интервала без запросов от пользователя.

Определяем дополнительно среднюю интенсивность запросов λ_i пользователя, время их начала t_i^{\min} и прекращения t_i^{\max} :

$$t_i^{\max} = \max_{j \in \mathfrak{S}_i} t_j, \quad t_i^{\min} = \min_{j \in \mathfrak{S}_i} t_j, \quad \lambda_i = \frac{|\mathfrak{S}_i|}{t_i^{\max} - t_i^{\min}}.$$

Заполнение списка потенциально возможных маршрутных поездок производится следующим образом:

Будем относить к вероятностным характеристикам каждой возможной корреспонденции на общественном ТС для поездки с остановки s_i^{start} набор вероятностных характеристик каждой такой корреспонденции в виде вероятностей $P_{i\ell}$, определяемых следующим образом:

$$p_i^+(t) = \begin{cases} \lambda_i \left(t - \left(t_i^{\max} - \frac{1}{3\lambda_i} \right) \right) \exp \left(- \left(t - \left(t_i^{\max} - \frac{1}{3\lambda_i} \right) \right) \right), & t - \left(t_i^{\max} - \frac{1}{3\lambda_i} \right) \geq 0, \\ 0, & \text{иначе.} \end{cases}$$

$$P_{i\ell}^+ = \frac{p_i^+(t(d, s_i^{start}, m_i^\ell, k_i^\ell))}{\sum_{j \in \mathfrak{S}_i} p_i^+(t(d, s_i^{start}, m_i^j, k_i^j))}.$$

Целью экспериментальных исследований является подтверждение работоспособности предложенного метода численного анализа данных о транспортных корреспонденциях пользователей общественного транспорта - то есть способности корректно реконструировать истинные транспортные корреспонденции соответствующего участника движения. В работе проводился анализ реальных данных с сервиса "Прибывалка-63".

Хотя сервис предоставляет огромное количество информации (общее число пользователей - более 200 тыс. человек, число запросов в день - около 1 млн.), основной проблемой оказался поиск заведомо известных транзакций, которые могли бы подтвердить корректность работы предложенного метода численного анализа. Для качественной верификации метода мы использовали только данные от пользователей сервиса, которые предоставили информацию о своих поездках.

Эксперимент проводился следующим образом: были отображены запросы пользователей, постоянно пользующихся predetermined маршрутами



общественного транспорта при поездках на работу и с работы. Запросы представлены на рисунке 1.

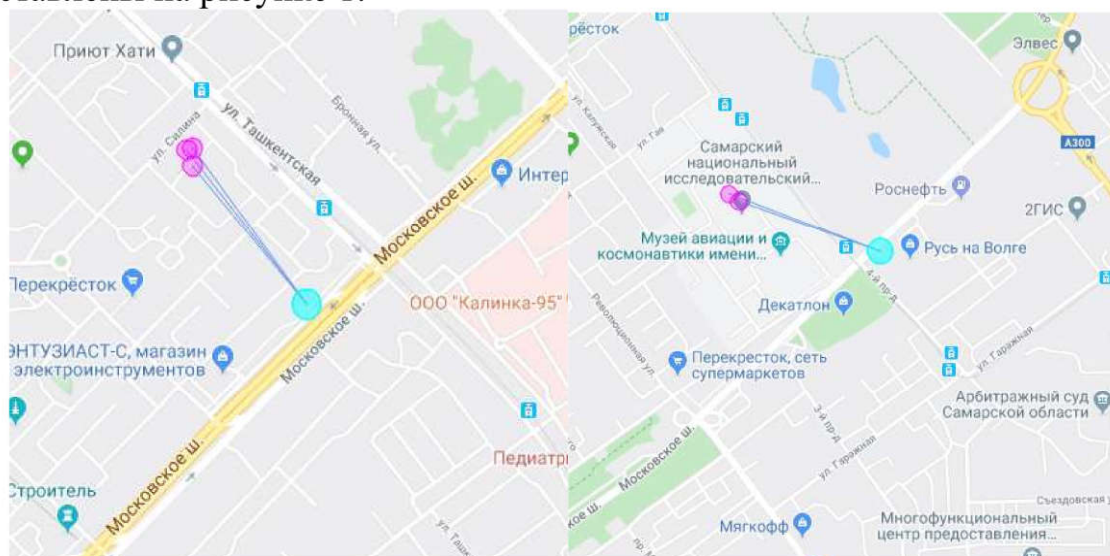


Рис. 1. Запросы участника проекта при поездке на общественном транспорте без пересадки: слева стартовая остановка отправления на работу, справа стартовая остановка отправления с работы домой

Запросы к мобильному приложению для данного пользователя, хранящиеся в базе данных, представлен на рисунке 1.4.

TimeStamp	DeviceID	Latitude	Longitude	Method	KS_ID	Stop_Latitude	Stop_Longitude
15.04.2019 8:13	f9344f04c90ff1ce9c	53.2629875	50.2330563	getFirstArrivalToStop	314	53.2601801307	50.2367014585211
15.04.2019 8:23	f9344f04c90ff1ce9c	53.2626987	50.2332317	getFirstArrivalToStop	314	53.2601801307	50.2367014585211
15.04.2019 8:28	f9344f04c90ff1ce9c	53.2630261	50.2332334	getFirstArrivalToStop	314	53.2601801307	50.2367014585211
15.04.2019 18:39	f9344f04c90ff1ce9c	53.2138125	50.1756480	getFirstArrivalToStop	872	53.2127417866	50.1806442642877
15.04.2019 18:40	f9344f04c90ff1ce9c	53.2139708	50.1752843	getFirstArrivalToStop	872	53.2127417866	50.1806442642877

Рис. 2. Запросы участника проекта при поездке на общественном транспорте без пересадки

Для поездки из дома до работы список вероятных корреспонденций пользователя представлен на рисунке 3.

KR_ID	ID_TransNavi	dateTime	P
625	403	2019-04-15T08:30:20	0.031966900365400235
1	389	2019-04-15T08:30:00	0.02046321597780766
53	109043	2019-04-15T08:32:01	0.018210441491160866
234	3024078	2019-04-15T08:34:53	0.009525851763204826
160	3024114	2019-04-15T08:34:31	0.00902011963856935
160	3024109	2019-04-15T08:41:16	0.00555184203669124
1	68	2019-04-15T08:42:57	0.0027284202526072143

Рис. 3. Список наиболее вероятных корреспонденций пользователя из дома до работы

Пользователь использует 625 маршрут для транспортной корреспонденции от дома до работы, что соответствует полученному списку



вероятных корреспонденций. Корреспонденция пользователя на 625 маршруте представлена на рисунке 4.

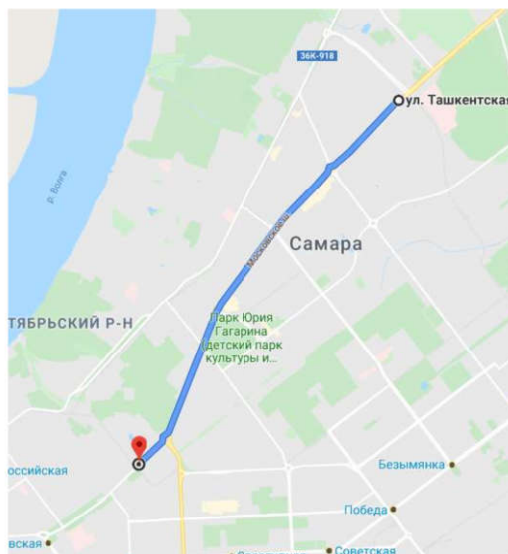


Рис. 4. Маршрут пользователя из дома до работы без пересадки

Благодарности

Работа выполнена при финансовой поддержке Министерства науки и высшего образования РФ (уникальный идентификатор проекта RFMEFI57518X0177).

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. Arentze, T.A. Adaptive personalized travel information systems: A bayesian method to learn users' personal preferences in multimodal transport networks [Text] / T.A. Arentze // *IEEE Transactions on Intelligent Transportation Systems*. – 2013. – Vol. 14, № 4. – P. 1957-1966.
3. Nuzzolo, A. Individual behavioural models for personal transit pre-trip planners [Text] / A. Nuzzolo [et al.] // *Transportation Research Procedia*. – 2015. – Vol. 5. – P. 30-43.
4. Portugal, I. The use of machine learning algorithms in recommender systems: A systematic review [Text] / I. Portugal, P. Alencar, D. Cowan // *Expert Systems with Applications*. – 2018. – P. 205-227.