

# Анализ предпочтений участников движения на маршрутном общественном транспорте в задаче построения персонализированной рекомендательной системы

А.А. Бородинов<sup>1</sup>, В.В. Мясников<sup>1,2</sup>

<sup>1</sup>Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

<sup>2</sup>Институт систем обработки изображений РАН - филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001

**Аннотация.** В работе рассматриваются теоретические и алгоритмические аспекты построения персонализированной рекомендательной системы (мобильного сервиса), предназначенной для пользователей общественного маршрутного транспорта. Основной упор сделан на выявлении и формализации понятия "пользовательские предпочтения", лежащего в основе современных персонализированных рекомендательных систем. Представлены неформальные (вербальные) и формальные (математические) постановки соответствующих задач определения "пользовательских предпочтений" в определенном пространственно-временном контексте: определение предпочитаемых остановок и определение предпочитаемых "транспортных корреспонденций". Показано, что первая из задач может быть представлена как известная задача классификации, то есть может быть сформулирована и решена с использованием известных методов распознавания образов и машинного обучения. Вторая же сводится к нахождению оценок серии условных распределений. Представлены результаты экспериментального исследования работоспособности предложенных подходов, методов и алгоритмов на примере данных мобильного приложения "Прибывалка-63" сервиса [tosamara.ru](http://tosamara.ru), используемого в настоящее время для информирования жителей г. Самара о движении общественного транспорта.

## 1. Введение

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 может быть использована и 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 (англ. Intelligent personal assistant). Несмотря на то, что окончательное решение или выбор в подобных системах остается за человеком, варианты предлагаемых ими решений существенным образом оказываются зависимы не только от сценарных условий запроса, но и от предыдущих действий и решений пользователя [5, 6]. Учет всех обозначенных факторов возможен в «самоподстраивающихся» под индивидуальные предпочтения пользователей системах на основе методов машинного обучения [7]. Но быстрому возникновению подобных сервисов препятствует несовершенство существующих алгоритмов и отсутствие значительного опыта в использовании методов машинного обучения в системах подобного рода.

Анализ современной литературы, посвященной вопросам построения рекомендательных систем мультимодальной маршрутизации [5, 8, 9] (то есть планирования поездок, в каждой из которых могут использоваться сразу несколько видов транспорта) позволяет выделить ряд крупных проблем, стоящих перед разработчиками:

- существует проблема «холодного старта» при построении профиля персональных предпочтений [8, 10]: крайне важно добиться баланса между точностью рекомендованных маршрутов с самого начала и инициализацией системы, - то есть допустимое время настройки профиля персональных предпочтений и получения релевантных рекомендаций должно быть малым;

- способ получения информации от пользователя не является формализованным [11, 12];

- индивидуальные характеристики, такие как личный доход, возраст, пол, размер семьи, доступ к общественному транспорту влияют на решения о выборе маршрута даже для одной и той же цели поездки [13];

- существует проблема изменения предпочтений пользователя во времени, а также влияние контекста ситуации (например, цель поездки) на выбор пользователя [14, 15];

- типовые существующие решения используют в основном байесовский подход с последовательной схемой пересчета параметров [5, 16];

- возможно использование методов передачи знаний и дообучения для улучшения рекомендаций [17];

- существует проблема определения плотности транспортного потока на пути следования транспортного средства [18].

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. Второй раздел работы направлен на формализацию основных понятий и введение основных обозначений для всех объектов взаимодействия, возникающих в рассматриваемой задаче. В третьем разделе представлен конкретный перечень информации, возникающей в процессе взаимодействия пользователей общественного транспорта с мобильным приложением "Прибывалка-63" сервиса tosamara.ru, используемого в настоящее время для информирования жителей г. Самара о движении общественного транспорта и прибытия его на остановки. Здесь же представлены варианты неформализованных (вербально описанных) определений "пользовательских предпочтений", целесообразных для дальнейшего рассмотрения. В четвертом разделе детально рассмотрены математические формулировки задач, соответствующих указанным вариантам определения "пользовательских предпочтений", а также представлены методы/алгоритмы их решения. Наконец, в пятом разделе представлены результаты экспериментальных исследований на реальных данных, полученных с использованием мобильного приложения "Прибывалка-63".

## 2. Основные понятия и обозначения

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:

$$W = W_0 \cup W_1, \quad (1)$$

$$W_0 \equiv \{MON, TUE, WEN, THU, FRI\}, \quad W_1 \equiv \{SAT, SUN\}.$$

Пусть, далее,  $V$  определяет множество объектов общественного транспорта (далее ТС - транспортных средств), каждый из которых  $v \in V$  характеризуется типом

$$type(v) \in \{BUS, TRAM, TROL, MARS\}, \quad (2)$$

а также имеет уникальный идентификатор  $ID(v)$  (на практике уникальный идентификатор может совпадать с государственным регистрационным номером ТС). Для каждого ТС в любой момент времени считаем определенными его пространственные координаты:

$$\mathbf{x}(v, d, t) \equiv (x(v, d, t), y(v, d, t), z(v, d, t)). \quad (3)$$

Множество маршрутов объектов общественного транспорта обозначим как  $M$ . При этом каждый маршрут  $m \in M$  охарактеризуем пятеркой объектов:

$$m \equiv (ID(m), N(m), \mathbf{s}(m), N^*(m), \mathbf{x}(m)), \quad (4)$$

где  $ID(m)$  - идентификатор маршрута (на практике - номер маршрута),  $N(m)$  - количество остановок в маршруте, а  $\mathbf{s}(m)$  - последовательность остановок в количестве  $N(m)$  вида:

$$\mathbf{s}(m) = (s_1^m, s_2^m, \dots, s_{N(m)}^m), \quad (5)$$

здесь  $s_n^m \in S$  ( $m \in M, n \in \overline{1, N(m)}$ ). Пусть также  $S(m) \equiv \{s_i^m\}_{i=1, N(m)} \subseteq S$  - множество остановок соответствующего маршрута,  $ind(s, m)$  - индекс остановки  $s$  маршрута  $m$ , то есть  $ind(s_n^m, m) = n$ . В случае, если  $s \notin S(m)$  соответствующий индекс считаем принимающим так называемое "неопределенное значение", обозначаемое далее  $\Delta$ :  $ind(s, m) = \Delta$ . Также для удобства введем обозначение для множества маршрутов, проходящих через одну или пару остановок следующим образом:

$$M(s) \equiv \{m \in M: s \in S(m)\}, \quad M(s_1, s_2) \equiv \{m \in M: s_1 \in S(m) \wedge s_2 \in S(m)\}. \quad (6)$$

Более детальная информация о геометрии маршрута представлена парой  $N^*(m), \mathbf{x}(m)$ , где первая величина определяет число узлов полилинии, описывающей маршрут, а вторая - есть вектор, определяющий координаты этих узлов:

$$\mathbf{x}(m) \equiv (\mathbf{x}_1^m, \mathbf{x}_2^m, \dots, \mathbf{x}_{N^*(m)}^m). \quad (7)$$

Для удобства далее пару  $(m, k)$ , ( $k = \overline{1, K(d, m)}$ ) будем называть реализацией маршрута (PM) в соответствующий день  $d$ .

Дополнительно обозначим  $t(d, m, k, s)$  - время прибытия ТС, назначенного на PM  $(m, k)$  в день  $d$ , на остановку  $s \in S(m)$  (в случае  $s \notin S(m)$  считаем значение времени неопределенным).

Обозначим далее ТС, назначенное на PM  $(m, k)$  в день  $d$ , как  $v(d, m, k) \in V$  ( $k = \overline{1, K(d, m)}$ ).

Дополнительно к ТС-ам, участниками уличного движения являются пешеходы/пассажиры, рассматриваемые в рамках настоящей работы как пользователи (некоторой мобильной услугой, связанной с транспортными сервисами). Обозначим величиной  $U$  их множество, и каждого конкретного пользователя  $u \in U$  охарактеризуем его уникальным идентификатором  $ID(u)$  (на

практике - идентификатор используемого мобильного устройства или хэш-код этого идентификатора) и пространственными координатами в конкретный момент времени  $d, t$ :

$$\mathbf{x}(u, d, t) \equiv (x(u, d, t), y(u, d, t), z(u, d, t)). \quad (8)$$

Если информация о координатах пользователя отсутствует, считаем, что они имеют "неопределенное значение"  $\Delta$  (одновременно все три).

### 3. Данные мобильного приложения и вспомогательного сервиса, неформальные варианты "пользовательских предпочтений"

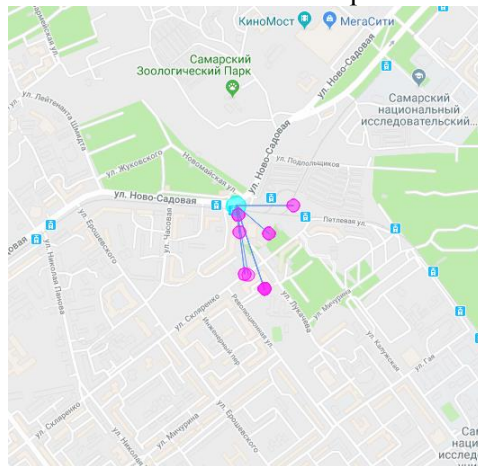
Для мобильного сервиса "Прибывалка-63" данные для анализа представлены следующим образом (из базы данных сервиса или логов функционирования):

- данные об остановках (идентификаторы и координаты);
- данные о маршрутах (идентификаторы и список идентификаторов остановок);
- данные о ТС (идентификаторы), координатах их местоположения (с частотой 2 раза в минуту) при нахождении на маршруте, назначения на маршруты;
- координаты пользователей фиксируются в момент формирования запросов, также как и параметры запроса (собственно результаты запроса не сохраняются, поскольку могут быть восстановлены по данным о движении ТС) в виде:  $ID(s), d, t, ID(u), \mathbf{x}(u, d, t)$ ;

- реакция пользователей на запрос не фиксируется.

Исходя из представленных данных, целесообразными вариантами "пользовательских предпочтений" 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.** К определению "индивидуальных предпочтений": голубой круг - местоположение остановки, фиолетовые круги - точки положения пользователя при формировании запроса.

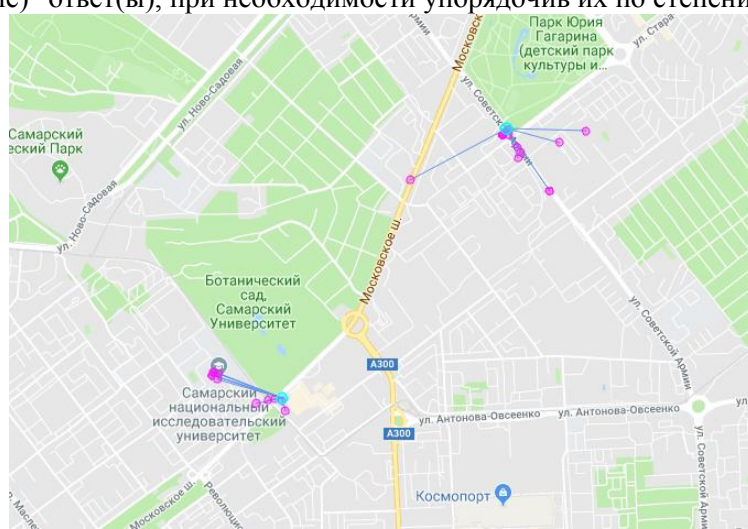
## 4. Задачи определения "пользовательских предпочтений" и методы их решения

### 4.1. Предпочитаемые пользователем остановки

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

Пусть для интересующего нас пользователя дано множество прецедентов (обучающие примеры), каждый из которых представляет собой совокупность вектора-описания

пространственно-временного контекста ситуации и соответствующего этой ситуации "ответа" (в нашем случае - идентификатор остановки). Необходимо по вновь поступившему вектору-описанию пространственно-временного контекста новой ситуации (который может отсутствовать в списке прецедентов-примеров), указать наиболее "подходящий(-ие)/релевантный(-ые)" ответ(ы), при необходимости упорядочив их по степени релевантности.



**Рисунок 2.** К определению "индивидуальных предпочтений": "стартовые" и "конечные" остановки конкретного пользователя.

Для случая одного выдаваемого ответа описанная задача является хорошо известной в теории распознавания [19, 20] или машинном обучении [7, 21] задачей классификации, где по описанию объекта/ситуации в виде вектора признаков необходимо указать класс объекта/ситуации. Тот же аппарат может быть применен и для случая получения упорядоченного множества ответов, поскольку любой алгоритм распознавания может быть представлен в виде двух последовательных операторов, первый из которых переводит описание объекта в некоторую числовую величину, характеризующую "степень принадлежности" к классу, а второй - по указанной величине "выдает" факт отнесения конкретному классу [22]. В статистических методах распознавания в качестве числовой величины выступает апостериорная вероятность [19], в алгебраических [23] - оценки, в нейросетевых - выход последнего слоя нейронов и т.д. Указанную числовую величину обозначим  $\Gamma(\text{вектор признаков}; \text{класс})$ .

Таким образом, формальная постановка задачи для конкретного пользователя  $u \in U$  (его обозначение далее опускаем) может быть представлена следующим образом (используемая ниже величина  $z \equiv ID(s)$ ).

Дано:

- а) множество  $\{\mathbf{x}_i, d_i, t_i; z_i\}_{i \in \mathcal{Z}}$  - прецедентов в виде (вектор признаков; ответ)
- б) вектор признаков новой ситуации  $\mathbf{x}, d, t$ .

Надо: для указанного вектора  $\mathbf{x}, d, t$  определить перестановку  $\sigma: N_{|S|} \rightarrow N_{|S|}$  объектов из упорядоченного множества (остановок)  $S = \{s_1, s_2, \dots, s_{|S|}\}$ , такую что

$$\Gamma(\mathbf{x}, d, t; ID(s_{\sigma(1)})) \geq \Gamma(\mathbf{x}, d, t; ID(s_{\sigma(2)})) \geq \dots \geq \Gamma(\mathbf{x}, d, t; ID(s_{\sigma(|S|)})). \quad (9)$$

Результатом для пользователя является упорядоченный список остановок:

$$s_{\sigma(1)}, s_{\sigma(2)}, \dots, s_{\sigma(|S|)}. \quad (10)$$

Формальным показателем качества итогового решения может выступать, например, следующий:

$$\Gamma_{\Sigma} = \sum_{i=1}^{|\mathcal{S}|} \frac{1}{i} \Gamma(\mathbf{x}, d, t; ID(s_{\sigma(i)})). \quad (11)$$

Решение:

Теория распознавания образов и машинное обучение предлагает разнообразный аппарат для решения указанной задачи. В настоящей работе мы используем подход, основанный на идеи алгоритма вычисления оценок, предложенного академиком Ю.И.Журавлевым [23], и метода непараметрического оценивания плотности вероятности Парзена [19]. А именно, зададим величину  $\Gamma(\text{вектор признаков; класс})$ , характеризующую принадлежность вектора признаков к классу, в виде

$$\Gamma(\mathbf{x}, d, t; z) = \sum_{i \in \mathcal{S}} \mu(\mathbf{x}, d, t; \mathbf{x}_i, d_i, t_i) I(z_i = z), \quad (12)$$

где

$$\mu(\mathbf{x}, d, t; \mathbf{x}_i, d_i, t_i) = I \left( \begin{array}{l} (w(d) \in W_0 \wedge w(d_i) \in W_0) \vee \\ (w(d) \in W_1 \wedge w(d_i) \in W_1) \end{array} \right) \cdot \exp(-\alpha |t - t_i|) \cdot \exp(-\beta \|\mathbf{x} - \mathbf{x}_i\|). \quad (13)$$

Здесь индикатор события

$$I(a) = \begin{cases} 1, & a = \text{true}; \\ 0, & a = \text{false}. \end{cases} \quad (14)$$

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  $\mathcal{S}$  рассчитываются величины (12):

$$\Gamma(\mathbf{x}, d, t; ID(s_i)), \quad i = \overline{1, |\mathcal{S}|}. \quad (15)$$

Шаг 2. Полученное в (15) множество значений упорядочивается по убыванию - формируется перестановка  $\sigma: \mathbb{N}_{|\mathcal{S}|} \rightarrow \mathbb{N}_{|\mathcal{S}|}$  (9).

Полученная перестановка и является решением искомой задачи. Пользователю предоставляется упорядоченный список остановок (10).

Решение проблемы "холодного старта":

Для решения проблемы "холодного старта" системы первоначально пустой список прецедентов дополняется множеством прецедентов следующего вида:

$$\{(\mathbf{x}_i, d0, t0; ID(s_i))\} \cup \{(\mathbf{x}_i, d1, t0; ID(s_i))\}, \quad i = \overline{1, |\mathcal{S}|}, \quad (16)$$

где  $t0 = "0\text{час}00\text{мин}"$ , в качестве  $d0$  и  $d1$  выступают даты, соответственно, выходного и рабочего дней, предшествующих дате запуска "системы", а  $\mathbf{x}_i (i = \overline{1, |\mathcal{S}|})$  - координаты остановок  $s_i (i = \overline{1, |\mathcal{S}|})$ . Анализ выражений (12)-(13) показывает, что при таких "стартовых" данных вклад временной составляющей  $\exp(-\alpha |t - t_i|)$  в выражение (12) будет одинаков, и различия в значениях  $\Gamma(\dots)$  будут полностью определяться различиями в евклидовых расстояниях от точки  $\mathbf{x}$  до координат расположения остановок  $\mathbf{x}_i (i = \overline{1, |\mathcal{S}|})$ . То есть чем ближе остановка  $s$  к точке  $\mathbf{x}$ , тем большее значение будет иметь величина  $\Gamma(\mathbf{x}, \dots; ID(s))$ .

#### 4.2 Предпочитаемые пользователем "транспортные корреспонденции"

Задача определения предпочитаемых пользователем "транспортных корреспонденций" может быть представлена как задача оценки вероятностных характеристик (относительной частоты) соответствующих корреспонденций, то есть перемещений от остановки  $s_1$  к остановке  $s_2$ , в пространственно-временном контексте. В частности, наибольший интерес представляют следующие величины (все характеристики связаны с поведением конкретного пользователя и рекомендательной системы):

- $p_u(t|s_1, s_2, m, W_a)$  ( $m \in M(s_1, s_2)$ ,  $a = \overline{0,1}$ ) - плотность распределения времени корреспонденции  $s_1 \rightarrow s_2$  с выбором маршрута  $m$  в день недели  $W_a$ ; функция плотности соответствует "времени посадки" на соответствующее маршрутное ТС, то есть указывается для остановки  $s_1$ ;
- $p_u(t|s_1, s_2, W_a)$  - плотность распределения времени корреспонденции  $s_1 \rightarrow s_2$  в день недели  $W_a$ ;
- $P_u(s_1, s_2|W_a)$  - вероятность корреспонденции  $s_1 \rightarrow s_2$  в день недели  $W_a$ ;
- $P_u(m|s_1, s_2, W_a)$  - вероятность выбора маршрута  $m$  для осуществления корреспонденции  $s_1 \rightarrow s_2$  в день недели  $W_a$ ;
- $P_u(m|W_a)$  - вероятность выбора маршрута  $m$  для осуществления корреспонденций в день недели  $W_a$ ;
- $P^*(s)$  - вероятность того, что остановка  $s$  является "конечной/начальной".

Дополнительная информация о поведении пользователей может быть получена из дополнительных данных:

- $p_u(\rho), p_u(\tau|\rho)$  - вероятностного распределения дистанций, которые пользователь способен преодолевать без использования маршрутных ТС, и
- $p_u(\tau|\rho)$  - распределения времени, которые показывает конкретный пользователь при преодолении соответствующей дистанции.

Все указанные величины могут быть рассчитаны на основании данных о потенциальных корреспонденциях пользователя, собираемых рекомендательной системой в виде:

$$\left\{ s_i^{start}, s_i^{end}, m_i^j, k_i^j, t(d, m_i^j, k_i^j, s_i^{start}), \tau_i^*, \sigma_i^* \right\}_{i \in I_d} \quad (17)$$

для каждого дня  $d$  и пользователя. Здесь  $s_i^{start}, s_i^{end}$  - данные о корреспонденции, данные о выбранном  $m_i^j, k_i^j$ ,  $t(d, m_i^j, k_i^j, s_i^{start})$  - время нахождения ТС с назначенной РМ на остановке  $s_i^{start}$ ,  $\tau_i^*, \sigma_i^*$  - среднее и среднеквадратическое отклонение потенциальной посадки на ТС.

### 5. Результаты экспериментальных исследований

Программная реализация представленного метода была выполнена на языке программирования Python. Визуализация результатов проводилась на основе сервиса Google Maps. В экспериментальном исследовании работоспособности предложенных методов и алгоритмов использовались данные мобильного приложения «Прибывалка-63» сервиса tosamaga.ru.

База данных, полученная из логов функционирования сервиса, содержит информацию о запросах 57190 пользователей. Каждый пользователь представлен уникальным идентификатором  $ID(u)$ , который определен хэш-кодом идентификационного номера его мобильного устройства и является обезличенным. База данных содержит в целом 4103161 запрос пользователей о предоставлении прогноза прибытия на остановку общественного транспорта. Из 1478 остановок сервиса tosamaga.ru пользователи делали запросы к 1417 остановкам.

Для проведения экспериментов были отобраны запросы типовых пользователей, которые представляют среднестатистического пользователя сервиса. 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 упорядоченного списка (10). 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.

Для получения численных показателей, характеризующих качество предложенного алгоритма прогнозирования предпочитаемых остановок, был применен контроль по отдельным запросам (Leave-One-Out CV). Количество разбиений  $C_{|\mathcal{S}|}^1$  множества всех запросов пользователя в таком случае равно  $|\mathcal{S}|$ , а точность классификации вычисляется следующим образом:

$$Accuracy = \frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} I(z_i \equiv ID(s_{\sigma_i(1)})) \cdot 100\% \quad (18)$$

где индикатор  $I(a)$  соответствует (14).

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.

## 6. Заключение

В работе представлены неформальные и математические постановки задач определения пользовательских предпочтений участников движения на маршрутном общественном транспорте в задаче построения персонализированной рекомендательной системы. Представлены результаты экспериментального исследования работоспособности метода определения предпочитаемых пользователем остановок, разработанного с использованием методов распознавания образов и машинного обучения. Формализована задача определения предпочитаемых пользователем транспортных корреспонденций, указан подход к ее решению.

## 7. Литература

- [1] Chorus, C.G. Use and effects of advanced traveller information services (ATIS): A review of the literature / C.G. Chorus, E.J.E. Molin, B. Van Wee // Transp. Rev. – 2006. – Vol. 26(2). – P. 127-149.



- [2] Агафонов, А.А. Исследование численного метода резервирования маршрутов в геоинформационной задаче маршрутизации автономных транспортных средств / А.А. Агафонов, В.В. Мясников // Компьютерная оптика. – 2018. – Т. 42, № 5. – С. 912-920. DOI: 10.18287/2412-6179-2018-42-5-912-920.
- [3] Агафонов, А.А. Анализ больших данных в геоинформационной задаче краткосрочного прогнозирования параметров транспортного потока на базе метода к ближайшим соседям / А.А. Агафонов, А.С. Юмаганов, В.В. Мясников // Компьютерная оптика. – 2018. – Т. 42, № 6. – С. 1101-1111. DOI: 10.18287/2412-6179-2018-42-6-1101-1111.
- [4] Agafonov, A. Traffic flow forecasting algorithm based on combination of adaptive elementary predictors / A. Agafonov, V. Myasnikov // Communications in Computer and Information Science. – 2015. – Vol. 542. – P. 163-174. DOI: 10.1007/978-3-319-26123-2\_16
- [5] Arentze, T.A. Adaptive personalized travel information systems: A Bayesian method to learn users' personal preferences in multimodal transport networks / T.A. Arentze // IEEE Transactions on Intelligent Transportation Systems. – 2013. – Vol. 14(4). – P. 1957-1966.
- [6] Nuzzolo, A. Individual behavioural models for personal transit pre-trip planners / A. Nuzzolo, U. Crisalli, A. Comi, L. Rosati // Transportation Research Procedia. – 2015. – Vol. 5. – P. 30-43.
- [7] Portugal, I. The use of machine learning algorithms in recommender systems: A systematic review / I. Portugal, P. Alencar, D. Cowan // Expert Systems with Applications. – 2018. – P. 205-227.
- [8] Campigotto, P. Personalized and situation-aware multimodal route recommendations: The FAVOUR algorithm / P. Campigotto, C. Rudloff, M. Leodolter, D. Bauer // IEEE Transactions on Intelligent Transportation Systems. – 2017. – Vol. 18(1). – P. 92-102.
- [9] Eiter, T. Semantically enriched multi-modal routing / T. Eiter, T. Krennwallner, M. Prandstetter, C. Rudloff, P. Schneider, M. Straub // International Journal of Intelligent Transportation Systems Research. – 2014. – Vol. 14(1). – P. 1-16.
- [10] Fonte, F.A.M. A tagging recommender service for mobile terminals / F.A.M. Fonte, M.R. Lopez, J.C. Burguillo, A. Peleteiro, A.B. Martinez // Information and communication, Technologies in Tourism. – 2013. – P. 424-435.
- [11] March, J.G. Bounded rationality, ambiguity, and the engineering of choice / J.G. March // The Bell Journal of Economics. – 1978. – Vol. 9(2). – P. 587-608.
- [12] Campigotto, P. Adapting to a realistic decision maker: Experiments towards a reactive multi-objective optimizer / P. Campigotto, A. Passerini // Learning and Intelligent Optimization. – 2010. – Vol. 6073. – P. 338-341.
- [13] Zhang, J. Design and implementation of a daily activity scheduler in the context of a personal travel information system / J. Zhang, T. Arentze // Progress in Location-Based Services. Lecture Notes in Geoinformation and Cartography. – Berlin, Heidelberg: Springer, 2013. – P. 407-433.
- [14] Braunhofer, M. Selective contextual information acquisition in travel recommender systems / M. Braunhofer, F. Ricci // Information Technology & Tourism. – 2017. – Vol. 17(1). – P. 5-29.
- [15] Braunhofer, M. User personality and the new user problem in a context-aware point of interest recommender system / M. Braunhofer, M. Elahi, F. Ricci // Information and Communication Technologies in Tourism: Proceedings of the International Conference in Lugano. – 2015. – P. 537-549.
- [16] Guo, S. Real-time multiattribute Bayesian preference elicitation with pairwise comparison queries / S. Guo, S. Sanner // Journal of Machine Learning Research. – 2010. – P. 289-296.
- [17] Pan, S.J. A survey on transfer learning / S.J. Pan, Q. Yang // IEEE Transactions on Knowledge and Data Engineering. – 2010. – Vol. 22(10). – P. 1345-1359.
- [18] Pan, S.J. A survey on transfer learning / S.J. Pan, Q. Yang // IEEE Transactions on Knowledge and Data Engineering. – 2010. – Vol. 22(10). – P. 1345-1359.
- [19] Myasnikov, V.V. Method for detection of vehicles in digital aerial and space remote sensed images. – Computer Optics. – 2012. – Vol. 36(3). – P. 429-438.
- [20] Fukunaga, K. Introduction to statistical pattern recognition. – San Diego: Academic Press, 1990.

- [21] Воронцов, К.В. Машинное обучение: курс лекций [Электронный ресурс]. Режим доступа: <http://www.machinelearning.ru/wiki/> (дата обращения 10.11.18).
- [22] Bishop, С.М. Pattern Recognition and Machine Learning. – Springer, 2006.
- [23] Журавлев, Ю.И. Алгоритмы распознавания, основанные на вычислении оценок / Ю.И. Журавлев, В.В. Никифоров // Кибернетика. – 1971. – Т. 3.
- [24] Журавлев, Ю.И. Распознавание образов и распознавание изображений / Ю.И. Журавлев, И.Б. Гуревич // Распознавание, классификация, прогноз. Математические методы и их применение: Ежегодник. – М.: Наука, 1989. - Вып. 2.

### **Благодарности**

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

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

**A.A. Borodinov<sup>1</sup>, V.V. Myasnikov<sup>1,2</sup>**

<sup>1</sup>Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

<sup>2</sup>Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

**Abstract.** The paper presents the theoretical and algorithmic aspects for making a personalized recommender system (mobile service) designed for public route transport users. The main focus is on identifying and formalizing the concept of "user preferences", which is the basis of modern personalized recommender systems. Informal (verbal) and formal (mathematical) formulations of the corresponding problems of determining "user preferences" in a specific spatial-temporal context are presented: the preferred stops definition and the preferred "transport correspondence" definition. The first task can be represented as a well-known classification problem. Thus, it can be formulated and solved using well-known pattern recognition and machine learning methods. The second is reduced to the construction of dynamic graphs series. The experiments were conducted on data from the mobile application "Pribyvalka-63". The application is the tosamara.ru service part, currently used to inform Samara residents about the public transport movement.