

АЛГОРИТМ ОЦЕНКИ ВРЕМЕНИ ПРИБЫТИЯ ОБЩЕСТВЕННОГО ТРАНСПОРТА С ИСПОЛЬЗОВАНИЕМ АДАПТИВНОЙ КОМПОЗИЦИИ ЭЛЕМЕНТАРНЫХ ПРОГНОЗОВ

Агафонов А.А.¹, Мясников В.В.^{1,2}

¹ Институт систем обработки изображений РАН,

² Самарский государственный аэрокосмический университет имени академика С.П. Королёва
(национальный исследовательский университет)

Аннотация

Работа посвящена решению задачи построения прогноза времени прибытия общественных транспортных средств на остановки общественного транспорта. Предложен оригинальный алгоритм прогнозирования, основанный на модели адаптивной композиции элементарных алгоритмов прогнозирования, каждый из которых характеризуется малым числом настраиваемых параметров. Адаптивность подразумевает зависимость параметров конструируемой композиции от ряда управляющих параметров модели, к которым относятся следующие актуальные (определённые на текущий момент) факторы: погодные условия, плотность транспортного потока, динамика движения, горизонт прогноза и др. Адаптивность достигается введением иерархического разбиения области значений управляющих параметров, применяемого в дереве регрессии. Проведено исследование предложенного алгоритма на данных движения городского пассажирского транспорта в г. Самаре, показавшее преимущество предлагаемого решения по сравнению с существующими.

Ключевые слова: общественное транспортное средство, прогнозирование времени прибытия, оценка времени прибытия, композиция алгоритмов, иерархическое разбиение, дерево регрессии.

Введение

Развитие современных средств глобального позиционирования сделало возможным решение целого ряда задач анализа транспортных систем, моделирования и прогнозирования [1]. Среди этого ряда задач одной из наиболее востребованных и понятных для конечного потребителя – участника дорожного движения – является задача прогнозирования времени движения транспортных средств. В настоящей работе рассматривается одна из возможных постановок такой задачи, заключающаяся в прогнозировании времени прибытия общественного транспортного средства (ОТС) на остановки. Решение этой задачи необходимо как для управления движением и внесения своевременных корректировок диспетчерскими службами, так и для оповещения пассажиров о времени прибытия ОТС на остановочные пункты.

Задаче прогнозирования времени прибытия ОТС и связанной с ней задаче оценки времени прохождения транспортными средствами (ТС) сегментов улично-дорожной сети (УДС) посвящено большое количество работ. Существующие методы можно условно разделить на две большие группы:

- статистические методы [2], использующие архивные данные и данные в реальном времени,
- методы моделирования [3], использующие модели транспортных потоков и модели распространения транспортных заторов.

Методы моделирования являются не столь распространёнными, как статистические методы, т.к. требуют актуальной модели транспортных потоков [3]. Статистические методы основываются на использовании различных моделей, учитывающих как архивные данные и данные в реальном времени о движении транспортных средств непосредственно, так и косвенную информацию, влияющую на дорожную ситуацию в целом. Ниже представлен их краткий обзор.

Модели на основе архивных данных строят прогноз скорости движения транспорта в определённый период времени по средней скорости в тот же период в предыдущие дни. Результаты этих моделей являются достоверными только тогда, когда схема движения транспорта является относительно стабильной в рассматриваемой области; в случае возникновения заторов и аварий точность этих моделей может сильно ухудшиться [4, 5].

Модели регрессии строятся как функции регрессии от набора независимых переменных, которые могут включать данные о прохождении дорожных сегментов в реальном времени, архивные данные, дорожные условия, пассажиропоток, погодные условия, задержки на остановках [6, 7]. Необходимое условие независимости переменных ограничивает применимость регрессионных моделей для транспортных систем, где переменные могут быть сильно коррелированы.

Также в оценке времени прибытия широко используются модели, основанные на *фильтрации Калмана* [8, 9, 10]. Хотя основной функцией моделей такого рода является прогноз текущего состояния системы, они могут служить основой для оценки будущих значений или для исправления предыдущих прогнозов. Модель может адаптироваться к колебаниям транспортного потока с зависящими от времени параметрами [9]; является эффективной для составления краткосрочных прогнозов.

Модели искусственных нейронных сетей применяются в транспортных задачах с начала 1990-х годов [11]. Популярность этих моделей объясняется их способностью моделировать сложные нелинейные отношения между временем прохождения сегментов сети и независимыми переменными, характеризующими дорожную ситуацию [10, 12].

Метод опорных векторов представляет собой набор алгоритмов вида «обучение с учителем», используемых для задач классификации и регрессии. Метод применялся для прогноза времени прибытия общест-

венного транспорта в работах [13, 14]; является вычислительно сложным, требует дальнейших исследований в вопросах выбора входных переменных и определения параметров алгоритма.

Гибридные модели являются объединением двух и более моделей для прогноза времени прибытия. В [15] используется объединение модели линейной регрессии и модели локально взвешенной линейной регрессии для повышения точности и надёжности прогнозов. В [16] показывается, что комбинация байесовской модели и модели нейронных сетей может давать хорошую оценку времени прохождения сегмента дорожной сети. Схожий подход к построению прогноза использовался в [17]. В работах [18, 19] предложен алгоритм прогноза, использующий фильтр Калмана для отслеживания местоположения транспортного средства и статистические оценки для прогнозирования времени прибытия. В [20] используется комбинированная схема из нейронной сети и фильтра Калмана.

В настоящей работе предложен *новый оригинальный алгоритм оценки (прогнозирования) времени прибытия ОТС на остановки общественного транспорта*, основанный на модели адаптивной композиции элементарных алгоритмов прогнозирования. Адаптивность подразумевает зависимость параметров конструируемой композиции от ряда управляющих параметров модели, к которым относятся следующие актуальные (определённые на текущий момент) факторы, влияющие на движение ТС и/или результат требуемого прогноза напрямую или косвенно: плотность транспортного потока, динамика движения, погодные условия и освещение, горизонт прогноза и др. В качестве используемых элементарных алгоритмов композиции могут быть выбраны алгоритмы достаточно произвольной природы, при этом в данной работе используются элементарные прогнозы по данным реального времени, прогнозы по архивным данным, расписанию движения транспорта и актуальному движению транспортного средства.

Работа построена следующим образом. В первом разделе вводятся основные понятия, выводятся основные соотношения, связывающие прогнозное время прибытия ОТС на остановку и время прохождения ОТС сегмента сети, даётся постановка задачи. Второй раздел посвящён описанию предлагаемой модели алгоритма оценки времени прохождения ОТС сегмента УДС, основанного на адаптивной комбинации элементарных алгоритмов прогноза. Третий и четвёртый разделы посвящены вопросам оценки параметров как элементарных алгоритмов, так и их композиции – агрегирующей функции. В пятом разделе приводится экспериментальное исследование разработанного алгоритма. В завершение работы приводятся выводы и благодарности, даётся список используемой литературы.

1. Основные обозначения и постановка задачи

Транспортную сеть определим как ориентированный граф, дуги которого соответствуют реальным участкам дорог (ниже – сегментам сети), а вершины

представляют собой разделяющие участки дорог узлы. Направление дуги определяет направления движения ТС на соответствующем участке сети. Введём дополнительные обозначения:

- w – конкретный сегмент сети, длину сегмента обозначим $|w|$;
- s – тип (сорт) ТС/ОТС; множество типов ТС/ОТС на конкретном сегменте w обозначим S^w ;
- m – номер маршрута ТС/ОТС; пара (s, m) определяет «маршрут» ТС/ОТС, при этом все маршруты считаются различными; на конкретном сегменте w множество номеров маршрутов с конкретным типом s ОТС обозначим M_s^w ;
- $W_{s,m}$ – множество сегментов сети, соответствующих (по которым проходит) конкретному маршруту ОТС (s, m) ;
- $v_{MAX(s,m)}^w(d, t)$ – максимальная скорость прохождения сегмента w дорожной сети ОТС конкретного маршрута m и конкретного типа s в определённый день d и определённое время t (определяется дорожными знаками);
- $j^w(d)$ – индекс, используемый для задания формального порядка прохождения конкретного сегмента w всеми ОТС. Далее для сокращения записи будет использоваться просто j ;
- $J^w(d)$ – множество индексов на сегменте w в конкретный день d ;
- $s(w, j)$, $m(w, j)$, – тип и маршрутный номер j -го ОТС из множества $J^w(d)$;
- ID – уникальный идентификатор ОТС. Может быть получен, например, по индексу прохождения ОТС конкретного сегмента: $ID(w, j)$. Тип и номер маршрута этого конкретного ОТС: $s(ID)$, $m(ID)$;
- d – день, однозначно идентифицируемый набором (год, день); всё множество дней обозначим Ψ ($d \in \Psi$);

– $\{\Psi_1, \Psi_2, \dots, \Psi_L\} \equiv \tilde{\Psi}$ – разбиение всего множества дней Ψ ($d \in \Psi$) на эквивалентные классы (например, дни недели). Должны выполняться соотношения: $\Psi = \bigcup_{\ell=1,L} \Psi_\ell$, $\Psi_i \cap \Psi_j = \emptyset$. Дополнительно опреде-

лим функцию $\psi(d) \in \tilde{\Psi}$, которая выдаёт класс/тип конкретного дня, называемого далее «типодень»;

– t_j – момент времени (начальный), в который j -ое ОТС из множества $J^w(d)$ появляется (въезжает) на сегмент w в день d ;

– $T_j(t)$ – длительность (в единицах времени) нахождения j -го ОТС из множества $J^w(d)$ на сегменте w в день d на момент t . Здесь возможны две принципиально различные ситуации. В первой ситуации в текущий момент t ОТС полностью проехало (миновало) сегмент и не находится на этом сегменте. В этой ситуации длительность $T_j(t)$ – это время прохождения сегмента. Во второй ситуации ОТС в момент t находится на сегменте. В этом случае длительность T_j – это время нахождения на сегменте (последнее из учтённых);

– $\Delta_j(t) \in [0, 1]$ – относительное положение j -го ОТС из множества $J^w(d)$ на сегменте w в день d на момент (текущий) времени t . В случае, если в текущий момент t ОТС полностью проехало сегмент, величина $\Delta_j(t) = 1$. Если ОТС в момент t находится на сегменте, величина $\Delta_j(t)$ показывает относительное положение ОТС на сегменте (последнее из учтённых);

– $J_s^w(d) = \{j \in J^w(d) : s(j) = s\}$ – множество индексов из $J^w(d)$ ОТС конкретного типа s ;

– $J_{s,m}^w(d) = \{j \in J^w(d) : s(j) = s \wedge m(j) = m\}$ – множество индексов ОТС конкретного типа и маршрута;

– $J_{id}^w(d) = \{j \in J^w(d) : ID(w, j) = id\}$ – множество индексов конкретного ОТС;

– t_c – текущий момент времени (момент, на который составляется прогноз);

– $v_j(t)$ – средняя скорость j -го ОТС в момент времени t (рассчитывается как средняя скорость за последние 30 минут);

– $T_{0(s,m)}^w(d, t)$ – нормативное время прохождения сегмента w дорожной сети, рассчитанное по расписанию движения ОТС конкретного маршрута m и конкретного типа s ;

– $T_{0(s)}^w(d, t)$ – среднее нормативное время прохождения сегмента w дорожной сети, рассчитанное по расписанию движения ОТС конкретного типа s . Определяется по формуле:

$$T_{0(s)}^w(d, t) = \frac{1}{|M_s^w|} \sum_{m \in M_s^w} T_{0(s,m)}^w(d, t),$$

где $|M_s^w|$ – мощность множества;

– $T_{\Sigma(s,m)}^w(d, t)$, $T_{\Sigma(s)}^w(d, t)$ – среднее (статистическое) время прохождения сегмента w дорожной сети ОТС (конкретного маршрута, или конкретного типа, или произвольного ТС) в определённый день d и определённое время t . Определяются по формулам:

$$T_{\Sigma(s,m)}^w(\tilde{d}, t) = \frac{1}{\sum_{d \in \Psi(\tilde{d})} \omega(d - \tilde{d})} \times \sum_{d \in \Psi(\tilde{d})} \omega(d - \tilde{d}) \frac{\sum_{j \in J_{s,m}^w(d)} \omega(t - t_j) T_j(t)}{\sum_{j \in J_{s,m}^w(d)} \omega(t - t_j)},$$

$$T_{\Sigma(s)}^w(\tilde{d}, t) = \frac{1}{\sum_{d \in \Psi(\tilde{d})} \omega(d - \tilde{d})} \times \sum_{d \in \Psi(\tilde{d})} \omega(d - \tilde{d}) \frac{\sum_{j \in J_s^w(d)} \omega(t - t_j) T_j(t)}{\sum_{j \in J_s^w(d)} \omega(t - t_j)}.$$

Здесь $\omega(t)$ – весовая функция по времени со свойствами: симметричная, положительная, невозрастающая по мере роста модуля аргумента, ограниченная по значению и носителю. Играет роль временного «окна». Например, можно задать в виде:

$$\omega(t) = \begin{cases} 1 - \frac{|t|}{\Delta_{\max}}, & |t| \leq \Delta_{\max}, \\ 0, & |t| > \Delta_{\max}. \end{cases}$$

Функция $\varpi(t)$ – это весовая функция по дням с необходимыми свойствами: симметричная, положительная, невозрастающая по мере роста модуля аргумента. Примеры: $\varpi(t) = 1$; $\varpi(t) = \exp(-\alpha|t|)$. В первом случае на результат собственно конкретный день не влияет, влияет только типодень.

Положение произвольного объекта на транспортной сети задаётся парой (сегмент, относительное положение в сегменте) (w, Δ) , где $\Delta \in [0, 1]$.

Постановка задачи

Пусть в конкретный момент t_c конкретное j -ое ОТС находится в положении (w_0, Δ_0) . Необходимо определить время, через которое оно будет находиться в заданном положении (на остановке) (w_K, Δ_K) .

Считаем, что маршрут между двумя положениями (w_0, Δ_0) и (w_K, Δ_K) включает в себя следующие сегменты из множества $W_{s(w_0, j), m(w_0, j)}$

$$(w_0, \Delta_0) \rightarrow w_1 \rightarrow w_2 \rightarrow \dots \rightarrow w_{K-2} \rightarrow w_{K-1} \rightarrow (w_K, \Delta_K),$$

при этом следующие положения считаются равными $(w_K, 1) = (w_{K+1}, 0)$.

Для представления формальной записи окончательного выражения используем ещё два обозначения:

$$T_{ID(w_0, j)}^{(w_{k1}, \Delta_{k1}) \rightarrow (w_{k2}, \Delta_{k2})}(d, t_c, t), \quad (1)$$

$$T_{ID(w_0, j)}^{w_{k1}}(d, t_c, t) \equiv T_{ID(w_0, j)}^{(w_{k1}, 0) \rightarrow (w_{k1}, 1)}(d, t_c, t),$$

первое из которых обозначает прогнозное время прохождения интервала транспортной сети $(w_{k1}, \Delta_{k1}) \rightarrow \dots \rightarrow (w_{k2}, \Delta_{k2})$ конкретным ОТС с $ID(w_0, j)$. Прогноз рассчитывается в момент времени t_c в предположении, что ОТС попадёт в положение (w_{k1}, Δ_{k1}) в момент времени t . Вторая величина есть прогнозное время прохождения конкретного сегмента w_{k1} транспортной сети конкретным ОТС с $ID(w_0, j)$. Параметры t_c и t имеют то же значение, что и выше.

Нас интересует прогнозное время прихода ОТС $ID(w_0, j)$ на остановку (w_K, Δ_K) , рассчитываемое на момент t_c и обозначаемое далее:

$$T_{ID(w_0, j)}^{(w_0, \Delta_0) \rightarrow (w_K, \Delta_K)}(d, t_c) \equiv T_{ID(w_0, j)}^{(w_0, \Delta_0) \rightarrow (w_K, \Delta_K)}(d, t_c, t_c).$$

Справедливы следующие рекуррентные соотношения, связывающие прогнозные времена «вхождения в сегменты» для ОТС:

$$\begin{aligned}
T_{ID(w_0, j)}^{(w_0, \Delta_0) \rightarrow (w_r, 0)}(d, t_c) = \\
= T_{ID(w_0, j)}^{(w_0, \Delta_0) \rightarrow (w_1, 0)}(d, t_c, t_c) + \\
+ \sum_{k=1}^{r-1} T_{ID(w_0, j)}^{(w_k, 0) \rightarrow (w_{k+1}, 1)}(d, t_c, t_c + T_{ID(w_0, j)}^{(w_0, \Delta_0) \rightarrow (w_k, 0)}(d, t_c)).
\end{aligned}$$

Используя обозначение для прогнозируемого времени прохождения сегмента, получаем:

$$\begin{aligned}
T_{ID(w_0, j)}^{(w_0, \Delta_0) \rightarrow (w_r, 0)}(d, t_c) = \\
= T_{ID(w_0, j)}^{(w_0, \Delta_0) \rightarrow (w_1, 0)}(d, t_c, t_c) + \\
+ \sum_{k=1}^{r-1} T_{ID(w_0, j)}^{w_k}(d, t_c, t_c + T_{ID(w_0, j)}^{(w_0, \Delta_0) \rightarrow (w_k, 0)}(d, t_c)).
\end{aligned} \quad (2)$$

Окончательно искомое время прихода заданного ОТС на остановку имеет вид:

$$\begin{aligned}
T_{ID(w_0, j)}^{(w_0, \Delta_0) \rightarrow (w_K, \Delta_K)}(d, t_c) = \\
= T_{ID(w_0, j)}^{(w_0, \Delta_0) \rightarrow (w_K, 0)}(d, t_c) + \\
+ T_{ID(w_0, j)}^{(w_K, 0) \rightarrow (w_K, \Delta_K)}(d, t_c, t_c + T_{ID(w_0, j)}^{(w_0, \Delta_0) \rightarrow (w_K, 0)}(d, t_c)).
\end{aligned} \quad (3)$$

Предполагая монотонность движения (движение без ускорения) на каждом из сегментов и/или их частей, выражения (2) и (3) (в части первого и последнего слагаемых) преобразуются к виду:

$$\begin{aligned}
T_{ID(w_0, j)}^{(w_0, \Delta_0) \rightarrow (w_r, 0)}(d, t_c) = \\
= (1 - \Delta_0) T_{ID(w_0, j)}^{w_0}(d, t_c, t_c) + \\
+ \sum_{k=1}^{r-1} T_{ID(w_0, j)}^{w_k}(d, t_c, t_c + T_{ID(w_0, j)}^{(w_0, \Delta_0) \rightarrow (w_k, 0)}(d, t_c)).
\end{aligned} \quad (4)$$

$$\begin{aligned}
T_{ID(w_0, j)}^{(w_0, \Delta_0) \rightarrow (w_K, \Delta_K)}(d, t_c) = \\
= T_{ID(w_0, j)}^{(w_0, \Delta_0) \rightarrow (w_K, 0)}(d, t_c) + \\
+ \Delta_K T_{ID(w_0, j)}^{w_K}(d, t_c, t_c + T_{ID(w_0, j)}^{(w_0, \Delta_0) \rightarrow (w_K, 0)}(d, t_c)).
\end{aligned} \quad (5)$$

Как видно из приведённых выражений, их основными составляющими являются слагаемые, характеризующие время прохождения отдельных сегментов $T_{ID(w_0, j)}^{w_k}(\dots)$.

Частный случай (однородная по времени модель)

Если допустить независимость прогнозируемого времени прохождения сегмента (что в общем случае неверно) от момента «вхождения» в этот сегмент (для обозначения такой модели будем использовать понятие «однородная по времени»), то есть допустить равенство

$$T_{ID(w_0, j)}^{w_k}(d, t_c, t_c) = T_{ID(w_0, j)}^{w_k}(d, t_c, t_c + t) \quad \forall t > 0, \quad (6)$$

тогда выражения (4)–(5) преобразуются в простую взвешенную сумму прогнозных величин по сегментам (без рекурсивной зависимости по задержкам вхождения):

$$T_{ID(w_0, j)}^{(w_0, \Delta_0) \rightarrow (w_r, 0)}(d, t_c) = \sum_{k=0}^K \tilde{\Delta}_k T_{ID(w_0, j)}^{w_k}(d, t_c), \quad (7)$$

где $\tilde{\Delta}_0 = 1 - \Delta_0$, $\tilde{\Delta}_K = \Delta_K$, $\tilde{\Delta}_k = 1$, $k = \overline{1, K-1}$.

Модель (4)–(5) в отличие от (7) будем называть «неоднородной по времени».

Таким образом, основной задачей при прогнозировании времени прибытия (4)–(5) или (7) с использованием предложенной модели является оценка времени прохождения конкретного сегмента w_k конкретным ОТС $ID(w_0, j)$ в конкретное время

$$T_{ID(w_0, j)}^{w_k}(d, t_c, t), \quad t \geq t_c, \quad (8)$$

то есть построение следующей оценки

$$\hat{T}_{ID(w_0, j)}^{w_k}(d, t_c, t), \quad t \geq t_c. \quad (9)$$

2. Оценка времени прохождения ОТС конкретного сегмента. Модель адаптивной композиции элементарных прогнозов

Конструируемая оценка (9) должна учитывать следующие специфические величины (8). Во-первых, эта величина характеризует время прохождения сегмента w_k совершенно конкретным транспортным средством с идентификатором $ID(w_0, j)$. При этом на том же сегменте может выполняться движение других транспортных средств, в частности:

– ОТС того же маршрута $s(w_0, j)$, $m(w_0, j)$,

– ОТС того же типа $s(w_0, j)$,

– ОТС других типов и другие ТС, то есть все ТС из множества $J^w(\dots)$, причём произвольные ТС (в том числе индивидуальные ТС) для упрощения могут рассматриваться как ТС некоторого специального типа – маршрутными номерами в этом случае могут являться их регистрационные номера.

Замена прогнозного времени для конкретного ОТС с идентификатором $ID(w_0, j)$ на величину прогнозного времени другого ОТС того же типа или ТС другого типа сопровождается, с одной стороны, потенциальным увеличением используемых в построении оценки (9) числа ТС и, с другой стороны, игнорированием в конструируемой оценке специфики движения, присутствующей у конкретного ОТС, конкретного маршрута и т. д. Ниже даны некоторые комментарии, отражающие такую специфику.

1) Замена прогнозного времени для конкретного ОТС с идентификатором $ID(w_0, j)$ на прогнозное время произвольного ОТС конкретного маршрута $s(w, j)$, $m(w, j)$ игнорирует в конструируемой оценке (9) специфику конкретного ОТС. Поскольку различными ОТС управляют различные водители и сами ОТС могут быть сконструированы на различных платформах (например, автобусы разных производителей), их движение происходит по-разному. Фактически это выражается в различиях в скорости движения различных ОТС одного маршрута. Соответствующее изменение может быть модельно описано следующим образом:

$$T_{ID(w_0, j)}^{w_k}(d, t_c, t) = a_{ID(w_0, j)}(d, t) \cdot T_{s(w_0, j), m(w_0, j)}^{w_k}(d, t_c, t), \quad (10)$$

где $a_{ID}(d, t) \in \mathbb{R}_+$ – некоторый коэффициент, учитывающий специфику движения конкретного ОТС (специфику конкретного ОТС и водителя) по сравнению с ОТС того же маршрута, а $T_{s(w_0, j), m(w_0, j)}^{w_k}(\dots)$ – среднее время прохождения конкретного сегмента ОТС конкретного маршрута. Выражение (10) дополнительно показывает, что различия в этих величинах не зависят от сегмента, а зависят только от идентификатора ОТС.

2) Замена прогнозного времени для ОТС конкретного маршрута $s(w, j)$, $m(w, j)$ на прогнозные время произвольного ОТС конкретного типа $s(w, j)$ игнорирует в конструируемой оценке (9) специфику конкретного маршрута. Такой спецификой для конкретного сегмента может являться:

– наличие дополнительных остановок, приводящих к дополнительной задержке по времени на конкретном сегменте сети (пример: автобусные маршруты «обычный» и «скорый», где у второго исключён ряд остановок первого);

– различия в нагрузке на маршрут (его полезности), выражающиеся в различном числе пассажиров на остановках, осуществляющих посадку на ОТС соответствующих маршрутов, также приводят к дополнительной задержке по времени на конкретном сегменте сети.

Вышеназванные различия могут быть отражены в математической модели следующим образом:

$$T_{s(w_0, j), m(w_0, j)}^{w_k}(d, t_c, t) = b_{m(w_0, j)/s(w_0, j)}^{w_k}(\psi(d), t) T_{s(w_0, j)}^{w_k}(d, t_c, t), \quad (11)$$

где $b_{m(w_0, j)/s(w_0, j)}^{w_k}(\psi(d), t)$ – коэффициент, мультипликативно учитывающий специфику движения ОТС конкретного маршрута по сравнению с ОТС конкретного типа, а $T_{s(w_0, j)}^{w_k}(\dots)$ – среднее время прохождения конкретного сегмента ОТС конкретного типа.

3) Величина $T_s^w(d, t_c, t)$ определяет прогнозные время прохождения конкретного сегмента w ОТС типа s в момент времени t дня d при условии, что прогноз вычисляется в момент времени t_0 . Очевидно, построение такой прогнозной величины может производиться с использованием различных моделей и различной информации. Перечислим вначале состав информации, которая (неочевидным образом и в неочевидной степени) оказывает влияние на значение $T_s^w(d, t_c, t)$. Для удобства введём индекс «последнего» по порядку ОТС на сегменте w конкретного типа s :

$$\kappa \equiv \kappa_s^w(d, t) = \arg \max \{j : j \in J_s^w(d) \wedge t_j \leq t\}. \quad (12)$$

В частности, в момент построения прогноза индекс $\kappa_s^w(d, t_c)$ совпадает с индексом последнего ОТС того же типа s , прошедшим сегмент w .

Тогда состав искомой информации может быть определён следующим образом:

1) $\tau_s^w(d, t_c) \equiv t_c - (t_{\kappa_s^w(d, t_c)} + T_{\kappa_s^w(d, t_c)}(t_c))$ – диапазон времени, прошедший с момента прохождения сегмента последним ОТС того же типа (на момент построения прогноза);

2) $t - t_c$ – требуемый горизонт прогноза;

3) $T_{\kappa_s^w(d, t_c)}(t_c)$ – время нахождения на сегменте последнего (на момент построения прогноза) ОТС;

4) $\tilde{T}_{\kappa_s^w(d, t_c)}(t_c) \equiv T_{\kappa_s^w(d, t_c)}(t_c) / \Delta_{\kappa_s^w(d, t_c)}$ – «наивная» оценка времени прохождения сегмента ОТС с порядковым номером $\kappa_s^w(d, t_c)$, совпадающая с реальным временем нахождения указанного ОТС на сегменте в случае, если ОТС уже оказалось на другом сегменте, и элементарным образом экстраполирующая время прохождения сегмента в противном случае;

5) $T_{0(s)}^w(d, t)$ – нормативное время прохождения сегмента w ОТС того же типа s ;

6) $T_{\Sigma(s)}^w(d, t)$ – статистическое время прохождения сегмента w ОТС того же типа s ;

7) $T_{v(s)}^w(d, t) \equiv |w| / \min(v_j(t), v_{MAX(s, m)}^w(d, t))$ – предположительное время прохождения сегмента с учётом средней скорости конкретного ОТС;

8) $\{(t - t_j, T_j(t_c), \Delta_j(t_c))\}_{j \in J^w(d)}$ – множество прецедентов, отражающих моменты и время прохождения сегмента различными ОТС к моменту прогноза t_c ;

9) $\gamma(d, t) \in [0, 1]$ – условное значение, интегрально характеризующее на момент прогноза t сложность вождения при данном освещении и при данных погодных условиях;

10) $\rho^w(d, t) \in [0, 1]$ – условное значение, характеризующее плотность потока транспортных средств на участке сети в конкретное время;

11) s – тип ОТС.

Указанные величины несут различную содержательную информацию, влияющую на способ её использования, а именно: величины 3–8 отражают отдельные или статистически агрегированные «прецеденты» временных затрат. Очевидно, что указанные величины могут быть использованы (вместе или отдельно, напрямую или косвенно) при построении искомой оценки как некоторые реализации искомой оценки (неискажённые или искажённые заранее определённым образом).

Величины 1–2 и 9–11 непосредственно оказывают влияние на искомую величину. Однако ни одна из этих величин не может быть рассмотрена как прецедент временных затрат. Как следствие, указанные величины можно рассматривать как *управляющие*, характеризующие алгоритм композиции и/или его параметры.

Учитывая сделанные замечания, величину $\hat{T}_s^w(d, t_c, t)$ целесообразно рассматривать в виде некоторой функции следующего вида:

$$T_s^w(d, t_c, t) = f_{s, \tau_s^w(d, t_c), t-t_c, \gamma(d, t), \rho^w(d, t)} \left(T_{\kappa_s^w(d, t_0)}(t_c), \bar{T}_{\kappa_s^w(d, t_0)}(t_c), T_{0(s)}^w(d, t), T_{\Sigma(s)}^w(d, t), T_{\nu(s)}^w(d, t), \left\{ (t-t_j, T_j(t_c), \Delta_j(t_c)) \right\}_{j \in J^w(d)} \right). \quad (13)$$

Вид этой функции оказывается, естественно, неизвестным. Выбор конкретного вида этой функции – это вопрос выбора соответствующей математической модели с последующим решением вопросов идентификации и верификации. В данной работе предлага-

ется следующий её вид, удовлетворяющий формальному соотношению (13) и основанный на композиции алгоритмов, выполняющих элементарные прогнозы для каждого типа ОТС. Схематический вид модели выглядит следующим образом (рис. 1):

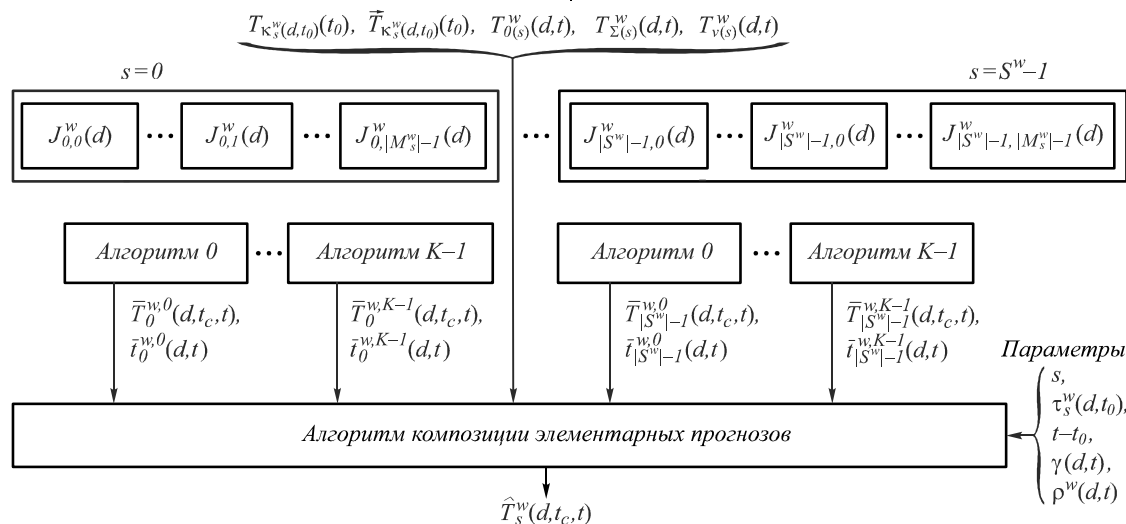


Рис. 1. Схематический вид модели адаптивной композиции элементарных прогнозов

Представленная на рисунке модель определена не полностью. В частности, нет уточнений о том, как именно реализуются алгоритмы элементарных прогнозов, рассчитывающие оценки средних $\bar{T}_s^{w,k}(d, t_c, t)$, и не определена функция агрегации этих величин. Ниже представлен их вид.

Алгоритмы элементарных прогнозов реализуют отображения множества прецедентов $\{(t-t_j, T_j(t_c), \Delta_j(t_c))\}_{j \in J_s^w(d, t_c)}$ в различные оценки средних $\bar{T}_s^{w,k}(d, t_c, t)$. В качестве подобных отображений в работе предлагается использовать следующие:

$$\bar{T}_s^{w,k}(d, t_c, t) = \frac{\sum_{j \in J_s^w(d, t_c)} \omega_k(t-t_j) T_j(t_c)}{\sum_{j \in J_s^w(d, t_c)} \omega_k(t-t_j)}, \quad (14)$$

$$\bar{T}_s^{w,k}(d, t) = \frac{\sum_{j \in J_s^w(d, t_c)} \omega_k(t-t_j) t_j}{\sum_{j \in J_s^w(d, t_c)} \omega_k(t-t_j)}, \quad k = \overline{0, K-1}.$$

Предлагаемые оценки относятся к классу ядерных, при этом различия в алгоритмах заключаются в используемых ядрах $\omega_k(t)$. Дополнительное требование, предъявляемое к используемому набору ядер $\omega_k(t)$, – это их линейная независимость на рассматриваемой области (прогнозном горизонте). Для определенности предлагается использовать следующий набор из четырёх ядер:

– прямоугольное ядро ($k=0$):

$$\omega_0(t) = \begin{cases} \frac{1}{\Delta_{\max}}, & |t| \leq \Delta_{\max}, \\ 0, & |t| > \Delta_{\max}; \end{cases}$$

– треугольное ядро ($k=1$)

$$\omega_1(t) = \begin{cases} 1 - \frac{|t|}{\Delta_{\max}}, & |t| \leq \Delta_{\max}, \\ 0, & |t| > \Delta_{\max}; \end{cases}$$

– экспоненциальное ядро ($k=2$)

$$\omega_2(t) = \begin{cases} e^{-\alpha \frac{t}{\Delta_{\max}}}, & |t| \leq \Delta_{\max}, \\ 0, & |t| > \Delta_{\max}; \end{cases}$$

– рациональное ядро ($k=3$)

$$\omega_3(t) = \begin{cases} \frac{1}{1 + \alpha \frac{t}{\Delta_{\max}}}, & |t| \leq \Delta_{\max}, \\ 0, & |t| > \Delta_{\max}. \end{cases}$$

Область определения финитных ядер должна совпадать. Дополнительно следует заметить, что в целом предлагаемая модель допускает использование произвольного набора алгоритмов оценки среднего, не обязательно совпадающих с выбранными.

Алгоритм композиции элементарных прогнозов реализует отображение набора оценок средних $\{\bar{T}_s^{w,k}(d, t_c, t)\}_{s=0, |S^w|-1, k=0, K-1}$ и величин $T_{\kappa_s^w(d, t_c)}(t_c)$, $\bar{T}_{\kappa_s^w(d, t_c)}(t_c)$,

$T_{0(s)}^w(d, t)$, $T_{\Sigma(s)}^w(d, t)$, $T_{v(s)}^w(d, t)$ в итоговую оценку $\hat{T}_s^w(d, t_c, t)$ с учётом значений управляющих параметров $s, \tau_s^w(d, t_c), t - t_c, \gamma(d, t), \rho^w(d, t)$. Представляется целесообразным, чтобы этот алгоритм удовлетворял двум основным требованиям:

- был адаптивен по отношению к изменениям дорожной ситуации,
- был адаптивен по отношению к значению параметров.

Первое требование подразумевает, что алгоритм выполняет композиции временных оценок по-разному, в зависимости от динамики транспортного потока (изменения скорости движения ОТС). В данном случае, поскольку вся информация о динамике содержится в прецедентах $\{(t - t_j, T_j(t_c) \Delta_j(t_c))\}_{j \in J_{s,m}^w(d, t_c)}$, в качестве дополнительного параметра, влияющего на алгоритм композиции оценок, можно использовать следующую величину:

$$\eta(d, t) = \frac{\sum_{j \in J^w(d, t_c)} \omega_1(t - t_j) T_j(t_c) \Delta_j^{-1}(t_c)}{\sum_{j \in J^w(d, t_c)} \omega_1(t - t_j)} - \frac{\sum_{j \in J^w(d, t_c)} (2\omega_0(t - t_j) - \omega_1(t - t_j)) T_j(t_c) \Delta_j^{-1}(t_c)}{\sum_{j \in J^w(d, t_c)} (2\omega_0(t - t_j) - \omega_1(t - t_j))}. \quad (15)$$

Уменьшаемое в этом выражении соответствует треугольному ядру, спадающему по мере «удаления» по временной шкале от момента прогноза, а вычитаемое – наоборот возрастающему. Результирующая разность оказывается положительной, если время прохождения сегмента для транспорта имеет динамику к увеличению, отрицательной – если имеет динамику к уменьшению. Абсолютная величина характеризует выраженность этой динамики: при нахожде-

$$\hat{T}_s^w(d, t_c, t) = \frac{\sum_{k=0}^{K-1} \sum_{r \in S^w} a_{rs}^k \omega(t - \bar{t}_r^{w,k}(d, t)) \bar{T}_r^{w,k}(d, t) + a_{1s}^* T_{\kappa_s^w(d, t_c)}^w(t_c) + a_{2s}^* \bar{T}_{\kappa_s^w(d, t_c)}^w(t_c) + a_{3s}^* T_{0(s)}^w(d, t) + a_{4s}^* T_{\Sigma(s)}^w(d, t) + a_{5s}^* T_{v(s)}^w(d, t)}{\sum_{k=0}^{K-1} \sum_{r \in S^w} a_{rs}^k \omega(t - \bar{t}_r^{w,k}(d, t)) + a_1^* + a_2^* + a_3^* + a_4^* + a_5^*}. \quad (16)$$

Параметры этого выражения

$$\{a_{rs}^k\} \cup \{a_{1s}^*, a_{2s}^*, a_{3s}^*, a_{4s}^*, a_{5s}^*\}$$

определяются отдельно для каждого типа s ОТС и подобластей параметров

$$\tau_s^w(d, t_c), t - t_c, \gamma(d, t), \rho^w(d, t), \eta(d, t)$$

алгоритма агрегации.

Определение областей для пяти параметров $\tau_s^w(d, t_c), t - t_c, \gamma(d, t), \rho^w(d, t), \eta(d, t)$ осуществляется автоматически в процессе конструирования иерархической регрессионной конструкции, подобной дереву регрессии, а именно: для каждого из параметров (используем далее для указанных пяти параметров обо-

нии разности вблизи нуля изменение скорости на участке (динамику) можно считать несущественным.

Второе требование подразумевает, что алгоритм композиции выполняет агрегацию временных оценок по-разному, в зависимости от значения параметров. В частности:

- в зависимости от искомого типа ОТС s ;
- в зависимости от «удалённости» по времени между моментами получения информации о движении $(\tau_s^w(d, t_c))$ и моментом, на который прогноз вычисляется $(t - t_c)$ (например, при малой «удалённости» между этими моментами факт остановки предшествующего рельсового ОТС приводит к остановкам последующих),
- в зависимости от сложности вождения, характеризующей величиной $\gamma(d, t)$;
- в зависимости от плотности потока транспортных средств на участке сети, характеризующей величиной $\rho^w(d, t)$ (например, высокая плотность транспортного потока приводит практически к полному совпадению скоростей движения ОТС различных типов).

Подобная адаптивность может быть достигнута путём разбиения области определения параметров алгоритма агрегации на подобласти, в каждой из которых используется агрегация средних $\{\bar{T}_s^{w,k}(d, t, t_c)\}_{s=0, \overline{S^w}-1, k=0, K-1}$ и величин

$$T_{\kappa_s^w(d, t_c)}^w(t_0), \bar{T}_{\kappa_s^w(d, t_c)}^w(t_c), T_{0(s)}^w(d, t), T_{\Sigma(s)}^w(d, t), T_{v(s)}^w(d, t)$$

в итоговую оценку $\hat{T}_s^w(d, t_c, t)$ с использованием заранее заданного аналитического выражения, определённого с точностью до некоторого набора параметров (этого выражения), которые и определяются для указанной подобласти параметров алгоритма в некотором смысле оптимальным образом.

Предлагается в качестве агрегирующей функции использовать следующее выражение:

значение p_i) заранее определяется интервал его допустимых значений $[p_i^{\min}, p_i^{\max}]$. Комментарии относительно значения границ интервалов для параметров приведены в табл. 1.

Далее для области

$$[p_0^{\min}, p_0^{\max}] \times [p_1^{\min}, p_1^{\max}] \times [p_2^{\min}, p_2^{\max}] \times [p_3^{\min}, p_3^{\max}] \times [p_4^{\min}, p_4^{\max}],$$

соответствующей единственной терминальной вершине дерева на первом шаге, производится оценка параметров агрегирующей функции с использованием подхода, описанного ниже. Далее для отобранных

на текущем шаге терминальных вершин производится их разбиение на 2^5 новые терминальные вершины путём деления области значений каждого из параметров на две, например:

$$\left[p_0^{\min}, p_0^{\max} \right] \rightarrow \left[p_0^{\min}, p_0^{\min} + \frac{p_0^{\max} - p_0^{\min}}{2} \right], \left[p_0^{\min} + \frac{p_0^{\max} - p_0^{\min}}{2}, p_0^{\max} \right].$$

Отличием от известного дерева регрессии в данном случае является то, что значения параметров, по которым производится иерархическое разбиение, не используются в расчёте функции регрессии (16).

Таблица 1. Значение границ интервалов параметров

Параметр	Обозначение	Левая граница	Правая граница
$\tau_s^w(d, t_0)$	p_0	0	Определяется по данным
$t - t_0$	p_1	0	Определяется заданным прогнозным горизонтом
$\gamma(d, t)$	p_2	0	1
$\rho^w(d, t)$	p_3	0	1
$\eta(d, t)$	p_4	Определяется по данным	Определяется по данным

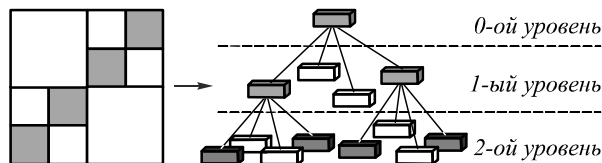


Рис. 2. Пример иерархического построения адаптивной композиции для двух управляющих параметров

3. Оценка параметров элементарных алгоритмов прогноза

Для реализации элементарных прогнозных функций необходимо оценить их параметры: Δ_{\max} (область определения ядра) и α (для экспоненциального и рационального ядер).

Параметр Δ_{\max} определяется по минимуму среднеквадратической ошибки элементарной прогнозных функций с треугольным ядром. В качестве обучающей выборки использовалась информация о прохождении всех сегментов сети всеми ОТС за день. На рис. 3 показан график зависимости СКО от значения параметра Δ_{\max} .

Нужно учесть, что при малых значениях Δ_{\max} множество прецедентов для составления элементарных прогнозов будет пустым. Для реализации элементарных функций было выбрано значение параметра $\Delta_{\max} = 2700$ (секунд).

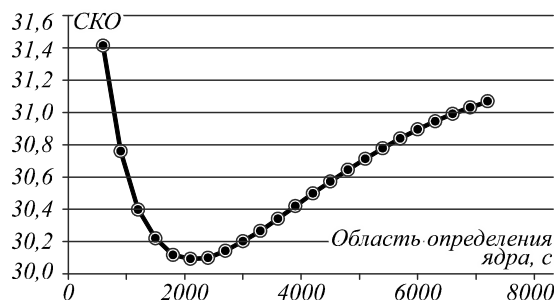


Рис. 3. Зависимость СКО от области определения ядра

Выбор значения параметра α для экспоненциальной и рациональной функции проводился по минимуму СКО при фиксированном параметре Δ_{\max} . Зависимость СКО от α для экспоненциальной функции приведено в табл. 2, для рациональной – в табл. 3.

Таблица 2. Зависимость СКО для экспоненциальной функции

α	50	25	12,5	6,25	3,125
СКО	33,238	32,33	31,08	29,95	29,497
α	1,57	2,35	2,73	2,55	2,63
СКО	29,54	29,483	29,482	29,4802	29,4804

Таблица 3. Зависимость СКО для рациональной функции

α	50	25	37,5	31,25
СКО	29,5087	29,509	29,50832	29,50834
α	34,39	35,9		
СКО	29,5083	29,5083		

Для экспоненциальной функции выбрано $\alpha = 2,55$, для рациональной – $\alpha = 35,9$

4. Оценка параметров модели адаптивной композиции

Для определения параметров адаптивной композиции для каждой терминальной вершины конструируемой иерархической регрессии необходимо сформировать обучающую выборку. Ниже представлен предложенный способ её формирования.

Зафиксируем сегмент w , тип ТС s и типодень \tilde{d} . Для этих величин по обучающим данным (дни с выборками) для каждой текущей терминальной вершины иерархической регрессии формируем выборки следующим образом.

ЦИКЛ по $d \in \Psi(\tilde{d})$

ЦИКЛ по номерам ОТС $i \in J_s^w(d)$

$$t^* = t_i; T^* = T_i (\Delta = 1); id = ID(w, i);$$

ЦИКЛ по временам прогноза $t_c = t^* - n\Delta t$ ($\Delta t = 30c$, $n = 1, 2, \dots$)

Определяем местоположение ОТС с id в это время. Формируем набор параметров прогнозирования для данного ОТС с моментом построения прогноза t_c и моментом прогноза $t^* (\equiv t)$

для сегмента w :

$$\tau_s^w(d, t_c), t - t_c, \gamma(d, t), \rho^w(d, t), \eta(d, t).$$

Формируем набор прецедентов

$$\left\{ \left(t - t_j, T_j(t_c), \Delta_j(t_c) \right) \right\}_{j \in J_{s,m}^w(d,t_c)} \text{ и считаем по ним}$$

$$\text{элементарные прогнозы } \left\{ \bar{T}_s^{w,k}(d,t,t_c) \right\}_{s=0, \lfloor S^w \rfloor - 1, k=0, K-1}.$$

Результаты расчётов элементарных прогнозов

$$\left\{ \bar{T}_s^{w,k}(d,t,t_c) \right\}_{s=0, \lfloor S^w \rfloor - 1, k=0, K-1}, \text{ величины}$$

$$T_{\kappa_s^w(d,t_c)}(t_0), \bar{T}_{\kappa_s^w(d,t_c)}(t_c), T_{0(s)}^w(d,t), T_{\Sigma(s)}^w(d,t), T_{v(s)}^w(d,t)$$

и требуемый результат T^* добавляется в список терминальной вершины.

КОНЕЦ_ЦИКЛА

КОНЕЦ_ЦИКЛА

КОНЕЦ_ЦИКЛА

По завершении формирования обучающих выборок для каждой терминальной вершины производится оценка параметров адаптивной композиции по методу минимума СКО.

Если размер обучающей выборки терминальной вершины недостаточен для оценки параметров композиции, используются оценки, полученные на родительской вершине.

5. Экспериментальные исследования

Экспериментальные исследования разработанного алгоритма проводились на улично-дорожной сети г. Самары. Дорожная сеть состоит из 3387 сегментов, трамвайная сеть – из 409 сегментов. Количество ОТС, подключённых к системе мониторинга, – более 1500, новые координаты положения ОТС поступают с усреднённой периодичностью в 30 секунд. Подробнее система мониторинга движения описана в работе [5].

При составлении прогнозов проводилась дополнительная фильтрация данных, попадающих в список прецедентов $\left\{ \left(t - t_j, T_j(t_c), \Delta_j(t_c) \right) \right\}_{j \in J^w(d)}$, а именно:

не рассматривались ТС, находящиеся на конечных точках маршрута либо не изменявшие своего положения в течение долгого промежутка времени (больше 5 минут). Для упрощения нахождения решения специфики движения конкретного ТС и ТС определённого маршрута не учитывались, т.е. предполагалось, что $a_{ID}(d,t) = 1$ и $b_{m/s}^w(\psi(d),t) = 1$. Кроме того, не проводилась агрегация прогнозов от ТС разных типов.

Проводилось сравнение предложенного алгоритма адаптивной композиции (с элементарными прогнозами $\bar{T}_s^{w,k}(d,t_c,t)$ для каждого типа ядра, прогнозом по статистике $T_{\Sigma(s)}^w(d,t)$ и прогнозом по средней скорости движения $T_{v(s)}^w(d,t)$) с моделью линейной регрессии, представленной в работе [6]. Сравнение проводилось по критериям СКО, средней абсолютной ошибки и средней относительной ошибки.

Результаты сравнения алгоритмов на обучающей и контрольной выборках для ТС трамвайных маршрутов приведены в таблице 4. Размер выборок составлял порядка 2500000 записей. Стоит отметить, что контрольная выборка содержала большее число прогнозов в статистике по сравнению с обучающей выборкой, что снизило ошибку прогноза по статистическим данным.

Таблица 4. Сравнение алгоритмов для ОТС трамвайных маршрутов

Алгоритм прогноза	СКО		Средняя абсолютная ошибка		Средняя относительная ошибка	
	Обучающая	Контрольная	Обучающая	Контрольная	Обучающая	Контрольная
Прямоугольное ядро	28,492	28,990	17,526	17,420	0,342	0,327
Треугольное ядро	28,605	28,949	17,650	17,582	0,346	0,330
Экспоненциальное ядро	28,672	28,995	17,704	17,643	0,347	0,331
Рациональное ядро	29,971	30,164	18,530	18,424	0,361	0,343
Статистика	40,060	32,514	21,884	18,708	0,392	0,343
Скорость	37,812	39,003	25,197	25,505	0,380	0,370
Регрессия	25,955	28,552	16,111	16,767	0,315	0,314
Композиция	23,361	27,963	14,991	16,350	0,287	0,307

Графики зависимости СКО, средней абсолютной ошибки и средней относительной ошибки от горизонта прогноза на обучающей выборке для ОТС трамвайных маршрутов показаны на рис. 4, 5 и 6 соответственно.

По графикам видно, что элементарные прогнозы дают очень близкую ошибку прогнозирования для разных типов ядер. Предложенная модель с компози-

цией различных типов прогнозов даёт лучшие результаты по каждому критерию.

На рис. 7, 8 и 9 показаны графики зависимости СКО, средней абсолютной ошибки и средней относительной ошибки от горизонта прогноза на контрольной выборке для ТС трамвайных маршрутов.

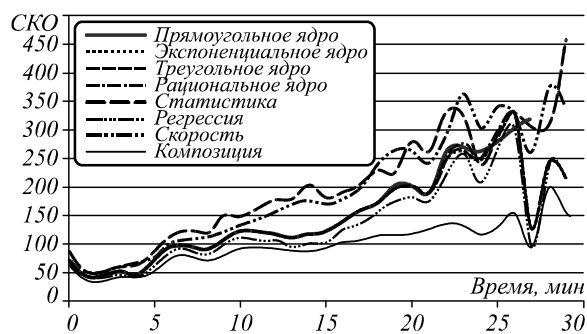


Рис. 4. Зависимость СКО от горизонта прогноза на обучающей выборке

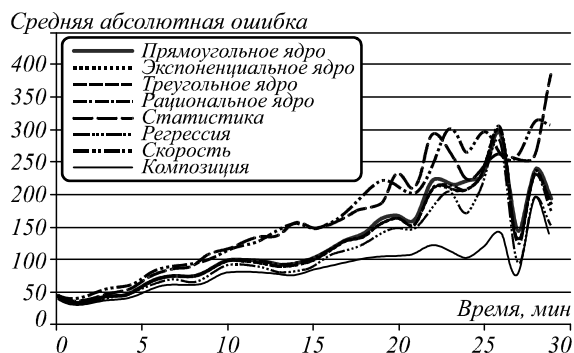


Рис. 5. Зависимость средней абсолютной ошибки от горизонта прогноза на обучающей выборке

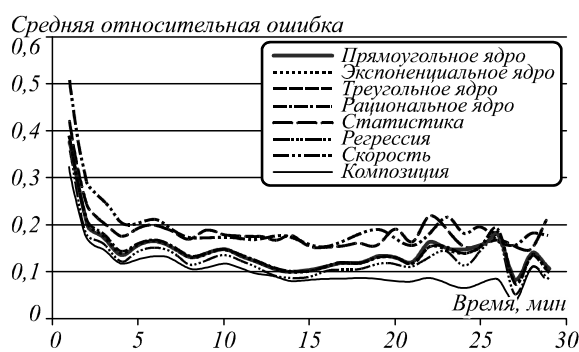


Рис. 6. Зависимость средней относительной ошибки от горизонта прогноза на обучающей выборке

На контрольной выборке разработанная модель для некоторых прогнозных горизонтов проигрывает элементарным прогнозам, что, судя по поведению графиков, связано с достаточно неудачным статистическим прогнозом по историческим данным (предыдущим дням).

Результаты сравнения алгоритмов на обучающей и контрольной выборках для ТС автобусных маршрутов приведены в табл. 5. Размер выборок составлял порядка 8 млн. записей.

Графики зависимости СКО, средней абсолютной ошибки и средней относительной ошибки от горизонта прогноза на обучающей выборке для ТС автобусных маршрутов показаны на рис. 10, 11 и 12 соответственно.

Элементарные прогнозы для ТС автобусных маршрутов дают большую ошибку по сравнению с прогнозами для ТС трамвайных маршрутов. Это объясняется сильной изменчивостью транспортной ситуации на дорожных сегментах в течение дня, в то время как трамвайные сегменты часто обособлены и не так сильно зависят от транспортной ситуации в городе.

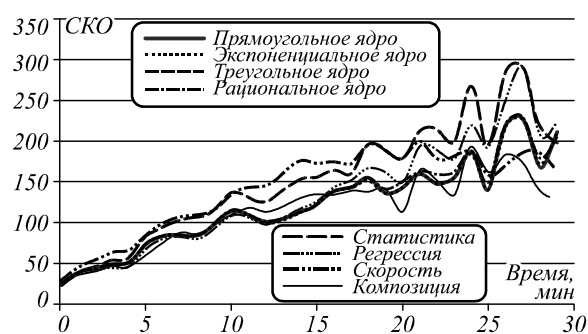


Рис. 7. Зависимость СКО от горизонта прогноза на контрольной выборке

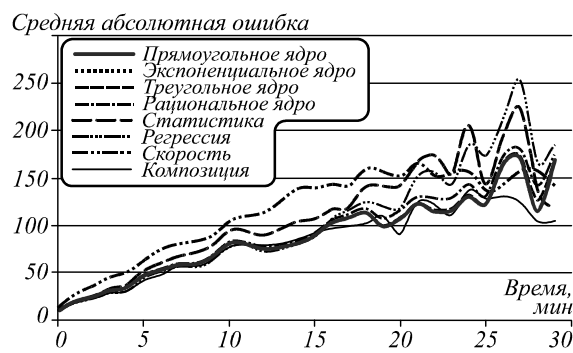


Рис. 8. Зависимость средней абсолютной ошибки от горизонта прогноза на контрольной выборке

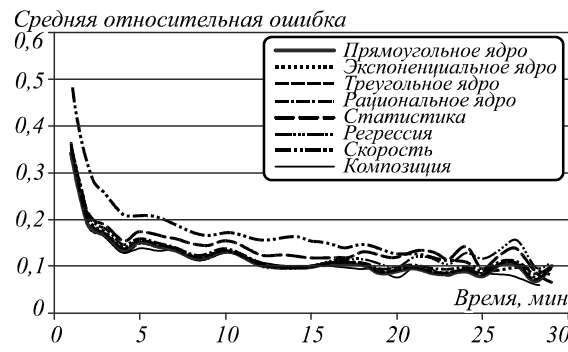


Рис. 9. Зависимость средней относительной ошибки от горизонта прогноза на контрольной выборке

На рис. 13, 14 и 15 показаны графики зависимости СКО, средней абсолютной ошибки и средней относительной ошибки от горизонта прогноза на контрольной выборке для ТС автобусных маршрутов.

На контрольной выборке разработанная модель также показала лучшие результаты среди рассмотренных алгоритмов прогнозирования.

Выводы

В работе предложен новый оригинальный алгоритм построения прогноза времени прибытия общественных транспортных средств на остановки общественного транспорта, основанный на модели адаптивной композиции элементарных алгоритмов прогнозирования, каждый из которых характеризуется малым числом настраиваемых параметров.

Адаптивность подразумевает зависимость параметров конструируемой композиции от ряда управляющих параметров модели, к которым относятся следующие актуальные (определённые на текущий

момент) факторы: погодные условия, плотность транспортного потока, динамика движения, горизонт прогноза и др. Адаптивность достигается вве-

дением иерархического разбиения области значений управляющих параметров, применяемого в дереве регрессии.

Таблица 5. Сравнение алгоритмов ОТС для автобусных маршрутов

	СКО		Средняя абсолютная ошибка		Средняя относительная ошибка	
	Обучающая	Контрольная	Обучающая	Контрольная	Обучающая	Контрольная
Прямоугольное ядро	27,1347	25,1064	12,7607	12,4590	0,4626	0,4636
Треугольное ядро	26,5599	24,9221	12,7034	12,5126	0,4646	0,4670
Экспоненциальное ядро	26,4379	24,9060	12,6992	12,5315	0,4652	0,4677
Рациональное ядро	26,7537	25,5068	13,0073	12,9006	0,4766	0,4797
Статистика	28,1978	26,1688	13,3373	13,3127	0,4906	0,5054
Скорость	32,7356	31,8209	17,1002	17,0614	0,5848	0,6004
Регрессия	22,1130	24,2553	11,1114	12,0814	0,4243	0,4604
Композиция	19,0246	23,2880	10,2108	11,9968	0,3963	0,4620

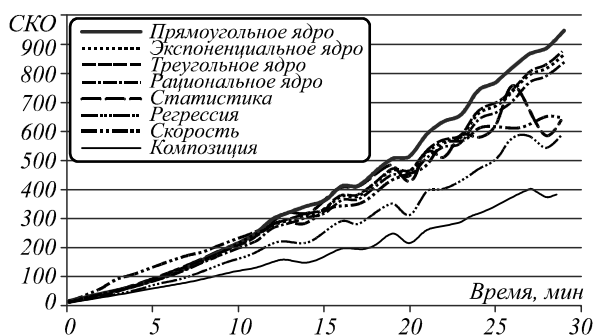


Рис. 10. Зависимость СКО от горизонта прогноза на обучающей выборке

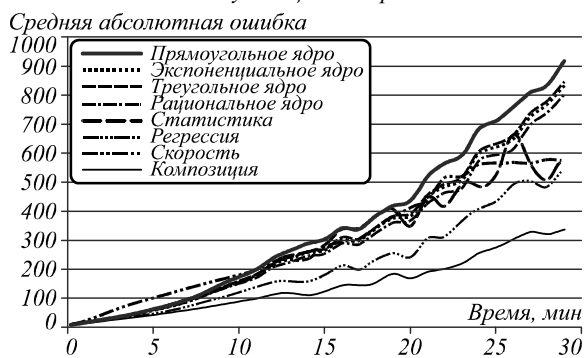


Рис. 11. Зависимость средней абсолютной ошибки от горизонта прогноза на обучающей выборке

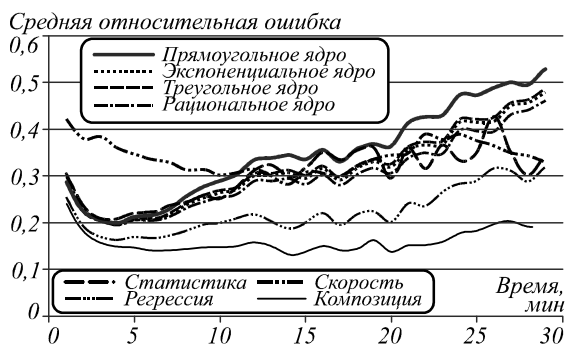


Рис. 12. Зависимость средней относительной ошибки от горизонта прогноза на обучающей выборке



Рис. 13. Зависимость СКО от горизонта прогноза на контрольной выборке

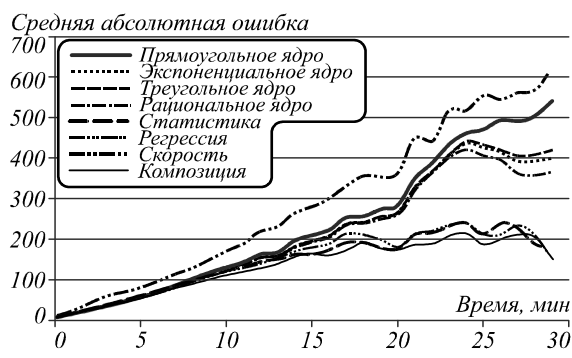


Рис. 14. Зависимость средней абсолютной ошибки от горизонта прогноза на контрольной выборке

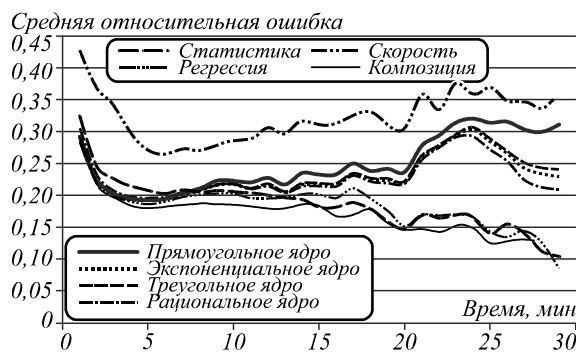


Рис. 15. Зависимость средней относительной ошибки от горизонта прогноза на контрольной выборке

В исследованиях, проведённых на данных движения городского пассажирского транспорта в г. Самаре, предложенный алгоритм прогнозирования показал лучший результат по сравнению с элементарными прогнозами и разработанной ранее моделью линейной регрессии.

Учитывая, что предложенный алгоритм адаптивной композиции обладает свойствами, которые совместно не присущи ни одному из представленных в литературе, а именно:

- позволяет включать в композицию алгоритмы прогнозирования достаточно произвольного типа,

- обладает адаптивностью по отношению к изменению дорожной ситуации (т.е. учитывает факторы, оказывающие непосредственное влияние на движение – плотность транспортного потока, динамику движения и т. п.),

- обладает адаптивностью по отношению к актуальным (определённым на текущий момент) факторам, прямым и/или косвенным образом влияющим на движение и/или результат прогноза: погодным условиям и освещённости, требуемому горизонту прогноза и другим,

– предлагаемый алгоритм и использованная в нём модель адаптивной композиции прогнозов представляются наиболее современными и наилучшим образом подходящими для решения рассмотренной задачи.

Дальнейшие направления работ включают в себя исследования, связанные с выбором наилучшего множества элементарных алгоритмов для конструируемой композиции, а также исследования, связанные с анализом и прогнозированием состояния (параметров) транспортных потоков.

Благодарности

Работа выполнена при частичной финансовой поддержке:

- грантов РФФИ, проекты № 13-07-12103-офи-м, 13-01-12080-офи-м, 12-07-00021-а;

- программы фундаментальных исследований Президиума РАН «Фундаментальные проблемы информатики и информационных технологий», проект 2.12;

- Министерства образования и науки Российской Федерации (в рамках постановления Правительства Российской Федерации от 09.04.2010 г. № 218: договор № 02.Г36.31.0001 от 12.02.2013).

Литература (References)

1. **Hall, R.** Handbook of transportation science / Randolph W. Hall. – Dordrecht: Kluwer Academic Publishers, 2003. – 737 p.
2. **Altinkaya, M.** Urban Bus Arrival Time Prediction: A Review of Computational Models / M. Altinkaya, M. Zontul // International Journal of Recent Technology and Engineering (IJRTE). – 2013. – V. 2, Issue 4. – P. 164-169.
3. **Hoogendoorn, S.P.** State-of-the-art of vehicular traffic flow modeling / S.P. Hoogendoorn, P.H.L. Bovy // Proceedings of the Institution of Mechanical Engineers. Part I: Journal of Systems and Control Engineering. – 2001. – V. 215(4). – P. 283-303.
4. **Padmanaban, P.** Estimation of Bus Travel Time Incorporating Dwell Time for APTS Applications / R.P.S. Padmanaban, L. Vanajakshi, S.C. Subramanian // IEEE Intelligent Vehicles Symposium. – 2009. – V. 2. – P. 955-959.
5. **Агафонов, А.А.** Прогнозирование параметров движения городского пассажирского транспорта по данным спутникового мониторинга / А.А. Агафонов, А.В. Сергеев, А.В. Чернов // Компьютерная оптика. – 2012. – Т. 36, № 3. – С. 453-489. (**Agafonov, A.A.** Forecasting of the motion parameters of city transport by satellite monitoring data / A.A. Agafonov, A.V. Sergeyev, A.V. Chernov // Computer Optics. – 2012. – V. 36 (3). – P. 453-458.)
6. **Agafonov, A.** City transport motion parameters forecasting by satellite monitoring data and statistics / A. Agafonov, A. Chernov, A. Sergeyev // PRIA-2013. – 2013. – V. 2. – P. 489-491.
7. **Sun, H.** Use of Local Linear Regression Model for Short-term Traffic Forecasting / H. Sun, H.X. Liu, H. Xiao, R.R. He, B. Ran // Transportation Research Record. – 2003. – Issue 1836. – P. 143-150.
8. **Vanajakshi, L.** Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses / L. Vanajakshi, S.C. Subramanian, R. Sivanandan // IET Intelligent Transport Systems. – 2009. – V. 3. – P. 1-9.
9. **Shalaby, A.** Prediction Model of Bus Arrival and Departure Times Using AVL and APC Data / A. Shalaby, A. Farhan // Journal of Public Transportation. – 2004. – V. 7(1). – P. 41-63.
10. **Chen, M.** A dynamic bus-arrival time prediction model based on APC data / M. Chen, X. Liu, J. Xia, S.I. Chien // Computer-Aided Civil and Infrastructure Engineering. – 2004. – V. 19(5). – P. 364-376.
11. **Chang, G.-L.** Predicting intersection queue with neural network models / G.-L. Chang, C.-C. Su // Transportation Research Part C. – 1995. – V. 3(3). – P. 175-191.
12. **Jeong, R.** Bus arrival time prediction using artificial neural network model / R. Jeong, L.R. Rilett // IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. – 2004. – P. 988-983.
13. **Bin, Y.** Bus arrival time prediction using support vector machines / Y. Bin, Y. Zhongzhen, Y. Baozhen // Journal of Intelligent Transportation Systems: Technology, Planning, and Operations. – 2007. – V. 10, Issue 4. – P. 151-158.
14. **Wu, C.-H.** Travel-time prediction with support vector regression / C.-H. Wu, J.-M. Ho, D.T. Lee // IEEE Transactions on Intelligent Transportation Systems. – 2004. – V. 5(4). – P. 276-281.
15. **van Lint, J.W.C.** Accurate freeway travel time prediction with state-space neural networks under missing data / J.W.C. van Lint, S.P. Hoogendoorn, H.J. van Zuylen // Transportation Research Part C: Emerging Technologies. – 2005. – V. 13(5-6). – P. 347-369.
16. **Park, T.** A bayesian approach for estimating link travel time on urban arterial road network / T. Park, S. Lee // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). – 2004. – V. 3043. – P. 1017-1025.
17. **Zheng, W.** Short-term freeway traffic flow prediction: Bayesian combined neural network approach / W. Zheng, D.-H. Lee, Q. Shi // Journal of Transportation Engineering. – 2006. – V. 132(2). – P. 114-121.
18. **Yang, J.-S.** Travel time prediction using the GPS test vehicle and Kalman filtering techniques / J.-S. Yang // Proceedings of the American Control Conference. – 2005. – V. 3. – P. 2128-2133.
19. **Wall, Z.** An Algorithm for Predicting the Arrival Time of Mass Transit Vehicles Using Automatic Vehicle Location

Data / Z. Wall, D. J. Dailey // 78th Annual Meeting of the Transportation Research Board, Washington D.C., 1999.
20. **Zaki, M.** Online Bus Arrival Time Prediction Using Hybrid Neural Network and Kalman filter Techniques /

M. Zaki, I. Ashour, M. Zorkany, B. Hesham // International Journal of Modern Engineering Research. – 2013. – V. 3, Issue 4. – P. 2035-2041.

AN ALGORITHM FOR CITY TRANSPORT ARRIVAL TIME ESTIMATION USING ADAPTIVE ELEMENTARY PREDICTIONS COMPOSITION

A.A. Agafonov, V.V. Myasnikov
Image Processing Systems Institute, Russian Academy of Sciences,
Samara State Aerospace University

Abstract

The problem of precise arrival time of public transport is considered in this paper. There is proposed a new prediction algorithm based on adaptive composition model using elementary prediction. A small number of adaptive parameters characterizes each elementary prediction algorithm. Adaptability means that parameters of the constructed compositions depend on a number of control parameters of the model, which includes the following factors: weather conditions, traffic density, driving dynamics, prediction horizon, etc. Adaptability is achieved by introducing a hierarchical decomposition range of control parameters used in regression tree. We made experimental investigations on real routes of city public transport in Samara to evaluate the prediction accuracy of the proposed algorithm. We also explain the advantages of the proposed solution in comparison with existing ones.

Key words: city public transport, arrival time prediction, arrival time estimation, algorithms composition, hierarchical decomposition, regression tree.

Сведения об авторах



Агафонов Антон Александрович, 1988 года рождения. В 2011 году окончил Самарский государственный аэрокосмический университет (СГАУ). В настоящее время работает стажёром-исследователем в Федеральном государственном бюджетном учреждении науки Институте систем обработки изображений РАН и по совместительству инженером-математиком в ОАО «Самара-Информспутник». Круг научных интересов включает геоинформационные технологии, веб-технологии.

E-mail: ant.agafonov@gmail.com.

Anton Aleksandrovich Agafonov (1988 b.), graduated from Samara State Aerospace University (SSAU) at 2011. At present he is intern-researcher at the Image Processing Systems Institute of the Russian Academy of Sciences, holding a part-time positions of engineer-mathematician at JSC "Samara-Informsputnik". The area of interests includes geoinformatics and web-technologies.

Сведения об авторе **Мясников Владислав Валерьевич** смотри стр. 296 этого номера.

Поступила в редакцию 18 марта 2014 г.