

АНАЛИЗ ПАТТЕРНОВ МОБИЛЬНОСТИ ОБЩЕСТВЕННОГО
ТРАНСПОРТА И ПЕРЕНОС АВТОБУСНЫХ ОСТАНОВОК© 2019 г. Э. Фаббиани^a, С. Несмачнов^a, Д. Тутух^b,
А. Черных^{c,d,e,*}, А. Аветисян^{d,f}, Г. Радченко^e^a Государственный университет в Монтевидео (Монтевидео, Уругвай)^b Малагский университет (Малага, Испания)^c Научно-исследовательский центр CICESE (Энсенада, Баха Калифорния, Мексика)^d Институт системного программирования РАН (Москва, Россия)^e Южно-Уральский государственный университет (Челябинск, Россия)^f Московский государственный университет им. М.В. Ломоносова (Москва, Россия)

* E-mail: chernykh@cisece.mx

Поступила в редакцию 11.05.2018 г.

После доработки 11.05.2018 г.

Принята к публикации 12.06.2018 г.

Знание моделей мобильности граждан, использующих общественный транспорт, является актуальной проблемой для современных умных городов. Информация о мобильности имеет решающее значение для проектирования и планирования системы городского транспорта, способной обеспечить хорошее обслуживание граждан. В этой статье рассматриваются две актуальные проблемы, связанные с системами общественного транспорта: анализ моделей мобильности пассажиров и перенос автобусных остановок в городской среде. Для первой проблемы применяется подход обработки больших объемов информации, получения ключевых показателей и создания матриц спроса и отправления-назначения, путем оценки мест назначения пассажиров, анализа информации о проданных билетах и местонахождения автобусов. Несколько соответствующих показателей вычисляются и анализируются, чтобы охарактеризовать модели мобильности, используя данные из системы общественного транспорта в Монтевидео, Уругвай. Предлагается распределенная реализация, достигающая значительных улучшений времени выполнения (ускорение до 17.10 при использовании 24 вычислительных ресурсов). Для второй проблемы предлагается использовать многокритериальный эволюционный алгоритм для перемещения автобусных остановок с целью повышения качества обслуживания за счет оптимизации времени движения и минимизации эксплуатационных расходов системы. Алгоритм оценивается на реальных данных полученных в 2015 году. Экспериментальные результаты показывают, что алгоритм может добиться улучшения до 16.7% и 33.9% по времени и стоимости, соответственно, по сравнению с реальной ситуацией в 2015 году.

DOI: 10.1134/S0132347419010023

1. ВВЕДЕНИЕ

Городские транспортные системы играют фундаментальную роль, обеспечивая решения для населения в современных городах [8]. В настоящее время, основные проблемы, связанные с городской мобильностью в крупных городах, связаны с неспособностью систем общественного транспорта удовлетворять потребности растущего числа пользователей.

Парадигма “умных городов” предполагает использование информационных и коммуникационных технологий с целью повышения качества и эффективности городских служб [3]. Опираясь на информационные технологии, в процессе проек-

тирования и планирования систем городского транспорта, важно обеспечить качественный сервис для граждан.

Интеллектуальная транспортная система (ИТС) является фундаментальным компонентом умного города. ИТС позволяет собирать большой объем данных о транспорте и мобильности в городах [7]. В крупных городах ИТС генерирует огромные объемы данных, которые могут быть обработаны для получения ценной информации о мобильности граждан. Эта информация может быть предоставлена пользователям транспортной системы, проектировщикам и лицам, принимающим решения.

В этом контексте город Монтевидео, Уругвай, представил в 2010 году план городской мобильно-

сти для реорганизации и модернизации общественного транспорта [10]. В рамках этого плана была внедрена столичная транспортная система (Metropolitan transportation system – STM) с целью централизации различных компонентов системы общественного транспорта. Одним из первых шагов было оснащение каждого автомобиля глобальной системой позиционирования (GPS) и введение смарт-карт для оплаты билетов (STM-карта). Эти устройства позволяют собирать огромный объем данных о местонахождении транспорта, продажах билетов, переходах между разными линиями и т.д.

В этой статье мы представляем методологию обработки данных из ИТС для извлечения информации о мобильности пассажиров. Результаты базируются на данных, собранных в течение 2015 года (более 13 миллионов поездок). Описан метод расчета соответствующих показателей системы общественного транспорта и оценки матрицы отправления-назначения (ОН).

Используя вышеупомянутую информацию, предлагается использовать многокритериальный эволюционный алгоритм (МКЭА) для переопределения расположения автобусных остановок с целью улучшения качества обслуживания пассажиров, путем оптимизации времени поездки и минимизации эксплуатационных расходов для транспортных компаний. Предлагаемый МКЭА тестируется при различных сценариях, с использованием реальных данных. Итоги тестирования показывают, что предлагаемый эволюционный подход эффективно выполняет поиск высококачественных решений, позволяя добиться уменьшения времени поездки до 16.7% и уменьшения стоимости обслуживания до 33.9% по сравнению с реальной ситуацией, соответствующей 2015 году.

В разделе 5 представлены основные результаты анализа мобильности. Экспериментальная оценка предлагаемого МКЭА представлена в разделе 6. Наконец, в разделе 7 представлены выводы и основные направления будущей работы.

2. АНАЛИЗ МОБИЛЬНОСТИ И ПРОБЛЕМЫ ПЕРЕМЕЩЕНИЯ АВТОБУСНЫХ ОСТАНОВОК

В этом разделе описываются две проблемы: анализ моделей мобильности для общественного транспорта и перемещение автобусных остановок.

2.1. Анализ моделей мобильности

Анализ мобильности имеет решающее значение для повышения качества обслуживания транспортных систем. Ряд релевантных показателей можно вычислить из статической информации и обследований. Однако эти методы предполагают дорогостоящие процессы (как по време-

ни, так и по необходимым ресурсам), и они предоставляют только частичную информацию о мобильности граждан. Более эффективным вариантом является применение алгоритмов обработки для извлечения соответствующей информации из необработанных данных, как правило, из современных репозиториях ИТС. Учитывая большой объем доступных данных, для ускорения их обработки применяются техники работы с большими данными и технологии распределенных вычислений. Это позволяет вычислять полезные индикаторы и показатели в режиме близком к реальному времени.

Нами применяется методология, объединяющая методы анализа больших данных и распределенные вычисления для сбора и обработки соответствующих данных из STM. Полезная информация включает в себя: профили спроса, ОН-матрицы, профили использования автобусов, модели мобильности, продажи билетов, характеристики системы общественного транспорта и т.д.

Автобусные компании отправляют информацию городским властям, в том числе геолокацию автобусов и данные о продажах билетов. Сеть автобусов сложна: она имеет 144 маршрута (и несколько вариантов для многих маршрутов) и 4835 остановок [9].

Данные геолокации содержат положение каждого автобуса с частотой от 10 до 30 секунд. Каждая запись GPS содержит следующую информацию:

ID маршрута, код, который идентифицирует маршрут; *ID поездки*, код, который однозначно идентифицирует поездку для маршрута; *широта*; *долгота*; *скорость* автобуса (в км/ч) *временная метка*; и *ID остановки*, если GPS находится рядом с остановкой.

Данные продаж соответствуют поездкам с смарт-картами и без них. Для билетов, оплаченных смарт-картой, хранится уникальный идентификатор, чтобы идентифицировать поездки, оплачиваемые с помощью одной и той же карты. По соображениям конфиденциальности, идентификатор карты маскируется *хеш*-функцией для скрытия конфиденциальной информации пассажиров, использующих эту услугу.

Существует два разных способа купить билет: одночасовые билеты, которые позволяют перейти на другую линию в течение одного часа с момента покупки билета; и двухчасовые билеты, позволяющие неограниченные переходы в течение двух часов. Не обязательно использовать карту STM для оплаты билета, но пассажиры, платящие наличными, теряют право на переходы.

Каждая запись продажи, сделанная с или без смарт-карты, содержит следующую информацию:

ID поездки, как и в GPS записи; *широта*; *долгота*; *ID остановки*, как и в GPS записи; *количество пассажиров*, так как карта позволяет покупать би-

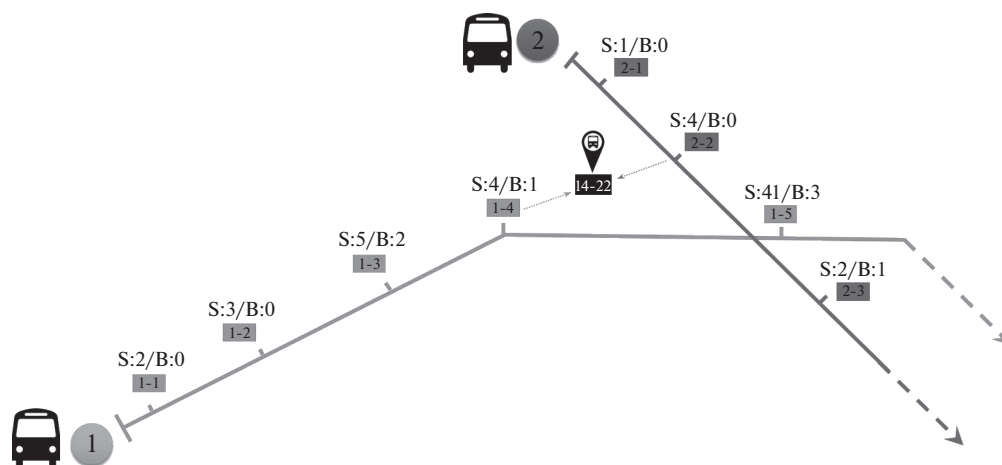


Рис. 1. Пример проблемы перемещения автобусных остановок.

леты для нескольких пассажиров; и *временную метку*.

Если продажа осуществляется с помощью смарт-карты, также сохраняется *идентификатор смарт-карты*, *количество транзакций* выполненных с одной и той же STM-карты, и *последняя оплаченная транзакция*. Эти данные позволяют идентифицировать поездку с пересадкой: в этом случае количество транзакций увеличивается, но последняя транзакция не оплачивается.

2.2. Проблема перемещения автобусных остановок

Введение. Создание грамотной конфигурации автобусных остановок является дорогостоящим процессом из-за сложности идентификации тех остановок, которые не являются полезными в отношении качества обслуживания, предоставляемого пользователям. Проблема перемещения автобусных остановок заключается в нахождении конфигурации автобусных остановок, которая обеспечивает хороший компромисс между качеством обслуживания (критерий оптимизации, ориентированный на пользователя) и эксплуатационными расходами (оптимизация, ориентированная на поставщиков услуг).

Несколько факторов ограничивают возможные конфигурации для перемещения автобусной остановки: *i)* максимальное количество пассажиров, которые могут одновременно находиться в автобусе, *ii)* расстояние, на которое пассажир допускает перемещение пешком, *iii)* расходы на зарплату, топливо и др. Улучшение качества обслуживания заключается в минимизации времени в пути, с учетом времени прибытия, времени ожидания на остановке и продолжительности поездки. Стоимость компаний рассчитывается как разница между прибылью за количество проданных билетов и расходами, необходимыми для ор-

ганизации поездок. В рамках этого проекта предполагается, что расходами являются только затраты на заработную плату водителей и топливо для транспортных средств.

На рисунке 1 представлен пример проблемы для двух маршрутов с одним и тем же пунктом назначения и максимальной загрузкой в 10 пассажиров на автобус. Автобусные остановки идентифицируются кодом в цветном прямоугольнике, соответствующем идентификатору маршрута (первый номер, перед тире) и порядковым номером остановки (второе число после тире). Числовые значения, показанные на каждой остановке, соответствуют количеству пассажиров, желающих сесть (S) и сойти (B) с автобуса. Например, на остановке 1—4 (зеленая линия) четыре человека хотят сесть, но свободное пространство в этот момент составляет всего лишь три места. Однако на остановке 2—2 (синяя линия) четыре человека хотят сесть, при том что в автобусе в этот момент девять свободных мест. Поэтому возможным вариантом является устранение обоих остановок и создание новой в середине между ними, так что оба маршрута могут использовать новую остановку, а человек с маршрута 1, который не смог сесть на автобус 1, сможет сесть на автобус 2.

Все используемые данные параметризуются, тем самым обеспечивая большую гибкость алгоритмов, чтобы их можно было легко применить в разных сценариях и ситуациях.

Математическая постановка. В математической постановке задачи перемещения остановки автобуса учитываются следующие элементы:

- (непересекающееся) множество *зон* $Z = \{z_1, \dots, z_n\}$; каждая зона представляет собой географическую зону, например, район в городе;
- набор возможных локаций для автобусных остановок, $S = \{s_1, \dots, s_p\}$. Каждая локация опреде-

ляется ее географическими координатами в городе, широтой и долготой ($lat_s, long_s$). Активные локации представляют собой остановки для одного или нескольких маршрутов. Кроме того, в некоторых локациях потенциально могут быть установлены новые автобусные остановки. Например, новая автобусная остановка может быть определена в новой локации после устранения соседних остановок в окрестности;

- набор *маршрутов* общественного транспорта $R = \{r_1, \dots, r_m\}$. Каждый маршрут проходит через множество зон, заданных функцией $RZ : R \rightarrow Z$. Каждый маршрут имеет фиксированную частоту в минутах, заданную функцией $RF : R \rightarrow \mathbb{N}$ (например, автобусы, которые работают на маршруте r_j прибывают на каждую остановку регулярно, каждые $RF(r_j)$ минут). Каждый маршрут имеет упорядоченный список остановок, заданный функцией $RS : R \rightarrow S$ (порядок имеет значение);

- функция *спроса пассажиров* $D : S \times Z \rightarrow N$. $D(s_i, z_j)$, указывающая количество пассажиров, которые осуществляют посадку в автобус на остановке s_i и отправляются в пункт назначения z_j . Функция спроса пассажиров определяется от остановок до зон, потому что пассажиры подтверждают поездку (по смарт-карте или за наличный платеж), когда они садятся на автобус, а не когда они выходят. По этой причине необходимо оценить остановку в пределах зоны назначения z_j , где каждый пассажир выходит из автобуса. Стратегия оценки остановки назначения является частью процесса формирования матрицы ОН, описанной в разделе 2.1;

- набор транспортных средств (*автобусов*) $B = \{b_1, \dots, b_q\}$. Каждый автобус имеет фиксированную вместительность K , которая определяет максимальное количество пассажиров, которым разрешено перемещаться в автобусе, в соответствии с городскими правилами. Каждое транспортное средство проходит через набор остановок, связанных с маршрутом в определенном порядке, и для завершения каждой поездки требуется определенное время. *Время в пути* между последовательными остановками связано с географическим расстоянием между ними (в соответствии с маршрутом, определенным для каждого маршрута), дорожными условиями (трафик, максимальная скорость и т.д.) и количеством пассажиров, которые осуществляют посадку и высадку из автобуса на каждой остановке. Время в пути задается функцией $TT : S \times S \rightarrow N$, где $TT(s_i, s_j)$ — это время, необходимое для проезда автобуса от остановки s_i до следующей остановки s_j . Каждый раз, когда автобус b_h останавливается на остановке s_i , предлагаемая модель различает два случая: *i*) по меньшей мере один пассажир осуществляет посадку на автобус b_h на остановке s_i ; тогда время в

пути будет увеличено на $t_s + \sum_{k=1}^{k=n} D(s_i, z_k \times t_b)$, где t_s — это время, необходимое для остановки автобуса и t_b — время, необходимое одному пассажиру для посадки на автобус; *ii*) автобус b_h останавливается на остановке s_i только для того, чтобы позволить пассажирам выйти, тогда время в пути будет увеличено на $t_s + t_a$, где t_a — это (постоянное) время необходимое для высадки пассажиров. Модель предполагает, что высадка пассажиров значительно быстрее, чем посадка (т.к. не требуется покупка билетов или проверка карты). Для определения количества пассажиров, которые выходят из автобуса на остановке s_i , модель использует оценочную матрицу ОН;

- функция *подсчета перемещающихся пассажиров* $C : B \times S \rightarrow \mathbb{N}$, где $C(b_i, s_i)$ обозначает число пассажиров, которые находятся на борту автобуса b_i на остановке s_i . Количество пассажиров на каждом автобусе обновляется каждый раз, когда пассажиры осуществляют посадку или высадку с автобуса на каждой остановке. Число пассажиров не может превышать вместительности K ;

- функция *операционных затрат*, связанная с автобусными компаниями, вычисляет стоимость перемещения между автобусными остановками по конкретному маршруту. Она определяется как $BC : S \times S \rightarrow \mathbb{N}$, где $CC(s_i, s_j) = (C_g + C_s) \times dist(s_i, s_j)$ — это стоимость проезда от s_i до s_j , C_g — это стоимость топлива, а C_s — это зарплата водителя за каждый 1 км.

Также рассматриваются две переменные: x_{ij} — это i -я остановка автобуса, определенная на маршруте r_j и u_{ij}^k (бинарная) обозначает что маршрут k соединяет зоны z_i и z_j .

Когда остановка s_i удаляется из маршрута r_j , пассажиры, перемещающиеся из s_i в зону z_k распределяются между соседними остановками, которые хотят сесть в автобус, который едет до того же пункта назначения. Например, когда остановка автобуса s_i удаляется из маршрута r_j , пассажиры должны пройти к другой остановке автобуса s_h , который проверяет $s_h \in RZ(r_j)$ и $s_i \in RZ(r_j)$. Когда доступно более одного варианта, количество пассажиров распределяется обратно пропорционально расстоянию, которое пассажиры должны пройти, чтобы перейти к ближайшей новой остановке. В этом случае пассажирам необходимо перейти на новую остановку, а необходимое время ходьбы определяется функцией $WT : S \times Z \rightarrow \mathbb{N}$, где $WT(s_i, z_k)$ определяется $WT(s_i, z_k) = \overline{ws} \times dist(s_i, s_h)$, где $\overline{ws}$ — это средняя скорость ходьбы для пешеходов и s_h — это ближайшая остановка маршрута r_j , та-

кая что $z_k \in RZ(r_j)$. Аналогичный подход используется, когда новая остановка добавляется между двумя существующими остановками. В этом случае пассажирам необходимо будет перейти на новую локацию.

Это проблема *многокритериальной оптимизации*, поскольку она включает одновременную оптимизацию двух функций, находящихся в конфликте друг с другом. Основные задачи — это минимизация времени движения автобуса, что напрямую влияет на качество обслуживания, предлагаемого пассажирам, которое определяется функцией QT (Equation 2.1), одновременно максимизируя прибыль компаний за счет уменьшения стоимости, заданной функцией СТ (Equation 2.2). Прибыль определяется как разница между количеством проданных билетов и затратами необходимыми для предоставления услуг. Функция QT учитывает точку зрения пассажиров и администрации города, в то время как функция СТ отвечает интересам транспортных компаний, эксплуатирующих автобусные линии.

$$QT = \sum_{j=1}^{j=m} \sum_{i=1}^{i=\#RS(r_j)} (t_s + D(s_i) \times t_b) + TT(x_{ij}, x_{i-1j}) + \sum_{j=1}^{j=m} \sum_{i=1}^{i=\#RS(r_j)} + \sum_{k=1}^{k=\#RZ(r_j)} WT(s_i, z_k) \quad (2.1)$$

$$CT = \sum_{j=1}^{j=m} \sum_{i=1}^{i=\#RS(r_j)} (G \times D(x_{ij}) - CC(x_{ij}, x_{i-1j})) \quad (2.2)$$

Сложность проблемы. Предполагая, что в каждом квартале может быть только одна автобусная остановка, количество возможных мест для расположения остановок равно количеству кварталов в городе. Кроме того, каждый маршрут имеет свои собственные остановки, которые должны одновременно использоваться и другими маршрутами. Это является причиной сложности проблемы определения мест для автобусных остановок, которые были бы эффективны с точки зрения системы общественного транспорта. Предполагая что мы имеем набор из K маршрутов и P автобусных остановок, количество возможных комбинаций для их расположения равно K^P . Таким образом, проблема распределения имеет вычислительную сложность, которая экспоненциальна по числу остановок ($O(K^P)$).

При решении больших задач на реальных данных, традиционные точные алгоритмы не применимы для решения проблемы перемещения автобусных остановок за приемлемое время. Использование метаэвристик позволяет получать решения приемлемого качества в разумные сроки выполнения [21]. Метаэвристики — это стратегии высокого уровня, которые объединяют наборы более

простых методов обработки (обычно эвристических), чтобы получить более мощную процедуру. В этой статье предлагается применить МКЭА для решения проблемы. Этот подход позволяет получать быстрые кандидатные решения благодаря эффективному изучению пространства решений.

3. ОБЗОР РАБОТ ПО ТЕМАТИКЕ ПРОЕКТА

В этом разделе рассматриваются ключевые работы, связанные с анализом мобильности и эволюционной оптимизации систем общественного транспорта.

3.1. Анализ мобильности и генерация матрицы ОН

В работе Pelletier et al. [23] рассматривается использование смарт-карт для анализа мобильности в системах общественного транспорта. В ходе исследования были проанализированы основные требования к внедрению решений для оплаты посредством смарт-карт, а также вопросы конфиденциальности и юридические вопросы, возникающие при использовании данных из таких систем. Кроме того, авторы определили основные виды использования этих данных, например, долгосрочное планирование, корректировки предоставляемых услуг и показатели эффективности транспортных систем. Наконец, были представлены примеры реального использования таких данных.

В работе Trépanier et al. [26] представлена модель для оценки мест назначения пассажиров общественного транспорта с использованием смарт-карт. Были рассмотрены две гипотезы, которые используются в других связанных работах: *i*) происхождение новой поездки — это пункт назначения предыдущей поездки, и *ii*) в конце дня, пассажиры возвращаются к пункту отправления первой поездки дня. Исходя из этих предположений, метод позволяет исследовать последовательность поездок каждого пользователя в системе. Поездки, для которых сборка последовательности невозможна (например, одна поездка за день для данного пользователя), сравниваются со всеми другими поездками месяца для одного и того же пользователя для того, чтобы найти подобные поездки с известным пунктом назначения. Экспериментальная оценка была проведена с реальной информацией от транзитных властей в городе Гатино, Квебек.

Результаты показали, что существует возможность оценить пункт назначения вплоть до 66% поездок. В статье также показано, что большинство поездок, где пункт назначения невозможно установить, совершаются в часы низкого спроса. Рассматривая только часы с высоким спросом, процент поездок, по которым можно оценить их назначения, составляет 80%. Однако точность не

была оценена из-за отсутствия второго источника реальных данных для сравнения.

Группа Wang et al. в статье [27] предложила методологию последовательности путешествий для определения пунктов отправления и назначения пассажиров автобусов с использованием смарт-карт и данных автоматизированной системы сбора местоположения транспортных средств (Automatic Vehicle Location — AVL) в Лондоне, Великобритания. Пункты отправления были точно рассчитаны с использованием *временных меток*, хранившихся в записях смарт-карт AVL, для определения автобусной остановки. Для оценки пунктов назначения, авторы использовали подход, аналогичный тому, который был представлен в [26], следуя, по возможности, последовательностью поездки. Результаты были сопоставлены с данными лондонского пассажирского опроса, проводимого каждые пять-семь лет для каждого автобусного маршрута и включающего информацию от людей, входящих и выходящих на каждой остановке автобуса. Анализ показал, что пункт отправления может быть оценен более чем в 90% от поездок, тогда как пункт отправления и назначения может быть оценен в 57% от поездок. При сравнении с результатами опроса, разница в оценке пунктов назначения составила ниже 4%. Наконец, были представлены два практических приложения: i) исследование суточного изменения потока для определения местоположений вдоль маршрута автобуса, на которые наблюдается высокий спрос среди пассажиров, а также недоиспользуемые местоположения и ii) анализ времени пересадок, оценка времени, которое пассажиры должны потратить на ожидание пересадки, в зависимости от конечной остановки и данных AVL.

В работе [18] представили подход, аналогичный подходу группы [27] для оценки ОН-матриц в мультимодальной транспортной системе Сантьяго, Чили. В этом случае пассажиры могли использовать свои смарт-карты для оплаты билетов в метро, автобусах и на автобусных остановках.

В различных исследованиях применялись распределенные вычисления для обработки больших объемов данных трафика, хотя очень немногие работы касались оценки спроса или ОН-матриц с использованием распределенных вычислений.

Ранние работы применяли распределенные вычисления для сбора данных о трафике. В работе Sun [24] была предложена клиент-серверная модель, реализованная на основе CORBA, для сбора информации о трафике в реальном времени и использования ее для динамического построения матриц спроса.

Предлагаемое решение включает клиента для извлечения данных и сервер для хранения данных. Информация обрабатывается и используется для построения стратегий распределения динами-

ческого трафика для транспортной сети посредством применения двухуровневой схемы оптимизации, которая включает в себя информацию о транспортных потоках.

В статье [25] комбинируются различные источники информации (т.е. записи звонков сотовых телефонов, переписи и опросы) для расчета полезных показателей общественного транспорта и вывода матриц ОН. Несколько существующих алгоритмов были объединены для создания ОН-матриц, назначения поездок для определенных маршрутов и для расчета показателей качества использования маршрутов. Кроме того, было реализовано веб-приложение, предоставляющее дружелюбный интерфейс для просмотра данной информации. Это исследование страдает отсутствием информации о применяемых методах параллельных вычислений.

В работе Mellegård [17] предложена реализация, использующая Hadoop для создания массивов ОН при сохранении конфиденциальности пользователей посредством синтетического мобильного сбора данных (симуляции). Однако показатели производительности не сообщаются, поэтому преимущества применения Hadoop не ясны.

В работе Massobrio et al. [16] представлен анализ использования распределенных вычислительных технологий для анализа данных GPS с автобусов. Представлен подход с использованием MapReduce для обработки исторических данных для изучения показателей для оценки качества обслуживания транспортных систем в Монтевидео, Уругвай. Авторы использовали стратегию расчета средней скорости автобусов и выявления проблемных зон движения в соответствии с задержками и отклонениями при прибытии на каждую остановку. Параллельное развертывание масштабируется должным образом при обработке больших объемов данных.

3.2. Эволюционные подходы к оптимизации общественного транспорта

Командой Johar et al. [13] представлен исчерпывающий обзор, касающийся проектирования и планирования сетей общественного транспорта. Авторы классифицировали проблемы транспортной сети на четыре больших блока: маршрутизация, планирование, комбинированная маршрутизация и планирование, а также интеграция планирования общественного транспорта. Большая часть работ была посвящена оптимизации общей стоимости маршрута с учетом как пользователя, так и поставщика услуг. Авторы предложили использовать эволюционные алгоритмы (ЭА) вместо классических алгоритмов оптимизации для решения таких проблем, из-за сложности и коли-

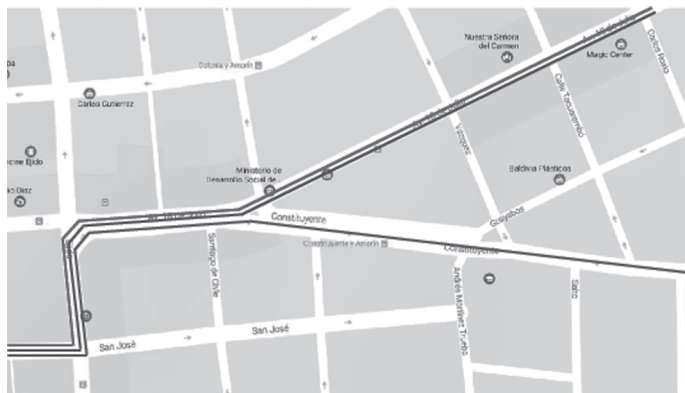
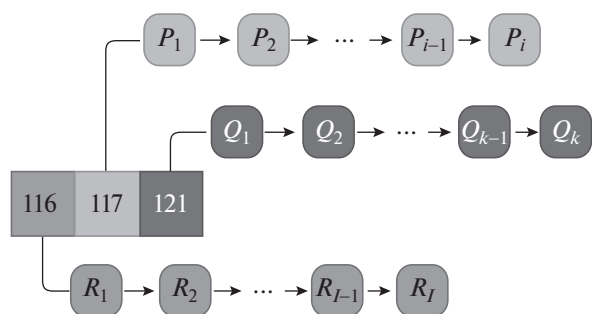


Рис. 2. Пример кодировки.

чества переменных, необходимых для принятия решения.

В работе Deb и др. [5] представлен ЭА для оптимального планирования расписаний и маршрутов автобусов или поездов. В работе также представлено его сравнение с классическими подходами (например, с методом ветвей и границ). Методология основана на минимизации общего времени, учитывая как время в пути, так и время, необходимое пассажирам для прибытия в пункт отправления и посадки в транспортное средство. Экспериментальная оценка проводилась на основе модели, основанной на некоторых предположениях. Как и Johar et al. [13], авторы предлагают использовать ЭА для планирования коллективных транспортных систем из-за количества переменных, которые необходимо учитывать при принятии решений и вычислительных ресурсов, необходимых для нахождения оптимального решения.

Król [14] представил ЭА для решения проблемы проектирования транспортной сети и проблемы настройки частоты, используя в качестве информации матрицу ОН и информацию о существующих маршрутах. Экспериментальный анализ проводился в сети общественного транспорта Бельско-Бяла в Польше, которая имеет 46 маршрутов, которые обслуживают около 90 тысяч человек в день. Работа не предоставляет подробной информации о том, как была создана матрица ОН.

В работе Bielli et al. [1] предложен ЭА для улучшения автобусной сети, путем оптимизации количества транспортных средств. Экспериментальный анализ был проведен в четырех сетях общественного транспорта: реальной сети в Парме (Италия) с 22 маршрутами и 459 остановками и тремя синтетическими сценариями, определенными авторами. Предлагаемый метод позволил достигнуть улучшение 90% показателей относительно исходной сети.

Xiong et al. [28] предложили использование ЭА для обучения нейронной сети для получения ре-

шений проблемы проектирования дискретной транспортной сети, что является NP-сложной проблемой. Предлагаемый метод, обозначенный кумулятивным генетическим алгоритмом (cumulative genetic algorithm – CGA), обнаруживает особей с большей пригодностью, которые будут использоваться в сочетании с новым поколением при воспроизведении. Нейронная сеть обучается у CGA. Сеть имеет три уровня, в общей сложности 62 нейронов и 1261 связи. Для обучения сети использовалось 2500 случайно выбранных решений, вычисляющих их затраты с помощью алгоритма Флойда–Варшалла. Экспериментальный анализ проводился по сценариям, определенным авторами, с 20 различными моделями сети общественного транспорта и матрицей фиксированного спроса. Никаких числовых результатов не сообщается.

3.3. Резюме

При обзоре литературы можно выделить большой интерес к изучению различных подходов к анализу сетей общественного транспорта. В некоторых исследованиях были использованы аналогичные предложения, представленные в этой статье для оценки ОН-матриц. Однако ни один из них не использовал распределенный подход для генерации их из данных местоположения GPS и проданных билетов. Многие подходы, которые ищут улучшения при проектировании и планировании сетей общественного транспорта, используют ЭА. Тем не менее, не было обнаружено никакой предыдущей работы о многокритериальном перемещении автобусных остановок. В существующих статьях основное внимание уделяется оптимизации общего времени в пути, в основном с помощью однокритериальных методов оптимизации. В отличие от представленных работ, мы предлагаем многокритериальный подход для предоставления набора эффективных решений, которые представляют собой различные компо-

миссы между общим временем и транспортными расходами.

4. МНОГОКРИТЕРИАЛЬНАЯ ПРОБЛЕМА ПЕРЕМЕЩЕНИЯ АВТОБУСНЫХ ОСТАНОВОК

В этом разделе описывается предлагаемый многокритериальный эволюционный алгоритм (МКЭА) для решения проблемы перемещения автобусных остановок.

4.1. Многоцелевые эволюционные алгоритмы

Эволюционные алгоритмы – это недетерминированные методы, которые эмулируют естественную эволюцию для решения задач оптимизации, поиска и смежных проблем [20]. За последние тридцать лет, ЭА были успешно применены для решения задач оптимизации, лежащих в основе многих реальных приложений высокой сложности.

ЭА – итеративный метод (каждая итерация называется *поколением*), который работает, применяя стохастические операторы на множестве *индивидуумов* (популяции P), чтобы улучшить их *приспособленность*, то есть меру, связанную с целевой функцией, которая оценивает, насколько это решение хорошее. Каждый индивидуум в популяции является кодированной версией кандидата решения для проблемы. Первоначальная популяция генерируется случайным методом или с использованием определенной эвристики для проблемы. Функция оценки присваивает значение приспособленности каждому индивидууму, указывая на его пригодность для решения про-

блемы. Поиск основан на вероятностном выборе наилучшего метода (как для родителей, так и для потомства) для предварительных решений более высокого качества. Итерационно, решения модифицируются вероятностным применением *операторов вариации*, в том числе *рекомбинацией* частей из двух особей или случайных изменений (*мутаций*) в их содержании, которые применяются для создания новых решений во время поиска. Эволюционный алгоритм возвращает лучшее решение, когда-либо найденное в итерационном процессе, с учетом функции приспособленности.

Многокритериальные эволюционные алгоритмы (МКЭА) [2, 4] – это специальные ЭА предназначенные для решения проблем с двумя или более противоречивыми целями. МКЭА оказались эффективными в решении сложных проблем оптимизации в реальном времени во многих областях исследований. МКЭА предназначены для одновременного выполнения двух целей: *i*) обеспечивают аппроксимации границы Парето, используя эволюционный поиск на основе Парето и *ii*) поддерживают разнообразие, а не сходятся к определенному участку границы Парето, используя методы, также применяемые при оптимизации мультимодальных функций (например, посредством совместного использования или переполнения).

Нами предлагается использовать Недоминируемый Генетический Алгоритм Сортировки, версии II (Non-dominated Sorting Genetic Algorithm, version II – NSGA-II) [4] для решения проблемы многокритериального перемещения автобусных остановок. NSGA-II – это современный

Таблица 1. Входные данные для предлагаемого МКЭА

Входные данные	Значение	Описание
Вместительность автобуса (K)	57	Максимальное число пассажиров, которые могут одновременно перемещаться в автобусе
Количество автобусных остановок ($\#S$)	6150	Количество автобусных остановок в Монтевидео
Новое максимальное расстояние	500	Максимальное расстояние для перемещения остановки (в метрах)
Стоимость топлива	5	Стоимость топлива для проезда 1 км (in \$U, 1 USD $\approx$ 30 \$U))
Заработная плата	100	Заработная плата водителя за каждый 1 км вождения (in \$U, 1 USD $\approx$ 30 \$U)
Время остановки	5	Время, необходимое для остановки автобуса (в секундах)
Время посадки	5	Время, необходимое пассажиру для посадки на автобус (в секундах)
Время высадки	3	Время, необходимое пассажиру для высадки из автобуса (в секундах)
Скорость ходьбы	5	Средняя скорость ходьбы пешеходов (в км/ч)
Расстояния	Haversine	Расстояние между двумя точками на сфере, учитывая их долготы и широты

МКЭА, который успешно применяется во многих областях.

Алгоритм 1 представляет псевдокод NSGA-II. Расчет приспособленности основан на доминировании Парето, построении *границ* не доминирующих решений. Основными особенностями NSGA-II являются: *i*) не доминирующая элитарная сортировка, которая снижает сложность проверки доминирования; *ii*) метод переполнения для сохранения разнообразия; и *iii*) назначение приспособленности, которое учитывает значения расстояния переполнения.

Algorithm 1. Псевдо-код NSGA-II

Require: N , размер популяции
 1: $t \leftarrow 0$ ▷ счетчик поколения
 2: $\text{offspring} \leftarrow \emptyset$
 3: **initialize**($P(0)$) ▷ инициализация популяции
 4: **while** not критерий_остановки **do**
 5: **evaluate**($P(t)$) ▷ оценка популяции
 6: $R \leftarrow P(t) \cup \text{потомки}$
 7: границы $\leftarrow$ **non-dominated sorting**(R)
 8: $P(t+1) \leftarrow \emptyset$; $i \leftarrow 1$
 9: **while** $P(t+1) + \text{границы}(i) \leq N$ **do**
 10: **расстояние переполнения** (границы(i))
 11: $P(t+1) \leftarrow P(t+1) \cup \text{границы}(i)$
 12: $i \leftarrow i + 1$
 13: **end while**
 14: **сортировка по расстоянию** (границы(i))
 15: $P(t+1) \leftarrow P(t+1) \cup \text{границы}(i) [1:(N - |P(t+1)|)]$
 16: выбранные $\leftarrow$ **выборка** ($P(t+1)$)
 17: потомки $\leftarrow$ **операторы вариации** (выбранные)
 18: $t \leftarrow t + 1$
 19: **end while**
 20: **return** вычисленная граница Парето

4.2. Кодирование решения

Проблема перемещения автобусных остановок имеет несколько особенностей, которые требуют специальных методов решений. Поскольку определение проблемы основано на наборе маршрутов, предлагаемая кодировка использует список маршрутов. Каждый маршрут имеет список остановок, упорядоченных от пункта отправления до пункта назначения. Пример этой кодировки показан на рисунке 2 для маршрутов 116, 117 и 121. Пример решения показывает соответствие между набором маршрутов и их остановками с их соответствующим представлением. Ограничение проблемы состоит в том, что начальные

и конечные остановки каждого маршрута должны оставаться неизменными.

4.3. Входные данные

В предлагаемом МКЭА используются данные из реальных сценариев. Эти данные используются в качестве входных данных для алгоритма и предоставляют информацию о расходах и доходах от поездок, требуемом времени, скоростях и других значениях, находящихся в файле параметров. Это позволяет сделать алгоритм более гибким и расширить возможности его применения к другим исходным данным. Соответствующие входные данные описаны в таблице 1.

4.4. Эволюционные операторы

Инициализация популяции. Популяция инициализируется механизмом, который учитывает реальную информацию о системе общественного транспорта в Монтевидео в 2015 году и матрицу ОН, рассчитанную на предыдущем этапе.

Чтобы обеспечить разнообразие для первоначальной популяции, выполняются случайные изменения базового решения, например, путем изменения местоположения остановок автобусов и количества пассажиров, осуществляющих посадку и посадку на автобус в случайно выбранных остановках. Пассажиры перераспределяются на остановки непосредственно рядом с изменяемой остановкой. Расположение выбранных остановок также изменяется. Они перемещаются на переменное расстояние (максимум 500 метров) от их первоначального местоположения. Исходные и конечные остановки не перемещаются, поскольку они определяют каждый маршрут и не могут быть изменены.

Отбор. Используется стандартный турнирный отбор, который включен в первоначальное предложение NSGA-II. Результат предварительного анализа, выполненный на небольшом сценарии, показал, что турнирный отбор обеспечивает необходимое давление выбора.

Рекомбинация. Для рекомбинации маршрутов двух решений использовался оператор одноточечного кроссовера (Single point crossover — SPX). Решения, которые нарушают установленные ограничения (например, те, которые превышают вместительность автобуса), просто отбрасываются. Рисунок 3 показывает пример рекомбинации для упрощенного экземпляра проблемы с тремя маршрутами и их соответствующими остановками.

Мутация. Для этой проблемы использовался особый оператор мутации. Выбирается остановка (используется равномерное распределение) для каждого маршрута решения. Затем выбирается

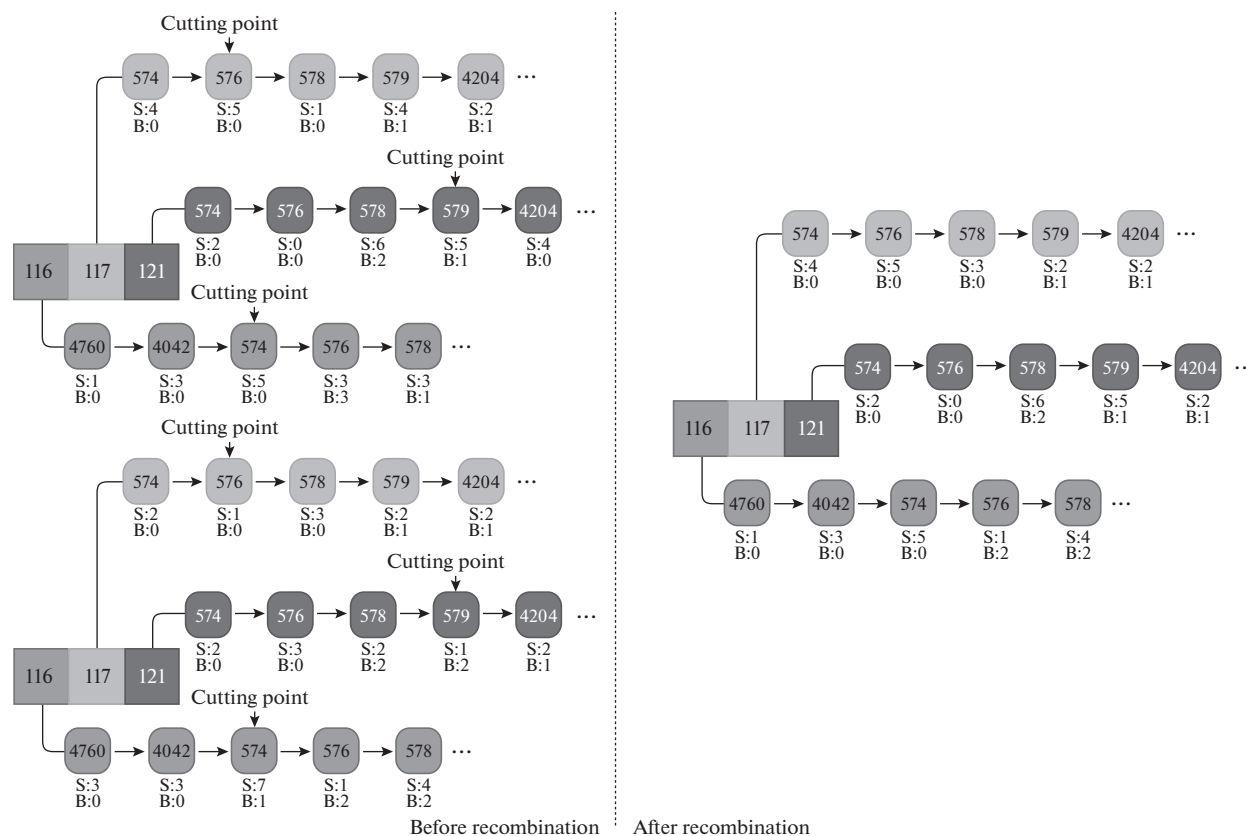


Рис. 3. Пример оператора рекомбинации.

действие: остановка остается неизменной, она исключается из маршрута или перемещается на случайное расстояние в метрах (не более максимального разрешаемого расстояния). В случае, если остановка удаляется, пассажиры, которые осуществляют посадку и высадку на этой остановке, распределяются равномерно между ближайшей следующей и предыдущей остановками. Когда остановка перемещается, пассажиры, осуществ-

ляющие посадку и высадку на этой остановке, равномерно распределяются между перемещенной остановкой и ближайшими следующей и предыдущей остановками.

Рисунок 4 показывает пример мутации для экземпляра проблемы с тремя маршрутами и их соответствующими остановками. Для каждого маршрута показано одно из возможных действий: остановка 4042 маршрута 116 остается неизменной,

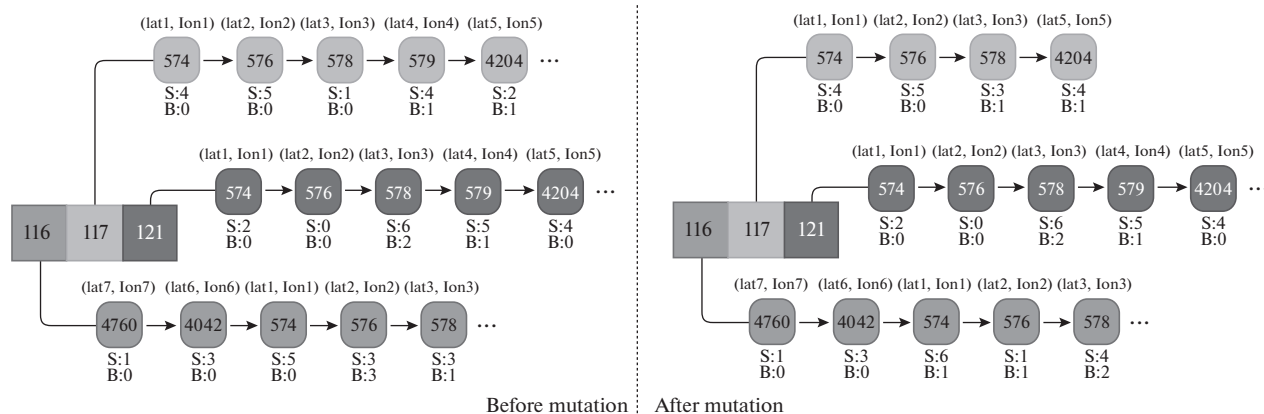


Рис. 4. Пример предлагаемого оператора мутации.

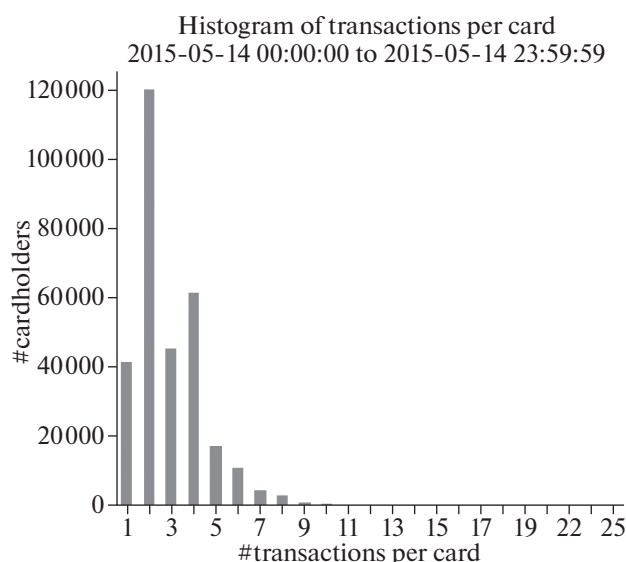


Рис. 5. Транзакции по карте, 2015.

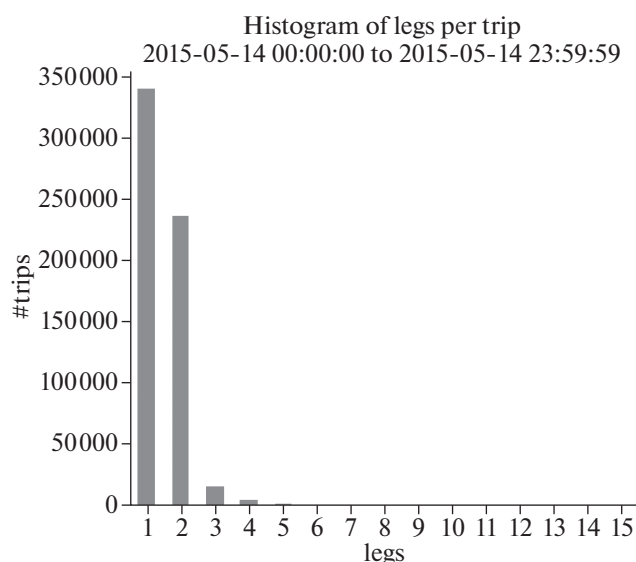


Рис. 6. Ног на поездку, 2015.

остановка 579 маршрута 117 устраняется, а остановка 576 маршрута 121 перемещается в новое место.

Корректирующая функция. Во время эволюционного поиска некоторые из решений, возникающих после применения эволюционных операторов, могут включать бесполезную информацию; например, они могут иметь маршруты с остановками, где пассажиры не производят посадку в автобус. Была реализована корректирующая функция: если есть две последовательные остановки, где пассажиры не производят посадки на автобус, то обе остановки объединяются в одну остановку посередине между ними. Исходные и конечные остановки исключаются в этом процессе поиска, потому что они должны оставаться неизменными для всех маршрутов.

5. РЕЗУЛЬТАТЫ АНАЛИЗА МОБИЛЬНОСТИ

В этом разделе описаны некоторые наиболее релевантные результаты анализа мобильности, выполненные с использованием методологии, описанной в разделе 1 в Монтевидео, Уругвай.

5.1. Модели мобильности, тепловые карты и другая полезная информация о системе общественного транспорта

Несколько показателей были рассчитаны при анализе больших данных, выполненном над информацией из системы общественного транспорта в Монтевидео.

Гистограммы на рисунках 5–6 представляют четыре примера соответствующих показателей, рассчитанных для рабочих дней в мае 2015 года.

Анализ показал, что большинство людей использовали смарт-карту для трансферных поездок. Сегодня ситуация может поменяться, так как в 2017 году власти Монтевидео начали программу поощрения использования смарт-карт, предложив снижение цен. Что касается транзакций по карте (рисунок 5), анализ показал, есть два четких максимума в двух и четырех случаях, что соответствует одной или двум поездкам, которые большинство людей делают в рабочие дни.

Анализ количества перегонов на поездку (Рисунок 6) показал, что почти все пассажиры используют смарт-карту поотдельности, как указано численно в таблице 2. 98.13% всех оплат по картам соответствует только одному пассажиру, тогда как всего 1.76% соответствуют двум или более лицам, путешествующим вместе. Этот результат свидетельствует о том, что граждане не знают о возможностях карты STM оплачивать сразу несколько билетов.

Что касается транзакций на автобус (Рисунок 7), анализ показал, что некоторые автобусы имеют значительно большее количество проверок смарт-карт, что, возможно, указывает на то, что общая длина каждого маршрута довольно велика (ситуация, которая признана аналитиками дорожного движения), а также некоторые линии с небольшими количествами валидаций карт указывают на меньшее количество пассажиров и, возможно, на недоиспользование ресурсов.

Наконец, в отношении транзакций в час (рисунок 8) анализ показал, что в рабочие часы спрос несколько меняется и что можно выделить 3 основных пика в часы высокой загрузки транспорта (7:00–8:00, 12:00–13:00 и 17:00–18:00). Анализ также позволил определить часы, в которых вы-

Table 2. Валидации карт STM относительно количества пассажиров

Пассажиры	Валидации	Процент
1	865066	98.13%
2	15488	1.76%
3	890	0.10%
4	94	0.01%
5	26	0.00%
6	2	0.00%
8	2	0.00%
9	2	0.00%

полняется самое низкое количество транзакций (3:00–4:00). Этот час использовался как точка разрезания для метода оценки конечной остановки, применяемого при построении матрицы ОН.

Профили загрузки. Профили загрузки полезны для определения областей, которые чрезмерно либо недостаточно используются в рамках определенного маршрута. Например, на Рисунке 9 представлено число людей, путешествующих по маршруту 121 (вариант 2399) с 17:00 до 19:00 часов (час пик) на неделе 16–20 Марта. Максимальная емкость линии представлена оранжевым цветом (32 пассажира сидят и 25 стоят), а количество пассажиров, которые садятся в автобус по маршруту, представлено синим. Средняя частота – один автобус каждые пять минут, согласно администрации города [12].

Рисунок 10 показывает профиль загрузки для маршрута 121 между 13:00 и 17:00 часами в ту же неделю марта 2015 года. Результаты показывают, что этот маршрут не используется чрезмерно, по-

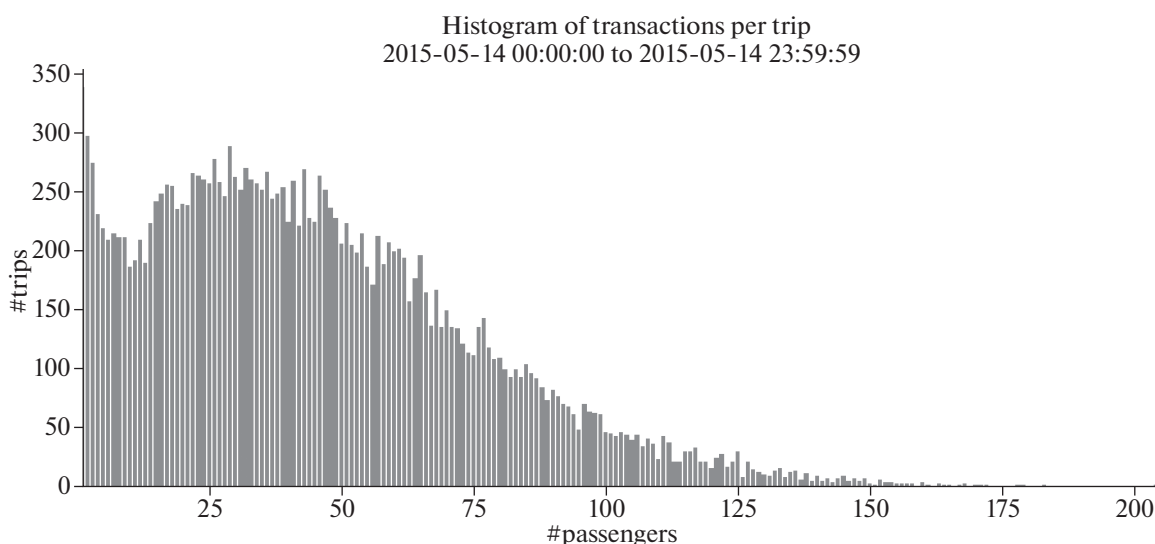
тому что, хотя автобус близок к максимальной загрузке в пиковое время, заполнение автобуса не превышает вместительность в каждой из зон. Это может быть связано с тем, что он имеет относительно короткий маршрут, поскольку он использует только 30 остановок через центр города.

Рисунок 11 показывает профиль загрузки для маршрута 300 (вариант 2065) с 13:00 до 17:00 и с 17:00 до 19:00, в мае 2015 года. Эта линия имеет более длинный маршрут, чем маршрут 121, с 74 остановками. Средняя частота этой линии составляет один автобус каждые 10 минут.

Результаты показывают, что маршрут 300 используется с 13:00 до 19:00 часов с более высокой потребностью в час пик. Этот паттерн повторяется на протяжении всех недель. В отличие от маршрута 121, маршрут 300 перемещается по более чем в два раза большему количеству районов и имеет частоту, уменьшенную наполовину, что приводит к избытку пассажиров.

Данный анализ не является исчерпывающим. Он представлен для иллюстрации шаблонов, которые могут быть обнаружены путем отображения профилей загрузки, полученных из обработанных данных мобильности. Эти показатели могут применяться для анализа спроса на конкретные проблемные линии с целью повышения качества обслуживания.

Тепловые карты. Тепловые карты – это визуальный инструмент, позволяющий анализировать спрос на определенные линии. Рисунки 12 и 13 показывают тепловые карты для районов города и Старого города, с 17:00 до 19:00 в марте 2015 года. Тепловые карты показывают, что наибольшие требования предъявляются в точках, где сосредоточено большое количество офисов или ра-

**Рис. 7.** Транзакции на автобус, 2015.

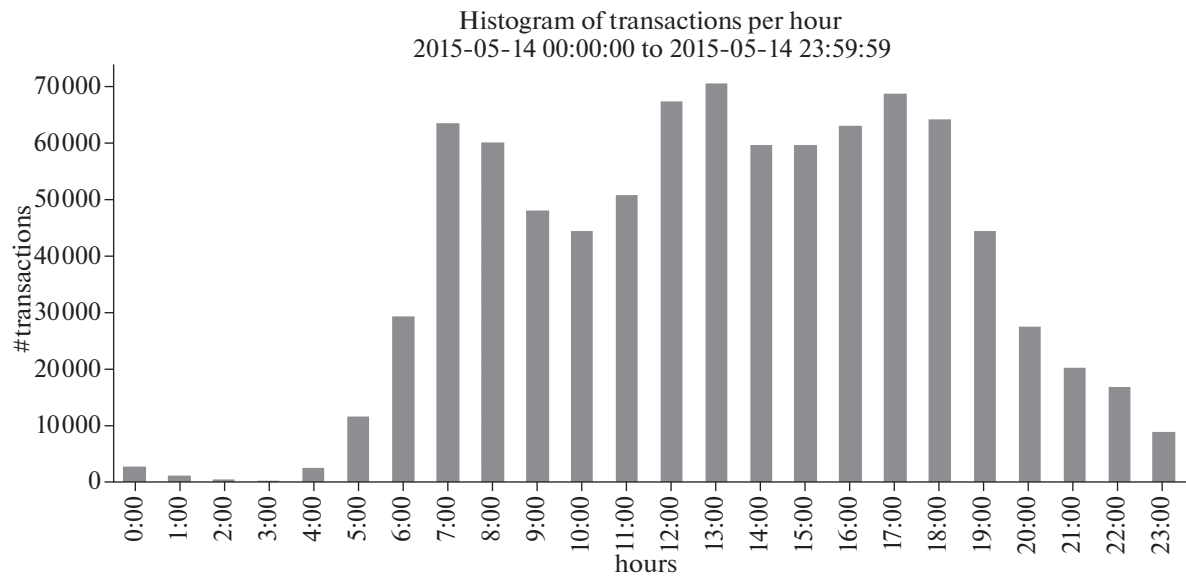


Рис. 8. Транзакций в час, 2015.

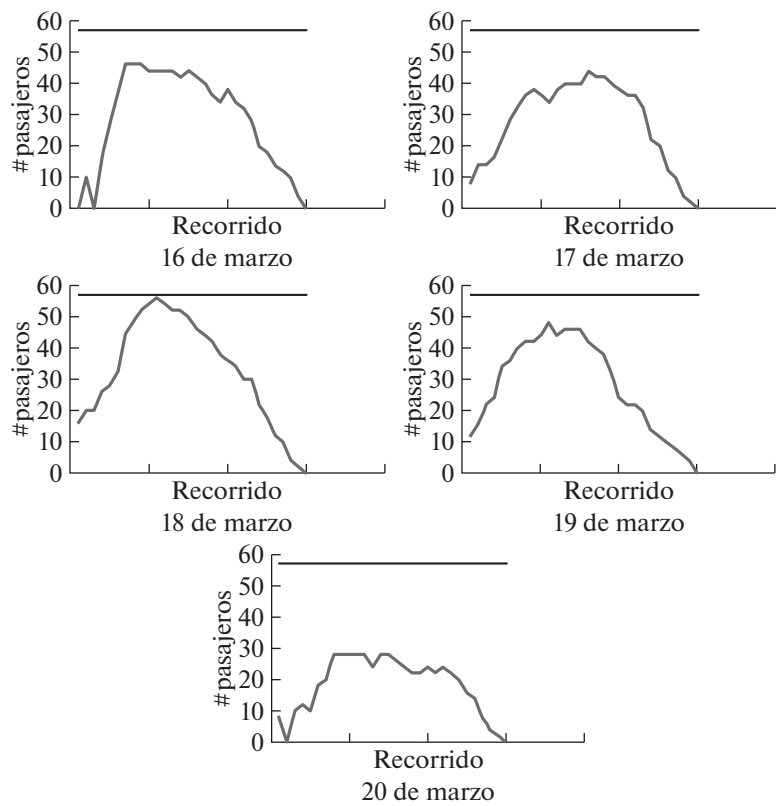


Рис. 9. Профиль загрузки для маршрута 121 с 17:00 до 19:00, 16–20 марта 2015 года.

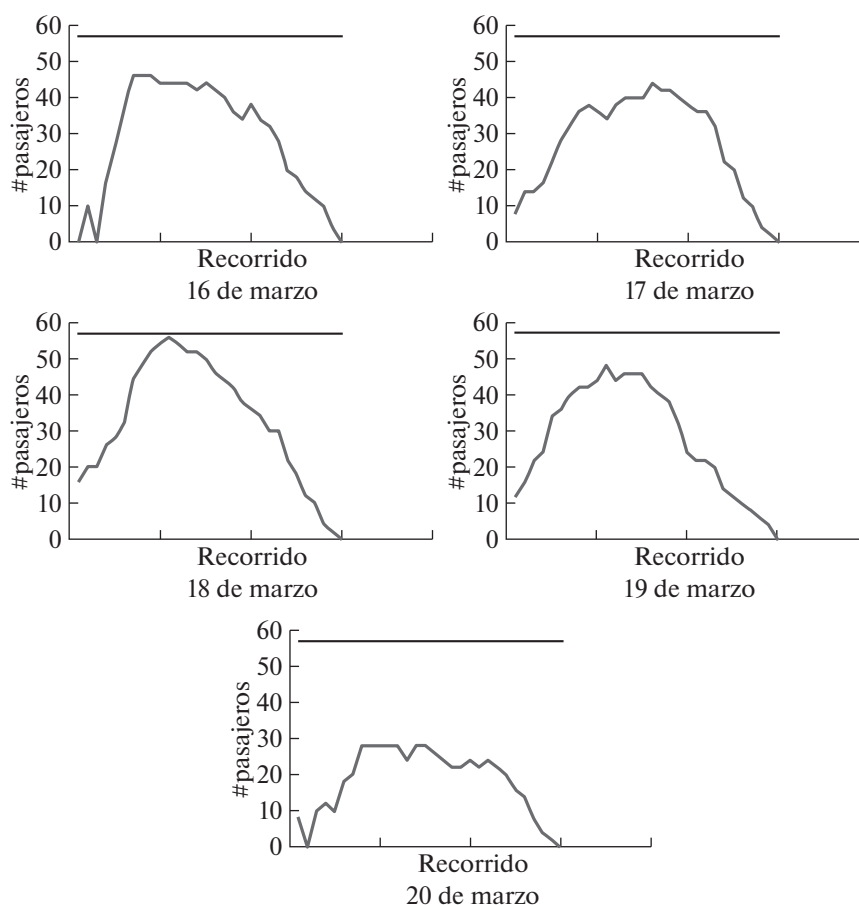


Рис. 10. Профиль загрузки маршрута 121 между 13:00 и 17:00 часами 16–20 марта, 2015.

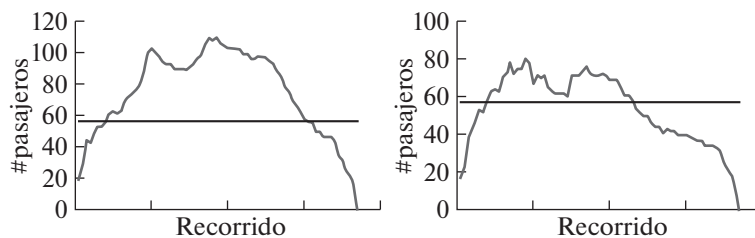


Рис. 11. Профиль загрузки маршрута 300 между 13:00 и 17:00 и между 17:00 и 19:00 часами 11 мая.

бочих мест, например, вблизи площади Независимости. Количество пассажиров уменьшается в противоположном направлении. Кроме того, на улицах с большим количеством маршрутов число пассажиров также увеличивается из-за того, что люди возвращаются в свои дома после рабочего дня, например, на автобусных остановках на проспекте 18 июля.

5.2. Матрицы спроса и отправления-назначения

Предлагаемая методология оценки спроса и ОН-матриц с использованием анализа больших

данных и распределенных вычислений была применена для системы общественного транспорта Монтевидео.

Время выполнения предлагаемого метода исследовалось при различном количестве вычислительных ресурсов (ядер) и различных рабочих нагрузках, назначенных каждому процессу в процессе выполнения, при реализации модели портфеля задач (Bag-of-Tasks – BoT).

Анализ эффективности представлен в таблицах 3 (матрица спроса) и 4 (матрица отправления-назначения, основанная на основных результа-



Рис. 12. Тепловая карта для центра города.

тах, опубликованных в наших предыдущих работах [6, 15, 22]). Значения ускорения – это оценка улучшения времени выполнения, полученного распределенным алгоритмом, по сравнению с последовательной реализацией с использованием только одного вычислительного ресурса.

Результаты показывают, что использование распределенного подхода значительно повышает эффективность обработки данных. Было достигнуто значительное улучшение времени выполнения: ускорение до 12.8 для прямых поездок и 16.4 для трансфертных поездок (матрица спроса) и ускорение до 17.1 (матрица ОН). Размер портфеля задач также влияет на время выполнения алгоритма. В обоих случаях лучшие результаты эффективности были получены с использованием портфеля задач в 5000 поездок и 24 ядер.

6. ЭКСПЕРИМЕНТАЛЬНЫЙ АНАЛИЗ: МКЭА ДЛЯ ПЕРЕМЕЩЕНИЯ АВТОБУСНЫХ ОСТАНОВОК

В этом разделе представлена экспериментальная оценка предложенного МКЭА для решения проблемы перемещения автобусных остановок.

6.1. Методика

Мультиобъективные показатели оптимизации. В существующей литературе предложено несколько показателей для оценки результатов полученных МКЭА [4, 2]. В этой статье рассматриваются несколько релевантных показателей для оценки предлагаемого МКЭА для проблемы перемещения автобусных остановок:

- генерационное расстояние (Generational Distance – GD): (нормированная) сумма расстояний между не доминируемыми решениями на границе Парето, найденными алгоритмом (P) и набором равномерно распределенных точек v на истинной границе Парето (P^*) (уравнение 6.1). Меньшие значения GD означают лучшее приближение к границе Парето

$$GD = \frac{1}{|P^*|} \sum_{v \in P^*} d(v, P) \quad (6.1)$$

- интервал (spacing): метрика, которая оценивает дисперсию неосновных решений в вычисленной границе Парето, определяемых уравнением 6.2;

- Распределение (spread): в отличие от интервала метрика распределения включает в себя информацию о крайних точках истинной границы



Рис. 13. Тепловая карта для центра города.

Парето, чтобы вычислить более точное значение распределения, как показано в уравнении 6.3 [4]. Меньшие значения метрики распределения означают лучшее распределение неориентированных решений в вычисленной границе Парето, а метрика принимает значение нуль для идеального равномерного распределения

$$spacing = \sqrt{\frac{\sum_{i=1}^{ND} (\bar{d} - d_i)^2}{ND - 1}} \quad (6.2)$$

$$spread = \frac{\sum_{h=1}^k d_h^e + \sum_{i=1}^{ND} (\bar{d} - d_i)^2}{\sum_{h=1}^k d_h^e + ND \times \bar{d}} \quad (6.3)$$

В уравнениях 6.2 и 6.3, d_i обозначает расстояние между i -м решением в вычисленной границе Парето и его ближайшем соседе, тогда как $\bar{d}$ — это среднее значение для всех d_i . В уравнении 6.3 d_h^e — это расстояние между крайним значением h -й функции в истинной границе Парето и ближайшей точкой в вычисленной границе Парето;

- относительный гиперобъем (Relative hypervolume — RHV): отношение объема (в пространстве объективных функций), покрытого вычисленной границей Парето, и объемом, покрытым реальной границей Парето.

В экспериментальном анализе истинная граница Парето — которая неизвестна для изученных экземпляров проблемы — была аппроксимирована путем сбора недоминированных решений, найденных в 30 независимых исполнениях, выполненных для каждого экземпляра проблемы.

Экземпляры проблемы. Предложенный МКЭА для перемещения автобусных остановок был оценен на основе реальных случаев указанной проблемы. Информация, полученная из разных источников, использовалась для создания реалистичных сценариев. Статическая информация о точках высадки, маршрутах и автобусных остановках была публично получена с помощью Открытого каталога данных в Intendencia de Montevideo [11]. Матрица ОН была сгенерирована посредством методологии, описанной в разделе 5. Для того, чтобы определить функции времени в пути, были рассмотрены три периода для создания разных случаев: утром (с 07:00 до 13:00), днем

Таблица 3. Время выполнения и ускорение для генерации матрицы спроса с использованием различного количества ядер и размера портфеля задач

#ядер	#ВоТ	Прямые поездки		Поездки с пересадками	
		время (мин)	ускорение	время (мин)	ускорение
		сред. $\pm \sigma$ (лучшее)		сред. $\pm \sigma$ (лучшее)	
1	1	25920	—	30240	—
16	5000	2092.1 $\pm$ 3.4 (2089.6)	12.4	2648.9 $\pm$ 3.2 (2645.5)	11.4
16	10000	2372.4 $\pm$ 1.8 (2371.1)	10.9	3068.8 $\pm$ 3.5 (3063.2)	9.9
24	5000	1582.7 $\pm$ 2.4 (1579.4)	16.4	2371.1 $\pm$ 2.5 (2368.1)	12.8
24	10000	1858.2 $\pm$ 2.1 (1855.9)	14.0	2617.9 $\pm$ 3.3 (2614.3)	11.6

(с 13:00 до 19:00) и ночью (с 19:00 до 7:00) с учетом различных условий движения в городе.

Всего было создано 30 экземпляров проблемы:

- 11 — *малые экземпляры*, в том числе 10 строк с указанием точек отправления и конечных станциях в населенных районах Монтевидео и использованием матрицы ОН для января 2015 года;
- 9 — *средние экземпляры*: между 30 и 60 маршрутов с точками отправления в соседних окрестностях Монтевидео и пунктах назначения в каждом другом районе с использованием матрицы ОН в период с января по май 2015 года;
- 10 — *большие экземпляры*: в том числе более 100 маршрутов из разных районов Монтевидео с использованием матрицы ОН за весь 2015 год.

Все сведения о экземплярах проблем общедоступны на веб-сайте проекта www.fing.edu.uy/inco/grupos/cescal/hpc/IOTU.

Вычислительная платформа. Экспериментальный анализ был выполнен на сервере HP Proliant DL385 G7 с 2 процессорами AMD Opteron 6172 на частоте 2,66 ГГц, по 12 ядер каждый и на 72 ГБ оперативной памяти от кластера FING, платформы HPC в Universidad de la República, Уругвай [19].

Статистический анализ. Из-за стохастической природы ЭА, разные результаты могут быть получены для разных запусков одного и того же экземпляра проблемы. Для получения значимых результатов, для каждого экземпляра проблемы было произведено 30 независимых запусков.

Таблица 4. Время выполнения и ускорение для генерации матрицы ОН с использованием различного количества ядер и размера портфеля задач

#ядер	#ВоТ	время (min) $\pm \sigma$ (лучшее)	ускорение
1	1	63972	—
16	5000	5070.3 $\pm$ 2.5 (5656.2)	12.58
16	10000	5553.1 $\pm$ 2.4 (5525.9)	11.19
24	5000	3730.2 $\pm$ 2.1 (3728.6)	17.10
24	10000	4433.4 $\pm$ 2.0 (4401.5)	14.39

6.2. Настройка параметров

Был выполнен анализ конфигурации для определения наилучших значений для каждого соответствующего параметра в предлагаемом МКЭА. Анализ параметров был выполнен на трех малых экземплярах проблемы, которые отличались от тех, которые использовались в экспериментальной оценке, чтобы избежать смещения.

В экспериментах изменялся размер популяции (#P, значениями кандидатов 50, 100 и 150), вероятность рекомбинации (p_R , значения кандидата 0.6, 0.75 и 0.9) и вероятность мутации (p_M , значения кандидата 0.001, 0.01 и 0.05). Для каждой комбинации параметров выполнялось 30 независимых запусков из 5000 поколений. Чтобы выбрать наилучшую конфигурацию, была исследована метрика относительного гиперобъема, так как она учитывает как дисперсию найденных решений, так и близость к границе Парето.

В таблице 5 представлены полученные результаты, из которых можно сделать вывод, что использование размера популяции 100, вероятности рекомбинации 0.9 и мутации 0.05 позволяет найти наилучшие решения для оцениваемых экземпляров проблемы.

6.3. Численные результаты

В таблицах 6–8 приведены численные результаты для мультиобъективных показателей оптимизации. Среднее, стандартное отклонение и наилучшее значение (в скобках), рассчитанное в 30 независимых исполнениях предлагаемого МКЭА, представлено для каждого экземпляра проблемы.

Таблица 9 суммирует полученные результаты, сгруппированные по размеру экземпляра.

Результаты показывают, что значения расстояний поколений (GD) очень малы, что говорит о хорошей сходимости к реальной границе Парето. Достигнутые значения распространения также малы, что свидетельствует о хорошем распределении не доминирующих решений на границе Па-

Таблица 5. Результаты RHV для разных комбинаций значений параметров

$(\#P, p_R, p_M)$	<i>instance #1</i>	<i>instance #2</i>	<i>instance #3</i>	$(\#P, p_R, p_M)$	<i>instance #1</i>	<i>instance #2</i>	<i>instance #3</i>
(50, 0.6, 0.01)	0.85 ± 0.08	0.79 ± 0.10	0.90 ± 0.08	(100, 0.75, 0.01)	0.75 ± 0.09	0.80 ± 0.08	0.95 ± 0.06
(50, 0.6, 0.05)	0.80 ± 0.11	0.50 ± 0.28	0.85 ± 0.10	(100, 0.9, 0.01)	0.82 ± 0.11	0.70 ± 0.12	1.00 ± 0.32
(50, 0.6, 0.001)	0.68 ± 0.11	0.70 ± 0.13	0.72 ± 0.10	(100, 0.9, 0.05)	1.00 ± 0.27	1.00 ± 0.17	0.63 ± 0.18
(50, 0.75, 0.05)	0.87 ± 0.10	0.68 ± 0.12	1.00 ± 0.22	(100, 0.9, 0.001)	0.77 ± 0.08	0.85 ± 0.07	1.00 ± 0.10
(50, 0.75, 0.001)	0.78 ± 0.11	0.66 ± 0.15	0.71 ± 0.11	(150, 0.6, 0.01)	0.65 ± 0.11	0.74 ± 0.11	0.81 ± 0.13
(50, 0.75, 0.01)	0.85 ± 0.11	0.88 ± 0.09	0.96 ± 0.11	(150, 0.6, 0.05)	0.79 ± 0.10	0.93 ± 0.10	0.86 ± 0.10
(50, 0.9, 0.01)	0.89 ± 0.14	0.73 ± 0.10	0.72 ± 0.17	(150, 0.6, 0.001)	0.88 ± 0.08	0.87 ± 0.08	0.84 ± 0.07
(50, 0.9, 0.05)	0.79 ± 0.16	0.75 ± 0.12	0.74 ± 0.10	(150, 0.75, 0.05)	0.84 ± 0.06	0.83 ± 0.10	0.75 ± 0.12
(50, 0.9, 0.001)	0.80 ± 0.09	0.66 ± 0.140	0.79 ± 0.12	(150, 0.75, 0.001)	0.93 ± 0.67	0.93 ± 0.10	0.80 ± 0.13
(100, 0.6, 0.01)	0.73 ± 0.10	0.74 ± 0.10	0.86 ± 0.13	(150, 0.75, 0.01)	0.81 ± 0.07	0.79 ± 0.08	0.85 ± 0.09
(100, 0.6, 0.05)	0.70 ± 0.10	1.00 ± 0.08	0.77 ± 0.10	(150, 0.9, 0.01)	0.68 ± 0.14	0.83 ± 0.15	0.63 ± 0.18
(100, 0.6, 0.001)	0.78 ± 0.11	0.91 ± 0.07	0.98 ± 0.07	(150, 0.9, 0.05)	0.55 ± 0.13	0.44 ± 0.15	0.30 ± 0.18
(100, 0.75, 0.05)	0.87 ± 0.08	0.89 ± 0.05	0.75 ± 0.14	(150, 0.9, 0.001)	0.74 ± 0.08	0.81 ± 0.08	0.94 ± 0.08
(100, 0.75, 0.001)	0.89 ± 0.12	0.85 ± 0.08	0.74 ± 0.11				

Таблица 6. Метрики многокритериальной оптимизации для экземпляров малых проблем

Экземпляр	GD	spacing	spread	RHV
small #1	0.66 ± 0.53 (0.0)	0.09 ± 0.05 (0.01)	0.97 ± 0.14 (0.72)	1.0 ± 0.2 (1.0)
small #2	0.02 ± 0.01 (0.0)	0.01 ± 0.01 (0.0)	0.71 ± 0.08 (0.62)	1.0 ± 0.2 (1.0)
small #3	0.66 ± 0.53 (0.0)	0.09 ± 0.05 (0.01)	0.97 ± 0.14 (0.72)	1.0 ± 0.2 (1.0)
small #4	0.02 ± 0.01 (0.0)	0.04 ± 0.10 (0.0)	1.02 ± 0.17 (0.75)	0.9 ± 0.1 (1.0)
small #5	0.01 ± 0.01 (0.0)	0.04 ± 0.05 (0.0)	1.04 ± 0.13 (0.83)	0.8 ± 0.1 (1.0)
small #6	0.01 ± 0.0 (0.0)	0.05 ± 0.04 (0.0)	1.03 ± 0.16 (0.80)	0.9 ± 0.1 (1.0)
small #7	0.01 ± 0.0 (0.0)	0.03 ± 0.02 (0.0)	0.99 ± 0.08 (0.84)	0.8 ± 0.1 (1.0)
small #8	0.02 ± 0.01 (0.01)	0.05 ± 0.10 (0.0)	1.11 ± 0.13 (0.94)	0.8 ± 0.1 (1.0)
small #9	0.02 ± 0.01 (0.01)	0.04 ± 0.01 (0.0)	1.06 ± 0.12 (0.86)	0.8 ± 0.2 (1.0)
small #10	0.02 ± 0.01 (0.0)	0.03 ± 0.02 (0.0)	1.0 ± 0.11 (0.84)	1.0 ± 0.2 (1.0)
small #11	0.07 ± 0.04 (0.0)	0.81 ± 0.07 (0.0)	1.31 ± 0.18 (0.85)	1.0 ± 0.2 (1.0)

Таблица 7. Метрики многокритериальной оптимизации для экземпляров средних проблем

Экземпляр	GD	spacing	spread	RHV
medium #1	0.08 ± 0.02 (0.03)	0.07 ± 0.05 (0.03)	0.86 ± 0.07 (0.74)	1.0 ± 0.1 (1.0)
medium #2	0.07 ± 0.01 (0.03)	0.05 ± 0.01 (0.03)	0.82 ± 0.05 (0.71)	1.0 ± 0.1 (1.0)
medium #3	0.06 ± 0.02 (0.02)	0.83 ± 0.06 (0.02)	0.83 ± 0.06 (0.02)	1.0 ± 0.1 (1.0)
medium #4	0.05 ± 0.01 (0.03)	0.04 ± 0.02 (0.02)	0.88 ± 0.05 (0.74)	1.0 ± 0.1 (1.0)
medium #5	0.07 ± 0.02 (0.02)	0.06 ± 0.03 (0.03)	0.88 ± 0.06 (0.71)	1.0 ± 0.1 (1.0)
medium #6	0.76 ± 0.2 (0.0)	1.10 ± 0.20 (0.0)	0.88 ± 0.06 (0.71)	1.0 ± 0.1 (1.0)
medium #7	0.06 ± 0.02 (0.01)	0.07 ± 0.04 (0.02)	0.91 ± 0.07 (0.73)	1.0 ± 0.1 (1.0)
medium #8	0.06 ± 0.01 (0.03)	0.07 ± 0.04 (0.02)	0.88 ± 0.07 (0.75)	1.0 ± 0.1 (1.0)
medium #9	0.07 ± 0.02 (0.03)	0.05 ± 0.02 (0.02)	0.81 ± 0.06 (0.67)	1.0 ± 0.1 (1.0)

Таблица 8. Метрики многокритериальной оптимизации для экземпляров больших проблем

Экземпляр	GD	spacing	spread	RHV
large #1	0.08 ± 0.02 (0.03)	0.05 ± 0.02 (0.02)	0.83 ± 0.07 (0.70)	1.0 ± 0.1 (1.0)
large #2	0.09 ± 0.02 (0.02)	0.06 ± 0.01 (0.03)	0.86 ± 0.07 (0.68)	1.0 ± 0.1 (1.0)
large #3	0.08 ± 0.02 (0.03)	0.07 ± 0.03 (0.02)	0.91 ± 0.08 (0.76)	1.0 ± 0.1 (1.0)
large #4	0.12 ± 0.02 (0.05)	0.08 ± 0.03 (0.03)	0.90 ± 0.06 (0.80)	1.0 ± 0.1 (1.0)
large #5	0.08 ± 0.02 (0.02)	0.07 ± 0.03 (0.02)	0.74 ± 0.09 (0.74)	1.0 ± 0.1 (1.0)
large #6	0.09 ± 0.01 (0.05)	0.06 ± 0.02 (0.02)	0.89 ± 0.07 (0.72)	1.0 ± 0.1 (1.0)
large #7	1.19 ± 0.28 (0.28)	2.12 ± 1.50 (0.0)	0.89 ± 0.07 (0.72)	1.0 ± 0.3 (1.0)
large #8	0.08 ± 0.02 (0.02)	0.07 ± 0.02 (0.02)	0.90 ± 0.05 (0.75)	1.0 ± 0.1 (1.0)
large #9	0.09 ± 0.02 (0.02)	0.07 ± 0.02 (0.03)	0.90 ± 0.06 (0.77)	1.0 ± 0.1 (1.0)
large #10	0.08 ± 0.02 (0.01)	0.05 ± 0.02 (0.02)	0.82 ± 0.07 (0.65)	1.0 ± 0.1 (1.0)

Таблица 9. Мультиобъектная метрика оптимизации для предлагаемого NSGA-II

Размер экземпляра	GD	spacing	spread	RHV
малый	0.14 ± 0.10 (0.0)	0.12 ± 0.05 (0.02)	1.02 ± 0.13 (0.80)	0.9 ± 0.1 (1.0)
средний	0.14 ± 0.04 (0.01)	0.17 ± 0.05 (0.02)	0.86 ± 0.06 (0.64)	1.0 ± 0.1 (1.0)
большой	0.20 ± 0.04 (0.05)	0.27 ± 0.17 (0.02)	0.86 ± 0.07 (0.73)	1.0 ± 0.1 (1.0)

рето. Наконец, достигнутые значения RHV подтверждают хорошую конвергенцию и разнообразие найденных решений.

Очень важный результат вытекает из сравнения решений МКЭА с реальной ситуацией. Это сравнение выполняется для каждого экземпляра проблемы с номером 1, который соответствует реальным данным и сценариям в 2015 году.

В таблице 10 представлены численные значения функций времени поездки и *прибыли*, определяемой в виде разницы между доходом от проданных билетов и стоимостью предоставления услуги. Это важная функция с точки зрения компаний общественного транспорта. Время в пути выражается в часах, а прибыль выражается в уругвайских песо.

Результаты показывают, что предлагаемый МКЭА способен находить решения, которые улучшают время в пути до 16.70% и прибыль до 33.92% от реальной ситуации.

Для небольших экземпляров, где есть несколько маршрутов для изменения, процент улучшения значительно выше. Эти результаты могут быть объяснены тем, что чем больше количество маршрутов, перемещение одной остановки автобуса, оказывает большее влияние на остальные маршруты. В этом случае МКЭА может сосредоточиться на соответствующих улучшениях для сети общественного транспорта.

7. ВЫВОДЫ И ПРОДОЛЖЕНИЕ РАБОТЫ

Нами было рассмотрены две важные проблемы, связанные с системами общественного транспорта в современных умных городах: анализ моделей мобильности и перенос автобусных остановок.

Для анализа моделей мобильности был применен подход анализа больших данных с GPS и автоматов продажи билетов с целью получения информации и расчета показателей для характеристики системы общественного транспорта.

Таблица 10. Сравнительные результаты: реальная ситуация в 2015 году и решения, предложенные предлагаемым МКЭА

Размер экземпляра	Реальные значения в 2015		Значения МКЭА		Улучшение	
	время	прибыль	время	прибыль	время	прибыль
small	70.46	303611.63	58.69	459443.96	16.70%	33.92%
medium	386.15	550285.98	365.19	666476.22	5.43%	17.43%
large	687.37	1029411.29	669.85	1226279.14	2.55%	16.05%

Предлагаемый подход позволяет вычислять соответствующие показатели и сгенерировать матрицы спроса и ОН, применяя метод маршрутизации для оценки точек назначения пассажиров.

Нами были вычислено несколько показателей для оценки моделей мобильности с использованием данных из системы общественного транспорта. Для обработки больших данных была предложена распределенная реализация, достигающая значительных улучшений времени выполнения по сравнению со стандартной последовательной реализацией (ускорение до 17.10 при использовании 24 вычислительных ресурсов).

Для решения проблемы переноса автобусных остановок был предложен многокритериальный эволюционный подход. Он обеспечивает оптимизацию качества обслуживания, минимизируя время в пути и эксплуатационные расходы системы. Этот подход учитывает как оптимизационные критерии, ориентированные как на пользователя, так и на поставщика услуг.

Предлагаемый МКЭА был оценен на основе набора примеров, созданных с использованием реальных данных и матриц спроса/ОН, рассчитанных на предыдущем этапе. Экспериментальные результаты показали, что предлагаемый МКЭА способен обеспечить улучшенные решения по обему задач. Улучшения в решениях МКЭА по реальной ситуации достигли 16.7% по времени поездок и 33.9% по их стоимости. Эти результаты подтверждают, что предлагаемый подход полезен для лиц, принимающих решения, и органов власти в целях повышения качества обслуживания системы общественного транспорта.

Основные направления будущей работы связаны с улучшением предлагаемого решения. Мы разработаем другие эффективные алгоритмы для лучшего изучения основных характеристик систем общественного транспорта в соответствии с распределенным подходом для обработки больших данных. Что касается переноса автобусных остановок, модель проблемы может быть расширена, посредством включения в нее актуальной информации и данных реального времени. Это является важным для предложения временных модификаций для решения непредвиденных ситуаций, которые влияют на систему общественного транспорта. Предлагаемая модель и алгоритм также могут быть применены к исследованиям, проводимым в других городах и сценариях.

СПИСОК ЛИТЕРАТУРЫ

1. Bielli M., Carambia M., Carotenuto P. Genetic algorithms in bus network optimization. *Transportation Research Part C: Emerging Technologies*. 2002. V. 10. P. 19–34.
2. Coello C., Van Veldhuizen D., Lamont G. *Evolutionary algorithms for solving multiobjective problems*. Kluwer, 2002.
3. Deakin M., Al Waer H. From intelligent to smart cities. *Intelligent Buildings International*. 2011. V. 3. P. 133–139.
4. Deb K. *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons, 2001.
5. Deb K., Chakroborty P., Subrahmanyam P.S. Optimal scheduling of urban transit systems using genetic algorithms. *Journal of Transportation Engineering*. 1995. V. 121. P. 544–553.
6. Fabbiani E., Vidal P., Massobrio R., Nesmachnow S. Distributed big data analysis for mobility estimation in intelligent transportation systems. In *High Performance Computing*. 2016. P. 146–160.
7. Figueiredo L., Jesus I., Tenreiro Machado J., Rui Ferreira J., Carvalho J.L. Towards the Development of Intelligent Transportation Systems. In *IEEE Intelligent Transportation Systems*. 2001. P. 1206–1211.
8. Grava S. *Urban transportation systems: choices for communities*. McGraw-Hill, 2002.
9. Intendencia de Montevideo. STM: Sistema de Transporte Metropolitano. Disponible en: <http://www.montevideo.gub.uy/transito-y-transporte>. [Online: July 2017].
10. Intendencia de Montevideo. STM: Sistema de Transporte Metropolitano. Disponible en: <http://www.montevideo.gub.uy/transito-y-transporte>. [Online: July 2017].
11. Intendencia de Montevideo. Plan de movilidad urbana: hacia un sistema de movilidad accesible, democrático y eficiente. Disponible en: http://www.montevideo.gub.uy/sites/default/files/plan_de_movilidad.pdf, 2010. [Online: July 2017].
12. Intendencia de Montevideo. Catálogo de Datos Abiertos. Available: <https://catalogodatos.gub.uy>, 2014. [Online: July 2016].
13. Intendencia de Montevideo. Horarios de ómnibus. available: <http://www.montevideo.gub.uy/horariosSTM/>, 2017. [Online; último acceso: Julio 2017].
14. Johar A., Jain S., Garg P. Transit network design and scheduling using genetic algorithm – a review. *An International Journal of Optimization and Control: Theories & Applications*. 2015. V. 6. P. 9–22.
15. Król A. Application of the Genetic Algorithm for Optimization of the Public Transportation Lines, pages 135–146. Springer International Publishing, Cham, 2017.
16. Massobrio R., Pias A., Vázquez N., Nesmachnow S. Map-Reduce for Processing GPS Data from Public Transport in Montevideo, Uruguay. In *2nd Argentinian Symposium on Big Data*. 2016. P. 41–54.
17. Mellegård E. Obtaining origin/destination matrices from cellular network data. <http://publications.lib.chalmers.se/records/fulltext/154702.pdf>, 2011. [Online: September 2017].
18. Munizaga M., Palma C. Estimation of a disaggregate multimodal public transport origin-destination matrix

- from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies*. 2012. V. 24. P. 9–18.
19. *Nesmachnow S.* Computación científica de alto desempeño en la Facultad de Ingeniería, Universidad de la República. *Revista de la Asociación de Ingenieros del Uruguay*. 2010. V. 61 (1). P. 12–15. Text in Spanish.
 20. *Nesmachnow S.* An overview of metaheuristics: accurate and efficient methods for optimisation. *International Journal of Metaheuristics*. 2014. V. 3 (4). P. 320–347.
 21. *Nesmachnow S.* Using metaheuristics as soft computing techniques for efficient optimization. *An Encyclopedia of Information Science and Technology*, Third Edition. 2015. P. 7390–7399.
 22. *Nesmachnow S., Bana S., Massobrio R.* A distributed platform for big data analysis in smart cities: combining intelligent transportation systems and socioeconomic data for montevideo, uruguay. *EAI Endorsed Transactions on Smart Cities*. 2017. V. 2 (5). P. 1–18.
 23. *Pelletier M., Trépanier M., Morency C.* Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*. 2011. V. 19 (4). P. 557–568.
 24. *Sun C.* Dynamic Origin/Destination Estimation Using True Section Densities. *California PATH Research Report*, 2000.
 25. *Toole J.L., Colak S., Sturt B., Alexander L.P., Eysukoff A., González M.C.* The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*. 2015. V. 58. Pt B. P. 162–177.
 26. *Trépanier M., Tranchant N., Chapleau R.* Individual trip destination estimation in a transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems*. 2007. V. 11 (1). P. 1–14.
 27. *Wang W., Attanucci J., Wilson N.* Bus passenger origin-destination estimation and related analyses using automated data collection systems. *Journal of Public Transportation*. 2011. V. 14 (4). P. 131–150.
 28. *Xiong Y., Schneider J.B.* Transportation network design using a cumulative genetic algorithm and neural network. *Transportation Research Record*. 1993. V. 1364.