

Лебедева Ольга Анатольевна,

к.т.н., доцент, ФГБОУ ВО «Ангарский государственный технический университет»,
e-mail: kravhome@mail.ru,

Баянов Семен Владимирович,

студент гр. ТТП-23-1, ФГБОУ ВО «Ангарский государственный технический университет»,
e-mail: cyja642@yandex.ru

ОЦЕНКА МАТРИЦЫ КОРРЕСПОНДЕНЦИЙ НА ОСНОВЕ ОПРЕДЕЛЕНИЯ ПРИЗНАКОВ В ДАННЫХ СМАРТ-КАРТЫ

Lebedeva O.A., Bayanov S.V.

ASSESSMENT OF THE CORRESPONDENCE MATRIX BASED ON THE DEFINITION OF FEATURES IN SMART CARD DATA

Аннотация. В общественном транспорте смарт-карты в основном используются для автоматического сбора платы за проезд, что генерирует огромные объемы данных. Одним из применений данных смарт-карт является оценка матрицы «пункт отправления-назначения» общественного транспорта. Основная цель этой статьи – критический анализ существующей литературы по основным этапам процесса оценки. Эти этапы включают процессы очистки данных, оценку неизвестных параметров, обнаружение пересадок, проверку разработанных алгоритмов и, в конечном итоге, оценку пункта отправления-назначения.

Ключевые слова: общественный транспорт, данные смарт-карт, оценка матриц в общественном транспорте.

Abstract. In public transport, smart cards are mainly used to automatically collect fares, which generate huge amounts of data. One of the applications of these smart cards is to evaluate the "point of departure-destination" matrix of public transport. The main purpose of this article is a critical analysis of the existing literature on the main stages of the assessment process. These steps include data purification processes, evaluation of unknown parameters, detection of transfers, verification of developed algorithms, and, ultimately, evaluation of the destination point.

Keywords: public transport, smart card data, evaluation of matrices in public transport.

Развитие технологий способствует внедрению интеллектуальных транспортных систем в сектор общественного транспорта. Многие перевозчики используют системы автоматического сбора данных (ADC), включая автоматический сбор платы за проезд (AFC) с использованием смарт-карт, автоматическое определение местоположения транспортных средств (AVL) и автоматический подсчет пассажиров (APC). Данные AFC в основном используются для сбора платы за проезд, что генерирует огромные объемы, которые легко доступны перевозчикам и транспортным организациям. В последние двадцать лет наблюдается тенденция к увеличению применения больших данных для извлечения информации, которая используется в транспортных приложениях, охватывающих как долгосрочное планирование, так и оперативное управление и контроль в реальном времени.

Матрица OD предоставляет информацию о спросе на мобильность в сети, которая является важным входным параметром для крупномасштабного транспортного моделирования. Матрица мобильности наряду с матрицей транспортных потоков являются входными данными для моделирования, результаты которого могут иметь множество применений, таких как:

- оценка дорожной инфраструктуры, проектирование светофорного регулирования;
- оценка заторов, ценообразование в связи с заторами и функционирование городской территории помогает в планировании общественного транспорта на стратегическом, тактическом и оперативном уровнях.

Потенциальные области применения матрицы мобильности включают, проектирование маршрутов и сетей общественного транспорта; и распределение подвижного состава по сети; определение приоритетов финансирования проектов, связанных с транспортом; выявление разрывов между спросом и предложением транспортных услуг; перегруженность; планирование технического обслуживания; и проверка крупномасштабных моделей распределения.

Оценка матрицы транспортных потоков осуществляется с использованием ряда традиционных и продвинутых наборов данных. Традиционные наборы данных включают выборочные опросы о поездках и подсчет транспортных средств с помощью метода петель или вручную. Современные наборы – включают данные Bluetooth, систему распознавания номерных знаков транспортных средств, AVL, APC, данные о местоположении телефонов и данные социальных сетей. Хотя указанные системы имеют более высокую первоначальную стоимость, она остается низкой на протяжении всего срока службы.

Типичная структура оценки может быть разделена на четыре части:

1. Очистка данных.
2. Оценка местоположения посадки и высадки.
3. Оценка пересадочности между зонами.
4. Проверка предложенной методологии.

В дополнение к вышеуказанным четырем шагам, система AFC только для входа требует дополнительного алгоритма для определения остановки высадки, поскольку в ней отсутствует такая информация. Необходимость оценки информации о посадке обычно зависит от полей, доступных в данных смарт-карты. Если данные смарт-карты не содержат информацию о месте или времени посадки, или отсутствуют из-за неисправности системы, потребуется оценить неизвестную величину.

Системы AFC можно классифицировать следующим образом:

- ✓ система «на входе» – в такой системе пассажирам необходимо приложить смарт-карту только при посадке в транспортное средство. В системах с фиксированной оплатой проезда, где стоимость проезда не зависит от расстояния;
- ✓ система «на входе и выходе» – в этой системе пассажиру необходимо приложить смарт-карту при посадке и при выходе из транспортного средства; является предпочтительной в районах с тарифами, зависящими от расстояния.

Иногда в литературе – систему только для входа и систему для выхода/входа называют открытой и закрытой системами соответственно. В целом, закрытой системой считается та, в которой пассажиры могут входить и выход-

дять только после прикладывания смарт-карт к считывателю, тогда как открытой системой считается система, в которой пассажиры имеют неограниченный доступ в подвижном составе.

Типичные наборы данных включают идентификатор транспортного средства, остановку посадки, остановку высадки, время посадки и высадки. Существует множество причин неправильного считывания данных со смарт-карт, включая проблемы, связанные с интеграцией больших наборов данных, неточность GPS [1].

Очистка данных требует информации об источниках и типах ошибок для уточнения данных. Двумя основными источниками ошибок являются сбои оборудования и человеческий фактор. Сбои оборудования могут быть вызваны неисправностью считывателя смарт-карт, рассинхронизацией часов устройств, установленного GPS и/или системы в целом. Человеческий фактор – не прикладывание карты, прикладывание карты другой платежной системы.

В целом, такие сбои приводят к ошибкам в данных смарт-карт: отсутствие записи о времени и/или месте посадки/высадки; совпадение зарегистрированного времени и/или места посадки и/или высадки; более раннее время посадки для той же поездки; отсутствие идентификатора смарт-карты; дублирование события транзакции на остановке, местоположение которой невозможно отследить. На более позднем этапе анализа могут быть обнаружены поездки со временем в пути, превышающим типичное – поездка продолжительностью несколько часов в средних городах. Минимальное время в пути и расстояние соответственно, могут указывать на посадку в неправильный автобус и выход из него после осознания этого факта, безбилетный проезд или другие человеческие ошибки (многократное прикладывание карты, когда пассажир думает, что списание не произошло). Необходимо установить верхний и нижний пределы времени и расстояния поездки для каждого этапа, подлежащего анализу.

Тип ошибки и ее правильная интерпретация имеют решающее значение при оценке OD. Неправильная интерпретация ошибки может привести к неточному значению OD. Следовательно, данные должны быть проверены на все возможные несоответствия, которые должны быть устранены до и во время анализа, если таковые обнаружены.

После того, как данные смарт-карт будут уточнены и не содержат ошибок, их можно использовать для оценки stOD.

Количество неизвестных параметров зависит от типа данных и атрибутов, зарегистрированных в соответствии с транзакцией, совершенной в общественном транспорте.

Для оценки места назначения качество данных играет жизненно важную роль. На основе доступных полей в записанных данных системы, работающей только с входами, оценку остановки посадки можно разделить на три категории, показанные в таблице 1.

Таблица 1 – Категории задач оценки остановочного пункта посадки на основе доступных признаков в данных смарт-карты

Время посадки	Остановка посадки	Примечания
✓	✓	для оценки места посадки не требуется никаких вычислений
✓	✗	место посадки можно рассчитать, интегрировав данные смарт-карты с другим набором данных, таким как AVL, данные расписания, если таковые имеются
✗	✗	место посадки можно определить, используя данные GTFS, APC и AFC

Первая категория – это когда в данных записаны и остановка, и время посадки, следовательно, их можно использовать напрямую для приложения. Вторая категория – это когда записано время посадки, но отсутствует информация об остановке посадки. В этом случае необходимое поле можно определить, объединив данные смарт-карты с другими источниками данных (AVL, данные расписания), если таковые имеются. Третья категория – система смарт-карт, в которых не регистрируется ни время, ни место посадки. В таком исследовании могут быть объединены данные AFC, APC и GTFS. Обычно третья категория встречается редко.

В исследовании изложены существующие знания, касающиеся использования данных смарт-карт для оценки корреспонденций. Хотя большая часть процесса оценки считается общепринятой, существует необходимость представить этот процесс в более надежной и понятной форме для использования специалистами-практиками. Основные темы включают: очистку данных; оценку неизвестных параметров (место посадки/высадки и обнаружение пересадок); проверку предложенных алгоритмов и преобразование. Обязательной процедурой при работе с большими массивами данных является их очистка от несоответствий, вызванных инструментальными погрешностями и влиянием человеческого фактора.

ЛИТЕРАТУРА

1. **Hussain, E.** Transit OD matrix estimation using smartcard data: Recent developments and future research challenges / E. Hussain, A. Bhaskar, E. Chung // *Transportation Research Part C: Emerging Technologies*. – Volume 125. – 2021.