

Лебедева Ольга Анатольевна,
к.т.н., доцент, Ангарский государственный технический университет,
e-mail: kravhome@mail.ru,

Ерофеев Ермак Владимирович,
студент гр. ИТБ-23-1, Ангарский государственный технический университет,
e-mail: ermak.080@mail.ru

ОЦЕНКА МАТРИЦЫ КОРРЕСПОНДЕНЦИЙ С ИСПОЛЬЗОВАНИЕМ ДАННЫХ СМАРТ-КАРТ: ПРОЦЕДУРА ОЧИСТКИ ДАННЫХ

Lebedeva O.A., Erofeev E.V.

CORRESPONDENCE MATRIX ASSESSMENT USING SMART CARDDATA: DATA CLEANSING PROCEDURE

Аннотация. В общественном транспорте смарт-карты применяются для автоматического сбора платы за проезд, что генерирует огромные объемы данных. Основная цель этой статьи – критический анализ основных этапов процесса оценки, а, в частности, процедуры очистки данных.

Ключевые слова: общественный транспорт, смарт-карта, анализ данных, очистка данных.

Abstract. In public transportation, smart cards are used for automated fare collection, generating enormous amounts of data. The main objective of this article is to critically analyze the key stages of the evaluation process, particularly the data cleansing procedure.

Keywords: public transport, smart card, data analysis, data cleaning.

Типичная структура оценки матрицы корреспонденций может быть разделена на следующие части: очистка данных; поиск пересадок; оценка мест посадки и высадки; оценка пересадочности между зонами и остановками, и проверка предложенной методологии (рис. 1). Необходимость очистки зависит от сбоев в работе оборудования и человеческих ошибок. Для очистки необходима информация об источниках и типах ошибок. Сбои в работе оборудования могут быть вызваны неисправностью считывателя смарт-карт, рассинхронизацией часов в устройствах сбора данных, установленного GPS и/или системы в целом. Такие несоответствия в транзакциях могут составлять до двух процентов от общего числа [1]. Сбои приводят к следующим ошибкам в данных смарт-карт: отсутствие записи или совпадение времени и/или места посадки/высадки; более раннее время посадки для той же поездки; отсутствие идентификатора смарт-карты; дублирование события; транзакции на остановке, которую невозможно отследить. Тип ошибки и ее правильная интерпретация имеют решающее значение при оценке транспортного спроса. Неправильная интерпретация ошибки может привести к неточному значению матрицы корреспонденций.

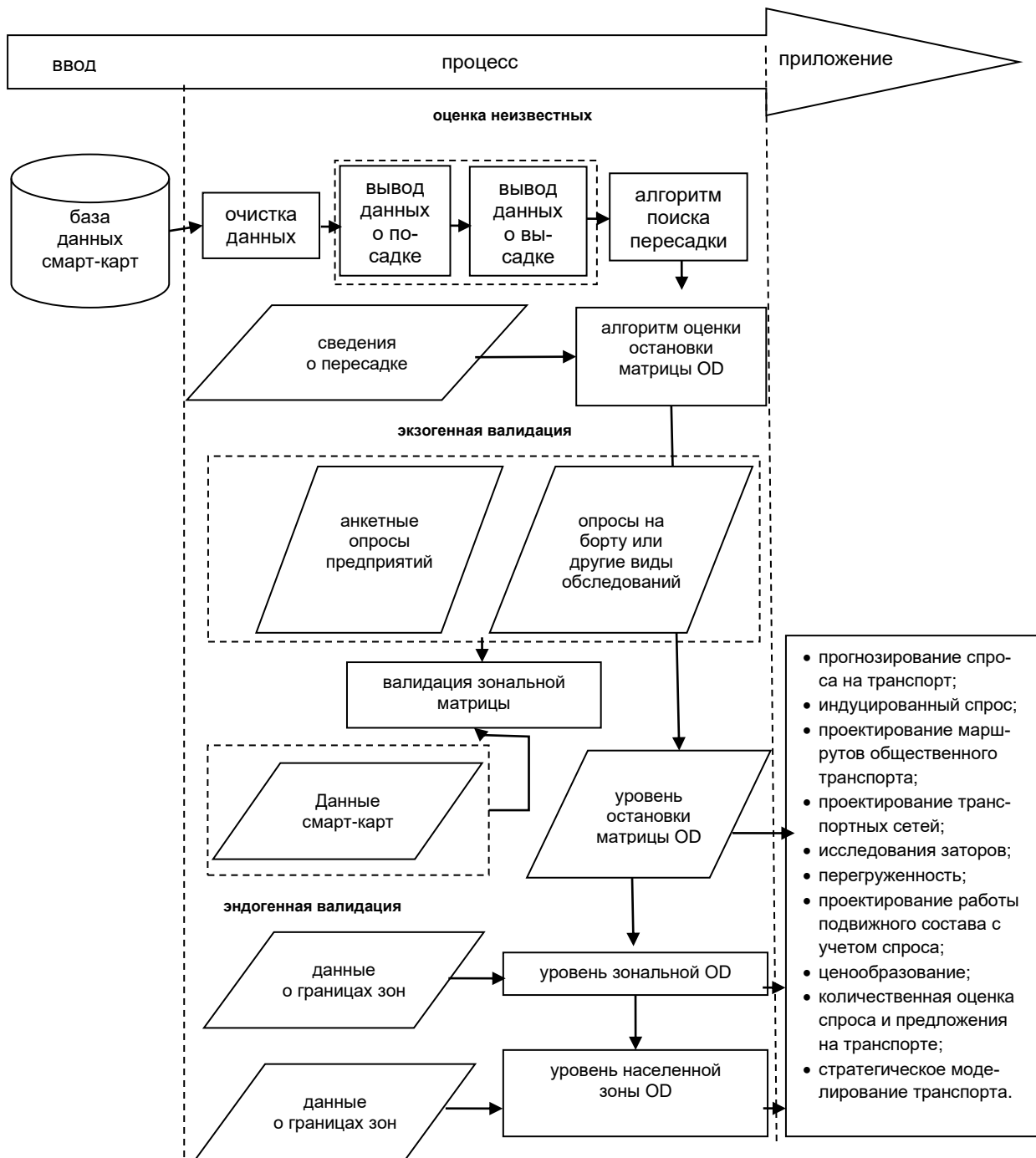


Рис. 1. Этапы решения задачи оценки транспортного спроса

Следовательно, данные должны быть проверены на все возможные несоответствия, которые необходимо устранить во время анализа, если таковые обнаружены, после этого – использовать для процедуры оценки.

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.