

УДК 338.47 DOI: 10.14451/1.240.128

Применение алгоритмов обработки и анализа больших данных в процессе реализации проекта «Цифровой след пассажира»*

© 2024 **Гулый Илья Михайлович**

Кандидат экономических наук, доцент. Петербургский государственный университет путей сообщения Императора Александра I, Санкт-Петербург.

E-mail: ilya.guliy@mail.ru

© 2024 **Корниенко Елизавета Николаевна**

Обучающийся бакалавриата. Петербургский государственный университет путей сообщения Императора Александра I, Санкт-Петербург.

E-mail: elizavetakornienko201003@gmail.com

© 2024 **Светлов Всеволод Дмитриевич**

Обучающийся бакалавриата. Петербургский государственный университет путей сообщения Императора Александра I, Санкт-Петербург.

E-mail: svetlov.seva@gmail.com

© 2024 **Задорожин Владимир Вячеславович**

Обучающийся бакалавриата. Петербургский государственный университет путей сообщения Императора Александра I, Санкт-Петербург.

E-mail: ilya.guliy@mail.ru

Ключевые слова: цифровой след пассажира, кластерный анализ, нейросетевой анализ.

Цель: обосновать применение методов кластерного анализа и построения нейронных сетей для реализации технологий анализа больших данных при реализации проекта «Цифровой след пассажира». Методы: систематизация статистических массивов данных, нейросетевой анализ, кластерный анализ. Результаты: на примере поездки пассажиров в купейном вагоне вечернего поезда по маршруту Санкт-Петербург – Москва выполнена группировка пассажиров по однородным параметрам и получена предсказательная функция на основе результатов обучения нейронной сети с последующим достижением наибольшей эффективности адресации дополнительных услуг. Сформулированы эффекты технологий больших данных, реализованных с применением сервиса «Цифровом следе пассажира».

* Настоящая статья (работа) опубликована (выполнена) при поддержке Федерального государственного бюджетного образовательного учреждения высшего образования «Петербургский государственный университет путей сообщения Императора Александра I» инициативных научных работ, выполняемых студенческими научными коллективами. (Р)

Введение

Актуальность исследования связана с тем, что сегодня каждая транспортная компания конкурирует на рынке за клиента-пассажира на основе предложения ценности транспортной услуги, которая связана с такими элементами, как гибкость, динамичное тарифообразование, цифровые сервисы, возможность построения и заказа смешанной перевозки, внедрения компонентов бизнес-модели «Мобильность как услуга».

На современном этапе происходят значительные изменения в потребительской модели пассажира:

- удобство и комфорт поездки для пассажира становится определяющим фактором;
- снижается уровень физической вовлеченности в процесс поездки (то есть потребление услуг в поезде через физические средства уменьшается, а через цифровые – растет);
- реализуется индивидуализация опыта конкретного пассажира;
- перераспределяются предпочтения клиентов в сторону онлайн-покупок и заказов;
- появляются новые сегменты пассажиров: миллениалы (молодежь-цифра), «цифровые кочевники» (блогеры, путешественники «езде», без привязки к конкретному месту) пожилые люди (в том числе маломобильные категории) и т.д. [7].

В электронных системах постоянно регистрируются данные об активности человека в виртуальном пространстве. Развитие методов сбора, очистки и структурирования данных предоставляет возможность быстрого поиска информации о клиенте-пассажире в разных источниках и ее объединение для комплексного анализа [1].

В электронных системах взаимодействия перевозчика с клиентами появляются сервисы и технологии, которые получили название «Цифровой след пассажира» [9].

Результаты исследований

Под «цифровым следом» мы понимаем уникальный набор действий в Интернете или на цифровых устройствах. Также встречаются определе-

ния цифрового отпечатка или цифровой тени, которые также употребляются для обозначения цифрового следа.

В стратегии цифровой трансформации холдинга «РЖД» до 2025 года и последующую перспективу проект цифровой след пассажира представлен в числе пула приоритетных проектов [2]. Цифровой след пассажира – это комплекс технологий по оперативному управлению качеством транспортных услуг, оказываемых пассажирам; это удобная система информирования пассажиров; а также это мультимодальные услуги пассажирской перевозки, а также различные дополнительные наборы опций сервисов и дополнительных услуг.

Реализация цифрового следа пассажира возможна, благодаря внедрению и широкому использованию в транспортных компаниях технологий обработки и аналитики больших данных. Большие данные и основанные на их использовании технологии формируют колоссальный потенциал для дальнейшего применения, их монетизации [3; 4; 10].

На рынке пассажирских перевозок возникает необходимость применения единой цифровой экосистемы, которая объединяет множество участников и будет функционировать в различных сегментах рынка как «бесшовная цифровая среда» с возможностью реализации собственных и партнерских сервисов разных компаний. Интеграция всех сервисов по предоставлению транспортных услуг в рамках единой цифровой экосистемы обеспечивает появление «бесшовного клиентского опыта», который подразумевает мгновенное переключение пассажира между различными сервисами, входящими в такую экосистему, через единый цифровой профиль [8].

Цифровой след пассажира способствует решению таких проблем развития пассажирских перевозок, как увеличение пассажиропотока, повышение заполняемости поездов, рост посещаемости цифровых платформ и мобильных приложений, увеличение конверсии рекламных кампаний.

Цифровой след, алгоритмизация обработки и анализа данных о пассажирах и их профилях позволяют реализовать концепцию новой ценности транспортной услуги, которая связана с гибкостью предложения, динамичным тарифообразованием, возможностями оформления онлайн различных опций дополнительных сервисов, заказа смешанной перевозки. В итоге происходят значительные изменения в потребительской модели пассажира. Внедрение цифрового следа пассажира позволяет решить ряд проблем: дальнейший прирост пассажиропотока, повышение заполняемости поездов, рост посещаемости цифровых платформ и мобильных приложений, увеличение конверсии рекламных кампаний [5; 6].

Покажем применение методов кластерного анализа и построения нейронных сетей в проекте «Цифровой след пассажира».

1. Кластерный анализ – метод, используемый для классификации объектов или событий в относительно однородные (имеющие множества сходств) группы-кластеры. Применение метода мы продемонстрировали на примере поездки пассажиров в купейном вагоне вечернего поезда по маршрут Санкт-Петербург – Москва. Данные представлены в разрезе показателей: пол, демография (дети), возраст, цель поездки, частота поездок в год, конечная стоимость приобретения билетов (билетная и внебилетная выручка с пассажира) – таблица 1.

Таблица 1. Исходные данные для формирования предсказательной функции нейросети.

№	Пол	Демография (число детей)	Возраст	Месячный доход, тыс. р.	Цель поездки	Частота поездок в год	Price (Cost)
1	1	1	42	140	1	19	7500
2	1	1	28	90	3	9	6200
3	0	3	51	185	1	21	9000
4	1	0	42	130	7	20	7000
5	1	1	32	85	5	8	7300
6	0	2	53	35	1	8	6400
7	1	1	55	150	1	16	9500
8	0	2	37	150	1	1	7500
9	1	0	42	85	2	24	2000
10	0	2	54	64	7	12	8500
11	0	0	17	32	3	10	6200
12	0	3	72	24	4	2	7000
13	1	1	37	45	1	14	7000
14	0	2	61	37	7	5	7000
15	1	0	30	140	7	15	2500
16	0	0	57	217	7	15	7000
17	1	1	74	30	4	6	4800
18	1	0	19	24	3	5	6200
19	0	1	76	27	8	6	7400
20	1	1	38	65	2	5	9000
21	0	0	31	45	1	7	6500

Продолжение на следующей странице

Таблица 1. Исходные данные для формирования предсказательной функции нейросети. (Продолжение таблицы)

№	Пол	Демография (число детей)	Возраст	Месячный доход, тыс. р.	Цель поездки	Частота поездов в год	Price (Cost)
22	1	1	88	21	4	2	5200
23	1	1	38	110	6	9	6200
24	0	2	57	130	1	12	9300
25	0	0	51	48	1	10	7000
26	1	0	18	22	8	6	7000
27	0	1	53	230	5	24	10500
28	0	1	57	82	7	4	7000
29	1	0	25	135	7	21	7000
30	1	0	43	38	5	10	8000
31	1	3	52	94	6	14	8000
32	0	1	35	145	2	10	8000

Источник: данные авторов, систематизированы из открытого источника.

Основой проведения анализа и разбиения на однородные кластеры является программа IBM SPSS Statistics.

В таблице 2 приведен результат кластеризации.

Сформированы 6 кластеров (табл. 2). Пассажиры кластеризованы в основном по возрасту, частоте поездок, конечной стоимости приобретения билета и дополнительных услуг. Наибольшее число пассажиров (19) приходится на кластер не склонных к приобретению дополнительных услуг и сервиса. Также 2 пассажира приобрели билет по акции, у них тоже не выявлена склонность к дополнительным тратам. Кластеры 2, 4 и 6 – это целевые области для адресации в мобильном приложении пассажира дополнительных предложений. Во втором и шестом кластере целесообразно направить весь спектр товаров в пути, туристические программы и варианты формирования мультимодальных маршрутов.

Таким образом, кластерный анализ позволит группировать пассажиров по однородным параметрам для наибольшей эффективности адресации дополнительных услуг.

2. Построение нейросетей и получения значений предсказательной функции был применен

алгоритм SPSS по обучению нейронной сети.

Благодаря свойству нейросети «связи всех параметров со всеми» при вложении в нее различных данных с использованием программы мы можем получить данные по адресации предложений для конкретного пассажира в моменты его будущих поездок

Нейросеть построена на примере данных о поездке в пассажирском купейном вагоне 32 человек. Основные параметры, которые структурированы в массивы данных: пол, демография (число детей), возраст, месячный доход, цель поездки, данные о полученном цифровом следе по прошлым потреблением дополнительных услуг, частота поездок в год. Путем ввода данных в SPSS, ввода параметров, уточнения настроек программы, получена функция предсказания потребления конкретного пассажира в зависимости от базового тарифа (рис. 1).

Получены данные по тем пассажирам, для которых возможны адресные предложения (№№: 3, 24, 8, 27, 32, 10, 16, 7, 12, 1, 6, 20, 22, 19). При этом построенная нейросеть дополнила кластерный анализ, рекомендовав пассажирам с №№ 1, 6, 8, 12, 16 и 19 направлять адресные предложения,

Таблица 2. Пример кластеризации.

№	Распределение пассажиров в вагоне, доли, %	Описания групп
1	59,40%	Пассажиры приобретают билет по базовому тарифу или с небольшой скидкой (невозвратный тариф), частота поездок 4–21 в год, пассажиры не склонны к приобретению дополнительных услуг вообще
2	12,50%	Пассажиры в возрасте от 38 до 57 лет, с детьми, высокий доход, склонны к приобретению питания, сувениров и принадлежностей в пути, услуг для детей
3	6,30%	Пассажиры в возрасте 32–40 лет, без детей, приобрели билет по программе «РЖД-Бонус» (билет бесплатно + платный набор стандартных услуг, питание), пассажиры не склонны к приобретению дополнительных услуг в поезде
4	12,50%	Пассажиры в возрасте 35–54 года, с детьми и без, цель поездки – отпуск и поездки к родным приобретают в вагоне только питание
5	6,30%	Пассажиры старшего возраста, приобрели билеты по скидке для инвалидов, цель поездки – лечение и отдых, приобрели завтрак (небольшой пакет питания)
6	3,10%	Пассажир среднего возраста, с самым высоким доходом, 1 ребенок, приобрел наибольший пакет услуг, сувениры и товары в пути (на сумму 2500 рублей)

Источник: составлено авторами.

в то время как кластерный анализ поместил этих пассажиров в сегмент без адресации. Благодаря различным надстройкам нейросети при вложении в нее различных параметров можно получить данные по адресации предложений для конкретного пассажира в моменты его будущих поездок.

В рамках исследования разработана программа на языке Python «Смысловая модель обработки и анализа больших данных, отражаемых «Цифровым следом» пассажиров, при формировании им адресных предложений в поездах на поездах дальнего следования». Программа предназначена для расчета потенциальной вероятности восприимчивости пассажира железнодорожного транспорта при поездках в дальнем сообщении к рассылке ему отдельных групп адресных предложений, позволяет оценить склонность пассажира к приобретению дополнительных услуг в поездке по категориям услуг: меню питания, товаров для детей, развлекательного контента для взрослых и для детей, сувенирной продукции, туристических программ, мультимодальных вариантов маршрутов на основе железнодорож-

ной поездки (рис. 2).

Заключение

По итогам создания нейросетей в Цифровом следе пассажира до 80–90% потенциальных запросов пассажира возможно генерировать для него заблаговременно путем адресации push-уведомлений.

Основными эффектами применения технологий больших данных, реализованными в «Цифровом следе пассажира», являются:

- обеспечение высокой конкурентоспособности предлагаемых сервисов и продуктов;
- генерирование доходности непрофильных видов деятельности (например, внебилетная выручка для перевозчиков);
- существенное расширение клиентских баз и снижение расходов для привлечения клиентов;
- увеличение рыночной капитализации транспортной компании, на базе которой реализованы сервисы предсказательной аналитики;
- развитие инструментов и механизмов, содействующих инновациям пассажирской компании.

№	S	F	A	R	P	TR	Dloc	Price	V9	V10	M:P_Predict_ edValue
1	1.0	1.0	42.0	140.0	1.0	2.0	19.0	7500.0	-	-	8159.5
2	1.0	1.0	28.0	90.0	3.0	0	9.0	6200.0	-	-	5289.3
3	0	3.0	51.0	195.0	1.0	2.0	21.0	9000.0	-	-	8970.3
4	1.0	0	42.0	130.0	7.0	2.0	20.0	7000.0	-	-	6652.1
5	1.0	1.0	32.0	85.0	5.0	1.0	8.0	7300.0	-	-	6133.2
6	0	2.0	53.0	35.0	1.0	0	8.0	6400.0	-	-	8171.2
7	1.0	1.0	55.0	150.0	1.0	2.0	16.0	9000.0	-	-	9212.0
8	0	2.0	37.0	150.0	1.0	2.0	1.0	7500.0	-	-	8546.3
9	1.0	0	42.0	85.0	2.0	1.0	24.0	2000.0	-	-	6202.9
10	0	2.0	54.0	64.0	7.0	2.0	12.0	8500.0	-	-	8663.9
11	0	0	17.0	32.0	3.0	0	10.0	6200.0	-	-	6520.4
12	0	3.0	72.0	24.0	4.0	0	2.0	7000.0	-	-	8183.6
13	1.0	1.0	37.0	45.0	1.0	1.0	14.0	7000.0	-	-	6854.5
14	0	2.0	61.0	37.0	7.0	0	5.0	7000.0	-	-	7300.3
15	1.0	0	30.0	140.0	7.0	1.0	15.0	2000.0	-	-	5307.1
16	0	0	57.0	217.0	7.0	2.0	15.0	7000.0	-	-	8448.2
17	1.0	1.0	74.0	30.0	4.0	1.0	6.0	4800.0	-	-	6526.2
18	1.0	0	19.0	24.0	3.0	0	5.0	6200.0	-	-	4765.3
19	0	1.0	76.0	27.0	8.0	1.0	6.0	7400.0	-	-	7703.8
20	1.0	1.0	38.0	65.0	2.0	2.0	5.0	9000.0	-	-	7896.2

Вывод функции предсказания M:P_Predict_ edValue

Рис. 1. Пример построения функции нейросети по данным цифрового следа. Источник: построено авторами.

Тело программы

```

import pandas as pd
from numpy import *
from sklearn.preprocessing import StandardScaler

def predict_ticket_price(x):
    scaler = StandardScaler()
    x_scaled = scaler.fit_transform(x)
    price = model.predict(x_scaled)
    return price

# Example usage
x = [1, 1, 42, 140, 1, 2, 19, 7500]
predicted_price = predict_ticket_price(x)
print(predicted_price)
    
```

Рис. 2. Интерфейс разработанной программы. Источник: разработано авторами.

Библиографический список

- Егоров Ю. В., Сакс Н. В. Особенности экономики цифровой железной дороги: отечественный и зарубежный опыт // Развитие экономической науки на транспорте: создание методологической основы для развития компетенций цифровизации транспортных систем : сборник научных статей VI международной научно-практической конференции, Санкт-Петербург, 26 апреля 2018 года. – Санкт-Петербург : ПГУПС, 2018. – С. 132–138.
- Об утверждении Транспортной стратегии Российской Федерации до 2030 года с прогнозом на период до 2035 года: распоряжение Правительства РФ от 27 ноября 2021 г. № 3363-р. – URL: https://www.consultant.ru/document/cons_doc_LAW_402052 (дата обр. 07.05.2024).
- Слепцова Ю. А., Качалов Р. М. Интеграционная стратегия предприятия в условиях цифровой трансформации экономики // Научно-технические ведомости СПбГПУ. Экономические науки. – 2018. – Т. 11, № 5. – С. 7–21.
- Цифровой след приведет к любому. 22 мая 2019. – URL: <https://www.rosbalt.ru/moscow/2019/05/22/1782565.html>.
- Чеченова Л. М. Повышение инвестиционной привлекательности транспортных организаций, осуществляющих интермодальные контейнерные перевозки в период пандемического кризиса // Транспортные системы и технологии. – 2020. – Т. 6, № 4. – С. 98–112.
- Чеченова Л. М., Егоров Ю. В. К вопросу об использовании экспертных оценок для анализа рисков, влияющих на эффективность реализации инвестиционного проекта // Финансовые аспекты структурных преобразований экономики. – 2019. – № 5. – С. 437–444.
- Bauer F., Dichter A., Rothkopf M. Travel and logistics: data drives the race for customers // Travel, Transport & Logistics. – 2018. – May. – P. 5–7.
- Identifying the impact of parking policy on road transport economics / M. Poliak [et al.] // Mobile Networks and Applications. – 2021.
- The Gartner Supply Chain Top 25. – URL: <https://www.gartner.com>.
- Tretyak V. P., Lyakina M. A., Volkova E. The ways of business digitalization in global corporations // SHS Web of Conferences : The 20th International Scientific Conference Globalization and its Socio-Economic Consequences. – 2021. – P. 0502–7.