



УДК: 004.89: 004.67: 004.422

МОДЕЛИ И АЛГОРИТМЫ, ПРИМЕНЯЕМЫЕ ДЛЯ ОПТИМИЗАЦИИ МАРШРУТОВ И ЦЕНООБРАЗОВАНИЯ ПАССАЖИРСКИХ ЖЕЛЕЗНОДОРОЖНЫХ ПЕРЕВОЗОК

А.А. Кацупеев, andreykatsupееv@gmail.com, **Э.О. Комбарова**, ms.ekom@mail.ru
Южно-Российский Государственный Политехнический Университет (НПИ) имени М.И. Платова,
г. Новочеркасск

В статье рассматриваются нейросетевые модели и алгоритмы, используемые для решения задачи оптимизации маршрутов и ценообразования пассажирских железнодорожных перевозок. Описаны данные, оказывающие влияние на принятие решений в этой сфере. Перечислены существующие подходы к решению подобных задач. Сформулированы методы и модели анализа данных, построены нейросетевые модели и алгоритмы, которые могут быть использованы при построении маршрутов как в железнодорожной сфере, так и в других областях пассажирских перевозок. Описан подход к динамическому ценообразованию, основанный на анализе данных. Описан программный продукт, реализующий предложенные модели.

MODELS AND ALGORITHMS APPLICABLE FOR OPTIMIZING OF ROUTES AND PRICING POLICY FOR PASSENGER RAILWAY TRANSPORT

A.A. Katsupееv, E.O. Kombarova
Platov South-Russian State Polytechnic University (NPI), Novocherkassk

The article discusses neural network models and algorithms used to solve the problem of optimizing of routes and pricing policy of passenger rail traffic. The data that influence on decision making in this area is described. The existing approaches to solving such problems are listed. Methods and models of data analysis are formulated, neural network models and algorithms are built, which can be used in optimizing of routes both in the railway sector and in other areas of passenger traffic. A dynamic pricing approach based on data analysis is described. A software product that implements the proposed models is described.

Анализ данных занимает сейчас важное место в деятельности различных компаний. В сфере пассажирских железнодорожных перевозок актуальными являются вопросы оптимизации маршрутов транспорта и ценовой политики. Для этого собирается информация о пассажирах железнодорожного транспорта, о билетах, приобретенных пассажирами, о самих поездах. Анализ проводится для визуализации статистики и последующей выдаче рекомендаций. Данные для анализа вносятся автоматически из расписания или при покупке пассажиром билета на поезд. Самим анализом в предлагаемом подходе занимается аналитик, который обрабатывает поступающие данные с помощью программы, реализующей статистические методы. На выходе эти данные представляются в виде графиков или диаграмм для удобства и наглядности, предоставляются руководству компании для принятия мер по улучшению качества пассажирских перевозок.

Для прогнозирования заполняемости могут использоваться линейная регрессия [1], при которой количество людей в поезде зависит от цены проезда, коэффициента сезонности и рейтинга, выставленного пассажирами, и нейросетевой подход [2].

Коэффициент сезонности принимает следующие значения:

- 2, если дата поездки приходится на июль, август, новогодние или майские праздники;
- 1, если дата поездки приходится на выходной день;



– 0, в остальных случаях.

Таким образом уравнение линейной регрессии принимает следующий вид:

$$y = ax_1 + bx_2 + dx_3 + c,$$

где y – количество людей в поезде, x_1 – цена проезда за километр, x_2 – коэффициент сезонности, x_3 – рейтинг поезда, a, b, d – коэффициенты регрессии, c – свободный член.

Для решения уравнения регрессии используется метод наименьших квадратов, который позволяет найти такие значения коэффициентов, что сумма квадратов отклонений будет минимальной. Для нахождения коэффициентов решается система нормальных уравнений:

$$\begin{cases} c + a \sum x_1 + b \sum x_2 + d \sum x_3 = \sum y \\ c \sum x_1 + a \sum x_1^2 + b \sum x_1 x_2 + d \sum x_1 x_3 = \sum x_1 y \\ c \sum x_2 + a \sum x_1 x_2 + b \sum x_2^2 + d \sum x_2 x_3 = \sum x_2 y \\ c \sum x_3 + a \sum x_1 x_3 + b \sum x_2 x_3 + d \sum x_3^2 = \sum x_3 y \end{cases}$$

Пассажиры поезда могут выставять рейтинг своим поездкам. Для определения того, как на рейтинг влияют какие-либо параметры, используется коэффициент корреляции Пирсона, который определяется как:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{j=1}^n (y_j - \bar{y})^2}}$$

где \bar{x} , \bar{y} – выборочные средние, определяющиеся следующим образом:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

где n – количество наблюдений, x – выборка рейтинга пользователей, y – выборка другого показателя (возраст пассажира, наличие кондиционера в вагоне и т.д.).

Структурная схема программной реализации проекта на клиенте представлена на рисунке 1.

На структурной схеме представлены язык разработки данного программного продукта, среда разработки, обеспечение для создания шаблонов форм, библиотеки, используемые для анализа данных, графического представления данных в виде графиков, диаграмм, гистограмм, подключения к базе данных, с хранящейся в ней информацией по железнодорожным пассажирским перевозкам.

Графический интерфейс разработан с помощью PyQt Designer, являющимся главным окном программы. В данном модуле расположены следующие функции:

- `retranslateUi` – функция, которая задает кодировку и название программы;
- `load_page` – описывает первую вкладку, на которой происходит загрузка данных из базы данных;
- `passengers_page` – описывает вкладку с пассажирами, содержит статистику в виде круговых диаграмм, несколько таблиц, где отображены все пассажиры поезда и всех их поездки с выставленным рейтингом, приведены результаты корреляционного анализа Пирсона;

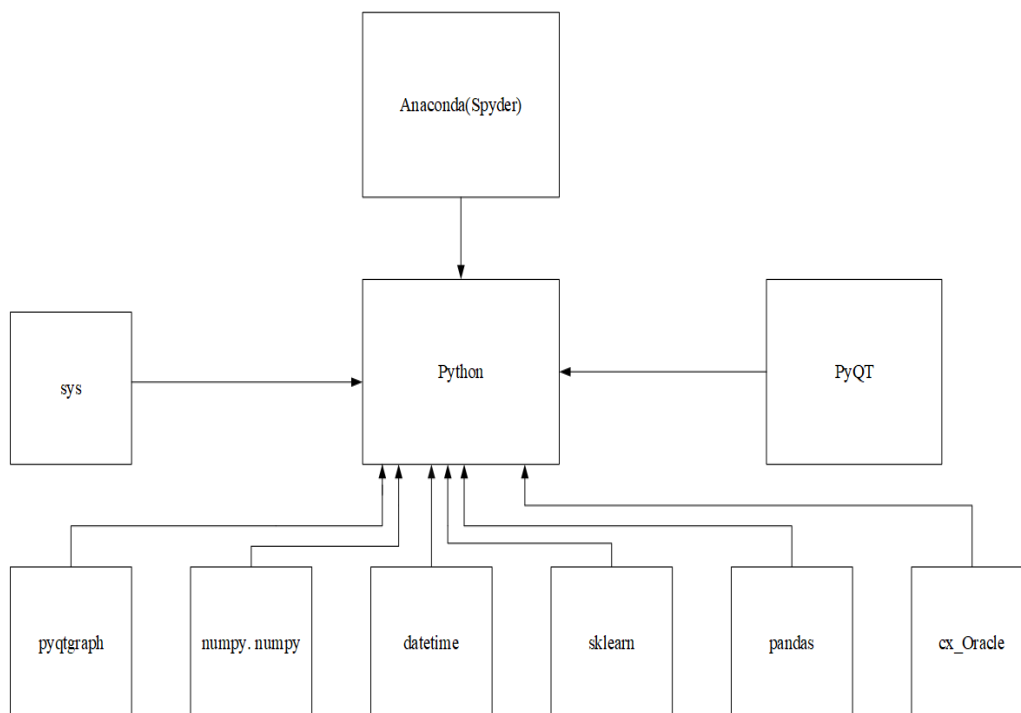


Рис. 1 – Структурная схема программной реализации

– `trains_page` – описывает вкладку с поездами, где можно выбрать один из поездов, посмотреть его расписание с соответствующим ему рейтингом за поездку, спрос на билеты во время года представлен на графике, также выводится список станций, по которым следует поезд;

– `stations_page` – описывает вкладку со станциями, где содержится таблица станций, по которым проходит данный поезд, и график, показывающий количество пассажиров, севших на поезд на какой-либо из станций;

– `analytics_page` – описывает вкладку с анализом данных, на которой производится анализ с помощью линейной регрессии, где можно составить прогноз на спрос билетов о той или иной цены на них, посмотреть, как на этот спрос будет влиять сезонность;

– `regression` – описывает формулу линейной регрессии для высчитывания спроса на билеты для прогнозирования результатов;

– `station` – данная функция выводит расписание поезда на вкладку с расписанием;

– `roezd` – данная функция выводит на вкладку с поездами все поезда и их расписание;

– `passenger` – данная функция выбирает из базы всех пассажиров, когда-либо пользовавшихся услугами данного поезда;

– `date_evaluation` – функция, определяющая по дате коэффициент сезонности для использования в формуле линейной регрессии.



При нейросетевом подходе на вход многослойного перцептрона подаются данные о рейтинге поездки, времени года, цене билета. Выходным вектором является количество людей, которые покупают билет на этот поезд. Сравнение нейросетевого подхода с линейной регрессией на тестовых данных показало, что нейросеть показывает лучший результат с меньшим среднеквадратическим отклонением, чем линейная регрессия.

В данной работе описаны данные, влияющие на пассажирские железнодорожные перевозки, приведён подход, позволяющий спрогнозировать количество людей, которые воспользуются поездом, описано программное обеспечение, реализующее нейросетевой подход и алгоритм линейной регрессии.

Список цитируемой литературы

1. И. Ю. Выгодчикова. Алгоритм оценки параметров линейной множественной модели регрессии по минимаксному критерию. – 2015. – №4.
2. Goodfellow, Y. Bengio, A. Courville. Deep Learning (Adaptive Computation and Machine Learning series). – 2016.

© А.А. Кацупеев, Э.О. Комбарова, 2019