

СИСТЕМА ПРОГНОЗИРОВАНИЯ ПАССАЖИРОПОТОКА НА МЕТРОПОЛИТЕНЕ

к.т.н. А.И. Левтеров, к.э.н. В.Г. Сословский, И.В. Скрипина
(представил д.т.н., проф. Е.В. Нагорный)

Для эффективного управления перевозочным процессом на метрополитене предложена система прогнозирования входящего потока пассажиров, использующая модели нейронных сетей и программы для их разработки.

Постановка проблемы. Выбор стратегии развития любого предприятия зависит от величины и закономерностей изменения спроса на его продукцию. На Харьковском метрополитене величина спроса на продукцию отслеживается по количеству входящих на каждой станции пассажиров, учет которых производится с помощью комплекса технических средств. Полученная информация в настоящее время используется преимущественно для контроля полноты сбора выручки от перевозки пассажиров, оценки потерь от перевозки льготного контингента и разработки планов по выручке для каждой линии метро. Получаемые сведения могут также служить основой для подготовки решений по организации перевозок. Табличный способ представления данных затрудняет поиск закономерностей в изменении пассажиропотока, поскольку используются простейшие приемы их выявления – визуализация с помощью графиков, сравнительный анализ. Проведенный нами анализ показал, что на основе существующего комплекса технических средств можно построить многофункциональную систему прогнозирования пассажиропотока. В статье рассмотрены вопросы выбора программных средств и методов прогнозирования для создаваемой системы.

Анализ существующих методов прогнозирования. Поскольку получаемая информация о пассажиропотоках является набором данных наблюдений за входящими пассажирами, фиксируемых ежедневно по каждой станции метро, то анализ ограничим методами прогнозирования на основе временных рядов, в которых уровни ряда удалены друг от друга на одинаковые промежутки времени, равные одному дню.

Прогнозы по времени принято классифицировать на краткосрочные и долгосрочные. В нашем случае речь идет о краткосрочных прогнозах.

Существуют две основные цели анализа таких временных рядов – это

определение природы ряда и прогнозирование [2, 3, 5, 6]. Обе эти цели требуют, чтобы модель ряда была идентифицирована и формально описана. По своей природе временной ряд может соответствовать различным моделям поведения данных. Принято выделять четыре основных типа моделей данных: горизонтальная, трендовая, сезонная и циклическая [6]. На основе этих типов можно разработать комбинированную модель временного ряда, например, стационарный ряд можно представить горизонтальной моделью (средним значением), дополненной циклической и сезонной моделями. Аналогичные дополнения могут иметь ряд с линейным или нелинейным трендом. К сожалению, не существует четких рекомендаций относительно того, каким образом эти дополнения должны быть объединены в обобщенной модели ряда – аддитивно или мультипликативно.

Для стационарных данных используются скользящие средние [2], простое экспоненциальное сглаживание [5] и методы авторегрессионного скользящего среднего [3]. Для рядов, имеющих тренд, метод линейного экспоненциального сглаживания Хольта, возрастающие кривые, экспоненциальные модели и методы авторегрессионных интегрированных скользящих средних. Для анализа рядов с сезонностью используют спектральный анализ, метод Census X-12, экспоненциальное сглаживание Винтера, многомерную регрессию и метод Бокса – Дженкинса [5, 6]. Для анализа циклических рядов рекомендуется использовать экономические индикаторы и эконометрические модели.

Анализ показал, что временные ряды о потоках ежедневно входящих пассажиров на каждой станции и линии метро, могут быть стационарными или трендовыми, но, как правило, содержат сезонность и неравномерность по дням недели, искажения периодов и амплитуды колебаний, вызываемые праздничными днями и их переносами на рабочие дни недели. Специфика работы с такими рядами в литературе не описана.

Цель статьи. Показать результаты анализа пассажиропотоков на Харьковском метрополитене, а также выбора методов и программных средств для создаваемой системы прогнозирования пассажиропотоков в разрезе линий и станций метро, которая позволит разрабатывать наилучшие модели временных рядов с тремя и более колебательными составляющими (циклическость, сезонность, неравномерность по дням недели, по часам суток и т.д.).

Анализ исходной информации. Сведения о входящем потоке пассажиров на Харьковском метрополитене получают с использованием технических средств, которые ведут подсчет входящих пассажиров. Для этого турникеты всех станций оборудованы фотоэлементами и устройствами считывания информации с многоцветных проездных билетов, приборами накопления сведе-

ний о количестве опущенных жетонов и предъявленных на считывание проездных билетов. Для фиксации количества пассажиров, имеющих льготы, служебные проходы оборудованы "вертушками".

Поскольку часть вошедших пассажиров совершает поездки по разным линиям метрополитена с пересадкой, то каждый зафиксированный проход "платного" или "льготного" пассажира нельзя приравнять к одной поездке.

На точность учета входящего пассажиропотока влияет погрешность приборов и ошибочные действия пассажиров (проход "платных" пассажиров через турникеты для льготного контингента). В ходе предварительного анализа первичной информации были обнаружены дни, когда из-за выхода приборов из строя по отдельным станциям отсутствовали сведения о количестве "льготных" пассажиров. По нашему мнению общая погрешность оценки количества входящих пассажиров составляет 5...7%.

Для прогнозирования пассажиропотоков использовались сведения из базы данных метрополитена о количестве ежедневно входящих на 26 станциях "платных" и "льготных" пассажиров за период с 1 января 2001 по 30 ноября 2003 года. Каждый год был представлен отдельным файлом с электронной таблицей, в которой для записи анализируемых данных использовалось шесть полей: шифр линии (1, 2, 3); дата; шифр станции (1, 2, ..., 26); название станции; всего платные (количество "платных" пассажиров); льготные (количество "льготных" пассажиров). Общее количество подлежащих анализу первичных наблюдений составило 55692.

Используемая метрополитеном систематизация данных неудобна для поиска закономерностей в рядах динамики. В связи с этим была произведена перегруппировка данных в разрезе линий метрополитена и разработаны три электронных таблицы. В каждой таблице в первом столбце был показан день наблюдения, а далее по станциям этой линии были представлены сведения о количестве входящих "платных" и "льготных" пассажиров. Такая организация данных позволяет подсчитать суммарные пассажиропотоки по линии в целом за день, за месяц, квартал и год, визуализировать данные наблюдений в виде параллельных временных рядов.

Представление данных в виде параллельных временных рядов способствовало выявлению закономерностей в изменениях пассажиропотоков. Например, на рис. 1 приведены данные о платном пассажиропотоке на линии "Метростроителей – Научная", свидетельствующие о наличии сезонности и тренда в изменении платного пассажиропотока.

В ходе предварительного анализа временных рядов было установлено, что "платные" и "льготные" пассажиропотоки на линиях и станциях метро изменяются по-разному и для их описания нужно разрабаты-

вать специфические модели временных рядов.

Выбор программных средств и модели прогнозирования. Возможности современных статистических пакетов неоднократно анализировались разными авторами [1]. Все пакеты реализуют стандартный подход к анализу временного ряда: визуализация, определение его типа (стационарный или нестационарный) и наличия в нем компонент (тренд, сезонность); разработка моделей компонент; построение обобщенной модели ряда и оценка ее качества. Наличие нескольких колебательных компонент в изучаемых временных рядах заставило нас обратиться к новому направлению статического моделирования, использующего идеи искусственных нейронных сетей.

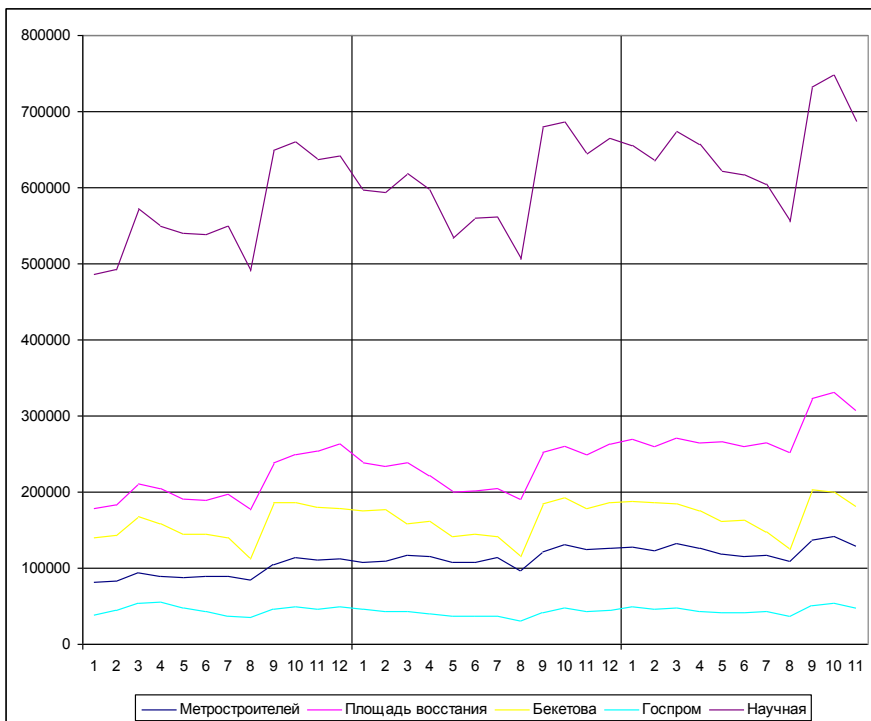


Рис. 1. Изменение входящего "платного" пассажиропотока по месяцам на станциях линии "Метроstroytelей – Научная"

Для разработки моделей временных рядов выбран пакет STATISTICA Neural Networks [4, 7]. Для каждой линии метро и станции с помощью этого пакета необходимо было найти наилучшую модель временного ряда. В качестве примера приведем результаты моделирования ежедневных наблюдений

ний за "платным" пассажиропотоком на станциях линии "Холодная гора – Пролетарская". Рассматривались разные виды моделей. Первой была апробирована модель двухслойного персептрона. Она имела неплохую среднюю относительную ошибку предсказания, равную 7,27%, но недостаточно хорошо описывала поведение пассажиропотока в выходные и праздничные дни, "спотыкалась" в оценке дней пассажиропотока, включаемых в так называемый праздничный цикл. Так, относительные ошибки прогноза в эти дни достигали до 20% , а 1 января, 1 мая и другие праздничные дни ошибка достигала 80 и более процентов. При этом ошибка прогноза пассажиропотока в дни недели, не содержащие праздничных дней и переносов дней отдыха, не превышала 5%.

В связи с этими особенностями временного ряда была разработана модель нейронной сети по укороченному временному ряду, из которого были исключены недели, содержащие праздничные циклы. В результате этого средняя относительная ошибка предсказания данных уменьшилась до 4,8%, а разброс значений ошибок лежал в интервале от 0 до 15%. Этот результат позволил сделать вывод о целесообразности разработки специальной методики корректировки результатов для праздничных дней.

Для поиска лучшей нейронной модели мы увеличивали число типов моделей, число слоев, пересматривали соотношение между количеством данных в обучающей, тестирующей и контрольной выборках. В результате работы генетического алгоритма лучшей оказалась нейросетевая модель четырехслойного персептрона (рис. 2) со средней относительной ошибкой 2,7%. На рис. 3 приведена гистограмма относительных отклонений предсказанных данных от отчетных. Из 1064 наблюдений в диапазон от -15 до +15% попадает 1000 наблюдений, и 64 наблюдения выходят за его границы. Анализ расхождений показал, что наибольшие расхождения между отчетными и предсказанными по модели значениями наблюдаются в дни праздничные, предпраздничные и послепраздничные (от 15 до 73%). Ошибка предсказания от 5 до 10% отмечена для 121 дня. По остальным 908 дням ошибка была менее 5%.

Процесс предсказания отдельных наблюдений представлен графически (рис. 4). На рис. 4 видно, что модель неплохо "уловила" характер колебаний пассажиропотока по дням. Некоторые расхождения между отчетными и предсказанными данными вызваны влиянием фактора сезонной неравномерности по месяцам года. Влияние фактора сезонности хорошо видно при увеличении длины ряда в окне просмотра результатов прогнозирования (рис. 5).

Результаты прогнозирования входящего "платного" пассажиропотока на рассматриваемой линии метрополитена на первые две недели де-

кабря 2003 года приведены на рис. 6.

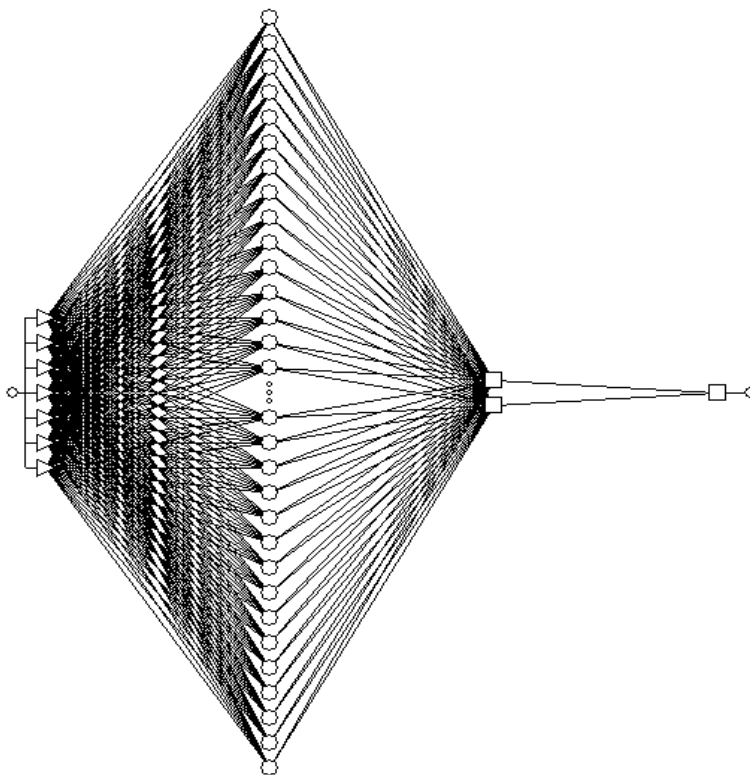


Рис. 2. Модель четырехслойной нейронной сети, описывающая временной ряд ежедневного входящего платного пассажиропотока на линии метро "Холодная гора – Пролетарская"

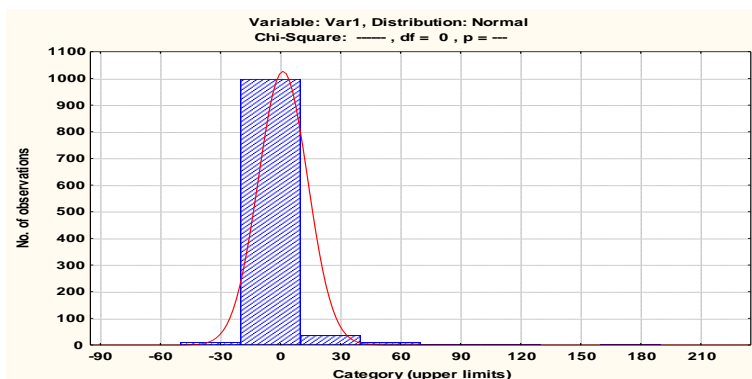


Рис. 3. Гистограмма отклонений предсказанных значений

пассажиropотока от отчетных

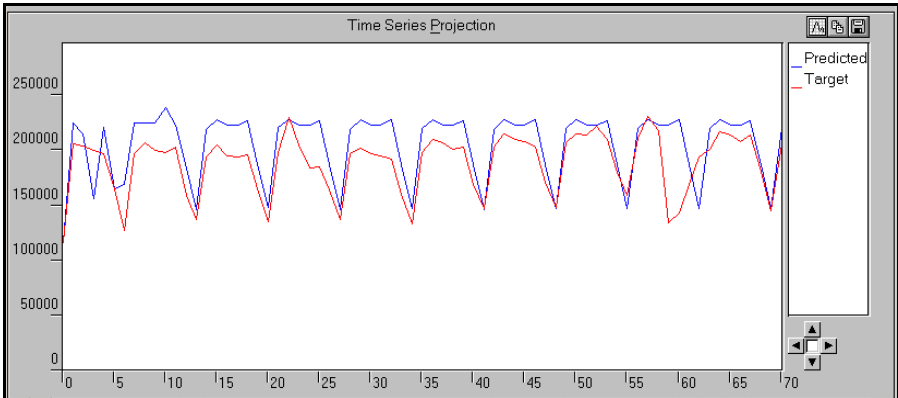


Рис. 4. Характер колебаний пассажиропотока по дням недели предсказанных (Predicted) и отчетных (Target)

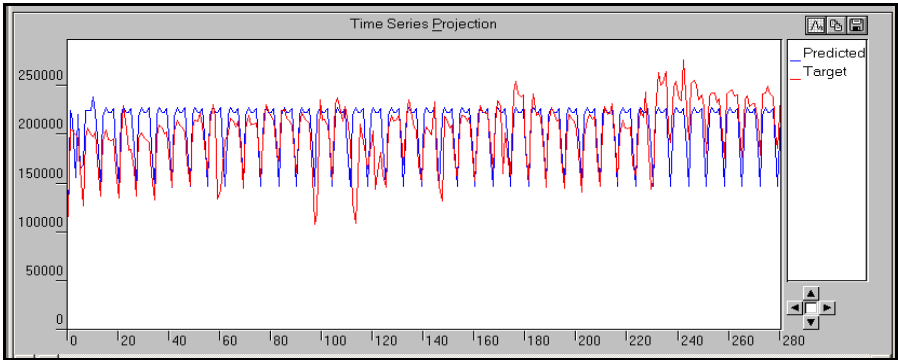


Рис. 5. Сезонная волна, характерная для изменения пассажиропотока по месяцам года

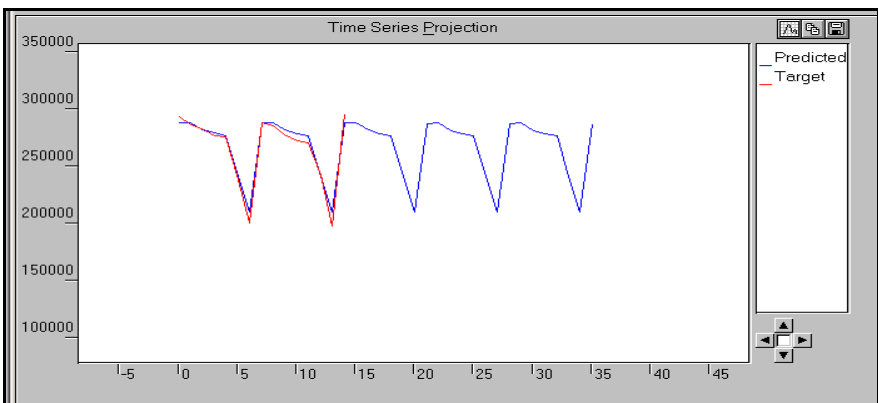


Рис. 6. Прогноз пассажиропотока на первые две недели декабря 2003 года

Поскольку прогноз на декабрь был дан до получения фактических данных, то было большим успехом узнать о том, что прогноз сбился, и средняя относительная ошибка предсказания данных была менее 1%. Дальнейшее улучшение модели возможно разными путями, например, можно попробовать разделить ряд на типичные и нетипичные рабочие недели, разложить уровни ряда на компоненты и моделировать каждую компоненту в отдельности с последующим обобщением полученных вариантов. Возможны и другие подходы. Выявление причин, повлекших за собой резкие отклонения пассажиропотока в некоторые дни от общих закономерностей, позволит разработать меры по улучшению предсказывающих свойств модели.

Выводы. Результаты исследований показали о необходимости совершенствования системы сбора информации. Использование предложенных моделей возможно при решении проблем учета количества пассажиров в неделях, в которых имеются праздничные дни. Для более эффективного прогноза кроме закономерностей временных рядов необходимо учитывать также различные факторы, связанные с развитием наземного транспорта, перспективами самого метрополитена, в дальнейшем преобразовании транспортной отрасли в целом, что указывает о целесообразности построения многофакторной модели.

ЛИТЕРАТУРА

1. Айвазян С.А., Степанов В.С. Программное обеспечение по статистическому анализу данных: методология сравнительного анализа и выборочный обзор рынка. – <http://ts1.cemi.rssi.ru/rus/publicat/e-pubs/ep97001t.htm>.
2. Андерсон Т. Статистический анализ временных рядов. – М.: Мир, 1976. – 426 с.
3. Бокс Дж., Дженкинс Г. Анализ временных рядов. – М.: Мир, 1974. – Вып. 1. – 406 с. – Вып. 2. – 196 с.
4. Боровиков В.В. STATISTICA: искусство анализа данных на компьютере. Для профессионалов. – СПб.: Питер, 2001. – 656 с.
5. Кендал М. Временные ряды. – М.: Финансы и статистика, 1981. – 199 с.
6. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования. – М.: Статистика, 1979. – 254 с.
7. Нейронные сети. STATISTICA Neural Networks: Пер. с англ. – М.: Горячая линия, Телеком, 2000. – 182 с.

Поступила 4.02.2004

ЛЕВТЕРОВ Андрей Иванович, канд. техн. наук., доцент, зав. каф. информатики ХНАДУ. Область научных интересов – математическое моделирование, информационные технологии.

СОСЛОВСКИЙ Владимир Георгиевич, канд. экон. наук, доцент каф. транспортных технологий ХНАДУ. Область научных интересов – статистическое моделирование и оптимизация производственно экономических систем.

СКРИПИНА Ирина Валентиновна, ст. преп. кафедры информатики ХНАДУ. Об-

ласть научных интересов – математическое моделирование, информационные технологии.