

АНАЛИЗ ВРЕМЕННЫХ РЯДОВ ДЛЯ ПРОГНОЗИРОВАНИЯ ДОСТУПНОСТИ ЗАРЯДНЫХ СТАНЦИЙ

Голубева И.А.

Кафедра информационных технологий автоматизированных систем, Белорусский государственный университет информатики и радиоэлектроники

Объединённый институт проблем информатики Национальной академии наук Беларуси

Минск, Республика Беларусь

E-mail: holubeva.irina@gmail.by

Рассматривается анализ временных рядов для предсказания доступности зарядных станций для электромобилей по настоящим и прошлым значениям таких факторов, как спрос на зарядные станции в зависимости от времени, дня недели, месяца, погоды в определенный промежуток времени. Построенную модель временного ряда можно обучать и тестировать на данных о зарядных станциях из открытых источников для исследования влияния параметров на точность прогнозов. Однако каждую конкретную сеть зарядных станций необходимо обучать на данных, учитывающих ее локальные особенности.

Существуют две основные цели анализа временных рядов – это определение природы ряда и прогнозирование (предсказание будущих значений временного ряда по настоящим и прошлым значениям). Для достижения этих целей требуется построение математической модели временного ряда.

После определения модели ряда становится возможным с ее помощью интерпретировать рассматриваемые данные (использовать для понимания временных или сезонных изменений популярности тех или иных зарядных станций). Затем, на основе построенной модели, можно экстраполировать ряд, то есть предсказать его будущие значения.

Часто существует дополнительная информация, которая помогает в прогнозировании. Данные, известные на некоторый промежуток времени в будущем, называются регрессорами. Это дополнительные временные ряды, значения которых известны не только в прошлом. Такие данные можно использовать для извлечения большей информации из процесса и улучшения прогноза. Так, календарь праздников поможет предсказать рост продаж шампанского перед Новым годом и тюльпанов перед 8 Марта, а вместе с этим накануне праздников вырастет трафик движения и, как следствие, спрос на зарядные станции. Использование регрессоров позволит улучшить модель и сможет предсказывать такое поведение. Как и большинство других видов анализа, анализ временных рядов предполагает, что данные содержат закономерную составляющую и случайный шум, затрудняющий обнаружение регулярных компонент. Большинство методов исследования временных рядов включает различные способы фильтрации шума, позволяющие увидеть регулярную составляющую более отчетливо. Например, на качество прогноза не должен повлиять единичный случай ро-

ста спроса на зарядные станции в районе X из-за сбоя работы зарядных станций в районе Y.

Большинство регулярных составляющих временных рядов являются либо трендом, либо сезонной составляющей. Тренд представляет собой общую систематическую линейную или нелинейную компоненту, которая может изменяться во времени. Сезонная составляющая – это периодически повторяющаяся компонента. Оба эти вида регулярных компонент часто присутствуют в ряде одновременно. Например, в часы пик, выходные и праздники, а также летом спрос на зарядные станции может возрасти, так как большее количество людей пользуется автомобилями, но также спрос на зарядные станции будет возрасти и каждый год, так как будет увеличиваться количество владельцев электромобилей. Эту общую модель можно понять на ряде, представляющем месячные международные авиаперевозки в течение 12-ти лет с 1949-го по 1960-й годы. График месячных перевозок показывает почти линейный тренд, имеется устойчивый рост перевозок из года в год. В то же время характер месячных перевозок повторяется, они имеют почти один и тот же характер в каждом годовом периоде (например, перевозок больше в отпускные периоды, чем в другие месяцы). Этот пример показывает довольно определенный тип модели временного ряда, в которой амплитуда сезонных изменений увеличивается вместе с трендом, – модель с мультипликативной сезонностью (см. рис. 1).

Если тренд является монотонным (устойчиво возрастает или устойчиво убывает), то анализировать такой ряд нетрудно. Если временные ряды содержат значительную ошибку, то первым шагом выделения тренда является сглаживание. Сглаживание всегда включает некоторый способ локального усреднения данных, при котором несистематические компоненты взаимно погашают друг друга. Таким образом, сглажи-

вание позволяет отфильтровать шумы и преобразовать данные в относительно гладкую кривую, используя методы скользящего среднего, медианного сглаживания или метод наименьших квадратов. Ряды с относительно небольшим количеством наблюдений и систематическим расположением точек могут быть сглажены с помощью бикубических сплайнов.

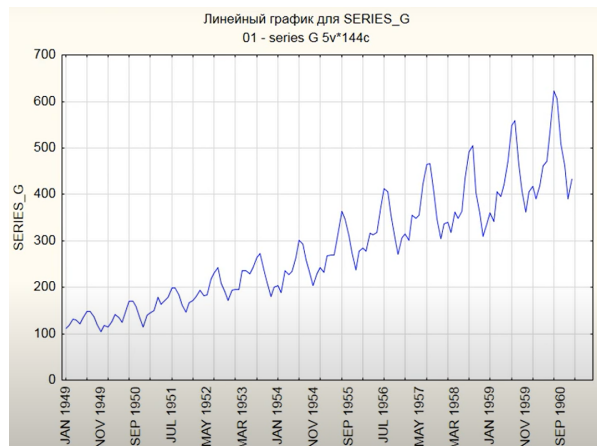


Рис. 1 – График, демонстрирующий ежегодный рост количества авиаперевозок

Многие монотонные временные ряды хорошо описываются линейной функцией. Если же имеется явная монотонная нелинейная компонента, то данные вначале следует преобразовать, чтобы устранить нелинейность. Обычно для этого используют логарифмическое, экспоненциальное или полиномиальное преобразование данных. Периодическая и сезонная зависимости представляют собой другой общий тип компонент временного ряда. Это понятие было проиллюстрировано ранее на примере авиаперевозок пассажиров. Можно заметить, что каждое наблюдение очень похоже на соседнее; имеется повторяющаяся сезонная составляющая, означающая, что каждое наблюдение похоже на наблюдение в том же самом месяце, но год тому назад. Периодическая зависимость может быть формально определена как корреляционная зависимость порядка k между каждым i -м элементом ряда и $(i-k)$ -м элементом, где k обычно называют лагом. Ее можно измерить с помощью автокорреляции, т.е. корреляции между самими членами временного ряда. Если ошибка измерения не слишком большая, то сезонность мож-

но определить визуально, рассматривая поведение членов ряда через каждые k временных единиц. Сезонные составляющие временного ряда могут быть найдены с помощью коррелограммы. Коррелограмма (автокоррелограмма) показывает численно и графически автокорреляционную функцию (АКФ), иными словами, коэффициенты автокорреляции (и их стандартные ошибки) для последовательности лагов. Корреляционный анализ произвольного временного ряда позволяет сделать следующие выводы:

1. Если ряд содержит тренд, то коэффициент автокорреляции значителен для первых нескольких сдвигов ряда, а в дальнейшем – убывает до нуля.
2. Если действие случайной компоненты велико, то коэффициенты автокорреляции для любого значения сдвига будут близки к нулю.
3. Если ряд содержит сезонную компоненту, то коэффициент автокорреляции будет большим для значений сдвига, равных периоду сезонной составляющей или кратных ему.

Таким образом, корреляционный анализ позволяет выявить во временном ряду тренд и сезонную компоненту, а также определить, насколько поведение ряда обусловлено его случайной компонентой. Знание данных свойств временного ряда помогает строить более адекватные модели и выбирать методы прогнозирования. Использование реальных данных о спросе на зарядные станции в зависимости от времени, дня недели, месяца, погоды и т.д.; а также регрессоров и методов фильтрации шума делает возможным прогнозировать спрос на зарядные станции, а значит, и предсказывать их доступность.

СПИСОК ЛИТЕРАТУРЫ

1. Time series analysis: „Forecasting and control” / Box, G. E. and Jenkins, G.M./ Holden-Day, San Francisco., 1976. – 500 p.
2. The Advanced Theory of Statistics Volume 3: Design and analysis, and Time-Series / Kendall, M.G. and Stuart, A. / Charles Griffin and Co. Ltd., London., 1976. – 557 p.
3. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. — 3-е изд., перераб. и доп. — СПб.: ВХВ-Петербург, 2009. — 512 с.