

УДК 004.891.2

Гайдукова Кристина Михайловна, аспирантка,
Алтайский государственный технический университет
им. И.И. Ползунова, г. Барнаул

**МОДЕЛЬ ВРЕМЕННЫХ РЯДОВ
ДЛЯ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ДАННЫХ
В АВТОМАТИЗИРОВАННЫХ СИСТЕМАХ ПРИНЯТИЯ РЕШЕНИЙ
A TIME SERIES MODEL FOR DATA ANALYSIS
AND FORECASTING IN AUTOMATED DECISION-MAKING SYSTEMS**

Аннотация: Данная статья представляет собой обзор российских и зарубежных научных работ, и публикаций, связанных с применением методов анализа и прогнозирования временных рядов. Рассмотрены материалы за последние 10 лет, расположенные на Интернет-ресурсах. Приведены примеры применения таких методов как регрессионный анализ, экспоненциальное сглаживание, а также метод ARIMA. Новое направление Time Series Data Mining для решения проблемы анализа ВР в настоящее время активно развивается.

Abstract: This article is a review of Russian and foreign scientific papers and publications related to the use of time series analysis and forecasting methods. The materials for the last 10 years, located on Internet resources, are considered. Examples of the application of such methods as regression analysis, exponential smoothing, as well as the ARIMA method are given. A new direction of Time Series Data Mining to solve the problem of BP analysis is currently actively developing.

Ключевые слова: временной ряд, анализ и прогнозирование, автоматизированные системы принятия решений, гибридные модели, ARIMA (SARIMA).

Keywords: time series, analysis and forecasting, automated decision-making systems, hybrid models, ARIMA (SARIMA).

В век информационных технологий анализ и прогнозирование временных рядов является важным вопросом во многих научных исследованиях. Основной целью анализа и прогнозирования временных рядов является получение важной информации, извлечение значимой статистики и других данных [3].

Целью прогнозирования временных рядов является возможность предсказать будущее значение или классификацию в определенный момент времени.

Экспериментальное обоснование постановки задачи

Ряд точек данных, проиндексированных во временном порядке, представляют собой временной ряд. Временной порядок может быть ежедневным, еженедельным, ежемесячным или годовым. Данные временных рядов могут отображать множество шаблонов, поэтому необходимо разделить временной ряд на компоненты, каждый из которых представляет основную категорию шаблонов.

Одна из самых востребованных и популярных видов декомпозиции – классическая декомпозиция. Существует два типа классической декомпозиции: декомпозиция на основе скорости изменения и декомпозиция на основе предсказуемости.

При разложении временного ряда на компоненты, ряд состоит из трех компонентов: компонента тренда, сезонного компонента и остатков или «шума» (всего, что находится вне тренда или сезонности во временном ряду) [3].

Регрессия временных рядов и экспоненциальное сглаживание

Один из наиболее распространенных методов анализа и прогнозирования временных рядов являются регрессионные модели. Они описывают математическую связь между



прогнозируемой переменной и одной переменной-предиктором. Самая известная регрессионная модель – это линейная модель. Несмотря на популярность линейных моделей, также популярны модели нелинейной регрессии, они предназначены для понимания более сложных методов прогнозирования и анализа временных рядов.

Общие типы экспоненциального сглаживания включают одинарное экспоненциальное сглаживание (SES), которое представляет собой средневзвешенные значения самого временного ряда. Двойное экспоненциальное сглаживание (DES) представляют собой средневзвешенные значения как тенденции, так и временного ряда. И тройное экспоненциальное сглаживание (TES, также известное как метод Холта-Винтерса) представляют собой средневзвешенные значения сезонности, тренда и временных рядов.

Модели ARIMA

Авторегрессионная интегрированная скользящая средняя (ARIMA) – это модель прогнозирования временных рядов, которая включает меры автокорреляции для моделирования временных структур в данных временных рядов для прогнозирования будущих значений [1].

Поскольку линейная независимость и отсутствие коллинеарности являются одним из фундаментальных допущений линейной регрессионной модели, поэтому для того, чтобы сделать шаблоны данных стационарными или свести к минимуму очевидную корреляцию с прошлыми данными необходимо эти различия измерить и интегрировать [2].

Модель ARIMA делится на два типа: несезонные модели ARIMA и сезонные модели ARIMA (SARIMA). Для того, чтобы определить ARIMA и SARIMA, необходимо сначала определить авторегрессию.

Авторегрессия – это модель временных рядов, которая использует наблюдения из предыдущих временных шагов в качестве входных данных в уравнение регрессии для прогнозирования значения на следующем временном шаге [4].

Модели ARIMA объединяют два подхода. Одно из основных предположений модели ARIMA состоит в том, что временной ряд является стационарным. Стационарный временной ряд – это временной ряд, компоненты которого не зависят от того, когда этот временной ряд наблюдается (то есть, временной ряд не показывает тенденции или сезонности). Поскольку модели ARIMA требуют, чтобы временные ряды были стационарными, дифференцирование может быть необходимым этапом предварительной обработки перед использованием модели ARIMA для прогнозирования. Модель SARIMA расширяет ARIMA, добавляя линейную комбинацию прошлых сезонных значений и / или ошибок прогноза.

Прогнозирование временных рядов с помощью ARIMA в Python

Для отслеживания сезонности используется сезонная модель ARIMA – ARIMA (p,d,q) (P,D,Q)s.

Здесь (p, d, q) – несезонные параметры, описанные выше, а (P, D, Q) следуют тем же определениям, но применяются к сезонной составляющей временного ряда. Параметр s определяет периодичность временного ряда (4 – квартальные периоды, 12 – годовые периоды и т.д.).

Подбор параметров для модели ARIMA

Главное при подборе данных временных рядов в сезонной модели ARIMA – найти значения ARIMA (p, d, q) (P, D, Q) s, которые оптимизируют требуемый показатель. Функция SARIMAX () из модуля statsmodels для каждой комбинации параметров может подобрать новую сезонную модель ARIMA и оценить ее общее качество. Оптимальным набором параметров будет тот, в котором нужные критерии наиболее производительны. Для того, чтобы сгенерировать различные параметры необходимо:



```
# Определить p, d и q в диапазоне 0-2
p = d = q = range (0, 2)
# Сгенерировать различные комбинации p, q и q
pdq = list (itertools.product (p, d, q))
# Сгенерировать комбинации сезонных параметров p, q и q
seasonal_pdq = [(x [0], x [1], x [2], 12) for x in list (itertools.product (p, d, q))] print
('Examples of parameter combinations for Seasonal ARIMA...')
print ('SARIMAX: {} x {}'.format (pdq [1], seasonal_pdq [1]))
print ('SARIMAX: {} x {}'.format (pdq [1], seasonal_pdq [2]))
print ('SARIMAX: {} x {}'.format (pdq [2], seasonal_pdq [3]))
print ('SARIMAX: {} x {}'.format (pdq [2], seasonal_pdq [4]))
Examples of parameter combinations for Seasonal ARIMA...
SARIMAX: (0, 0, 1) x (0, 0, 1, 12)
SARIMAX: (0, 0, 1) x (0, 1, 0, 12)
SARIMAX: (0, 1, 0) x (0, 1, 1, 12)
SARIMAX: (0, 1, 0) x (1, 0, 0, 12)
```

Для автоматизации процесса оценки моделей ARIMA по различным комбинациям (сетчатый поиск, или оптимизация гиперпараметров) нужно использовать определенные выше триплеты параметров.

Прогнозирование временных рядов

Для прогнозирования данных, а также понимания точности прогнозов, необходимо сравнить прогнозируемые значения с реальными значениями временного ряда.

При обычном анализе данных порядок собираемых наблюдений обычно не имеет значения. Большая часть данных, собранных в реальном мире, имеет определенный встроенный порядок. Эта структура может быть использована для лучшего прогнозирования с помощью модели временных рядов. Первым шагом в анализе модели временных рядов является определение и понимание встроенной структуры данных с течением времени. Эти базовые закономерности можно разделить на четыре категории:

Trend – Долгосрочный паттерн в данных.

Сезонность – краткосрочные паттерны, которые возникают в рамках общей временной структуры и повторяются бесконечно.

Циклическая составляющая – долгосрочные колебания (большие волны) в рамках общей временной структуры.

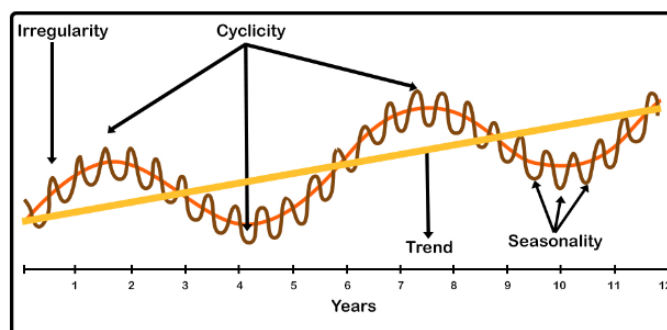


Рис. 1. Компоненты временных рядов

Вторым шагом является проверка исходных предположений:

Стационарность – это означает, что ряды нормально распределены, а среднее значение и дисперсия постоянны в течение длительного периода времени.



Некоррелированная случайная ошибка – среднее значение и дисперсия постоянны в течение определенного периода времени.

Заключение

Таким образом, изучив научные работы, и публикации за последние 10 лет, был сделан вывод, о том, что для прогнозирования и анализа временных рядов можно применять различные методы.

Самым удобным методом исследования временных рядов является интегрированная модель и методология анализа временных рядов ARIMA.

Чтобы привести данные временного ряда к стационарному виду и выявить все темпоральные зависимости необходимо применять данный способ с вариантом SARIMA.

Список литературы:

1. Антипов С. Г. Методы диагностики динамических объектов на основе анализа временных рядов [Текст] / С. Г. Антипов, В. Н. Вагин, М. В. Фомина // ИТНОУ: информационные технологии в науке, образовании и управлении, 2017. – №2 (2). – С. 3-12.
2. Ярушкина, Н. Г. Интеллектуальный анализ временных рядов [Текст]: учеб. пособие / Н. Г. Ярушкина, Т. В. Афанасьева, И. Г. Перфильева. – Ульяновск: УлГТУ, 2010. – 320 с.
3. Староверова К. Ю. Мера различия временных рядов, основанная на их характеристиках [Текст] / К. Ю. Староверова, В. М. Буре // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления, 2017. – Т. 13. Вып. 1. – С. 51–60.
4. Нечеткие гибридные системы. Теория и практика [Текст] / И.З. Батыршин [и др.]; под ред. Н.Г. Ярушкиной. – М.: Физматлит, 2007. – 208 с.

