электронный журнал
МОЛОДЕЖНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
Издатель Общероссийская общественная организация "Академия инженерных наук им. А.М. Прохорова"
ISSN 2307-0609

 # 12, декабрь 2017

УДК 004.021

Обзор современных подходов к прогнозированию временных рядов

Семиохин С. И., *магистр*
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Компьютерные системы и сети»
drstep321@mail.ru

Научный руководитель: Самарев Р.С., к.т.н., доцент
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Компьютерные системы и сети»
samarev@acm.org

Аннотация: В данной статье рассмотрены наиболее популярные и эффективные подходы к прогнозированию временных рядов. Приведен общий итеративный подход к построению математической модели прогнозирования. Рассмотрены различные прогнозные модели. Проведена оценка достоинств и недостатков различных типов моделей. Сформулированы выводы относительно наиболее многообещающих подходов.

Ключевые слова: временные ряды (time series), прогноз (forecast), линейная регрессия (linear regression), экспоненциальное сглаживание (exponential smoothing), авторегрессия (autoregressive), скользящее среднее (moving average), модель Хольта-Винтерса (Holts-Winters model).

Введение

Анализ временных рядов — совокупность математико-статистических методов анализа, предназначенных для выявления структуры временных рядов и для их прогнозирования. Сюда относятся, в частности, методы регрессионного анализа. Выявление структуры временного ряда необходимо для того, чтобы построить математическую модель того явления, которое является источником анализируемого временного ряда. Прогноз будущих значений временного ряда используется для эффективного принятия решений. Прогнозирования временных рядов имеет широкое применение: от прогноза температуры (рис. 1) до предсказания значения биржевых котировок и валютных пар.

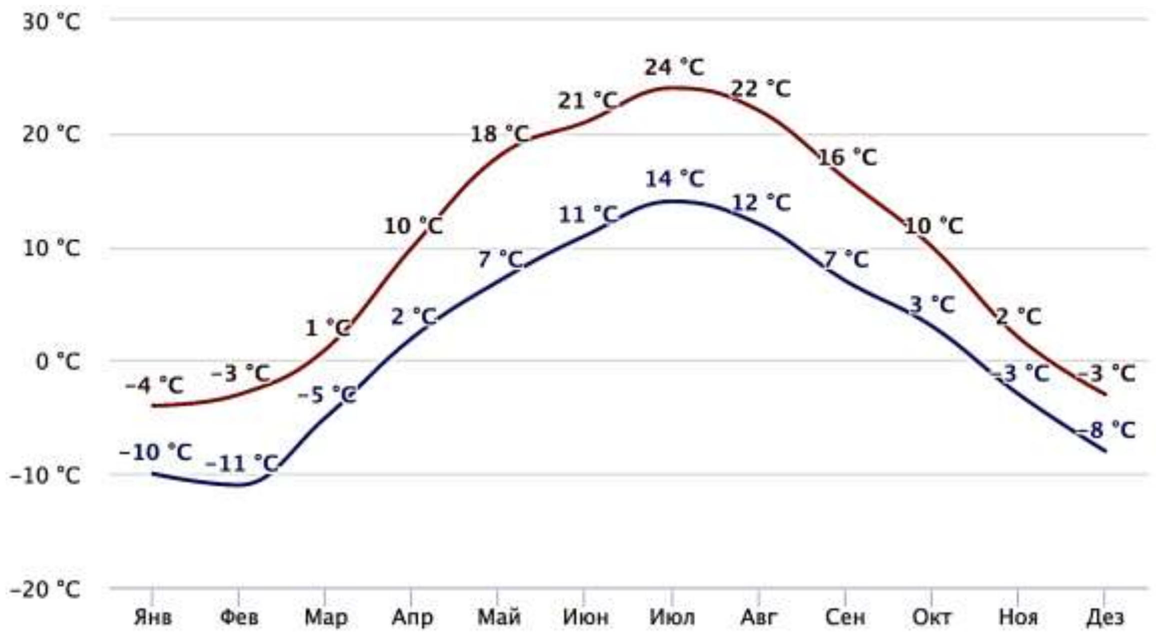

Рис. 1. Пример временных рядов: максимальные и минимальные значения среднесуточной температуры

Временные ряды состоят из двух элементов:

– периода времени, за который или по состоянию на который приводятся числовые значения;
– числовых значений того или иного показателя, называемых уровнями ряда.

Временные ряды классифицируются по следующим признакам:

– по форме представления уровней:
  - ряды абсолютных показателей;
  - относительных показателей;
  - средних величин.
– по количеству показателей, для которых определяются уровни в каждый момент времени: одномерные и многомерные временные ряды;
– по характеру временного параметра: моментные и интервальные временные ряды. В моментных временных рядах уровни характеризуют значения показателя по состоянию на определенные моменты времени. В интервальных рядах уровни характеризуют значение показателя за определенные периоды времени. Важная особенность интервальных временных рядов абсолютных величин заключается в возможности суммирования их уровней. Отдельные же уровни моментного ряда абсолютных величин содержат элементы повторного счета. Это делает бессмысленным суммирование уровней моментных рядов;
– по расстоянию между датами и интервалами времени выделяют равноотстоящие — когда даты регистрации или окончания периодов следуют друг за другом с равными интервалами и неполные (неравноотстоящие) — когда принцип равных интервалов не соблюдается;
– по наличию пропущенных значений: полные и неполные временные ряды;

– временные ряды бывают детерминированными и случайными: первые получают на основе значений некоторой неслучайной функции (ряд последовательных данных о количестве дней в месяцах); вторые есть результат реализации некоторой случайной величины.
– в зависимости от наличия основной тенденции выделяют стационарные ряды, в которых среднее значение и дисперсия постоянны, и нестационарные, содержащие основную тенденцию развития.

Временные ряды, как правило, возникают в результате измерения некоторого показателя. Это могут быть как показатели (характеристики) технических систем, так и показатели природных, социальных, экономических и других систем (например, погодные данные). Типичным примером временного ряда можно назвать биржевой курс, при анализе которого пытаются определить основное направление развития (тенденцию или тренд).

1. Методы прогнозирования временных рядов

Метод прогнозирования представляет собой последовательность действий, в результате выполнения которой определяется модель прогнозирования конкретного временного ряда.

На первом этапе классификации методы можно условно поделить на две группы: интуитивные и формализованные (рис. 1).

Рис. 2. Классификация методов прогнозирования

Интуитивные методы прогнозирования имеют дело с суждениями и оценками экспертов. На сегодняшний день они часто применяются в маркетинге, экономике, политике, так как система, поведение которой необходимо спрогнозировать, или очень сложна и не поддается математическому описанию, или очень проста и в таком описании не нуждается.

Формализованные методы — описанные в литературе методы прогнозирования, в результате которых строят модели прогнозирования, то есть определяют такую математическую зависимость, которая позволяет вычислить будущее значение процесса, то есть сделать прогноз.

В данной работе рассматриваются формальные методы. Метод прогнозирования содержит действия по оценке качества прогнозных значений. Общий итеративный подход к построению математической модели прогнозирования состоит из следующий шагов [1].

Шаг 1. На первом шаге на основании предыдущего собственного или стороннего опыта выбирается общий класс моделей для прогнозирования временного ряда на заданный горизонт.

Шаг 2. Определенный общий класс моделей обширен. Для непосредственной подгонки к исходному временному ряду, развиваются грубые методы идентификации подклассов моделей. Такие методы идентификации используют качественные оценки временного ряда.

Шаг 3. После определения подкласса модели, необходимо оценить ее параметры, если модель содержит параметры, или структуру, если модель относится к категории структурных моделей. На данном этапе обычно используется итеративные способы, когда производится оценка участка (или всего) временного ряда при различных значениях изменяемых величин. Как правило, данный шаг является наиболее трудоемким в связи с тем, что часто в расчет принимаются все доступные исторические значения временного ряда.

Шаг 4. Далее производится диагностическая проверка полученной модели прогнозирования. Чаще всего выбирается участок или несколько участков временного ряда, достаточных по длине для проверочного прогнозирования и последующей оценки точности прогноза. Выбранные для диагностики модели прогнозирования участки временного ряда называются контрольными участками (периодами).

Шаг 5. В случае если точность диагностического прогнозирования оказалась приемлемой для задач, в которых используются прогнозные значения, то модель готова к использованию. В случае если точность прогнозирования оказалось недостаточной для последующего использования прогнозных значений, то возможно итеративное повторение всех описанных выше шагов, начиная с первого.

2. Модели прогнозирования временных рядов

На основе формализованных методов базируется две группы моделей: модели предметной области и модели временных рядов.

Модели предметной области — такие математические модели прогнозирования, для построения которых используют законы предметной области. Например, модель, на которой делают прогноз погоды, содержит уравнения динамики жидкостей и термодинамики. Прогноз развития популяции делается на модели, построенной на дифференциальном уравнении. То есть, в таких моделях используются зависимости, свойственные конкретной предметной области, модель, к которой требуется индивидуальный подход.

Модели временных рядов — математические модели прогнозирования, которые стремятся найти зависимость будущего значения от прошлого внутри самого процесса и на этой зависимости вычислить прогноз. Эти модели универсальны для различных предметных областей, то есть их общий вид не меняется в зависимости от природы временного ряда. Таким образом, можно говорить, что это обобщенные модели.

Составить общую классификацию моделей предметной области не представляется возможным. Для временных рядов модели могут быть разделены на статистические и структурные.

В статистических моделях зависимость будущего значения от прошлого задается в виде некоторого уравнения. К ним относятся:

1) регрессионные модели (линейная регрессия, нелинейная регрессия);
2) авторегрессионные модели (ARIMAX, GARCH, ARDLM);
3) модель экспоненциального сглаживания;
4) модель по выборке максимального подобия.

В структурных моделях зависимость будущего значения от прошлого задается в виде некоторой структуры и правил перехода по ней. К ним относятся:

1) нейросетевые модели;
2) модели на базе цепей Маркова;
3) модели на базе классификационно-регрессионных деревьев.

Все методы и модели были сведены в обобщенную классификацию (рис.3).

Рис. 3. Общая классификация методов и моделей прогнозирования

Рассмотрим приведенные выше модели подробнее.

2.1. Регрессионные модели прогнозирования

Данные модели являются одними из старейших, однако сейчас находят не самое большое применение.

Существует много задач, требующих изучения отношения между двумя и более переменными. Для решения таких задач используется регрессионный анализ. В настоящее время регрессия получила широкое применение, включая задачи прогнозирования и управления. Целью регрессионного анализа является определение зависимости между исходной переменной и множеством внешних факторов (регрессоров). При этом коэффициенты регрессии могут определяться по методу наименьших квадратов или методу максимального правдоподобия.

Основные виды регрессионной модели:

- простая линейная регрессия (linear regression);
- множественная регрессия;
- нелинейная регрессия.

Рассмотрим эти виды регрессии.

2.1.1. Линейная регрессионная модель

Самым простым вариантом регрессионной модели является линейная регрессия. В основу модели положено предположение, что существует дискретный внешний фактор X(t), оказывающий влияние на исследуемый процесс Z(t), при этом связь между процессом и внешним фактором линейна. Модель прогнозирования на основании линейной регрессии описывается уравнением: $Z(t) = \alpha_0 + \alpha_1 X(t) + \varepsilon_t$, где α_0 и α_1 — коэффициенты регрессии; ε_t — ошибка модели.

Для получения прогнозных значений Z(t) в момент времени t необходимо иметь значение X(t) в тот же момент времени t, что редко выполнимо на практике.

2.1.2. Множественная регрессионная модель

На практике на процесс Z(t) оказывают влияние целый ряд дискретных внешних факторов $X_1(t)$,…, $X_s(t)$. Тогда модель прогнозирования имеет вид $Z(t) = \alpha_0 + \alpha_1 X_1(t) + \alpha_2 X_2(t) + \cdots + \alpha_s X_s(t) + \varepsilon_t$, где α_0 и α_1 — коэффициенты регрессии; ε_t — ошибка модели.

Недостатком данной модели является то, что для вычисления будущего значения процесса Z(t) необходимо знать будущие значения всех факторов $X_1(t)$,…, $X_s(t)$. что почти невыполнимо на практике.

2.1.3. Нелинейная регрессионная модель

В основу нелинейной регрессионной модели положено предположение о том, что существует известная функция: $Z(t) = F(X(t), A)$, где $Z(t)$ – исходный процесс, $X(t)$ –

внешний фактор, от которого зависит процесс $Z(t)$, A – функция, параметры которой необходимо определить в рамках построения модели прогнозирования.

Например, можно предположить, что $Z(t) = \alpha_1 \cos(X(t)) + \alpha_0$. Тогда для построения модели достаточно определить параметры $A = |\alpha_0, \alpha_0|$, где α_0 и α_1 — коэффициенты регрессии. Однако на практике редко встречаются процессы, для которых вид функциональной зависимости между процессом Z(t) и внешним фактором X(t) заранее известен. В связи с этим нелинейные регрессионные модели применяются редко.

2.2. Авторегрессионные модели

В основу авторегрессионных моделей заложено предположение о том, что значение процесса *Z(t)* линейно зависит от некоторого количества предыдущих значений того же процесса *Z(t-1),…,Z(t-p)*.

Рассмотрим основные виды авторегрессионных моделей:

- модель скользящего среднего;
- модель с условной гетероскедастичностью;
- модель с распределенным лагом.

2.2.1. Авторегрессионная модель скользящего среднего

В области анализа временных рядов модель авторегрессии (autoregressive, AR) и модель скользящего среднего (moving average, MA) является одной из наиболее используемых [1].

Согласно работе [1], модель авторегрессии является исключительно полезной для описания некоторых встречающихся на практике временных рядов. В этой модели текущее значение процесса выражается как конечная линейная совокупность предыдущих значений процесса и импульса, который называется «белым шумом». Рассмотрим следующую формулу:

$$Z(t) = C + \varphi_1 Z(t-1) + \varphi_2 Z(t-2) + \dots + \varphi_p Z(t-p) + \varepsilon_t ,$$

где $Z(t)$ – процесс авторегрессии порядка p (часто в литературе AR(p)),

C – вещественная константа,

φ_1,.., φ_p – коэффициенты модели,

ε_t – ошибка модели.

Для определения φ_i и C используют метод наименьших квадратов или метод максимального правдоподобия.

Другой тип модели имеет большое значение в описании временных рядов и часто используется совместно с авторегрессией называется моделью скользящего среднего порядка q (MA(q)) и описывается уравнением:

$$Z(t) = \frac{1}{q}(Z(t-1) + Z(t-2) + \dots + Z(t-p)) + \varepsilon_t ,$$

где q – порядок скользящего среднего,

ε_t – ошибка прогнозирования.

Модель скользящего среднего является по сути дела фильтром низких частот. Существуют разные типы скользящего среднего: простые, взвешенные, кумулятивные, экспоненциальные модели скользящего среднего.

Для достижения большей гибкости в подгонке модели часто целесообразно объединить в одной модели авторегрессию и скользящее среднее. Общая модель обозначается *ARMA(p,q)* соединяет в себе фильтр в виде скользящего среднего порядка q и авторегрессию фильтрованных значений процесса порядка p.

Если в качестве входных данных используются не сами значения временного ряда, а их разность d-того порядка (на практике d необходимо определять, однако в большинстве случаев $d \leq 2$), то модель носит название авторгерессии проинтегрированного скользящего среднего. В литературе данную модель называют *ARIMA(p,d,q)* (autoregression integrated moving average).

Развитием модели ARIMA(p,d,q) является модель ARIMAX(p,d,q), которая описывается уравнением:

$$\mathrm{Z(t)} = \mathrm{AR(p)} + \alpha_1 \mathrm{X}_1(\mathrm{t}) + \alpha_2 \mathrm{X}_2(\mathrm{t}) + \cdots + \alpha_\mathrm{s} \mathrm{X}_\mathrm{s}(\mathrm{t}) ,$$

где α_1,..., α_s — коэффициенты внешних факторов $X_1(t),\ldots,X_S(t)$.

В данной модели чаще всего процесс $Z(t)$ является результатом модели MA(q), то есть отфильтрованными значениями исходного процесса. Далее для прогнозирования $Z(t)$ используется модель авторегрессии, в которой введены дополнительные регрессоры внешних факторов $\mathrm{X}_1(\mathrm{t})$,…, $\mathrm{X}_\mathrm{s}(\mathrm{t})$.

2.2.2. Авторегрессионная модель с условной гетероскедастичностью

Авторегрессионная модель с условной гетероскедастичностью (autoregressive conditional heteroskedasticity, GARCH) была разработана в 1986 году Тимом Петером Борреслевом и является моделью остатков для модели AR(p) [2].

Особенностью данной модели является предположение, что условное значение дисперсии зависит от предыдущих значений ряда и от предыдущих значений дисперсии. В таком случае дисперсия модели описывается следующим уравнением:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 \sigma_{t-1} + \cdots + \beta_p \sigma_{t-p} ,$$

где α_0,..., α_q и β_0,..., β_p — неотрицательные коэффициенты модели,

ε_t – ошибка модели.

Приведенное уравнение называется моделью GARCH(p,q) и имеет два параметра: p характеризует порядок авторегрессии квадратов остатков (вклад предыдущих значений ряда); q — количество предшествующих оценок остатков (вклад предыдущих значений дисперсии).

Наиболее частое применение данная модель получила в финансовом секторе, где с помощью нее моделируется волатильность. На сегодняшний день существует ряд модификаций модели под названиями NGARCH, IGARCH, EGARCH, GARCH-M и другие [2].

2.2.3. Авторегрессионная модель с распределенным лагом

Авторегрессионная модель с распределенным лагом (autoregressive distributed lag models, ARDLM) преимущественно описывается в книгах по эконометрике.

Часто при моделировании процессов на изучаемую переменную влияют не только текущие значения процесса, но и его лаги, то есть значения временного ряда, предшествующие изучаемому моменту времени. Модель авторегрессии распределенного лага описывается уравнением:

$$Z(t) = \varphi_0 + \varphi_1 Z(t - l - 1) + \varphi_2 Z(t - l - 2) + \cdots + \varphi_p Z(t - l - p) + \varepsilon_t \qquad (5)$$

где φ_0,..., φ_p – коэффициенты модели,

l — величина лага.

Данная модель обозначается как ARDLM(p,l) и чаще всего применяется для моделирования экономических процессов [3].

2.3. Модели экспоненциального сглаживания

Модели экспоненциального сглаживания разработаны в середине XX века и до сегодняшнего дня являются широко распространенными в силу их простоты и наглядности.

Модель экспоненциального сглаживания (exponential smoothing, ES) чаще всего применяется для моделирования финансовых и экономических процессов. В основу экспоненциального сглаживания заложена идея постоянного пересмотра прогнозных значений по мере поступления фактических. Модель ES присваивает экспоненциально убывающие веса наблюдениям по мере их старения. Таким образом, последние доступные наблюдения имеют большее влияние на прогнозное значение, чем старшие наблюдения.

Функция модели ES имеет вид:

$$Z(t) = S(t) + \varepsilon_t,$$

$$S(t) = \alpha * Z(t - 1) + (1 - \alpha) * S(t - 1) + \varepsilon_t,$$

где α — коэффициент сглаживания, $0 < \alpha < 1$;

начальные условия определяются как $S(1) = Z(0)$.

В данной модели каждое последующее сглаженное значение $S(t)$ является взвешенным средним между предыдущим значением временного ряда $Z(t)$ и предыдущего сглаженного значения $S(t\text{-}1)$.

2.3.1. Модель Хольта или двойное экспоненциальное сглаживание

Данная модель применяется для моделирования процессов, имеющих тренд. В этом случае в модели необходимо рассматривать две составляющие: уровень и тренд [4]. Уровень и тренд сглаживаются отдельно:

$$S(t) = \alpha * Z(t-1) + (1-\alpha) * (S(t-1) + B(t-1)) ,$$

где α – коэффициент сглаживания уровня,

$B(t)$ – функция линейного тренда.

$$B(t) = \gamma * \big(S(t-1) + S(t-2)\big) + (1+\gamma) * B(t-1) ,$$

где γ –коэффициент сглаживания тренда.

2.3.2. Модель Хольта-Винтерса или тройное экспоненциальное сглаживание

Метод Близок к методу Хольта, применяется для процессов, которые имеют тренд и сезонную составляющую:

$$Z(t) = \big(R(t) + G(t)\big) * S(t) ,$$

где $R(t)$ — сглаженный уровень без учета сезонной составляющей,

$G(t)$ — сглаженный тренд,

$S(t)$ — сезонная составляющая.

2.4. Модели на основе нейронных сетей

В целом, нейронная сеть – это система, способная изменять свою структуру под воздействием внешних факторов. Нейронная сеть состоит из нейронов.

Нейрон – это переключатель, получающий и передающий импульсы, или сигналы. Если нейрон получает достаточно сильный импульс, то говорят, что нейрон активирован, то есть передает импульсы связанным с ним нейронам. Не активированный нейрон остается в состоянии покоя и не передает импульс.

Нейрон состоит из нескольких компонентов: синапсов, соединяющих нейрон с другими нейронами и получающих импульсы от соседних нейронов, аксона, передающего импульс другим нейронам, и дендрита, получающего сигналы из различных источников, в т.ч. от синапсов.

Когда нейрон получает импульс, превышающий определенный порог, он передает импульс последующим нейронам (активирует импульс).

Искусственные нейронные сети имитируют работу нейронной сети. Они состоят из искусственных нейронов – упрощенной вариацией биологического нейрона (рис.4). На вход нейрона передается вектор из n компонентов, каждый элемент вектора – один из импульсов получаемый нейроном.

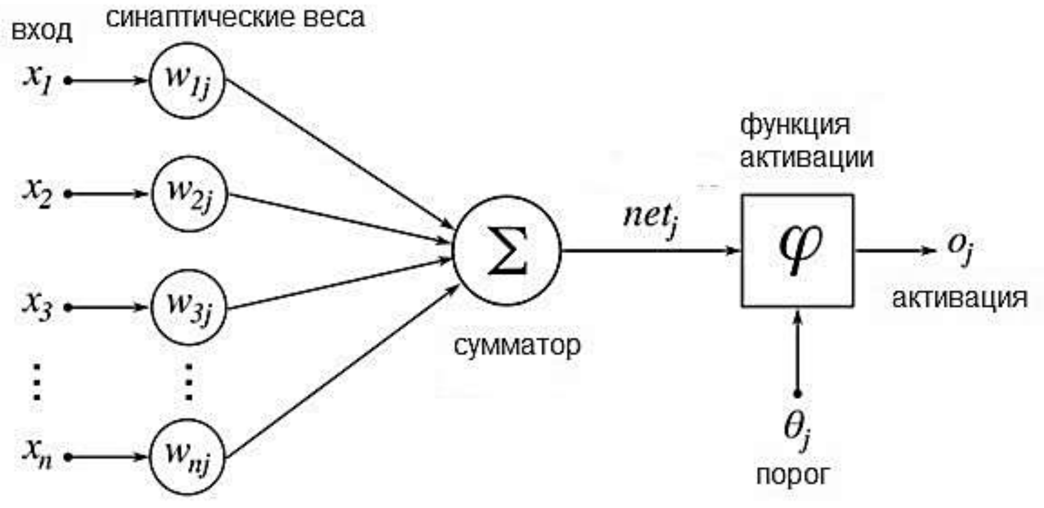

Рис. 4. Искусственный нейрон

Принцип работы следующий: каждый приходящий импульс умножается на соответствующий вес из вектора весов W (w_1, w_2, … , w_m), аналогичный синаптической силе. Далее все произведения суммируются, определяя уровень активации нейрона. В зависимости от функции активации и порога активации нейрон либо активируется и передает сигнал дальше (другим нейронам в сети), либо нет.

Соединяя такие нейроны (на вход одного нейрона i-го слоя идут импульсы других нейронов (i-1)-го слоя), получают нейронную сеть.

При помощи нейронных сетей возможно моделирование нелинейной зависимости будущего значения временного ряда от его фактических значений и от значений внешних факторов посредством обучения ее на уже имеющихся данных. Более подробно устройство нейронных сетей будет рассмотрено далее.

2.4.1. Другие модели прогнозирования

Можно использовать ряд других моделей, таких как модели на основе цепей Маркова, модель на классификационно-регрессионных деревьях, модель на основе генетического алгоритма для задач прогнозирования временных рядов, но по ряду причин было принято решение не рассматривать их подробно [6]. Например, модели на цепях Маркова основываются на том, что прогноз будущего значения зависит только от текущего значения, то есть модель будет гораздо менее гибкая и с большой вероятностью менее точная. Так же ряд подобных моделей не обладает достаточной методологической базой, то есть имеет недостаточно подробное описание как моделей, так и их возможностей.

3. Сравнение моделей прогнозирования

Основные достоинства и недостатки различных моделей были сведены в таблицу 1 [5].

Таблица 1.

Сравнение моделей прогнозирования

Модель и метод	Достоинства	Недостатки
Регрессионные модели и методы	Простота, гибкость, прозрачность моделирования; единообразие анализа и проектирования.	Сложность определения функциональной зависимости; трудоемкость нахождения коэффициентов зависимости; отсутствие возможности моделирования нелинейных процессов (для нелинейной регрессии)
Авторегрессионные модели и методы	Простота, прозрачность моделирования; единообразие анализа и проектирования; множество примеров применения	Трудоемкость и ресурсоемкость идентификации моделей; невозможность моделирования нелинейностей; низкая адаптивность
Модели и методы экспоненциального сглаживания	Простота моделирования; единообразие анализа и проектирования	Недостаточная гибкость; узкая применимость моделей
Нейросетевые модели и методы	Нелинейность моделей; масштабируемость, высокая адаптивность; единообразие анализа и проектирования; множество примеров применения	Отсутствие прозрачности; сложность выбора архитектуры; жесткие требования к обучающей выборке; сложность выбора алгоритма обучения; ресурсоемкость процесса обучения

Нелинейность нейронных сетей позволяет устанавливать нелинейные зависимости между будущими и фактическими значениями процессов. Другими важными достоинствами является масштабируемость – параллельная структура искусственных нейронных сетей ускоряет вычисления, что является крайне актуальным в промышленных масштабах, когда необходимо обрабатываться терабайты данных.

Заключение

В статье были рассмотрены различные подходы к анализу временных рядов. Приведены и описаны наиболее популярные прогнозные модели, из использующихся для решения задач предиктивной аналитики.

На основе проведенного обзора сформулированы основные преимущества и недостатки различных типов моделей, на основе которых можно принимать решения об использовании того или иного вида моделей для решения конкретных задач.

Было выявлено, что наиболее перспективным подходом на данный момент можно считать использование нейронных сетей, так как они позволяют построить наиболее гибкую и полную с точки зрения отслеживаемых закономерностей модель.

Список литературы

[1]. Бокс Дж., Дженкинс Г.М. Анализ временных рядов, прогноз и управление. М.: Мир, 1974. 406 с.

[2]. Autoregressive conditional heteroskedasticity // The free encyclopedia «Wikipedia». Режим доступа: http://en.wikipedia.org/wiki/Autoregressive_conditional_heteroskedastici (дата обращения 20.05.2017) .

[3]. Эконометрия: Учебное пособие / В.И. Суслов [и др.] Новосибирск: Издательство СО РАН, 2005. 744 с.

[4]. Prajakta S.K. Time series Forecasting using Holt-Winters Exponential Smoothing // Kanwal Rekhi School of Information Technology Journal 2004. 13 p. Режим доступа :http://www.it.iitb.ac.in/~praj/acads/seminar/04329008_ExponentialSmoothing.pdf (дата обращения 12.05.2017).

[5]. Чучуева А.И. Модель прогнозирования временных рядов по выборке максимального правдоподобия: дис. … канд. техн. наук. М., 2012. 155 с.

[6]. Афанасьев В. Н., Юзбашев М. М. Анализ временных рядов и прогнозирование //М.: Финансы и статистика. – 2001. – Т. 228. – С. 2.