

На правах рукописи

ЧУЧУЕВА ИРИНА АЛЕКСАНДРОВНА

**МОДЕЛЬ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ
ПО ВЫБОРКЕ МАКСИМАЛЬНОГО ПОДОБИЯ**

Специальность 05.13.18 – Математическое моделирование,
численные методы и комплексы программ

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Москва – 2012

Работа выполнена на кафедре информационных измерительных систем и телекоммуникаций Московского государственного технического университета имени Н.Э. Баумана

- Научный руководитель: доктор технических наук, профессор
Павлов Юрий Николаевич
- Официальные оппоненты: доктор физико-математических наук,
профессор Павлов Игорь Валерианович
- кандидат физико-математических наук
Масалович Андрей Игоревич
- Ведущая организация: Федеральное государственное унитарное
предприятие «Научно-производственное
объединение имени С.А. Лавочкина»

Защита диссертации состоится «20» марта 2012 г. в 11 час. 00 мин. на заседании диссертационного совета Д 212.141.15 при Московском государственном техническом университете имени Н.Э. Баумана по адресу: 105082, Москва, Рубцовская наб., д. 2/18, ауд. 1006Л.

Отзывы на автореферат в двух экземплярах, заверенные печатью, просим направлять по адресу: 105005, Москва, 2-я Бауманская ул., д. 5, МГТУ им. Н.Э. Баумана, ученому секретарю диссертационного совета Д 212.141.15.

С диссертацией можно ознакомиться в библиотеке Московского государственного технического университета имени Н.Э. Баумана.

Автореферат диссертации разослан «__» _____ 2012 г.

Ученый секретарь диссертационного совета,
кандидат технических наук,
старший научный сотрудник, доцент

Аттетков А.В.

НТБ МГТУ им. Н.Э. Баумана



1997144

Чучуева И.А. Модель прогноз

-145
24891

Общая характеристика работы

Актуальность темы. Задача прогнозирования будущих значений временного ряда, используя его исторические значения, является основой для финансового планирования в экономике и торговле, планирования, управления и оптимизации объемов производства, складского контроля.

В настоящее время компаниями осуществляется накопление исторических значений экономических и физических показателей в базах данных, что существенно увеличивает объемы входной информации для задачи прогнозирования. Вместе с тем развитие аппаратных и программных средств предоставляет все более мощные вычислительные платформы, на которых возможна реализация сложных алгоритмов прогнозирования. Кроме того, современные подходы к экономическому и техническому управлению предъявляют все более жесткие требования к точности прогнозирования. Таким образом, задача прогнозирования временных рядов усложняется одновременно с развитием информационных технологий.

Задача прогнозирования различных временных рядов актуальна и ее решение является неотъемлемой частью ежедневной работы многих компаний. Данная задача решается созданием модели прогнозирования, адекватно описывающей исследуемый процесс.

К сегодняшнему дню разработано множество моделей прогнозирования временных рядов: регрессионные и авторегрессионные модели, нейросетевые модели, модели экспоненциального сглаживания, модели на базе цепей Маркова, классификационные модели и др. Наиболее популярными и широко используемыми являются классы авторегрессионных и нейросетевых моделей. Существенным недостатком авторегрессионного класса является большое число свободных параметров, идентификация которых неоднозначна и ресурсоемка. Основным недостатком класса нейросетевых моделей является недоступность промежуточных вычислений, выполняющихся в «черном ящике», и, как следствие, сложность интерпретации результатов моделирования. Кроме того, слабой стороной данного класса моделей является сложность выбора алгоритма обучения нейронной сети.

Диссертация посвящена разработке новой авторегрессионной модели прогнозирования, которая имеет сравнимую с другими моделями эффективность прогнозирования различных временных рядов и при этом устраняет основной и наиболее существенный недостаток авторегрессионного класса моделей — большое число свободных параметров.

Целью работы является разработка новой модели и соответствующего ей метода прогнозирования, относящейся к классу авторегрессионных моделей и устраняющей основной недостаток данного класса — большое число свободных параметров. Новая модель и соответствующий ей метод должны иметь высокую скорость вычисления прогнозных значений и сравнимую с другими моделями точность прогнозирования различных временных рядов.

Для достижения этой цели были поставлены и решены следующие задачи.

1. Осуществить обзор моделей и методов прогнозирования временных рядов, выявить достоинства и недостатки каждого класса моделей. Выявить наиболее используемые классы моделей прогнозирования и их основные недостатки, определить перспективные подходы, позволяющие устранить недостатки авторегрессионного класса моделей.

2. Разработать новую модель прогнозирования временных рядов, устраняющую указанный недостаток авторегрессионного класса моделей.

3. Разработать новый метод прогнозирования на основании предложенной модели и выполнить программную реализацию алгоритмов.

4. Оценить эффективность предложенной модели прогнозирования при решении задачи прогнозирования различных временных рядов.

Методы исследования. При решении поставленных задач в работе использованы методы математического моделирования, анализ временных рядов, регрессионный анализ, методы объектно-ориентированного программирования.

Научная новизна. В диссертации получены следующие основные результаты, которые выносятся на защиту.

1. Модель экстраполяции временных рядов по выборке максимального подобия, относящаяся к классу авторегрессионных моделей и имеющая единственный параметр.

2. Метод прогнозирования временных рядов на основании разработанной модели, содержащий набор алгоритмов для экстраполяции временных рядов, идентификации модели и построения доверительного интервала прогнозных значений.

3. Результаты прогнозирования временных рядов показателей энергорынка РФ, а также временных рядов из других предметных областей, подтверждающие эффективность разработанной модели.

Достоверность и обоснованность выносимых на защиту результатов прогнозирования показателей энергорынка РФ документально подтверждается ЗАО «РусПауэр», использующего разработанные алгоритмы на ежедневной основе. Достоверность результатов прогнозирования временного ряда уровня сахара крови человека, больного диабетом первого типа, обеспечивается строгостью применяемого математического аппарата и подтверждается приведенным сравнительным анализом. Достоверность результатов прогнозирования скорости движения транспорта по г. Москва обеспечивается условиями открытого конкурса, проводимого компанией «Яндекс». Результаты конкурса опубликованы в открытом доступе по адресу <http://imat2010.yandex.ru/results>.

Практическая ценность. Разработанная модель и метод прогнозирования по выборке максимального подобия могут применяться для прогнозирования временных рядов различных предметных областей. Разработанные алгоритмы экстраполяции временных рядов с учетом и без учета внешних факторов наглядны для программной реализации. Скорость вычисления прогнозных значений при использовании модели высока. Задача идентификации модели упрощена в сравнении с другими моделями авторегрессионного класса.

Реализация и внедрение результатов работы. Результаты работы реализованы по заказу Закрытого акционерного общества «РусПауэр» в виде серверного приложения для прогнозирования показателей энергорынка РФ на ежедневной основе. Приложение работает в автоматическом режиме и предоставляет прогнозные значения показателей без вмешательства эксперта.

Апробация работы. Основные результаты диссертационной работы докладывались на I Международной научно-практической конференции ученых, аспирантов и студентов «Наука и современность 2010» (Новосибирск, 2010), на научно-технической конференции «Студенческая научная весна» (Москва, 2010), на III Международной конференции «Математическое моделирование социальной и экономической динамики (MMSED-2010)» (Москва, 2010).

Публикации. Основные результаты диссертации опубликованы в 8 научных статьях, в том числе в 5 статьях, опубликованных в журналах из Перечня рецензируемых ведущих научных журналов и изданий, и 2 тезисов докладов.

Личный вклад соискателя. Все исследования, результаты которых изложены в настоящей работе, получены лично соискателем в процессе научных исследований. Из совместных публикаций в диссертацию включен лишь тот материал, который непосредственно принадлежит соискателю.

Структура и объем работы. Диссертационная работа состоит из введения, четырёх глав, заключения, списка литературы и приложения, занимающих 154 страниц текста, в том числе 33 рисунка на 29 страницах, 37 таблиц на 29 страницах, список использованной литературы из 75 наименований на 10 страницах.

Содержание работы

Во введении обоснована актуальность темы, сформулированы цель и задачи исследования, научная новизна, указана достоверность и обоснованность, а также практическая значимость полученных результатов, основные положения, выносимые на защиту.

Первая глава является обзорной. В ней приведены определения основных понятий, сформулированы два вида постановки задачи прогнозирования временного ряда — с учетом и без учета внешних факторов.

В результате обзора установлено, что к текущему времени разработано множество моделей прогнозирования временных рядов. Подробнее рассмотрены следующие классы моделей прогнозирования:

- регрессионные модели;
- авторегрессионные модели;
- модели экспоненциального сглаживания;
- нейросетевые модели;
- модели на базе цепей Маркова;
- модели на базе классификационно-регрессионных деревьев.

Среди множества моделей прогнозирования наиболее широкую применимость имеют классы авторегрессионных и нейросетевых моделей.

По итогам обзора выявлены достоинства и недостатки всех рассмотренных классов моделей. Установлено, что существенным недостатком авторегрессионного класса моделей является большое число свободных параметров, требующих идентификации. Основным недостатком класса нейросетевых моделей является недоступность промежуточных вычислений и, как следствие, сложность интерпретации результатов моделирования. Кроме того, к слабой стороне нейросетевых моделей следует отнести сложность выбора алгоритма обучения сети.

На основании обзора моделей прогнозирования установлено, что наиболее перспективным направлением развития моделей с целью повышения точности является создание комбинированных моделей, выполняющих на первом этапе кластеризацию, а затем прогнозирование временного ряда внутри установленного кластера. В рамках данного направления подробно рассмотрен ряд моделей, на основании которых

сформулированы подходы к разработке новой модели.

Во второй главе предложены модели экстраполяции временных рядов по выборке максимального подобия для двух видов постановки задачи — с учетом и без учета внешних факторов.

Пусть дан временной ряд $Z(t) = Z(1), Z(2), \dots, Z(T)$. Набор последовательных значений $Z_t^M = Z(t), Z(t+1), \dots, Z(t+M-1)$, лежащий внутри исходного временного ряда, назовем выборкой длины M с моментом начала отчета t ; $M \in \{1, 2, \dots, T\}$, $t \in \{1, 2, \dots, T-M+1\}$. Фактически выборкой является фрагмент временного ряда, имеющий точку начала отсчета и длину. Две выборки одинаковой длины, принадлежащие одному временному ряду, обозначим через временную задержку k : $Z_t^M = Z(t), \dots, Z(t+M-1)$ и $Z_{t-k}^M = Z(t-k), \dots, Z(t-k+M-1)$, где $k \in \{1, 2, \dots, t-1\}$. Если речь идет о векторах, соответствующих указанным выборкам, то используется обозначение Z_t^M .

По результатам обзора моделей прогнозирования было сформулировано следующее свойство временных рядов: «Моделирование временных рядов при помощи выборок основано на предположении, что временной ряд представляет собой последовательность выборок. При этом на практике часто наблюдается, что фактические выборки временного ряда могут иметь подобие с будущими выборками». Указанное свойство выборок использовано для построения модели прогнозирования.

Используя свойство выборок повторяться, аппроксимируем более позднюю выборку Z_t^M при помощи более ранней выборки Z_{t-k}^M

$$Z_t^M = \alpha_1 Z_{t-k}^M + \alpha_0 I^M + E^M. \quad (1)$$

Перепишем выражение (1) следующим образом

$$\hat{Z}_t^M = \alpha_1 Z_{t-k}^M + \alpha_0 I^M. \quad (2)$$

Здесь α_1 и α_0 — коэффициенты, I^M — единичный вектор, E^M — вектор значений ошибок аппроксимации.

Пусть дана линейная зависимость (1), тогда функция ошибки аппроксимации S_k^M для выборок Z_t^M и Z_{t-k}^M с задержкой k имеет вид

$$S_k^M(\alpha_1, \alpha_0) = \sum_{i=0}^{M-1} (Z(t+i) - \alpha_1 Z(t-k+i) - \alpha_0)^2. \quad (3)$$

Функция $S_k^M(\alpha_1, \alpha_0)$ называется функцией суммы квадратов. Задача состоит в том, чтобы подобрать такие значения α_1 и α_0 , чтобы при подстановке их в (3) было получено минимальное возможное значение

$S_k^M(\alpha_1, \alpha_0)$. В работе задача решена методом наименьших квадратов.

Пусть дан временной ряд $Z(t)$ и некоторая выборка Z_t^M , принадлежащая данному временному ряду. Определим все значения $S_k^M(\alpha_1, \alpha_0)$ для $k \in \{1, 2, \dots, t-1\}$, $M = const$. Далее в множестве значений S_k^M найдем минимальное

$$S_{kmin}^M = \min(S_1^M, S_2^M, \dots, S_{t-1}^M) \quad (4)$$

В работе показано, что минимум ошибки регрессии S_{kmin}^M соответствует максимуму модуля коэффициента линейной корреляции между исходной выборкой Z_t^M и модельной $\hat{Z}_t^M$. То есть если для $k \in \{1, 2, \dots, t-1\}$ и $M = const$ определить множество значений модуля линейной корреляции

$$\rho_k^M = |\rho(\hat{Z}_t^M, Z_t^M)| = \frac{\left| \sum_{i=1}^M (\hat{Z}(t+i) - \bar{Z})(Z(t+i) - \bar{Z}) \right|}{\sqrt{\sum_{i=1}^M (\hat{Z}(t+i) - \bar{Z})^2 \cdot \sum_{i=1}^M (Z(t+i) - \bar{Z})^2}} \in [0, 1], \quad (5)$$

а после определить максимальное значение полученного множества

$$\rho_{kmax}^M = \max(\rho_1^M, \rho_2^M, \dots, \rho_{t-1}^M), \quad (6)$$

то задержка $kmin$ из выражения (4) и задержка $kmax$ из выражения (6) будут равны между собой, т.е. $kmin = kmax$. Проведенные в работе расчеты подтверждают данное утверждение.

Определенную в (4) или (6) задержку, соответствующую минимуму ошибки регрессии S_{kmin}^M и максимуму модуля корреляции ρ_{kmax}^M обозначим $kmax$, а выборку Z_{t-kmax}^M назовем выборкой максимального подобия (most similar pattern). Выборка максимального подобия Z_{t-kmax}^M является выборкой, которая при подстановке в уравнение (2) дает в результате значения выборки $\hat{Z}_t^M$, которая максимально точно описывает исходную выборку Z_t^M .

В работе сформулирована следующая гипотеза подобия: если исходная выборка Z_t^M и модельная выборка $\hat{Z}_t^M$, полученная на основании (2) с использованием выборки Z_{t-kmax}^M , имеют значение величины ρ_{kmax}^M , близкое к единице, то для некоторых значений P и выборки Z_{t-kmax}^{M+P} , Z_t^{M+P} значение величины ρ_{kmax}^{M+P} также близко к единице.

Задача экстраполяции временного ряда без учета внешних факторов состоит в том, что в момент времени T необходимо определить P значений временного ряда в моменты времени $Z(T+1)$, $Z(T+2)$, ...,

$Z(T+P)$.

Используя сформулированное выше свойство выборок, выразим выборку $\hat{Z}_{T+1}^P$ через некоторую выборку Z_τ^P , лежащую внутри исходного временного ряда Z_1^T

$$\hat{Z}_{T+1}^P = \alpha_1 Z_\tau^P + \alpha_0 I^P. \quad (7)$$

Решение задачи экстраполяции временного ряда состоит из следующих шагов.

Шаг 1. Определим выборку максимального подобия для выборки новой истории Z_{T-M+1}^M (выборка Новая история, рис. 1).

Поиск выборки максимального подобия осуществляем перебором всех возможных значений задержек $k \in \{1, 2, \dots, T-M-1\}$. Для каждого значения k из указанного диапазона решаем задачу аппроксимации (1), в результате которой определяем коэффициенты α_1 и α_0 , соответствующие k . Далее для найденной пары коэффициентов определяем значения модельной выборки $\hat{Z}_{T-M+1}^M$, на основании которых вычисляем значение ρ_k^M (5). После того, как множество значений ρ_k^M для $k \in \{1, 2, \dots, T-M-1\}$ получено, определяем значение ρ_{kmax}^M по выражению (6) и соответствующую выборку максимального подобия $Z_{T-M+1-kmax}^M = Z_{kmax}^M$.

Шаг 2. Вычислим выборку Z_τ^P (выборка Базовая история, рис. 1).

В соответствии с гипотезой подобия, в качестве выборки Z_τ^P используем выборку Z_{kmax+M}^P , то есть выборку, расположенную на оси времени сразу за выборкой максимального подобия.

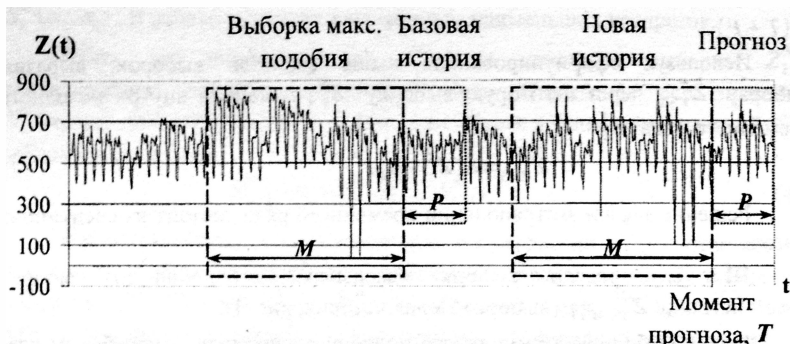
Шаг 3. Вычислим выборку $\hat{Z}_{T+1}^P$ (выборка Прогноз, рис. 1).

Экстраполированные значения выборки $\hat{Z}_{T+1}^P$ определим по формуле

$$\hat{Z}_{T+1}^P = \alpha_1 Z_{kmax+M}^P + \alpha_0 I^P = EMMSP(M), \quad (8)$$

которая представляет собой модель экстраполяции временных рядов по выборке максимального подобия (extrapolation model on most similar pattern, EMMSP(M)).

Положения упомянутых выборок представлены на рисунке 1.



Задача экстраполяции временного ряда с учетом внешних факторов состоит в том, чтобы определить будущие значения временного ряда $Z(t)$, т. е. выборку $\hat{Z}_{T+1}^*$, учитывая влияние на них внешних факторов, представленных в виде временных рядов $X_1(t), \dots, X_s(t)$.

Аналогично случаю без учета внешних факторов определим модель аппроксимации следующим образом

$$\hat{\mathbf{Z}}_t^M = \alpha_{S+1} \mathbf{Z}_{t-k}^M + \alpha_S X_{(S)t}^M + \dots \alpha_1 X_{(1)t}^M + \alpha_0 \mathbf{I}^M. \quad (9)$$

Пусть модельная выборка $\hat{Z}_t^M$ вычисляется на основании указанной зависимости (9). Вычислим все значения $S_k^M(\alpha_{s+1}, \dots, \alpha_0)$ для задержек $k \in \{1, 2, \dots, t-1\}$ и $M = \text{const}$. Далее в множестве значений S_k^M найдем минимальное по выражению (4). Как и в предыдущем случае минимум ошибки регрессии $S_{k_{\min}}^M$ соответствует максимуму модуля коэффициента линейной корреляции $\rho_{k_{\max}}^M$ (6).

Аналогично случаю без учета внешних факторов определяется выборка максимального подобия Z_{i-kmax}^M . Подстановка значений Z_{i-kmax}^M в уравнение аппроксимации (9) дает в результате значения выборки $\hat{Z}_i^M$, которая максимально точно описывает исходную выборку Z_i^M с учетом выборок внешних факторов $X_{(1)}^M, \dots, X_{(s)}^M$. Кроме того, аналогично предыдущему случаю в работе сформулирована гипотеза подобия для случая учета внешних факторов.

Для определения значений прогнозной выборки $\hat{Z}_{T+1}^P$, учитывая

влияние на них внешних факторов, представленных в виде временных рядов $X_1(t), \dots, X_s(t)$, был разработан алгоритм аналогичный случаю без учета внешних факторов. Выразим прогнозную выборку $\hat{Z}_{T+1}^P$ через некоторую выборку Z_τ^P , лежащую внутри исходного временного ряда Z_1^T , и выборки $X_{(1)T+1}^P, \dots, X_{(s)T+1}^P$ следующим образом

$$\hat{Z}_{T+1}^P = \alpha_{s+1} Z_\tau^P + \alpha_s X_{(s)T+1}^P + \dots + \alpha_1 X_{(1)T+1}^P + \alpha_0 I^P. \quad (10)$$

Решение задачи экстраполяции временного ряда состоит из следующих шагов.

Шаг 1. Определим выборку максимального подобия для выборки новой истории Z_{T-M+1}^M (выборка Новая история, рис. 2).

Выборку максимального подобия Z_{kmax}^M определяем аналогично предыдущему случаю.

Шаг 2. Вычислим выборку Z_τ^P (выборка Базовая история, рис. 2).

Базовую выборку Z_τ^P определяем аналогично предыдущему случаю.

Шаг 3. Вычислим выборку $\hat{Z}_{T+1}^P$ (выборка Прогноз, рис. 2).

Экстраполированные значения выборки $\hat{Z}_{T+1}^P$ определим по формуле

$$\hat{Z}_{T+1}^P = \alpha_{s+1} Z_{kmax}^P + \alpha_s X_{(s)T+1}^P + \dots + \alpha_1 X_{(1)T+1}^P + \alpha_0 I^P = EMMSPX(M), \quad (11)$$

которая представляет собой расширенную модель экстраполяции временных рядов по выборке максимального подобия (extrapolation model on most similar pattern extended, EMMSPX(M)).

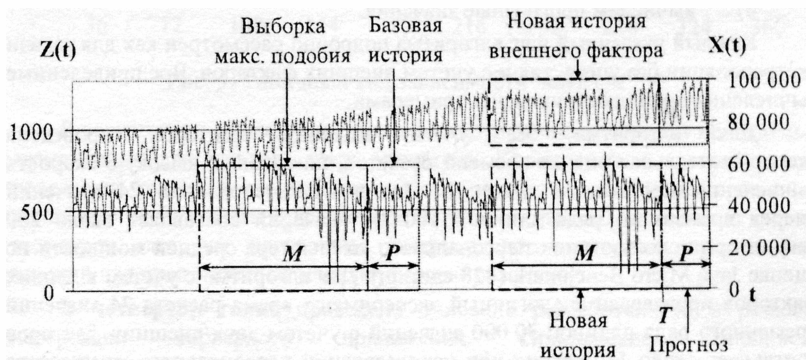


Рис. 2. Положение выборок на оси времени при экстраполяции временного ряда с учетом одного внешнего фактора

Главные особенности моделей прогнозирования по выборке максимального подобия с учетом и без учета внешних факторов:

- обе модели относятся к классу авторегрессионных моделей прогнозирования;
- модели имеют единственный параметр M ;
- экстраполяция P значений временного ряда для обеих моделей производится за одну итерацию.

В данной главе указано, что точность аппроксимации и экстраполяции временных рядов оценивается значением средней абсолютной ошибки в процентах (mean absolute percentage error, MAPE)

$$MAPE = \frac{1}{M} \sum_{i=t}^{t+M-1} \frac{|Z(i) - \hat{Z}(i)|}{Z(i)} \cdot 100\%. \quad (12)$$

В дополнение во второй главе приведены ограничения применимости разработанных моделей, а также предложены варианты моделей, применение которых может повысить точность прогнозирования временных рядов.

В третьей главе разработан метод прогнозирования на базе модели экстраполяции по выборке максимального подобия.

Экстраполяция временного ряда с учетом и без учета внешних факторов выполняется по единому алгоритму:

- 1) определим выборку новой истории;
- 2) определим выборку максимального подобия;
- 3) определим выборку базовой истории;
- 4) вычислим прогнозные значения.

Каждый указанный шаг алгоритма подробно рассмотрен как для задачи экстраполяции без учета, так и с учетом внешних факторов. Все приведенные вычисления проиллюстрированы примерами.

Для алгоритма без учета внешних факторов проведена экспериментальная оценка времени расчетов, показавшая высокую скорость вычислений прогнозных значений. Время прогнозирования 24 значений вперед временного ряда длиной 100 000 значений составляет около 200 секунд при использовании персонального компьютера средней мощности по оценке Java Micro Benchmark (828 единиц). Для алгоритма с учетом внешних факторов произведен аналогичный эксперимент: время расчета 24 значений временного ряда длиной 30 000 значений с учетом двух внешних факторов составляет около 170 секунд при использовании персонального компьютера указанного типа.

Предложен алгоритм идентификации разработанных моделей экстраполяции, состоящий из следующих шагов:

- 1) определим тестовый и контрольный периоды временного ряда;
- 2) определим время упреждения P , определим диапазон возможных значений параметра M ;
- 3) прогнозируем тестовый период на P значений вперед при всех значениях параметра M из установленного диапазона;
- 4) строим зависимость абсолютной ошибки прогнозирования (mean absolute error, MAE) от M , выделяем область устойчивого минимума ошибки (на рисунке 3 данная область выделена штриховой линией);
- 5) экспертно определяем окончательное значение параметра M .

Все указанные шаги алгоритма идентификации подробно рассмотрены в диссертации и проиллюстрированы примерами. На рисунке 3 приведена зависимость $MAE(M)$ для временного ряда энергопотребления европейской территории РФ. Подобный вид зависимости $MAE(M)$ является типичным.

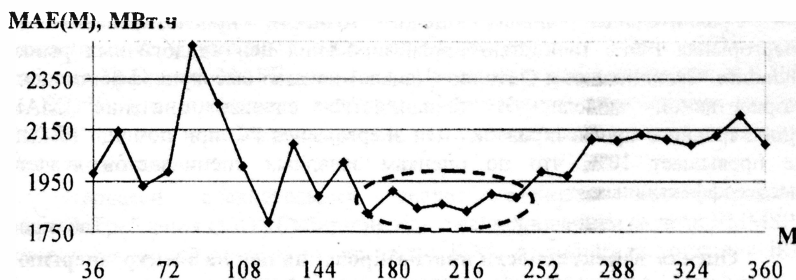


Рис. 3. Типичный вид зависимости $MAE(M)$

В завершении описания алгоритма идентификации моделей приведены рекомендации для каждого шага алгоритма. Кроме того, в третьей главе предложена схема формирования набора моделей для прогнозирования временного ряда, позволяющая повысить точность прогнозирования, а также рассмотрен алгоритм построения доверительного интервала прогнозных значений для предложенной модели прогнозирования.

В четвертой главе приведено описание разработанного в рамках диссертации серверного приложения. Приложение выполняет прогнозирование 19 временных рядов показателей энергорынка РФ на разработанной модели экстраполяции без участия эксперта на ежедневной основе. Приложение разработано по заказу компании ЗАО «РусНауэр»,

являющейся поставщиком аналитических отчетов на энергорынке РФ.

Разработанные модели EMMSP и EMMSPX применялись для краткосрочного (на сутки вперед) и среднесрочного (на неделю вперед) прогнозирования 19 временных рядов цен энергорынка РФ в почасовом разрешении общей длиной более 500 000 значений. Оценки точности прогнозирования цен на электроэнергию энергорынка РФ, представленные в диссертации, являются первыми опубликованными в открытом доступе.

Проведен сравнительный анализ точности прогнозирования цен на электроэнергию с использованием разработанной в диссертации модели и нейросетевой модели. Нейросетевая модель реализована компанией BIGroup Labs в виде программного комплекса для прогнозирования цен энергорынка РФ. Результаты совместного прогнозирования показали сравнимую эффективность модели EMMSP и нейросетевой модели для временного ряда цен на электроэнергию европейской зоны РФ. Значения MAPE составили 5.97% для EMMSP и 6.10% для нейросетевой модели (таблица 1).

Сравнительный анализ оценок точности прогнозирования цен энергорынка РФ с точностью прогнозирования цен аналогичных рынков Испании, Скандинавии и Онтарио (Канада) показал высокую эффективность разработанной модели. В большинстве случаев значение MAPE краткосрочного прогнозирования цен энергорынка РФ при помощи EMMSP не превышает 10%, что по оценкам западных специалистов является высокоэффективным.

Таблица 1.

Оценка эффективности прогнозирования цен на электроэнергию с использованием модели EMMSP

Времен- ной ряд	Время упреж- дения	Значение MAPE для EMMSP	Значение MAPE в российских публикациях	Значение MAPE в западных публикациях
Цены на электро- энергию энерго- рынка РФ (19 вр. рядов)	24	4.21% – 15.85%	Точности прогнозирования являются первыми опубликованными в открытом доступе по энергорынку РФ	Рынок Испании: 2.90% – 19.93%
	168	5.07% – 17.61%		Рынок Канады: 8.60% – 16.10% Рынок Скандинавии: 2.54% – 13.40%

Времен- ной ряд	Время упреж- дения	Значение MAPE для EMMSP	Значение MAPE в российских публикациях	Значение MAPE в западных публикациях
Цена европей- ской зоны РФ	24	5.97% (31.27 Руб/МВт·ч)	Оценка точности нейросетевой модели от компании BIGroupLabs: 6.10% (31.88 Руб/МВт·ч)	

В рамках диссертации были исследованы 8 временных рядов энергопотребления территорий РФ в почасовом разрешении, общая длина которых составляет более 300 000 значений.

Полученные значения MAPE для краткосрочного прогнозирования энергопотребления лежат в диапазоне от 0.91% до 1.83%. Значения MAPE для среднесрочного прогнозирования энергопотребления лежат в диапазоне от 1.26% до 3.30%. Полученные оценки точности сравнимы со значениями MAPE прогнозирования энергопотребления, приведенными в ряде новейших отечественных и западных работ, опубликованных в 2006 – 2011 годах (таблица 2). Оценки MAPE, указанные в работах, лежат в диапазоне от 0.80% до 5.27%, что доказывает высокую эффективность применения модели EMMSP для краткосрочного и среднесрочного прогнозирования энергопотребления.

Проведен сравнительный анализ точности прогнозирования энергопотребления ОАО «Сибирьэнерго» с использованием модели EMMSP и авторегрессионной модели от компании iRM. Австрийская компания iRM является одним из европейских лидеров по разработке программных продуктов для автоматизации торговли, прогнозирования, управления рисками на энергорынках Европы и РФ. Оценка точности прогнозирования энергопотребления на двух моделях показала сравнимую эффективность EMMSP и авторегрессионной модели компании iRM. Значения MAPE для исследуемого временного ряда энергопотребления составили 4.32% для EMMSP и 4.21% для авторегрессионной модели.

Таблица 2.

Оценка эффективности прогнозирования энергопотребления с использованием модели EMMSP

Временной ряд	Время упреж- дения	Значение MAPE для EMMSP	Значение MAPE в российских публикациях	Значение MAPE в западных публикациях
Энерго- потребление территорий РФ	24	0.91% – 1.83%	Ошибка прогнозирования согласно четырем научным публ. за 2007 – 2011 г.г.: 1.43% – 5%	Ошибка прогнозирования согласно 6 научным публ. за 2006 – 2010 г.г.: 0.99% – 5.27%
	168	1.26% – 3.30%		
Энергопотреб- ление ОАО «Сибирьэнерго»	24	4.32%	Значение MAPE для модели ARIMAX от компании iRM (Австрия): 4.21%	

Компания «РусПауэр» при помощи разработанного в диссертационной работе серверного приложения ежедневно формирует прогнозы по 12 временным рядам цен на электроэнергию и 8 временным рядам энергопотребления в виде аналитического продукта.

Реализация разработанной модели прогнозирования с использованием математического пакета MATLAB показала высокую точность прогнозирования и других временных рядов, в частности, временного ряда уровня сахара крови человека, больного сахарным диабетом первого типа. Проведено сравнение точности прогнозирования на один и полтора часа вперед на модели EMMSP и специализированной нейросетевой модели. Значения MAPE при прогнозировании на один час вперед составили 11.33% для EMMSP и 12.21% для нейросетевой модели, а при прогнозировании на полтора часа вперед — 14.70% для EMMSP и 17.02% для нейросетевой модели. Результаты прогнозирования доказывают сравнимую эффективность EMMSP и специализированной нейросетевой модели.

Реализация предложенной модели с использованием программного пакета MATLAB для прогнозирования скорости движения транспорта по дорогам г. Москвы проводилась в рамках открытого конкурса компании «Яндекс». Модель EMMSP показала сравнимую со специализированными моделями точность прогнозирования и получила оценку «Яндекса» 64.93 единиц (38 место из 191 участника). Оценка точности победителя конкурса

составила 58.92 единиц. Оценки точности прогнозирования для всех участников доступны в открытом доступе на сайте конкурса.

Проведенные эксперименты по формированию консенсус-прогноза на основании линейной комбинации двух независимых прогнозов в трех исследуемых случаях приводили к существенному повышению точности прогнозирования.

Основные результаты работы

В ходе выполнения диссертационной работы автором были получены следующие результаты.

1. Задача прогнозирования временных рядов актуальна и решается на основании модели прогнозирования. Одним из наиболее широко применяемых классов моделей прогнозирования является класс авторегрессионных моделей. Установлено, что основным недостатком данного класса является большое число свободных параметров, требующих идентификации. Определено перспективное направление развития моделей прогнозирования, позволяющее устранить указанный недостаток.

2. Разработана новая модель прогнозирования временных рядов по выборке максимального подобию для двух видов постановки задачи прогнозирования временного ряда — с учетом и без учета внешних факторов. Новая модель относится к авторегрессионному классу моделей и имеет единственный параметр, что упрощает задачу идентификации модели, устраняя основной недостаток моделей данного класса.

3. Разработан новый метод прогнозирования на основе предложенной модели, содержащий набор алгоритмов для экстраполяции временных рядов, идентификации модели и построения доверительного интервала прогнозных значений. Произведена оценка времени последовательных вычислений при решении задач экстраполяции временного ряда и идентификации модели. Предложена схема параллельных вычислений, позволяющая сократить время расчета при решении задачи идентификации.

4. Выполнена программная реализация разработанных алгоритмов средствами математического пакета MATLAB. По заказу компании «РусПауэр» разработано специализированное серверное приложение для прогнозирования показателей энергорынка РФ на ежедневной основе, работающее в автоматическом режиме и предоставляющее прогнозные значения показателей без вмешательства эксперта.

5. Произведена оценка эффективности новой модели прогнозирования. Ее применение для прогнозирования временных рядов энергорынка РФ показало высокую эффективность предложенной модели. Применение новой

модели для прогнозирования временных рядов уровня сахара крови больных сахарным диабетом первого типа и скорости движения транспорта по дорогам г. Москвы показали эффективность, сравнимую с эффективностью специализированных моделей для данных областей.

Основные результаты диссертации опубликованы в работах

1. Чучуева И.А. Модель экстраполяции по максимуму подобия (ЭМП) для временных рядов цен и объемов на рынке на сутки вперед ОРЭМ (Оптовом рынке электроэнергии и мощности) // Наука и образование [электронный ресурс]. 2010. № 1. URL: <http://technomag.edu.ru/doc/135870.html> (дата обращения 28.08.2011).

2. Чернецов С. А., Чучуева И. А. Прогнозирование уровня глюкозы в крови больных инсулинозависимым диабетом нейронными сетями и методом экстраполяции по выборке максимального подобия // Наука и образование [электронный ресурс]. 2010. №11. URL: <http://technomag.edu.ru/doc/162847.html> (дата обращения 28.08.2011).

3. Павлов Ю. Н., Чучуева И. А. Экстраполяция псевдослучайных процессов по максимуму подобия // Наука и образование [электронный ресурс]. 2009. №7. URL: <http://technomag.edu.ru/doc/129712.html> (дата обращения 28.08.2011).

4. Pavlov J. N., Chuchueva I. A. Extrapolation of pseudorandom number sequence on maximum likeness // Наука и образование [электронный ресурс]. 2009. №7. URL: <http://technomag.edu.ru/en/doc/129712.html> (дата обращения 28.08.2011).

5. Чучуева И. А. Модель экстраполяции временных рядов по выборке максимального подобия // Информационные технологии. 2010. №12. С. 43 – 47.

6. Чучуева И. А., Павлов Ю. Н. Сезонно-регрессионная модель прогнозирования в решении задачи прогнозирования цен РСВ (рынок на сутки вперед) // Энерго-Info. 2009. №4. С. 46 – 49.

7. Чучуева И. А. Прогнозирование временных рядов при помощи модели экстраполяции по выборке максимального подобия // Наука и современность: сборник материалов Международной научно-практической конференции. Новосибирск, 2010. С. 187 – 192.

8. Chuchueva I. The time series extrapolation model based on maximum likeness set // Математическое моделирование социальной и экономической динамики: труды III Международной конференции. М., 2010. С. 281–283.