

На правах рукописи

ЕВСЕЕВА Анна Владимировна

ГИБРИДНЫЙ АЛГОРИТМ ПРОГНОЗИРОВАНИЯ
И КОМПЛЕКС РЕАЛИЗУЮЩИХ ЕГО ПРОГРАММ

Специальность 05.13.18 – Математическое моделирование,
численные методы и комплексы
программ

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата технических наук

Иваново 2011

Работа выполнена в ФГБОУВПО «Ивановский государственный энергетический университет имени В. И. Ленина»

Научный руководитель: доктор физико-математических наук, профессор Ясинский Федор Николаевич

Официальные оппоненты: доктор технических наук, профессор Староверов Борис Александрович

доктор технических наук, профессор
Косяков Сергей Витальевич

Ведущая организация: ОАО «Зарубежэнергопроект», г. Иваново

Защита диссертации состоится 2 марта 2012 г. в 11 часов в ауд. Б-237 на заседании диссертационного совета Д 212.064.03 при ИГЭУ (153003, г. Иваново, ул. Рабфаковская, 34).

Отзывы на автореферат в двух экземплярах, заверенные печатью, просим присылать по адресу: 153003, г. Иваново, ул. Рабфаковская, 34, Учёный совет ИГЭУ. Тел.: (4932) 38-57-12, 26-98-61, факс: (4932) 38-57-01. E-mail: uch_sovet@ispu.ru

С диссертацией можно ознакомиться в библиотеке ИГЭУ.

Автореферат разослан « ___ » _____ 201__ г.

Ученый секретарь
диссертационного совета

А.А. Шульпин

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы

Построение точного прогноза для некоторого наблюдаемого процесса или явления – задача, нахождение решения которой является заветной целью для многих исследователей в самых разных областях человеческой деятельности. Независимо от предметной области (медицина, рынок услуг или технологическое производство) получение оптимального с точки зрения задаваемого критерия прогноза всегда принесет желаемую выгоду. В одном случае она может быть выражена в сэкономленных денежных средствах компании, в другом – в спасенных человеческих жизнях.

При всем многообразии существующих методов прогнозирования каждый из них решает задачу определения состояния объекта в будущем по значениям некоторых показателей, известных в настоящем. Так, например, в метеорологии прогнозированию подвергаются атмосферные явления, в гидрологии – паводки, в техносфере – режимы работы механизмов, в экономике – финансовые показатели, в энергетике – потребление ресурсов.

Результат прогнозирования зависит от многих параметров, таких как объем необходимой статистической информации, ее качество и достоверность, точность анализа предмета исследования, корректность в постановке задачи прогнозирования, доступность необходимых вычислительных средств. Одним из определяющих факторов также является обоснованный выбор необходимого метода прогнозирования.

Основы теории построения прогноза с помощью различных подходов заложены в работах Н. Винера, Л. Фогеля, А. Айфичера, Г. Лэма, Ф. Уоссермана, А.Я. Городецкого, В.А. Головки, В.А. Бесекерского, А.Б. Сергиенко и др.

Значительный вклад в решение вопросов прогнозирования на основе анализа временных рядов внесли Б.А. Староверов, Ф.Н. Ясинский, Б.И. Макоклюев, А.И. Орлов, В.Г. Царегородцев.

На сегодняшний день рынок программных продуктов, позволяющих строить прогноз с учетом тех или иных требований, достаточно разнообразен. В качестве примера можно привести некоторые из числа самых распространенных программ как универсальных: Microsoft Excel, Statistica, ForecastPro, Simulink, SPSS Statistics – так и специальных: NeuroShell Trader, Энергостат, IBM Traffic Prediction Tool, Stepwise Linear Search.

Несмотря на многочисленные разработки, которые ведутся в настоящее время в данном направлении, проблема продолжает быть актуальной по ряду причин. Во-первых, у каждого алгоритма есть как свои достоинства, так и недостатки. В зависимости от поставленной задачи не всегда является возможным применение одного из уже имеющихся подходов, без учета наличия его слабых сторон. Во-вторых, для каждой конкретной практической задачи характерно наличие своей специфики, что зачастую приводит к необходимости разработки алгоритма для отдельно взятой проблемы в рассматриваемой предметной области.

Цель работы

Цель проводимого исследования – разработка нового гибридного алгоритма для построения прогноза на основе анализа временных рядов, позволяющего повысить точность вычислений, а также разработка комплекса программ для решения задач прогнозирования с его использованием.

Задачи исследования

В соответствии с целью диссертации были **поставлены и решены следующие задачи:**

1. Выбрать и программно реализовать оптимальные подходы для решения поставленной задачи.
2. Разработать и реализовать программно гибридный метод расчета прогнозных значений на основе выбранных подходов.
3. Разработать методику оптимизации вычислительного процесса, позволяющую повысить точность решения данной задачи гибридным методом.
4. Разработать стратегию распараллеливания для решения поставленной задачи с учетом предложенной методики оптимизации на суперкомпьютере с графическими ускорителями.
5. Разработать программный комплекс для использования разработанных методов при решении задачи прогнозирования.
6. Провести апробацию разработанной системы на примере решения задачи прогнозирования электропотребления.

Объект исследования – временные ряды с наличием периодической составляющей в характере поведения.

Предмет исследования – математические методы построения прогноза, базирующиеся на основе анализа временных рядов.

Методы исследования

Для решения поставленных задач были использованы методы математического моделирования и прогнозирования временных рядов, понятия и методы теории нейронных сетей, основы теории фильтрации, методы эволюционного моделирования, методы вычислительной математики, теория параллельных вычислений.

Научная новизна работы

1. Разработан принципиально новый гибридный алгоритм, объединяющий результаты прогнозирования трех методов: фильтра Винера, нейросетевого подхода и эволюционного моделирования, позволивший повысить точность прогнозирования.
2. Составлена и решена с помощью метода наименьших квадратов и метода градиентного спуска задача оптимизации для нахождения оптимального набора коэффициентов, позволяющая максимально повысить точность прогнозирования.

3. С использованием технологии CUDA на суперкомпьютере с графическими ускорителями выполнена параллельная реализация базовых методов, а также самого гибридного алгоритма для уменьшения времени построения прогноза.

Обоснованность научных положений

Обоснованность основных научных положений и выводов обеспечивается корректным использованием методов математического моделирования и прогнозирования временных рядов, теории искусственных нейронных сетей и теории фильтрации, методов эволюционного моделирования и вычислительной математики.

Достоверность результатов

Достоверность полученных результатов компьютерного моделирования подтверждается их согласованностью с фактическими данными.

Практическая ценность работы

Практическая значимость диссертационной работы заключается в создании методики, позволяющей повысить эффективность существующих подходов прогнозирования на основе анализа временных рядов.

Разработанный программный комплекс может быть использован для решения задачи прогнозирования на сутки вперед для показателей потребления электроэнергии, что подтверждается актом внедрения результатов исследования в деятельность Ивановской энергосбытовой компании.

Указанная методика может быть использована при построении прогноза для временных рядов показателей из других предметных областей, отличающихся наличием периодической составляющей в характере поведения. В частности, для прогнозирования вероятности возникновения лесных пожаров, о чем свидетельствует акт внедрения результатов в процесс выполнения комплексного проекта по созданию высокотехнологичного производства «Разработка программно-технического комплекса обнаружения и прогнозирования крупномасштабных природных пожаров» (2010-218-02-139) при поддержке Министерства образования и науки РФ (ГК № 13.G25.31.0077) сотрудниками ГОУ ВПО «Ивановский институт ГПС МЧС России».

Написаны программы для построения прогноза с помощью следующих методов: фильтра Винера, нейронной сети, эволюционного моделирования, гибридного алгоритма. Разработаны параллельные версии указанных методов. Представленный гибридный алгоритм и его параллельная версия зарегистрированы в Федеральной службе по интеллектуальной собственности, патентам и товарным знакам. Получены два свидетельства о регистрации программ для ЭВМ № 2011617027 и № 2011616154.

Использование в учебном процессе

Разработанная методика и программный комплекс используются в учебном процессе кафедры высокопроизводительных вычислительных систем «Ивановского государственного энергетического университета имени В.И. Ленина», что подтверждается справкой о внедрении результатов исследования в учебный процесс.

Апробация работы

Основные положения диссертационной работы докладывались и обсуждались на IX международной научно-практической конференции «Исследование, разработка и применение высоких технологий в промышленности» (Санкт-Петербург, 2010 г.), X международной конференции «Высокопроизводительные параллельные вычисления на кластерных системах» (Пермь, 2010 г.), V международной научно-практической конференции «Пожарная и аварийная безопасность» (Иваново, 2010 г.), межвузовской научно-технической конференции аспирантов и студентов «Молодые ученые – развитию текстильной и легкой промышленности» ПОИСК-2011 (Иваново, 2011 г.), на расширенном заседании кафедры автоматики и микропроцессорной техники Костромского государственного технологического университета (Кострома, 2011 г.).

Публикации

По материалам диссертации опубликовано 11 печатных работ, в том числе четыре статьи в изданиях, рекомендованных ВАК министерства образования и науки РФ, два свидетельства о регистрации программ для ЭВМ.

Структура и объём работы

Работа состоит из введения, четырех глав, заключения, библиографического списка, включающего 146 наименований, и четырех приложений. Полный объем диссертации составляет 173 страницы, включая 29 рисунков и 13 таблиц.

СОДЕРЖАНИЕ РАБОТЫ

Введение

Во введении обосновывается актуальность исследования, ставятся цели и задачи исследования, раскрывается его научная новизна и практическая ценность, излагаются основные положения, выносимые на защиту.

В первой главе сформулированы основные понятия и терминология. Представлен обзор современных методик прогнозирования. Поставлена задача, которая будет решаться на протяжении последующих глав.

Математически задача прогнозирования имела следующую формулировку.

Пусть значения случайного вектора \vec{X} размерности k известны, а случайная величина Y недоступна для наблюдения. Задача состоит в том, чтобы определить значение Y по \vec{X} .

Любая функция ϕ , заданная на k -мерном пространстве, принимающая действительные значения и такая, что вместо Y возможно использование $\phi(\vec{X})$, называется предиктором Y по \vec{X} .

Таким образом, решить задачу прогнозирования – определить вид предиктора, позволяющего предсказывать значение случайной величины Y по \vec{X} с отклонением из допустимого интервала.

Поиск решения столь актуальной проблемы нашел свое отражение в работах многих авторов. Каждый из них выбирает свой подход к поставленной задаче. Однако проведенный анализ методов решения поставленной задачи: экстраполяция и тренды, метод экспоненциального сглаживания, метод скользящего среднего, регрессионный анализ, искусственные нейронные сети, экспертные системы, метод нечеткой логики – свидетельствует о том, что несмотря на все многообразие, каждый из рассматриваемых подходов в конечном итоге сводится к использованию и обработке временных рядов.

В терминологии временных рядов задача построения прогноза приобрела следующую формулировку.

Временной ряд $X = x(t_1), x(t_2), x(t_3), \dots, x(t_N)$ обозначим $X_t^N = x_1, x_2, x_3, \dots, x_N$. Набор последовательных значений $X_t^M = x_i, x_{i+1}, x_{i+2}, \dots, x_{i+M-1}$, лежащих внутри временного ряда X_t^N , назовем выборкой из этого ряда длины M с моментом начала отсчета t_i , $M \in [1, N-1]$, $i \in [1, N-M]$. Разность начала отсчетов выборок X_t^M , X_{t-k}^M назовем задержкой k , $k \in [1, t-1]$. Тогда задача прогнозирования на k тактов вперед заключается в построении выборки X_t^M на основании значений элементов выборки с задержкой k X_{t-k}^M .

В результате проведенного анализа было решено для повышения точности прогнозируемых значений разработать гибридный алгоритм на базе уже существующих. Для его создания предлагается объединение методов, базирующихся на анализе временных рядов.

В качестве способов обработки временных рядов решено использовать три алгоритма. Выбор первого метода – фильтра Винера – обусловлен наличием влияния случайных составляющих на характер изменения показателей временного ряда с течением времени и необходимости разделения искомым показателей и шума. Второй – нейросетевой подход – позволяет учесть нелинейность наблюдаемых процессов и сложный характер внутренних закономерностей между элементами, составляющими временной ряд. Третий подход – эволюционное моделирование – призван решить задачу оптимизации.

Во второй главе изложены теоретические основы и методология трех базовых методов (фильтрации Винера, нейронной сети и эволюционного моделирования). Приведен алгоритм решения задачи прогнозирования с помощью указанных методов. Предложен гибридный алгоритм построения прогноза.

Первым из трех выбранных базовых методов исследован метод прогнозирования с помощью алгоритмов теории фильтрации.

В общем виде фильтр – это система, которая характеризуется набором пар функций типа вход – выход (рис. 1):

$$y(t) = h(t - \tau) \otimes x(\tau), \quad (1)$$

где $y(t)$ – выходной сигнал; \otimes – обозначение свёртки; $x(\tau)$ – входной сигнал.

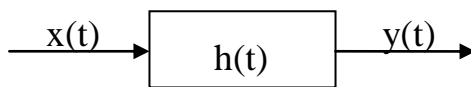


Рис. 1. Фильтр с заданными характеристиками типа вход-выход

Функция $h(t)$ – импульсный отклик фильтра (функция отклика) – связывает вход и выход фильтра по реакции на единичный импульсный входной сигнал и однозначно определяется оператором преобразования фильтра.

В случае предсказывающего фильтра: $x(t)$ – сигнал на входе фильтра, $x(t + \Delta t) = h(t - \tau) \otimes x(\tau)$ – сигнал на выходе фильтра, $\Delta t > 0$.

Задача прогнозирования в данном случае свелась к нахождению функции отклика фильтра $h(t)$, оптимальной по критерию минимума среднеквадратичной ошибки ε^2 :

$$\varepsilon^2 = \overline{[x(t) - x(t + \Delta t)]^2}. \quad (2)$$

После того, как коэффициенты фильтра будут вычислены, необходимый прогноз может быть получен путем подачи на вход фильтра следующей последовательности величин временного ряда.

Далее был реализован метод нейросетевого программирования.

Для решения задачи прогнозирования был использован двухслойный персептрон с множеством входов и множеством выходов (рис. 2).

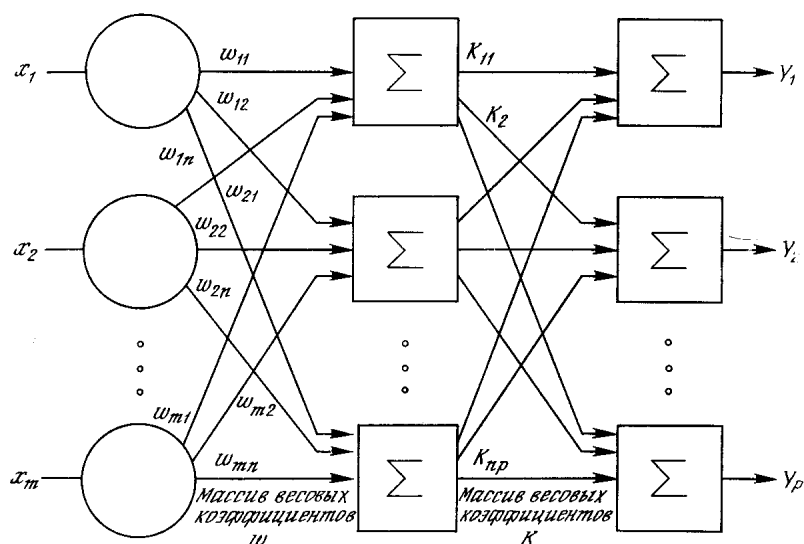


Рис. 2. Двухслойная нейронная сеть

Общее количество нейронов, а также входов сети были подобраны экспериментально. Количество входов сети зависело от данных о показателях за прошлый период, на основании которых строился прогноз. Количество выходов сети совпадало с числом периодов, для которых должны быть получены

прогнозные значения. В данном случае, если на вход была подана последовательность случайной величины X в моменты $t_0, t_1, t_2, \dots, t_k$, то значение на выходе должно было быть равно $X(t_{k+1}), X(t_{k+2}), X(t_{k+3}), \dots, X(t_{k+p})$, где p – период упреждения.

Для решения поставленной задачи прогнозирования в качестве активационной была применена сигмоидная функция:

$$y = \frac{1}{1 + e^{-cs}}, \quad (3)$$

где $c > 0$ – коэффициент, характеризующий ширину сигмоидной функции по оси абсцисс.

Её выбор обусловлен, в первую очередь, простотой вычисления производной и, как следствие, высокой скоростью обучения сети.

Благодаря непрерывности первой производной сигмоидной функции активации возможно обучение сети градиентными методами. Решено было использовать один из методов данного вида – метод обратного распространения ошибки.

Последним из трех базовых подходов был исследован и реализован метод эволюционного моделирования.

Для решения задачи методом эволюционного моделирования дискретный временной ряд случайной величины $X(t)$ был заменен на дискретный временной ряд случайной величины $Y(t)$, элементами которого будут приращения между двумя последовательными величинами ряда $X(t)$, то есть $Y(t_k) = X(t_{k+1}) - X(t_k)$. Таким образом, величина $Y(t)$ характеризует поведение величины $X(t)$, а именно ее возрастание или убывание. Тогда последовательность символов, составляющих историю функционирования автомата, будут представлять элементы временного ряда случайной величины $Y(t)$.

После рассмотрения трех базовых подходов к прогнозированию был предложен алгоритм построения гибридного метода на их основе. В основу построения гибридного метода был положен следующий принцип.

При помощи определенного набора коэффициентов $\alpha_i \geq 0$, $\sum_i \alpha_i = 1$, $i = 1, 2, 3$ результаты работы трёх базовых методов были объединены в единый синтетический алгоритм (рис. 3).

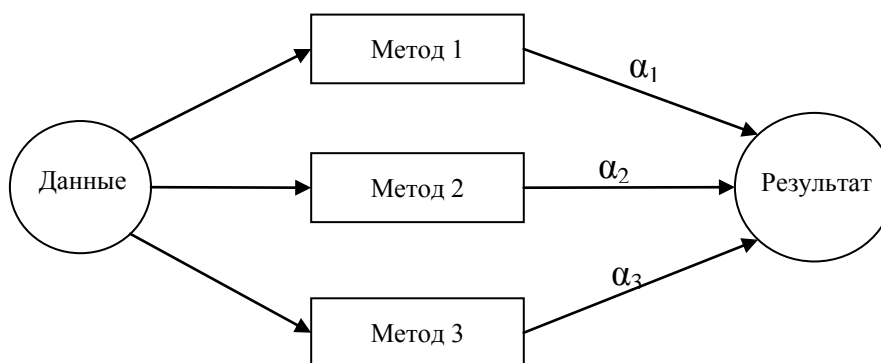


Рис. 3. Схема гибридного алгоритма

Таким образом, принцип действия гибридного алгоритма следующий:

$$p(t) = \alpha_1 \cdot f(t) + \alpha_2 \cdot n(t) + \alpha_3 \cdot m(t); \quad (4)$$
$$\alpha_i \geq 0; \quad \sum_i \alpha_i = 1; \quad i = 1, 2, 3,$$

где $p(t)$ – спрогнозированное значение наблюдаемой величины, полученное с применением гибридного алгоритма; $f(t)$ – значение прогнозируемой величины, полученное с применением фильтра Винера; $n(t)$ – значение прогнозируемой величины, полученное с применением нейронной сети; $m(t)$ – значение прогнозируемой величины, полученное с применением эволюционного моделирования.

Одной из принципиальных и сложных задач является подбор искомого оптимального набора коэффициентов α_i , $i = 1, 2, 3$ для получения наименьшей ошибки прогнозирования синтетического алгоритма.

Чтобы избежать полного перебора параметров, ставится линейная задача оптимизации для нахождения указанных коэффициентов:

$$r(t) - p(t) \rightarrow \min, \quad (5)$$

где $r(t)$ – реальное значение величины; $p(t)$ – спрогнозированное значение наблюдаемой величины, полученное с применением гибридного алгоритма. Задача оптимизации позволяет снизить количество итераций при нахождении необходимого набора коэффициентов α_i , $i = 1, 2, 3$.

Для повышения точности при решении задачи оптимизации будем применять метод наименьших квадратов.

Перейдем от непрерывного изменения времени к дискретному. Пусть в соответствии с гибридным алгоритмом получены прогнозные значения для моментов времени $t_1, t_2, t_3, \dots, t_k$.

Разность между значением наблюдаемой величины, вычисленным в результате применения метода прогнозирования, и истинным ее значением называется невязкой. В данном случае невязка будет определяться выражением $(r(t) - p(t))$.

Введём функцию $F(\alpha_1, \alpha_2, \alpha_3)$, равную сумме квадратов невязок:

$$F(\alpha_1, \alpha_2, \alpha_3) = \sum_{i=0}^n (r(t_i) - p(t_i))^2, \quad (6)$$

где k – число проведенных испытаний по построению прогноза.

Теперь задача оптимизации примет вид:

$$F(\alpha_1, \alpha_2, \alpha_3) \rightarrow \min. \quad (7)$$

Для решения поставленной задачи оптимизации предложено использовать метод градиентного спуска.

Зная коэффициенты α_i , $i = 1, 2, 3$, в дальнейшем для построения прогноза необходимо будет придерживаться следующего алгоритма:

1. Определить момент времени t_i , на который надо выполнить прогноз.

2. Получить прогнозные значения $f(t_i)$, $n(t_i)$ и $m(t_i)$, воспользовавшись базовыми методами: фильтрации Винера, нейронной сетью и эволюционным моделированием соответственно.

3. Вычислить прогнозируемое значение $p(t_i)$, полученное с помощью гибридного алгоритма, применив формулу (4).

Как было сказано выше, для решения задачи оптимизации был применен метод градиентного спуска. Выбор данного подхода обусловлен простотой производимых в данном случае вычислений. Однако данный способ обладает рядом недостатков, одним из которых является возможность попадания в локальный минимум и, как следствие, получение результата с меньшей точностью. Для преодоления подобного рода проблемы было принято решение использовать не одну, а целый ряд начальных точек, и выбор из полученных результатов того, который бы обладал минимальной погрешностью.

Было проведено тестирование трех базовых алгоритмов и разработанного гибридного подхода для построения прогноза. В качестве критерия оптимальности построенного прогноза был использован индекс Тейла (8):

$$U(X, Y) = \frac{\sqrt{\sum_{t=1}^n (X_t - Y_t)^2}}{\sqrt{\sum_{t=1}^n X_t^2 + \sum_{t=1}^n Y_t^2}}. \quad (8)$$

Индекс Тейла $U(X, Y)$ измеряет несовпадение временных рядов X_t и Y_t , и чем ближе он к нулю, тем ближе сравниваемые ряды. По результатам испытаний данный показатель при прогнозировании с помощью эволюционного моделирования составил 0,0436, фильтра Винера – 0,0329, нейронной сети – 0,0322, гибридного алгоритма – 0,0239.

Как видно из полученных результатов, разработка гибридного алгоритма позволила повысить точность при построении прогноза.

Однако использование данного подхода, а также необходимость в качестве базы для гибридного метода проводить вычисление прогнозных значений с помощью сразу трех базовых алгоритмов влекут за собой значительное увеличение временных затрат.

В третьей главе предложен метод параллельной организации с использованием системы CUDA разработанного алгоритма для уменьшения времени выполнения программы.

Для метода Винера была предложена параллельная реализация операции свертки-развертки, быстрого преобразования Фурье и предсказание будущих значений на основе рассчитанных коэффициентов и входных данных.

Обучение нейронной сети производится с помощью алгоритма обратного распространения ошибки. Данный алгоритм представляет собой набор векторных и матричных операций. Поэтому при составлении параллельной схемы для метода нейронных сетей была использована поставляемая вместе с CUDA библиотека cuRAND. Модификация функций, входящих в указанную библиотеку,

позволила реализовать эффективный алгоритм перемножения и транспонирования матриц.

Лучшим способом ускорения вычислений гибридного метода является использование нескольких графических ускорителей и многоядерного процессора. В данном подходе создаётся многопоточное окружение, при этом каждый поток связан с одним графическим ускорителем. Это позволяет использовать несколько поисковых точек для метода градиентного спуска (рис. 4).

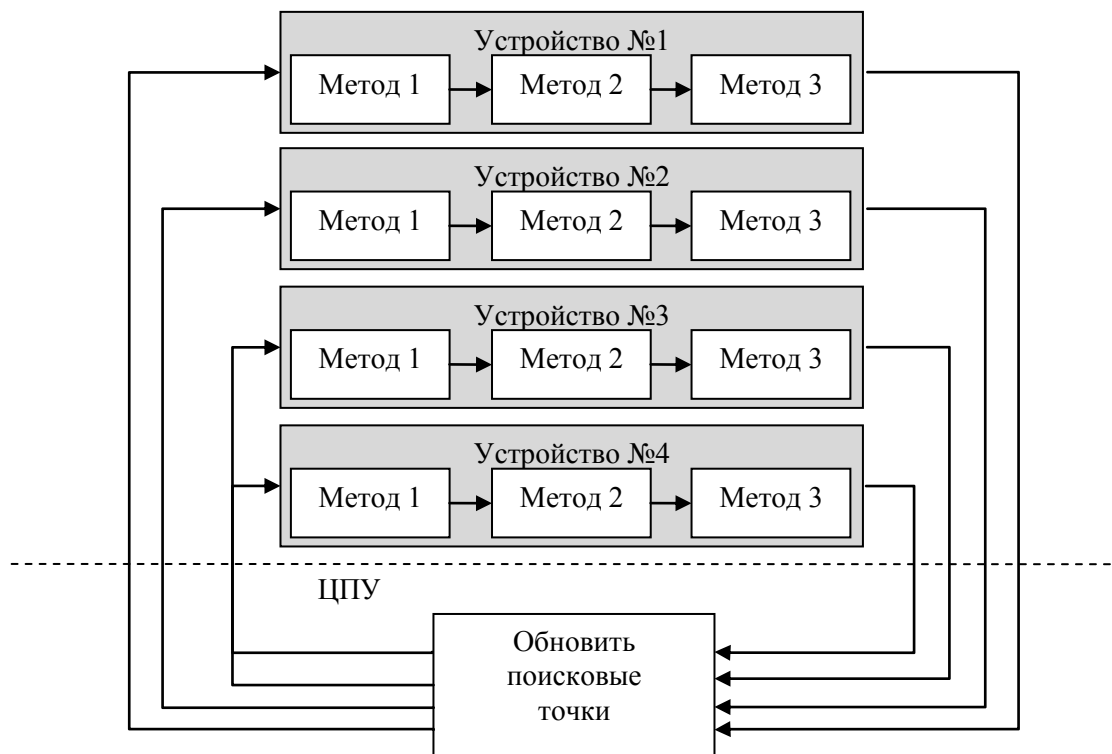


Рис. 4. Использование нескольких графических ускорителей

Каждый поток производит вычисление базовых методов на своём графическом ускорителе, а после этого находит коэффициенты гибридного метода на основе своей поисковой точки.

Четвертая глава посвящена экспериментальной проверке разработанной системы прогнозирования.

Представленная модель была применена для получения почасового прогноза на 24 часа вперед на основе данных о потреблении электроэнергии по Костромской области и г. Кострома и по г. Иваново.

Оценка полученных результатов проводилась с использованием следующих критериев.

1. Относительная ошибка прогноза:

$$\varepsilon_t = \frac{|y_t - \tilde{y}_t|}{y_t} \cdot 100\%, \quad (9)$$

где y_t – фактическое значение показателя энергопотребления на момент времени t ; \tilde{y}_t – значение показателя в момент времени t , полученное с применением прогнозной модели.

2. MAPE (the mean absolute percentage error) – средняя абсолютная ошибка в процентах:

$$MAPE = \frac{1}{Q} \sum_{i=1}^Q \frac{|y_t - \tilde{y}_t|}{y_t} \cdot 100\%, \quad (10)$$

где Q – количество значений временного ряда, попавших в контрольный период; y_t и \tilde{y}_t – соответственно фактическое и спрогнозированное значения временного ряда в момент времени t .

Костромская область и г. Кострома

Значения ошибки прогнозирования с помощью гибридного алгоритма по показателю MAPE, максимальная и минимальная относительная ошибка для каждого месяца года даны в табл. 1.

Таблица 1

Результаты прогнозирования электропотребления
Костромской области и г. Кострома

Месяц	MAPE, %			Минимальное отклонение, %	Максимальное отклонение, %
	будни	выходные	будни и выходные		
Январь	1,800	1,838	1,811	0,010	7,915
Февраль	1,579	2,650	1,885	0,022	7,249
Март	1,790	2,438	1,975	0,006	8,884
Апрель	1,564	3,008	1,977	0,008	7,860
Май	2,311	3,242	2,577	0,068	10,192
Июнь	2,386	3,004	2,606	0,010	10,303
Июль	2,373	2,561	2,427	0,030	8,011
Август	2,434	2,634	2,471	0,024	7,824
Сентябрь	2,057	2,575	2,205	0,076	9,956
Октябрь	2,012	2,347	2,122	0,001	9,242
Ноябрь	1,753	1,976	1,816	0,027	8,070
Декабрь	1,921	1,499	1,800	0,039	6,782
Среднее	1,998	2,481	2,139	0,027	8,530

Разработанный гибридный алгоритм показал высокую точность прогнозирования. Среднее значение абсолютной ошибки за месяц составило 2,139, что является высоким результатом для прогнозной модели данной предметной области. Минимальное отклонение за месяц наблюдается в декабре и январе, максимальное – в июне и августе. Среднее значение абсолютной ошибки в будни равно 1,998%, в выходные дни – 2,481%. В будни максимальное значение показателя MAPE пришлось на июнь (2,386%), минимальное – на апрель (1,564%). Максимальное значение абсолютной ошибки наблюдалось в выходные дни мая (3,242%).

Результаты прогнозирования для месяцев, в которых были показаны наименьшее и наибольшее значение показателя MAPE за месяц, наглядно представлены на рис. 5 и рис. 6 соответственно.

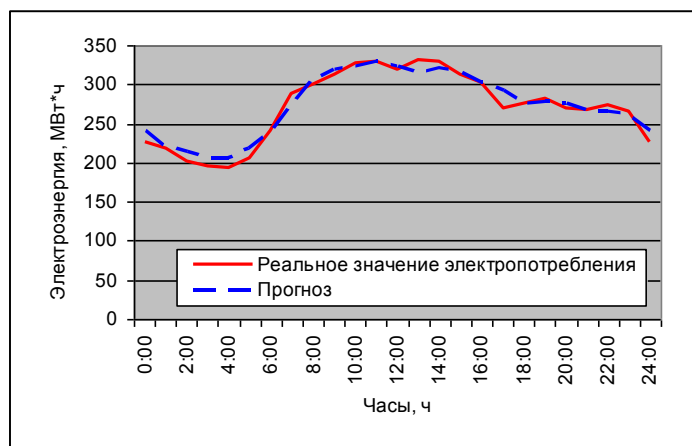


Рис. 5. Результаты прогноза электропотребления в июне (19.06.2006 г.)

Ошибка прогнозирования превышает допустимый 5%-ный интервал 7 раз из 24 предсказанных значений, что составляет около 30%. В 2%-ном интервале отклонения находятся около 50% полученных результатов.

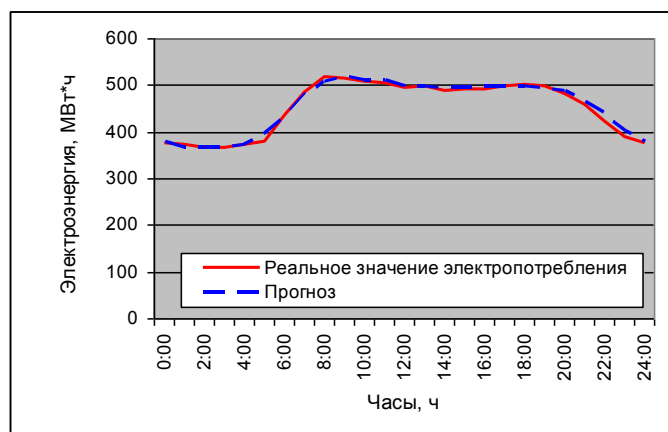


Рис. 6. Результаты прогноза электропотребления в январе (31.01.2006 г.)

Все полученные результаты попадают в 5%-й интервал ошибки прогнозирования. 2%-й порог ошибки превышен лишь для трех значений из 24-часового интервала.

Ивановская энергосбытовая компания

По Ивановской энергосбытовой компании были исследованы данные за три месяца: январь, февраль и март – для ряда фидеров, в том числе 607А, 609А, 612А и для системы в целом.

Значения ошибки прогнозирования с помощью гибридного алгоритма по показателю MAPE, максимальная и минимальная относительная ошибка для указанных месяцев даны в табл. 2.

Результаты построения прогноза для г. Иваново

Месяц	Фидер	МАРЕ, %		Минимальное отклонение, %		Максимальное отклонение, %	
		<i>будни</i>	<i>выходные</i>	<i>будни</i>	<i>выходные</i>	<i>будни</i>	<i>выходные</i>
Январь	607А	2,755	3,292	0,024	0,040	7,673	12,814
	609А	2,253	3,973	0,035	0,010	9,068	16,381
	612А	2,872	2,785	0,074	0,013	10,225	9,819
	система в целом	2,094	3,156	0,012	0,113	8,959	7,426
Февраль	607А	3,068	4,688	0,001	0,208	10,326	11,937
	609А	3,654	3,083	0,069	0,268	11,056	9,651
	612А	2,981	6,109	0,018	1,079	18,911	11,963
	система в целом	2,459	2,824	0,029	0,278	7,516	11,996
Март	607А	7,421	5,370	0,141	0,419	25,565	16,991
	609А	3,525	3,771	0,037	0,082	8,831	10,810
	612А	4,172	5,004	0,043	0,083	12,038	15,229
	система в целом	2,611	3,610	0,038	0,022	7,134	11,908

Как видно из приведенной таблицы, полученный результат удовлетворяет заявленным требованиям нахождения средней абсолютной ошибки в пределах 5%. Данный показатель превышен лишь для фидера 607А в будни и выходные дни марта, а также для фидера 612А – в выходные дни. Приведенный результат может быть объяснен, если рассмотреть изменение показателя энергопотребления, например, для фидера 607А за указанный период (рис. 7).

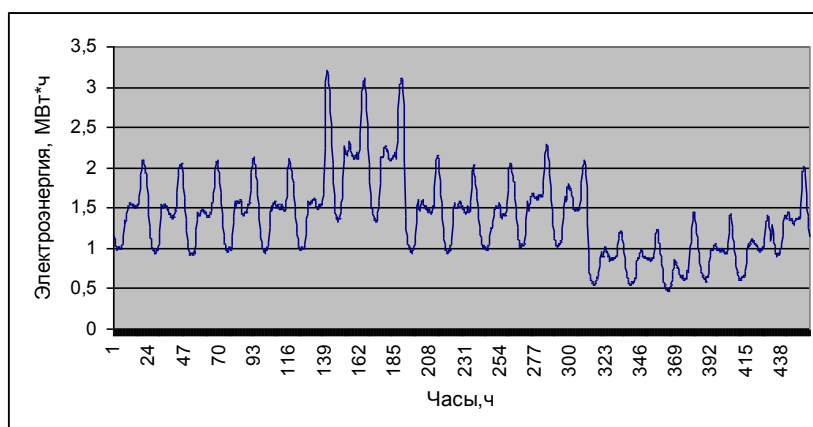


Рис. 7. График зависимости электропотребления от времени для фидера 607А

Представленный график наглядно демонстрирует чрезвычайно сложный характер наблюдаемых изменений. Минимальное значение относительной ошибки прогнозирования – 0,141%, максимальное – 25,565%, что является наибольшим для всех проанализированных данных. Показатель ошибки превысил уровень в 2% в 87% случаев, уровень в 5% – в 50%. Несомненно, представленные показатели достаточно высоки, однако, учитывая рассмотренный ха-

рактически для всего временного ряда, можно считать допустимыми полученные результаты и для данных дней.

Для дней, показавших наилучшие результаты – это январские будни, для фидера 609А показатели следующие: попадает в 2%-й интервал ошибки около 70% значений, 5%-й интервал превышен лишь для трех показателей, что является приемлемым.

Для системы в целом показатели для одного среднестатистического дня следующие (рис. 8).

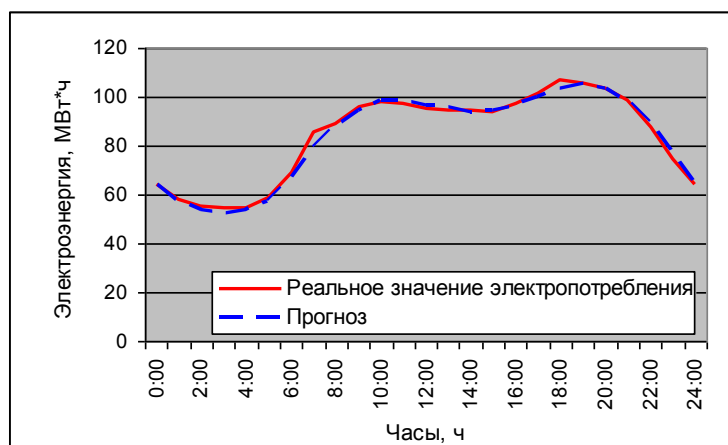


Рис. 8. График зависимости электропотребления от времени для системы в целом

В необходимый 5%-й интервал ошибки не попадает лишь одно значение из 24, 2%-й интервал ошибки превышен лишь для четырех значений. Таким образом, по Иванову для системы в целом получены столь же высокие результаты, что и для Костромской области. При этом увеличение ошибки при проведении испытаний по отдельным фидерам может быть объяснено сложным характером сложившейся энергосистемы: отдельные фидеры могут быть отключены или же переведены на другие объекты. Энергосистема в целом при этом значительных колебаний не претерпевает, что дает возможность получать прогноз с большей точностью. Однако результаты, полученные при прогнозировании временных рядов показателей энергопотребления по отдельным фидерам, могут быть использованы в качестве экспериментальных данных для временных рядов с высокой долей колебаний из других предметных областей.

В заключении подведены итоги и результаты проведенных исследований.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ И ВЫВОДЫ

1. Проведен анализ существующих на сегодняшний день методов построения прогноза. Выявлены основные закономерности при построении прогноза и наиболее перспективные направления улучшения его качества.

2. По результатам проведенного анализа выделены три базовых метода построения прогноза: фильтр Винера, нейронная сеть и эволюционное моделирование и реализован алгоритм решения задачи прогнозирования.

3. Разработан и реализован гибридный метод расчета прогнозных значений на основе вышеуказанных подходов.

4. Разработана методика оптимизации вычислительного процесса, позволившая повысить точность решения задачи гибридным методом по показателю индекса Тейла в среднем на 0,01.

5. Разработана стратегия распараллеливания для решения поставленной задачи с учетом предложенной методики оптимизации на суперкомпьютере с графическими ускорителями, позволившая сократить временные затраты в несколько раз.

6. Разработан программный комплекс для использования разработанных методов решения задачи прогнозирования.

7. Выполнено исследование разработанной системы на примере прогнозирования показателя потребления электроэнергии. Ошибка прогнозирования в среднем составила 2,1%.

Поставленные задачи диссертационного исследования успешно решены и цель достигнута.

Разработанная методика прогнозирования может быть использована для временных рядов показателей из предметных областей, отличающихся наличием периодической составляющей в характере поведения.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ РАБОТЫ

В изданиях, рекомендованных ВАК РФ:

1. Сидоров, С.Г. Анализ временных рядов как метод построения прогноза потребления электроэнергии / С.Г. Сидоров, А.В. Никологорская // Вестник ИГЭУ. – 2010. – Вып. 3. – С. 81-83.

2. Ясинский, Ф.Н. Построение гибридного метода прогнозирования энергопотребления с использованием суперкомпьютера на графических ускорителях / Ф.Н. Ясинский, А.В. Никологорская // Вестник ИГЭУ. – 2010. – Вып. 4. – С. 89-91.

3. Ясинский, Ф.Н. Прогнозирование вероятности возникновения лесных пожаров с помощью нейросетевого алгоритма на многопроцессорной вычислительной технике / Ф.Н. Ясинский, О.В. Потемкина, С.Г. Сидоров, А.В. Евсеева // Вестник ИГЭУ. – 2011. – Вып. 2. – С. 82-84.

4. Евсеева, А.В. Прогнозирование энергопотребления при помощи гибридного алгоритма с применением технологии CUDA / А.В. Евсеева // Вестник ИГЭУ. – 2011. – Вып. 3. – С. 47-49.

В прочих изданиях:

5. Никологорская, А.В. Опыт прогнозирования энергопотребления в энергосетях Костромской области / А.В. Никологорская, С.Г. Сидоров // Высокие технологии, исследования, промышленность: сб. тр. Девятой международной научно-практической конференции «Исследование, разработка и применение высоких технологий в промышленности». – СПб.: Изд-во Политехн. ун-та, 2010. – Т.4. – С. 266-270.

6. Никологорская, А.В. Временные ряды. Их анализ и прогнозирование : учеб. пособие / А.В. Никологорская, Ф.Н. Ясинский; ГОУВПО «Ивановский государственный энергетический университет имени В.И. Ленина». – Иваново, 2010. – 72 с.

7. Никологорская, А.В. Об использовании суперкомпьютера при построении гибридного метода прогноза потребления электроэнергии на основе анализа временных рядов / А.В. Никологорская, Ф.Н. Ясинский // «Высокопроизводительные параллельные вычисления на кластерных системах» (НПС-2010): материалы X международной конференции (Пермь, 1-3 ноября 2010 г.). – Пермь: Изд-во ПГТУ, 2010. – Т.2. – С. 148-153.

8. Евсеева, А.В. Разработка гибридного метода прогнозирования потребления электроэнергии с использованием суперкомпьютера / А.В. Евсеева, Ф.Н. Ясинский // Сборник трудов V Международной научно-практической конференции «Пожарная и аварийная безопасность». – Иваново, 2010. – Ч.2. – С. 171-172.

9. Евсеева, А.В. Использование суперкомпьютера с графическими ускорителями для решения задачи программирования на основе анализа временных рядов / А.В. Евсеева, Е.Л. Никологорская // Молодые ученые – развитию текстильной и легкой промышленности (ПОИСК – 2011): сб. материалов межвузовской научно-технической конференции аспирантов и студентов / ИГТА. – Иваново, 2011. – Ч.2. – С. 332-333.

Свидетельства на программные продукты:

10. Евсеева, А.В. Технология CUDA для системы прогнозирования энергопотребления «PowerSystemLoadForecastingCUDA» / А.В. Евсеева // Свидетельство об официальной регистрации программы для ЭВМ № 2011617027. Зарегистрировано в Реестре программ для ЭВМ 12.07.2011.

11. Евсеева, А.В. Система прогнозирования энергопотребления «PowerSystemLoadForecasting» / А.В. Евсеева // Свидетельство об официальной регистрации программы для ЭВМ № 2011616154. Зарегистрировано в Реестре программ для ЭВМ 05.08.2011.

