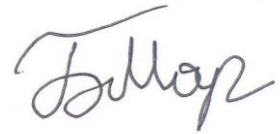


На правах рукописи



МАРТЫНЕНКОВ БОРИС ВИТАЛЬЕВИЧ

**МОДЕЛИ И АЛГОРИТМЫ ИНТЕЛЛЕКТУАЛЬНОЙ
ПОДДЕРЖКИ УПРАВЛЕНИЯ РАБОЧЕЙ НАГРУЗКОЙ
СИСТЕМ ОБРАБОТКИ ИНФОРМАЦИИ НА ОСНОВЕ
РЕТРОСПЕКТИВНЫХ ДАННЫХ**

Специальность 2.3.1. Системный анализ, управление
и обработка информации, статистика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Москва – 2026

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего образования «Воронежский государственный технический университет», г. Воронеж.

Научный руководитель: Цветков Александр Васильевич, доктор технических наук, доцент, профессор базовой кафедры кибернетики в системах организационного управления Федеральное государственное бюджетное образовательное учреждение высшего образования «Воронежский государственный технический университет», г. Воронеж.

Официальные оппоненты:

1) Рытов Михаил Юрьевич, доктор технических наук, доцент, заведующий кафедрой «Системы информационной безопасности» факультета информационных технологий Федерального государственного бюджетного образовательного учреждения высшего образования «Брянский государственный технический университет», г. Брянск.

2) Куцакин Максим Алексеевич, кандидат технических наук, старший преподаватель кафедры «Информатика и вычислительная техника» Федерального государственного казенного военного образовательного учреждения высшего образования «Академия Федеральной службы охраны Российской Федерации», г. Орёл.

Ведущая организация:

Федеральное государственное бюджетное образовательное учреждение высшего образования «Дагестанский государственный технический университет», Республика Дагестан, г. Махачкала.

Защита диссертации состоится «20» мая 2026 г. в 12:00 часов на заседании диссертационного совета 75.1.026.01, созданного на базе АО «Научно-исследовательский институт вычислительных комплексов им. М.А. Карцева», по адресу: 117437, г. Москва, ул. Профсоюзная, д. 108.

С диссертацией можно ознакомиться в библиотеке и на сайте АО «Научно-исследовательский институт вычислительных комплексов им. М.А. Карцева» <https://niivk.ru/>.

Автореферат разослан «___» _____ 2026 г.

Ученый секретарь
диссертационного совета 75.1.026.01
канд. техн. наук, доцент



Е.И. Мутина

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Развитие и совершенствование распределенных систем обработки информации, а также активное развитие систем искусственного интеллекта стало основой для разработки и широкого использования центров обработки данных (ЦОД), в частности, виртуализированных ЦОД (ВЦОД), инфраструктура которых основана на парадигме виртуализации вычислительных ресурсов в составе виртуальных машин и/или контейнеров.

Важной задачей управления ВЦОД является динамическая реконфигурация их инфраструктуры в зависимости от текущей рабочей нагрузки. Показатели рабочей нагрузки могут носить, как периодический (сезонный: время суток, время года), так и случайный (реакция на события и т.д.) характер. В общем случае можно говорить об управлении рабочей нагрузкой ВЦОД в нормальных и специальных условиях их функционирования.

Процесс динамической реконфигурации является задачей оптимизации и основан на анализе результатов мониторинга использования (утилизации) вычислительных ресурсов ВЦОД, которые сохраняются и именуется ретроспективной информацией о рабочей нагрузке. Эта информация позволяет реализовать функцию проактивного управления, которая базируется в классе задач прогнозирования рабочей нагрузки. Одной из проблем эффективного решения задачи прогнозирования рабочей нагрузки является «зашумление» значений временного ряда ретроспективной информации. Источниками зашумления могут выступать, как сами приложения или сервисы, так и другие приложения и сервисы, функционирующие в составе ВЦОД.

Известные подходы к решению задачи прогнозирования рабочей нагрузки ВЦОД не в полной мере рассматривают возможные условия ее зашумления, а также модели, методы и алгоритмы снижения влияния шумовых факторов на значения показателей временного ряда ретроспективных данных.

Таким образом, актуальность исследования обосновывается необходимостью разработки и исследования новых эффективных моделей и алгоритмов прогнозирования рабочей нагрузки ВЦОД на основе ретроспективной информации с учетом ее зашумления.

Тематика диссертационной работы соответствует научному направлению ФГБОУ ВО «Воронежский государственный технический университет» «Вычислительные комплексы и проблемно-ориентированные системы управления».

Степень разработанности темы. Фундаментальными исследованиями в теории прогнозирования временных рядов методами статистического и машинного обучения посвящены работы Дж. Бокса (George Box), Г. Дженкинса (Gwilym Jenkins), Я. Лекуна (Yann LeCun), З. Хохрайтера (Sepp Hochreiter), Ю. Шмидхубера (Jürgen Schmidhuber). Теоретическими и прикладными исследованиям обработки зашумленных сигналов в различных предметных областях посвящены работы Д. Гильберта (David Hilbert),

Н. Хуанга (N. Huang), Г. де Мораеса (G. de Moraes), М.Д. Сенюка, С.А. Ерошенко, Х. Джанга (H. Zhang).

Задача прогнозирования рабочей нагрузки на основе анализа значений временного ряда ее ретроспективной информации в указанных исследованиях рассматривается в основном для контролируемых условий функционирования ВЦОД и носит узкий характер за счет получения специальных решений, или рассматривается в общем случае, что не позволяет получить эффективные модели и алгоритмы принятия решения.

Вопросы разработки моделей и алгоритмов снижения влияния факторов зашумления на значения временного ряда рабочей нагрузки ВЦОД для решения задачи ее прогнозирования в рассматриваемых исследованиях, либо выносятся в область ограничений, либо рассматриваются недостаточно полно. В связи с этим предлагается разработка моделей и алгоритмов предварительной обработки зашумленного временного ряда ретроспективной информации рабочей нагрузки, а также моделей и алгоритмов машинного обучения для решения задачи прогнозирования рабочей нагрузки на основе предварительно обработанного временного ряда.

Объектом исследования является временной ряд ретроспективных данных рабочей нагрузки системы обработки информации в условиях влияния факторов зашумления и процедура получения его прогнозных значений.

Предметом исследования являются модели представления временных рядов, алгоритмы их разложения, процедуры машинного обучения для получения их прогнозных значений.

Цель и задачи исследования. Целью диссертационного исследования является разработка модели временного ряда рабочей нагрузки и алгоритмов, снижающих факторные зашумления, а также моделей, алгоритмов и специального программного обеспечения управления системой обработки информации для прогнозирования рабочей нагрузки.

Для достижения поставленной цели необходимо решить следующие частные научные задачи:

1. Разработать модель модовой декомпозиции временного ряда рабочей нагрузки системы обработки информации, на основе анализа формирования временных рядов ее ретроспективной информации, влияния на нее факторов зашумления и существующих методов модовой декомпозиции сигналов.

2. Создать комплексный алгоритм предварительной обработки временного ряда рабочей нагрузки на основе анализа способов интеграции эмпирического и вариационного методов декомпозиции сигнала.

3. Разработать гибридный алгоритм системы управления прогнозированием временного ряда рабочей нагрузки, на основе анализа методов глубокого машинного обучения, поддерживающих анализ временных рядов.

4. Модифицировать существующую архитектуру системы управления прогнозированием рабочей нагрузки на основе анализа функциональных возможностей существующих фреймворков модовой декомпозиции сигналов и гибридных сетей глубокого машинного обучения.

Методология и методы исследования. Для решения поставленных задач используются методы математической статистики, модовой декомпозиции сигналов, методы искусственного интеллекта, теория прогнозирования, теория сложных систем. Общей методологической основой является системный подход.

Научная новизна. В диссертации получены следующие результаты, характеризующиеся научной новизной:

– модель модовой декомпозиции временного ряда рабочей нагрузки, отличающаяся совместным использованием эмпирического и вариационного подходов получения множества колебательных модовых функций и обеспечивающая снижение влияния факторов зашумления на значения временного ряда;

– комплексный алгоритм предварительной обработки временного ряда рабочей нагрузки, отличающийся наличием этапа вторичной вариационной модовой декомпозиции базовой колебательной модовой функции, полученной методом эмпирической модовой декомпозиции, и позволяющий формировать множества обучающей и тестовой выборки для системы управления прогнозированием значений временного ряда на основе методов глубокого обучения;

– гибридный алгоритм прогнозирования временного ряда рабочей нагрузки для системы глубокого обучения, отличающийся ансамблевым способом выделения значимых признаков шаблонов рабочей нагрузки и каскадным режимом ее прогнозирования, обеспечивающий получение разномасштабных прогнозных значений временного ряда рабочей нагрузки;

– архитектура системы управления прогнозированием рабочей нагрузки, отличающаяся интеграцией модуля предварительной обработки временного ряда рабочей нагрузки и обеспечивающая прогнозирование рабочей нагрузки с пригодной точностью прогноза.

Теоретическая значимость заключается в том, что предлагаемые новые подходы к предварительной обработке временного ряда рабочей нагрузки и разработанные алгоритмы прогнозирования его значений, основанные на методах модовой декомпозиции и глубокого обучения, могут быть применены при формировании экспериментальных моделей временных рядов в других предметных областях и для разработки программных комплексов их прогнозирования.

Практическая значимость работы заключается в повышении точности прогноза значений временного ряда, необходимого в различных областях человеческой деятельности. Разработано специальное программное обеспечение, позволяющее осуществлять прогнозирование временного ряда рабочей нагрузки (на примере использования процессорных ядер). Предложены рекомендации по исследованию ретроспективной информации временных рядов вычислительных ресурсов систем обработки информации для решения задачи прогнозирования шаблонных типов рабочей нагрузки.

Достоверность результатов подтверждается использованием при разработке моделей известных математических методов и результатами вычислительных экспериментов.

Положения, выносимые на защиту:

1. Модель модовой декомпозиции временного ряда обеспечивает снижение влияния факторов зашумления на значения временного ряда.

2. Комплексный алгоритм предварительной обработки временного ряда рабочей нагрузки обеспечивает формирование множества обучающих и тестовых выборок для системы прогнозирования элементов временного ряда на основе методов глубокого обучения

3. Гибридный алгоритм прогнозирования временного ряда рабочей нагрузки для системы глубокого обучения обеспечивает получение разномасштабных прогнозных значений временного ряда рабочей нагрузки.

4. Архитектура системы прогнозирования рабочей нагрузки, обеспечивает прогнозирование рабочей нагрузки с пригодной точностью прогноза.

Апробация работы. Основные результаты диссертационной работы докладывались и обсуждались на следующих конференциях и семинарах: Международная научно-практическая конференция «Общество – Наука Инновации (Уфа, 2021), Материалы круглого стола «Хроники цифровых трансформаций» (Волгоград, 2022), III ежегодная национальная научно-практическая конференция «Кибербезопасность: технические и правовые аспекты защиты информации» (Москва, 2024), Международная научно-практическая конференция «Стратегии успеха: инновационные методы, технологии и практики в науке для достижения глобального прогресса» (Уфа, 2025), XXXI International Open Science Conference «Modern informatization problems (Yelm, WA, USA, 2026), а также на научных семинарах кафедры кибернетики в системах организационного управления ВГТУ (2023-2026 гг.).

Реализация и внедрение результатов работы. Результаты диссертации внедрены в практическую деятельность ФГУП «Научно-технический центр «Орион» (г. Москва) и ООО «Научно-технический центр «Разработка сложных систем» (г. Орёл), а также в образовательный процесс Московского технического университета связи и информатики (МТУСИ).

Соответствие паспорту специальности. Содержание диссертации соответствует п. 4 «Разработка методов и алгоритмов решения задач системного анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта», п. 5 «Разработка специального математического и алгоритмического обеспечения систем анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта», п. 6 «Методы идентификации систем управления на основе ретроспективной, текущей и экспертной информации» паспорта специальности 2.3.1. Системный анализ, управление и обработка информации, статистика.

Публикации. По материалам диссертации опубликовано 24 печатные работы, в том числе 4 статьи опубликовано в изданиях, рекомендованных ВАК

при Минобрнауки России, 1 статья в издании, индексируемом в WoS, 12 свидетельств о регистрации программ для ЭВМ.

В работах, опубликованных в соавторстве и лично соискателем предложены: [1, 3, 6-8, 16, 23] – подход к предварительной обработке ретроспективных данных об уровне загрузки вычислительных ресурсов; [2, 4, 5, 9, 12, 13, 21] – гибридной модели глубокого обучения для прогнозирования рабочей нагрузки в виртуализированных центрах обработки данных; [10, 11, 14, 15, 22] – алгоритмы системы прогнозирования рабочей нагрузки, функционирующей в условиях зашумления ее ретроспективных данных; [18, 19, 20, 24] – архитектура системы прогнозирования рабочей нагрузки; [17] – процедура интеграции модуля предварительной обработки временного ряда рабочей нагрузки в систему прогнозирования рабочей нагрузки.

Структура и объем работы. Диссертация состоит из введения, четырех глав, заключения, списка литературы из 130 наименований. Работа изложена на 154 страницах машинописного текста (основной текст занимает 125 страниц, содержит 71 рисунок, 4 таблицы).

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертационного исследования, сформулированы цель и частные задачи, показана научная новизна и практическая значимость полученных результатов, выносимых на защиту.

В первой главе решается задача разработки модели модовой декомпозиции временного ряда рабочей нагрузки. На основании анализа исследований, посвященных организации и функционированию службы администрирования ВЦОД: рассмотрена актуальность задач реактивного и проактивного управления рабочей нагрузкой ВЦОД; определена структурная схема подсистемы мониторинга вычислительных ресурсов, отображающих рабочую нагрузку, методы и средства их анализа (рисунок 1); рассмотрен подход к сохранению ретроспективных данных рабочей нагрузки в виде временных рядов значений заданных показателей утилизации вычислительных ресурсов; обобщенно сформулирована задача прогнозирования

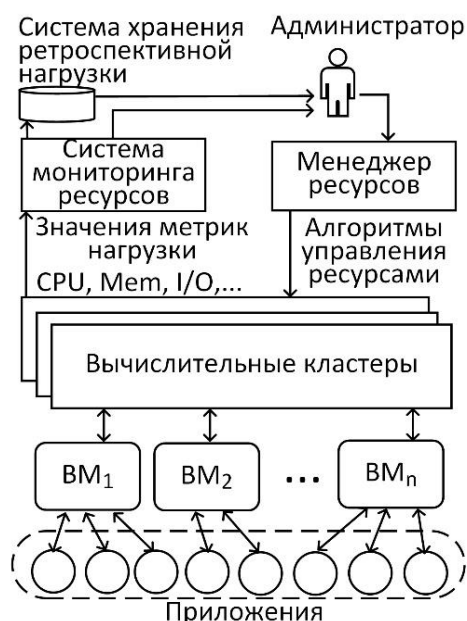


Рисунок 1 – Структурная схема системы управления рабочей нагрузкой ВЦОД

рабочей нагрузки ВЦОД на основе анализа временных рядов ее показателей; выявлены проблемы их искажения за счет внешних и внутренних факторов зашумления; рассмотрены особенности проблем «шумные соседи» (Noisy Neighbours) и «старение программного обеспечения» (Software Aging), как

факторов зашумления, наиболее влияющих на значения временного ряда рабочей нагрузки (рисунок 2).

В результате проведенного анализа было сформулировано противоречие между необходимостью получения приемлемых (заданных) прогнозных значений рабочей нагрузки ВЦОД и отсутствием в существующих службах их администрирования методов и алгоритмов, учитывающих факторы зашумления значений временных рядов заданных показателей производительности вычислительных ресурсов.

На основании этого противоречия сделана постановка научной задачи по разработке моделей, алгоритмов и архитектуры системы прогнозирования рабочей нагрузки ВЦОД в условиях зашумления значений временных рядов ее ретроспективных данных и предложена

модель модовой декомпозиции временного ряда, которая обеспечивает снижение влияния факторов зашумления на значения временного ряда за счет двухэтапной процедуры его разложения на множество колебательных модовых функций (КМФ) – амплитудно-частотных модуляций в заданных узких полосах частот, что позволяет связать их с определенным процессом и выявлять области локальности.

Проводится исследование использования модели модовой декомпозиции для анализа временных рядов сигналов различной природы и ее применение для снижения влияния факторов зашумления за счет дифференцированного анализа элементов множества КМФ.

Исследуются теоретические основы модовой декомпозиции сигналов, в частности, преобразование Гильберта-Хуанга, а также методы ее реализации. Выделяются: семейство методов декомпозиции на эмпирические моды (ДЭМ) и методы декомпозиции на вариационные моды (ДВМ). Рассматриваются их особенности, достоинства и недостатки.

В качестве варианта ДЭМ рассматривается метод комплементарной множественной декомпозиции на эмпирические моды с адаптивным шумом (КДЭМАШ), основанный на:

– объединении временного ряда $w(t)$ с гауссовым белым шумом $n_i(t)$ для создания N_e зашумленных реализаций сигнала $D_i(t)$:

$$D_i(t) = w(t) + en_i(t), \quad (1)$$

где, $i=1,2,\dots,N_e$, а значение e определяет амплитуду $n_i(t)$ с нормальным гауссовским распределением $N(\mu, \sigma^2)$, где коэффициент сдвига $\mu=0$;

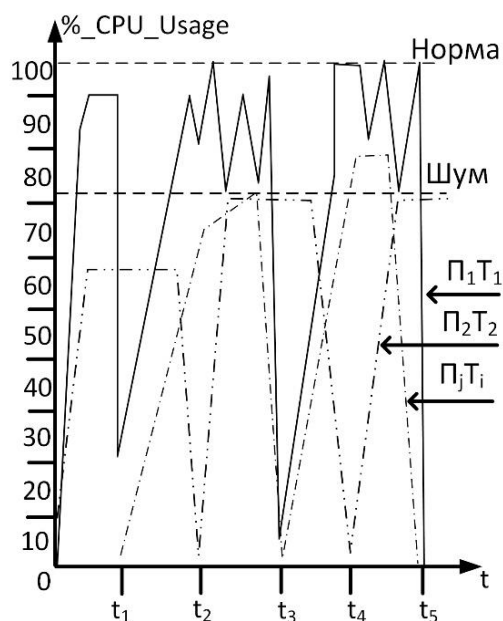


Рисунок 2 – Пример зашумления временного ряда показателя CPU_Usage потока влиянием других активных потоков

– выделении из сигнала $D_1(t)$ первой (базовой) функции КМФ:

$$\overline{\text{КМФ}}_1(t) = \frac{1}{N_e} \sum_{i=1}^{N_e} \text{КМФ}_1^i(t) \quad (2)$$

– вычислении значения первого остатка $r_1(t)$:

$$r_1(t) = w(t) - \overline{\text{КМФ}}_1(t) \quad (3)$$

– итерационном повторе этих операции для вычисления значений последующих $(k+1)$ функции КМФ с обновлением остатка на каждой итерации до условия останова:

$$\overline{\text{КМФ}}_{k+1}(t) = \frac{1}{N_e} \sum_{i=1}^{N_e} E_1(r_k(t) + \varepsilon_k E_k(n_k(t))), \quad (4)$$

где, E_k - процедура получения k -й КМФ, включающая определение ее локальных экстремумов и построение огибающих, а также получение средней огибающей и ее вычитание из исходного сигнала;

- получение модели временной ряда $w(t)$ в виде суммы множества КМФ и конечного остатка $R(t)$:

$$w(t) = \sum_{k=1}^K \overline{\text{КМФ}}_k(t) + R(t) \quad (5)$$

Обосновывается предположение компоненты разложения (КМФ) в достаточной мере сохраняют статистические свойства данных временного ряда и не теряют информацию о долгосрочных временных тенденциях. Также обосновывается факт, что поскольку гауссовский шум $n_1(t)$ имеет распределение $N(\mu, \sigma^2)$ с нулевым коэффициентом сдвига μ и благодаря усреднению по нескольким зашумленным реализациям сигнала шумовая компонента будет существенно снижена, а сами значения временного ряда $w(t)$ не изменятся.

В силу эмпирического характера получения элементов множества КМФ за счет выполнения ряда итераций, методы семейства ДЭМ, являясь вычислительно простыми, могут в ряде случаев не обеспечивать требуемой точности разложения.

С целью устранения этого недостатка рассматривается использование метода ДВМ – решение варианта вариационной задачи вида:

$$\min_{(f_k), (c_k)} \begin{cases} \sum_{k=1}^K \left\| \partial_t \left[\sigma(t) + \frac{i}{\pi t} f_k(t) \right] e^{-j c_k t} \right\|_2^2, \\ \text{при условии } \sum_{k=1}^K f_k(t) = w(t) \end{cases} \quad (6)$$

где, ∂_t - частная производная t , а элементы множества $c_K = c_1, c_2, \dots, c_k$ являются значениями центральных частот каждого элемента множества функций КМФ. Для получения оптимального решения этой вариационной задачи применяется метод множителей Лагранжа с квадратичным перемножением штрафной функции:

$$L(f_k, c_k, \lambda) = \alpha \sum_{k=1}^K \left\| \partial_t \left[\sigma(t) + \frac{i}{\pi t} f_k(t) \right] e^{-j c_k t} \right\|_2^2 + \left\| w(t) - \sum_{k=1}^K f_k(t) \right\|_2^2 + \left\langle \lambda(t), w(t) - \sum_{k=1}^K f_k(t) \right\rangle, \quad (7)$$

где, λ - множитель Лагранжа, а α - коэффициент штрафной функции.

В дальнейшем получается экстремальное решение для частот каждого элемента множества КМФ и их центральных частот c_k :

$$f_k^{n+1}(c) = \frac{\hat{w}(c) - \sum_{k=1, k \neq K}^k f_k(c) + \frac{\hat{\lambda}(c)}{2}}{1+2} \alpha (c - c_k)^2, \quad (8)$$

$$c_k^{n+1} = \frac{\int_0^\infty c |f_k(c)|^2 dc}{\int_0^\infty |f_k(c)|^2 dc}. \quad (9)$$

Для получения оптимального решения для всех элементов множества КМФ используется известный алгоритм ADMM (Alternating Direction Method of Multipliers), который декомпозирует сложную вариационную задачу на частные подзадачи. Его этапами являются:

1. Установка нулевых значений параметров $f_k(t)$, c_k , λ_k^1 и n .
2. Обновление значений $f_k^{n+1}(t)$ и c_k^{n+1} (выражения 8 и 9).
3. Обновление значения оператора множителя Лагранжа λ^{n+1} :

$$\hat{\lambda}^{n+1}(c) = \hat{\lambda}^{n+1}(c) + \tau \left(\hat{f}(c) - \sum_{k=1}^K \hat{f}_k^{n+1}(c) \right). \quad (10)$$

4. Повтор этапов 2-3 до получения значения $f_k^{n+1}(c)$ при выполнении

следующего условия $\sum_{k=1}^K \frac{\|f_k^{n+1} - f_k^n\|_2^2}{\|f_k^n\|_2^2} < \varepsilon$.

Таким образом, применение метода ДВМ сохраняет основные частотные характеристики временного ряда, минимизируя шумовую компоненту, и при этом полученное множество функций КМФ является полностью реконструктивным (обеспечивающим восстановление значений исходного временного ряда). Однако, являясь численным методом, требующим дополнительной итерационной процедуры ADMM, метод ДВМ имеет высокую алгоритмическую вычислительную сложность.

В связи с этим в работе предложена модель модовой декомпозиции временного ряда, получаемая двухэтапным применением методов ДЭМ (КДМЭАШ) и ДВМ (рисунок 3).

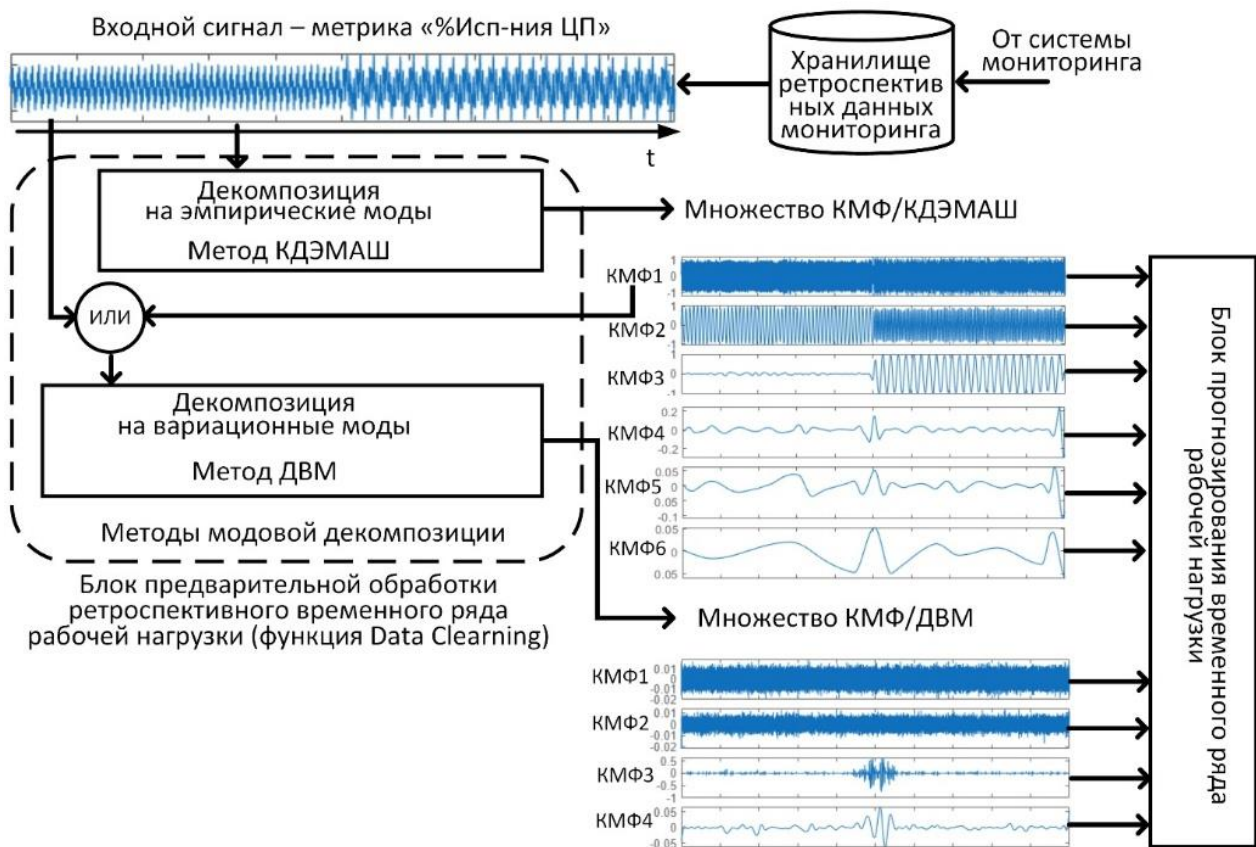


Рисунок 3 – Двухэтапная процедура модовой декомпозиции формирования модели временного ряда рабочей нагрузки

Основным этапом является метод КДМЭАШ, формирующий множество функций КМФ/КДМЭАШ. В случае, если результаты прогнозирования временного ряда $w(t)$ на основе этого множества будут недостаточно точными, на втором этапе модовой декомпозиции используется метод ДВМ, который, в зависимости, от заданных условий, применяется либо к исходному временному ряду $w(t)$, либо к базовой функции $\overline{КМФ}_1(t)$, полученной методом КДМЭАШ.

Таким образом, на основе представления зашумленного временного ряда рабочей нагрузки в виде сигнальной конструкции, теоретических подходов к разложению сигнала на колебательные модовые функции и двухэтапного использования методов эмпирической и вариационной модовой декомпозиции разработана модель модовой декомпозиции временного ряда рабочей нагрузки ВЦОД, обеспечивающая снижение влияния факторов зашумления на значения временного ряда.

Вторая глава посвящена разработке комплексного алгоритма предварительной обработки временного ряда рабочей нагрузки. В соответствии с предложенной моделью модовой декомпозиции временного ряда рабочей нагрузки был разработан комплексный алгоритм модовой декомпозиции, основанный на двухэтапной процедуре (рисунок 3) и включающий следующие разработанные частные алгоритмы: реализации метода КДМЭАШ – эмпирического разложения временного ряда на множество КМФ-функций (рисунок 4); реализации метода ДВМ для вариационного разложения

на множество КМФ-функций, в зависимости от тестовых результатов прогнозирования, либо базовой КМФ-функции, полученной методом КДМЭАШ, либо исходного временного ряда.

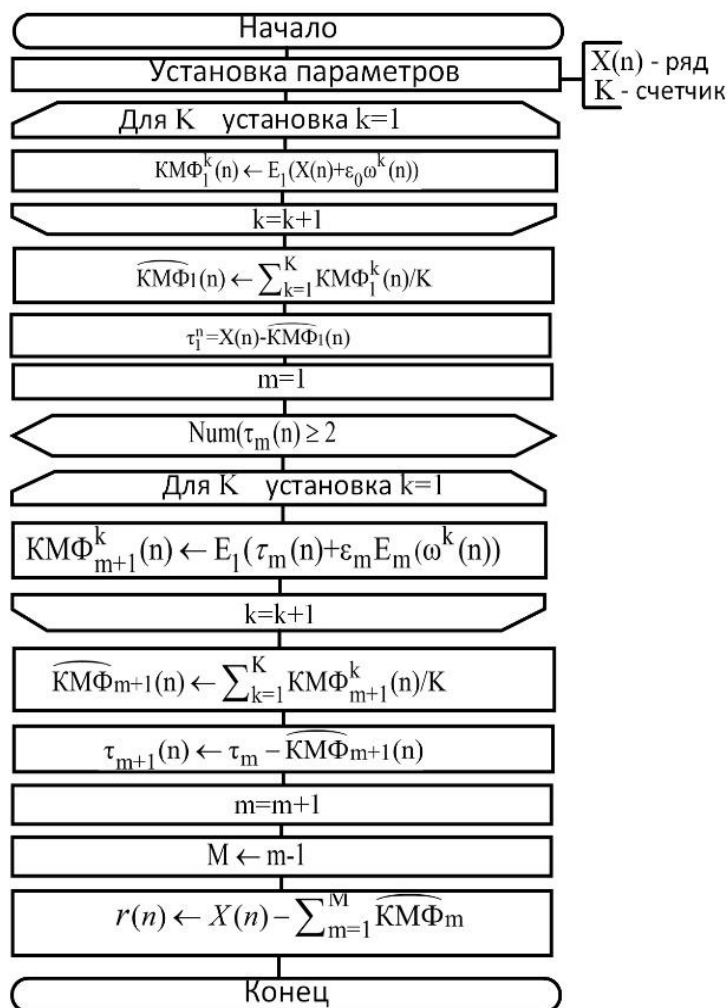


Рисунок 4 – Схема частного алгоритма КДМЭАШ для первого этапа модовой декомпозиции временного ряда

Из рисунка 4 следует, что предлагаемый алгоритм КДМЭАШ реализует итеративную процедуру получения множества КМФ, каждый элемент которого находится эмпирически на основе циклического поиска экстремальных точек предварительно зашумленных входных значений. При этом базовая функция $КМФ_1(n)$, являясь основой получения значений остатков для последующих итераций, фактически совпадает с входным временным рядом, представляя усредненное значение его зашумленного варианта.

Разработанный алгоритм ДВМ (рисунок 5) используется на втором этапе комплексного алгоритма для устранения основного недостатка алгоритма КДМЭАШ – эмпирического характера получения значений множества КМФ, что может давать достаточно большую погрешность разложения относительно входных данных. Особенностью предлагаемого алгоритма ДВМ является использование штрафного параметра α в сочетании с множителем Лагранжа

$\lambda(t)$ для преобразования сложной вариационной задачи с ограничениями (выражение 6) в простую без ограничений:

$$\omega_k^{n+1} \leftarrow \arg_{\omega_k} \min L(\{u_{i < k}^{n+1}\}, \{u_{i \geq k}^n\}, \{\omega_i^n\}, \lambda^n). \quad (11)$$

$$\omega_k^{n+1} \leftarrow \arg_{\omega_k} \min L(\{u_i^{n+1}\}, \{u_{i < k}^n\}, \{\omega_{i \geq k}^n\}, \lambda^n). \quad (12)$$

$$\lambda^{n+1}(\omega) = \lambda^n + \tau(f - \sum_k u_k^{n+1}), \quad (13)$$

где, τ - показатель устойчивости к сходимости.

Из рисунка 5 видно, что важным фактором в реализации алгоритма ДВМ является первоначальное определение значений и штрафного параметра, существенно влияющих на последующие этапы декомпозиции. Включение в алгоритм цикла подбора этих параметров существенно усложняет его и увеличивает и без того высокую вычислительную сложность.

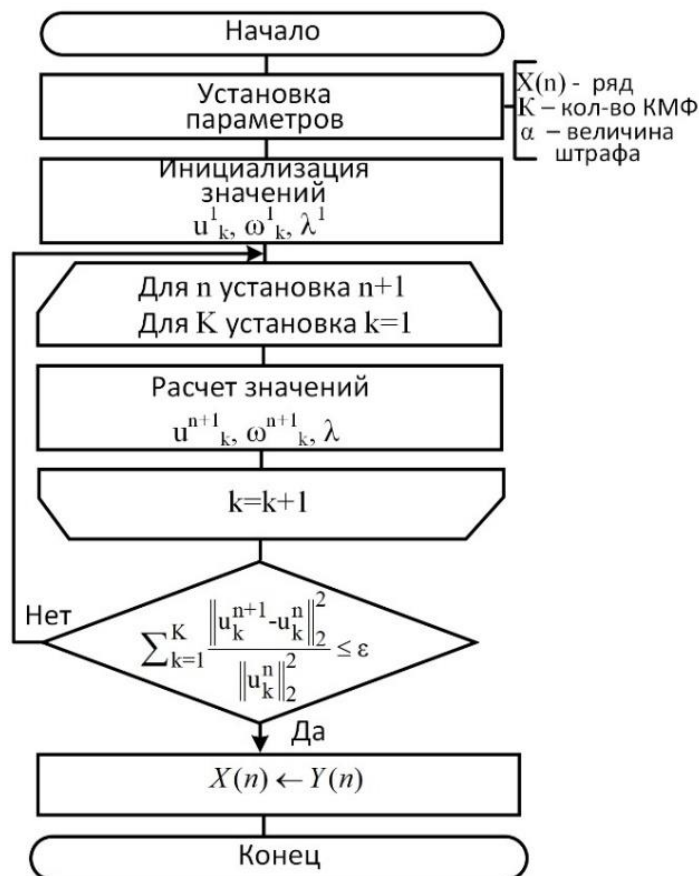


Рисунок 5 – Схема частного алгоритма ДВМ для второго этапа модовой декомпозиции временного ряда

Поэтому наиболее оптимальным способом является подбор этих параметров, основанный на эмпирическом опыте администратора системы прогнозирования, хорошо знакомого с особенностями рабочей нагрузки конкретного ВЦОД.

Полученные в результате реализации разработанного комплексного алгоритма модовой декомпозиции множества функций КМФ используются

в качестве входных последовательностей гибридного алгоритма прогнозирования.

Таким образом, разработан комплексный алгоритм предварительной обработки временного ряда рабочей нагрузки, отличающийся наличием этапа вторичной вариационной модовой декомпозиции базовой колебательной модовой функции, полученной методом эмпирической модовой декомпозиции, который обеспечивает формирование множеств обучающей и тестовой выборок для системы прогнозирования элементов временного ряда.

Третья глава посвящена разработке гибридного алгоритма прогнозирования рабочей нагрузки. Поскольку полученные при реализации алгоритмов КДМЭАШ и ДВМ множества КМФ функций являются частными реализациями исходного временного ряда рабочей нагрузки, а значит содержат зависимости, связанные с интересующими службу администрирования ВЦОД шаблонами нагрузки, процессу непосредственно прогнозирования предшествует процесс выявления указанных зависимостей. В связи с этим разработанный алгоритм является гибридным – объединяющим частные алгоритмы глубокого обучения для: поиска и выявления зависимостей параметров временного ряда, соответствующих шаблонам рабочей нагрузки; прогноза появления указанных зависимостей в последующих отсчетах временного ряда на основе анализа его предыдущих отсчетов.

Решение задачи поиска зависимостей в одномерных массивах данных, к которым относятся временные ряды, наиболее эффективно осуществляют одномерные сверточные сети (СНН-ОМ, англ. 1D-CNN), ядро которых представляет одномерный массив – фильтр, скользящий вдоль значений отсчетов временного ряда. Ее основой является два и более сверточных слоя, решающих задачу поиска зависимостей, чередующихся с ними слоев подвыборки, предназначенных для снижения размерности признакового пространства с целью обеспечения инвариантности к небольшим искажениям, а также полносвязный слой, формирующий выходные значения. Существенную роль при структурно-параметрической разработке СНН-ОМ является выбор размера ее ядра, от которого зависит масштаб охвата поиска зависимостей. В разработанном гибридном алгоритме предложено использование ансамбля СНН-ОМ, отличающихся размером ядра в масштабах «малый», «средний» и «большой». Полученные выходные данные ансамбля объединяются путем выполнения операции конкатенации и являются входом сети, обеспечивающей прогнозирование, основанной на двунаправленной сети с долгой краткосрочной памятью (СДКП-ДН, англ. Bi-LSTM). Скрытые слои сети СДКП-ДН состоят из ячеек (cell), содержащих элементы памяти и триггеры (gates), обеспечивающие сохранение/удаление предшествующих значений для их сравнения с текущими значениями. Поскольку сеть СДКП-ДН имеет в скрытых слоях ячейки, в которых анализируемые данные распространяются в противоположных направлениях, то сравнительному анализу подвергаются как предшествующие, так и последующие значения, что позволяет формировать выходные значения временного ряда и последующие. Глубина прогноза при этом зависит от выбора количества скрытых слоев и гиперпараметров сети.

В общем виде схема взаимодействия моделей глубокого обучения в гибридном алгоритме представлена на рисунке 6 в виде ансамбля трех СНН-ОМ, отличающихся размерами ядер (c_1, c_2, c_3), выход которого – реализация временного ряда $X' = \{x'_1, x'_2, \dots, x'_n\}$ является входом сети СДКП-ДН, которая содержит: два скрытых двунаправленных слоя ячеек, attention-слой, объединяющий найденные СНН-ОМ зависимости и выбирающий критические важные из них путем перераспределения весовых коэффициентов h_n , и полносвязный слой, формирующий выходные данные прогноза – ряд значений $Y = \{y_1, y_2, \dots, y_n\}$.

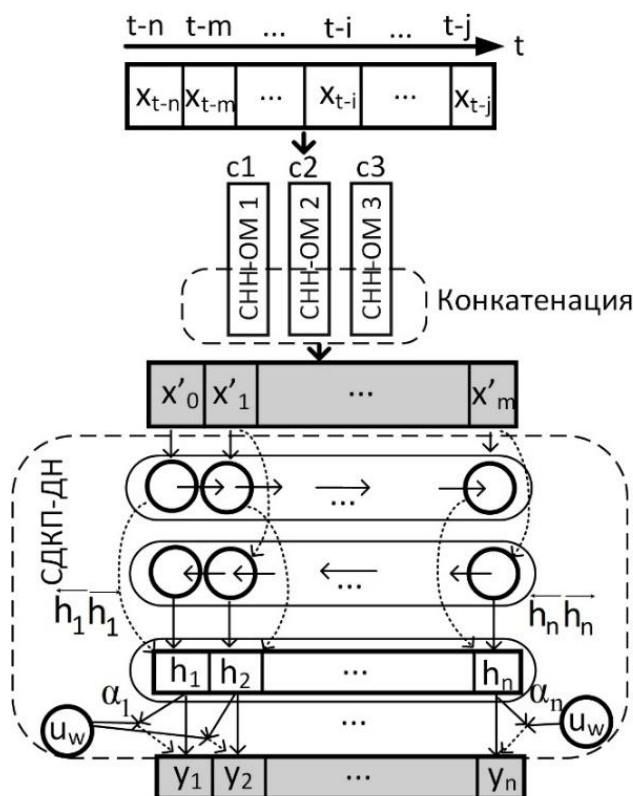


Рисунок 6 – Предлагаемая гибридная схема взаимодействия моделей глубокого обучения

На рисунке 7 представлена схема разработанного гибридного алгоритма прогнозирования временного ряда рабочей нагрузки для системы глубокого обучения.

Поскольку анализ всех элементов множества КМФ функций выполняется ансамблем СНН-ОМ, каждый экземпляр которого отличается размером ядра c_n , в разрабатываемом обобщенном алгоритме гибридной МГО предлагается

этап параллельной обработки $\text{СНН}_n^{\text{ВЫХ}} \leftarrow \text{СНН}_n(X, c_n)$ данных X , включенный

во вложенные циклы со счетчиками: K – количества эпох обучения; M – количества обучающих выборок (batch). Конкатенированный результат

$СДКП_2^{ВЫХ} \leftarrow СДКП_1^{ВЫХ} \leftarrow СНН^{ВЫХ}$ – поступает в каскад двух сетей СДКП-ДН, отличающихся числом скрытых слоев для решения задачи прогнозирования в разных временных масштабах.



Рисунок 7 – Схема гибридного алгоритма прогнозирования рабочей нагрузки

Предложенный алгоритм гибридной МГО является одним из вариантов, поскольку решение вопроса о количестве СНН-ОМ в ансамбле сетей СНН, а также количество СДКП-ДН в каскаде сетей СДКП существенно зависит от характеристик обрабатываемого временного ряда ретроспективных данных рабочей нагрузки, которые, в свою очередь зависят от особенностей организации ВЦОД, а также особенностей их потребительской нагрузки.

Таким образом, разработан гибридный алгоритм прогнозирования временного ряда рабочей нагрузки для системы глубокого обучения, который обеспечивает получение разномасштабных прогнозных значений временного ряда рабочей нагрузки.

В четвертой главе предложена и реализована в виде комплекса программ архитектура системы прогнозирования рабочей нагрузки. Предлагаемая архитектура базируется на совокупности процессов, реализуемых разработанными алгоритмами: предварительной обработки временного ряда ретроспективных данных рабочей нагрузки, разделенных на обучающее и тестовое подмножества в соотношении 75/25, и процессов выделения значимых признаков и прогнозирования значений временного ряда (рисунок 8).

Для указанной совокупности процессов, с учетом особенностей разработанных алгоритмов модовой декомпозиции и прогнозирования были рассмотрены и выбраны соответствующие фреймворки: MathLab Signal Processing для реализации модуля предварительной обработки временного ряда и Python 1dcnn-bilstm для модуля прогнозирования.

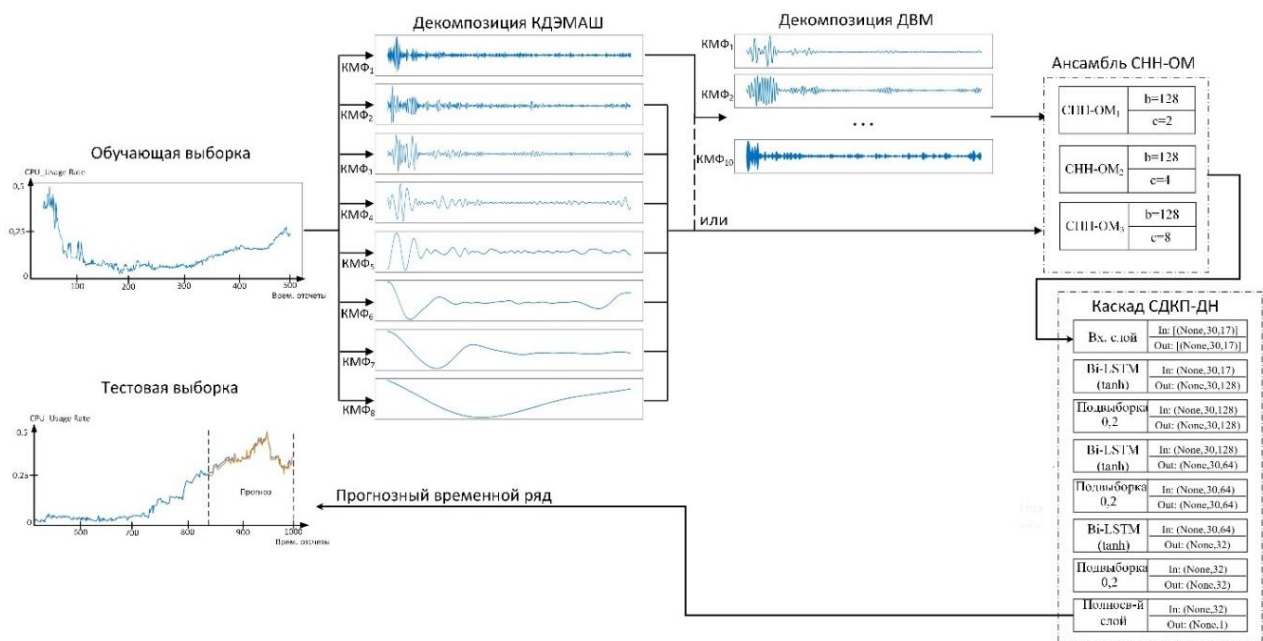


Рисунок 8 – Совокупность процессов прогнозирования рабочей нагрузки

На основе выбранных фреймворков был реализован основной программный компонент системы Hybrid-VMD-CEEMDAN-LSTM Preprocessor. Для администрирования среды виртуализации был разработан компонент управления «ПК СВ «БРЕСТ»_Config» и компонент визуализации BRAIN_STORM_VIK. Дополнительно была разработана структура файловых хранилищ обучающего и тестовых подмножеств, а также результатов модовой декомпозиции, для управления хранимыми в них данными были реализованы программные компоненты: COM_BOX, DOC.EU.CLASSIFIER, STEGO_JPEG_ANALIZER, STEGOMP4. Разработанная архитектура системы прогнозирования рабочей нагрузки представлена на рисунке 9.

Сохраняемый в хранилище набор исходных временных рядов предварительно преобразуется в совокупность векторов - скользящих окон фиксированного размера. Дополнительно данные каждого из векторов нормализуются с целью приведения их характеристик к единому масштабу.

В качестве метода нормализации предложено использовать минимаксное масштабирование:

$$X_{\text{норм}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}. \quad (14)$$

Для оценивания предложенных модели, алгоритмов и архитектуры системы прогнозирования по критерию пригодности был спланирован и проведен сравнительный эксперимент по прогнозированию рабочей нагрузки на основе ретроспективных данных ВЦОД GoogleCluster, находящимся в открытом доступе для проведения исследований. В качестве сравниваемой альтернативы использован статистический метод прогнозирования временных на основе машины опорных векторов в совокупности с авторегрессией интегрированной модели скользящего среднего (SVM-ARIMA), реализованный в системе администрирования GoogleCluster. В качестве показателя нормализованных значений временного ряда был выбран показатель CPU_UsageRate – коэффициент загрузки процессора, процент времени, в течение которого он занят обработкой задач, рассчитываемый путём деления времени работы процессора на общее время за заданный период. Размер вектора - скользящего окна обрабатываемых данных представлен 500 условными отсчетами. Исследовались: эффективность модовой декомпозиции зашумленного временного ряда эмпирическим (метод КДМЭАШ) и вариационным (ДВМ) способом, а также сравнительная эффективность процесса прогнозирования значений временного ряда на основе реализации SVM-ARIMA и предложенного гибридного алгоритма прогнозирования.

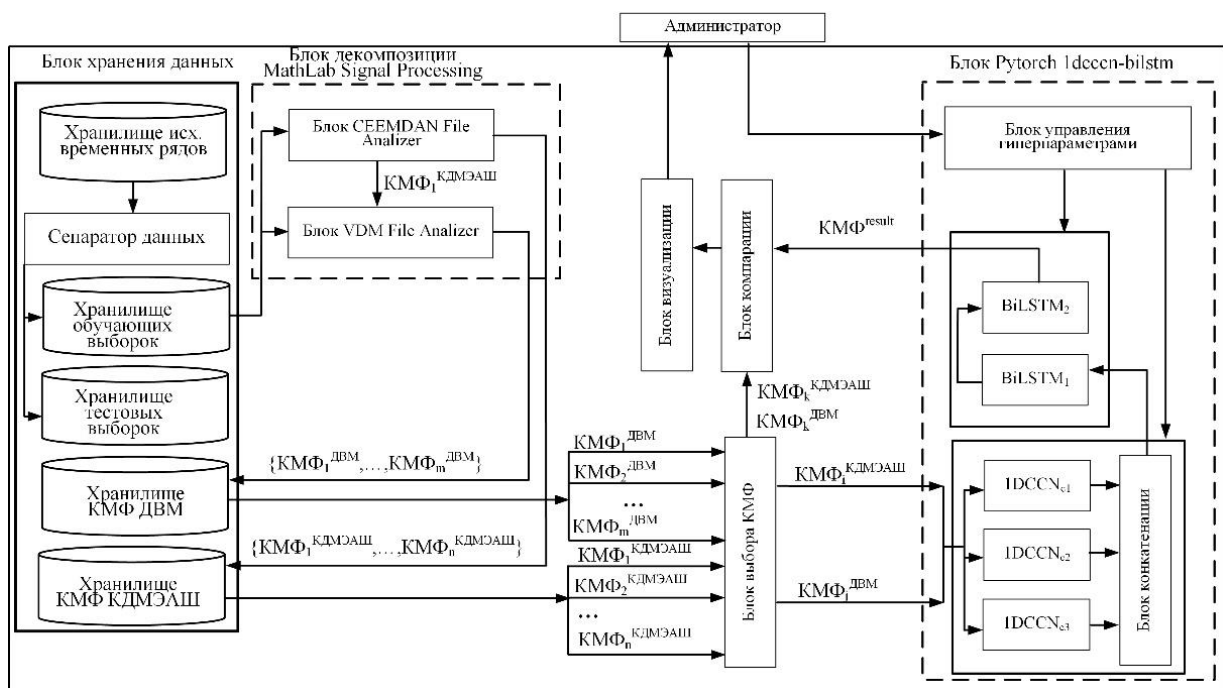


Рисунок 9 – Архитектура системы прогнозирования рабочей нагрузки

В обобщенном виде результаты указанных экспериментов представлены на рисунках 10 и 11.

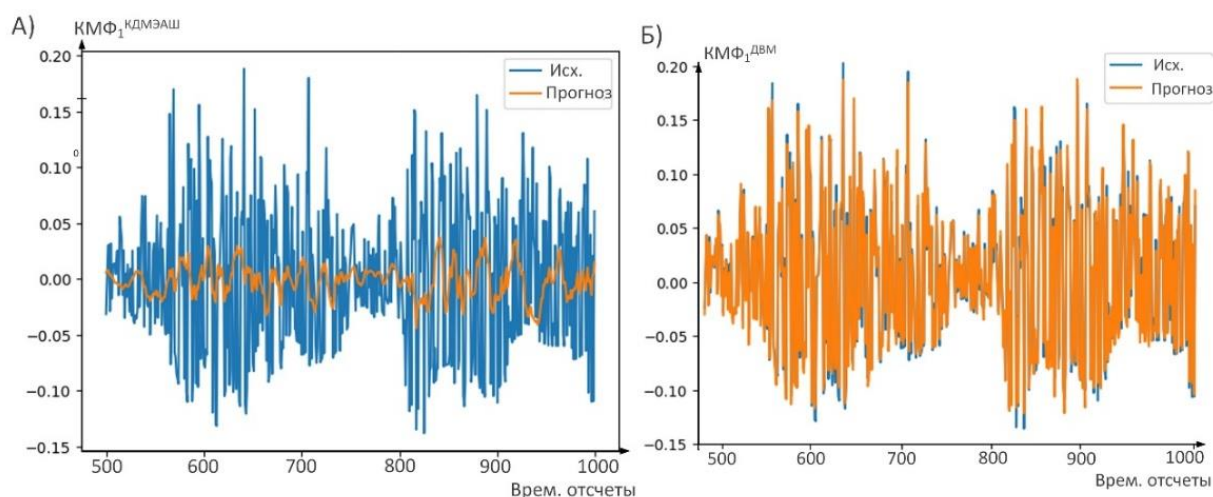


Рисунок 10 – Сравнительная оценка эффективности эмпирической (а) и вариационной модовой декомпозиции (б) временного ряда (на примере базовой функции КМФ)

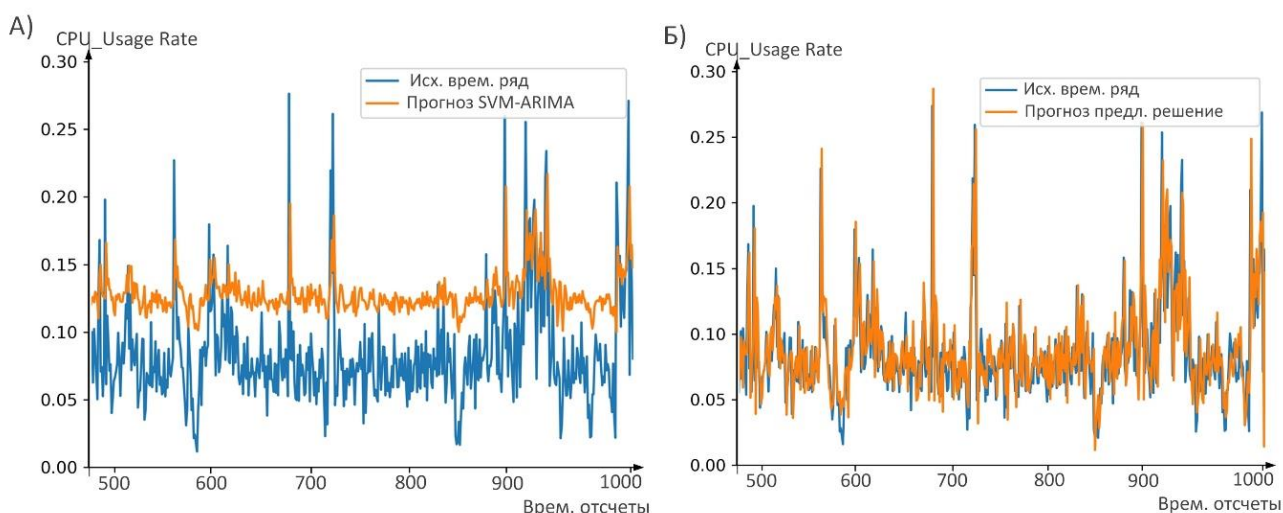


Рисунок 11 – Сравнительная оценка эффективности прогнозирования временного ряда способом SVM-ARIMA(а) и предлагаемым решением (б)

Таким образом, показана возможность применения предложенных модели и алгоритмов в составе программного обеспечения систем прогнозирования рабочей нагрузки на основе временных рядов ее ретроспективных данных в условиях воздействия факторов зашумления.

В заключении подведены итоги диссертации, сформулированы основные результаты и выводы.

ЗАКЛЮЧЕНИЕ

На основании теоретических исследований и вычислительных экспериментов сделаны следующие выводы:

1. На основе анализа формирования временных рядов ретроспективных данных рабочей нагрузки, влияния на нее факторов зашумления и существующих методов модовой декомпозиции сигналов разработана модель модовой декомпозиции временного ряда рабочей нагрузки, включающая этапы

декомпозиции на эмпирические моды и декомпозиции на вариационные моды и применение полученных множеств колебательных модовых функций в зависимости от результатов тестового прогона алгоритмов прогнозирования.

2. Разработан комплексный алгоритм предварительной обработки временного ряда рабочей нагрузки, основанный на двухэтапной процедуре реализации метода комплементарной декомпозиции на эмпирические моды с адаптивным шумом и метода декомпозиции на вариационные моды, обеспечивающей получение множеств колебательных модовых функций с приемлемой вычислительной сложностью.

3. Разработан гибридный алгоритм прогнозирования временного ряда рабочей нагрузки, сочетающий: ансамблевый подход к использованию одномерных сверточных нейронных сетей с разным размером ядра, для выявления значимых признаков шаблонов рабочей нагрузки в разных временных масштабах; алгоритм прогнозирования, основанный на каскаде двунаправленных сетей с долгой краткосрочной памятью с разным числом скрытых слоев, обеспечивающий формирование прогнозного временного ряда рабочей нагрузки на основе конкатенированных выходов ансамбля одномерных сверточных нейронных сетей.

4. Разработана архитектура системы прогнозирования рабочей нагрузки ВЦОД на основе значений временного ряда ее ретроспективных данных, включающая модуль предобработки временного ряда, снижающий влияние шумовых факторов за счет разложения ряда на множество колебательных модовых функций и модуля прогнозирования значений временного ряда рабочей нагрузки на основе выделенных значимых признаков ее шаблонов. Для разработанной архитектуры обосновано выбраны фреймворки указанных модулей и программно реализованы разработанные алгоритмы.

5. Проведен анализ возможностей использования предложенных модели, алгоритмов и архитектуры системы прогнозирования на примере решения задачи прогнозирования сохраненных ретроспективных данных рабочей нагрузки ВЦОД Google Cluster. Проведен эксперимент по решению задачи прогнозирования в сравнении с методом статистического прогнозирования SVM-ARIMA. Результаты сравнения полученных экспериментальных серий предлагаемого решения и альтернативного решения (SVM-ARIMA и SVM-FARIMA) показывают, что в анализируемом временном окне (850-1000 отсчетов) предложенное решение дает прогнозные значения, в среднем отклоняющиеся от значений исходного временного ряда на участках стационарности менее чем на 2 %, в то время как прогнозные значения альтернативного способа имеют отклонение от 5 % (FARIMA) до 15 % (ARIMA).

Дальнейшие разработки по данному направлению могут быть связаны с повышением точности прогнозирования рабочей нагрузки ВЦОД для разных масштабов ее временного ряда, а также особенностей факторов его зашумления выполнением в рамках ВЦОД специфических приложений и сервисов.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в изданиях, рекомендованных ВАК

1. Подход к предварительной обработке зашумленных ретроспективных данных об уровне загрузки вычислительных ресурсов для системы прогнозирования рабочей нагрузки виртуализированного центра обработки данных / Мартыненко Б.В., Кравец О.Я., Белецкая С.Ю., Скурихин А.А. // Системы управления и информационные технологии, №3 (101), 2025. С. 41-50.

2. Сравнительный анализ методов аутентификации: от традиционных паролей до биометрических решений / Мартыненко Б.В., Охотников К.А. // Динамика сложных систем - XXI век, №3, 2025. С. 52-57.

3. Разработка алгоритмов системы прогнозирования рабочей нагрузки виртуализированного центра обработки данных, функционирующей в условиях зашумленных ретроспективных данных об уровне загрузки вычислительных ресурсов / Мартыненко Б.В. // Системы управления и информационные технологии, № 4 (102), 2025. С. 14-20.

4. Разработка гибридной модели глубокого обучения для прогнозирования рабочей нагрузки в виртуализированных центрах обработки данных / Мартыненко Б.В. // Моделирование, оптимизация и информационные технологии. 2025, Т. 3, № 4. Доступно по: <https://moitvvt.ru/ru/journal/pdf?id=2097> DOI: 10.26102/2310-6018/2025.51.4.054.

Публикации в изданиях, индексируемых в 1 квартале базы РИНЦ, базах «Scopus» и «Web of Science»

5. Algorithms for achieving mutual information coordination / O. Ja. Kravets, B. V. Martynenkov, A. V. Tsvetkov, E. O. Puzhanova, D. I. Mutin, P. V. Besspalov, Yu. V. Redkin // International Journal on Information Technologies and Security, vol. 18, no. 1, 2026.

Свидетельства о регистрации программы для ЭВМ

6. Мартыненко Б.В., Мартыненко В.В., Покормяко В.В., Петрищев Е.В., Иванов В.А., Иванов И.В. DOC.EU.CLASSIFIER. Свидетельство о регистрации программы для ЭВМ № 2021617173 от 11.05.2021. – М: Роспатент, 2021.

7. Мартыненко Б.В., Родин Я.М., Иванов В.А., Иванов И.В. STEGO_JPEG_ANALIZER. Свидетельство о регистрации программы для ЭВМ № 2022616936 от 18.04.2022. – М: Роспатент, 2022.

8. Мартыненко Б.В., Гусейнов М.А., Иванов В.А., Иванов И.В. STEGOMP4. Свидетельство о регистрации программы для ЭВМ № 2022614274 от 18.03.2022. – М: Роспатент, 2022.

9. Мартыненко Б.В., Чернышов Д.Д., Шпаков А.А., Антонова А.М., Иванов В.А., Иванов И.В. Программа по настройке ПК СВ «БРЕСТ». Свидетельство о регистрации программы для ЭВМ № 2023661041 от 26.05.2023. – М: Роспатент, 2023.

10. Мартыненко Б.В., Иванов В.А., Иванова И.А., Куземкина А.О., Лаврентьев Ю.С., Кабешов И.Ю. Программа обнаружения криптографических контейнеров в файловой системе ОС Windows Scanner_Cryptocontainer. Свидетельство о регистрации программы для ЭВМ № 2023619566 от 12.05.2023. – М: Роспатент, 2023.

11. Мартыненко Б.В., Полков А.А., Енютин Д.Г., Гибовский А.А., Чуксеев Н.М., Иванов В.А. Программа шифрования «СОМ ВОХ». Свидетельство о регистрации программы для ЭВМ № 2023617526 от 11.04.2023. – М: Роспатент, 2023.

12. Мартыненко Б.В., Дорофеев Д.А., Махинов С.Н., Пивоваров Н.Р., Иванов В.А., Мартыненко И.И. Программа визуального отображения вопросов в многопользовательском режиме «BRAIN_STORM_VIK». Свидетельство о регистрации программы для ЭВМ № 2023614395 от 01.03.2023. – М: Роспатент, 2023.

13. Мартыненко Б.В., Мысин Н.О., Иванова И.А., Иванов В.А., Попов А.И., Дибиров К.Г. Программа настройки безопасной конфигурации средства ЭВТ под управлением ОССН AstraLinux SE. Свидетельство о регистрации программы для ЭВМ № 2024618031 от 08.04.2024. – М: Роспатент, 2024.

14. Мартыненко Б.В., Иванов В.А., Антонова В.М., Воронков Р.Д., Иванов И.В., Овсянкин С.В., Рублев В.С., Майбурд С.В. Программа для управления устройством защиты на ЭВМ с помощью почерка. Свидетельство о регистрации программы для ЭВМ № 2025616380 от 17.03.2025. – М: Роспатент, 2025.

15. Мартыненко Б.В., Иванов В.А., Мартыненко В.В., Самоаев А.В., Потапов Г.А., Климин Д.Г., Елизарова Н.А. Блувштейн А.А. Программа идентификации инцидентов информационной безопасности на основе анализа данных оперативной памяти (RAM). Свидетельство о регистрации программы для ЭВМ № 2025664302 от 03.06.2025. – М: Роспатент, 2025.

16. Мартыненко Б.В., Иванов В.А., Иванов И.В., Овсянкин С.В., Худайназарова Д.Р., Тараканов О.В., Дементьев А.Н., Маркин А.В. Программа идентификации детерминированных цифровых последовательностей в не циклических цифровых потоках. Свидетельство о регистрации программы для ЭВМ № 2025694366 от 04.12.2025. – М: Роспатент, 2025.

17. Мартыненко Б.В., Иванов В.А. Гибридный модуль предварительной обработки временных рядов на основе VMD и CEEMDAN для обучения LSTM. Свидетельство о регистрации программы для ЭВМ № 2026613698 от 09.02.2026. – М: Роспатент, 2026.

Статьи и материалы конференций

18. Новые вызовы цифровизации политической системы государства // Мартыненко Б.В., Мартыненко В.В. // Международный академический вестник, № 3 (47), 2020. С. 84-88.

19. Сети 5G / Мартыненко Б.В., Гатилов О.В. // Общество - Наука – Инновации // Международная научно-практическая конференция, Уфа, 2021. С. 74-77.

20. Виртуализация инфраструктуры программно-определяемых центров обработки данных на базе программного комплекса «Средство виртуализации «Брест» ОССН AstraLinux SE» / Мартыненко Б.В., Чернышов Д.Д., Шпаков А.А. // Хроники цифровых трансформаций. Матер. круглого стола. - Волгоград, 2022. С. 88-92.

21. Применение нейронной сети для интеллектуальной обработки больших объемов информации / Мартыненко Б.В., Беляков Е.В., Мансурова А.А. // Кибербезопасность: технические и правовые аспекты защиты информации. – Сб. науч. тр. по итогам III ежегодной нац. НПК. – М., 2024. С. 200-203.

22. Модель угроз и модель нарушителя безопасности маршрутизаторов / Мартыненко Б.В., Антонова В.М., Тихомиров В.А. // Стратегии успеха: инновационные методы, технологии и практики в науке для достижения глобального прогресса. Междунар. НПК. - Уфа, 2025. С. 7-10.

23. К вопросу о прогнозировании рабочей нагрузки в виртуализированных центрах обработки данных в условиях зашумления ретроспективных данных мониторинга / Мартыненко Б.В. // Гибкое производство, цифровая трансформация, информационная безопасность и экология: сборник трудов по материалам I Всероссийской научно-практической конференции, Москва, 2025. С. 486 – 495.

24. Control the workload prediction of a virtualized data center for noisy retrospective data / Martynenkov B.V., Tsvetkov A.V. // Modern informatization problems: Proc. of the XXXI-th Int. Open Science Conf. - Yelm, WA, USA: Science Book Publishing House, 2026.