

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

кандидата физико-математических наук Воеводина Вадима Владимировича на диссертационную работу **ЮРТИНА Алексея Артемьевича** «Нейросетевые методы восстановления потоковых данных», представленную на соискание ученой степени кандидата физико-математических наук по специальности 2.3.5 – математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Актуальность диссертационного исследования А.А. Юртина, направленного на разработку новых методов восстановления потоковых данных, представленных в форме временных рядов, обусловлена следующими факторами. Во-первых, в связи со стремительным ростом объема потоковых данных в современных информационных системах задача их автоматизированного анализа является актуальной и относится к числу ключевых направлений интеллектуального анализа данных. Во-вторых, реальные временные ряды часто содержат пропуски, возникающие вследствие технических сбоев, ошибок регистрации и иных факторов. В-третьих, значительные вычислительные затраты на обучение нейросетевых моделей делают актуальной задачу повышения эффективности их жизненного цикла, которая предполагает разработку методов предварительного прогнозирования качества моделей, сокращающих временные и ресурсные затраты на этапе выбора архитектуры.

Цель диссертационного исследования А.А. Юртина заключалась в разработке и исследовании новых методов восстановления потоковых данных, представленных в форме многомерных временных рядов, на основе совместного использования нейросетевых моделей и поведенческих шаблонов. Для выполнения поставленной цели А.А. Юртиным были решены следующие задачи.

1. Разработаны следующие нейросетевые модели, методы и алгоритмы, предназначенные для восстановления потоковых данных:
 - нейросетевые методы восстановления многомерных временных рядов в офлайн и онлайн режимах;
 - функция потерь для обучения нейросетевых моделей восстановления временных рядов;

– метод прогнозирования ошибки и времени обучения нейросетевых моделей восстановления временных рядов.

2. Проведены вычислительные эксперименты, результаты которых продемонстрировали эффективность нейросетевых моделей, методов и алгоритмов.

Диссертация состоит из введения, четырех глав, заключения, списка литературы и трех приложений. Объем диссертации составляет 193 страницы, объем библиографии 168 наименований.

В **первой главе** проведен анализ существующих методов и инструментальных средств, применяемых для восстановления пропущенных значений временных рядов. Выполнена систематизация подходов к решению данной задачи. Проанализированы используемые функции потерь и методы предварительной оценки качества нейросетевых моделей восстановления временных рядов. На основе проведенного анализа определены ключевые проблемы существующих решений и сформулированы направления дальнейших исследований, направленных на повышение точности восстановления потоковых данных.

Во **второй главе** представлены новые нейросетевые методы восстановления потоковых данных для различных режимов обработки временных рядов. Предложен оригинальный метод SANNI для онлайн восстановления, обеспечивающий заполнение пропусков на основе предыдущих наблюдений. Разработан метод SAETI для офлайн восстановления, позволяющий обрабатывать пропущенные значения во временных рядах независимо от расположения пропуска. Представлены архитектуры нейросетевых моделей, реализующих ключевые этапы обработки данных в рамках предложенных методов.

В **третьей главе** представлены новые инструментальные средства для оценки и повышения точности нейросетевых моделей восстановления потоковых данных, представленных в форме временных рядов. Предложена новая функция потерь MPDE, учитывающая поведенческое сходство при сравнении подпоследовательностей временного ряда. Представлен новый метод

прогнозирования ошибки и времени обучения tsGAP, использующий архитектуру нейросетевой модели восстановления пропусков потоковых данных, представленной в виде ориентированного ациклического графа, для предварительной оценки точности восстановления.

В четвертой главе представлены результаты вычислительных экспериментов, подтверждающие эффективность предложенных нейросетевых методов и функции потерь. Сравнительный анализ показал, что метод SANNI обеспечивает высокую точность онлайн восстановления временных рядов с разнообразными активностями, превосходя существующие аналоги. Метод SAETI демонстрирует преимущество в офлайн восстановлении. Нейросетевые модели, обученные с использованием функции потерь MPDE, демонстрируют точность восстановления, превышающую точность моделей, обученных с применением других аналогичных функций потерь. Метод tsGAP обеспечивает прогнозирование ошибки и времени обучения нейросетевых моделей восстановления потоковых данных, превосходя по точности аналогичные методы. Проведенные эксперименты подтверждают высокую точность и практическую применимость всех инструментальных средств, полученных в ходе исследования.

Научная новизна проведенного диссертационного исследования заключается в следующих ключевых результатах. Разработаны новые нейросетевые методы восстановления потоковых данных, представленных в форме временных рядов, для офлайн и онлайн режимов, впервые использующие поведенческие шаблоны для повышения точности восстановления. Предложена новая функция потерь, учитывающая поведенческое сходство подпоследовательностей в процессе обучения. Предложен метод прогнозирования ошибки и времени обучения нейросетевых моделей восстановления временных рядов, основанный на представлении обучаемых моделей в виде направленного ациклического графа.

В автореферате корректно представлены основные положения диссертации. Его содержание соответствует материалам диссертационной работы.

Теоретическая значимость диссертационной работы обусловлена следующим. В разработанных методах восстановления потоковых данных предложены оригинальные архитектуры нейросетевых моделей, позволяющие использовать шаблоны, извлекаемые из исходных данных, в качестве дополнительного источника информации. Разработанная функция потерь учитывает поведенческую схожесть временных рядов во время обучения. Разработана нейросетевая модель, позволяющая прогнозировать ожидаемую ошибку и время обучения моделей восстановления многомерных временных рядов без необходимости их непосредственного обучения.

Практическая значимость диссертационной работы заключается в том, что разработанные нейросетевые модели и алгоритмы могут применяться для восстановления временных рядов в интеллектуальных системах обработки потоковых данных. Использование предложенных в работе методов позволит автоматизировать процесс подготовки данных, содержащих пропуски. Предложенные методы могут использоваться в системах AutoML и MLOps для повышения точности и устойчивости обработки потоковых данных.

Достоверность и обоснованность полученных результатов подтверждаются проведенными вычислительными экспериментами с использованием как реальных, так и синтетических данных. Эксперименты выполнены в соответствии с общепринятыми методологическими стандартами. Произведено сравнение полученных результатов с существующими аналогами.

Публикации и апробация. Основные результаты диссертационной работы отражены в 5 научных публикациях автора. Из них 4 статьи опубликованы в российских научных журналах, входящих в Ядро РИНЦ, в том числе 1 статья в зарубежном журнале, приравненном к журналам категории K1 Перечня ВАК, индексируемом в квартиле Q2 библиографической базы данных Scopus. Опубликованные работы в достаточной степени отражают содержание и ключевые результаты проведенного исследования.

Тематика проведенного исследования и полученные научные результаты соответствуют паспорту научной специальности 2.3.5 — «Математическое и

программное обеспечение вычислительных систем, комплексов и компьютерных сетей»: п. 4. Интеллектуальные системы машинного обучения, управления базами данных и знаний, инструментальные средства разработки цифровых продуктов.

В качестве **замечаний** к диссертационной работе А.А. Юртина, которые, однако, не снижают общей значимости полученных результатов диссертационного исследования, необходимо отметить следующие.

1. В разделах 2 и 3, посвященных описанию архитектуры разрабатываемых методов (а именно, SANNI, SAETI и tsGAP), не дано пояснения, почему в каждом случае была выбрана именно такая архитектура нейросети (например, почему автоэнкодер был выбран для SAETI, но не для SANNI). Также не указано, пробовались ли другие виды нейросетей и/или другие варианты их архитектур.
2. В разделе 4.3.1 не пояснено, почему метод SANNI проверялся на четырех сценариях формирования пропусков, а SAETI – только на одном. Аналогичный вопрос относительно функции потерь MDPE, которая проверялась только на сценарии Blackout (раздел 4.4.1).
3. Для функции потерь MDPE необходимо вручную подбирать значения параметров α и β , что несколько затрудняет работу с ней. Стоило привести в работе рекомендации по подбору этих значений либо реализовать возможности для автоматизации этого процесса.
4. В тексте диссертации присутствуют небольшие неточности. Например, в Заключение сказано, что «Вычислительные эксперименты, проведенные на реальных и синтетических данных, продемонстрировали превосходство метода tsGAP над передовыми аналогами в среднем на 38.8%», однако ранее нигде в тексте диссертации это значение не встречается. Также в тексте присутствуют пунктуационные ошибки.

ЗАКЛЮЧЕНИЕ

Диссертационная работа А.А. Юртина представляет собой самостоятельное и завершенное научно-квалификационное исследование. В диссертационной

работе представлены результаты, имеющие важное научное и практическое значение для систем интеллектуальной обработки потоковых данных. Результаты исследований, представленные в диссертации, вносят существенный вклад в решение актуальной проблемы повышения точности восстановления пропусков в потоковых данных, представленных в форме многомерных временных рядов. По своему содержанию, уровню выполненных исследований и полученным результатам диссертация соответствует требованиям, установленным Положением о порядке присуждения ученых степеней, в том числе пункту 9. На основании изложенного можно сделать вывод, что А.А. Юртин заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 2.3.5 — «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Официальный оппонент:



Вадим Владимирович Воеводин

«15» апреля 2026 г.

Кандидат физ.-мат. наук

заведующий Лабораторией анализа суперкомпьютерных систем и приложений
Научно-исследовательского вычислительного центра Федерального
государственного бюджетного образовательного учреждения высшего
образования «Московский государственный университет имени
М.В. Ломоносова», г. Москва

Адрес организации: 119234, Москва, Ленинские Горы, д. 1, стр. 4

Телефон: + 7 (495) 939-52-16

Email: vadim@parallel.ru

