

**ОТЗЫВ официального оппонента
на диссертацию на соискание ученой степени
кандидата физико-математических наук
Шайхисламова Дениса Ильгизовича на тему: «Исследование и
разработка методов для сравнительного анализа суперкомпьютерных
приложений на основе технологий интеллектуального анализа данных»
по специальности 2.3.5. Математическое и программное обеспечение
вычислительных систем, комплексов и компьютерных сетей**

Актуальность диссертационного исследования

Диссертационная работа Шайхисламова Д. И. посвящена актуальной проблеме повышения эффективности эксплуатации суперкомпьютерных систем. Современные вычислительные комплексы генерируют огромные объемы данных мониторинга и системного журналирования, однако ручной анализ этих данных крайне затруднителен. В этих условиях разработка автоматизированных методов, основанных на интеллектуальном анализе данных (Data Mining) и способных выявлять скрытые закономерности в потоке заданий, является важной научно-технической задачей.

Особую значимость работе придает фокус на задаче поиска схожих приложений. Как справедливо отмечает автор, способность идентифицировать задания, обладающие общими свойствами, открывает широкие возможности для прогнозирования поведения новых заданий, выявления аномальных прогонов, кластеризации пользовательских задач и оптимизации использования суперкомпьютерных ресурсов. Разработка таких методов, их программная реализация и интеграционное тестирование на реальном суперкомпьютере петафлопсного уровня «Ломоносов-2» определяет высокую актуальность диссертационного исследования.

Степень обоснованности научных положений, выводов и рекомендаций

Научные положения, выводы и рекомендации, сформулированные в диссертации, являются обоснованными. Это подтверждается следующими факторами:

1. *Корректным применением математического аппарата.* Автор использует апробированные методы машинного обучения (Doc2Vec, Dynamic Time Warping), статистического анализа и кластеризации, адаптируя их для решения специфических задач суперкомпьютерного мониторинга.
2. *Масштабной экспериментальной проверкой.* Все предложенные методы прошли тщательное тестирование на реальных данных суперкомпьютера «Ломоносов-2». Эксперименты проводились на выборках, насчитывающих тысячи заданий, что обеспечивает статистическую значимость полученных результатов.
3. *Сравнением с существующими подходами.* Качество разработанных методов оценивалось путем сравнения с ручной разметкой (для кластеризации) и с данными существующей системы XALT (для обнаружения пакетов прикладных программ), что позволяет объективно судить о значимости вклада автора.
4. *Практической реализацией.* Созданное программное решение внедрено в эксплуатацию и используется администраторами суперкомпьютерного центра для ежедневного анализа потока заданий, что является наилучшим подтверждением обоснованности и практической ценности работы.

Достоверность и новизна научных положений, выводов и рекомендаций

Достоверность полученных результатов не вызывает сомнений и обеспечивается использованием общепризнанных метрик качества (ARI, MAE, MSE), воспроизводимостью экспериментов на репрезентативных

выборках данных, а также успешным функционированием разработанного программного обеспечения на действующей суперкомпьютерной системе.

Научная новизна диссертационного исследования заключается в следующем:

1. Разработаны новые методы выявления схожих суперкомпьютерных приложений:
 - Статический метод, основанный на векторном представлении имен функций и переменных, извлеченных из исполняемых файлов. Новизна заключается в адаптации модели для анализа именно НРС-приложений и демонстрации ее эффективности для решения прикладных задач.
 - Динамический метод, базирующийся на алгоритме Dynamic Time Warping, для сравнения временных рядов. Новизна подхода состоит в комплексной многоэтапной предобработке данных (обработка пропусков, логарифмирование, масштабирование), позволившей достичь высокой точности (0.95) и устойчивости к временным сдвигам и различной длительности заданий.
2. Предложены новые подходы к решению практически важных задач на основе разработанных метрик схожести:
 - Разработан новый метод обнаружения используемых прикладных программных пакетов, который позволил выявить на 40% больше заданий с пакетами по сравнению с существующей системой XALT, а также обнаружить новый пакет XAMG для решения систем линейных алгебраических уравнений.
 - Разработан комбинированный (статический и динамический) метод предсказания метрик качества использования суперкомпьютерных ресурсов, позволивший получать оценки для 87% заданий со средней ошибкой 2.5%, что существенно снижает накладные расходы на сбор данных во время запусков приложений.
 - Предложена двухэтапная методика кластеризации приложений (сначала по статике, затем по динамике), на основе которой созданы

алгоритмы выделения аномалий в сериях запусков похожих заданий и анализа крупных групп заданий.

Практическая значимость

Практическая значимость работы не вызывает сомнений. Автором создано программное решение, которое интегрировано в инфраструктуру суперкомпьютера «Ломоносов-2». Решение позволяет:

- Ежедневно формировать отчеты о случаях использования программных пакетов, не выявляемых системой XALT.
- Предсказывать оценки эффективности использования суперкомпьютерных ресурсов для заданий без больших затрат на динамический сбор информации об особенностях их выполнения.
- Проводить регулярную кластеризацию потока заданий для выявления крупных групп заданий и аномальных запусков.

Разработанные алгоритмы и архитектура программного обеспечения могут быть тиражированы и адаптированы для использования на других вычислительных системах, что подтверждает их высокую практическую ценность.

В качестве **замечаний** к диссертационной работе можно отметить следующие моменты.

1. Не ясно, что понимается под термином «схожие» в отношении заданий на рис. 3.1 (стр. 29 диссертации). Схожие задания попадают в один кластер по статическим характеристикам или у них похожие графики промахов в кэш-память и загрузки ЦПУ? Если второе, то как измеряется «похожесть» графиков?
2. В разделе 3.3.3, на стр. 44-45 рассматривается применение автокодировщика для уменьшения размерности данных. Автор приводит результаты его работы (таблица 2), но не описывает

архитектуру сети (количество слоев, функции активации, параметры обучения), что затрудняет воспроизведение этого результата. Более того, обоснованный отказ от автокодировщика в пользу ручного масштабирования на стр. 48 ставит вопрос о целесообразности обсуждения его использования в главе 3.

3. В разделе 4.1.3 для оценки точности предложенного метода на реальных данных в таблицах 4 и 5 (стр. 56, 58) используются метрики MAE и MSE. Почему также не использовать популярную метрику MAPE (Mean absolute percentage error), которая измеряет среднюю абсолютную процентную ошибку между прогнозируемыми и реальными значениями?
4. На стр. 63 диссертации читаем «После сравнения результатов предложенного метода с системой XALT, была получена точность 0.77». Точность чего показывает эта цифра? Откуда взята формула 4.6?
5. В таблицах 3, 4 и 5 (стр. 50, 56, 58) используется аббревиатура AE, которая нигде не расшифровывается.

Указанные замечания не снижают общей значимости результатов диссертационного исследования, выполненного Д.И. Шайхисламовым.

Заключение

Диссертация отвечает требованиям, установленным Московским государственным университетом имени М.В.Ломоносова к работам подобного рода. Содержание диссертации соответствует специальности 2.3.5. Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей (по физико-математическим наукам), а также критериям, определенным пп. 2.1-2.5 Положения о присуждении ученых степеней в Московском государственном университете имени М.В.Ломоносова. Диссертационное исследование оформлено согласно требованиям Положения о совете по защите диссертаций на соискание

ученой степени кандидата наук, на соискание ученой степени доктора наук Московского государственного университета имени М.В.Ломоносова.

Таким образом, соискатель Шайхисламов Денис Ильгизович 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:

Доктор физико-математических наук, профессор,
заведующий кафедрой системного программирования
Федерального государственного автономного образова-
тельного учреждения высшего образования «Южно-
Уральский государственный университет (национальный
исследовательский университет)»

Соколинский Леонид Борисович

24 февраля 2026 г.

Контактные данные:

тел.: 7(351)2723080, e-mail: leonid.sokolinsky@susu.ru

Специальность, по которой официальным оппонентом защищена диссертация: 05.13.18 Математическое моделирование, численные методы и комплексы программ

Адрес места работы:

454080, г. Челябинск, просп. В.И. Ленина, д. 76,
Южно-Уральский государственный университет

Тел.: 7(351) 267-99-00; e-mail: leonid.sokolinsky@susu.ru