

ОТЗЫВ

официального оппонента на диссертацию

Воробьевой Гульнары Равилевны

на тему

«Методологические основы обработки неоднородной пространственно-временной информации в системах поддержки принятия решений на основе технологий больших данных (на примере геомагнитных данных)»
по специальности 05.13.01 – Системный анализ, управление и обработка информации (информационные и технические системы)
на соискание ученой степени доктора технических наук

- **Актуальность темы исследования**

Исследование, результаты которого приведены в диссертационной работе, направлено на решение проблемы информационной перегруженности лиц, принимающих решения в ситуациях, требующих мониторинга и анализа состояния процессов различной природы происхождения. Во многих прикладных сферах, например, в области техносферной безопасности (в работе в качестве сквозного примера рассматривается геомагнитный мониторинг) обозначенная проблема может стать критической, поскольку может серьезно повлиять на скорость принимаемых решений. Ситуация еще более осложняется тем, что указанные данные в значительной степени описывают процесс, характеризующиеся выраженной пространственной и (или) временной неоднородностью, содержат многочисленные пропуски и велики по хранимым объемам. В этой связи разработка методологических основ обработки неоднородной пространственно-временной информации в системах поддержки принятия решений представляется актуальной.

- **Оценка структуры и содержания работы**



Структура и содержание диссертационной работы находятся в логическом единстве и соответствуют поставленным целям и задачам исследования. Диссертационная работа состоит из введения, шести глав, заключения, списка терминов, списка литературы и приложения с актами внедрения.

Во введении обоснована актуальности, описаны практическая и теоретическая ценности диссертационной работы. Здесь же формулируются цель и задачи работы; представлены положения, выносимые на защиту; изложены краткая характеристика и сведения об апробации работы.

В первой главе обсуждается проблема обработки пространственно-временной информации из распределенных неоднородных источников, обуславливающая как информационную перегруженность принимающих решения лиц, так и снижение эффективности соответствующих систем поддержки принятия решений (ориентированных на данные). Для решения обозначенной проблемы автором предложены и сформулированы методологические основы обработки неоднородной пространственно-временной информации, которые позволяют повысить оперативность соответствующих систем поддержки принятия решений за счет выделения критериев пространственной зависимости и гетерогенности для источников данных и их кластеризации на этой основе.

Вторая глава посвящена разработке моделей и методов обработки данных при интеграции гетерогенных источников в составе системы поддержки принятия решений. Автором предлагается концепция создания единого информационного пространства с гибридным информационным хранилищем, которое представлено статической и динамической составляющей, переход между которыми осуществляется по принципу старения информации по модели Бартона-Кеблера. Предлагаемые модели и методы обработки информации при реализации предполагают интеграцию данных из разнородных источников и позволяют ускорить доступ пользователей к данным без необходимости предварительной загрузки, обработки и фильтрации из больших наборов.

В третьей главе обсуждаются предложенные автором методы восстановления временных рядов данных (на примере геомагнитной информации) в информационных системах на основе принципов машинного обучения и информационного резервирования источников информации, обеспечивающих возможность импутации пропусков с ошибкой в пределах допустимого нормативами отклонения. Предложены и сформулированы два метода импутации пропусков. Первый – метод резервной станции – основан на составлении доверительных списков источников данных по результатам теоретико-множественного, теоретико-информационного и статистического анализа синхронно регистрируемых ими данных. Второй – индуктивный – сочетает особенности методов машинного обучения и прецедентов для восстановления пропусков на основании зарегистрированных в тех же условиях данных источника. Восстановленные одним из предложенных способов данные размещаются в едином информационном пространстве и могут быть использованы для обработки, анализа и визуализации в процессе принятия решений.

Четвертая глава посвящена обсуждению особенностей предложенной автором модели хранения данных в системах поддержки принятия решений. Модель основана на сочетании реляционного и нереляционного подходов. При этом она отличается тем, что, с целью повышения реактивности программных средств и сокращения затрат физической памяти, реляционная, иерархическая и колончатая модели данных объединены на базе правил ссылочной целостности, а также комбинирования текстового и бинарного форматов описания как собственно данных, так и их метаданных.

В пятой главе на примере геомагнитных данных рассматривается алгоритм визуализации пространственно-временного распределения данных на основе компьютерных методов обработки информации. Автором предлагается метод построения локальной координатной сетки, суть которого заключается в том, что ограниченное регионом пространство заполняется пространственными объектами, разделенные равными интервалами и описанные вдоль собственных

границ упорядоченным множеством размещенных с одинаковым разрешением пространственных точек. Также предлагается усовершенствовать известные подходы к визуальному представлению геопространственных данных применительно к рендерингу геомагнитных данных посредством разработки адаптивного алгоритма, формирующего изображение на основании пространственной привязки данных.

В шестой главе приводятся результаты оценки эффективности предложенных теоретических основ, методов, алгоритмов, а также специального математического и алгоритмического обеспечения на основе разработанного прототипа веб-ориентированного инструментально-программного средства (www.geomagnet.ru) и выделенных показателей эффективности. Эмпирически на примере геомагнитных данных известных магнитных сетей (в частности, INTERMAGNET) установлена эффективность полученных в ходе диссертационного исследования результатов, в частности, достигнуто повышение вычислительной скорости обработки, анализа и визуализации данных, сокращены вычислительные затраты на их хранение, а также сокращено время выполнения запросов к разнородным источникам за счет использования единого информационного пространства.

В заключении представлены основные результаты и выводы и определены перспективы дальнейшего развития.

• Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации

Научные положения диссертационной работы соответствуют целям и задачам исследования, обоснованию теоретических и методических положений по обработке неоднородной пространственно-временной информации в системах поддержки принятия решений.

Обоснованность и достоверность выводов и результатов диссертации подтверждается использованием широко апробированного научного инструментария: положений системного анализа, управления, обработки

информации, а также анализом трудов отечественных и зарубежных ученых по тематике диссертационного исследования.

Кроме того, ознакомление с изложенным в диссертации научно-практическим материалом позволяет сделать вывод о том, что результаты исследования, выводы и рекомендации являются научно обоснованными.

Результаты исследований опубликованы автором более чем 100 научных работах, в том числе 13 статьях в рецензируемых журналах из списка ВАК; 22 статьях в изданиях, индексируемых международными системами Scopus / Web of Science; 3 монографиях, изданных в России и за рубежом; 2 патентах на изобретение; 5 свидетельствах о государственной регистрации программы для ЭВМ, трудах конференций и др.

- **Достоверность и новизна полученных результатов;**

Достоверность полученных научных результатов обеспечивается строгостью применяемого математического аппарата, результатами математического и компьютерного моделирования, подтверждается при обработке и анализе оригинальных геомагнитных данных, предоставляемых территориально распределенными гетерогенными источниками на примере магнитных станций и обсерваторий международной сети INTERMAGNET, а также результатами обработки и визуализации геомагнитных данных в рамках разработанного прототипа веб-ориентированного инструментально-программного средства Geomagnet (<https://www.geomagnet.ru>).

Научная новизна результатов заключается в следующем.

1. Методологические основы обработки неоднородной пространственно-временной информации отличаются тем, что, с целью повышения оперативности доступа к информации, необходимой для принятия решений, выделяются критерии пространственной зависимости и пространственной гетерогенности для групп источников данных и на этой основе подстраиваются процессы сбора, анализа и визуализации информации.

2. Комплекс моделей и методов обработки информации при интеграции гетерогенных источников данных в гибридные хранилище систем поддержки принятия решений отличается тем, что, с целью повышения вычислительной скорости сбора и обработки данных, преобразование оперативной информационной составляющей в постоянную определяется адаптированной моделью старения информации Бартона–Кеблера с исключенным динамическим компонентом.

3. Методы восстановления временных рядов данных, включающие:

– индуктивный метод, отличающийся тем, что, с целью повышения точности и скорости восстановления данных, наиболее вероятные значения определяются на базе статического сходства между массивом, образованным предшествующими и последующими за пропущенным фрагментом значениями, и массивами, построенными аналогично из известных значений;

– метод информационного резервирования источников данных, отличающийся тем, что, с целью обеспечения полноты временных рядов, наиболее вероятные значения определяются посредством формирования доверительного списка на основании оценки пространственной гетерогенности и зависимости синхронно регистрируемых данных, а также сравнительной оценки фрагментов рядов, зарегистрированных в предшествующий момент.

4. Модель хранения данных отличается тем, что, с целью повышения реактивности программных средств и сокращения затрат физической памяти, реляционная, иерархическая и колончатая модели данных объединены на базе правил ссылочной целостности, а также комбинирования текстового и бинарного форматов описания как собственно данных, так и их метаданных.

5. Алгоритм визуализации пространственно-временного распределения геомагнитных данных отличается тем, что для клиентского веб-рендеринга больших пространственных данных учитывается их пространственная анизотропия посредством комбинирования подходов, демонстрирующих наилучшие показатели реактивности в соответствующих пространственных областях.

- **Теоретическая и практическая значимость полученных автором результатов**

Теоретическая значимость полученных автором результатов подтверждается тем, что в работе предложен подход к построению подсистем извлечения данных для информационных систем поддержки принятия решений (преимущественно ориентированных на данные), обобщающий результаты, полученные для отдельных моделей, методов и алгоритмов решения задач обработки, анализа и визуализации пространственно-зависимых данных, и позволяющий получить качественно новые решения для некоторых из них. В частности, предложены методологические основы обработки неоднородной пространственно-временной информации, необходимой для принятия решения, включающие в себя особенности теоретико-множественных, статистических и теоретико-информационных зависимостей как между элементами самих данных, так и их источниками.

Практическая значимость полученных автором результатов подтверждается представленными в диссертационной работе актами их внедрения на предприятия и организации различного профиля.

- **Замечания по диссертационной работе**

1. В теме диссертации заявлены: обработка пространственной информации (координаты, топология и т.д.) и больших данных, а в автореферате эти понятия не раскрыты.

2. Во второй главе диссертации заявлено повышение оперативности обработки, но при этом не приведено никаких оценок улучшения этого показателя.

3. В главе 3 непонятно, в чем новизна структурного сдвига фрагментов T (как и на стр. 20 автореферата).

4. В главе 4 отсутствуют оценки сокращения временных затрат на вычисления и повышения скорости обработки запросов.

5. В главе 1 данные разделены на глобальные, региональные и локальные. Непонятно, для чего во второй главе они же разделены на аналитические и оперативные?

Указанные замечания не снижают ценности диссертационного исследования и не влияют на общую положительную оценку.

• **Заключение** о соответствии диссертации критериям, установленным в «Положении о присуждении ученых степеней»:

Диссертация Воробьевой Гульнары Равиловны на соискание ученой степени доктора технических наук является научно-квалификационной работой, в которой на основании выполненных автором исследований решена научная проблема повышения эффективности процессов обработки информации в системах поддержки принятия решений, имеющая важное хозяйственное значение, что соответствует требованиям п. 9 «Положение о присуждении ученых степеней», а ее автор заслуживает присуждения искомой ученой степени.

Официальный оппонент:

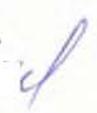
доктор технических наук, доцент
Андрианов Дмитрий Евгеньевич,
заместитель директора по учебной работе,
заведующий кафедрой информационных систем,
Муромского института (филиала)
ФГБОУ ВО «Владимирский государственный
университет имени Александра Григорьевича
и Николая Григорьевича Столетовых»



16.10.20

Докторская диссертация защищена по специальности 05.13.01 – Системный анализ, управление и обработка информации

Подпись д.т.н., доцента Андрианова Д.Е. заверяю.
Секретарь Ученого совета института



О.Н. Полулях



Адрес места основной работы: Россия, 602264, Владимирская область, г. Муром, ул. Орловская, д.23
Рабочий телефон: +7(49234) 77-1-12
Адрес эл. почты: zamdir_ur@mivlgu.ru