

ОТЗЫВ

официального оппонента на диссертацию

Воробьевой Гульнары Равиленовны

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 обусловлена острой необходимостью решения проблемы информационной избыточности лиц, принимающих решения в прикладных областях, в частности, техносферной безопасности (в работе рассматривается в качестве примера частный случай – геомагнитный мониторинг). Непрерывно растущий объем данных, задействованных в системах поддержки принятия решений, повышение сложности их структуры, появление новых гетерогенных источников данных – одни из многих факторов, обуславливающих обозначенную проблему. В работе Воробьевой Г.Р. предложен вариант ее решения, что в целом говорит об актуальности и значимости проведенных для этого научных исследований.

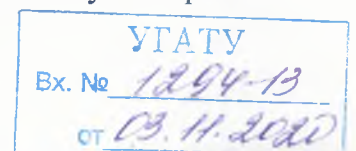
Оценка структуры и содержания работы

Диссертационная работа Г.Р. Воробьевой обладает целостностью, внутренним единством, а также характеризуется логической последовательностью изложения.

В работе представлены введение, 6 глав, список используемых терминов, заключение и приложение с актами внедрения результатов проведенных исследований.

Введение посвящено обоснованию актуальности, практической и теоретической ценности диссертационной работы. Здесь формулируются цель и задачи работы; представлены положения, выносимые на защиту; изложены краткая характеристика и сведения об апробации работы.

Первая глава посвящена анализу проблемы обработки информации из распределенных гетерогенных источников в процессе принятия решений. Автором предложены методологические основы подхода к анализу и обработке



пространственного распределения данных в информационных системах поддержки принятия решений на основе теоретико-множественного, теоретико-информационного и статистического подходов. Предложенные решения отличаются тем, что выделяются критерии пространственной зависимости и пространственной гетерогенности для групп источников данных и на этой основе подстраиваются процессы сбора, анализа и визуализации информации.

Во второй главе обсуждаются модели и методы обработки данных при интеграции гетерогенных источников, целью реализации которых является повышение оперативности получения данных, необходимых для принятия решений, в частности, в области техносферной безопасности. Предложен подход к построению единого информационного пространства при интеграции гетерогенных источников данных в гибридные хранилище систем поддержки принятия решений на основе принципов консолидации и федерализации. Представленные решения отличаются 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.

Третья глава посвящена обсуждению предложенных методов 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. Рассмотрены два метода. Первый из них основан на принципах информационного резервирования источников данных со сходными теоретико-множественными, теоретико-информационными и статистическими характеристиками. Второй способ предполагает восстановление пропущенного фрагмента по известной паре предшествующего и последующего за ним, рассматриваемых в качестве прецедентов.

В четвертой главе обсуждаются особенности предложенной автором модели хранения данных в системах поддержки принятия решений, обеспечивающего сокращение вычислительных затрат на их физическое размещение и повышение скорости обработки запросов к данным. Модель сочетает в себе реляционный и нереляционные подходы, что обеспечивает как отделение метаданных от хранимой информации, так и повышение вычислительной скорости доступа к данным за счет использования колоночно-ориентированного формата.

В пятой главе на примере геомагнитных данных рассматривается алгоритм визуализации пространственно-временного распределения данных на

основе компьютерных методов обработки информации. Отличительной особенностью алгоритма является то, что для клиентского веб-ориентированного рендеринга больших пространственных данных учитывается их пространственная анизотропия посредством комбинирования подходов, демонстрирующих наилучшие показатели реактивности в соответствующих пространственных областях.

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 и выделенных показателей эффективности.

Показано достижение поставленной в работе цели повышения эффективности процессов обработки информации в системах поддержки принятия решений посредством разработки единых методологических основ обработки, анализа и визуализации больших объемов пространственных данных, полученных из территориально распределенных гетерогенных источников.

В заключении приводятся основные результаты и выводы по проведенной работе.

Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации. В процессе ознакомления с диссертацией и авторефератом установлено, что диссертант провел большую работу по исследованию и анализу научно-методических работ по проблематике диссертации, практических материалов по обработке неоднородной пространственно-временной информации.

В ходе исследования Г.Р. Воробьевой было использовано большое количество источников, включающих научные труды отечественных и зарубежных авторов, нормативные документы и публикации в электронных ресурсах. Это позволило автору сделать научно-обоснованные выводы и осуществить разработку практически значимых методических основ обработки неоднородной пространственно-временной информации.

Научная обоснованность положений диссертационного исследования также определяется структурой и логикой изложения материала. Содержание диссертации соответствует поставленной цели, а именно – повышению эффективности процессов обработки информации в системах поддержки принятия решений посредством разработки единых методологических основ обработки, анализа и визуализации больших объемов пространственных

данных, полученных из территориально распределенных гетерогенных источников.

Сформулированные в диссертации задачи исследования, примененные методы, представленные на защиту основные научные положения полностью соответствуют поставленной цели, что обеспечило безусловное ее выполнение и подтверждает обоснованность научных положений и выводов.

Достоверность и новизна полученных 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 по существу результатов:

1. Методологические основы обработки неоднородной пространственно-временной информации, отличающиеся критериями пространственной зависимости и пространственной гетерогенности для групп источников данных, что позволило модернизировать процессы сбора, анализа и визуализации информации и повысить оперативность доступа к информации, необходимой для принятия решений.

2. Комплекс моделей и методов обработки информации при интеграции гетерогенных источников данных в гибридное хранилище систем поддержки принятия решений отличающийся способом преобразования оперативной информационной составляющей в постоянную на основе адаптированной модели старения информации Бартона–Кеблера с исключенным динамическим компонентом, что позволило снизить время на сбор и обработку исходных данных.

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 и метаданных, отличающаяся объединением реляционной, иерархической и колончатой моделей данных на базе правил ссылочной целостности, а также комбинированием текстового и бинарного форматов, позволяющая сократить физическую память для хранения данных и метаданных.

5. Метод визуализации пространственно-временного распределения геомагнитных данных, отличающийся включением результатов пространственной анизотропии и позволяющий снизить время рендеринга при полной визуализации геомагнитных данных.

Теоретическая и практическая значимость полученных автором результатов. В подтверждение теоретической значимости в работе предложены методологические основы обработки неоднородной пространственно-временной информации как парадигмы обработки и анализа данных на основе их пространственного распределения, базирующейся на совокупности согласованных методов и подходов формирования единого информационного пространства в виде множества восстановленных пространственных данных заданного формата, а также результатов их графической и аналитической обработки. Практическая значимость подтверждается имеющимися в диссертационной работе актами внедрения в организации и на предприятия различного профиля.

Замечания по диссертационной работе

1. В первой главе приводится по сути описание метода и результатов разведочного анализа исходных данных. Автору стоило убедительнее обосновать необходимость использования теоретическо-множественного описания как инструмента системного анализа, так как фактически рассматривалось статистическое представление объекта исследования.
2. В работе рассматриваются интересные особенности исследуемых источников данных – разделение на уровни (региональные, глобальные) и различное географическое расположение (различные часовые пояса). Несмотря на приведенную формализацию в диссертации опущены очень важные вопросы, возникающие в процессе сбора исходных данных. В частности: 1) каким образом осуществлялась консолидация данных с различных источников в какой-либо географической окрестности, при этом отнесенным к разным группам? 2) Каким образом осуществлялась

оркестрация процессов опроса источников данных с учетом различных временных поясов и описанных в работе недостатков в исходных данных? 3) Каким образом осуществлялась оркестрация процессов восстановления данных на основе описанного подхода?

3. При выборе источников данных для резервирования используются анализ пространственной однородности, анализ взаимных энтропийных характеристик данных, анализ информационной корреляции. Автор не приводит условий и требований к исходным данным при котором этот анализ возможен (при условии 30% пропусков в данных в среднем может сложиться ситуация, когда выбрать достоверные источники будет затруднительно из-за короткой длины временного ряда или из-за «непересекающихся» по времени временных рядов или в случае наличия нескольких источников-кандидатов).
4. В работе не хватает детальной информации об исходных данных и проведенных экспериментах: объеме исходных данных, интенсивности поступления, времени задержек при сборе данных. Указанные в диссертации результаты, обосновывающие эффективность, получены только для 1-2 –х станций при однократном запросе. Не приведены в должной мере количественные оценки временных задержек, анализа изменения времени выполнения запроса в зависимости от объема данных и уровня детализации данных. Работу несомненно усилило включение данной информации.
5. В диссертации много повторений уже сформулированного ранее. Это относится как к тексту (особенно про важность решаемой задачи и теоретико-множественного подхода), так и к формулам. Также приводится в излишнем объеме уже известная информация, хотя достаточно было бы ссылок на литературу. Кроме этого, автору стоило внимательнее отнестись к результатам формализации: в формулах присутствуют опечатки, одни и те обозначения вводятся для разных сущностей.

Указанные замечания не влияют на общую положительную оценку диссертации Г.Р. Воробьевой.

Заключение о соответствии диссертации критериям, установленным «Положение о присуждении ученых степеней»:

Диссертация Воробьевой Гульнары Равилевны на соискание ученой степени доктора технических наук является научно-квалификационной работой, в которой решена научная проблема повышения эффективности процессов обработки информации в системах поддержки принятия решений посредством разработки единых методологических основ обработки, анализа и визуализации

больших объемов пространственных данных, полученных из территориально распределенных гетерогенных источников, 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
Щербаков Максим Владимирович,
заведующий кафедрой,

кафедра "Системы автоматизированного проектирования
и поискового конструирования",
Федеральное государственное бюджетное образовательное
учреждение высшего образования «Волгоградский
государственный технический университет»

15.10.2020

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

Адрес места основной работы: 400005, г. Волгоград,
проспект им. В.И. Ленина, д. 28.

Рабочий телефон: +7(8442) 24-81-00

Адрес эл. почты: maxim.shcherbakov@vstu.ru



Подпись	<i>Щербаков М.В.</i>
УДОСТОВЕРЯЮ	<i>15.10.2020</i>
Нач. общего отдела	<i>А. А. А.</i>
	(подпись)