

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

доктора технических наук, профессора, заведующей кафедрой информационных технологий в менеджменте Института «Высшая школа менеджмента Санкт-Петербургского государственного университета»

Гавриловой Татьяны Альбертовны

на диссертационную работу Тушкановой Ольги Николаевны на тему «Семантические структуры и причинные модели больших данных для принятия решений с приложением к рекомендательным системам», представленную на соискание ученой степени кандидата технических наук по специальности 05.13.01 – «Системный анализ, управление и обработка информации (технические системы)»

Актуальность темы

Тематика диссертационного исследования Тушкановой Ольги Николаевны актуальна и востребована в связи с тем, что в настоящее время в самых разнообразных прикладных областях накоплен значительный объем разнообразных данных, которые представляют коммерческую и научную ценность, и объем этих данных продолжает расти по экспоненциальному закону. В то же время большинство существующих статистических методов, методов машинного обучения и других алгоритмов интеллектуального анализа данных не в состоянии справиться с задачами обработки больших данных из-за вычислительной неустойчивости, вычислительной сложности и ряда других причин, обусловленных особенностями этих данных. Особенно сложной эта задача оказывается в связи с необходимостью семантической интерпретации результатов обработки, что в настоящее время является общим требованием в этой области. Как известно, семантические аспекты представления моделей данных и результатов обработки достаточно часто решаются с использованием онтологий или концептуальных моделей.

Однако, для больших данных традиционно используемые методы построения онтологий, в которых основная роль отводится экспертам, также практически нереализуемы по ряду причин. Одна из них - это трудоемкость этой задачи для экспертов. Присутствие в данных текстов на естественном языке также усложняет задачу экспертов, причем многократно. Существуют и другие сложные проблемы обработки больших данных, однако перечисленные проблемы должны решаться в первую очередь и они составляют содержание исследования рецензируемой работы.

Целью диссертационной работы Тушкановой О.Н. является разработка алгоритмов машинного обучения и принятия решений в задачах классификации на основе семантических структур и причинных моделей больших данных, что следует квалифицировать как актуальную научную задачу, которая имеет существенное значение для науки и практики.

Степень обоснованности научных положений, выводов и рекомендаций

Тема диссертации, направленность выполненных исследований и результатов соответствуют пункту 4 «Разработка методов и алгоритмов решения задач системного анализа, оптимизации, управления, принятия решений и обработки информации» и пункту 5 «Разработка специального математического и алгоритмического обеспечения систем анализа, оптимизации, управления, принятия решений и обработки информации» паспорта специальности 05.13.01 - «Системный анализ, управление и обработка информации (технические системы)».

Автором сформулированы и вынесены на защиту следующие научные положения:

1. Теоретически и экспериментально обоснованная мера оценки силы причинной связи обеспечивает вычислительную эффективность и масштабируемость алгоритмов поиска причинных связей в больших данных.

2. Вычислительно эффективный алгоритм автоматического построения онтологии позволяет построить семантическую модель больших данных, пригодную для обучения и минимизации причинных моделей в задачах ассоциативно-причинной классификации.

3. Семантическая модель больших данных представляет мета-свойства данных, их синтаксис и семантику в рамках единой структуры и обеспечивает доступ к большим данным в задачах обучения и принятия решений на основе ассоциативно-причинной классификации.

4. Алгоритм поиска причинных связей в данных является вычислительно эффективным и масштабируемым, что обеспечивается выбранной мерой оценки силы причинной связи и использованием семантической модели данных.

5. Алгоритм минимизации размерности пространства причинных правил модели АПК реализует снижение её избыточности и минимизацию размерности практически без потери точности.

Обоснованность сформулированных Тушкановой О.Н. научных положений, а также выводов и результатов, полученных в ходе диссертационного исследования, подтверждается корректностью используемых и разработанных методов и методик, которые детально описаны в работе, тщательной экспериментальной проверкой полученных результатов, а также апробацией полученных результатов на международных конференциях с высоким рейтингом и их публикацией в рецензируемых изданиях.

Оценка новизны и достоверности

При решении задач, сформулированных в работе, автором получены новые научные результаты, которые обладают рядом существенных отличий.

1. Обоснован выбор меры оценки силы причинных связей в данных. Выбранная мера является вычислительно эффективной и семантически кор-

ректной, что подтверждается теоретическими и экспериментальными исследованиями.

2. Разработан алгоритм автоматической генерации семантической модели больших данных, в основу которого положена новая методика семантического анализа понятий. Семантический анализ понятий построен как комбинация методов извлечения релевантных данным понятий глобальной онтологии данных и генерации двойственных им формальных понятий.

3. Предложена новая семантическая модель больших данных. Модель строится в ходе семантического анализа понятий и позволяет отразить метасвойства данных и их семантику в рамках одной структуры.

4. Разработан новый алгоритм поиска причинных зависимостей в данных. Алгоритм использует семантическую модель данных, а именно, метасвойства данных, что обеспечивает его вычислительную эффективность.

5. Разработан алгоритм минимизации количества извлеченных причинных правил, основанный на методах кластеризации. Алгоритм позволяет повысить вычислительную эффективность ассоциативно-причинной классификации путем уменьшения избыточности, несмотря на некоторую потерю точности.

Достоверность основных научных положений и результатов диссертационной работы обеспечены критическим обзором исследований в области построения онтологий, ассоциативного и причинного анализа, обработки больших данных, апробацией в докладах на российских и международных конференциях и в печатных трудах, согласованностью теоретических положений и выводов с экспериментальными результатами.

Теоретическая и практическая значимость

Теоретическая значимость диссертационного исследования заключается в разработке новых, обоснованных и экспериментально проверенных моделей и алгоритмов обучения и принятия решений в области ассоциативно-

причинной классификации, которые эффективно работают применительно к большим данным, а также в разработке семантического анализа понятий – методики и алгоритма автоматического построения семантической модели данных.

Практическая значимость работы заключается в реализации всех разработанных методик, моделей и алгоритмов в виде программной библиотеки на языке Java с ориентацией на решение задач в области рекомендательных систем. Программная библиотека также позволяет повторное использование разработанных программных компонент при решении других задач обработки больших данных и принятия решений. Кроме того, по результатам тестирования, выполненного в ходе решения поставленных задач диссертационного исследования, предложены практические рекомендации по выбору алгоритмов реализации современных рекомендательных систем.

Апробация и внедрение полученных результатов

Результаты диссертационного исследования были представлены в виде докладов на пяти российских и международных научных конференциях, где получили положительную оценку. По теме диссертационной работы опубликовано 9 печатных работ, в том числе, 3 публикации в рецензируемых научных изданиях из перечня ВАК, а именно, «Труды СПИИРАН», «Информационные технологии и вычислительные системы», 4 публикации в изданиях, индексируемых в международных базах данных Web of Science и Scopus.

Результаты исследования использованы при реализации проектов по нескольким контрактам, что подтверждается соответствующими актами внедрения.

Замечания по диссертации

По диссертационной работе имеется ряд замечаний:

1. Разработанный алгоритм семантического анализа понятий в полностью автоматическом режиме может быть использован только для по-

строения онтологии текстовых данных. В остальных случаях, по всей видимости, все же требуется привлечение экспертов.

2. Работа содержит большое количество разделов описательного характера, не связанных непосредственно с разработкой алгоритмов.

3. В диссертации в явном виде не описано, как именно разработанная семантическая структура понятий позволяет представить синтаксис данных.

4. В разделах 1.3 и 1.6 сформулированы основные проблемы в области построения моделей принятия решений на основе больших данных, однако в работе явно не показано, каким образом разработанные алгоритмы позволяют преодолеть эти проблемы хотя бы частично.

5. Не слишком удачно выбран способ цитирования, без фамилий и года, что заставляет при чтении все время сверяться с библиографией.

6. Вызывает удивление выбор примера для экспериментов с моделью – данные Amazon 2006, т.е. десятилетней давности.

7. В автореферате не описан пример работы алгоритмов, что существенно снижает уровень понимания при чтении.

Приведенные замечания не снижают общий высокий научный уровень проведенных исследований и носят рекомендательный характер.

Заключение

Диссертационная работа Тушкановой Ольги Николаевны представляет собой законченную научно-квалификационную работу, выполненную на актуальную тему. Полученные в ходе исследования результаты обладают теоретической и практической значимостью и научной новизной.

В диссертационной работе Тушкановой Ольги Николаевны сформулирована и решена актуальная научная задача разработки алгоритмов обработки больших данных для построения моделей ассоциативно-причинной класси-

фикации с демонстрацией их работоспособности в приложениях из области рекомендательных систем.

Оформление диссертации отвечает требованиям, предъявляемым к работам, направляемым в печать. Основное содержание работы достаточно полно отражено в автореферате.

Диссертационная работа Тушкановой Ольги Николаевны отвечает требованиям, установленным п. 9 Положения о присуждении ученых степеней, утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 № 842, предъявляемым к кандидатским диссертациям, а ее автор заслуживает присуждение ученой степени кандидата технических наук по специальности - 05.13.01 – «Системный анализ, управление и обработка информации (технические системы)».

06/09/2016

Официальный оппонент

Гаврилова Татьяна Альбертовна,

доктор технических наук, профессор,

зав. кафедрой информационных технологий в менеджменте

Института «Высшая школа менеджмента

Санкт-Петербургского государственного университета»

(ВШМ СПбГУ)

Тел. / Fax: +7 812 329 32 34

gavrilova@gsom.spbpu.ru

www.gsom.spbpu.ru/professors/gavrilova

199004, Санкт-Петербург Волховский пер., 3

