



МИНОБРНАУКИ РОССИИ  
федеральное государственное автономное образовательное учреждение  
высшего образования

«Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)»

(СПбГЭТУ «ЛЭТИ»)

ул. Профессора Попова, д.5, Санкт-Петербург, 197376

Телефон: (812) 346-44-87

Факс: (812) 346-27-58

E-mail: [eltech@eltech.ru](mailto:eltech@eltech.ru)

[http:// www eltech.ru](http://www.eltech.ru)

ОКПО 02068539

ОГРН 1027806875381

ОКВЭД 80.3, 73.1

ОКТМО 40392000000

ИНН/КПП 7813045402/781301001

№ 2074/707 №Т 09.09.2016

На № \_\_\_\_\_ от \_\_\_\_\_

отзыв ведущей организации

199178, Россия, Санкт-Петербург,  
14 линия ВО, дом 39

Федеральное государственное бюджетное  
учреждение науки Санкт-Петербургский  
институт информатики и автоматизации  
Российской академии наук (СПИИРАН)

по научной работе

А.В. Жарковский

#### ОТЗЫВ

ведущей организации Федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)» 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

Актуальность направления диссертационного исследования Тушкановой О.Н. определяется следующими факторами:

- проблема обработки больших данных в настоящее время оценивается экспертами, как одна из самых важных проблем информационных технологий;
- существующие методы обработки больших данных зачастую не отвечают ожиданиям и потребностям специалистов в этой области;
- ведущие специалисты в области больших данных указывают, что для задач обработки больших данных необходимо разрабатывать специализированные методы и программные инструменты, использующие новые принципы построения моделей данных;
- в центре внимания разработчиков методов и программных средств обработки больших данных должны быть вопросы вычислительной эффективности.

Фактически методы обработки больших данных формируют новую область научных исследований и разработок в области извлечения знаний из данных. Диссертационная работа Тушкановой О. Н. посвящена теоретическому и экспериментальному исследованию ряда новых актуальных задач в этой области. В частности, в ней исследуются модели и алгоритмы извлечения причинных зависимостей в больших данных гетерогенного характера и их использование для построения семантически интерпретируемых вычислительно эффективных моделей принятия решений в задачах классификации и, в частности, в задачах разработки рекомендательных систем. Тесная связь тематики исследования работы Тушкановой О.Н. с наиболее сложными задачами

семантической обработки больших данных с целью извлечения знаний позволяет считать данную работу актуальной.

### **Степень обоснованности научных положений, выводов и рекомендаций**

Тема диссертационной работы и полученные в ходе исследования результаты соответствуют пунктам 4 и 5 паспорта специальности 05.13.01 - «Системный анализ, управление и обработка информации (технические системы)».

Основные научные положения, приведённые в автореферате и вынесенные на защиту, а также выводы и результаты диссертационного исследования обоснованы и аргументированы. Поставленная в работе научная задача исследована и решена путем корректного использования концепций и подходов из области семантических моделей на основе онтологии, ассоциативного и причинного анализа, статистики и теории вероятностей, методов анализа формальных понятий и кластерного анализа.

### **Оценка новизны и достоверности**

Научной новизной обладают следующие результаты диссертационного исследования:

- теоретически и экспериментально обоснованный выбор меры оценки причинных связей в данных. Выбранная мера предназначена для применения в разработанном соискателем алгоритме поиска причинных связей в данных;
- семантический анализ понятий – разработанный алгоритм автоматического построения семантической модели данных, который использует глобальную онтологию DBpedia, NLP-средство DBpedia Spotlight Service и методы анализа формальных понятий. Алгоритм также отличается тем, что ориентирован на использование в моделях ассоциативно-причинной классификации;
- предложенная семантическая модель данных, которая включает в себя онтологию данных и структуру двойственных формальных понятий, что позволяет описывать синтаксис и семантику данных, а также хранить мета-информацию об этих данных;
- алгоритм поиска причинных структур в обрабатываемых данных. Алгоритм отличается тем, что использует выбранную в работе семантически корректную меру оценки силы причинной связи и предложенную соискателем семантическую модель данных;
- алгоритм минимизации причинных правил, используемых в ходе ассоциативно-причинной классификации. Алгоритм отличается тем, что основан на методах кластеризации и позволяет устранить избыточность модели принятия решений.

Кроме того, соискателем получены более частные результаты в области рекомендательных систем: предложена новая форма представления семантического причинного профиля пользователя в форме дерева решений; предложена новая метрика, позволяющая оценить семантическое сходство интересов пользователей, и алгоритм гибридной коллаборативной фильтрации, использующий профиль пользователя и эту метрику; предложен алгоритм выработки кросс-доменных рекомендаций, основанный на алгоритме гибридной коллаборативной фильтрации.

Степень достоверности научных результатов, полученных в диссертационной работе, подтверждается результатами разностороннего критического анализа состояния исследований в области затрагиваемых проблем, корректным использованием научно-методического аппарата, положительными результатами экспериментальной проверки разработанных алгоритмов, апробацией в печатных работах и докладах на конференциях высокого уровня, положительным опытом использования результатов в работах по коммерческим проектам.

### **Теоретическая и практическая значимость**

Теоретическая значимость диссертационной работы определяется:

- предложенной новой семантической моделью больших данных и алгоритмом её автоматического построения, который назван семантическим анализом понятий;
- обоснованием семантически корректной меры оценки «силы» причинной связи между атрибутами больших данных;

- разработанным масштабируемым алгоритмом поиска причинных зависимостей между атрибутами данных с использованием выбранной меры причинной связи атрибутов данных.

**Практическая значимость диссертационной работы** определяется:

- программной реализацией разработанных алгоритмов и моделей в виде библиотеки переиспользуемых Java-классов, применимых в задачах принятия решений из области рекомендательных систем, а также в других задачах из области анализа и обработки больших данных;

- формулировке практических рекомендаций по применению разработанных алгоритмов в области рекомендательных систем третьего поколения на основе развернутого экспериментального исследования.

#### **Апробация и внедрение полученных результатов**

Результаты диссертационной работы были апробированы на 5 всероссийских и международных научно-практических конференциях. По теме диссертации опубликовано 9 научных работ, из них 3 статьи в печатных изданиях, входящих в перечень ВАК («Информационные технологии и вычислительные системы», «Труды СПИИРАН») и 4 в изданиях, индексируемых в WebofScience и Scopus. Полученные в ходе исследования практические результаты были использованы при выполнении работ по нескольким контрактам и договорам, в том числе, с такими компаниями, как EMCInternationalCompany и Московское подразделение Samsung Electronics – Samsung Research Center, что подтверждено актами внедрения.

#### **Замечания по диссертации**

К диссертационной работе имеется ряд замечаний:

– диссертация включает чрезмерное количество объемных разделов описательного характера, которые не относятся напрямую к разработанным моделям и алгоритмам;

– из текста диссертации не следует, изучалась ли в ходе исследований возможность использования других глобальных онтологий, кроме DBpedia, в алгоритме семантического анализа понятий;

– в разделе 1.4 диссертационной работы упоминаются контекстно-зависимые рекомендательные системы. Далее в тексте работы не поясняется, применимы ли разработанные алгоритмы в подобных рекомендательных системах;

– из текста диссертации непонятно, какая именно мера оценки силы причинных связей (мера Клозгена или коэффициент регрессии) является в итоге наиболее эффективной и точной;

– в работе не приведены результаты сравнения эффективности и/или точности разработанных алгоритмов с аналогами;

– псевдокод алгоритмов содержит чрезмерное количество текстовых комментариев и плохо воспринимается;

– в тексте диссертации присутствуют опечатки.

Указанные замечания носят частный характер и не снижают общей ценности диссертационного исследования и значимости полученных в ходе его выполнения научных результатов.

#### **Заключение**

Диссертация Тушкановой О.Н. представляет собой законченную научно-квалификационную работу на актуальную тему, результаты которой обладают научной новизной, теоретической и практической значимостью.

В работе Тушкановой О.Н. решена важная и актуальная научная задача разработки семантических моделей и алгоритмов ассоциативно-причинной классификации для принятия решений в области обработки больших данных.

Представленная работа характеризуется полнотой изложения рассматриваемых проблем, обладает внутренним единством, содержит аргументированные выводы по каждой главе, свидетельствует о личном вкладе соискателя в соответствующую отрасль науки. Работа отвечает общепринятым критериям научного стиля изложения материала.

Основное содержание работы, а также выводы и результаты диссертационного исследования, достаточно полно отражены в автореферате.

Диссертация Гушкановой О.Н. отвечает требованиям, 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)».

Настоящий отзыв на диссертационную работу рассмотрен и утвержден на заседании кафедры информационных систем федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)», протокол № 10 от 08 сентября 2016 г., присутствовало: 34 чел., результаты голосования: «за» – 34 чел., «против» – нет, воздержалось – нет.

Профессор канд. техн. наук кафедры информационных систем федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)», заведующий кафедрой

Ученый секретарь кафедры информационных систем старший преподаватель



Владислав Владимирович Цехановский



Владимир Петрович Коробкин