

## ОТЗЫВ

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

Работа Тушкановой О.Н. посвящена разработке алгоритмов обучения и принятия решений в задачах классификации на основе семантических и причинных моделей больших данных. Инструменты big data используются в различных секторах современной экономики, объёмы инвестиций в эту область составляют миллиарды долларов. Проблемы обработки больших объемов данных являются наиболее сложными по сравнению с проблемами их передачи, хранения и пр. Для большинства приложений целью обработки больших данных является построение эмпирической модели целевых переменных, учитывающей некоторые атрибуты данных, при этом ключевым требованием является минимизация числа используемых атрибутов при условии обеспечения заданной точности модели. В данной работе задачи такого типа являются основным предметом исследований и разработок. Отметим также, что при них решении наиболее трудными являются проблемы алгоритмического характера, связанные с обеспечением точности, устойчивости и вычислительной эффективности, что также находится в фокусе внимания автора. Сказанное определяет актуальность темы исследования.

Судя по автореферату, к наиболее существенным новым научным результатам, полученным в работе Тушкановой О.Н., относятся следующие:

1. Мера оценки «силы» причиной связи между атрибутами данных, выбор которой основан на обширном экспериментальном исследовании и построен на численных оценках. Семантическая корректность и вычислительная эффективность меры делает ее привлекательной для широкого круга приложений.

2. Масштабируемый и вычислительно эффективный алгоритм поиска множества причинных зависимостей между атрибутами данных, использующий оригинальную семантическую модель данных.

3. Алгоритм автоматической генерации семантической модели, ориентированной на построение и минимизацию множества причинных правил модели. Алгоритм отличается тем, что построен как комбинация методов и средств автоматизированной генерации иерархии понятий онтологии данных и генерации двойственных формальных понятий, определяющих условия останова процесса генерации понятий.

Практическая ценность работы определяется, в частности, тем, что предложенные модели и алгоритмы реализованы в виде программной библиотеки, включающей множество

Java-классов, построенных с учетом возможности повторного использования в широком круге задач обработки больших данных и принятия решений. Разработанные алгоритмы протестированы на данных из области кросс-доменных рекомендательных систем. На основании результатов тестирования формулируются практические рекомендации по выбору алгоритмов и технологий для разработки современных рекомендательных систем.

Результаты работы полно и своевременно опубликованы в 9 печатных работах, в том числе в 7 статьях из перечня ведущих журналов, утвержденного ВАК РФ.

Имеются замечания по содержанию автореферата:

1. Автор использует достаточно специфическую терминологию без определения или хотя бы объяснения этих терминов – например рекомендательные системы «третьего поколения» (с. 4 и ниже).

2. В автореферате упоминается, что работоспособность алгоритмов проверена на примерах из экспериментального набора данных Amazon (с. 12, первый абзац), однако не приведено никаких результатов такого эксперимента.

3. Имеются огрехи в оформлении текста – например, символ в конце предпоследней строки на с. 14.

4. Один из важнейших вопросов фильтрации правил профиля – порог отсечения – описан в очень общих словах (с. 15, первый абзац) и не достаточен для понимания логики автора.

5. Интересно было бы увидеть результаты сравнения работы предложенных алгоритмов с имеющимися на известных тестовых наборах данных, достаточно хорошо представленными в интернете.

В целом, судя по автореферату, диссертационная работа Тушкановой О.Н. соответствует требованиям положения ВАК РФ, а ее автор заслуживает присуждения ей ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (технические системы).

Доцент  
кафедры математической теории интеллектуальных систем  
механико-математического факультета  
МГУ им. М.В. Ломоносова  
д.т.н., профессор

Подпись А.П. Рыжова заверяю  
Начальник отдела кадров

Рыжов А.П.

Ткачева Л.Л.