

ОТЗЫВ

на автореферат диссертации Тушкановой Ольги Николаевны по теме
«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)»

Как это следует из автореферата, диссертационная работа О.Н. Тушкановой посвящена проблеме больших данных (big data). Основным предметом исследований в работе является задача построения эмпирической модели целевых переменных в больших данных с минимизацией числа используемых атрибутов данных. Используется семантико-причинный подход, связанный с поиском причинных связей в структуре понятий больших данных, основанный на ассоциативно-причинном анализе данных, что дает возможность использования ассоциативной классификации при принятии решений. Конкретная цель работы – разработка алгоритмов обучения и принятия решений в задачах классификации на основе семантических и причинных моделей больших данных. Для реализации этой цели использованы современные методы корреляционного, ассоциативного и причинного анализа, методы машинного обучения системы распределенных классификаторов, методы анализа формальных понятий и кластерного анализа. Разработанные алгоритмы реализованы в форме программного прототипа. Проведена экспериментальная оценка их масштабируемости, вычислительной эффективности и точности в задачах принятия решений. Проведенные исследования **актуальны** в плане повышения эффективности обработки больших данных для принятия решений. Нужно отметить, что в литературе до проведения этих исследований отсутствовали вычислительно эффективные алгоритмы поиска ассоциативных правил классификации и построения механизмов принятия решений на основе этих правил.

Из автореферата следует, что в диссертационной работе Тушкановой О.Н. получены **новые научные результаты**, в частности, впервые введена обоснованная и вычислительно эффективная мера оценки силы причинной связи между атрибутами данных; предложена семантическая модель больших данных с определенной структурой для ее представления, а также разработан новый алгоритм для ее автоматической генерации и оригинальные вычислительно эффективные алгоритмы поиска множества причинных связей между атрибутами данных и минимизации мощности множества причинных правил ассоциативной классификации путем устранения избыточных правил.

Анализ краткого содержания работы, приведенного в автореферате, показал, что эти результаты соответствуют основным положениям, выносимым на защиту, и доказывают их корректность.

Теоретическая значимость полученных в диссертации результатов состоит в том, что введенная мера оценки силы причинной связи атрибутов и предложенная семантическая модель больших данных являются теоретически корректными, а разработанные алгоритмы - экспериментально проверенными на задачах обучения и принятия решений в рекомендательных системах. Это позволило по-новому взглянуть на процесс обработки больших данных.

В практическом плане разработанные новые алгоритмы обучения и принятия решений ориентированы на задачи рекомендательных систем третьего поколения и 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 и на основании результатов тестирования даны практические рекомендации по выбору алгоритмов и технологий для

разработки современных рекомендательных систем третьего поколения. Нужно отметить, что, семантико-причинный анализ данных может дать толчок к разработке искусственных когнитивных (познающих) информационно-управляющих систем, которые могут быть близки по эффективности к человеческим. Это, в частности, важно, например, для развития современной робототехники, направленной на создание гуманоидных роботов, не только подобных человеку по форме, но и способных к человекоподобному принятию решений в сложных проблемных областях.

Работа Тушкановой О.Н. прошла **апробацию** на 6-ти международных и российских научно-технических конференциях и ее результаты использованы в 3-х проектах, причем на некоторые результаты имеются акты внедрения. По теме диссертации опубликовано 9 печатных работ, включая 3 публикации в журналах из перечня ВАК, а также 4 публикации в изданиях, индексируемых в базах данных Web of Science и Scopus.

При анализе автореферата возникло следующее **замечание**: в разделе «Краткое содержание работы», глава 4, приведена схема построения семантического профиля пользователя (рисунок 4), где предложено использовать материалы Википедии, в которой, как правило, имеется множество не Достоверной информации, которая может исказить семантический профиль. Кроме того, из материалов автореферата трудно понять, что такое коллаборативная фильтрация, и как она позволяет улучшить результаты по точности классификации.

Вывод. Диссертационная работа О.Н. Тушкановой «Семантические структуры и причинные модели больших данных для принятия решений с приложением к рекомендательным системам» имеет большое теоретическое и практическое значение, поскольку предложенная семантико-причинная модель больших данных, а также предложенная методика и алгоритмы ее автоматического построения на основе семантического анализа данных открывают новый путь обработки информации в больших данных и могут быть полезными при разработке рекомендательных систем третьего поколения, а также когнитивных информационно-управляющих систем для сложных проблемных областей.

В целом автореферат содержит ответы на все положения, выдвинутые на защиту, достаточно полно отражает содержание диссертации и характеризует личный вклад автора в их разработку.

По актуальности, объему выполненных исследований, методическому уровню, научной новизне и практической значимости полученных результатов работа полностью соответствует требованиям п.9 «Положения о порядке присуждения ученых степеней...», утвержденного постановлением Правительства Российской Федерации от 24.09.2013 г. № 842, предъявляемым к кандидатским диссертациям, а ее автор – Тушканова Ольга Николаевна, заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.01 – «Системный анализ, управление и обработка информации (технические системы)».

Станкевич Лев Александрович,

к.т.н., профессор кафедры «Системный анализ и управление»

Института компьютерных наук и технологий

Санкт-Петербургского политехнического университета Петра Великого

2014 г.