

Отзыв официального оппонента

о диссертационной работе

Юрия Савельевича Кашницкого

«Методы замкнутых описаний

в задаче классификации данных со сложной структурой»,

представленной к защите на соискание ученой степени кандидата технических наук по специальности 05.13.17 — «Теоретические основы информатики».

Диссертация Ю. С. Кашницкого посвящена актуальной проблеме интерпретации результатов алгоритмов классификации. Рассматриваются данные со сложной структурой, для которых нет удобного корректного способа векторизации, в контексте аппарата Анализа Формальных Понятий и, в частности, узорных структур и их проекций.

Диссертация состоит из 111 страниц, включающих введение, четыре главы, заключение, список литературы из 107 наименований и одно приложение.

Введение посвящено краткому описанию области изысканий и базовым сведениям о проведённом исследовании: объект и предмет, актуальность и новизна, теоретические и практические достоинства. Также перечислены научные события, на которых автор представлял результаты диссертации.

В первой главе даются понятия теории решёток и математического аппарата Анализа Формальных Понятий (АФП), в контексте которых проделано исследование, описываются методы машинного обучения, основанные на ассоциативных правилах, рассматриваются методы отбора правил. Важная особенность заключается в том, что уделяется внимание возможности интерпретации получаемых правил, на не только качеству классификации. Показывается, как алгоритм построения решётки формальных понятий позволяет произвести отбор правил, гарантирующий классификацию не хуже, чем при использовании дерева решений. В качестве критериев качества используются неопределённость Джини и прирост информации.

Вторая глава посвящена расширению методов АФП за счёт применения узорных структур для случая объектов со сложными описаниями, не задающимися бинарными признаками. Сравниваются подходы к классификации на основе ядерных функций, методом опорных векторов и с использованием узорных структур и их проекций. Рассматриваются проекции интервальных узорных структур и показывается результаты, аналогичный приведённому в первой главе — как для данных с количественными признаками получать правила классификации не хуже, чем результаты дерева решений. В качестве критериев качества снова используются неопределённость Джини и прирост информации.

В третьей главе предлагается алгоритм классификации данных со сложной структурой, для которых можно определить полурешёточную функцию сходства. Такой алгоритм предоставляет общий подход к классификации разнородных данных. Приводятся примеры для данных, описываемых бинарными, количественными и интервальными признаками, а также помеченными графами.

Четвёртая глава посвящена реализации разработанных алгоритмов в виде программного комплекса и экспериментам с наборами данных. Для исследования показателей разработан-

ного алгоритма на разнородных данных используется репозиторий UCI, включающий объекты, описываемые количественными, бинарными и категориальными признаками. Также рассматриваются результаты для задач классификации последовательностей и графов. Качество решения задачи определяется средней долей ответов с применением кросс-валидации. Возможность интерпретации полученных правил оценивается средним размером посылок правил, используемых для классификации. Результаты экспериментов подтверждают состоятельность предложенных методов, определяемую лучшим качеством классификации и лучшей возможностью интерпретации.

Среди других работ, посвящённых классификации данных со сложной структурой в контексте АФП, диссертация Ю. С. Кашницкого обладает следующими особенностями, делающими исследование уникальным:

- Подход, предлагаемый в работе, универсален по отношению к структуре классифицируемых данных: это проверено как на данных с категориальными и количественными признаками, так и на последовательностях и графах. Более того, с некоторыми ограничениями, но он может быть применён к для классификации произвольных структур данных. В то же время, большинство работ по теме ориентируются только на категориальные и количественные признаки, как на самый распространённый формат данных в реальных задачах.
- При сравнении алгоритмов наряду с качеством классификации и быстродействием в работе исследуются возможности интерпретации полученных правил и работы данных с различной структурой. При этом показано, что без потери точности можно использовать простые правила для классификации.

Подходы, предложенные в работе, внедрены в проекте компании «Гемотест» при исследовании прогнозирования концентрации химических веществ в венозной крови, что подтверждает практическую ценность исследования.

Тем не менее, стоит отметить, что работа не лишена ряда недостатков:

- В формулировке цели исследования отождествляются понятия «подход» и «алгоритм», что, вообще говоря, не корректно: один и тот же подход может быть реализован различными алгоритмами.
- В формулировке достоверности полученных результатов используется понятие строгости математических моделей. Однако в работе не приводится указаний на проверку критериев подобной строгости, в том числе применимо к предлагаемым алгоритмам.
- Несколько нестрогим образом вводится ключевое для исследования понятие локальной интерпретируемости: с одной стороны, определение 19 задаёт посылку правила как множество, с другой стороны, в определении 21 используется понятие длины посылки правила, которое остаётся неопределённым.
- Также в работе недостаёт примеров объяснения полученных правил классификации. А это — та неформальная цель, ради которой проводилось исследование.
- Описание отдельных классов и характеристик программного решения стоило вынести в инструкции для разработчиков или приложение, оно не является непосредственной частью проведённого фундаментального исследования.

- Из рассмотренных подходов к классификации графов ядерными методами сравнение проводится только с одним.
- В тексте диссертации довольно распространены неаккуратности оформления. Некоторые из них носят типографический характер: перенос знаков тире на новую строку (например, строка 3 в определении 17 и строка 2 в определении 34), висячие предлоги (например, строка 8 на странице 10, строка 3 на странице 20 и строка 6 на странице 25), висячие строки (например, последняя строка на странице 13 и строка 14 на странице 45), отрыв номера от соответствующего понятия (например, строки 5 и 10 на странице 11), кавычки, отличные от «ёлочек», принятых для публикаций на русском языке (например, строка 2 на странице 25 и третья снизу строка на странице 26) и т. п. Другие могут сказаться на корректности восприятия материалов работы: например, разрыв тройки формального контекста в определении 20 или многочисленные разрывы других формул без корректного оформления переноса.

Указанные замечания не носят принципиального характера и не снижают общей положительной характеристики диссертационной работы.

Результаты рецензируемого научного исследования соответствуют паспорту специальности 05.13.17 — «Теоретические основы информатики» в пунктах 5 — «Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечение, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений» и 7 — «Разработка методов распознавания образов, фильтрации, распознавания и синтеза изображений, решающих правил. Моделирование формирования эмпирического знания».

Диссертационная работа Ю. С. Кашницкого «Методы замкнутых описаний в задаче классификации данных со сложной структурой» содержит важные результаты, обладающие научной новизной, теоретической значимостью и практической ценностью и отвечает требованиям пунктов 9–14 Положения о защите ученых степеней, утвержденного Постановлением Правительства Российской Федерации от 24.09.2013 г. №842 и требованиям ВАК, Минобрнауки РФ, предъявляемым к диссертациям на соискание ученой степени кандидата технических наук по специальности 05.13.17 — «Теоретические основы информатики», а её автор, Ю. С. Кашницкий, заслуживает присвоения ему ученой степени кандидата технических наук по указанной специальности.

Кандидат технических наук, доцент кафедр
математических основ управления
и высшей математики
Московского физико-технического института

К. В. Чувилин
24.09.2018

ЗАВЕРЯЮ
УЧЕНЫЙ СЕКРЕТАРЬ
МФТИ
Ю. И. СКАЛЬКО

