

ОТЗЫВ

официального оппонента кандидата технических наук Игнатова Дмитрия Игоревича на диссертацию Прокофьева Петра Александровича «Корректное распознавание по прецедентам: построение логических корректоров общего вида и вычислительные аспекты», представленную на соискание ученой степени кандидата физико-математических наук по специальности 05.13.17 – «Теоретические основы информатики».

Диссертация Прокофьева П.А. посвящена исследованию алгебро-логических подходов к классификации объектов, описываемых (или сводящихся к ним) булевыми векторами признаков, на основе так называемых логических корректоров.

В настоящее время в сообществе машинного обучения довольно сильно подогрет интерес к таким подходам как глубинное (глубокое) обучение (deep learning) в связи с успехами применения технологий на основе нейронных сетей в ряде задач, например, распознавании образов в изображениях и видеоданных. Однако фундаментальные вопросы, связанные с возможностью построения корректных алгоритмов классификации, остаются открытыми. В наиболее простой базовой постановке задачу классификации объектов (прецедентов) можно решать для случая их описания булевыми векторами признаков. Случай с вещественными признаками и категориальными естественно сводится путем шкалирования к булеву. В классических работах по обучению понятиям на основе гипотез (элементарных классификаторов) рассматриваются конъюнкции признаков, например, в работах Т. Митчелла (Т. Mitchell) по пространствам версий; другой пример – ДСМ-гипотезы в работах Финна В.К., Кузнецова С.О. и др. В данном диссертационном исследовании автор отходит от довольно жестких классификационных схем (классификация только на основе элементарных классификаторов в виде конъюнкций, запрет на контр-пример и др.) и рассматривает (в теории произвольные) особые булевы функции над ними, называемые логическими корректорами. Автору не только удастся построить логические корректоры, превосходящие по качеству классификации ранее известные экспериментально (при определенных условиях), но и предложить более эффективные алгоритмические решения комбинаторных задач (дуализация как задача поиска покрытия особого вида булевой матрицы), связанных с построением таких корректоров.

Учитывая вышесказанное, актуальность диссертационной работы

Прокофьева П.А., направленной на построение эффективных логических корректоров, поиск эффективных способов их построения, а также расширение асимптотически оптимальных семейств алгоритмов дуализации, сомнений не вызывает.

Научная новизна работы Прокофьева П.А. определяется тем, что впервые предложена модель логического корректора POLAR на основе т.н. поляризуемой булевой функции в качестве корректора. Предложена новая методика повышения эффективности (в терминах как времени вычислений, так и качества классификации) логических корректоров. Построены новые асимптотически оптимальные алгоритмы дуализации.

Теоретическая значимость диссертационной работы связана, прежде всего с успешной попыткой рассмотрения ранее существующих логических подходов к классификации с позиций общей схемы синтеза логических корректоров. **Практическая значимость** подтверждается экспериментальными результатами, в частности, для случая данных, сводимых к булевым с большим числом признаков, наблюдается превосходство над классическими методами классификации. Помимо этого, при решении задачи дуализации, благодаря изменению схемы перебора, удалось превзойти алгоритмы японских коллег, считающиеся лучшими для решения задачи дуализации.

Обоснованность научных положений и выводов диссертации обеспечена корректным доказательством математических утверждений, а также убедительными и значимыми результатами, полученными на реальных и модельных данных.

Результаты диссертационного исследования Прокофьева П.А. могут быть использованы в ФИЦ ВЦ РАН, ИПУ РАН, ИППИ РАН, ИММ УрО РАН, ВМиК МГУ, МФТИ, ФКН НИУ ВШЭ и других ведущих научных учреждениях Российской Федерации и коллективах, где проводятся исследования требующие создания эффективных методов классификации.

Диссертация состоит из введения, 3 глав, заключения и списка литературы.

В первой главе дан краткий экскурс в теорию корректного распознавания по прецедентам и описан логический корректор POLAR.

Во второй главе автором рассматриваются способы повышения эффективности логических корректоров на основе бустинга и т.н. локальных базисов классов. Экспериментально исследуется POLAR.

Третья глава посвящена новым алгоритмам дуализации двух типов.

В заключении автор резюмирует основные результаты диссертационной работы.

Основные результаты диссертации обоснованы с помощью доказательств соответствующих утверждений и подкреплены эмпирически. Содержание диссертации в достаточной мере покрывается работами автора в научной печати, среди которых имеются публикации в пяти журналах из списка ВАК. Результаты докладывались автором на российских и международных научных конференциях и семинарах, автореферат диссертации в полной мере соответствует ее содержанию.

Тем не менее, стоит привести некоторые замечания и высказать пожелания по дальнейшим исследованиям.

1. Несмотря на то, что работа в целом написана достаточно аккуратно, имеется несколько неточностей и опечаток. Например, на странице 15 алгоритм $A^0_T(S)$ некорректно работает в случае объектов с одинаковыми описаниями; на странице 62 квантор всеобщности применен после использования индекса, который он связывает и т.п.
2. В работе остались без рассмотрения упомянутые выше пространства версий и ДСМ-метод на основе синтеза логических гипотез в виде конъюнкций признаков как логические методы классификации. По-видимому, данные методы также возможно вписать в общую схему работы с логическими корректорами, предлагаемую в этой работе.

Имеется пожелание. В работе особо подчеркивается, что рассматриваемые методы работают с бинарными матрицами объекты-признаки. В алгебраическом анализе данных сходство и различие объектов, описанных бинарными признаками, хорошо изучены в рамках анализа формальных понятий (Formal Concept Analysis) и неточных множеств (Rough Sets), соответственно. Например, в рамках первой дисциплины возможно находить максимальные подматрицы объектов, обладающих заданным подмножеством признаков (и наоборот), а в рамках второй – т.н. редукты, которые описывают попарно-различные объекты в терминах подвыборки признаков наименьшего размера с возможностью различения классов объектов. Возможно, что такие алгебро-комбинаторные объекты могут быть полезны как при решении задачи дуализации (покрытие максимальными подматрицами из единиц (т.е. понятиями, факторами) исходной матрицы), так и при непосредственном порождении элементарных классификаторов (можно брать только замкнутые множества признаков) или при построении ансамблей (логических корректоров) на основе редуктов.

Приведенные замечания не снижают общей высокой оценки диссертационной работы, ее теоретической и практической значимости. Работа полностью соответствует требованиям ВАК РФ, предъявляемым к диссертациям на соискание учёной степени кандидата физико-математических наук по специальности 05.13.17 – «Теоретические основы информатики», а её автор, Прокофьев П.А., заслуживает присуждения ему учёной степени кандидата физико-математических наук по данной специальности.

Официальный оппонент

Доцент департамента анализа данных и
искусственного интеллекта

Федерального государственного автономного
образовательного учреждения высшего
профессионального образования
Национальный исследовательский
университет «Высшая школа экономики»

Д. И. Игнатов

18 сентября 2016 г.

Адрес: 125319, Москва, Кочновский проезд, д. 3
Тел.: +7(495) 772-9590*22930
e-mail: dignatov@hse.ru

Подпись заверяю

