

ОТЗЫВ

**официального оппонента к.ф.-м.н. Г.О. Маслякова
на диссертацию Ишкиной Шауры Хабировны
«Комбинаторные оценки переобучения пороговых решающих правил»,
представленную на соискание ученой степени
кандидата физико-математических наук по специальности
1.2.1 - «Искусственный интеллект и машинное обучение»**

Актуальность темы диссертации. Рассматриваемая проблема построения оценок обобщающей способности пороговых решающих правил занимает важное место в современном машинном обучении. Обобщающая способность модели показывает ее возможности по экстраполяции информации, полученной на основе обучающей выборки данных, на новые случаи, встречающиеся в реальных условиях эксплуатации системы. Когда модель становится слишком сложной и начинает запоминать детали конкретного набора данных, теряя способность правильно обрабатывать новую информацию, ее точность на новых данных становится намного ниже, чем на обучающей выборке, и говорят, что произошло переобучение. Оценки обобщающей способности необходимы для предотвращения переобучения, например, с помощью регуляризации или увеличения объема обучающей выборки, выбора оптимальной архитектуры модели и повышения надежности ее решений. Таким образом, в связи с ростом объемов данных и усложнением моделей, тема диссертации соответствует актуальным научным направлениям.

Основные результаты и их научная новизна. В диссертационной работе получены новые оценки обобщающей способности. В частности, определен класс методов обучения, называемых финитными, и для произвольного семейства классификаторов и финитных методов обучения доказаны теоремы о представлении оценок обобщающей способности в виде произведения числа разбиений двух непересекающихся множеств объектов генеральной совокупности. Доказано, что методы минимизации эмпирического риска и максимизации переобученности являются финитными. Для двух данных методов обучения и семейства одномерных пороговых решающих правил доказаны теоремы о вычислении точных (достижимых) оценок обобщающей способности. Автором разработан оригинальный метод вычисления оценок на основе рекуррентного подсчета числа траекторий при блуждании по трехмерной сетке между двумя заданными точками с ограничениями специального вида.

Получен новый алгоритм построения дерева решений. В этом алгоритме в качестве критерия ветвления используются разработанные достижимые верхние оценки обобщающей способности одномерных пороговых решающих правил.

Построена новая суррогатная модель, позволяющая аппроксимировать с высокой точностью оценки обобщающей способности для рассмотренных семейств одномерных пороговых решающих правил и сократить сложность вычислений.

Теоретическая значимость работы обосновывается тем, что ранее в рамках комбинаторного подхода рассматривались семейства одномерных пороговых решающих правил в частном случае, когда значения признака были попарно различны на генеральном множестве объектов. В диссертационной работе рассмотрен общий случай без указанного ограничения, и для данного семейства пороговых классификаторов получены достигаемые верхние оценки обобщающей способности. Кроме того, стоит отметить, что разработанный метод вычисления оценок на основе случайных блужданий по трехмерной сетке позволяет рассмотреть широкий набор функционалов обобщающей способности: вероятность переобучения, полный скользящий контроль и ожидаемую переобученность. Доказанная теорема о представлении данных функционалов в виде произведения числа разбиений двух множеств объектов генеральной совокупности рассматривает также достаточно широкий класс финитных методов обучения и произвольное семейство классификаторов. Таким образом, доказанные теоремы и разработанные методы предполагают возможность продолжения исследований по данной теме, рассмотрение новых методов обучения и функционалов обобщающей способности. Проведенный анализ значимости признаков в построенной суррогатной модели выделяет набор характеристик семейства пороговых классификаторов, которые могут быть использованы при построении новых аналитических оценок переобучения.

Практическая значимость работы обосновывается тем, что разработанные комбинаторные оценки применимы в качестве критерия ветвления в решающих деревьях, при отборе признаков в логических закономерностях, при построении линейных классификаторов методом покоординатной оптимизации, подборе порога в дискриминантном классификаторе. Предложенный в работе способ выбора пар скважин для включения в программу трассерных (маркерных) исследований применим для оптимизации процесса трассерных исследований на месторождениях нефти и газа.

Достоверность и обоснованность результатов диссертационного исследования подтверждена строгими математическими доказательствами и проверкой на практике путём применения разработанных методов в прикладной задаче составления программы трассерных (маркерных) исследований в месторождениях нефти и газа. Основные результаты диссертационного исследования опубликованы в 16 работах, из которых 7 — в рецензируемых научных изданиях из перечня ВАК и приравненных к ним международных наукометрических баз. Отдельно следует подчеркнуть получение патента на изобретение и наличие акта внедрения основных результатов, подтверждающих успешную апробацию разработок при решении практических производственных задач. Основные результаты диссертации докладывались на восьми конференциях в 2013—2023 гг.

Замечания. Имеются следующие замечания по тексту диссертации.

Согласно проведенному обзору, рассматриваемые семейства одномерных пороговых правил возникают в решающих деревьях, логических закономерностях, алгоритмах вычисления оценок, а также при построении линейных классификаторов методом покоординатной оптимизации. При этом в диссертации рассматривается только одно

приложение предложенных оценок в качестве критерия ветвления в решающих деревьях. Хорошо было бы провести более широкое исследование, насколько комбинаторная оценка полезна в каждом из перечисленных в обзоре алгоритмов.

В главе 3 не проведено сравнение с алгоритмом QUEST построения дерева решений, который, как известно, использует статистические тесты для выбора атрибута для разделения узла на подузлы и может быть более устойчивым к изменению набора примеров, попавших в узел, чем рассмотренные критерии (индекс Джини и коэффициент прироста информации).

Указанные замечания не являются существенными и не влияют на общее положительное впечатление от работы.

Заключение. Диссертационная работа Ишкиной Шауры Хабировны представляет собой законченное научное исследование, выполненное на высоком уровне, оформленное в виде научно-квалификационной работы. Диссертация соответствует паспорту специальности 1.2.1 — «Искусственный интеллект и машинное обучение», а именно пунктам 1 «Естественно-научные основы и методы искусственного интеллекта», 2 «Исследования в области оценки качества и эффективности алгоритмических и программных решений для систем искусственного интеллекта и машинного обучения. Методики сравнения и выбора алгоритмических и программных решений при многих критериях» и 12 «Исследования в области «доверенных» систем класса ИИ, включая проблемы формирования тестовых выборок прецедентов, надежности, устойчивости, переобучения и т.д.».

В работе представлен ряд новых научных результатов, имеющих как фундаментальное, так и прикладное значение.

Автореферат в полной мере отражает содержание диссертации и дает представление о полученных автором результатах.

Основываясь на изложенных выше выводах об актуальности, научной новизне, теоретической и практической значимости работы, апробации результатов можно заключить, что диссертационная работа Ишкиной Шауры Хабировны «Комбинаторные оценки переобучения пороговых решающих правил» соответствует требованиям и критериям Положения ВАК РФ «О порядке присуждения ученых степеней», утвержденного постановлением Правительства РФ от 24.09.2013 № 842, предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук, а её автор, Ишкина Шаура Хабировна, заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 1.2.1 — «Искусственный интеллект и машинное обучение».

Официальный оппонент, к.ф.-м.н.,
старший разработчик
программного обеспечения
ООО «Яндекс.Технологии»

Глеб Олегович Масляков

Эл.почта: gleb-mas@mail.ru

Подпись *Маслякова Г.О.*
ЗАВЕРЯЮ
ПРЕДСТАВИТЕЛЬ
ПО ДОВЕРЕННОСТИ
ОТ 01.10.2023
ГОЛОВКОВА Т.А.

