



УТВЕРЖДАЮ  
Директор ИПУ РАН  
академик РАН  
Д. А. Новиков  
«2» марта 2026 г.

## ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

Федерального государственного бюджетного учреждения науки  
Институт проблем управления им. В. А. Трапезникова Российской академии наук  
на диссертацию Ишкиной Шауры Хабиловны  
«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

Темой диссертации является построение оценок обобщающей способности одномерных пороговых классификаторов. Проблема переобучения, возникающая вследствие разницы качества классификатора на обучающей и тестовой выборках, представляет серьезную трудность в теории машинного обучения. Надёжная оценка обобщающей способности позволяет строить более устойчивые и точные классификаторы, обладающие хорошей способностью экстраполировать обнаруженные по обучающей выборке закономерности на неизвестные объекты. Задача построения оценок обобщающей способности семейства классификаторов на основе информации об обучающей выборке и структуре семейства остается одной из основных задач теории статистического обучения более 50 лет, начиная с работ В.Н. Вапника и А.Я. Червоненкиса. Эксперименты показывают, что оценки Вапника-Червоненкиса завышены на 6–12 порядков, вследствие чего происходит некорректный подбор структурных параметров семейства классификаторов с целью контроля переобучения. Несмотря на наличие ряда существующих подходов, многие из них оказываются излишне консервативными, приводя к завышенным оценкам обобщающей способности.

Комбинаторная теория переобучения, предложенная К.В. Воронцовым, представляет собой альтернативный подход к решению проблемы переобучения в машинном обучении и обладает рядом преимуществ перед традиционными методами анализа сложности моделей. Комбинаторная теория предлагает строгий математический аппарат для вычисления вероятности случайного события на основе предположения о конечности множества объектов и классификаторов. Также в комбинаторной теории учитывается внутренняя структура семейства классификаторами и связи между объектами. Это позволяет более точно анализировать зависимость между сложностью модели и ее обобщающей способностью. В рамках комбинаторного подхода были получены оценки переобучения для ряда модельных семейств, однако, предположения о структуре данных семейств практически не выполнимо в реальных задачах, что ограничивает применимость разработанных оценок. В диссертационной работе Ш.Х. Ишкиной рассмотрено семейство одномерных пороговых классификаторов, которое используется в качестве элементарных строительных блоков в алгоритмах классификации, таких как, например, деревья решений и логические закономерности, и представляет теоретический и практический интерес для контроля переобучения данных алгоритмов.

## Структура и основное содержание диссертации

Диссертация состоит из введения, четырех глав, заключения, списка иллюстраций, списка таблиц, списка литературы и приложения.

В первой главе приводятся основные определения комбинаторной теории переобучения, методов обучения. Задаются функционалы для характеристики обобщающей способности семейства классификаторов: вероятность переобучения, полный скользящий контроль и ожидаемая переобученность. Рассматриваются различные методы обучения, которые по заданной обучающей выборке по определенному правилу выбирают классификатор из семейства: методы минимизации эмпирического риска, пессимистичной минимизации эмпирического риска и максимизации переобученности. Вводится понятие финитного метода обучения и доказывается теорема о представлении рассматриваемых функционалов обобщающей способности для финитного метода обучения в виде произведения числа разбиений двух непересекающихся подмножеств полного множества объектов, называемого генеральной совокупностью, которое в предположении комбинаторной теории переобучения конечно. Данная теорема справедлива для произвольного семейства классификаторов, которое также конечно в предположении комбинаторной теории конечно. Доказывается, что свойством финитности обладают рассматриваемые методы обучения. Для основного объекта исследования — семейства одномерных пороговых классификаторов — доказывается теорема и приводится псевдокод алгоритма вычисления точных (достижимых) верхних оценок функционалов обобщающей способности. Доказывается его полиномиальная вычислительная сложность.

Во второй главе проводится сравнение известных оценок обобщающей способности для пороговых решающих правил, полученных ранее в других подходах (в классической теории Вапника—Червоненкиса, комбинаторной теории Воронцова и подходах на основе Радемахеровской сложности) с достижимыми верхними оценками, рассчитанными с помощью алгоритма, описанного в первой главе. Показано, что оценки Гуза, полученные в рамках комбинаторной теории переобучения, обладают высокой точностью и быстро вычислимы, откуда следует вывод о применимости данных оценок в прикладных задачах в частных случаях.

В третьей главе полученные оценки переобучения применяются в прикладной задаче построения программы трассерных (маркерных) исследований в месторождениях нефти и газа. Описан алгоритм построения программы исследования, основанный на классификации пар скважин нагнетательная—добывающая: если классификатор выдает единицу, пара скважин включается в программу. Цель такого алгоритма заключается в сокращении пар скважин в программе, между которыми нет гидродинамической связи, поскольку в этом случае невозможно получить количественные оценки фильтрационных параметров пласта. Классификатором является дерево решений, которое, как известно, склонно к переобучению. Для повышения обобщающей способности дерева решений применяются комбинаторные оценки переобучения пороговых решающих правил в качестве критерия ветвления. Проводятся вычислительные эксперименты на промысловых данных, результаты которых показывают, что данный подход приводит к повышению точности и снижению переобученности итогового классификатора.

В четвертой главе решается проблема больших вычислительных затрат при использовании алгоритма, разработанного в первой главе. Для этого строится суррогатная модель, которая с большой точностью (средняя относительная ошибка MARE равна 2.8%) вычисляет приближенные оценки переобучения, при этом снижая вычислительную сложность с  $O(L^5)$  до  $O(L^2)$ , где  $L$  — общее количество объектов генеральной совокупности.

В заключении приведен основные результаты диссертации. Список литературы включает 83 публикации.

В приложении приведен акт внедрения основных результатов диссертационной работы в производственные процессы ООО «РН-БашНИПИнефть».

### **Основные результаты и их новизна**

В диссертационной работе Ш.Х. Ишкиной получены следующие новые результаты:

1. Доказаны теоремы о представлении достигаемых верхних оценок обобщающей способности произвольного семейства классификаторов в виде произведения числа разбиений двух непересекающихся множеств объектов генеральной совокупности для финитного метода обучения.

2. Доказаны теоремы и разработан алгоритм полиномиальной сложности для вычисления достигаемых верхних оценок обобщающей способности прямых последовательностей классификаторов, порождаемых одномерными пороговыми решающими правилами при варьировании параметра порога, для финитного метода обучения.

3. Разработан алгоритм для построения программы трассерных исследований с применением деревьев решений.

4. Разработан алгоритм построения дерева решений с использованием полученных достигаемых верхних оценок полного скользящего контроля и ожидаемой переобученности в качестве критерия выбора атрибута в узле.

5. Разработан алгоритм вычисления приближенных оценок обобщающей способности одномерных пороговых решающих правил с использованием суррогатных моделей.

### **Достоверность полученных результатов**

Научные положения и выводы обоснованы строгими математическими доказательствами. Проведена экспериментальная проверка разработанных алгоритмов на верифицированных промышленных данных трассерных (маркерных) исследований на двух месторождениях Западной Сибири. Эксперименты сопровождаются подробным описанием методики, что допускает их воспроизводимость. Выводы по итогам экспериментов обоснованы корректной проверкой статистических гипотез.

Основные положения, выносимые на защиту, получили апробацию на международных и российских конференциях и опубликованы в рецензируемых научных изданиях из перечня ВАК и приравненных к ним международных наукометрических баз (Scopus/Web of Science).

### **Значимость полученных результатов**

Теоретическая значимость диссертационной работы заключается в том, что сформулированные и доказанные теоремы расширяют класс задач и семейства отображений, для которых можно эффективно вычислять достигаемые комбинаторные оценки обобщающей способности. Кроме того, представления для функционалов обобщающей способности в виде произведения числа разбиений множеств объектов генеральной совокупности являются достаточно общими результатами, применимыми при построении оценок обобщающей способности для новых семейств классификаторов.

Практическую ценность представляют предложенный полиномиальный алгоритм вычисления достигаемых оценок и суррогатная модель для вычисления приближенных оценок обобщающей способности, которые могут быть применены при построении алгоритмов классификации, где возникает задача выбора оптимального порога, например, при принятии решения на основе значения вероятности принадлежности к классу в логистической регрессии, методе опорных векторов, нейронных сетях; при отборе

признаков в деревьях решений или в методе покоординатной оптимизации при построении алгоритмов классификации для контроля переобучения.

Практическая значимость полученных результатов подтверждена регистрацией патента на изобретение и актом внедрения основных результатов в производственные процессы ООО «РН-БашНИПИнефть».

### **Рекомендации по использованию результатов**

Результаты диссертационной работы Ишкиной Ш.Х. могут быть использованы:

- В системах защиты информации и информационной безопасности для настройки механизмов обнаружения атак и мошенничества с целью предотвращения ложных срабатываний и пропуска важных угроз.

- В финансовых технологиях и страховании для разработки надежных скоринговых моделей кредитования, алгоритмов оценки рисков наступления страховых случаев с целью снижения финансовых потерь вследствие неверных прогнозов.

- В медицинских исследованиях при диагностике сложных заболеваний путем тщательного подбора критериев деления пациентов на группы риска для увеличения чувствительности и специфичности диагностики.

- В разработке автономных транспортных систем и беспилотников для снижения частоты аварийных ситуаций благодаря лучшей адаптации системы к разнообразным условиям окружающей среды и поведения участников движения.

- В области промышленного производства при построении эффективных систем автоматического контроля дефектов продукции с целью уменьшения доли брака и улучшения экономических показателей предприятий.

- В научных исследованиях и образовательных программах: методология оценки обобщающей способности может стать частью учебных курсов по машинному обучению, позволяя студентам глубже понимать принципы работы классификаторов и анализировать реальные практические случаи.

### **Замечания**

1. Одномерные пороговые классификаторы имеют узкую применимость в современных задачах с большим количеством признаков. Интереснее было бы получить оценки переобучения многомерного порогового классификатора и сравнить их, например, с методом покоординатной оптимизации. Интересен вопрос, достаточно ли проводить оптимизацию на каждом признаке последовательно для контроля переобученности итогового алгоритма или необходимо анализировать признаки в совокупности.

2. Приведенный пример прикладной задачи (около 10 признаков и 300 объектов в выборке) выглядит не сопоставимым с современными масштабами задач. Интересно было бы рассмотреть более реальную задачу с большим количеством признаков и провести сравнение с широко распространенными алгоритмам, такими, как ансамбли деревьев или нейронные сети.

3. Выбор решающих деревьев в рассмотренной прикладной задаче недостаточно обоснован, поскольку автором не рассмотрены более сильные модели машинного обучения и современные методы интерпретации, например, построение вектора Шепли. Не проведено масштабное исследование моделей машинного обучения для данной задачи.

Отмеченные замечания носят преимущественно рекомендательный характер и не влияют на общую положительную оценку диссертации.

### **Заключительная оценка**

Диссертация Ишкиной Шауры Хабиловны на тему «Комбинаторные оценки переобучения пороговых решающих правил» является научно-квалификационной работой, выполненной под руководством д.ф.-м.н., профессора РАН Воронцова Константина Вячеславовича.

Представленные в работе результаты и выводы обоснованы корректно, обладают новизной, достоверностью и значимостью для развития соответствующей области знаний. Результаты работы своевременно и полно опубликованы в 16 печатных трудах, из которых 8 — в рецензируемых научных изданиях, определяемых в соответствии с требованиями пп. 11—13 «Положения о присуждении ученых степеней». Результаты докладывались автором на российских и международных конференциях. Автореферат верно и полно отражает содержание диссертационной работы.

Диссертация соответствует всем критериям, установленным «Положением о присуждении ученых степеней», предъявляемым к диссертациям на соискание ученой степени кандидата наук по специальности 1.2.1. «Искусственный интеллект и машинное обучение», а ее автор — Ишкина Шаура Хабиловна — заслуживает присуждения ученой степени кандидата физико-математических наук по указанной специальности.

Диссертация и автореферат заслушаны и обсуждены на расширенном заседании семинара Федерального государственного бюджетного учреждения науки Институт проблем управления им. В.А. Трапезникова Российской академии наук, протокол № 1 от 16 февраля 2026 года.

Председатель семинара,  
доктор технических наук,  
ведущий научный сотрудник

Д.А. Губанов

#### Сведения о составителе отзыва

Фамилия, имя, отчество: Губанов Дмитрий Алексеевич  
Ученая степень: доктор технических наук

Наименование отрасли наук, научных специальностей, по которым защищена диссертация: 05.13.10 — «Управление в социальных и экономических системах (технические науки)»

Должность: Заведующий лабораторией 25, ведущий научный сотрудник  
Контактный телефон: 7 495 198-17-20 доб. 1402  
Адрес электронной почты: gubanov@ipu.ru

#### Сведения о ведущей организации

Федеральное государственное бюджетное учреждение науки «Институт проблем управления им. В.А. Трапезникова Российской академии наук» (ИПУ РАН)

Адрес: 117997, ГСП-7, г. Москва, Варшавское шоссе, д. 45

Сайт: [www.ipu.ru](http://www.ipu.ru)

Телефон: +7 (495) 334-89-10

E-mail: [dan@ipu.ru](mailto:dan@ipu.ru)

