

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА Д002.017.02 НА БАЗЕ
ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО УЧРЕЖДЕНИЯ
НАУКИ ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР ИМ. А.А. ДОРОДНИЦЫНА
РОССИЙСКОЙ АКАДЕМИИ НАУК
ПО ДИССЕРТАЦИИ НА СОИСКАНИЕ УЧЁНОЙ СТЕПЕНИ ДОКТОРА НАУК

аттестационное дело № _____

решение диссертационного совета от 13.11.2014, № 12

О присуждении Стрижову Вадиму Викторовичу, гражданину РФ, учёной степени доктора физико-математических наук. Диссертация «Порождение и выбор моделей в задачах регрессии и классификации» по специальности 05.13.17 – «Теоретические основы информатики» принята к защите 19 июня 2014 г., Протокол № 4 диссертационным советом Д002.017.02 на базе Федерального государственного бюджетного учреждения науки вычислительный центр им. А.А. Дородницына Российской академии наук, 119333, Москва, ул. Вавилова, 40, приказ 192/НК от 09.04.2013 г.

Соискатель Стрижов Вадим Викторович 1967 года рождения. Диссертацию на соискание учёной степени кандидата физико-математических наук «Согласование экспертных оценок при построении интегральных индикаторов» защитил в 2002 году, в диссертационном совете, созданном на базе Вычислительного центра им. А.А. Дородницына РАН, работает научным сотрудником ФГБУН Вычислительного центра им. А.А. Дородницына РАН. Диссертация выполнена в ФГБУН Вычислительный центр им. А.А. Дородницына РАН в Отделе интеллектуальных систем.

Официальные оппоненты:

1. Двоенко Сергей Данилович, доктор физико-математических наук, доцент, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования Тульский государственный университет, профессор Кафедры автоматике и телемеханики;

2. Сметанин Юрий Геннадиевич, доктор физико-математических наук, Российский фонд фундаментальных исследований, начальник Отдела инфокоммуникационных технологий и вычислительных систем;

3. Хачай Михаил Юрьевич, доктор физико-математических наук, старший научный сотрудник, Федеральное государственное бюджетное учреждение науки Институт математики и механики им. Н. Н. Красовского УрО Российской академии наук, заведующий Отделом математического программирования; –

дали положительные отзывы о диссертации.

Ведущая организация Федеральное государственное бюджетное учреждение науки Институт системного анализа Российской академии наук (ФГБУН ИСА РАН), г. Москва, в своем положительном заключении, подписанном Геннадием Семеновичем Осиповым, доктором физико-математических наук, профессором, заместителем директора ФГБУН ИСА РАН по научной работе, указала, что

«Диссертация соответствует критериям, установленным настоящим Положением о присуждении ученых степеней: совокупность результатов, содержащихся в ней, следует квалифицировать как научное достижение, имеющее существенное значение для развития теоретических основ информатики, она написана автором самостоятельно, обладает внутренним единством и содержит новые научные результаты, полученные автором лично. В диссертации приведены рекомендации по использованию научных выводов. Автореферат диссертации адекватно отражает содержание работы. Работа полностью соответствует требованиям ВАК МОН РФ, предъявляемым к диссертациям на соискание ученой степени доктора физико-математических наук по специальности 05.13.17 — теоретические основы информатики, а ее автор, В. В. Стрижов, заслуживает присуждения ему ученой степени доктора физико-математических наук по данной специальности».

Соискатель имеет 98 опубликованных работ, в том числе по теме диссертации 45 работ, опубликованных в рецензируемых научных изданиях, 28. Общий объем опубликованных по теме диссертации работ составляет 33 печатных листа. 5 статей опубликовано в изданиях, выходящих на английском языке. Соискателем опубликована 21 работа по теме диссертации в материалах всероссийских и меж-

дународных конференций и симпозиумов. Все вынесенные на защиту результаты получены автором самостоятельно и опубликованы.

Наиболее значимые научные работы по теме диссертации из журналов, рекомендованных ВАК:

1. Kuznetsov M.P., Strijov V.V. Methods of expert estimations concordance for integral quality estimation // Methods of expert estimations concordance for integral quality estimation// // Expert Systems with Applications, 2014. Vo. 41(4)/ Pp. 1988-1996.
2. Motrenko A.P., Strijov V.V., Weber G.-W. Bayesian sample size estimation for logistic regression // Journal of Computational and Applied Mathematics, 2014. Vol. 255. Pp. 743-752.
3. Strijov V., Krymova E.A., Weber G.W. Evidence optimization for consequently generated models // Mathematical and Computer Modelling, 2013. Vol. 57(1-2). Pp. 50-56.
4. Strijov V., Granic G. Integral Indicator of Ecological Footprint for Croatian Power Plants // Energy, 2011. Vol. 36(7). Pp. 4144-4149.
5. Strijov V., Weber G.-W. Nonlinear regression model generation using hyperparameter optimization // Computers and Mathematics with Applications, 2010. Vol. 60(4). Pp. 981-988.
6. Цыганова С.В., Стрижов В.В. Построение иерархических тематических моделей коллекции документов // Прикладная информатика, 2013. N 1. С. 109-115.
7. Стрижов В.В. Функция ошибки в задачах восстановления регрессии // Заводская лаборатория, 2013, Т. 79(5). С. 65-73.
8. Медведникова М.М., Стрижов В.В. Построение интегрального индикатора качества научных публикаций методами ко-кластеризации // Известия Тульского государственного университета, Естественные науки, 2013. No 1. С. 154-165.
9. Адуенко А.А., Стрижов В.В. Алгоритм оптимального расположения названий коллекции документов // Программная инженерия, 2013, No 3. С. 21-25.
10. Будников Е.А., Стрижов В.В. Оценивание вероятностей появления строк в коллекции документов // Информационные технологии, 2013. No 4 С. 40-45.
11. Зайцев А.А., Стрижов В.В., Токмакова А.А. Оценка гиперпараметров регрес-

- сионных моделей методом максимального правдоподобия // Информационные технологии, 2013. No 2. С. 11-15.
12. Иванова А.В., Адуенко А.А., Стрижов В.В. Алгоритм построения логических правил при разметке текстов // Программная инженерия, 2013. No 6. С. 41-48.
 13. Кузьмин А.А., Стрижов В.В. Проверка адекватности тематических моделей коллекции документов // Программная инженерия, 2013. No 4. С. 16-20.
 14. Медведникова М.М., Стрижов В.В. Построение интегрального индикатора качества научных публикаций методами ко-кластеризации // Известия Тульского государственного университета, Естественные науки, 2013. No 1. С. 154-165.
 15. Рудой Г.И., Стрижов В.В. Алгоритмы индуктивного порождения суперпозиций для аппроксимации измеряемых данных // Информатика и её применения, 2013. Т. 7(1). С. 17-26.
 16. Адуенко А.А., Кузьмин А.А., Стрижов В.В. Выбор признаков и оптимизация метрики при кластеризации коллекции документов // Известия Тульского государственного университета, Естественные науки, 2012. No 3. С. 119-131.
 17. Мотренко А.П., Стрижов В.В. Многоклассовая логистическая регрессия для прогноза вероятности наступления инфаркта // Известия ТулГУ, 2012. No 1. С. 153-162.
 18. Стрижов В.В., Кузнецов М.П., Рудаков К.В. Метрическая кластеризация последовательностей аминокислотных остатков в ранговых шкалах // Математическая биология и биоинформатика, 2012. Т. 7(1). С. 345-359.
 19. Сандуляну Л.Н., Стрижов В.В. Выбор признаков в авторегрессионных задачах прогнозирования // Информационные технологии, 2012. No 7. С. 11-15.
 20. Токмакова А.А., Стрижов В.В. Оценивание гиперпараметров линейных и регрессионных моделей при отборе шумовых и коррелирующих признаков // Информатика и её применения, 2012, Т. 6(4). С. 66-75.
 21. Кузнецов М.П., Стрижов В.В., Медведникова М.М. Алгоритм многоклассовой классификации объектов, описанных в ранговых шкалах // Научно-технический вестник С.-Пб.ПГУ. Информатика. Телекоммуникации. Управление 2012. No 5. С. 92-95.
 22. Крымова Е.А., Стрижов В.В. Алгоритмы выбора признаков линейных регрес-

- сионных моделей из конечного и счетного множеств, Заводская лаборатория, 2011. Т. 77(5). С. 63-68.
23. Стрижов В.В., Крымова Е.А. Выбор моделей в линейном регрессионном анализе // Информационные технологии, 2011. №. 10. С. 21-26.
24. Стрижов В.В. Уточнение экспертных оценок, выставленных в ранговых шкалах, с помощью измеряемых данных // Заводская лаборатория, 2011. Т. 77(7). С. 72-78.
25. Стрижов В.В., Сологуб Р.А. Индуктивное порождение регрессионных моделей предполагаемой волатильности для опционных торгов // Вычислительные технологии, 2009. Т. 14(5). С. 102-113.
26. Стрижов В.В., Казакова Т.В. Устойчивые интегральные индикаторы с выбором опорного множества описаний // Заводская лаборатория, 2007. Т. 73(7). С. 72-76.
27. Стрижов В.В. Поиск параметрической регрессионной модели в индуктивно заданном множестве // Вычислительные технологии, 2007. № 1. С. 93-102.
28. Стрижов В.В. Уточнение экспертных оценок с помощью измеряемых данных // Заводская лаборатория, 2006. Т. 72(7). С. 59-64.

На диссертацию поступили отзывы:

В отзыве д.ф.-м.н. С.Д. Двоенко отмечено, что

«Данная работа является актуальной: для получения структурно простых моделей, устойчиво и точно приближающих результаты измерений необходимы методы, позволяющие выбирать адекватные модели из некоторого множества допустимых моделей-претендентов. Задача настройки моделей со многими параметрами является многоэкстремальной оптимизационной задачей. Автор предлагает упростить эту задачу путем анализа наборов последовательно порождаемых устойчивых моделей. [...] В работе предложен новый подход для решения задачи автоматического последовательного выбора моделей регрессионного анализа и классификации. Предложенные автором диссертации решения аргументированы и оценены по сравнению с другими известными решениями. [...] Форма представления материала работы, полученные выводы представляются вполне достоверными. Научные положения, выводы и рекомендации, сформулированные в

диссертации, являются обоснованными в достаточной степени».

«Замечание. Во всяком исследовании системы случайных величин очевидно, что предполагается наличие их совместного распределения. В данной работе такие величины предполагаются не совсем тривиальными. По-видимому, было бы методически более корректным в данной работе ввести сначала явно их совместное распределение и охарактеризовать его в рамках проблематики работы.

Высказанное замечание не снижает ценности работы и не влияет на ее общую положительную оценку».

В отзыве д.ф.-м.н. Ю.Г. Сметанина отмечено, что

«Задача выбора моделей, решаемая в диссертационной работе, является одной из центральных и актуальных проблем машинного обучения и интеллектуального анализа данных. В работе предложен новый подход к автоматическому построению моделей регрессии и классификации. Центральной проблемой, исследуемой работе, является проблема выбора из большого количества порождаемых моделей. Для решения этой проблемы разработаны новые методы направленного поиска в пространстве моделей. Приводится теоретическое обоснование предлагаемых методов. [...]

Работа не свободна от недостатков, не снижающих, впрочем, положительного впечатления о ней: недостаточно подробно изучены свойства оценок вероятностей связей между элементами суперпозиции, не проведен сравнительный анализ предложенных способов получения оценок гиперпараметров моделей, в работе не исследуется проблема многоэкстремальности пространства параметров моделей.»

В отзыве д.ф.-м.н. М.Ю. Хачая отмечено, что

«В диссертационной работе В.В. Стрижова получены следующие основные результаты. Разработаны новые способы построения моделей как суперпозиций параметрических функций из экспертно-задаваемого множества для решения задач регрессии и классификации. Предложен ряд новых алгоритмов получения оценок параметров и гиперпараметров регрессионных моделей, основанных на свойствах связанного байесовского вывода. [...]

По сути представленных результатов замечаний нет. В качестве пожелания к дальнейшим исследованиям хотелось бы отметить следующее. Большая часть полученных в работе результатов получена в предположении о нормальности законов распределения параметров (гиперпараметров) и помех. Интересна возможность распространения этих результатов на случай более общих параметрических систем распределений».

Выбор официальных оппонентов и ведущей организации обосновывается известностью их достижений в области машинного обучения и анализа данных.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

- Формализованы и исследованы методы выбора моделей для основных классов моделей: линейных, обобщенно-линейных и нелинейных.
- Предложен способ оценки информативности элементов суперпозиций путем анализа пространства параметров моделей.
- Предложены алгоритмы оптимизации параметров и гиперпараметров параметров функций распределения параметров моделей.
- Исследованы ограничения, накладываемые на множество суперпозиций, при которых порождаемые суперпозиции являются допустимыми, предложены методы порождения допустимых суперпозиций.
- Предложен метод последовательного порождения и выбора моделей.
- Предложен метод анализа ковариационной матрицы параметров нелинейных моделей. Предложен критерий отыскания мультиколлинеарности.

Новизна работы. Выносимые на защиту результаты являются новыми; также новыми являются следующие результаты, ранее опубликованные автором в рецензируемых журналах: 1) метод индуктивного порождения регрессионных моделей как суперпозиций гладких функций из заданного множества; 2) алгоритм выбора наиболее информативных элементов суперпозиции с помощью вектора гиперпараметров; 3) метод выбора опорного множества объектов как альтернатива процедурам регуляризации при построении интегральных индикаторов; 4) алгоритм согласования экспертных оценок в ранговых шкалах: используется линейная комбинация конусов экспертных оценок в пространстве интегральных

индикаторов и в пространстве весов показателей, 5) алгоритм решения прямой и обратной задачи при нахождении оптимального управления.

Теоретическая значимость. Впервые связаны методы порождения и методы выбора моделей. При этом снята проблема оценки параметров и их ковариационных матриц моделей большой структурной сложности, так как для этой оценки параметров последующих моделей используются результаты анализа ранее порожденных моделей. Такой подход позволяет получать устойчивые оценки параметров в условиях большого числа мультикоррелирующих и шумовых признаков. Для выбора конкурирующих моделей используется байесовский подход, что позволяет получить модель оптимальной статистической сложности.

Значение полученных соискателем результатов исследования **для практики** подтверждается тем, что предложены решения ряда прикладных задач, приведенные в последней главе работы. В частности, получены и проанализированы решения задачи прогнозирования объемов потребления электроэнергии, обратной задачи управления в макроэкономическом анализе, задачи кластеризации аминокислотных последовательностей, задачи классификации состояния физической активности человека, задачи классификации пациентов с сердечными заболеваниями. Решения были использованы в работе российских и зарубежных компаний.

Для иллюстрации возможных практических применений в последней главе работы приведены математические постановки и анализ прикладных задач, при решении которых были использованы полученные результаты.

Достоверность и обоснованность результатов подтверждена строгостью и корректностью математических высказываний и доказательств. Выполнена экспериментальная проверка полученных результатов на задачах с модельными и реальными данными. Результаты исследований неоднократно обсуждались на российских и международных научных конференциях.

Личный вклад соискателя: все основные результаты диссертационной работы получены автором самостоятельно. Вклад автора в получение результатов, опубликованных в наиболее значимых работах: [1] Предложен способ построения и выбора модели, определенной на описаниях объектов, выполненных в ранговых шкалах. [2] Предложен метод оценки необходимого объема выборки для задач ло-

гистической регрессии. [3] Предложен метод последовательного порождения регрессионных моделей. [4] Предложен ряд алгоритмов построения и согласования интегральных индикаторов в линейных шкалах. [5] Предложен способ оценки гиперпараметров, ковариационных матриц распределения параметров нелинейных регрессионных моделей. [6] Исследована проблема построения иерархических моделей коллекций коротких документов. [7] Описаны основные теоретические положения построения функции ошибки регрессионных моделей с помощью связанного байесовского вывода. Приведены сравнения с базовыми функциями ошибки. [8] Предложен новый структурный подход к построению ранжирующих моделей. [9] Предложен оптимизационный алгоритм для решения многоэкстремальных задач высокой размерности. [10] Исследованы модели построения программ на классах. [11] Предложен и теоретически обоснован метод оценки гиперпараметров линейных моделей. [12] Поставлена и решена задача построения структурных моделей. [13] Предложен способ верификации тематических моделей больших коллекций текстовых документов с иерархической структурой. [14] Предложена новая структурная ранжирующая модель для построения кластерных интегральных индикаторов. [15] Предложен новый алгоритм порождения и упрощения структур суперпозиций регрессионных моделей. [16] Разработан и обоснован алгоритм выбора тематических моделей иерархических коллекций коротких документов. [17] Решена прикладная задача многоклассовой классификации. [18] Предложен и теоретически обоснован алгоритм быстрой кластеризации выборок большого объема. [19] Предложен метод выбора признаков для существенно вырожденной матрицы плана. [20] Исследованы свойства выборок, для которых получение адекватных моделей затруднительно. [21] Предложен новый подход к построению интегральных индикаторов путем последовательного построения ранжирующих функций. [22] Исследована проблема мультиколлинеарности при выборе признаков. [23] Выполнен обзор и аналитическое сравнение методов выбора линейных регрессионных моделей. [24] Впервые предложено использовать выпуклые многогранные конусы для построения моделей при согласовании экспертных оценок. [25] Предложен метод последовательного построения нелинейных интерпретируемых регрессионных моделей. [26] Предложен и

теоретически обоснован способ прореживания выборки с целью получения устойчивых регрессионных моделей. [27] Предложен алгоритм индуктивного построения и оценки параметров нелинейных регрессионных моделей. [28] Предложен способ согласования экспертных оценок, выставленных в линейных шкалах путем решения регрессионной задачи с регулируемыми структурными параметрами.

На заседании 13 ноября 2014 г. диссертационный совет принял решение присудить Стрижову Вадиму Викторовичу учёную степень доктора физико-математических наук.

При проведении тайного голосования диссертационный совет в количестве 27 человек, из них 11 докторов наук по специальности защищаемой диссертации, участвовавших в заседании, из 35 человек, входящих в состав совета, дополнительно введены на разовую защиту 0 человек. Проголосовали: за присуждение учёной степени 27, против присуждения учёной степени 0, недействительных бюллетеней 0.

Председатель диссертационного совета
академик РАН, д.ф.-м.н.



[Signature]
Журавлёв Ю.И.

Учёный секретарь диссертационного совета
д.ф.-м.н., профессор

[Signature]
Рязанов В.В.

« ___ » _____ 2014 г.