

Отзыв
официального оппонента на диссертацию Стрижова Вадима Викторовича
«Порождение и выбор моделей в задачах регрессии и классификации»,
представленную на соискание ученой степени доктора физико-
математических наук по специальности 05.13.17 – теоретические основы
информатики

Проблематика диссертационной работы. Работа посвящена проблемам построения моделей в задачах регрессии и классификации. Предлагается новый подход к решению задачи выбора моделей.

Выбор производится из счетного множества моделей – суперпозиций экспертно-заданных порождающих функций. Для нахождения оптимальной модели предлагается оптимизировать ее правдоподобие – комбинированный критерий, включающий в себя понятия сложности и точности модели. Модели выбираются из последовательно порождаемых наборов. При выборе на каждом шаге с помощью оценок параметров и структурных параметров анализируется структура модели, выбирается множество моделей-претендентов субоптимальной сложности. Для получения оценок параметров предлагается использовать связанный байесовский вывод. Предполагается, что модели могут быть проинтерпретированы экспертами в рамках предметной области.

Актуальность. 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.

Новизна. 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.

Достоверность полученных результатов. Форма представления материала работы, полученные выводы представляются вполне достоверными. Научные положения, выводы и рекомендации, сформулированные в диссертации, являются обоснованными в достаточной степени.

Содержание работы. Во введении обосновывается актуальность работы, определяются цели и задачи исследования.

В первой главе ставится задача выбора регрессионных моделей. Вводятся основные определения функции регрессии, параметров, структурных параметров модели. Предлагаются альтернативные методы оценки параметров и структурных параметров модели. Вводится понятие гипотезы порождения данных, описываются методы тестирования адекватности модели.

Во второй главе рассмотрены методы порождения моделей из заданного множества порождающих функций. В качестве алгоритмов, позволяющих сократить перебор всех допустимых суперпозиций, предлагаются методы символьной регрессии для поиска оптимального дерева в графе суперпозиции. Предлагается метод прогнозирования структуры суперпозиции, основанный на оценке элементов матрицы инцидентности графа.

В третьей главе предлагается метод выбора моделей, основанный на сравнении сложности элементов последовательно порождаемых моделей. В частности, предлагается метод последовательного добавления и удаления элементов модели для достижения оптимального значения функции качества. Производится сравнение различных методов выбора признаков как элементов линейных моделей.

В четвертой главе исследуется и развивается метод связанного байесовского вывода для выбора моделей. Предлагаются методы оценки параметров и структурных параметров модели, основанные на аппроксимации функции ошибки, а также на семплировании пара-

метров модели. Для удаления неинформативных элементов модели предлагается использовать информацию об оценках параметров и структурных параметров. Предлагается процедура выбора многоуровневых моделей.

В пятой главе рассмотрен выбор моделей для данных в разнородных шкалах. Рассматривается задача построения интегральных оценок качества объектов. На множестве объектов и признаков рассматриваются экспертные оценки в линейных и ранговых шкалах. Предлагается метод согласования экспертных оценок для построения устойчивых ранжирующих моделей. В качестве аналитического описания ранговой шкалы рассматриваются конусы в пространстве объектов и признаков. Предлагается метод многоклассовой Парето-классификации объектов с ранговым экспертным описанием.

В шестой главе выполнен анализ некоторых прикладных задач. Рассматриваются основные этапы постановки прикладной задачи как задачи регрессионного анализа. Рассматривается задача прогнозирования квазипериодических временных рядов на основе построения матрицы авторегрессии, задача сглаживания временного ряда, задача восстановления форм геометрических фигур, задача оценки волатильности случайных процессов, задача кластеризации структур белка. Для решения этих задач используются предлагаемые в работе методы выбора нелинейных моделей.

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.

Заключение. Результаты диссертационной работы изложены в 28 научных статьях, опубликованных в журналах, входящих в перечень ВАК. Результаты докладывались на российских и международных конференциях.

Диссертация соответствует критериям, установленным Положением о присуждении ученых степеней: автором исследований разработаны теоретические положения, совокупность которых можно квалифицировать как научное достижение. В ней содержится решение задачи, имеющей значение для развития методов анализа данных, она написана автором самостоятельно, обладает внутренним единством, содержит новые научные результаты, полученные автором лично, и свидетельствует о личном вкладе автора в науку. В диссертации приведены рекомендации по использованию научных выводов.

Работа полностью соответствует требованиям ВАК 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.

Официальный оппонент

д.ф.-м.н., доцент, профессор кафедры «Информационная безопасность»

Института прикладной математики и компьютерных наук

Тульского государственного университета (ИБ ИПМKN ФГБОУ ВПО ТулГУ с 1.10.2014)

Адрес: 300012, Тула, пр. Ленина, 92

Тел. 8 4872 357984

E-mail: danilov@tsu.tula.ru

Двоенко Сергей Данилович

