

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертационную работу Адуенко Александра Александровича

«Выбор мультимodelей в задачах классификации»,

представленную на соискание ученой степени

кандидата физико-математических наук по специальности

01.01.09 – «Дискретная математика и

математическая кибернетика»

Оценка актуальности темы диссертационной работы

Проблематика диссертационной работы. Работа посвящена исследованию проблемы выбора адекватных мультимodelей, содержащих попарно статистически различимые модели, в задачах классификации. Эта задача является актуальной, так как наличие похожих моделей в мультимodelи ведет к неинтерпретируемости и низкому качеству прогноза. Разработаны методы прореживания мультимodelей, основанные на статистическом подходе к задаче сравнения моделей. Предложена функция сходства апостериорных распределений параметров моделей, позволяющая решать эту задачу. Рассмотрена задача оптимизации структуры моделей в мультимodelи и предложен метод учета мультиколлинеарности признаков.

Актуальность. Диссертационная работа является актуальной: разработанные модели и методы могут быть использованы для решения задач классификации, регрессии, информационного поиска.

Структура и содержание работы

Диссертационная работа состоит из введения, пяти глав, заключения и списка литературы.

Во введении обоснована актуальность работы, определены цели и задачи исследования.

В первой главе введены основные понятия и обозначения, включая описание мультимodelьного подхода, а также понятия оптимальности и обучения вероятностных моделей классификации.

Во второй главе рассмотрены задачи отбора признаков и учета мультиколлинеарности между ними. Показано, что принцип максимума обоснованности приводит к асимптотически вырожденной оценке недиагональной ковариационной матрицы параметров логистической модели. Для учета мультиколлинеарности предложена оптимальная схема линейного комбинирования признаков.

В третьей главе рассмотрена задача обучения многоуровневых моделей и смесей моделей. Для смесей моделей предложен алгоритм совместного обучения и оптимизации моделей в смеси.

В четвертой главе поставлена задача построения адекватных мультимodelей, содержащих попарно статистически различимые модели. Для решения задачи статистического сравнения моделей предложена функция схождения апостериорных распределений параметров моделей. Показано, что она удовлетворяет требованиям задачи сравнения моделей. Доказаны теоремы об асимптотическом распределении значений функции схождения для совпадающих обобщенно-линейных моделей. На основании этих результатов предложены методы прореживания мультимodelей для построения адекватных мультимodelей. Получены верхняя и нижняя оценки на максимальное число моделей в адекватной мультимodelи.

Пятая глава посвящена анализу результатов работы методов на реальных и синтетических данных, а также установлению границ применимости методов. Приведено описание разработанного программного комплекса.

Основные результаты и их научная новизна

1. Разработан статистический подход к задаче сравнения моделей в мультимodelи. Для решения этой задачи предложена функция схождения плотностей апостериорных распределений параметров моделей. Показано, что она удовлетворяет требованиям к функции схождения для решения задачи сравнения моделей. Исследованы свойства предложенной функции схождения и получено ее асимптотическое распределение для пары совпадающих обобщенно-линейных моделей.
2. Предложены методы прореживания мультимodelей для построения адекватных мультимodelей, содержащих попарно статистически различимые модели. Получены верхняя и нижняя оценки на максимальное число моделей в адекватной мультимodelи. Предложен метод совместного обучения и отбора признаков для смеси моделей.
3. Доказана асимптотическая вырожденность недиагональной оценки максимума обоснованности для ковариационной матрицы параметров логистической модели. Предложен метод детектирования и учета мультиколлинеарности между признаками путем их комбинирования.
4. Разработан программный комплекс для построения и обучения адекватных мультимodelей, оптимизации структуры моделей в мультимodelи и учета мультиколлинеарности признаков при решении задач классификации. Проведены вычислительные эксперименты, подтверждающие значимость улучшения качества классификации на реальных данных, а также существенное сокращение числа моделей в мультимodelях.

Новизна. Предложенная функция схождения апостериорных распределений параметров моделей и метод статистического сравнения моделей в мультимodelи на основании этой функции являются новыми. Утверждение об асимптотической вырожденности недиагональной оценки максимума обоснованности для ковариационной матрицы весов признаков и предложенный метод комбинирования

мультиколлинеарных признаков являются новыми. Предложенный метод совместного обучения и отбора признаков для смесей моделей обобщает известные методы раздельного обучения и отбора признаков и является новым.

Достоверность полученных результатов. Достоверность подтверждается публикациями результатов в высокоцитируемых научных журналах, а также докладами на российских и международных конференциях по анализу данных и исследованию операций. Научные положения, выводы и рекомендации, сформулированные в диссертации, являются обоснованными и подкреплены необходимыми математическими утверждениями.

Замечания

1. В тексте диссертации встречаются незначительные опечатки (например, стр. 110, 145, 147), оформление некоторых рисунков (например, 5.20, 5.21) неудобно, так как они располагаются на следующей, а иногда и через одну страницу от ссылки на них.
2. Во второй части раздела 4.5, посвященной сравнению моделей линейной регрессии, отсутствует прямое сравнение требований к моделям для наличия асимптотического распределения хи-квадрат у функции сходства моделей с требованиями к обобщенно-линейным моделям, хотя и указано, что во втором случае они сильнее.
3. В работе отсутствуют оценки быстродействия разработанных алгоритмов.

Заключительная оценка

Приведенные замечания не снижают общей высокой оценки диссертационной работы.

Результаты диссертационной работы изложены в 14 научных статьях, из которых 9 опубликованы в журналах из Перечня ведущих рецензируемых научных журналов, включенных Высшей аттестационной комиссией России в список изданий, рекомендуемых для опубликования основных научных результатов диссертации на соискание ученой степени кандидата и доктора наук. Результаты докладывались на Всероссийских и международных конференциях. Автореферат достаточно полно и правильно отражает содержание диссертации.

Диссертационная работа представляет собой законченное научное исследование, содержит 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 (п. 9 «Положения о присуждении учёных степеней», утверждённого Постановлением Правительства РФ от 24 сентября 2013 года № 842), предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук по специальности 01.01.09 – «Дискретная математика и математическая кибернетика», а ее автор, А.А. Адуенко, заслуживает присуждения ему ученой степени кандидата физико-математических наук по данной специальности.

Официальный оппонент

Кандидат физико-математических наук, доцент,
доцент центра по научным и инженерным вычислительным технологиям для задач с большими массивами данных

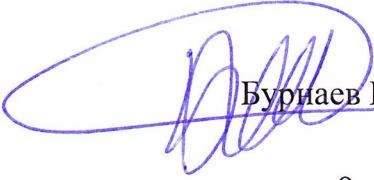
Автономной некоммерческой организации высшего профессионального образования «Сколковский институт науки и технологий»

<http://www.skoltech.ru/>

Адрес: 143026 Москва, Территория Инновационного Центра “Сколково”, улица Нобеля, д. 3

Тел.: +7 (495) 280 14 81

E-mail: e.burnaev@skoltech.ru


Бурнаев Евгений Владимирович

9 марта 2017 г.

*Министерство по административным
и кадровым вопросам*

