

МИНОБРНАУКИ РОССИИ  
Федеральное государственное бюджетное  
образовательное учреждение высшего  
образования



«Тульский  
государственный  
университет»  
(ТулГУ)



Проспект Ленина, д. 92, г.Тула, 300012  
Тел. (4872) 33-24-10, факс (4872) 35-81-81  
e-mail: info@tsu.tula.ru, http://www.tsu.tula.ru

№ \_\_\_\_\_

УТВЕРЖДАЮ

Проректор по научной работе  
Тульского государственного университета  
д.т.н., профессор  
Кухарь Владимир Денисович



21 апреля 2017 года

## ОТЗЫВ

ведущей организации – Федерального государственного  
бюджетного образовательного учреждения высшего образования  
«Тульский государственный университет»  
на диссертационную работу Адуенко Александра Александровича  
на тему **«Выбор мультимоделей в задачах классификации»**,  
представленную на соискание ученой степени  
кандидата физико-математических наук по специальности  
**01.01.09 – «Дискретная математика и  
математическая кибернетика»**

### Актуальность темы

Задача классификации является базовой в машинном обучении, при этом задачи многоклассовой классификации могут быть эффективно сведены к решению одной или нескольких задач двухклассовой. Модель логистической регрессии, являющаяся стандартом во многих практических областях, была предложена в 1944 году Берксоном, а обобщенно-линейные модели – в 1972 году Нельдером и Веддербурном. Однако эти модели не позволяют учесть неоднородности в данных. Для учета неоднородностей в данных были предложены композиции моделей, что привело к появлению проблемы интерпретируемости композиции, поскольку модели в композиции могут быть близкими или совпадать. Для решения этой проблемы были предложены разные методы прореживания композиций, включая генетические алгоритмы (Жу: 2003), жадные стратегии выбора (Martínez-Muñoz: 2006), кластеризацию (Бэккер: 2003), введение априорного распределения, поощряющего разреженность и выбор структуры модели из принципа максимума обоснованности (МакКэй: 1992, Блей: 2003). Однако эти методы прореживания композиций прямо не учитывают близости между моделями в мультимодели, а потому мультимодель по-прежнему может содержать близкие модели. Для получения статистически различных моделей в мультимодели используют внешнюю процедуру прореживания, основанную на статистическом сравнении моделей путем подсчета расстояний между апостериорными распределениями параметров для разных моделей, например, с помощью дивергенций Брегмана или f-дивергенций (Basseville: 2013, Petz: 2007).



В работе А.А. Адуенко развивается статистический подход к задаче построения адекватных композиций, содержащих попарно различимые модели. Показано, что существующие меры сходства отличают неинформативную модель от совпадающей, но информативной модели, а потому не позволяют решить задачу сравнения моделей. С помощью построения новой функции сходства апостериорных распределений параметров А.А. Адуенко удалось решить задачу статистического сравнения моделей. На основании этого предложен метод прореживания мультимodelей для получения адекватной мультимodelи. Получены верхняя и нижняя оценки на максимальное число моделей в адекватной мультимodelи. Предлагаемый в работе А.А. Адуенко подход позволяет решить актуальную проблему получения адекватной мультимodelи в задачах классификации, содержащей небольшое число попарно различных моделей и обладающей лучшим качеством классификации по сравнению с исходной.

## **Содержание**

Диссертационная работа состоит из введения, пяти разделов и заключения. Во введении приведена актуальность работы, установлены цели и задачи, перечислены основные выносимые на защиту положения, приведен краткий обзор методов исследования, установлена научная новизна, теоретическая и практическая значимость, приведена достоверность результатов, перечислены основные публикации по теме диссертации.

В первой главе введены основные понятия и определения. Рассмотрены задача классификации, ее решение в общем виде, а также понятие оптимальности и обучения вероятностных моделей классификации. Изложен мультимodelьный подход к задаче классификации.

Вторая глава посвящена задаче построения оптимальной мультимodelи и комбинированию признаков с учетом их мультиколлинеарности. Доказана теорема об оптимальной схеме линейного комбинирования признаков. Доказана асимптотическая вырожденность недиагональной оценки максимума обоснованности ковариационной матрицы параметров логистической регрессии.

В третьей главе рассматривается задача обучения мультимodelей. Предложен алгоритм совместного обучения и оптимизации смесей моделей.

Четвертая глава посвящена задаче статистического сравнения моделей и построения адекватных мультимodelей. Построена функция сходства апостериорных распределений параметров моделей и показано, что она отвечает требованиям к функции сходства, то есть является корректной. Доказана некорректность существующих функций сходства. Получены верхняя и нижняя оценки на максимальное число попарно различных моделей в мультимodelи. Исследованы статистические свойства распределения функции сходства в условиях истинности гипотезы о совпадении моделей и доказаны теоремы об ее асимптотическом распределении для обобщенно-линейных моделей. На основании этих результатов предложен метод прореживания мультимodelей для построения адекватной мультимodelи.

В пятой главе приведено описание разработанного программного комплекса для решения задач классификации и комбинирования признаков и анализ результатов на синтетических и реальных данных в сравнении с известными методами.

## **Основные результаты и их новизна**

В рамках диссертационной работы А.А. Адуенко получены следующие новые результаты:

1. Предложена функция сходства плотностей апостериорных распределений параметров моделей, удовлетворяющая требованиям к функции сходства для решения задачи сравнения моделей. Показано, что известные функции сходства требованиям не удовлетворяют.



2. Получены верхняя и нижняя оценки на максимальное число моделей в адекватной мультимодели.
3. Разработаны методы выбора адекватных оптимальных обученных мультимodelей, содержащих попарно статистически различимые модели.
4. Предложен метод комбинирования мультиколлинеарных признаков. Доказана асимптотическая вырожденность недиагональной оценки максимума обоснованности ковариационной матрицы параметров логистической модели.
5. Разработан программный комплекс для построения адекватных мультимodelей при решении задач классификации. Модели протестированы на синтетических и реальных данных.

### **Достоверность результатов**

Математические результаты диссертационной работы А.А. Адуенко сформулированы в виде утверждений и теорем с корректными доказательствами. Проведена экспериментальная проверка методов на задачах классификации реальных и синтетических данных. Положения и выводы, сформулированные в диссертации, получили квалифицированную апробацию на международных и российских конференциях и семинарах. Основные результаты по теме диссертации изложены в четырнадцати работах в рецензируемых научных журналах.

### **Значимость результатов**

Работа А.А. Адуенко носит преимущественно теоретический характер. Основной ее ценностью является разработка статистического подхода к сравнению моделей в мультимodelях и методов построения адекватных мультимodelей, содержащих попарно статистически различимые модели. Эффективность разработанных методов построения адекватных оптимальных обученных мультимodelей в терминах значительного повышения качества классификации и снижения числа моделей в мультимodelях продемонстрирована на реальных данных.

### **Замечания**

1. В тексте диссертации встречаются незначительные опечатки (с. 59, 110) и переполнение строк (с. 85).
2. При формулировании определения 21 опущено слово «нижняя», хотя в дальнейшем оценки такого вида называются в диссертации «вариационными нижними оценками».
3. В алгоритме комбинирования признаков не пояснено, как получено формулы итеративного пересчета оценок дисперсии шума в каждом из них.

### **Заключительная оценка**

Диссертационная работа Адуенко Александра Александровича «Выбор мультимodelей в задачах классификации» является законченной научно-квалификационной работой, выполненной под руководством д. ф.-м. н. В.В. Стрижова, содержащей новое решение актуальной научной задачи - построение адекватных мультимodelей для решения задач классификации с неоднородностями в данных. Представленные в работе результаты и выводы обоснованы корректно. Результаты работы своевременно и полно опубликованы в 14 печатных трудах, 9 из которых - в изданиях, рекомендованных ВАК РФ. Результаты докладывались автором на российских и международных конференциях. Автореферат правильно и полно отражает содержание диссертационной работы.

Диссертация соответствует всем критериям, установленным Положением о присуждении ученых степеней, предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук по специальности 01.01.09 – «Дискретная математика и математическая кибернетика», а ее автор – Адуенко Александр Александрович – заслуживает присуждения ученой степени кандидата физико-математических наук по указанной специальности.

Настоящий отзыв обсужден и одобрен на расширенном заседании кафедры информационной безопасности Федерального государственного бюджетного учреждения высшего образования «Тульский государственный университет» (протокол №10 от 04 апреля 2017 г.)

Заведующий кафедрой ИБ ФГБОУ ВО ТулГУ, к.т.н



Сычугов А.А.

Д.ф.-м.н, профессор



Двоенко С. Д.

