

УТВЕРЖДАЮ

Проректор по научной работе и
программам развития, к.ф.-м.н.

В.А. Баган

2019 г.



ЗАКЛЮЧЕНИЕ

федерального государственного автономного образовательного
учреждения высшего образования
«Московский физико-технический институт (национальный исследовательский
университет)» (МФТИ)

Диссертация «Выбор моделей прогнозирования мультикоррелирующих временных рядов» выполнена на кафедре интеллектуальных систем федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (национальный исследовательский университет)» Министерства науки и высшего образования Российской Федерации.

В период подготовки диссертации с 2014 г. по настоящее время соискатель **Мотренко Анастасия Петровна** работала в Федеральном исследовательском центре «Информатика и управление» Российской академии наук в должности инженера-исследователя.

В 2014 г. А.П. Мотренко с отличием окончила федеральное государственное автономное образовательное учреждение высшего образования «Московский физико-технический институт (государственный университет)» по направлению подготовки 03.04.01 Прикладные математика и физика. В 2014 г. поступила в очную аспирантуру МФТИ. В 2018 году окончила аспирантуру федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (национальный исследовательский университет)» по направлению подготовки 09.06.01 Информатика и вычислительная техника.

Справка о сдаче кандидатских экзаменов выдана в 2019 году в федеральном государственном автономном образовательном учреждении высшего образования «Московский физико-технический институт (национальный исследовательский университет)».

Научный руководитель – доктор физико-математических наук, Стрижов Вадим Викторович. Основное место работы – Федеральный исследовательский центр «Информатика и управление» Российской академии наук, отдел интеллектуальных систем, д.ф.-м.н., ведущий научный сотрудник.

По итогам обсуждения диссертации «Выбор моделей прогнозирования мультикоррелирующих временных рядов» *принято следующее заключение.*

Тема диссертации является актуальной для развития методологии прогнозирования временных рядов, в силу того, что подход, развиваемый в диссертации, учитывает структуру зависимостей между признаками в задачах анализа наборов коррелирующих временных рядов.

Цели и задачи диссертации:

1. Разработка метода выбора модели прогнозирования мультикоррелирующих временных рядов для решения задачи декодирования временного ряда.
2. Разработка метода выбора признаков, учитывающего многоиндексную структуру данных. Метод должен быть вычислительно простым и учитывать связь между признаками.
3. Разработка метода снижения размерности, сохраняющего метрическую структуру данных.
4. Формулировка критерия информативности объекта для построения выборки оптимального состава. Применение построенного критерия для оценки объема выборки и выбора оптимальной модели.
5. Разработка алгоритма построения модели и проведение вычислительного эксперимента для сравнения различных методов выбора признаков при решении задачи декодирования электрокортикограмм.

Основные результаты диссертации:

1. Предложен метод выбора признаков, учитывающий многоиндексное мультикоррелирующее представление данных. Метод позволяет учитывать связи между признаками независимо от модели прогнозирования, не перебирая всех комбинаций признаков.
2. Исследованы свойства предложенного метода. Сформулированы условия локальной сходимости предложенного алгоритма. Доказана приближенная оптимальность низкорангового релаксированного решения.
3. Разработан метод выделения сегментов временных рядов для интерпретируемого сегментирования квазипериодических сигналов.
4. Разработан метод снижения признакового пространства, сохраняющий локальную структуру близости между объектами. Решена проблема учета разметки при выполнении вложения для классификации.
5. Приведены рекомендации по практическому использованию разработанных алгоритмов. Рассмотрены практические задачи декодирования временных рядов электрокортикограмм, выделения внутреннего плагиата, задачи агрегирования и прогнозирования временных рядов грузовых железнодорожных перевозок.

Все *результаты диссертации получены лично соискателем* при научном руководстве д.ф.-м.н. Стрижова В.В.

Научная новизна работы заключается в том, что разработана теория выбора модели прогнозирования при неадекватности предположений о независимости признаков. Предложен способ выбора признаков, учитывающий многоиндексную структуру данных. Предложен способ снижения размерности признакового пространства в задаче классификации, учитывающий локальную структуру близости между объектами и целевую переменную. Предложен способ сегментирования квазипериодических временных рядов с помощью анализа фазовой траектории главных компонент временного ряда. Предложен способ оценки объема

достаточного выборки, основанный на сравнении оценок функции распределения данных, полученных на различных подвыборках исследуемой выборки.

Практическая ценность результатов диссертации заключается в том, что предложенные в работе методы выбора признаков для построения моделей прогнозирования повышают качество классификации и снижают сложность моделей, сохраняя их интерпретируемость.

Обоснованность и достоверность результатов и выводов.

Достоверность результатов подтверждена математическими доказательствами, экспериментальной проверкой полученных методов на реальных задачах; публикациями результатов исследования в рецензируемых научных изданиях, в том числе рекомендованных ВАК. Результаты работы докладывались и обсуждались на российских и международных научных конференциях по анализу данных и исследованию операций.

Материалы диссертации опубликованы автором достаточно полно в следующих работах, опубликованных в журналах, рекомендованных ВАК РФ:

1. A. Motrenko, V. Strijov. Multi-way Feature Selection for ECoG-based Brain-Computer Interface // Expert Systems with Applications. – 2018. – V. 114. – № 30. – P. 402-413.
 2. A. Aduenko, A. Motrenko, V. Strijov. Object selection in credit scoring using covariance matrix of parameters estimations // Annals of Operations Research. – 2018. № 1. – P. 3–21.
 3. K.V. Rudakov, V.V. Strizhov, D.O. Kashirin, M.P. Kuznetsov, A.P. Motrenko, M.M. Stenina. Selecting an Optimal Model for Forecasting the Volumes of Railway Goods Transportation // Automation and Remote Control. – 2017. – V. 78. – № 1. P. 74–87.
 4. A. Motrenko, K. Rudakov, V. Strijov. Combining endogenous and exogenous variables in a special case of non-parametric time series forecasting model // Moscow University Computational Mathematics and Cybernetics. – 2016. V. 40. – № 2. – P. 71—78.
 5. A. Motrenko, V. Strijov. Extracting fundamental periods to segment biomedical signals // Journal of Biomedical and Health Informatics. – 2016. V. 20. – № 6. – P. 1466–1476.
 6. И. О. Молибог, А. П. Мотренко. Повышение качества классификации в задаче обнаружения внутреннего плагиата // Информатика и её применения. – 2017, Т. 11. – № 3. – С. 59–71.
 7. А. П. Мотренко, В. В. Стрижов. Построение агрегированных прогнозов объёмов железнодорожных грузоперевозок с использованием расстояния Кульбака-Лейблера // Информатика и её применения. – 2014. – Т. 8. – № 2. – С. 86–97.
 8. A. Motrenko, V. Strijov, G.-W. Weber. Sample size determination for logistic regression // Journal of Computational and Applied Mathematics. – 2014. – № 255. – P. 743–752.
 9. А. П. Мотренко, В. В. Стрижов. Многоклассовая логистическая регрессия для прогноза вероятности наступления инфаркта // Известия ТулГУ. – 2012. № 1. – С. 153–162.
- и других изданиях:
10. Э.Э. Гасанов, А.П. Мотренко. Построение аппроксимирующего описания скалограммы в задаче прогнозирования движений по электрокортикограмме // Машинное обучение и анализ данных. – 2017. – Т. 3. – № 2. – С. 160–169.

11. M. Kuznetsov, A. Motrenko, R. Kuznetsova, V. Strijov Methods for Intrinsic Plagiarism Detection and Author Diarization // CEUR Workshop Proceedings. CLEF2016 Working Notes. – 2016. – V. 1609. – P. 912–919.
12. А.С. Вальков, Е.М. Кожанов, А.П. Мотренко, Ф.И. Хусаинов. Построение кросс-корреляционных зависимостей при прогнозе загруженности железнодорожного узла // Машинное обучение и анализ данных. – 2013. – № 5. – С. 503–517.
13. А. П. Мотренко Оценка плотности совместного распределения // Машинное обучение и анализ данных. – 2012. № 4. – С. 428–435.
14. А. П. Мотренко Оценка необходимого объема выборки пациентов при прогнозировании сердечно-сосудистых заболеваний // Машинное обучение и анализ данных. – 2012. № 3. – С. 354–366.
15. А.П. Мотренко. Использование теста Грейнджера при прогнозировании временных рядов // Машинное обучение и анализ данных. – 2011. – № 1. – С. 51–60.

Личный вклад соискателя в работах с соавторами заключается в следующем:

[1, 4, 7, 11, 12] – постановка задачи и подготовка теоретической части, проведение вычислительных экспериментов; [2] – обработка результатов вычислительных экспериментов, [3] – проведение вычислительных экспериментов, [5, 6, 10] – постановка задачи и разработка теоретической части; [8] – разработка и реализация метода оценки достаточного объема выборки, [9, 14] – разработка и реализация метода выбора оптимального набора признаков в логистической регрессии, проведение вычислительного эксперимента; [13] – реализация комбинированного метода оценки плотности многомерного распределения, проведение вычислительного эксперимента; [15] – реализация метода уточнения прогноза временного ряда на основе зависимостей, выявленных с помощью метода Грейнджера, проведение вычислительного эксперимента.

Основные результаты работы докладывались на следующих научных конференциях и семинарах:

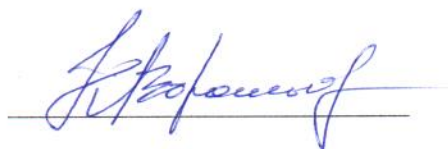
1. Sample size estimation in classification problems // International Digital Processing Conference, Barcelona, 2016.
2. Оценка объема выборки в задачах классификации // Математические методы распознавания образов, г. Светлогорск, Калининградская область, 2015
3. Bayesian Sample Size Estimation for Patient Classification Survey // 20th Conference of the International Federation of Operational Research Societies, Barcelona, 2014
4. Small CVD sample set classification: generative versus discriminative // 26th European conference on operation research, Rome, 2013.
5. Multiclass classification of cardio-vascular disease patients with sample size estimation // 25th European conference on operation research, Vilnius, 2012.

Содержание диссертации соответствует паспорту специальности 05.13.17 – теоретические основы информатики, в частности, пунктам:

- п. 5 – Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений.
- п. 6 – Разработка методов, языков и моделей человекомашинного общения; разработка методов и моделей распознавания, понимания и синтеза речи, принципов и методов извлечения данных из текстов на естественном языке.

Диссертация «Выбор моделей прогнозирования мультикоррелирующих временных рядов» **Мотренко Анастасии Петровны** рекомендуется к защите на соискание ученой степени кандидата физико-математических наук по специальности **05.13.17 – теоретические основы информатики**.

Заключение принято на заседании кафедры интеллектуальных систем МФТИ. Присутствовало на заседании 21 человек. Результаты голосования: «за» – 21 чел., «против» – нет, «воздержались» – нет. Протокол № 1 от 18 апреля 2019 года.



Воронцов Константин Вячеславович,
профессор РАН,
д.ф.-м.н.,
заместитель заведующего
кафедры интеллектуальных систем МФТИ