

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертационную работу Мотренко Анастасии Петровны

«Выбор моделей прогнозирования мультикоррелирующих временных рядов»,

представленную на соискание ученой степени

кандидата физико-математических наук по специальности

05.13.17 – «Теоретические основы информатики»

Оценка актуальности темы диссертационной работы

Проблематика диссертационной работы. Работа посвящена исследованию проблемы выбора признаков при построении моделей прогнозирования и классификации в условиях избыточного признакового пространства. Эта задача является актуальной, так как наличие коррелирующих признаков ведет к завышению сложности и снижению обобщающей способности модели. Предложены метод фильтрации признаков, описываемых многоиндексными матрицами, и метод снижения размерности признаков в метрических пространствах, сохраняющих структуру близостей между объектами. Рассмотрены этапы построения выборки при анализе временных рядов - сегментирования временного ряда, определения оптимального объема выборки.

Актуальность. Диссертационная работа является актуальной: разработанные модели и методы могут быть использованы для решения задач прогнозирования, классификации, и декодирования временных рядов.

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. Сформулированы задачи прогнозирования и декодирования временных рядов, формализована проблема мультикорреляций между признаками. Предложен алгоритм фильтрации признаков

в задачах с многоиндексным признаковым описанием. Доказана сходимость задачи оптимизации, лежащей в основе предложенного метода.

Во второй главе рассмотрена задача снижения размерности признакового пространства. Рассмотрен случай, когда признаковое пространство образует многообразие. Предложен стохастический метод снижения размерности, сохраняющий отношения соседства между близко расположенными объектами.

В третьей главе рассмотрена задача сегментирования временных рядов для построения признакового пространства. Предложен метод разбиения квазипериодических рядов на сегменты, соответствующие периодам.

В четвертой главе рассмотрена задача построения выборки оптимального состава. Определено понятие достаточного объема выборки. Предложен критерий сравнения выборок, основанный на сравнении распределений апостериорных параметров модели классификации и метод оценки достаточного объема выборки.

В пятой главе приведены результаты экспериментального анализа свойств предложенных алгоритмов и сравнения предложенных в работе алгоритмов с существующими.

Основные результаты и их научная новизна

1. Предложен многоиндексный метод фильтрации признаков. Задача выбора признаков сформулирована в виде задачи квадратичного программирования. Предложен алгоритм ее решения, доказана сходимость предложенного алгоритма.
2. Предложен метод снижения размерности с сохранением локальной структуры близости данных в задачах обучения с учителем. Разработана модификация непараметрического метода снижения размерности t-SNE. Решена проблема добавления ранее не просмотренных объектов.
3. Предложен способ оценки достаточного объема выборки. Метод основан на определении устойчивости распределение параметров модели. Введен и исследован критерий значимости изменения функции распределения параметров модели.
4. Приведены рекомендации по практическому использованию разработанных алгоритмов. Рассмотрены практические задачи декодирования временных рядов, выделения внутреннего плагиата, задачи агрегирования и прогнозирования временных рядов грузовых железнодорожных перевозок.

Новизна. Предложены методы построения моделей в случае, когда предположения о независимости признаков неадекватны. Метод фильтрации признаков, учитывающий многоиндексную структуру, предлагается впервые.

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. Обозначения в автореферате не всегда соответствуют обозначениям в тексте диссертации, что в нескольких случаях привело к использованию в автореферате не введенных ранее обозначений.
2. В главе 4 разделы 4.1 и 4.2 слабо связаны между собой в плане обозначений, параметрическая постановка сменяется непараметрическим алгоритмом.
3. Оценки быстродействия разработанных алгоритмов даны только для алгоритма фильтрации многоиндексных признаков.

Заключительная оценка

Приведенные замечания не снижают общей высокой оценки диссертационной работы.

Результаты диссертационной работы изложены в 15 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 ученых степеней», утвержденного Постановлением Правительства РФ от 24 сентября 2013 года № 842), предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук по специальности 05.13.17 – «Теоретические основы информатики», а ее автор, А.П. Мотренко, заслуживает

присуждения ей ученой степени кандидата физико-математических наук по данной специальности.

Официальный оппонент

Кандидат физико-математических наук,
старший преподаватель центр по научным и инженерным
вычислительным технологиям для задач с большими массивами данных
Сколковского института 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: m.panov@skoltech.ru

Панов Максим Евгеньевич

27 августа 2019 г.

Егорьев *Панова М.Е. подтверждает.*

Руководитель отдела
Кадрового администрирования



Е. 27.08.2019