

УТВЕРЖДАЮ

Директор Федерального государственного
бюджетного учреждения науки
Института проблем управления
им. В.А. Трапезникова
Российской академии наук,
член-корреспондент РАН

__ Д. А. Новиков

› января 2020 г.

О Т З Ы В

**ведущей организации
на диссертационную работу Уварова Вадима Евгеньевича
«Разработка и исследование методов распознавания
последовательностей, описываемых скрытыми марковскими моделями,
при неполных данных» на соискание ученой степени кандидата
технических наук по специальности 05.13.17 – Теоретические основы
информатики**

Актуальность исследования. Представленное исследование имеет своей целью разработку методов обработки символьных и числовых последовательностей, порождающий механизм которых может быть формализован скрытой марковской моделью (СММ), однако в силу некоторых обстоятельств подвержен помехам извне. На практике такая ситуация возникновения помех, вызывающую пропажу или порчу данных, достаточно широко распространена. При этом существующее состояние теории таково, что случай неполноты описываемых СММ последовательностей не рассматривался исследователями в полной мере, как отдельное явление. Ранее не существовало единого, хорошо исследованного подхода к борьбе с возникновением пропусков в последовательностях, что и ставит целью исправить настоящая диссертация. Исходя из этого и учитывая эффективность такого инструмента машинного обучения, как скрытые марковские модели, актуальность данного исследования не вызывает сомнений.

Тема диссертации соответствует формуле и области исследований специальности 05.13.17 «Теоретические основы информатики» и согласуются с пунктом 5 научной специальности ВАК России «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».

Поставленную цель работы полно и всесторонне раскрывают задачи исследования, которые касаются всех аспектов данной научной проблемы.

Структура диссертации. Диссертация состоит из введения, 5 глав, заключения, списка условных обозначений, словаря терминов, списка литературы (100 наименований) и 2 приложений, изложенных на 134 страницах машинописного текста, включая 28 рисунков и 2 таблицы. Содержание диссертационной работы представлено исследованием состояния вопроса и изложением последовательности защищаемых научных положений.

В первой главе диссертации Уваровым В. Е. сделан подробный анализ существующей теории СММ, основных направлений её развития и ключевых исследований в этой области. При этом автором отмечается недостаточная разработанность теории в области анализа неполных данных, а также несовершенство имеющихся работ, касающихся этого направления.

Основой исследования послужили труды отечественных и зарубежных ученых по тематике скрытых марковских моделей, как касающихся их теоретического усовершенствования, так и их применения к определенной предметной области.

Во второй главе диссертации рассматривается решение классической задачи теории СММ, а именно - декодирования последовательностей, описываемых СММ, но в условиях неполноты данных. Дополнительно к ней, учитывая характер данных, рассматривается задача восстановления неполных последовательностей.

Особый интерес представляет описанная автором идея маргинализации пропусков, а также теоретическое обоснование её применения для решения проблемы нивелирования пропусков в последовательностях. Такое теоретическое обоснование было проведено впервые.

В главе предлагаются методы декодирования и восстановления неполных последовательностей, в основу которых легла указанная выше идея. Преимущество в эффективности данных методов в сравнении с другими подходами на основе СММ было подтверждено с помощью вычислительных экспериментов.

Также в этой главе предложенные методы были применены при разработке методики восстановления неполных данных двигательной активности, а также методики декодирования наиболее вероятной траектории пользователя мобильной связи на транспортном графе. Следует отметить, что вторая методика была внедрена компанией ООО «Т2 Мобайл» в свою платформу анализа геоданных, что подтверждает практическую значимость и достоверность обоснованных в диссертации выводов и рекомендаций.

В третьей главе диссертационной работы рассматривается решение задачи распознавания неполных последовательностей, описываемых СММ. Также экспериментально было продемонстрировано преимущество разработанного подхода, в основе которого лежит идея маргинализации пропусков, в сравнении с другими методами. Следует отметить, что все эксперименты рас-

сма тривали как СММ с дискретным, так и с непрерывным распределением наблюдений.

Практическая полезность предлагаемого метода показана при разработке методики идентификации личности по неполным данным двигательной активности.

В четвёртой главе разрабатывается метод обучения СММ по неполным данным. Решение этой задачи также крайне важно, так как не всегда удаётся подобрать идеальные обучающие примеры для настройки параметров модели. Предлагаемый метод базируется на модифицированном с помощью идеи маргинализации, известном алгоритме Баума-Велша, позволяющим итеративно обучать СММ по набору последовательностей.

Вычислительные эксперименты, проведенные автором, показывают повышенную эффективность данного метода в сравнении с другими подходами. Также продемонстрировано решение усложненной практической задачи из предыдущей главы при разработке методики идентификации личности по неполным данным акселерометра, но с наличием пропусков не только в тестовой, но и в обучающей выборке. В основе данной методики лежит предлагаемый метод.

Пятая глава диссертационной работы посвящена решению задачи распознавания неполных последовательностей, но в условиях, когда конкурирующие случайный процесс, породившие последовательности, максимально близки.

Ранее для решения такой задачи (для целых последовательностей) другими исследователями был предложен алгоритм вычисления производных от логарифма функции правдоподобия того, что СММ породила последовательность, по различным параметрам модели. Затем вычисленные производные использовались в качестве векторов признаков для обучения классификатора, например, основанного на методе опорных векторов. Автором предложена модификация алгоритма вычисления производных, дополняющая его формулы на класс последовательностей с пропусками. Модификация также основана на идее маргинализации пропущенных наблюдений.

Эффективность разработанного метода также демонстрируется с помощью вычислительных экспериментов, состоящих в распознавании последовательностей, порожденных максимально схожими по параметрам СММ. Наблюдается также, что метод на основе производных превосходит метод на основе модифицированного алгоритма forward-backward, предложенный в третьей главе, в таких условиях параметрической близости. Также в данной главе описывается методика распознавания личности по неполным данным акселерометра, в основу которой лёг предлагаемый метод, что показывает его практическую применимость.

В заключении подведены итоги исследования, изложена его значимость для науки, обоснованы выводы, которые полностью соответствуют решению

задач, поставленных в диссертации, и доказательству защищаемых научных положений.

Значимость результатов исследования для науки заключается в том, что теория скрытых марковских моделей дополнена в области анализа неполных последовательностей. Теоретически установлены зависимости точности анализа неполных последовательностей от количества пропусков в них, что позволяет прогнозировать применимость алгоритмов СММ в разных условиях качества исходных данных. Выводы и результаты, полученные диссертантом, обоснованы и достоверны, так как опираются на существующую теоретическую базу и результаты вычислительных экспериментов.

Практическое значение результатов работы определяется тем, что они использованы при разработке методики анализа неполных данных двигательной активности, а также методики определения положения абонента в пространстве и времени по последовательности его событий в мобильной сети. Методика определения траектории абонента используется в компании ООО «Т2 Мобайл» и прошла апробацию на реальных данных сотового оператора.

Основные положения диссертации нашли отражение в публикациях автора, а также в докладах на российских и международных научно-практических конференциях. Авторство разработанных методов подтверждено свидетельством о государственной регистрации программы для ЭВМ, то есть результаты исследований известны широкому кругу научной общественности.

Содержание диссертации изложено в логически последовательной форме, приведен табличный, графический и иллюстративный материалы, текст оформлен в соответствии с требованиями ВАК, читается с интересом. Стиль изложения четкий и ясный, разделы диссертации раскрывают защищаемые научные положения. В конце каждого раздела сделаны содержательные выводы, облегчающие понимание диссертации. Имеются полные списки использованных терминов, сокращений и обозначений. Автореферат соответствует содержанию диссертации.

Рекомендации по использованию результатов и выводов диссертации. Необходимо расширить исследования для обоснования адаптации разработанных методов для других практических предметных областей.

Замечания.

1. Не указано, как длина распознаваемых последовательностей, размерность элементов, размер обучающей выборки, количество и параметры конкурирующих случайных процессов, описываемых СММ, влияют на точность распознавания последовательностей. Почему были выбраны именно такие параметры СММ для проведения экспериментов?
2. Автором сделано заключение, что предложенные им методы и алгоритмы СММ работают корректно, и это достаточно убедительно показано. Но есть и другие методы решения задач анализа последовательностей, например, рекуррентные нейронные сети. Важно уточнить, в чем преимущество реали-

зованных автором подходов, либо яснее очертить область применения этих методов.

3. Во введении и приложении Б упоминается разработанная автором зарегистрированная программа для ЭВМ, однако в тексте диссертации отсутствует её описание: язык программирования, операционная система, тип интерфейса, функциональные возможности программы и т. д.
4. В тексте диссертации приводится мало конкретных примеров практического применения методик анализа неполных данных двигательной активности и определения траектории абонента на графе дорог.

Заключение. В целом указанные замечания не влияют существенно на положительную оценку выполненной работы и основные выводы и рекомендации, обоснованные в диссертации.

Диссертация представляет собой завершённую научно-квалификационную работу по актуальной теме, в которой на основании выполненных автором исследований решена научная проблема, имеющая существенное значение для развития теории СММ и её практического применения. Выводы и рекомендации научно обоснованы. Работа отвечает требованиям ВАК, изложенным в п. 9 «Положения о присуждении ученых степеней» и предъявляемым к кандидатским диссертациям с точки зрения актуальности, новизны, научной и практической значимости полученных результатов, а её автор Уваров Вадим Евгеньевич заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.17 – Теоретические основы информатики.

Отзыв на диссертацию обсужден и одобрен на заседании семинара лаборатории № 69 в Федеральном государственном бюджетном учреждении науки Институт проблем управления им. В.А. Трапезникова Российской академии наук (протокол № 2 от 16.01.2020г.).

Заведующий лабораторией № 69
Федерального государственного бюджетного
учреждения науки Института проблем управления
им. В. А. Трапезникова Российской академии наук
д. т. н., проф.

Вишневский Владимир Миронович

ГСП-7, В-342, г. Москва, Профсоюзная, 65.
e-mail vishn@inbox.ru, тел. +7 (495) 334-75-91

ого В. ил.

*Отзыв поступил
в совет 5.02.2020*

С отзывом ознакомлен

06.02.2020 Уваров