ОТЗЫВ официального оппонента
о диссертации на соискание ученой степени кандидата физико-математических наук Казачук Марии Андреевны
на тему: «Динамическая аутентификация пользователей на основе анализа работы с клавиатурой компьютера»
по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

Актуальность выбранной темы

Массовое внедрение информационно-коммуникационных технологий (ИКТ) в различных областях (государственное и муниципальное управление, промышленность, банки, торговля и многих других) и их активное использование, характерное для последних лет, привносит в нашу повседневную жизнь не только удобства, но и угрозы. Так, по оценкам Комиссии ООН по предупреждению преступности и уголовному правосудию, потери мировой экономики от преступлений, совершаемых с помощью ИКТ составили 1,5 триллиона долларов в 2018 году и достигнут 2 триллионов долларов в 2019 году, что сравнимо с бюджетами многих государств. Разработка инструментов защиты от таких угроз составляет значительную часть современной отрасли ИКТ. Одной из важнейших задач здесь является аутентификация пользователей. В настоящее время используется много различных решений этой задачи от ввода пароля до биометрии, однако потребность в новых подходах и решениях несомненна. Именно такое решение - динамическая аутентификация пользователей на основе анализа их работы с клавиатурой компьютера и рассматривается в диссертационной работе Казачук М.А.

Отметим, что современные стандарты по информационной безопасности декларируют уникальность клавиатурного почерка пользователей и возможность создания систем аутентификации на его основе.

Наиболее актуальными и массовыми областями применения динамической аутентификации пользователей по клавиатурному почерку являются системы обнаружения внутренних и внешних вторжений, а также системы обеспечения безопасности пользователей в Интернете (например, при осуществлении web-платежей). Существующие в данной области решения обладают недостаточно высоким качеством распознавания: используемые в них признаковые пространства не являются стабильными по времени, также является актуальной проблема большой размерности. Отсутствие примеров данных клавиатурного почерка злоумышленников обуславливает трудность построения модели пользователя в рассматриваемой задаче. Также существующие решения не предлагают способа оценки степени аномальности поведения пользователя за сессию работы в целом. В своей диссертационной работе Казачук М.А. формулирует решение указанных проблем: ею предложены новый подход к подготовке данных клавиатурного почерка пользователей, способный строить информативные и стабильные по времени признаковые характеристики, а также новый метод одноклассовой классификации, подбор оптимальных значений метапараметров которого осуществляется собственно разработанным алгоритмом. Дополнительно, предложен новый метод получения агрегированной оценки степени аномальности поведения пользователя за длительный период его работы за компьютером, показавший более точную оценку по сравнению с результатами распознавания отдельных событий. Создан экспериментальный образец программного комплекса, использующий предложенный комплекс алгоритмов для обнаружения аномального поведения пользователей на основе анализа их работы с клавиатурой компьютера. По результатам экспериментов, использование разработанных алгоритмов позволило превзойти качество работы существующих решений.

Перечисленное выше определяет актуальность темы диссертационной работы и полученных автором результатов.

Содержание диссертации

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

Во введении обосновывается актуальность работы, ставится цель и формулируются задачи исследования, определяется научная новизна и основные положения работы, выносимые на защиту.

Первая глава посвящена аналитическому обзору существующих решений по данной тематике. По результатам обзора формируются направления дальнейших исследований.

Во второй главе проводятся исследование и разработка алгоритмов обработки данных клавиатурного почерка пользователей. Предлагается новый подход к построению признакового пространства, сокращение размерности пространства признаков в котором производится путем выделения наиболее стабильных из них, а также подход дискретизации признаков по квантилям, решающий проблему мультимодального распределения признаков. Предлагаемый подход отбора признаков по уровню их стабильности позволил улучшить качество аутентификации в среднем на 4%. Предлагаемый подход дискретизации признаков по квантилям в среднем на 6% превзошел используемую в существующих работах стандартизацию признаков.

В третьей главе проводится исследование и разработка методов построения модели пользователя, способных превзойти качество работы современных методов машинного обучения в данной задаче. В рамках данного исследования автором предлагаются:

- Новый нечеткий метод поиска исключений в данных, строящий в пространстве высокой размерности эллиптические области с оптимальным центром для выявления аномалий;
- Новый метод подбора оптимальных значений метапараметров одноклассовых алгоритмов классификации, позволяющий строить стабильные к смене тестового набора данных модели без использования информации о специфике данных нелегитимного класса;

- Новый метод оценки степени аномальности набора последовательных событий на основе t-статистики Уэлша, позволяющий производить аутентификацию пользователя как за короткий, так и за длительный период (например, целую сессию) его работы за компьютером.

Четвертая глава посвящена разработке и реализации экспериментального образца программного комплекса, использующего предложенный комплекс алгоритмов для обнаружения аномального поведения пользователей на основе анализа динамики их работы с клавиатурой компьютера. Подробно описана архитектура разработанного решения, приведен пошаговый пример его использования. В ходе проведенного экспериментального исследования комбинация разработанных алгоритмов превзошла существующие решения и показала высокое качество работы как при классификации отдельных событий, так и при классификации целых сессий работы пользователей за компьютером.

В заключении приводятся основные результаты диссертационной работы и делаются выводы, на основе которых можно судить о достижении целей, сформулированных в ней.

Автореферат соответствует содержанию диссертации и полностью удовлетворяет требованиям Положения о порядке присуждения учёных степеней, так как отражает основные положения, результаты и выводы диссертации, научную новизну и практическую значимость результатов исследования, отражает структуру диссертации и личный вклад автора.

Научная новизна, обоснование, достоверность и практическая ценность полученных научных положений, выводов и рекомендаций

В работе получены следующие новые научные результаты:

1. Новый подход к сокращению размерности признакового пространства путем выделения наиболее стабильных признаков с использованием статистики Колмогорова-Смирнова;

2. Новый нечеткий метод поиска исключений в данных, строящий в пространстве высокой размерности эллиптические области с оптимальным центром для выявления аномалий, использующий разработанный автором метод подбора оптимальных значений метапараметров алгоритмов одноклассовой классификации, который позволяет строить стабильные к смене тестового набора данных модели без использования информации о данных нелегитимного класса;
3. Новый метод оценки степени аномальности набора событий на основе t-статистики Уэлша, сравнивающий средние значения аномальностей на валидационной выборке легитимного пользователя и тестовой выборке рассматриваемого пользователя.

Достоверность полученных результатов подтверждается их целостным, последовательным и подробным математическим обоснованием, корректным применением методов теории вероятностей, математической статистики и теории машинного обучения, большим количеством экспериментальных исследований, публикациями результатов в изданиях, рецензируемых Web Of Science, Scopus, RSCI, представлением результатов на шести международных и всероссийских научных конференциях, а также использованием полученных результатов в государственной научно-исследовательской работе «Разработка технологий биометрической идентификации пользователя по признакам, проявляющимся при использовании устройств ввода данных персональных ЭВМ».

Практическая ценность работы Казачук М.А. состоит в том, что результаты ее теоретических исследований составили основу для разработки экспериментальной системы для обнаружения аномального поведения пользователей по клавиатурному почерку. Данная система включает в себя модули сбора поведенческой информации, а также построения и применения индивидуальных моделей пользователей. Разработанные модули могут использоваться в перспективных системах аутентификации.

Диссертация Казачук Марии Андреевны является законченной научно-квалификационной работой, обладает актуальностью, научной новизной и практической значимостью. Предлагаемые теоретические положения и методы разработаны до практических методик, алгоритмов и программ.

Результаты диссертационной работы соответствуют паспорту специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей» (по физико-математическим наукам), а разработанные теоретические положения и полученные результаты имеют важное научное и прикладное значение.

Замечания по диссертации

Несмотря на полученные новые результаты и высокое качество выполненной соискателем работы, диссертация также имеет некоторые недостатки:

1. Существенным является использование расстояния Махаланобиса, однако выигрыш, получаемый при этом, никак не формулируется. Наверно, не просто доказать теорему, что данное расстояние дает нам всегда результат, хотя бы не хуже классических методов, однако провести сравнение качества работы алгоритмов при разных формализациях расстояния было бы интересно и убедительно.
2. Хотелось бы увидеть более подробное описание границ применимости разработанного решения: зависит ли качество работы предложенного подхода от смены времени суток, изменения состояния здоровья пользователя и типа используемого оборудования.
3. Не всегда корректно используются термины. Например, употребляется термин «оптимальные значения параметров» (с. 71, с. 77), однако доказательств оптимальности не приводится. Более правильно говорить о наилучших значениях. Термин «степень нечеткости» (с. 89) уже используется в научной литературе (степень нечеткости множества) и желательно его замена или уточнение.

4. Есть замечания по оформлению текста работы. В частности:

- Нет объяснений используемых на рис. 2 (с. 34) обозначений (X1, X2, Y1, Y2).

- «Также было установлено, что в собранных данных могут присутствовать непарные события нажатия или отжатия клавиш …» (с. 51), которые предлагается «удалять из дальнейшего рассмотрения» (там же). Хотелось бы понимать какова доля таких событий.

- На с. 53 читаем «Проведенный анализ показал…», «… как показали эксперименты…», однако ссылок на литературу или описания экспериментов не приводится. Та же ситуация на с. 54 «… в результате серии проведенных экспериментов было выявлено …»; с. 63 «По результатам предварительной серии экспериментов было выявлено…», с. 99 «Экспериментально было получено, что …».

- Разное форматирование абзацев (с. 60, первый и второй абзацы; с. 86, третий и четвертый абзацы; с. 110, второй и третий абзацы).

- Форматирование рисунков и таблиц (рис. 6, с. 63, 64; таблица 8, с. 71, с. 72).

- Фразы «… по сравнению с достигнутыми в соответствующих работах результатами …» (с. 76), «… рассмотренных в существующих научных работах …» (с. 78) даются без указания работ.

- Демонстрационные примеры (рис. 7 с. 84 и рис.8 с.85) не объясняются.

- На с. 90 используется обозначение $\langle \cdot, \cdot \rangle_C$, которое не расшифровывается в тексте.

- Факт того, что центр нечеткого кластера можно выразить через линейную комбинацию образов входных объектов (с. 89-90) можно оформить в виде утверждения.

- Фразы «… при ∀ начальном приближении…» (с. 94) не есть хороший стиль для текста диссертационной работы.

Заключение

Вместе с тем, указанные замечания не умаляют значимости диссертационного исследования. Диссертация отвечает требованиям, установленным Московским государственным университетом имени М.В. Ломоносова к работам подобного рода. Содержание диссертации соответствует паспорту специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей» (по физико-математическим наукам), а также критериям, определенным пп. 2.1–2.5 Положения о присуждении ученых степеней в Московском государственном университете имени М.В. Ломоносова, а также оформлена, согласно приложениям №5, 6 Положения о диссертационном совете Московского государственного университета имени М.В. Ломоносова.

Таким образом, соискатель Казачук Мария Андреевна заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Официальный оппонент:
доктор технических наук, профессор,
профессор кафедры Математической теории интеллектуальных систем
Механико-математического факультета
МГУ им. М.В. Ломоносова
РЫЖОВ Александр Павлович 11 декабря 2019 г.

Контактные данные:
Тел.: +7 495 939 4637, e-mail: ryjov@intsys.msu.ru
Специальность, по которой официальным оппонентом
защищена диссертация:
05.13.19 «Управление в социальных и экономических системах»

Адрес места работы:

119991, ГСП-1, г. Москва,

Ленинские горы, МГУ, д.1, Главное здание

Кафедра Математической теории интеллектуальных систем

Механико-математического факультета МГУ им. М.В. Ломоносова

Тел: 8(495)939-20-90; e-mail: office@mech.math.msu.su

Подпись сотрудника

Механико-математического факультета

МГУ им. М.В. Ломоносова

А.П. Рыжова удостоверяю:

Вед. спец. Морозова А.А. 17.12.2019 г.