

Экспериментальная оценка использования расширенного пространства признаков для детектирования автоматизированных средств подмены пользователя

Н. С. Афанасьева

Омский государственный университет путей сообщения

Ранее автором был предложен подход к формированию расширенного пространства признаков, позволяющих осуществлять детектирование автоматизированных средств подмены пользователя в информационной среде на основании передвижения компьютерной мыши. Рассматривался базовый набор, применяемый в известных работах по поведенческой биометрии мыши, состоящий из 56 параметров, который был дополнен 166 новыми признаками. Результирующее комбинированное пространство образовано 222 признаками. Целью настоящей работы является эмпирическая проверка гипотезы о том, что предложенное расширенное пространство признаков обеспечивает более высокое качество детектирования ботов по сравнению с базовым набором признаков. На данных, включающих датасеты BALABIT, BOGAZICI и внутренние наборы синтетических и браузерных ботов, обучен классификатор Random Forest с оптимизацией гиперпараметров методом GridSearchCV. Получены значения точности (accuracy) 0,99–0,999 и F1-меры по классу «бот» до 0,945–0,983. Дополнительно проведен контролируемый вычислительный эксперимент, в котором сопоставлены базовый набор признаков и комбинированное пространство. Переход к комбинированному пространству признаков сопровождается ростом полноты (recall) по классу «бот» с 0,800 до 0,912 и F1-меры с 0,887 до 0,945. Анализ важности признаков показал, что подавляющее большинство наиболее значимых для классификации характеристик относится к предложенным новым признакам, а не к базовому набору. Проведено сопоставление с результатами известных работ в области поведенческой биометрии. Результаты подтверждают целесообразность расширения признакового пространства для повышения точности детектирования автоматизированных средств подмены пользователя

Бот; вредоносный бот; пространство признаков; детектор ботов; поведенческая биометрия.

ВВЕДЕНИЕ

Детектирование автоматизированных средств подмены пользователя (ботов) по показателям поведенческой биометрии движения компьютерной мыши остается востребованным направлением защиты информационных систем, поскольку не требует дополнительного аппаратного обеспечения и характеризуется устойчивостью к прямой подделке со стороны злоумышленника. Ранее автором был проведен аналитический обзор существующих методов детектирования бота по движению мыши и общедоступных датасетов, используемых для решения данной задачи [Афа24]. На предшествующем этапе настоящего исследования [Афа26] была решена задача формирования признакового пространства, пригодного для указанной цели. Базовый набор из 56 характеристик, регулярно применяемых в известных научных работах по поведенческой биометрии мыши (координатные и временные признаки, скорость, ускорение, рывок, угол, угловая скорость, кривизна и производные от них статистики – среднее, стандартное отклонение, минимум, максимум) [Gam04, Zhe11, Feh12, Chu14], был дополнен 166 новыми признаками, объединенными в четыре содержательные группы.

Афанасьева Н. С. Экспериментальная оценка использования расширенного пространства признаков для детектирования автоматизированных средств подмены пользователя // СИИТ. 2026. Т. 8, № 5(29). С. 113-126. DOI: 10.54708/SIIT-2026-no5-p113. EDN: OCJLVK.

Afanaseva N. S. "Experimental evaluation of an extended feature space for detecting automated user substitution tools" // SIIT. 2026. Vol. 8, no. 5(29), pp. 113-126. DOI: 10.54708/SIIT-2026-no5-p113. EDN: OCJLVK. (In Russian).

Первая группа признаков основана на порядковых статистиках – квантилях (0,25; 0,5; 0,75), медианах и межквартильном размахе для основных кинематических параметров. Также данная группа включает признаки, отслеживающие аномальные скачкообразные смещения курсора и распределение абсолютных приращений координат. Вторая группа оценивает меру близости распределения параметров конкретного трека к распределениям, наблюдаемым в эталонных датасетах реальных пользователей BALABIT [Ant20], Bogazici [Kil21] и датасете синтетических ботов N_BOT. Третья группа основана на критерии хи-квадрат – статистике согласия распределения параметров трека с эталонным распределением, бинарном индикаторе статистического сходства, а также включает энтропийные показатели, в том числе отклонение энтропии от максимально возможного значения, характеризующее равномерность распределения перемещений. Четвертая группа образована угловыми характеристиками – долями углов, попадающих в узкие окрестности характерных значений, и показателями повторяемости координат вдоль траектории.

Результирующее комбинированное пространство образовано 222 признаками. В заключении предшествующего этапа исследования был сформулирован план эмпирической проверки, предусматривающий сравнение точности классификации при использовании базового и комбинированного наборов признаков с применением ансамблевых алгоритмов машинного обучения и оценку вклада каждого нового признака методом SHAP. Настоящая статья представляет результаты указанной проверки, полученные на эмпирическом материале, включающем данные о движении мыши реальных пользователей и различных категорий ботов.

Задачи исследования:

1. Провести классификацию треков движения мыши (человек / бот) с использованием комбинированного пространства признаков и ансамблевого классификатора Random Forest, оптимизированного методом GridSearchCV.
2. Провести контролируемый вычислительный эксперимент, сопоставляющий качество классификации при использовании исключительно базового набора признаков и комбинированного набора при идентичной выборке и идентичном протоколе обучения.
3. Оценить вклад каждой группы признаков в решение классификатора на основании встроенной важности признаков (feature importances) и метода SHAP.
4. Сопоставить полученные результаты и характеристики признакового пространства с известными работами по поведенческой биометрии мыши.

ИЗВЕСТНЫЕ ИССЛЕДОВАНИЯ В ОБЛАСТИ ПОВЕДЕНЧЕСКОЙ БИОМЕТРИИ, ОСНОВАННЫЕ НА ДВИЖЕНИИ МЫШИ

Анализ существующих исследований в области поведенческой биометрии позволяет сделать вывод об использовании только классических кинематических параметров (скорость, ускорение, кривизна, рывок и т. д.), не позволяющих полноценно описать вариативную природу человеческой моторики. В табл. 1 обобщены результаты и признаковые пространства наиболее значимых работ в данной области и представлены результаты, полученные в настоящем исследовании.

Следует отметить методологическое ограничение прямого сопоставления. Большинство приведенных работ решают задачу верификации, либо идентификации личности пользователя, по движению мыши, для которой типичной метрикой качества выступает равный уровень ошибок (EER) или площадь под ROC-кривой (AUC), тогда как настоящее исследование решает задачу бинарной классификации «человек – бот», для которой применяются метрики ассурасу, precision, recall и F1-мера. Указанные задачи различны по постановке и метрики оценки напрямую не сопоставимы, поэтому сравнение также учитывает глубину признакового пространства и соответствующее ему качество разделения классов.

Как показано в табл. 1, ранние работы (2004–2014 гг.) оперируют признаковым пространством существенно меньшей мощности по сравнению с предложенным комбинированным набором. Более поздние работы (2019–2024 гг.) демонстрируют иную тенденцию. Переход

к сквозным (end-to-end) архитектурам глубокого обучения (сверточные и рекуррентные нейронные сети), обрабатывающим сырую последовательность координат либо ее визуальное представление напрямую, без этапа ручного конструирования признаков [Liu25]. Такие методы достигают высоких значений ассигасы, однако решают задачу за счет автоматического извлечения признаков внутри нейронной сети, что затрудняет содержательную интерпретацию результата и требует существенно большего объема размеченных данных и вычислительных ресурсов для обучения по сравнению с ансамблевыми методами на основе заранее сконструированных признаков [Shi23].

Таблица 1

**Сопоставление результатов известных работ в области поведенческой биометрии
с результатами настоящего исследования**

Источник	Задача, метрика	Число признаков	Метод	Результат
Gamboa, Fred, 2004 [Gam04]	Верификация личности, EER	20 (скорость, ускорение, угол, рывок)	Статистическая модель	EER = 0,7% (после 100 штрихов)
Zheng, Paloski, Wang, 2011 [Zhe11]	Верификация личности, EER	20 (угловые метрики)	SVM	EER = 1,3% (после 20 кликов)
Feher et al., 2012 [Feh12]	Верификация личности (по действиям)	15–20 (по типу действия)	Random Forest (по типам действий)	Высокая точность верификации, без единого сводного показателя
Chuda, Kratky, 2014 [Chu14]	Идентификация по микроколебаниям курсора	10–15	Статистическая модель	Подтверждена информативность микроколебаний
Wei, Zhao, Cai, 2019 [Wei19]	Детектирование бота	Не заданы вручную (end-to-end подход)	CNN	Обнаружение в три раза больше ботов по сравнению с традиционными методами
Acien et al., 2022 (BeCAPTCHA) [Aci22]	Детектирование бота	Нейромоторные (Sigma-Lognormal)	SVM/ RF/ KNN/ MLP/ LSTM/ GRU	Accuracy 93%
Niu et al., 2024 [Niu24]	Детектирование бота	Не заданы вручную (end-to-end подход)	Lightweight (облегченная нейросеть)	Accuracy 99,78%
Настоящее исследование (базовый набор)	Детектирование бота, ассигасы/F1-мера	56	Random Forest	Accuracy 0,991; F1-мера(бот) 0,887 (контролируемый эксперимент)
Настоящее исследование (комбинированный набор)	Детектирование бота, ассигасы/F1-мера	222 (56 + 166)	Random Forest + GridSearchCV	Accuracy 0,99–0,999; F1-мера(бот) 0,945–0,983

Предлагаемый подход позволяет достичь точности того же уровня, что и современные нейросетевые методы, хотя прямое сравнение затруднено различием данных и постановки задач. Его преимущество заключается в интерпретируемости. Решение классификатора объясняется через важность конкретных признаков и явные пороговые правила, для моделей глубокого обучения это требует применения дополнительных методов апостериорного объяснения.

Источники данных, признаковое пространство и алгоритм классификации

Для экспериментальной проверки использовались следующие датасеты движений мыши, агрегированные в виде характеристических векторов для треков длительностью около трех секунд:

- BALABIT и BOGAZICI – общедоступные датасеты движений мыши реальных пользователей, класс «человек»;
- N_BOT – датасет синтетически сгенерированных ботов, класс «бот»;
- N_BROWSER – реальные браузерные сессии пользователей, класс «человек»;
- N_BROWSER_E и N_BROWSER_SCRAPERS_E – автоматизированные (скриптовые) и скрейперские сессии в браузере, класс «бот».

Итоговые объемы выборок варьировались по версиям вычислительного пайплайна от порядка 30 тысяч до более одного миллиона наблюдений, при выраженном дисбалансе классов в пользу «человек», компенсируемом параметром `class_weight = 'balanced'` при обучении классификатора. Для контролируемого эксперимента использовалась выборка из 53943 треков, сформированная на основе датасетов BALABIT, BOGAZICI и N_BOT (924 трека класса «бот», 53019 треков класса «человек»).

Для каждого трека рассчитывался характеристический вектор, включающий как базовые, так и предложенные новые признаки. Из итогового комбинированного пространства признаков дополнительно исключались неинвариантные к источнику данных характеристики, значения которых зависели не от того, человек это или бот, а от технических особенностей конкретного источника данных, такие как абсолютные значения скорости, ускорения и рывка. Для этого распределения каждого признака сравнивались между двумя датасетами реальных пользователей (BALABIT и BOGAZICI) с помощью статистических критериев Манна–Уитни, Крускала–Уоллиса, медианного теста и критерия хи-квадрат [Guy03]. Признак исключался, если его распределение значимо различалось между этими датасетами, несмотря на то что оба содержат данные реальных людей.

Базовые признаки описывают каждый кинематический параметр трека (скорость, ускорение, рывок, угол, угловую скорость, кривизну) средним значением, стандартным отклонением и экстремумами, то есть предполагают, что распределение параметра внутри трека полностью характеризуется первыми двумя моментами. Для движения руки человека это предположение не выполняется. Согласно модели минимизации рывка [Fla85] и кинематической теории быстрых движений [Pla95], целенаправленное движение человека складывается из последовательности перекрывающихся субдвижений с колоколообразными (логнормальными) профилями скорости, дополненных корректирующими микродвижениями и физиологическим тремором. Вследствие этого распределения скорости, ускорения и особенно рывка у человека асимметричны, имеют широкую центральную часть и тяжелые хвосты. Программно сгенерированное движение строится интерполяцией между опорными точками (линейной, кривыми Безье) с постоянным шагом по времени или пространству, иногда с добавлением синтетического шума. Распределения такого типа движения либо вырождены (центральная часть стянута к одному значению), либо имеют форму, заданную генератором. Поэтому информативным является не столько среднее значение параметра, сколько форма его распределения внутри трека. Каждая из четырех групп новых признаков описывает один из аспектов этой формы (табл. 2), формальные определения признаков приведены в [Афа26].

Порядковые статистики (квантили уровней 0,25; 0,5; 0,75 и межквартильный размах кинематических параметров и абсолютных приращений координат) описывают положение и ширину центральной части распределения. В отличие от среднего и стандартного отклонения они устойчивы к единичным выбросам, возникающим из-за дискретизации меток времени (малые интервалы между событиями приводят к аномально большим значениям ускорения и рывка). У человека межквартильный размах рывка и ускорения велик вследствие непрерывной коррекции движения, тогда как у траектории, полученной интерполяцией, центральная часть распределения рывка стягивается к нулю. К этой же группе отнесена доля скачкообразных смещений курсора – шагов, длина которых аномально велика по отношению к остальным шагам трека, что характерно для программной установки координат.

Таблица 2

Состав и обоснование групп новых признаков

Группа	Вычисляемые параметры	Обоснование различия «человек – бот», на котором основан признак
Порядковые статистики	Квантили 0,25; 0,5; 0,75 и межквартильный размах кинематических параметров и приращений координат; доля скачкообразных смещений	Широкая центральная часть распределений у человека (коррекционные субдвижения) и вырожденная – у интерполированных траекторий. Скачки координат у ботов
Близость к эталонам	Доля значений параметра трека в межквартильном и расширенном интервалах эталонов BALABIT, BOGAZICI, N_BOT	Доля точек трека в типичном диапазоне человека или бота
Согласие и энтропия	Статистика хи-квадрат и индикатор согласия с эталоном; энтропия и ее отклонение от максимума	Форма распределения. При постоянном шаге – вырожденная, при синтетическом шуме – избыточная равномерность
Углы и повторяемость	Доли направлений около 0°, 45°, 90°, 135°, 180°; доля повторяющихся координат	Направления, кратные 45°, при движении по прямой; повторы координат при таймерной генерации

Признаки близости к эталонным распределениям определяют долю значений параметра трека, попадающих в типичный интервал в эталонном датасете (межквартильный интервал и интервал, расширенный на полтора межквартильных размаха в соответствии с правилом Тьюки). В качестве эталонов использованы датасеты реальных пользователей BALABIT и BOGAZICI и датасет синтетических ботов N_BOT. Такой признак интерпретируется как вероятность того, что случайно выбранная точка трека типична для эталонного класса. То есть для трека, порожденного тем же механизмом, что и эталон, доля попадания в межквартильный интервал близка к 0,5, а в расширенный интервал – к единице, тогда как траектории иной природы дают смещенные значения. Использование эталонов обоих классов дает симметричные меры сходства с человеком и с ботом.

Признаки согласия и энтропийные признаки сравнивают форму гистограммы параметра трека с эталонной гистограммой. Статистика критерия хи-квадрат количественно характеризует отличие формы распределения от эталонной, а бинарный индикатор фиксирует, отвергается ли гипотеза о согласии на заданном уровне значимости. Энтропия гистограммы и ее отклонение от максимально возможного значения характеризуют равномерность распределения перемещений. У траекторий с постоянным шагом значения сосредоточены в одном-двух интервалах гистограммы и энтропия мала, у траекторий с равномерно распределенным синтетическим шумом энтропия близка к максимальной, тогда как движение человека занимает промежуточное положение. Таким образом, энтропийные признаки выявляют оба типичных способа генерации траекторий ботов.

Угловые признаки и показатели повторяемости координат определяют долю направлений движения, попадающих в узкие окрестности значений, и долю событий, повторно регистрирующих уже посещенную точку. Движение, сформированное программно по прямой между опорными точками, на пиксельной сетке дает направления, кратные 45° (горизонталь, вертикаль, диагональ), и соответствующие доли близки к единице. У человека распределение направлений непрерывно и такие доли малы. Повторы координат типичны для генераторов событий, работающих по фиксированному таймеру.

В качестве базового классификатора использован алгоритм Random Forest [Bre01] из библиотеки scikit-learn. Выборка делилась на обучающую и тестовую части в соотношении 67% к 33%, при этом стратификация по классу обеспечивала одинаковую долю ботов и пользователей в обеих частях. Гиперпараметры модели (criterion, max_depth, max_features,

`min_samples_leaf`, `min_samples_split`, `n_estimators`) подбирались методом GridSearchCV с пятикратной кросс-валидацией. Для интерпретации вклада отдельных признаков в решение модели дополнительно применялся метод SHAP (SHapley Additive exPlanations) [Lun17]. Дополнительно из деревьев ансамбля и отдельного дерева решений извлечен набор интерпретируемых экспертных правил (majority vote), позволяющий оценить вклад отдельных признаков в объяснимой форме.

Выбор Random Forest (случайного леса) обусловлен следующими причинами. Во-первых, предметом исследования является признаковое пространство, а не классификатор, поэтому классификатор выступает измерительным инструментом и должен давать результат, минимально зависящий от настройки. Random Forest малочувствителен к гиперпараметрам, устойчив к неинформативным и коррелированным признакам, которых в пространстве из 222 признаков заведомо много, и не требует нормализации признаков, имеющих различные единицы измерения [Bre01]. Благодаря этому различие результатов на базовом и комбинированном наборах может быть отнесено на счет признаков, а не различий в настройке модели. Во-вторых, для табличных данных с разнородными признаками ансамбли деревьев решений остаются наиболее сильным классом моделей. Систематические сравнения на разных наборах данных показывают, что они превосходят глубокие нейронные сети, включая специализированные табличные архитектуры [Gri22, Shw22]. В-третьих, деревья решений допускают точное вычисление значений SHAP и извлечение явных правил, что необходимо для содержательной проверки вклада новых признаков.

Сравнительная характеристика случайного леса и альтернативных классов моделей приведена в табл. 3.

Таблица 3

Сравнительная характеристика классов моделей для детектирования ботов

Критерий	Random Forest	Градиентный бустинг	MLP на признаках	End-to-end CNN / RNN
Входные данные	Признаки	Признаки	Нормализованные признаки	Сырые последовательности
Чувствительность к гиперпараметрам	Низкая	Высокая	Высокая	Высокая
Требования к объему размеченных данных	Умеренные	Умеренные	Повышенные	Высокие
Качество на признаковых (табличных) данных	Высокое	Высокое, часто наилучшее	Ниже ансамблей деревьев	Не применимо (работает с сырыми последовательностями)
Интерпретируемость	Важность признаков, точный SHAP, явные правила	Важность признаков, точный SHAP	Только апостериорные методы	Только апостериорные методы
Возможность оценить вклад отдельных признаков	Да	Да	Ограниченно	Нет
Вычислительные затраты на обучение	Низкие	Средние	Выше среднего	Высокие

Градиентный бустинг над деревьями решений (XGBoost [Che16], LightGBM [Ke17]) на табличных данных нередко обеспечивает несколько более высокое качество, чем Random Forest, однако заметно чувствительнее к выбору гиперпараметров (темп обучения, число и глубина

деревьев, регуляризация) и склонен к переобучению на малочисленном классе при сильном дисбалансе. Поскольку в данной работе обучение проводилось на выборке, содержащей менее тысячи треков класса «бот», это повышает риск того, что различие результатов на двух наборах признаков будет отражать различие в настройке модели, а не информативность признаков. Нейронные сети на табличных данных требуют нормализации признаков. Сквозные (end-to-end) сверточные и рекуррентные модели исключают этап конструирования признаков и потому не позволяют ответить на исследовательский вопрос настоящей работы о вкладе конкретных признаков. Также они требуют большего объема размеченных данных и вычислительных ресурсов.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТА

Качество классификации на комбинированном пространстве признаков

В табл. 4 представлена динамика метрик качества классификации по мере доработки признакового пространства и очистки данных в ходе итеративной разработки вычислительного пайплайна.

Таблица 4

Динамика качества классификации по мере доработки признакового пространства

Версия пайплайна	Accuracy	Precision (бот)	Recall (бот)	F1-мера (бот)
Ранняя версия (I)	0,952	0,670	0,135	0,225
Ранняя версия (II)	0,980	0,973	0,695	0,811
Промежуточная версия	0,991	0,983	0,865	0,920
Поздние версии (доработанные)	0,999	0,982	0,912	0,945
Расширенная выборка (N = 46310)	0,998	0,992	0,974	0,983
Полная объединенная выборка (X_all, N = 1114235)	1,000	0,988	0,892	0,937

Итоговое качество классификации на комбинированном признаковом пространстве для доработанных версий пайплайна составило accuracy 0,998–1,000, precision(бот) 0,982–0,992, recall (бот) 0,865–0,974, F1-мера (бот) 0,920–0,983.

Наилучшие показатели по F1-мера (бот) = 0,983 достигнуты на расширенной выборке объемом 46310 наблюдений. По мере расширения и уточнения признакового пространства в ходе разработки наблюдался существенный рост recall и F1-меры по классу «бот» (с 0,135 и 0,225 на ранних версиях пайплайна до 0,912 и 0,945 на поздних соответственно).

«Промежуточная версия» и «Поздние версии (доработанные)» протестированы на выборке одинакового объема и отличаются исключительно составом и качеством признакового пространства, тогда как два последних эксперимента соответствуют тому же итоговому пайплайну, проверенному на существенно больших выборках. Расширенная выборка показывает наилучшее достигнутое качество, а полная объединенная выборка результат на максимально широком и разнородном по составу источников ботов тестовом материале. Умеренное снижение recall(бот) (с 0,974 до 0,892) и F1-меры(бот) (с 0,983 до 0,937) на выборке X_all (1114235 наблюдений) при сохранении высокого precision(бот) (0,988) указывает на то, что модель реже ошибается на реальных пользователях, но несколько хуже обобщается на более разнообразные типы ботов по мере роста и усложнения тестового материала.

Приведенная динамика носит характер внутреннего наблюдения за ходом итеративной разработки и отражает совокупный эффект уточнения признакового пространства и очистки данных.

На рис. 1 представлена матрица ошибок на тестовой выборке (X_{test}). Классификатор корректно распознал подавляющее большинство наблюдений обоих классов. Основная масса ошибок приходится на пропуск отдельных треков ботов (ложноотрицательные срабатывания), тогда как ложные срабатывания на треках реальных пользователей единичны, что согласуется с высоким значением $precision(бот)$.

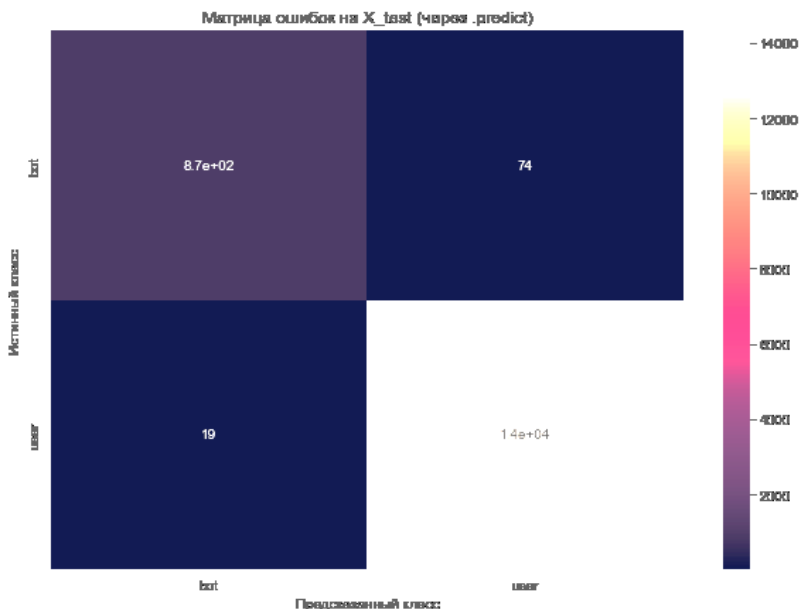


Рис. 1. Матрица ошибок классификатора Random Forest на тестовой выборке

На рис. 2 показана ROC-кривая по классу «бот», построенная отдельно для обучающей и тестовой выборок. Обе кривые проходят близко к верхнему левому углу графика, что соответствует высокой разделяющей способности классификатора; площадь под кривой (AUC) для обеих выборок практически не отличается, что указывает на отсутствие выраженного переобучения модели при подобранных методом GridSearchCV гиперпараметрах.

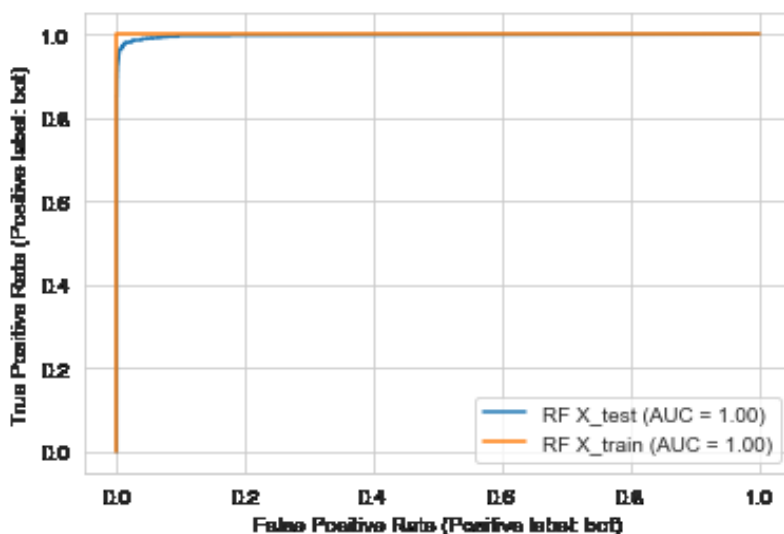


Рис. 2. ROC-кривая классификатора по классу «бот» на обучающей и тестовой выборках

Контролируемый эксперимент с использованием базового и комбинированного наборов признаков

Для получения количественной оценки прироста качества классификации, обусловленного непосредственно расширением признакового пространства, был проведен дополнительный контролируемый вычислительный эксперимент. В отличие от результатов, представленных в предыдущем подразделе, где версии пайплайна различались как составом признаков, так и составом обучающей выборки, настоящий эксперимент обеспечивает идентичность выборки, целевой переменной, алгоритма классификации и протокола разбиения данных, варьируя исключительно состав признакового пространства.

В качестве данных использовалась выборка из 53943 треков, которая включала в себя 924 наблюдения класса «бот» и 53019 наблюдений класса «человек». В качестве базового набора использовалось множество из 56 признаков, включающее среднее значение, стандартное отклонение, минимум и максимум для восьми кинематических параметров (скорость по осям X и Y, модуль скорости, ускорение, рывок, угол, угловая скорость, кривизна), а также скалярные признаки геометрии трека (координаты начала и конца, длительность, длина траектории, смещение, прямолинейность, число точек, суммарный угол поворота, число резких изменений направления, максимальное смещение, продолжительность первого сегмента с положительным ускорением).

Для обеих конфигураций использовался классификатор Random Forest из 100 деревьев с одинаковыми гиперпараметрами и одинаковым стратифицированным разбиением выборки (66,99 % треков для обучения и 33,01 % для тестирования). Таким образом, конфигурации различаются исключительно составом признаков. Результаты на тестовой выборке представлены в табл. 5.

Таблица 5

Результаты контролируемого эксперимента

Конфигурация признаков	Accuracy	Precision (бот)	Recall (бот)	F1-мера (бот)
Базовый набор	0,991	0,980	0,800	0,887
Комбинированный набор	0,999	0,982	0,912	0,945

Показатели комбинированного набора соответствуют воспроизведенной конфигурации пайплайна на сопоставимой по стратификации тестовой выборке.

Как показано в табл. 5, переход от базового набора признаков к комбинированному пространству сопровождается ростом полноты (recall) по классу «бот» с 0,800 до 0,912, то есть относительным сокращением доли пропущенных ботов, и ростом F1-меры с 0,887 до 0,945. При этом точность (precision) по классу «бот» на базовом наборе признаков оказывается сопоставимой с комбинированным набором. Данное наблюдение подтверждает ожидаемый эффект расширения признакового пространства — дополнительные признаки повышают чувствительность классификатора к более широкому спектру паттернов автоматизированного поведения (рост recall), одновременно с этим оставляя на одном уровне число ложноположительных срабатываний на объектах класса «человек» (сохранение precision). С точки зрения задачи детектирования пропуск бота (ложноотрицательная ошибка) представляет собой более критичный тип ошибки по сравнению с ложным срабатыванием на реальном пользователе, отмеченное изменение баланса precision и recall представляет собой практически значимый эффект.

Оценка вклада новых признаков в классификацию

Анализ важности признаков обученного классификатора Random Forest (встроенная метрика feature_importances) показал, что среди 20 наиболее значимых для разделения классов признаков более половины относится к предложенному расширенному набору. На рис. 3 представлена диаграмма важности 20 наиболее значимых признаков.

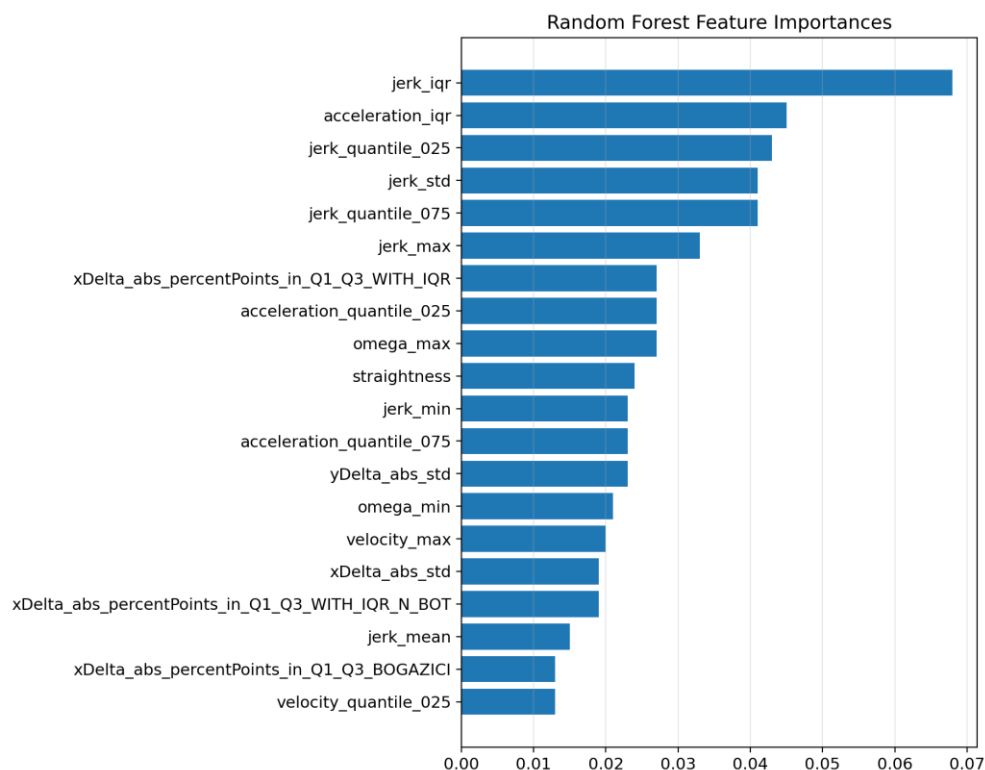


Рис. 3. Важность 20 наиболее значимых признаков по встроенной метрике Random Forest (feature_importances)

Как следует из рис. 3, наиболее значимым отдельным признаком является межквартильный размах рывка (`jerk_iqr`), являющийся новым признаком. Далее располагаются аналогичный размах ускорения (`acceleration_iqr`) и квантили рывка (`jerk_quantile_025`, `jerk_quantile_075`). Среди приведенных признаков новые характеристики представлены прежде всего порядковыми статистиками: межквартильными размахами и квантилями рывка, ускорения и скорости (`jerk_iqr`, `acceleration_iqr`, `jerk_quantile_025/075`, `acceleration_quantile_025/075`, `velocity_quantile_025`), а также признаками сравнения с эталонными распределениями (`xDelta_abs_percentPoints_in_Q1_Q3_WITH_IQR`, `xDelta_abs_percentPoints_in_Q1_Q3_WITH_IQR_N_BOT`, `xDelta_abs_percentPoints_in_Q1_Q3_BOGAZICI`).

Из базового набора в число наиболее значимых стабильно входят стандартное отклонение и экстремумы рывка, ускорения, угловой скорости и скорости (`jerk_std/max/min/mean`, `omega_max/min`, `velocity_max`), стандартное отклонение приращений координат по оси Y (`yDelta_abs_std`) и приращений по оси X (`xDelta_abs_std`), а также прямолинейность трека (`straightness`).

Приведенное распределение важности признаков служит альтернативным эмпирическим подтверждением результата, полученного в контролируемом эксперименте. Предложенные новые характеристики, основанные на сопоставлении с эталонными распределениями (критерий хи-квадрат, попадание в межквартильный интервал), а также на энтропийных и порядковых статистиках, вносят значимый вклад при отделении траекторий ботов от траекторий реальных пользователей, тогда как базовые кинематические характеристики выполняют вспомогательную роль, формируя систему координат, относительно которой новые признаки задают более тонкое описание движения.

Для проверки устойчивости приведенного ранжирования дополнительно применен метод SHAP (SHapley Additive exPlanations, TreeExplainer). На рис. 4 представлена диаграмма средних абсолютных значений SHAP для 16 наиболее значимых признаков, вклад которых показан отдельно для каждого из двух классов.

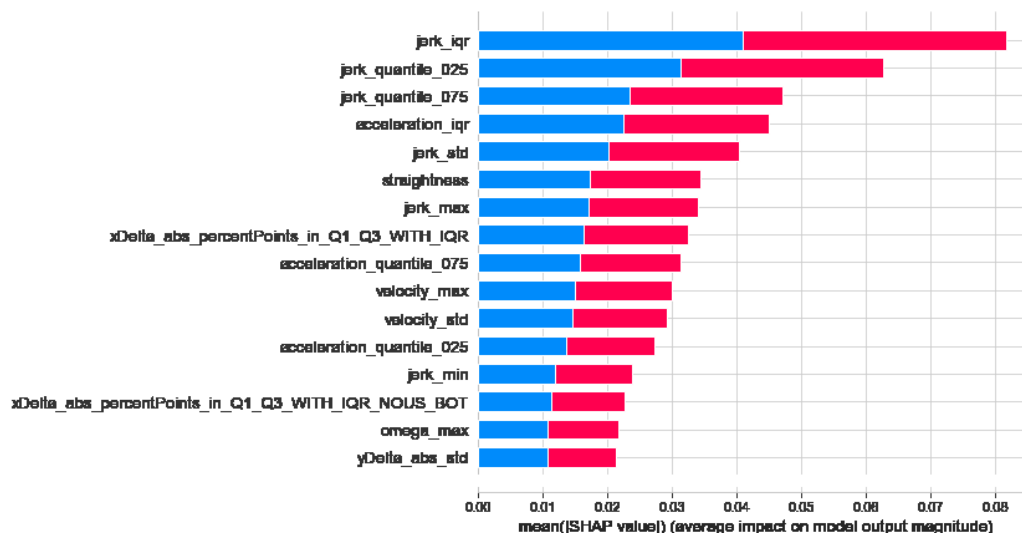


Рис. 4. Средние абсолютные значения SHAP для 16 наиболее значимых признаков, TreeExplainer

Полученные результаты практически полностью совпадают с ранжированием по встроенной метрике `feature_importances`. Верхние позиции отведены тем же признакам (`jerk_iqr`, `jerk_quantile_025`, `jerk_quantile_075`, `acceleration_iqr`, `jerk_std`, `straightness`, `jerk_max`), с незначительными перестановками порядка внутри группы.

Совпадение результатов двух способов оценки важности признаков `feature_importances` и SHAP снижает вероятность того, что наблюдаемое распределение важности признаков является артефактом конкретного метода оценки.

Интерпретируемые экспертные правила

Дополнительным подтверждением ценности новых признаков служит анализ интерпретируемых экспертных правил, извлеченных из деревьев Random Forest (30 деревьев глубины 5). Каждое правило представляет собой конъюнкцию пороговых условий вдоль пути от корня дерева решений к листу, относящему трек к классу «бот». Лучшее из проверенных правил (10-е правило, извлеченное из 18-го дерева ансамбля) на выборке из 2 182 175 наблюдений, содержащей 9 859 треков класса «бот», показало precision 96,0 %, recall 91,0 % и F1-меру 93,3 % по классу «бот». Условия правила приведены в табл. 6.

Таблица 6

Условия лучшего экспертного правила

№	Признак	Условие	Тип признака	Интерпретация
1	xDelta_abs_quantile_075	$> 16,875$	Новый (порядковая статистика)	75-й процентиль абсолютного смещения по оси X выше порога
2	jerk_iqr	$\leq 0,001$	Новый (порядковая статистика)	Межквартильный размах рывка мал (почти нулевая вариативность рывка)
3	curvature_std	$\leq 0,402$	Базовый	Стандартное отклонение кривизны траектории невелико
4	jerk_mean	$\leq 0,0$	Базовый	Средний рывок непожителен
5	chi2_xDelta_abs_with_BALABIT	$> 18,876$	Новый (сравнение с эталонным распределением)	Распределение приращений по X статистически значительно отличается от BALABIT (критерий хи-квадрат)

Трек относится к классу «бот», если одновременно выполняются все пять условий. Три из пяти условий опираются на признаки из нового набора (75-й перцентиль абсолютных приращений по оси X , межквартильный размах рывка и статистика хи-квадрат при сравнении с эталонным распределением BALABIT), что согласуется с результатами анализа важности признаков. Правило имеет содержательную интерпретацию – почти нулевой разброс рывка при крупных шагах по оси X и распределении приращений, статистически значимо отличающемся от распределения реальных пользователей, характерен для программной генерации движения.

ЗАКЛЮЧЕНИЕ

Совокупность полученных результатов формирует доказательную базу, подтверждающую исходную гипотезу исследования. Контролируемый эксперимент, проведенный на идентичной выборке и при идентичном протоколе обучения, демонстрирует прирост recall по классу «бот» при переходе от базового к комбинированному признаковому пространству. Указанный эффект также подтверждается анализом важности признаков, демонстрирующим доминирование новых признаков среди наиболее значимых для классификации, а также качеством отдельных экспертных правил, опирающихся на новые признаки.

Величину наблюдаемого прироста качества следует охарактеризовать как умеренную, но практически значимую. Расширение признакового пространства не приводит к кратному изменению метрик, поскольку базовый набор кинематических признаков уже обеспечивает высокое качество классификации на рассмотренной выборке. Эффект расширения признакового пространства проявляется прежде всего в снижении доли пропущенных ботов (рост recall), что представляет практический интерес для прикладных систем защиты, в которых цена ложноотрицательной ошибки, как правило, превышает цену ложноположительной.

Сопоставление с известными работами показывает две разнонаправленные тенденции в развитии предметной области. Ранние работы оперируют признаковым пространством на порядок меньшей мощности по сравнению с предложенным комбинированным набором. Более поздние работы реализуются в противоположном направлении – отказ от ручного конструирования признаков в пользу end-to-end глубокого обучения, достигая заявленной точности ценой потери интерпретируемости решения. Предложенный в настоящей работе подход занимает промежуточную позицию. Расширенное признаковое пространство остается априорно заданным и интерпретируемым, обеспечивая высокое качество детектирования и в отличие от end-to-end моделей сохраняя возможность оценить вклад отдельных признаков и обосновать принятое решение, что востребовано в прикладных системах информационной безопасности при разборе инцидентов и обосновании принятых решений.

Прямое количественное сопоставление с end-to-end методами и градиентным бустингом на единых данных и по единым метрикам составляет направление дальнейших исследований.

Исследование подтверждает практическую ценность предложенного расширенного признакового пространства и обосновывает целесообразность его дальнейшего применения в качестве основы для построения производственных систем защиты от автоматизированных средств подмены пользователя.

БЛАГОДАРНОСТИ И ПОДДЕРЖКА

Работа выполнена в рамках государственного задания на оказание государственных услуг (выполнение работ) № 109-03-2026-004 от 16.01.2026 г. по теме «Разработка технологии выявления автоматизированных средств подмены пользователя в информационной среде с помощью алгоритмов моторной идентификации поведенческих паттернов».

СПИСОК ЛИТЕРАТУРЫ | REFERENCES

- [Aci22] Acien A., Morales A., Fierrez J., Vera-Rodriguez R. "BeCAPTCHA-Mouse: Synthetic mouse trajectories and improved bot detection" // Pattern Recognition. 2022. Vol. 127, art. 108643. DOI: [10.1016/j.patcog.2022.108643](https://doi.org/10.1016/j.patcog.2022.108643).
- [Ant20] Antal M., Fejer N. "Mouse dynamics based user recognition using deep learning" // Acta Universitatis Sapientiae, Informatica. 2020. Vol. 12, no. 1, pp. 39–50. DOI: [10.2478/ausi-2020-0003](https://doi.org/10.2478/ausi-2020-0003).
- [Bre01] Breiman L. "Random forests" // Machine Learning. 2001. Vol. 45, no. 1, pp. 5–32. DOI: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- [Che16] Chen T., Guestrin C. "XGBoost: A scalable tree boosting system" // Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016, pp. 785–794. DOI: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).
- [Chu14] Chuda D., Kratyk P. "Usage of computer mouse characteristics for identification in web browsing" // Proc. of the 15th International Conference on Computer Systems and Technologies (CompSysTech '14). New York: ACM, 2014, pp. 218–225. DOI: [10.1145/2659532.2659645](https://doi.org/10.1145/2659532.2659645).
- [Feh12] Feher C., Elovici Y., Moskovitch R., Rokach L., Schclar A. "User identity verification via mouse dynamics" // Information Sciences. 2012. Vol. 201, pp. 19–36. DOI: [10.1016/j.ins.2012.02.066](https://doi.org/10.1016/j.ins.2012.02.066).
- [Fla85] Flash T., Hogan N. "The coordination of arm movements: an experimentally confirmed mathematical model" // Journal of Neuroscience. 1985. Vol. 5, no. 7, pp. 1688–1703. DOI: [10.1523/JNEUROSCI.05-07-01688.1985](https://doi.org/10.1523/JNEUROSCI.05-07-01688.1985).
- [Gam04] Gamboa H., Fred A. "A behavioral biometric system based on human-computer interaction" // Proc. SPIE. Vol. 5404: Biometric Technology for Human Identification. Bellingham: SPIE, 2004, pp. 381–392. DOI: [10.1117/12.542625](https://doi.org/10.1117/12.542625).
- [Gri22] Grinsztajn L., Oyallon E., Varoquaux G. "Why do tree-based models still outperform deep learning on typical tabular data?" // Advances in Neural Information Processing Systems (NeurIPS 2022), Datasets and Benchmarks Track. 2022. Vol. 35, pp. 507–520. DOI: [10.48550/arXiv.2207.08815](https://doi.org/10.48550/arXiv.2207.08815).
- [Guy03] Guyon I., Elisseeff A. "An introduction to variable and feature selection" // Journal of Machine Learning Research. 2003. Vol. 3, pp. 1157–1182. DOI: [10.1162/153244303322753616](https://doi.org/10.1162/153244303322753616).
- [Ke17] Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T.-Y. "LightGBM: A highly efficient gradient boosting decision tree" // Advances in Neural Information Processing Systems (NIPS 2017). 2017. Vol. 30, pp. 3146–3154.
- [Kil21] Kılıç A. A., Yıldırım M., Anarım E. "Bogazici mouse dynamics dataset" // Data in Brief. 2021. Vol. 36, art. 107094. DOI: [10.1016/j.dib.2021.107094](https://doi.org/10.1016/j.dib.2021.107094).
- [Liu25] Liu Y., Xu S., Li Y., Yu S. "Evolution of malicious social bot detection: From individual profiling to group analysis and beyond" // Journal of Social Computing. 2025. Vol. 6, no. 3, pp. 258–284. DOI: [10.23919/JSC.2025.0017](https://doi.org/10.23919/JSC.2025.0017).
- [Lun17] Lundberg S. M., Lee S.-I. "A unified approach to interpreting model predictions" // Advances in Neural Information Processing Systems (NIPS 2017). 2017. Vol. 30, pp. 4765–4774. DOI: [10.48550/arXiv.1705.07874](https://doi.org/10.48550/arXiv.1705.07874).
- [Niu24] Niu H., Zhou Y., Chen J., Cai Z. "Performance evaluation of lightweight network-based bot detection using mouse movements" // Engineering Applications of Artificial Intelligence. 2024. Vol. 135, art. 108801. DOI: [10.1016/j.engappai.2024.108801](https://doi.org/10.1016/j.engappai.2024.108801).
- [Pla95] Plamondon R. "A kinematic theory of rapid human movements. Part I. Movement representation and generation" // Biological Cybernetics. 1995. Vol. 72, no. 4, pp. 295–307. DOI: [10.1007/BF00202785](https://doi.org/10.1007/BF00202785).
- [Shi23] Shi Y., Wang X., Zheng K., Cao S. "User authentication method based on keystroke dynamics and mouse dynamics using HDA" // Multimedia Systems. 2023. Vol. 29, no. 2, pp. 653–668. DOI: [10.1007/s00530-022-00997-5](https://doi.org/10.1007/s00530-022-00997-5).
- [Shw22] Shwartz-Ziv R., Armon A. "Tabular data: Deep learning is not all you need" // Information Fusion. 2022. Vol. 81, pp. 84–90. DOI: [10.1016/j.inffus.2021.11.011](https://doi.org/10.1016/j.inffus.2021.11.011).
- [Wei19] Wei A., Zhao Y., Cai Z. "A deep learning approach to web bot detection using mouse behavioral biometrics" // Biometric Recognition (CCBR 2019). Lecture Notes in Computer Science. Vol. 11818. Cham: Springer, 2019, pp. 388–395. DOI: [10.1007/978-3-030-31456-9_43](https://doi.org/10.1007/978-3-030-31456-9_43).
- [Zhe11] Zheng N., Paloski A., Wang H. "An efficient user verification system via mouse movements" // Proc. of the 18th ACM Conference on Computer and Communications Security. New York: ACM, 2011, pp. 139–150. DOI: [10.1145/2046707.2046725](https://doi.org/10.1145/2046707.2046725).
- [Афа24] Афанасьева Н. С., Ложников П. С. Наборы данных, используемые для определения ботов на основании движения компьютерной мыши // Доклады Томского государственного университета систем управления и радиоэлектроники. 2024. Т. 27, № 3. С. 118–124. DOI: [10.21293/1818-0442-2024-27-3-118-124](https://doi.org/10.21293/1818-0442-2024-27-3-118-124). [Afanaseva N. S., Lozhnikov P. S. "Datasets used to detect bots based on computer mouse movement" // Proceedings of TUSUR University. 2024. Vol. 27, no. 3, pp. 118–124. DOI: [10.21293/1818-0442-2024-27-3-118-124](https://doi.org/10.21293/1818-0442-2024-27-3-118-124). (In Russian).]
- [Афа26] Афанасьева Н. С. Формирование пространства признаков, позволяющих осуществлять детектирование автоматизированных средств подмены пользователя в информационной среде // Безопасность информационных технологий. 2026 (в печати). [Afanaseva N. S. "Formation of a feature space enabling detection of automated user substitution tools in the information environment" // IT Security (Russia). 2026 (in press). (In Russian).]

ОБ АВТОРЕ | ABOUT THE AUTHOR

АФАНАСЬЕВА Наталья Сергеевна

Омский государственный университет путей сообщения,
Россия.

nati_dik@mail.ru ORCID: 0009-0003-7591-486X.

Ст. преп. каф. «Информационная безопасность».

Иссл. в обл. поведенческой биометрии.

AFANASEVA Natalia Sergeevna

Omsk State Transport University,
Russia.

nati_dik@mail.ru ORCID: 0009-0003-7591-486X.

Senior Lecturer of the Department of Information Security.

Research in the field of behavioral biometrics.

МЕТАДАННЫЕ | METADATA

Заглавие: Экспериментальная оценка использования расширенного пространства признаков для детектирования автоматизированных средств подмены пользователя.

Авторы: Афанасьева Н. С.

Аннотация: Ранее автором был предложен подход к формированию расширенного пространства признаков, позволяющих осуществлять детектирование автоматизированных средств подмены пользователя в информационной среде на основании передвижения компьютерной мыши. Рассматривался базовый набор, применяемый в известных работах по поведенческой биометрии мыши, состоящий из 56 параметров, который был дополнен 166 новыми признаками. Результирующее комбинированное пространство образовано 222 признаками. Целью настоящей работы является эмпирическая проверка гипотезы о том, что предложенное расширенное пространство признаков обеспечивает более высокое качество детектирования ботов по сравнению с базовым набором признаков. На данных, включающих датасеты BALABIT, BOGAZICI и внутренние наборы синтетических и браузерных ботов, обучен классификатор Random Forest с оптимизацией гиперпараметров методом GridSearchCV. Получены значения точности (accuracy) 0,99–0,999 и F1-меры по классу «бот» до 0,945–0,983. Дополнительно проведен контролируемый вычислительный эксперимент, в котором сопоставлены базовый набор признаков и комбинированное пространство. Переход к комбинированному пространству признаков сопровождается ростом полноты (recall) по классу «бот» с 0,800 до 0,912 и F1-меры с 0,887 до 0,945. Анализ важности признаков показал, что подавляющее большинство наиболее значимых для классификации характеристик относится к предложенным новым признакам, а не к базовому набору. Проведено сопоставление с результатами известных работ в области поведенческой биометрии. Результаты подтверждают целесообразность расширения признакового пространства для повышения точности детектирования автоматизированных средств подмены пользователя.

Ключевые слова: Бот; вредоносный бот; пространство признаков; детектор ботов; поведенческая биометрия.

Язык: Русский.

Статья поступила в редакцию 2 октября 2026 г.

Title: Experimental evaluation of an extended feature space for detecting automated user substitution tools.

Authors: Afanaseva N. S.

Abstract: An approach to forming an extended feature space for detecting automated user substitution tools (bots) in an information environment based on computer mouse movement was previously proposed by the author. The basic set of 56 parameters used in known works on mouse behavioral biometrics was supplemented with 166 new features, and the resulting combined space consists of 222 features. The aim of this work is to empirically test the hypothesis that the proposed extended feature space provides higher bot detection quality than the basic feature set. A Random Forest classifier with hyperparameters optimized by GridSearchCV was trained on data including the BALABIT and BOGAZICI datasets and in-house sets of synthetic and browser bots. Accuracy values of 0.99–0.999 and F1-scores for the bot class of up to 0.945–0.983 were obtained. An additional controlled computational experiment compared the basic feature set and the combined space. The transition to the combined feature space is accompanied by an increase in recall for the bot class from 0.800 to 0.912 and in the F1-score from 0.887 to 0.945. Feature importance analysis showed that the vast majority of the features most significant for classification belong to the proposed new features rather than to the basic set. The results are compared with those of known works in the field of behavioral biometrics. The results confirm the expediency of extending the feature space to improve the accuracy of detecting automated user substitution tools.

Key words: Bot; malicious bot; feature space; bot detector; behavioral biometrics.

Language: Russian.

The editors received the article on 2 October 2026.