DOI: 10.17323/2587-814X.2020.3.54.66

Методический подход к выявлению ботов в социальных сетях на основе специального объединения классификаторов

В.Н. Кузьмин
E-mail: kuzmindvn@ya.ru

А.Б. Менисов
E-mail: men.arty@yandex.ru

И.А. Шастун
E-mail: shastunivan1982@gmail.com

Военно-космическая академия имени А.Ф. Можайского
Адрес: 197198, г. Санкт-Петербург, ул. Ждановская, д. 13

Аннотация

В настоящее время применение ботов – автоматизированных учетных записей в социальных сетях, управляемых программным обеспечением, но замаскированных под обычных пользователей, имеет серьезные последствия. Например, боты использовались для влияния на политические выборы, искажения информации в сети интернет и манипуляции стоимостью акций на фондовой бирже. Выявлением таких аккаунтов занимаются многие научные коллективы, к направлениям исследований которых относится применение методов машинного обучения. Однако практические результаты выявления ботов в социальных сетях свидетельствуют о наличии существенных ограничений, так как используемый методический инструментарий обладает языковой ограниченностью и неэффективной критериальной базой определения ботов. В статье представлен методический подход к разработке универсального классификатора аккаунтов социальной сети, направленный на совершенствование способов противодействия ботам и позволяющий минимизировать средний риск ошибки выявления ботов. В основу формирования универсального классификатора аккаунтов социальных сетей положено использование ансамбля классификаторов, объединенных по критерию адаптации к исходным данным и дисперсии результатов каждой модели разрабатываемого ансамбля. Основными результатами, полученными авторами, являются предложенная система критериев выявления ботов и подход к преобразованию категориальных (номинальных) признаков для формирования специального ансамбля классификаторов. Практическое применение ансамбля моделей повысило результативность выявления ботов в сравнении с другими подходами.

Ключевые слова: выявление ботов; социальные сети; машинное обучение; ансамбль моделей; объединение классификаторов.

Цитирование: Кузьмин В.Н., Менисов А.Б., Шастун И.А. Методический подход к выявлению ботов в социальных сетях на основе специального объединения классификаторов // Бизнес-информатика. 2020. Т. 14. № 3. С. 54–66. DOI: 10.17323/2587-814X.2020.3.54.66

Введение

Выявление ботов в социальных сетях является предметом изучения уже более десяти лет [1] по причине их активного использования для достижения политических пропагандистских целей. Однако, до настоящего времени однозначного толкования термина «бот социальной сети» не сформировалось [1]. В настоящем исследовании под ботом будем понимать специальную страницу (аккаунт) социальной сети, замаскированную под обычного пользователя, которая автоматически и/или по расписанию выполняет действия по публикации, продвижению и комментированию материалов, направленных на достижение определенной пропагандистской или политической (экономической) цели. В зависимости от сферы применения можно выделить ряд характерных целей, представленных в *таблице 1.*

Таблица 1.

Информационные цели применения ботов социальных сетей

№	Сферы применения	Информационные цели
1.	политическая	продвижение идеологии; навязывание политических взглядов; агитация; пропаганда; привлечение электората
2.	экономическая	реклама товаров и услуг; повышение узнаваемости бренда
3.	социальная	повышение узнаваемости личности; черный, серый и белый PR
4.	духовная	пропаганда и изменение мировоззренческих стереотипов

В современных условиях применение ботов социальных сетей превратилось в угрозу, реализация которой направлена на дискредитацию легитимной власти и ухудшение управляемости общественными процессами и организациями [2]. В настоящее время по количеству ботов и последствий их действий выделяется социальная сеть Twitter, в которой насчитывают более миллиона таких аккаунтов [3], а отдельные сети ботов (ботнеты) содержат до полумиллиона аккаунтов [4].

Многие организации заинтересованы в развитии и совершенствовании средств противодействия применению ботов [5, 6], для выявления ботов, оценивания результатов их действий и нейтрализацию последствий. Анализ опубликованных результатов исследований [7–22] по данной тематике показал, что для выявления ботов социальных сетей широко применяются методы машинного обучения и нейронные сети. Можно выделить два основных подхода к выявлению ботов социальных сетей: на основе обработки публикуемых пользователями материалов (*таблица 2*) и на основе обработки количественно-качественных характеристик самих аккаунтов (*таблица 3*).

Несмотря на достижение хороших результатов разработанных подходов [1, 7–16], можно выделить следующие их недостатки:

- ✦ отсутствие достаточно полного набора данных для проверки качества выявления ботов социальных сетей;
- ✦ языковая ограниченность применяемых методов.

В связи с вышеизложенными недостатками, а также для обеспечения универсальности выявление ботов целесообразно проводить по результатам анализа количественно-качественных характеристик аккаунтов, которые у некоторых авторов трактуются как «метапризнаки» [3, 6–8, 11].

По мере того, как алгоритмы, управляющие ботами социальных сетей, становятся более сложными, развиваются и алгоритмы выявления ботов. В исследованиях [17–22] модели выявления ботов варьируются от самых простых [17–19], предназначенных для анализа одного фрагмента метапризнаков, до моделей, использующих ансамблевые подходы для анализа больших наборов данных, включая сочетание метапризнаков, действия аккаунтов в социальной сети и данных контента [20–22]. Ансамбли моделей могут обнаруживать новое

Таблица 2.

Результаты анализа методов и результативности выявления ботов социальных сетей (на основе обработки публикуемых пользователями материалов)

№	Авторы	Метод выявления	Результативность	Языки	Примечание
1.	A. Bacciu, M. La Morgia, A. Mei, E. Nerio Nemmi, V. Neri, J. Stefa [7]	латентно–семантический анализ твитов	точность больше 0,9	английский, испанский	Исследование проводилось на данных, предоставляемых на конференции PAN 2019
2.	P. Gamallo, S. Almatarneh [8]	байесовский классификатор	точность для английского языка – 0,81, для испанского языка – 0,88	английский, испанский	PAN 2019
3.	I. Vogel, P. Jiang [9]	метод главных компонент, применение N–грамм	точность для английского языка – 0,92, для испанского языка – 0,91	английский, испанский	PAN 2019
4.	A. Mahmood, P. Srinivasan [10]	TF–IDF	точность – 0,91	английский	
5.	M. Farber, A. Qurdina, L. Ahmedi [11]	нейронная сеть (CNN)	точность – 0,9	английский	PAN 2019

Таблица 3.

Результаты анализа методов и результативности выявления ботов социальных сетей (на основе обработки количественно-качественных характеристик аккаунтов)

№	Авторы	Метод выявления	Результативность
1.	J. Lundberg, J. Nordqvist, M. Laitinen [12]	алгоритм случайного леса	точность больше 0,9
2.	S.R. Sahoo, B.B. Gupta [13]	сети Петри	точность – 0,9916
3.	J. Novotny [14]	алгоритм случайного леса	точность – 0,9
4.	A. Davoudi, A. Z Klein, A. Sarker, G. Gonzalez–Hernandez [15]	Botometer	F–мера – 0,7
5.	M.Mazza, S. Cresci, M. Avvenuti, W. Quattrociocchi, M. Tesconi [16]	нейронная сеть (LSTM)	F–мера – 0,87

поведение ботов, которое не может быть выявлено отдельными моделями, поскольку последние способны обнаруживать только те боты, которые достаточно похожи на данные, которые использовались для обучения [17].

1. Постановка задачи исследования

Задано множество аккаунтов социальной сети A, множество их состояний $Y = \{0, 1\}$ и существует статистическая функция $y^*: A \to Y$, значения которой $y_i = y^*(a_i)$ известны только на конечном подмножестве объектов $\{a_1, a_2, ..., a_n\} \subset A$, причем при $y_i = 1$ аккаунт является ботом, а при $y_i = 0$ – обычным пользователем. Пары «объект–состояние» (a_i, y_i) являются прецедентами. Совокупность пар прецедентов $A^l = (a_i, y_i)_{i=1}^l$ является обучающей выборкой для восстановления зависимости y^*.

Задача выявления ботов социальных сетей заключается в том, чтобы построить решающую функцию $z: A \to Y$, максимально приближенную к $y^*(a)$, причем не только на объектах обучающей выборки, но и на всем множестве A. Иначе говоря, необходимо определить состояние произвольного аккаунта со-

циальной сети $a \in A$. При этом вероятность правильной классификации и вероятности ошибок задают средний риск ошибки выявления ботов:

$$H = \mathbb{E}[E] = p_o 0 + p_1 E_1 + ... + p_i E_i , \quad (1)$$

где H – риск ошибки выявления ботов;

$\mathbb{E}[E]$ – математическое ожидание ошибок выявления;

E – множество ошибок выявления;

$\langle E_1, ..., E_i \rangle$ – ошибки выявления;

i – количество классов состояний аккаунтов социальных сетей;

p_o – вероятность правильного решения;

$\langle p_1, ..., p_i \rangle$ – вероятности ошибок.

Таким образом, задача выявления ботов социальных сетей заключается в формировании решения о состоянии аккаунта социальной сети в наблюдаемый момент времени. В свою очередь, требования к качеству решения определяются в форме требований минимизации риска, связанного с принятием неправильного решения:

$$H = \mathbb{E}[E] = p_1 E_1 + p_2 E_2 \to \min \quad (2)$$

где H – риск ошибки выявления ботов;

$\mathbb{E}[E]$ – математическое ожидание ошибок выявления;

p_1 – вероятность ошибки первого рода, т.е. когда ошибочно классифицирован аккаунт пользователя как бот;

p_2 – вероятность ошибки второго рода, т.е. когда имеет место пропуск бота.

2. Методы

2.1. Критерии выявления ботов социальных сетей

Необходимо отметить, что не рекомендуется выявление ботов социальных сетей только по одному показателю (например, только по количеству публикаций или числу подписчиков) [22]. Важна комбинация таких признаков, как тематическая взаимосвязь аккаунтов, активность, анонимность и, в некоторых случаях, противоречивость данных.

Тематическая взаимосвязь аккаунтов. Наличие переходов, подписок или совершение других действий множеством аккаунтов на определенном тематическом кластере аккаунтов (или на одном аккаунте) является признаком применения ботов, так как одна из основных задач ботов – «усиливать сигнал» других пользователей, не только комментируя и цитируя их. В социальных сетях применяется система ранжирования, которая повышает степень распространения материалов аккаунта в зависимости от числа не только подписчиков, но и тех, кто пассивно просматривает материал или просто осуществляет переход на страницы.

Активность. Наиболее очевидный признак бота – его активность. Определить этот признак позволяют открытые данные (например, количество и частота постов и подписок с момента создания аккаунта).

Анонимность. Третий важный признак – степень анонимности аккаунта. В целом, чем меньше в аккаунте личной информации, тем вероятнее, что это бот. Также признаком бота является настроенная приватность для страницы.

Противоречивость данных может заключаться в несоответствии языка материалов и места основания аккаунта, места основания аккаунта и временной зоны работы аккаунта.

2.2. Формирование набора исходных данных

Применение ботов социальных сетей привлекло особое внимание общественности стран Южной Африки, когда PR-компания «Bell Pottinger» использовала социальные сети для распространения негативного контента [23]. Среди подписчиков учетных записей южноафриканских политиков имелись боты, которые следили и модерировали их твиты на своих страницах [24].

Анализ данных двух аккаунтов южноафриканских политиков (Пола Машалита, председателя Африканского национального конгресса (АНК) в провинции Гаутенг [25], и Айанды Длодло, члена АНК [26]), собранных в сентябре 2018 года, показал, что из 12 тысяч активных подписчиков 863 подписчика являются общими (*рисунок 1*). Из данной категории пользователей 121 аккаунт явно использовался для повышения рейтинга пропагандистского материала. Они были выбраны по следующим показателям: аккаунт распространял более 100 сообщений в сутки, а также отличался высокой активностью, анонимностью и противоречивостью данных.

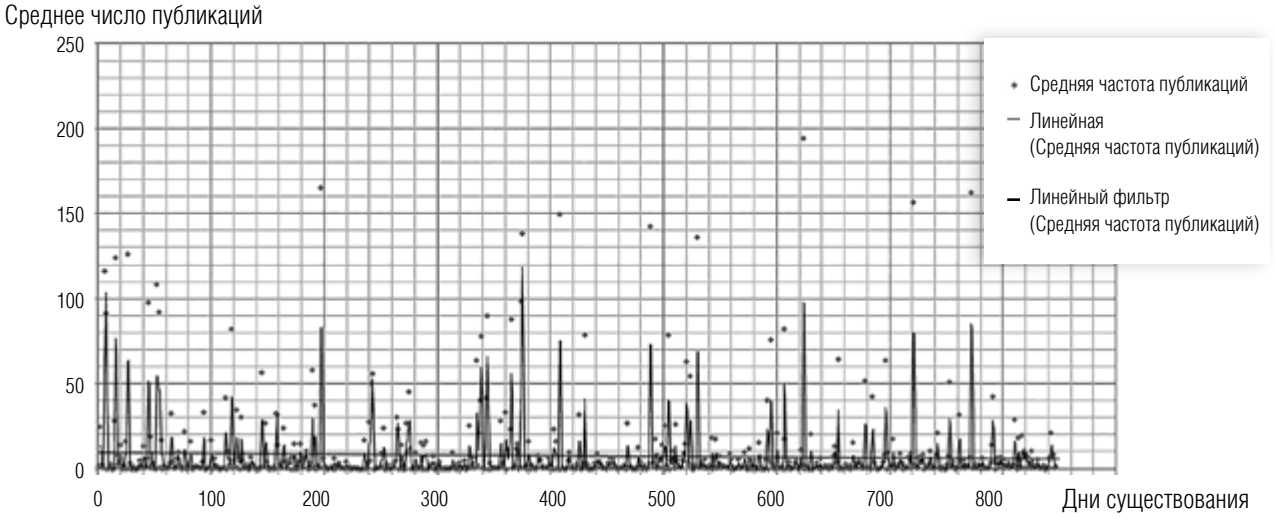

Рис. 1. Анализ частоты публикации у общих активных подписчиков двух членов Африканского национального конгресса

2.3. Описание методики применения специального объединения классификаторов

Методическая последовательность выявления ботов социальных сетей включает следующие этапы: поиск и сохранение данных аккаунтов социальной сети, предварительная обработка данных, выбор и обучение отдельных моделей машинного обучения, их объединение, определение состояния аккаунтов социальной сети.

Этап 1. Поиск и сохранение данных аккаунтов социальной сети**.** Эта фаза направлена на сбор всей доступной информации об аккаунтах социальной сети, с помощью методов API (application program interface) социальной сети Twitter [27]. Описание данных представлено в *таблице 4.*

Таблица 4.

Доступная информация об аккаунтах социальной сети Twitter

№	Название	Описание	Тип признака	Пример
1	ID	идентификационный номер	строка	4452841
2	Screen name	отображаемое имя	строка	sToneBirD
3	Date of creation	дата создания аккаунта	дата	2007–04–13 04:33:54
4	Favorites	страницы, на которые подписан	число	43
5	Followers	подписчики	число	386
6	Friends	друзья	число	798
7	GEO	географические настройки	категория	True
8	Lang	язык	категория	En
9	List	списки	число	18
10	Location	местоположение	категория	Doha, Qatar
11	Protected	приватность	категория	False
12	Status count	количество постов	число	3770
13	Time zone	часовой пояс	категория	Seoul
14	URL	адрес	строка	http://pbs.twimg.com//...
15	Verified	верификация	категория	False

Этап 2. Предварительная обработка данных. Данный этап включает заполнение пустых (некорректных) записей, а также преобразование категориальных (номинальных) признаков [28].

Признак f объекта $a \in A$ (где A – множество аккаунтов социальной сети) – это результат оценивания некоторой характеристики объекта [28]. Формально признаком называется отображение $f: A \to D_f$, где D_f – множество допустимых значений признака. Если D_f – конечное множество, то f – номинальный признак, а $f_1(a)$, ..., $f_n(a)$ – признаковое описание объекта $a \in A$. Для проведения исследования будем полагать, что $A = D_{f1} \times \ldots \times D_{fn}$.

Для преобразования категориальных (номинальных) признаков применим прием горячего кодирования [28], позволяющий представить все категории в виде дискретных значений. Пусть $d_i \in D_{f(cat)}$, $i = 1, ..., k$ – такая категориальная переменная, что при заданном сходстве категорий $sim(d_i, d_j)$: $D_{f(cat)} \times D_{f(cat)} \to [0, 1]$ определяется множество значений признака $f^{sim} \in R$ в виде:

$$f^{sim} = [sim(d_i, d_1), \ldots sim(d_i, d_k)], \qquad (3)$$

где $d_k \in D_{f(cat)}$ – набор всех категорий.

Подход преобразования категориальных признаков позволяет избежать выполнения одного из трудоемких этапов обучения моделей машинного обучения – нормализации записей в базах данных [29].

Этап 3. Выбор и обучение моделей. Разработанный подход ориентирован на применение ансамбля классификаторов. Ансамбли хорошо известны своим эффектом повышения точности и обобщающей способности решения, а также обеспечением параллелизма. Они успешно использовались в различных задачах бинарной классификации [30]. Поскольку деятельность ботов социальных сетей включает категориальные признаки, для использования ансамблей моделей необходима адаптация к исходным данным (этап 2). Особенностью подхода на основе специального объединения классификаторов является то, что объединение классификаторов представляет собой обертку для множества различных моделей, которые работают параллельно (*рисунок 2*), чтобы использовать различные достоинства каждой модели [31].

Создание специального объединения классификаторов заключается в выполнении следующих действий:

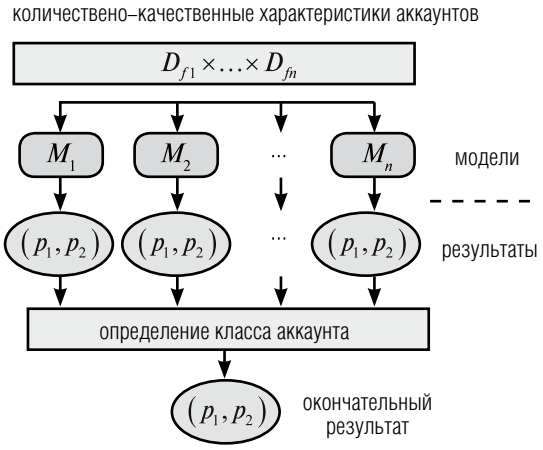

Рис. 2. Схема функционирования объединения классификаторов ($D_{f1} \times \ldots \times D_{fn}$ – множество признаков, $\langle M_1, ..., M_n \rangle$ – множество моделей, (p_1, p_2) – ошибки выявления)

1. Разработка N отдельных моделей, каждая из которых обладает своими значениями точности выявления. Поскольку обучение каждой модели производится с множеством разбиений выборки данных для скользящего контроля, значение количества моделей (N) зависит от статистической устойчивости результатов и улучшения решения.

2. Обучение каждой модели отдельно. Обучение по прецедентам каждой модели сводится к подбору наилучшего значения гиперпараметров модели (управляющих, внешних) [32]. Например, в модели полиномиальной регрессии попытка оптимизировать степень полинома по обучающей выборке приведет к выбору максимально возможной степени и переобучению.

3. Объединение моделей и улучшение значений в конечном классификаторе следующим приемом: вектор вероятности ошибок по всему множеству прецедентов для каждого прогнозируемого класса (для всех классификаторов) суммируется и усредняется. Значение класса выявляется при минимизации среднего риска ошибки выявления (выражение (2)):

$$y^*(a) \approx z(a), \text{при } \arg\min \frac{1}{N} \sum_{i=1}^{N} \left[\left(H(p_1, p_2) \right)_i \right], \qquad (4)$$

где $z(a)$ – значение класса;

N – количество моделей в конечном классификаторе;

$H(p_1, p_2)$ – риск, связанный с ошибками первого и второго рода.

Этап 4. Определение состояния аккаунтов социальной сети. Значение класса (бот или действительный аккаунт) выбирается путем противопоставления моделей друг другу с учетом весов. Для примера предположим, что в ансамбле находится три модели, выходными данными которых является следующие значения:

$$z_1(a) = 0, \text{ при } H = (0{,}1;\ 0{,}1),$$
$$z_2(a) = 0, \text{ при } H = (0{,}5;\ 0{,}5),$$
$$z_3(a) = 1, \text{ при } H = (0{,}9;\ 0{,}9).$$

В представленных выходных данных классификации модель, выполняющая объединение решений, может интерпретировать этот результат как $y^*(a) = 0$, однако присвоение моделям весов {0,1, 0,1, 0,8} даст прогноз $y^*(a) = 1$. Заметим, что возможность такого выбора является существенным отличием методов объединения решений, основанных на многоярусном обобщении, от других подходов, например, от методов, в которых финальное решение всегда выбирается из множества решений, предложенных базовыми классификаторами [33, 34].

Преимущество предложенного подхода заключается в том, что ансамбль моделей может быть легче обучен на небольших входных наборах данных и повысит результативность выявления ботов по сравнению с любой отдельной моделью.

3. Результаты

В научных статьях по машинному обучению [35, 36] принято приводить результаты тестирования предложенного нового метода обучения в сравнении с другими методами на представительном наборе задач. Сравнение должно проводиться в равных условиях по одинаковой методике (особенно если это скользящий контроль), при одном и том же количестве разбиений выборки данных на обучающие и валидационные части. В *таблице 5* представлен результат сравнения разработанного подхода с семью отдельными моделями машинного обучения, подходящими для бинарной классификации и входящими в состав конечного классификатора.

Такое повышение точности классификации можно объяснить тем, что каждая модель обладает весом, характеризующим важность вклада в общее решение, которое вычисляется по формуле (3). Вклад каждого классификатора может быть интерпретирован как оценка его компетентности, используемая для масштабирования выходов (результатов работы) классификаторов, тем самым увеличивая или уменьшая вклад каждого классификатора в общее решение.

На *рисунке 3* представлена дисперсия результатов классификации (точности) разработанного подхода и отдельных моделей, входящих в конечный ансамбль, полученная при скользящем контроле.

Таблица 5.

Сравнение результатов моделей для выявления ботов социальных сетей

№	Модель	Время обучения модели	Время вычисления	Точность	Точность (precision)	Полнота (recall)	F1-мера	AUC_ROC
1	Предлагаемая модель	0,003908	0,006760	0,994577	0,991750	0,986458	0,994471	0,999306
2	Логистическая регрессия	0,003907	0,003127	0,949845	0,921635	0,872000	0,946988	0,872000
3	Случайный лес	0,010161	0,008214	0,941092	0,925735	0,815792	0,936246	0,964767
4	K–ближайших соседей	0,000000	0,012502	0,939055	0,896358	0,864667	0,938098	0,936906
5	Линейный дискриминантный анализ	0,087689	0,012163	0,866311	0,433155	0,500000	0,804308	0,506250
6	Метод главных компонент	0,002342	0,005471	0,858846	0,753099	0,841125	0,872463	0,908184
7	Модель Байеса	0,00424	0,01271	0,807822	0,750054	0,803255	0,842993	0,807844
8	Квадратичный дискриминантный анализ	0,001763	0,004887	0,740661	0,678199	0,809625	0,777933	0,933076

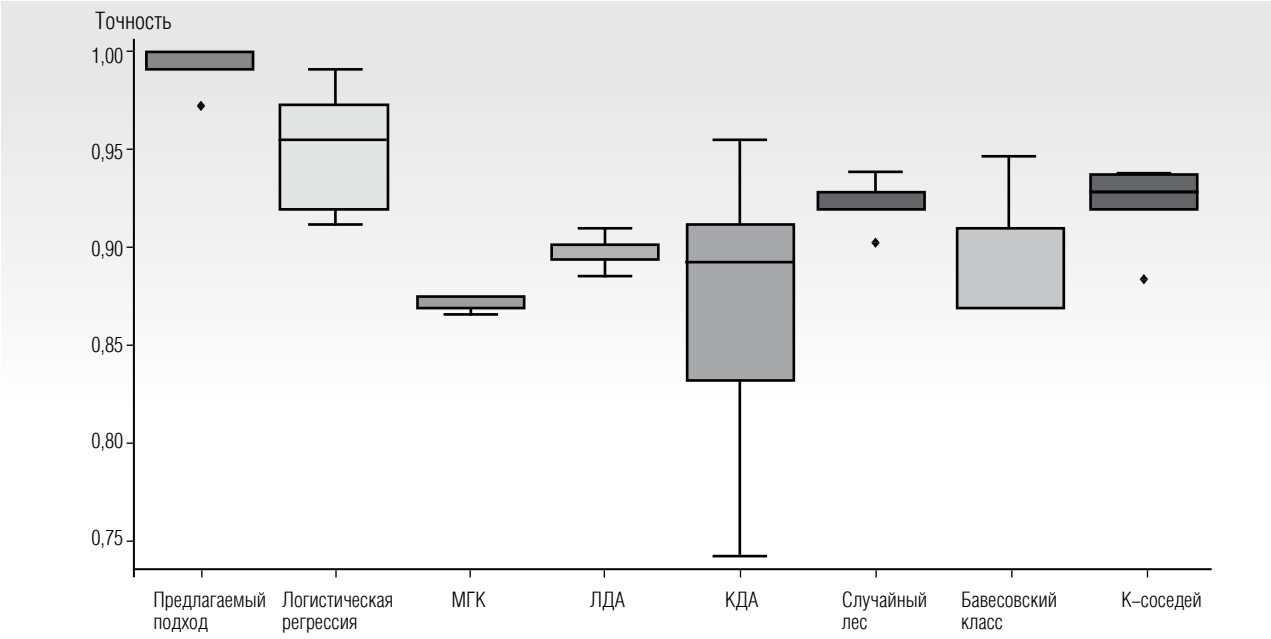

Рис. 3. Вклад классификаторов в общее решение

4. Дискуссия

Для оценки качества разработанного подхода определим валидационную часть как равную 0,3 от общего набора данных (пункт 1.2), включающих 47 записей для ботов и 242 записей обычных аккаунтов.

Для оценивания качества выходных данных построим матрицу несоответствия выявления ботов, представленную на *рисунке 4а*.

Матрица несоответствия отображает количество верных и ошибочных выявлений по сравнению с фактическими данными:

✦ (0,0) – правильно выявленные обычные аккаунты социальных сетей;

✦ (1,1) – правильно выявленные боты;

✦ (0,1) – для обычных аккаунтов принято решение о том, что они являются ботами;

✦ (1,0) – для ботов принято решение о том, что они обычные аккаунты.

Эти вероятности первого и второго рода можно вычислить как вероятность попадания случайной величины z в область допустимых значений классов аккаунтов социальных сетей, то есть $p_1 = P(0,1)$ и $p_2 = P(1,0)$. Подставив эти значения из матрицы несоответствия в формулу (1) получим:

$$H = \mathbb{E}[E] = \frac{2}{241}2 + \frac{1}{47}1 \approx 0,038,$$

где H – риск ошибки выявления ботов;

$\mathbb{E}[E]$ – математическое ожидание ошибок выявления.

Сравним полученный средний риск с результатами работы модели CatBoost [37], разработанной российской компанией Яндекс (*рисунок 4б*). Она основана на градиентном бустинге с реализацией подхода преобразования категориальных (номинальных) признаков:

$$H_{CatBoost} = \mathbb{E}[E_{CatBoost}] = \frac{1}{253}1 + \frac{11}{47}11 \approx 2,57,$$

где $H_{CatBoost}$ – риск ошибки выявления ботов модели CatBoost;

$\mathbb{E}[E_{CatBoost}]$ – математическое ожидание ошибок выявления модели CatBoost.

Также сравним полученный результат с результатом, полученным на основе среднего риска рандомного выбора:

$$H_{random} = \mathbb{E}[E_{random}] = \frac{126}{253}126 + \frac{24}{47}24 \approx 25,195,$$

где H_{random} – риск ошибки рандомного выявления ботов;

$\mathbb{E}[E_{random}]$ – математическое ожидание ошибок рандомного выявления.

Таким образом, предлагаемый подход показал лучший результат, что характеризует повышение качества выявления ботов социальных сетей.

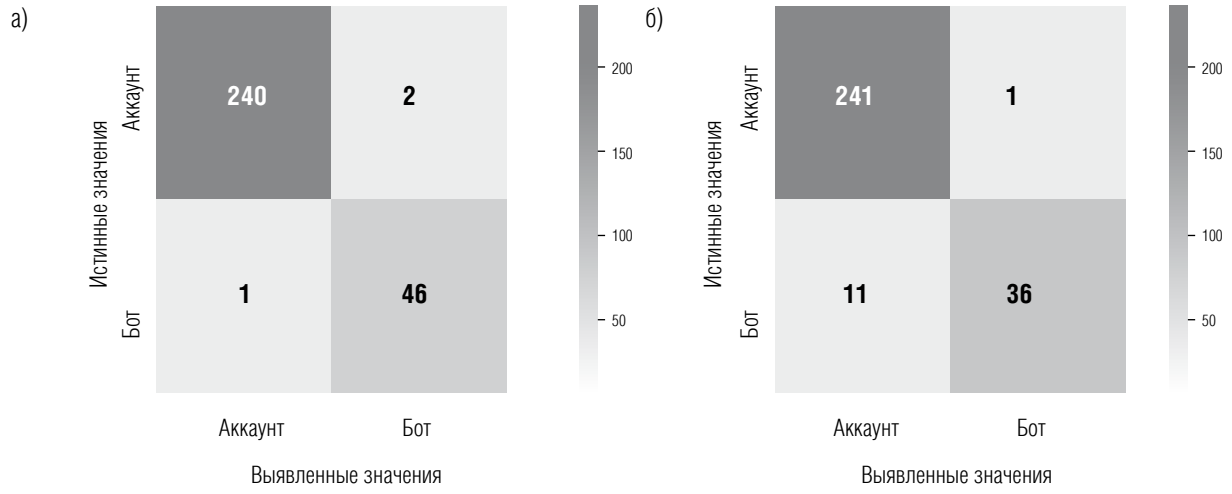

Рис. 4. Матрица несоответствия выявления ботов среди подписчиков аккаунтов южноафриканских политиков: а) с помощью предлагаемого подхода; б) с помощью подхода CatBoost

Заключение

Разработка новых подходов, позволяющих повысить защищенность государственных организаций и пользователей информационных web-систем, является постоянной и актуальной задачей.

Элементом научной новизны разработанного подхода к выявлению ботов социальных сетей является рекомендованная комбинация ряда признаков: тематической взаимосвязи аккаунтов, активности, анонимности и противоречивости данных. Особенностью данного подхода является учет возрастающей тенденции применения одного множества ботов для достижения разных информационных целей.

Практическая значимость исследования заключается в возможности применения предлагаемого подхода при обосновании и разработке технических решений информационной безопасности.

Разработанный подход к выявлению ботов в социальной сети Twitter на основе специального объединения классификаторов имеет преимущество по результативности по сравнению с современными алгоритмами машинного обучения и позволяет снизить ошибки выявления ботов. Поскольку деятельность ботов социальных сетей включает категориальные признаки, для использования ансамблей моделей необходима адаптация к исходным данным.

Однако, несмотря на достоинства машинного обучения, одним из основных недостатков разработанного подхода может оказаться его непрактичность при наличии слишком большого количества уникальных записей, например, если строковые представления категориальных признаков отображают опечатки или комбинации нескольких данных в одних и тех же записях.

В качестве направлений дальнейшего развития данного исследования можно выделить следующие:

✦ исследование вопросов сбора дополнительных данных об аккаунтах социальных сетей;

✦ анализ влияния дисбаланса данных на обучение моделей;

✦ исследование возможностей повышения производительности выявления ботов социальных сетей;

✦ разработка технических решений по совершенствованию сервисов выявления ботов разных типов. ■

Литература

1. Williamson W., Scrofani J. Trends in detection and characterization of propaganda bots // 52nd Hawaii International Conference on System Sciences. Honolulu, USA, 8–11 January 2019. P. 7118–7123. DOI: 10.24251/HICSS.2019.854.

2. Лукьянов Р.В. Методика контроля состояния информационной безопасности автоматизированных систем в условиях разнородно-массовых инцидентов // Труды Военно-космической академии имени А.Ф. Можайского. 2018. № 660. С. 111–115.

3. As many as 48 million Twitter accounts aren't people, says study // CNBC. [Электронный ресурс]: https://www.cnbc.com/2017/03/10/nearly-48-million-twitter-accounts-could-be-bots-says-study.html (дата обращения 01.12.2019).
4. Massive networks of fake accounts found on Twitter // BBC. [Электронный ресурс]: http://www.bbc.co.uk/news/technology-38724082 (дата обращения 01.12.2019).
5. Terdima D. Here's how Facebook uses AI to detect many kinds of bad content // Fast Company. [Электронный ресурс]: https://www.fastcompany.com/40566786/heres-how-facebook-uses-ai-to-detect-many-kinds-of-bad-content (дата обращения 05.12.2019).
6. Fighting disinformation online // RAND. [Электронный ресурс]: https://www.rand.org/research/projects/truth-decay/fighting-disinformation.html (дата обращения 05.12.2019).
7. Bot and gender detection of Twitter accounts using distortion and LSA / A. Bacciu [et al.] // Conference and Labs of the Evaluation Forum (CLEF 2019). Lugano, Switzerland, 9–12 September 2019. [Электронный ресурс]: http://ceur-ws.org/Vol-2380/paper_210.pdf (дата обращения 03.04.2020).
8. Gamallo P., Almatarneh S. Naive-Bayesian classification for bot detection in Twitter / // Conference and Labs of the Evaluation Forum (CLEF 2019). Lugano, Switzerland, 9–12 September 2019. [Электронный ресурс]: http://ceur-ws.org/Vol-2380/paper_194.pdf (дата обращения 03.04.2020).
9. Vogel I., Jiang P. Bot and gender identification in Twitter using word and character N-grams // Conference and Labs of the Evaluation Forum (CLEF 2019). Lugano, Switzerland, 9–12 September 2019. [Электронный ресурс]: http://ceur-ws.org/Vol-2380/paper_65.pdf (дата обращения 03.04.2020). DOI: 10.13140/RG.2.2.28481.71528.
10. Mahmood A., Srinivasan P. Twitter bots and gender detection using tf-idf // Conference and Labs of the Evaluation Forum (CLEF 2019). Lugano, Switzerland, 9–12 September 2019. [Электронный ресурс]: http://ceur-ws.org/Vol-2380/paper_253.pdf (дата обращения 03.04.2020).
11. Farber M., Qurdina A., Ahmedi L. Identifying Twitter bots using a convolutional neural network // Conference and Labs of the Evaluation Forum (CLEF 2019). Lugano, Switzerland, 9–12 September 2019. [Электронный ресурс]: http://ceur-ws.org/Vol-2380/paper_227.pdf (дата обращения 03.04.2020).
12. Lundberg J., Nordqvist J., Laitinen M. Towards a language independent bot detection // 4th Conference on Digital Humanities in the Nordic Countries (DHN 2019). Copenhagen, Denmark, 5–8 March 2019. P. 308–319.
13. Sahoo S.R., Gupta B.B. Hybrid approach for detection of malicious profiles in Twitter // Computers & Electrical Engineering. 2019. № 76. С. 65–81. DOI: 10.1016/j.compeleceng.2019.03.003.
14. Novotny J. Twitter bot detection & categorization – a comparative study of machine learning methods. Master's thesis in Statistics. Lund: Lund University, 2019.
15. Towards automatic bot detection in Twitter for health-related tasks / A. Davoudi [et al.] // AMIA Joint Summits on Translation Science. 23–26 March 2020. P. 136–141.
16. RTbust: Exploiting temporal patterns for botnet detection on Twitter / M. Mazza [et al.] // 10th ACM Conference on Web Science. Boston, MA, USA, 30 June – 3 July 2019. P. 183–192.
17. Beskow D.M., Carley K.M. Its all in a name: Detecting and labeling bots by their name // Computational and Mathematical Organization Theory. 2019. P. 1–12. DOI: 10.1007/s10588-018-09290-1.
18. Online human-bot interactions: Detection, estimation, and characterization / O. Varol [et al.] // Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017), Montreal, Quebec, Canada, 15–18 May 2017. [Электронный ресурс]: https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587/14817 (дата обращения 03.04.2020).
19. Minnich A., Chavoshi N., Koutra D., Mueen A. BotWalk: Efficient adaptive exploration of Twitter bot networks // 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2017), Sydney Australia, 31 July – 03 August 2017. P. 467–474. DOI: 10.1145/3110025.3110163.
20. Chavoshi N., Hamooni H., Mueen A. DeBot: Twitter bot detection via warped correlation // IEEE 16th International Conference on Data Mining (ICDM 2016). Barcelona, Spain, 12–15 December 2016. P. 817–822. DOI: 10.1109/ICDM.2016.0096.
21. The rise of social bots / E. Ferrara [et al.] // Communications of the ACM. 2016. Vol. 59. No 7. P. 96–104. DOI: 10.1145/2818717.
22. BotOrNot: A system to evaluate social bots / C. Davis [et al.] // 25th International Conference Companion on World Wide Web, Montreal, Canada, 11–15 May 2016. P. 273–275. DOI: 10.1145/2872518.2889302.
23. Thamm M. Analysis: Bell Pottinger more than just spin, its political interference in sovereign states // [Электронный ресурс]: https://www.dailymaverick.co.za/article/2017-07-05-analysis-bell-pottinger-more-than-just-spin-its-political-interference-in-sovereign-states/#gsc.tab=0 (дата обращения 05.12.2019).
24. Featherstone C. South African bot behaviour post the July 2018 Twitter account cull // 2019 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD). Winterton, South Africa, 5–6 August 2019. P. 1–6. DOI: 10.1109/ICABCD.2019.8851039.
25. Аккаунт Twitter Paul Mashatile. [Электронный ресурс]: https://twitter.com/PaulMashatile (дата обращения 09.09.2018).
26. Аккаунт Twitter Ayanda Dlodlo. [Электронный ресурс]: https://twitter.com/MinAyandaDlodlo (дата обращения 09.09.2018).
27. Документация API Twitter. [Электронный ресурс]: http://www.developer.twitter.com/docs (дата обращения 04.04.2019).
28. Воронцов К.В. Математические методы обучения по процедентам (теория обучения машин). [Электронный ресурс]: http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf (дата обращения 04.12.2019).
29. Zhang W., Du T., Wang J. Deep learning over multi-field categorical data // 38th European Conference on Information Retrieval Research (ECIR 2016). Padua, Italy, 20–23 March 2016. P. 45–57.
30. Менисов А.Б., Шастун И.А., Капицын С.Ю. Подход к выявлению вредоносных сайтов сети Интернет на основе обработки лексических признаков адресов (URL) и усредненного ансамбля моделей // Информационные технологии. 2019. Т. 25. № 11. С. 691–697. DOI: 10.17587/it.25.691-697.

31. Воронцов К.В. Лекции по методам оценивания и выбора моделей [Электронный ресурс]: http://www.ccas.ru/voron/download/Modeling.pdf (дата обращения 05.12.2019).

32. Городецкий В.И., Серебряков С.В. Методы и алгоритмы коллективного распознавания: обзор // Труды СПИИРАН. 2006. Т. 1. № 3 С. 139–171. DOI: 10.15622/sp.3.8.

33. Niyogi P., Pierrot J.-B., Siohan O. Multiple classifiers by constrained minimization // 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Istanbul, Turkey, 5–9 June 2000. P. 3462–3465. DOI: 10.1109/ICASSP.2000.860146.

34. Prodromidis A., Chan P., Stolfo S. Meta-learning in distributed data mining systems: Issues and approaches // Advances in Distributed Data Mining. 1999. No 3. P. 81–114.

35. Гнидко К.О., Макаров С.А., Сергеев А.С. Модель интеллектуальной системы поддержки принятия решений в целях выявления негативного информационно-психологического воздействия на обучающихся образовательных организаций Минобороны России и защиты от него // Труды Военно-космической академии имени А.Ф. Можайского. 2019. № 666. С. 142–147.

36. Качура Я.О., Сапрыкин Д.И., Фалеев П.А. Моделирование военнополитической деятельности государств методами ассоциативного анализа в системах поддержки принятия решений // Труды Военно-космической академии имени А.Ф. Можайского. 2018. № 660. С. 19–29.

37. Документация разработчика CatBoost. [Электронный ресурс]: https://tech.yandex.ru/catboost/ (дата обращения 07.12.2019).

Об авторах

Кузьмин Владимир Никифорович

доктор военных наук, профессор;

ведущий научный сотрудник военного института (научно-исследовательского)
Военно-космической академии имени А.Ф. Можайского, 197198, г. Санкт-Петербург, ул. Ждановская, д. 13;

E-mail: kuzmindvn@ya.ru

ORCID: 0000-0002-6411-4336

Менисов Артем Бакытжанович

кандидат технических наук;

научный сотрудник военного института (научно-исследовательского)
Военно-космической академии имени А.Ф. Можайского, 197198, г. Санкт-Петербург, ул. Ждановская, д. 13;

E-mail: men.arty@yandex.ru

ORCID: 0000-0002-9955-2694

Шастун Иван Анатольевич

кандидат технических наук;

преподаватель Военно-космической академии имени А.Ф. Можайского, 197198, г. Санкт-Петербург, ул. Ждановская, д. 13;

E-mail: shastunivan1982@gmail.com

ORCID: 0000-0002-1086-5345

An approach to identifying bots in social networks based on the special association of classifiers

Vladimir N. Kuzmin
E-mail: kuzmindvn@ya.ru

Artem B. Menisov
E-mail: men.arty@yandex.ru

Ivan A. Shastun
E-mail: shastunivan1982@gmail.com

Space Military Academy named after A.F. Mozhaysky
Address: 13, Zhdanovskaya Street, Saint Petersburg 197198, Russia

Abstract

Currently the use of bots, i.e. auto-accounts in social networks which are managed with special programs but disguised as ordinary users, has serious consequences. For example, bots have been used to influence political elections, distort information on the Internet and manipulate prices on the stock exchange. Many research teams concerned with the detection of such accounts have made use of machine learning methods. However, the practical results of detecting social network bots indicate significant limitations because the methodological tools used have language limitation and ineffective criteria for detection. This article presents improved countermeasures in a methodological approach to develop a universal social network account classifier for minimizing the average risk of errors in bot detection. The application of an assembly of classifiers united by a data adaptation criterion and results from the variance of each model found the formation of a universal classifier for social network accounts. The main results obtained by the authors consist of the criteria system and the categorical (nominal) features transformation approach for the formation of the special ensemble of classifiers. In practice, use of the ensemble of classifiers allows us to increase the effectiveness of bot detection compared to other approaches.

Key words: bot detection; social networks; machine learning; ensemble of models; association of classifiers.

Citation: Kuzmin V.N., Menisov A.B., Shastun I.A. (2020) An approach to identifying bots in social networks based on the special association of classifiers. *Business Informatics*, vol. 14, no 3, pp. 54–66.
DOI: 10.17323/2587-814X.2020.3.54.66

References

1. Williamson W., Scrofani J. Trends in detection and characterization of propaganda bots. Proceedings of the *52nd Hawaii International Conference on System Sciences. Honolulu, USA, 8–11 January 2019*, pp. 7118–7123. DOI: 10.24251/HICSS.2019.854.
2. Lukyanov R.V. (2018) Methodology for monitoring the state of information security of automated systems in the context of heterogeneous mass incidents. *Transactions of the Military Space Academy named after A.F. Mozhaysky*, no 660, pp. 111–115 (in Russian).
3. *As many as 48 million Twitter accounts aren't people, says study.* CNBC. Available at: https://www.cnbc.com/2017/03/10/nearly-48-million-twitter-accounts-could-be-bots-says-study.html (accessed 1 December 2019).
4. *Massive networks of fake accounts found on Twitter.* BBC. Available at: http://www.bbc.co.uk/news/technology-38724082 (accessed 1 December 2019).
5. Terdima D. *Here's how Facebook uses AI to detect many kinds of bad content.* Fast Company. Available at: https://www.fastcompany.com/40566786/heres-how-facebook-uses-ai-to-detect-many-kinds-of-bad-content (accessed 5 December 2019).
6. *Fighting disinformation online.* RAND. Available at: https://www.rand.org/research/projects/truth-decay/fighting-disinformation.html (accessed 5 December 2019).
7. Bacciu A., La Morgia M., Nemmi E., Neri V., Mei A., Stefa J. (2019) Bot and gender detection of Twitter accounts using distortion and LSA. Proceedings of the *Conference and Labs of the Evaluation Forum (CLEF 2019). Lugano, Switzerland, 9–12 September 2019.* Available at: http://ceur-ws.org/Vol-2380/paper_210.pdf (accessed 03 April 2020).
8. Gamallo P., Almatarneh S. (2019) Naive-Bayesian classification for bot detection in Twitter. Proceedings of the *Conference and Labs of the Evaluation Forum (CLEF 2019). Lugano, Switzerland, 9–12 September 2019.* Available at: http://ceur-ws.org/Vol-2380/paper_194.pdf (accessed 03 April 2020).
9. Vogel I., Jiang P. (2019) Bot and gender identification in Twitter using word and character N-grams. Proceedings of the *Conference and Labs of the Evaluation Forum (CLEF 2019). Lugano, Switzerland, 9–12 September 2019.* Available at: http://ceur-ws.org/Vol-2380/paper_65.pdf (accessed 03 April 2020). DOI: 10.13140/RG.2.2.28481.71528.
10. Mahmood A., Srinivasan P. (2019) Twitter bots and gender detection using tf-idf. Proceedings of the *Conference and Labs of the Evaluation Forum (CLEF 2019). Lugano, Switzerland, 9–12 September 2019.* Available at: http://ceur-ws.org/Vol-2380/paper_253.pdf (accessed 03 April 2020).
11. Farber M., Qurdina A., Ahmedi L. (2019) Identifying Twitter bots using a convolutional neural network. Proceedings of the *Conference and Labs of the Evaluation Forum (CLEF 2019). Lugano, Switzerland, 9–12 September 2019.* Available at: http://ceur-ws.org/Vol-2380/paper_227.pdf (accessed 03 April 2020).
12. Lundberg J., Nordqvist J., Laitinen M. Towards a language independent bot detection. Proceedings of the *4th Conference on Digital Humanities in the Nordic Countries (DHN 2019). Copenhagen, Denmark, 5–8 March 2019.* P. 308–319.
13. Sahoo S.R., Gupta B.B. (2019) Hybrid approach for detection of malicious profiles in Twitter. *Computers & Electrical Engineering*, no 76, pp. 65–81. DOI: 10.1016/j.compeleceng.2019.03.003.
14. Novotny J. (2019) *Twitter bot detection & categorization – a comparative study of machine learning methods.* Master's thesis in Statistics. Lund: Lund University.
15. Davoudi A., Klein A.Z., Sarker A., Gonzalez-Hernandez A. (2019) Towards automatic bot detection in Twitter for health-related tasks. Proceedings of the *AMIA Joint Summits on Translation Science. 23–26 March 2020*, pp. 136–141.
16. Mazza M., Cresci S., Avvenuti M.,Quattrociocchi W., Tesconi M. (2019) RTbust: Exploiting temporal patterns for botnet detection on Twitter. Proceedings of the *10th ACM Conference on Web Science. Boston, MA, USA, 30 June – 3 July 2019*, pp. 183–192.
17. Beskow D.M., Carley K.M. (2019) Its all in a name: detecting and labeling bots by their name. *Computational and Mathematical Organization Theory*, pp. 1–12. DOI: 10.1007/s10588-018-09290-1.
18. Varol O., Ferrara E., Davis C.A., Menczer F., Flammini A. (2017) Online human-bot interactions: Detection, estimation, and characterization. Proceedings of the *Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017), Montreal, Quebec, Canada, 15–18 May 2017.* Available at: https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587/14817 (accessed 3 April 2020).

19. Minnich A., Chavoshi N., Koutra D., Mueen A. (2017) BotWalk: Efficient adaptive exploration of Twitter bot networks. Proceedings of the *2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2017), Sydney Australia, 31 July – 03 August 2017*, pp. 467–474. DOI: 10.1007/s10588-018-09290-1.

20. Chavoshi N., Hamooni H., Mueen A. (2016) DeBot: Twitter bot detection via warped correlation. Proceedings of the *IEEE 16th International Conference on Data Mining (ICDM 2016). Barcelona, Spain, 12–15 December 2016*, pp. 817–822. DOI: 10.1109/ICDM.2016.0096.

21. Ferrara E., Varol O., Davis C., Menczer F., Flammini A. (2016) The rise of social bots. *Communications of the ACM*, vol. 59, no 7, pp. 96–104. DOI: 10.1145/2818717.

22. Davis C., Varol O., Ferrara E., Flammini A., Menczer F. (2019) BotOrNot: A system to evaluate social bots. Proceedings of the *25th International Conference Companion on World Wide Web, Montreal, Canada, 11–15 May 2016,* pp. 273–275. DOI: 10.1145/2872518.2889302.

23. Thamm M. (2019) *Analysis: Bell Pottinger more than just spin, its political interference in sovereign states*. Available at: https://www.dailymaverick.co.za/article/2017-07-05-analysis-bell-pottinger-more-than-just-spin-its-political-interference-in-sovereign-states/#gsc.tab=0 (accessed 5 December 2019).

24. Featherstone C. (2019) South African bot behaviour post the July 2018 Twitter account cull. Proceedings of the *2019 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD). Winterton, South Africa, 5–6 August 2019*, pp. 1–6. DOI: 10.1109/ICABCD.2019.8851039.

25. *Twitter account Paul Mashatile*. Available at: https://twitter.com/PaulMashatile (accessed 4 September 2019).

26. *Twitter account Ayanda Dlodlo*. Available at: https://twitter.com/MinAyandaDlodlo (accessed 4 September 2019).

27. *Twitter API Documentation*. Available at: http://www.developer.twitter.com/docs (accessed 4 April 2019).

28. Vorontsov K.V. (2019) *Mathematical methods of teaching by procedures (theory of machine learning)*. Available at: http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf (accessed 4 December 2019) (in Russian).

29. Zhang W., Du T., Wang J. (2016) Deep learning over multi-field categorical data. Proceedings of the *38th European Conference on Information Retrieval Research (ECIR 2016). Padua, Italy, 20–23 March 2016*, pp. 45–57.

30. Menisov A.B., Shastun I.A., Kapitsyn S.U. (2019) An approach to the identification of malicious Internet sites based on the processing of lexical signs of addresses (URLs) and an average ensemble of models. *Information Technologies*, vol. 25, no 11, pp. 691–697 (in Russian). DOI: 10.17587/it.25.691-697.

31. Vorontsov K.V. (2019) *Lectures on methods for evaluating and selecting models*. Available at: http://www.ccas.ru/voron/download/Modeling.pdf (accessed 5 December 2019) (in Russian).

32. Gorodetsky V.I., Serebryakov S.V. (2006) Collective recognition methods and algorithms: a review. *Transactions of SPIIRAS*, vol. 1, no 3, pp. 139–171 (in Russian). DOI: 10.15622/sp.3.8.

33. Niyogi P., Pierrot J.-B., Siohan O. (2000) Multiple classifiers by constrained minimization. Proceedings of the *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Istanbul, Turkey, 5–9 June 2000*, pp. 3462–3465. DOI: 10.1109/ICASSP.2000.860146.

34. Prodromidis A., Chan P., Stolfo S. (1999) Meta-learning in distributed data mining systems: Issues and approaches. *Advances in Distributed Data Mining*, no 3 pp. 81–114.

35. Gnidko K.O., Makarov S.A., Sergeev A.S. (2019) A model of an intellectual decision support system in order to identify the negative informational and psychological impact on students of educational organizations of the Ministry of Defense of Russia and to protect against it. *Transactions of the Military Space Academy named after A.F. Mozhaysky,* no 666, pp. 142–147 (in Russian).

36. Kachura Ya.O., Saprykin D.I., Faleev P.A. (2018) Modeling of the military-political activity of states by the methods of associative analysis in decision support systems. *Transactions of the Military Space Academy named after A.F. Mozhaysky,* no 660, pp. 19–29 (in Russian).

37. *Developer Documentation CatBoost*. Available at: https://tech.yandex.ru/catboost/ (accessed 7 December 2019).

About the authors

Vladimir N. Kuzmin

Dr. Sci. (Mil.), Professor;

Leading Researcher, Military Institute (Science and Researching),
Space Military Academy named after A.F. Mozhaysky, 13, Zhdanovskaya Street, Saint Petersburg 197198, Russia;

E-mail: kuzmindvn@ya.ru

ORCID: 0000-0002-6411-4336

Artem B. Menisov

Cand. Sci. (Tech.);

Researcher, Military Institute (Science and Researching),
Space Military Academy named after A.F. Mozhaysky, 13, Zhdanovskaya Street, Saint Petersburg 197198, Russia;

E-mail: men.arty@yandex.ru

ORCID: 0000-0002-9955-2694

Ivan A. Shastun

Cand. Sci. (Tech.);

Lecturer, Military Institute (Science and Researching),
Space Military Academy named after A.F. Mozhaysky, 13, Zhdanovskaya Street, Saint Petersburg 197198, Russia;

E-mail: shastunivan1982@gmail.com

ORCID: 0000-0002-1086-5345