УДК 372.8+004.9
DOI: http://dx.doi.org/10.21686/1818-4243-2025-4-64-74

Е.А. Останина[1], Е.В. Поколодина[2]

[1] Московский авиационный институт, Москва, Россия,
[2] Российский экономический университет им. Г.В. Плеханова, Москва, Россия

Оценка цифрового следа участника интернет-взаимодействий с применением методов математической статистики

Цель исследования. *В современном мире каждый человек является активным пользователем ресурсов, размещенных в глобальной сети. Заходя с разных устройств (компьютеров, планшетов, телефонов) мы посещаем множество электронных страниц, ведем переписку и просматриваем контент оставляя каждый раз свой электронный след в системе. Целью исследования является определение такого следа, его фиксация и обработка его с использованием методов математической статистики с целью получения персонифицированных и обобщенных данных для различных отраслей. Также на основе полученных результатов предполагается формирование рекомендаций по защите данных цифрового следа от возможного злонамеренного использования. В работе рассматривается понятие и подробно описаны источники цифрового следа участника интернет-взаимодействий, обозначены методы и инструменты сбора и предобработки информации. Раскрыто понятие активной и пассивной составляющих цифрового отпечатка. Акцентируется внимание на необходимости отнесения информации, содержащейся в цифровом следе, к персональным данным, так как она в совокупности может порой определять личность точнее документа, что, в свою очередь, требует законодательной защиты в части сбора, хранения, обработки и дальнейшего использования. В статье приведены примеры использования информации цифрового отпечатка в различных отраслях: криминалистике, маркетинговых исследованиях, формировании таргетированной рекламы и др.*

Материалы и методы. *В статье достаточно подробно рассмотрены различные методы математической статистики для оценки цифрового следа пользователя сети, приведены этапы, позволяющие осуществить сбор, предобработку и анализ информации, содержащейся в цифровом отпечатке, с применением языков программирования Python, R и различных программных инструментов. Рассмотрено использование статистических методов анализа, применение матрицы взаимодействий, возможности визуализации, алгоритмов машинного обучения, рекомендательных систем, сетевых метрик, временных рядов и методы анализа чувствительности.*

Результаты. *Несмотря на то, что данные полученные с использованием методов математической статистики могут быть полезны для бизнеса, правительственных организаций, научно-исследовательских институтов и даже правоохранительных органов, данные собираемые в процессе деятельности в глобальной сети и имеющие привязку к конкретной личности могут нести явную угрозу в случае их использования злоумышленниками. Эти методы оценки доступны как для конструктивной тенденции по улучшению нашей жизни путем подбора персонифицированного релевантного контента и улучшения сервиса, так и для деструктивной деятельности через манипуляторные действия и шантаж. В связи с этим знание оставляемого следа и возможностей использования этих данных в первоначальном и обработанном виде позволяет улучшить качество жизни, с одной стороны, и защитить персональную информацию с другой.*

Заключение. *Обращается внимание на необходимость информационного контроля своего цифрового следа, в связи с чем в статье приводится ряд рекомендаций по осуществлению безопасного интернет-взаимодействия.*

Ключевые слова: *цифровой след, цифровая тень, цифровой отпечаток, персональные данные, источники цифровой информации, методы анализа данных цифрового следа, защита данных.*

Elena A. Ostanina[1], Elena V. Pokolodina[2]

[1] Moscow Aviation Institute, Moscow, Russia,
[2] Plekhanov Russian University of Economics, Moscow, Russia

Assessment of the Digital Footprint of a Participant in Internet Interactions Using Mathematical Statistics Methods

The purpose of the study. *In the modern world, everyone is an active user of resources hosted on the global network. Logging in from different devices (computers, tablets, phones), we visit many electronic pages, correspond and view content, leaving our electronic footprint in the system each time. The purpose of the study is to identify such a trace, record it and process it using mathematical statistics methods in order to obtain personalized and generalized data for various industries. Also, based on the results obtained, recommendations are expected to be formed to protect digital footprint data from possible malicious use.*

The paper examines the concept and describes in detail the sources of a digital trace of a participant in Internet interactions, identifies methods and tools for collecting and pre-processing information. The concept of active and passive components of a digital fingerprint is disclosed. Attention is focused on the need to classify the information contained in the digital trace as personal data, since it can sometimes determine a person more accurately than a document, which, in turn, requires legislative protection in terms of collection, storage, processing and further use. The article provides examples of using digital fingerprint information in various industries: criminalistics, marketing research, targeted advertising, etc.

Materials and methods. *The article discusses in sufficient detail various methods of mathematical statistics for assessing the digital footprint of a network user, and provides steps that allow for the*

collection, preprocessing and analysis of information contained in a digital fingerprint using the Python, R programming languages and various software tools. The use of statistical analysis methods, the use of an interaction matrix, visualization capabilities, machine learning algorithms, recommendation systems, network metrics, time series, and sensitivity analysis methods are considered.

Results. *Despite the fact that data obtained using mathematical statistics methods can be useful for businesses, government organizations, research institutes, and even law enforcement agencies, data collected in the course of activities on a global network and linked to a specific individual can pose a clear threat if used by intruders. These assessment methods are available both for constructive tendencies to improve our lives by selecting personalized relevant content and improving the service, and for destructive activities through manipulative actions and blackmail. In this regard, knowledge of the trace left and the possibilities of using this data in its original and processed form makes it possible to improve the quality of life, on the one hand, and protect personal information, on the other.*

Conclusion. *Attention is drawn to the need for information control of one's digital trace, in connection with which the article provides a number of recommendations for the implementation of safe Internet interaction.*

Keywords: *digital trace, digital shadow, digital fingerprint, personal data, sources of digital information, methods of analyzing digital trace data, data protection.*

Практически любой человек в современном мире по желанию, а иногда вынужденно, обращается к возможностям предоставляемым всемирной паутиной, будь то получение новостной информации, взаимодействие с органами государственного управления и получение услуг, использование прочих Интернет-ресурсов или социальное сетевое общение. Все эти действия могут быть зафиксированы и использованы как во благо, так и во вред человеку. Во множестве электронных систем фиксируется все больше и больше данных о каждом из нас, о нашей деятельности, как активной, так и пассивной, в виртуальном пространстве [1]. Развитие вычислительных мощностей позволяет накапливать, структурировать и систематизировать эту информацию, что в конечном итоге позволяет оперативно получать ее о каждом человеке из разных источников и с целью осуществления комплексного анализа [2]. Совокупность подобной информации часто называют цифровым следом (отпечатком) или цифровой тенью человека. Под данными терминами понимается уникальный набор действий в Интернете или на любом из используемых цифровых устройств. Причем этот набор может включать массу действий и показателей [3]. Например,

в области электронной коммерции (онлайн покупки): совершение покупок, временной интервал действий и их регулярность, состав, стоимость, предпочитаемые интернет-магазины, создание учетных записей, подписки на купоны и участие в акциях, загрузка и использование приложений и др.;

применение интернет банкинга: установка и активность мобильного банковского приложения, открытие счетов и карт, кредитные истории, операции с акциями и валютой (в том числе крипто валютные операции) [4], подписки на финансовые публикации и блоги;

в сфере социального взаимодействия: регистрация, использование и его активность в социальных сетях и мессенджерах на компьютере и мобильных устройствах, временные рамки активности, переход на другие ресурсы с использованием учетных записей социальной сети, время и частота общения с друзьями и знакомыми, и даже содержание этого общения в результате обмена текстовой, аудиоинформацией и фото- и видеоматериалами. Так по сопровождающим файл метаданным можно узнать производителя и модель камеры на которую велась съемка, ее дату и время, местоположение съемки и другую пользовательскую информацию. Отметим, что в социальных сетях пользователи добровольно размещают информацию о себе заполняя анкеты о местах учебы и работы, своих интересах и предпочтениях, добавляют друзей и вступают в сообщества [5]. На основе чего посредством специальных алгоритмов производится анализ изображений, текстов, связей с другими людьми, анализ лайков и просмотров. Даже эмодзи (графический язык для передачи эмоций в сети, где вместо слов используются сочетания картинок) позволяют передать невербальную составляющую общения, хотя стоит отметить, что не всегда в полной мере;

интерес к новостной информации позволяет составить некоторый портрет человека определяя его интересы, зачастую гражданскую позицию [6]. Этому способствует отслеживание подписок на выбранные новостные издания в Интернете, тематика и выбор статей, их авторов, время и длительность просмотра, репосты и комментирование, подписки на рассылки;

поисковые запросы и переходы на сайты практически повсеместно сопровождаясь сообщениями о сборе cookies-файлов, и если пользователь перешел на сайт с использованием зарегистрированного профиля, то все запросы в поисковой системе, все опубликованное им войдет в состав активного цифрового следа [7];

использование умных устройств (фитнес-трекеров, IoT устройств) позволит оценить состояние здоровья, физическую активность, предпочтения в комфорте и даже распорядок дня [8]. Также информацию принесет использование приложения для получения медицинской помощи, регистрация адреса электрон-

ной почты и указание номера телефона в фитнес-клубе, подписки на блоги о здоровье, правильном питании и т.д.;

использование приложений с геолокацией позволит собрать информацию о всех местоположениях пользователя и различные моменты времени с учетом продолжительности нахождения, предпочитаемых маршрутах перемещений и видах транспорта [9].

Помимо, упомянутой в одном из примеров, активной составляющей цифрового следа, стоит обращать особое внимание на пассивную его составляющую, когда информация о пользователе собирается полностью без нашего ведома, когда собирается и анализируется информация о нашей активности с учетом всех переходов, времени, действий, IP-адресов и т.д.

Таким образом, информация о нас собирается постоянно и огромные массивы данных, позволяя выявить и установить разного рода закономерности, могут накапливаться, представлять собой определенную ценность и могут быть использованы, в том числе злоумышленниками.

Следует отметить значимость такой информации и обратиться к статье 3 Федерального закона от 27.07.2006 N 152-ФЗ (ред. от 08.08.2024) «О персональных данных», где определено понятие персональные данные. К ним причисляется любая информация, относящаяся к прямо или косвенно определенному, или определяемому физическому лицу (субъекту персональных данных). Многие аспекты цифрового следа содержат или могут содержать персональные данные [10]. И хотя, казалось бы, понятие цифрового следа, в части именно такой терминологии, строго законодательно не отнесено к персональным данным как, например, паспортные данные, номера СНИЛС и ИНН, эту информацию следует рассматривать именно как персональную. В совокупности она определяет личность даже точнее документа и требует законодательной защиты в части сбора, хранения и использования. Для пользователей важно быть осведомлёнными о том, какие данные они оставляют в сети, и как они могут управлять своими цифровыми следами для повышения уровня своей конфиденциальности.

В настоящее время цифровые отпечатки активно используются во многих научных исследованиях и различных отраслях: при приеме на работу, в психологии, социологии, юриспруденции, криминалистике, маркетинговых исследованиях и формировании таргетированной рекламы. Так, например, много статей в настоящее время посвящено применению данных цифрового следа в криминалистике. Эти примеры описаны А.В. Ростовцевым [11], Р.А. Дерюгиным [12], Н.Н. Ахтырской [13], О.Г. Карнауховой [14] и другими исследователями. Вопросам анализа цифровых следов в образовании посвящены работы Н.Л. Казариновой [15], А.А. Логиновой [16], С.А. Нестерова [17], А.С. Осановой [18], Е.Н. Черемисиной [19] и других. В тоже время с использованием данных наших поисковых запросов успешно работают всевозможные рекламные интернет агентства продвигая те или иные товары или услуги [20].

Отметим важность цифровых следов, обусловленную их относительным постоянством, что связано с нашими интересами, предпочтениями и практически неосуществимостью контроля над однажды опубликованной информацией (как в дальнейшем она будет использована другими пользователями Интернета). Также с цифровым следом в настоящее время связывают репутацию человека, что накладывает отпечаток на отношение к нему как потенциальных работодателей, так и коллег, знакомых. Опубликованные в сети фотографии могут иметь неоднозначную оценку в обществе, могут быть модифицированы злонамерено, что может иметь серьезные последствия как для человека, так и его окружения. В тоже время цифровое след может быть использован злоумышленниками для получения доступа к учетным записям или для создания фейковых профилей на основе полученных данных.

Таким образом, цифровой след позволяет воссоздать портрет практически каждого из участников интернет-взаимодействий. Основываясь на фиксируемых данных можно оценить уровень мышления и мировоззрение человека, составить представление о его интересах и потребностях, определить социальный статус и текущее психоэмоциональное состояние. В настоящее время уже можно составить цифровой портрет каждой личности, который возможно будет иметь некоторые отличия от реального, например, в том случае, когда человек намеренно будет достаточно часто «примерять на себя» несвойственную ему роль. Однако, в современном мире делать это становится все сложнее, и часто такая «созданная роль» ограничивается определенными площадками комментариев, обсуждений, форумами, однако алгоритмы использующие большие массивы данных уже способны определять ролевые моменты и текущее взаимодействие с подавляющим большинством ресурсов, характерное для той или иной личности.

К сожалению, постоянно обращаясь к ресурсам глобальной сети, используя все преимущества цифрового пространства мы не задумываемся об оставляемых отпечатках наших личностей, о том, что они могут стать интересны злоу-

мышленникам. В настоящее время отмечается рост числа инцидентов, связанных с утечками данных, в том числе связанных с цифровыми следами пользователей. Эти данные продаются и покупаются для самых разных целей, например, при оценке потребительских предпочтений, политической обстановки и социальной напряженности в обществе. Эти данные могут содержать персональные данные, в том числе которые могут позволить формировать наше отношение к тем или иным событиям (подбирая соответствующие факты и статьи), осуществить доступ к личным аккаунтам и банковским вкладам. Однако, даже с учетом всех законодательных инициатив по регламентации работы с персональными данными, проблема защиты цифрового следа, как одной из составляющих персональных данных пользователя глобальной сети, стоит достаточно остро. В результате, каждому из участников интернет взаимодействия необходимо учитывать возможность сбора данных о каждом из своих действий при использовании цифровых устройств [21].

№	Возраст	Сгруп. Возраст	Критерий	Количество сетей	ВК	ТТ	ЮТ	ТТ	ТВ
1	13-18	13-18	1	3	1	1	1	0	0
2	13-18	18-30	1	3	1	1	1	0	0
3	13-18		1	1	1	0	0	0	0
4	18-30		2	5	1	1	1	1	1
5	13-18		1	3	1	1	1	0	0
6	13-18		1	5	1	1	1	1	1
7	13-18		1	4	1	1	1	0	1
8	18-30		2	5	1	1	1	1	1
9	13-18		1	2	1	1	0	0	0
10	18-30		2	4	1	1	1	0	1
11	18-30		2	5	1	1	1	1	1
12	18-30		2	5	1	1	1	1	1
13	18-30		2	3	1	1	1	0	0
14	18-30		2	2	1	1	0	0	0
15	18-30		2	5	1	1	1	1	1
16	18-30		2	4	1	1	1	0	1
17	18-30		2	4	1	1	1	1	0
18	18-30		2	4	1	1	1	1	0
19	18-30		2	5	1	1	1	1	1
20	18-30		2	4	1	1	1	1	0
21	18-30		2	3	1	1	0	0	1
22	13-18		1	4	1	1	1	1	0
23	18-30		2	2	0	1	0	0	1
24	13-18		1	2	1	1	0	0	0
25	18-30		2	1	0	1	0	0	0
26	13-18		1	5	1	1	1	1	1
27	18-30		2	1	0	1	0	0	0
28	18-30		2	2	1	1	0	0	0
29	13-18		1	1	0	1	0	0	0
30	18-30		2	2	1	1	0	0	0
31	13-18		1	3	1	1	1	0	0

Рис. 1. Сведения о посещении социальных сетей (возрастные категории 13-18 лет, 18-30 лет)

Fig. 1. Information about visiting social networks (age categories 13-18 years, 18-30 years)

Цифровой след пользователя можно оценить с помощью различных методов, в том числе относящихся к методам математической статистики. Так исследования данного вопроса в части изучения освоения образовательного маршрута велись Е.В. Барановой и Г.В. Шевцовым [22], в соответствии с моделью деятельности человека Е.В. Поколодиной [23]. Обобщая эти исследования, а также исследования других авторов зафиксируем этапы сбора данных.

На первом этапе при осуществлении сбора данных определяется доступный сегмент данных, отвечающих поставленным целям. Например, это могут быть данные об активности пользователя на сайте, время, проведенное на страницах, клики и т.д. В работе в качестве примера рассмотрим данные посещения социальных сетей в разных возрастных группах. Данные сведены в таблицу, фрагмент которой представлен на рисунке 1.

Для сбора и предобработки данных цифрового следа важна цель работы и тип данных, что и предопределяет применение тех или иных программных инструментов. Эти процессы будут содержать следующие этапы:

отправку HTTP-запроса web-серверу поисковика с отметкой об интересующем объекте;

получение ответной ссылки на страницу в сети с информацией об интересующем объекте и отправка HTTP-запроса на получение кода этой страницы;

выявление из полученного кода информации с упоминанием объекта (вручную сбор в определенных сегментах или автоматически проверка на упоминание объекта);

непосредственная предобработка. Например, предобработка текста может включать в себя очистку от ненужных символов и токенизацию, а именно, разделение текста на отдельные слова, знаки, символы; предобработка числовых данных включать обработку пропусков (недостающих показателей) и нормализацию; при предобработке изображений использовать форматирование.

Отметим, что для автоматизированного сбора и систематизации информации из открытых источников с помощью скриптов (парсинга) в настоящее время существует достаточно много библиотек для разных языков программирования (Python, JavaScript, PHP и др. Очень простыми библиотеками Python для начинающих являются «Requests» и «Beautiful Soup». Результаты работы по обработке данных, представленных на рис.1 с использованием библиотек Numpy, Pandas, matplotlib представлены на рисунке 2, 3.

```python
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

df = pd.read_excel("AL1.xlsx")
df.columns = df.columns.str.strip()
social_networks = ['ВК', 'ТГ', 'ЮТ', 'ТТ', 'ТВ']
def get_average_age(age_group):
    age_range = age_group.split('-')
    return (int(age_range[0]) + int(age_range[1])) / 2
df['Средний возраст'] = df['Возраст'].apply(get_average_age)
plt.figure(figsize=(12, 6))
age_groups = df['Возраст'].unique()
age_group_counts = {group: {net: df[(df['Возраст'] == group) & (df[net] == 1)].shape[0] for net in social_networks} for group in age_groups}
width = 0.15
x = np.arange(len(age_groups))

for i, net in enumerate(social_networks):
    plt.bar(x + i * width, [age_group_counts[group].get(net, 0) for group in age_groups], width=width, label=net)
plt.xlabel('Возрастная группа')
plt.ylabel('Количество пользователей')
plt.title('Статистика посещений социальных сетей по возрастным группам')
plt.xticks(x + width * 2, age_groups)
plt.legend()
plt.tight_layout()
plt.show()
plt.figure(figsize=(8, 8))
network_counts = [df[df[net] == 1].shape[0] for net in social_networks]
plt.pie(network_counts, labels=social_networks, autopct='%1.1f%%', startangle=90)
plt.title("Диаграмма посещения социальных сетей")
plt.axis('equal')
plt.show()
```

Рис. 2. Код программы для систематизации данных о посещении социальных сетей
Fig. 2. Program code for systematizing data on visits to social networks

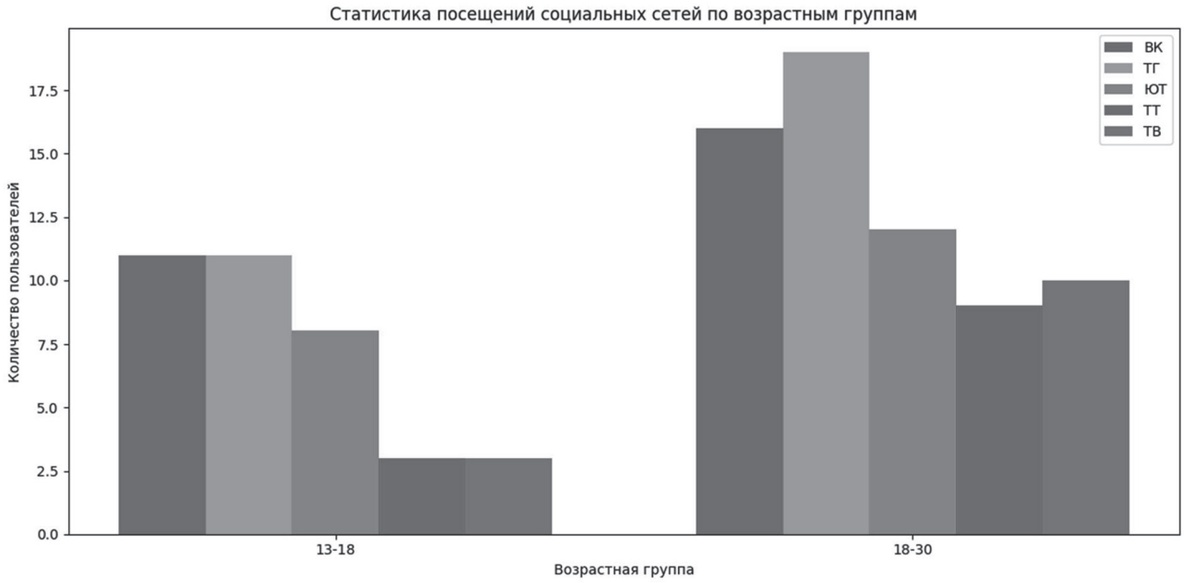

Рис. 3. Визуальное представление статистики посещений социальных сетей по исследуемым группам
Fig. 3. Visual representation of statistics of social network visits by study groups

Сбор цифрового следа также возможен при программном взаимодействии с другими сервисами через их API. Например, сбор данных из Google с помощью Google Search API, что позволяет выполнять поисковые запросы и получать результаты в структурированном формате, сбор данных из YouTube можно осуществить с помощью YouTube Data API. При предобработке также можно использовать специализированные библиотеки Python: для очистки текста можно использовать встроенную библиотеку «String», для токенизации удобно использовать «NLTK»; при работе с числовыми значениями для удаления пропущенных значений или замены их на среднее или медианное используют

«NumPy», для нормализации «Scikit learn», «Pandas»; простое форматирование изображений можно выполнять с помощью библиотеки «Pillow».

При анализе собранных данных может быть использован статистический анализ, позволяющий определить, например, среднее время, проведенное на сайте, число посещенных страниц, процент возврата пользователей на сайт, число переходов по тем или иным ссылкам. Широкое применение в области статистических вычислений и анализа данных имеет язык программирования R. Существует огромное количество библиотек и расширений, доступных для использования и обслуживания различных областей науки, которые предоставляют готовые функции для обработки и визуализации данных, проведения быстрых статистических операций и распознавания текстов.

Визуализировать и анализировать взаимодействия пользователя с контентом позволит создание матрицы взаимодействия. Так, например, матрица где строки будут представлять собой пользователей, а столбцы их соответствующие действия (клики, лайки, комментарии) позволит наглядно представить всю картину и дать оценку активности и в сочетании с другими данными о пользователях выявить целевую аудиторию, определить предпочтения и интересы. По количеству и типу лайков и комментариев можно понять, какие темы или продукты интересуют пользователя; их частота может указывать на уровень активности и вовлеченности, комментарии и реплики показать взаимодействие с другими пользователями, а также формировать представление о социальной динамике в сети; содержащие эмоциональную окраску комментарии позволят понять реакцию на определенный контент (позитивно, негативно, нейтрально); анализ кликов на рекламу или продукты помогает выявить, какие товары или услуги пользуются спросом; сопоставление активности пользователей с демографической информацией (например, возраст, пол, местоположение) помогает уточнить таргетирование аудитории; изучение времени активности может помочь в выборе наилучших периодов публикации контента; отслеживание кликов по ссылкам указывает на то, как пользователь исследует контент и какие источники или платформы он предпочитает.

Использование машинного обучения при оценке цифрового следа каждого пользователя предполагает реализацию кластеризации с помощью алгоритмов K-means и DBSCAN, когда пользователей группируются по схожим паттернам поведения. При этом K-means используется, когда заранее известно количество кластеров, которое нужно обнаружить в данных. Этот алгоритм хорошо работает с линейно разделимыми данными, когда можно провести линии, разделяющие кластеры. DBSCAN же эффективен, когда данные содержат объекты, не принадлежащие ни к одной группе, и кластеры имеют сложную форму. В отличие от K-means, DBSCAN не требует предварительного знания количества кластеров, оно устанавливается автоматически на основе плотности точек данных. Таким образом, если задача требует определения точного количества групп, K-means может быть более подходящим, но если данные не могут быть разделены линейно, то метод DBSCAN имеет преимущества.

Для оценки цифрового следа и предсказания поведения пользователя на основе его предыдущих действий можно использовать различные типы рекомендательных систем. Например, системы контентных рекомендаций (Content-Based Filtering) которые анализируют предпочтения пользователя, основанные на его предыдущих взаимодействиях и сопоставляют их со свойствами предметов (например, товаров, фильмов или статей). Метод же коллаборативной фильтрации (Collaborative Filtering) основывается на поведении других пользователей с похожими предпочтениями и предполагает два основных подхода: User-Based, при котором рекомендации основаны на действиях пользователей с похожими вкусами, и Item-Based когда рекомендации основываются на сходстве между предметами, которые пользователь ранее оценивал или с которыми взаимодействовал. В тоже время гибридные системы (Hybrid Models) комбинируют элементы контентного и коллаборативного подходов, что позволяет улучшить качество рекомендаций, минимизируя недостатки каждого из методов в отдельности. Выстраивание рекомендаций на основе моделей машинного обучения (Machine Learning-Based Recommendations) позволяет осуществлять предсказания предпочтений пользователя на основе более сложных паттернов. Это может включать анализ временных рядов, кластеризацию пользователей и предметов. Системы, основанные на контекстуальных данных (Contextual Recommendations) учитывают контекст, в котором происходит взаимодействие (например, время суток, местоположение, устройство), что может значительно повысить релевантность рекомендаций.

Системы, использующие технологии глубокого обучения (Deep Learning) могут использоваться для обработки больших объемов данных и выявления сложных паттернов в пользовательском поведении. Например, рекуррентные нейронные сети (RNN) могут обрабатывать последователь-

ности действий пользователя.

Системы, основанные на графах (Graph-Based Recommendations) используются для анализа сложных взаимосвязей между пользователями и предметами. Графы позволяют выявлять скрытые связи и паттерны взаимодействий. В этом случае цифровой след рассматривается как граф, где пользователи являются узлами, а ребрами их взаимодействия. Здесь можно применять такие метрики, как центральность (например, по степени, посредничеству, близости и собственному вектору), чтобы оценить влияние или активность пользователя в сети.

При наличии временных данных о действиях пользователя, целесообразно использовать модели временных рядов (например, ARIMA) для прогнозирования будущих действий на основе данных о прошлых.

Математическая оценка позволяет выявлять и определять различные индексы и коэффициенты. Например, «индекс активности пользователя», может учитывать такие параметры, как частота посещений, взаимодействия с контентом и т.д.

Анализ чувствительности в контексте цифрового следа позволяет оценить, каким образом изменения в поведении пользователей (например, увеличение или уменьшение активности в социальных сетях, частота публикаций или уровень взаимодействия с контентом) влияют на целостность оценки их цифрового следа. В качестве методов анализа чувствительности можно использовать:

регрессионный анализ. Например, проанализировав, как изменение количества публикаций и лайков повлияет на оценку репутации;

анализ временных рядов. Например, сравнив хронологию данных о поведении пользователей и оценках цифрового следа во времени, можно выявить тренды и аномалии в результате изменения активности;

сравнительный анализ. Например, осуществить сравнение пользователей с разной активностью путем создания групп с высокой и низкой активностью и изучения их цифрового следа;

методы машинного обучения. Например, использовать алгоритмы, такие как деревья решений или случайный лес, для анализа влияния различных факторов (количества постов, комментариев и т. д.) на итоговую оценку;

сценарный анализ. Например, смоделировав различные сценарии (например, уменьшение активности на 50%) оценить, как это повлияет на цифровой след.

Помимо рассмотренного ранее использования языка программирования Python с библиотеками «Pandas» для обработки данных, «Scikit learn» для машинного обучения, «Statsmodels» для статистического анализа и языка программирования R для статистического анализа и визуализации с использованием пакетов dplyr, ggplot2 и caret для регрессий и моделей можно использовать для базового анализа данных, построения графиков, создания простых моделей регрессии MS Excel или Open Office Calc.

Использование Business Intelligence (BI) систем, например, Tableau или Power BI, позволяет обрабатывать данные и представлять их в удобном для пользователя виде. Например, в виде отчетов, панелей мониторинга, диаграмм и графиков. Инструменты BI позволяют быстро и удобно получать доступ к различным типам данных – историческим и текущим, сторонним и собственным, а также к полуструктурированным (электронные письма, файлы в формате JavaScript Object Notation (JSON), XML, архивные файлы, цифровые фотографии) и неструктурированным данным (информация с сайтов и социальных сетей).

Специализированные инструменты для анализа социальных медиа, например, Hootsuite, Brandwatch, Sprout Social и т.д., могут предоставить специфическую информацию и метрики о поведении пользователей и его влиянии на показатели.

Системный подход к анализу чувствительности позволяет глубже понять влияние поведения пользователя на его цифровой след. Это позволяет достигать различных целей, например, при оптимизации деятельности и изменении стратегии маркетинга или управления репутацией.

Отметим, что вышеуказанные подходы можно комбинировать для более полной и точной оценки цифрового следа пользователя. Однако необходимо учитывать, что методы могут варьироваться в зависимости от доступных данных и целей анализа.

Таким образом, рассмотренные методы математической статистики позволяют с использованием современных мощностей вычислительной техники собрать достаточно полный цифровой след составить участника интернет-взаимодействий, проанализировать его и достаточно точно спрогнозировать поведение пользователя на основе его предыдущих действий. Однако достаточно важным в современном мире является безопасное использование получаемых данных. Как уже отмечалось ранее цифровой след содержит конфиденциальную информацию, а именно персональные данные пользователей, что может представлять интерес для злоумышленников и требует соответствующей защиты как со стороны государства, так и со стороны самой личности пользователя. В связи с этим, с учетом описанных возможных способах сбора и оценки циф-

ровых следов, с учетом опубликованных в [24] материалов, следует обозначить основные рекомендации по защите личной информации и управлению репутацией в сети:

целесообразно периодически отслеживать представленную общедоступную информацию о себе и в случае неудовлетворительного результата связываться с администраторами ресурсов с целью удаления данной информации;

минимизировать количество источников информации, содержащих личную информацию (например, номер телефона, электронную почту, адрес и т.д.), а также минимизировать объем такой предоставляемой информации (например, не заполнять необязательные поля в формах, критически оценивать необходимость их заполнения);

проверять параметры конфиденциальности (например, использовать возможность ограничения видимости публикаций в социальных сетях);

избегать публикаций излишней информации в социальных сетях (например, планов поездок, минимизировать персональные данные в разделе «Информация», не ставить лайки организациям, получившим ваши учетные записи);

избегать незащищенных web-сайтов;

не указывать личные данные при использовании публичных сетей Wi-Fi;

удалять старые учетные записи, что снижает вероятность утечки данных;

создавать надежные пароли, не использовать один пароль для всех учетных записей и регулярно их менять;

сохранять конфиденциальность медицинских и банковских документов;

поддерживать актуальность программного обеспечения для своевременного устранения возможных уязвимостей в нем;

настраивать используемые мобильные устройства (устанавливать пароли для активации, внимательно ознакомиться с пользовательскими соглашениями);

всегда оценивать материалы перед публикацией, в том числе, для создания положительного цифрового образа;

незамедлительно принимать меры при компрометации учетных данных.

В условиях стремительного развития информационного пространства, каждая наша активность оставляет след, создавая объемные базы данных, которые могут быть использованы для анализа поведения и предпочтений пользователей. Исследование данных, характеризующих цифровой след позволяют получить множество сведений, позволяющих расширить наши познания о современном человеке и обществе, больше узнать о поведении пользователей сети, об их интересах и предпочтениях. Эти данные можно использовать в дальнейшем для оптимизации процессов и улучшения качества обслуживания, выявления асоциальных личностей и раскрытии преступлений, подбора кадров и формирования установок в обществе.

Использование различных методов математической статистики при оценке цифрового следа участника интернет-взаимодействий позволяет получить достаточно полную информацию о каждом человеке в отдельности и о группах пользователей в целом, что позволяет не только анализировать поведение, но и выявляя скрытые закономерности и тренды предсказывать будущие действия. Однако, вместе с тем возникает необходимость в обеспечении конфиденциальности и защиты персональных данных пользователей. Отметим, что междисциплинарный подход, объединяющий математику, информатику и социологию, открывает новые горизонты для развития маркетинга, управления, безопасности и социального анализа.

Литература

1. Гайдаш О.В. Феномен цифрового следа в современном обществе [Электрон. ресурс] // Вестник магистратуры. 2020. № 6 (105). Режим доступа: https://cyberleninka.ru/article/n/fenomen-tsifrovogo-sleda-v-sovremennom-obschestve.

2. Фомичева Т.Л. Цифровой след и цифровая гигиена // Самоуправление. 2023. № 6 (139). С. 361–363.

3. Буйнов Д.О. Развитие научных подходов к понятию «цифровые следы» как объекта экспертного исследования // Законы России: опыт, анализ, практика. 2023. № 3. С. 17–22.

4. Семикаленова А.И. Цифровые следы: назначение и производство экспертиз [Электрон. ресурс] // Вестник Университета имени О.Е. Кутафина. 2019. № 5(57). Режим доступа: https://cyberleninka.ru/article/n/tsifrovye-sledy-naznachenie-i-proizvodstvo-ekspertiz.

5. Шестакова А.А. Цифровая личность: границы и барьеры коммуникативных практик в сетевом взаимодействии // VIII Международная социологической Грушинской конференции «Социолог 2.0: трансформация профессии». М.: Всероссийский центр изучения общественного мнения, 2018. С. 422–425.

6. Мещеряков В.А. «Виртуальные следы» под «Скальпелем Оккама» [Электрон. ресурс] // Информационная безопасность регионов. 2009. № 1. Режим доступа: https://cyberleninka.ru/article/n/virtualnye-sledy-pod-skalpelem-okkama.

7. Багмет А.М., Бычков В.В., Скобелин С.Ю., Ильин Н.Н. Цифровые следы преступлений. М.: Проспект, 2025. 168 с.

8. Бояркина Л.А., Бояркин В.В. Цифровой след и цифровая тень как производные персональных данных // Сборники конференций НИЦ Социосфера. 2016. № 62. С. 78–81.

9. Кизима С.В., Кидалов Ф.В., Пальцин Д.А. Телекоммуникационная среда и проблематика данных цифрового следа субъектов цифрового общества // Электросвязь. 2021. № 12. С. 33–37.

10. Карлаш Д.С. Правовое регулирование цифровых следов и цифровых теней при обработке больших данных // Вестник арбитражной практики. 2023. № 6 (109). С. 14–21.

11. Ростовцев А.В. Особенности судебно-экспертного исследования «Цифровых следов» [Электрон. ресурс] // Академическая мысль. 2019. № 3(8). Режим доступа: https://cyberleninka.ru/article/n/osobennosti-sudebno-ekspertnogo-issledovaniya-tsifrovyh-sledov.

12. Дерюгин Р.А., Файсханов И.Ф. О криминалистическом исследовании электронных носителей информации и цифровых следов [Электрон. ресурс] // Вестник Уральского юридического института МВД России. 2019. № 4. Режим доступа: https://cyberleninka.ru/article/n/o-kriminalisticheskom-issledovanii-elektronnyh-nositeley-informatsii-i-tsifrovyh-sledov.

13. Ахтырская Н.Н. Процессуальные вопросы признания цифрового следа надлежащим доказательством // Международная научно-практическая конференция «Цифровой след как объект судебной экспертизы» (Москва, 17 января 2020). М.: РГ-Пресс, 2020. 272 с.

14. Карнаухова О.Г. Цифровой след как профиль личности допрашиваемого // Международная научно-практическая конференция «Цифровой след как объект судебной экспертизы» (Москва, 17 января 2020). М.: РГ-Пресс, 2020. 272 с.

15. Казаринова Н.Л., Кудреватых В.А., Мишакина М.Г. Оценка фактического результата формирования цифрового профиля студента педагогического университета по цифровому следу // Педагогический журнал. 2022. Т. 12. № 6–2. С. 1140–1154.

16. Логинова А.А., Денисов А.Р. Применение технологии анализа цифрового следа для создания системы, формирующей индивидуальный цифровой профиль студента // VIII Международная научно-практическая конференция «BIG DATA и анализ высокого уровня» (Минск, 11–12 мая 2022). Минск: Бестпринт УП, 2022. С. 411–420.

17. Нестеров С.А., Смолина Е.М. Понятие цифрового следа и анализ цифрового следа в образовании // XXVI Международная научно-практическая конференция «Системный анализ в проектировании и управлении» (Санкт-Петербург, 13–14 октября 2022). СПб.: Санкт-Петербургский политехнический университет Петра Великого, 2023. С. 309–314.

18. Осанова А.С. Достоинства и недостатки цифровых следов: результаты анализа цифровых следов обучающегося // Международная научно-практическая конференция, посвященная Дню космонавтики «Актуальные проблемы авиации и космонавтики» (Красноярск, 11–15 апреля 2022). Красноярск: Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева, 2022. С. 287–289.

19. Черемисина Е.Н., Кирпичева Е.Ю., Токарева Н.А. и др. Цифровые следы: влияние на IT-образование и цифровую трансформацию // Физика элементарных частиц и атомного ядра. 2024. Т. 55. № 3. С. 519.

20. Гончарова Е.В. Цифровой след как качественная характеристика объема данных в цифровой экономике // Международная научно-практическая конференция «Цифровой след как объект судебной экспертизы» (Москва, 17 января 2020). М.: РГ-Пресс, 2023. С. 55–56.

21. Никишин В.Д. Цифровые и речевые следы в аспекте обеспечения информационной (мировоззренческой) безопасности в интернет-среде // Судебная экспертиза. 2020. № 1 (61). С. 131–139.

22. Баранова Е.В., Швецов Г.В. Методы и инструменты для анализа цифрового следа студента при освоении образовательного маршрута [Электрон. ресурс] // ПНиО. 2021. № 2 (50). Режим доступа: https://cyberleninka.ru/article/n/metody-i-instrumenty-dlya-analiza-tsifrovogo-sleda-studenta-pri-osvoenii-obrazovatelnogo-marshruta.

23. Поколодина Е.В. Обработка цифрового следа в соответствии с моделью деятельности человека (группы людей) и информационно-коммуникационных систем [Электрон. ресурс]. Режим доступа: https://academia-moscow.ru/catalogue/5532/681904/.

24. Поколодина Е.В. Володин С.М., Набойщиков Ф.П. Разработка рекомендаций по защите цифрового следа пользователей сети Интернет //: IV Международная научная конференция, посвящённой памяти доктора технических наук, профессора А.А. Тарасова и доктора технических наук, старшего научного сотрудника О.В. Казарина «Взаимодействие вузов, научных организаций и учреждений культуры в сфере защиты информации и технологий безопасности». (Москва, 19–21 апреля 2023). М.: Российский государственный гуманитарный университет, 2023. С. 218–224.

References

1. Gaydash O.V. The phenomenon of the digital trace in modern society [Internet]. Vestnik magistratury = Bulletin of the Magistracy. 2020: 6(105). Available from: https://cyberleninka.ru/article/n/fenomen-tsifrovogo-sleda-v-sovremennom-obschestve. (In Russ.)

2. Fomicheva T.L. Digital trace and digital hygiene. Samoupravleniye = Self-government. 2023; 6(139): 361-363. (In Russ.)

3. Buynov D.O. Development of scientific approaches to the concept of «digital traces» as an object of expert research. Zakony Rossii: opyt, analiz, praktika = Laws of Russia: experience, analysis, practice. 2023; 3: 17-22. (In Russ.)

4. Semikalenova A.I. Digital traces: purpose and production of examinations [Internet]. Vestnik Universiteta imeni O.Ye. Kutafina = Bulletin of the O.E. Kutafin University. 2019: 5(57). Available from: https://cyberleninka.ru/article/n/tsifrovye-sledy-naznachenie-i-proizvodstvo-ekspertiz. (In Russ.)

5. Shestakova A.A. Digital personality: boundaries and barriers of communication practices in network interaction. VIII Mezhdunarodnaya sotsiologicheskoy Grushinskoy konferentsii «Sotsiolog 2.0: transformatsiya professii» = VIII International Sociological Grushin Conference «Sociologist 2.0: transformation of the profession». Moscow: All-Russian Center for the Study of Public Opinion; 2018: 422-425. (In Russ.)

6. Meshcheryakov V.A. «Virtual traces» under «Occam's scalpel» [Internet]. Informatsionnaya bezopasnost' regionov = Information security of regions. 2009: 1. Available from: https://cyberleninka.ru/article/n/virtualnye-sledy-pod-skalpelem-okkama. (In Russ.)

7. Bagmet A.M., Bychkov V.V., Skobelin S.YU., Il'in N.N. Tsifrovyye sledy prestupleniy = Digital traces of crimes. Moscow: Prospekt; 2025. 168 p. (In Russ.)

8. Boyarkina L.A., Boyarkin V.V. Digital trace and digital shadow as derivatives of personal data. Sborniki konferentsiy NITS Sotsiosfera = Conference proceedings of the Sociosphere Research Center. 2016; 62: 78-81. (In Russ.)

9. Kizima S.V., Kidalov F.V., Pal'tsin D.A. Telecommunication environment and problems of digital trace data of subjects of digital society. Elektrosvyaz' = Electrosvyaz. 2021; 12: 33-37. (In Russ.)

10. Karlash D.S. Legal regulation of digital traces and digital shadows in big data processing. Vestnik arbitrazhnoy praktiki = Bulletin of arbitration practice. 2023; 6(109): 14-21. (In Russ.)

11. Rostovtsev A.V. Osobennosti sudebno-ekspertnogo issledovaniya «Tsifrovykh sledov» = Features of forensic examination of «Digital traces» [Internet]. Akademicheskaya mysl' = Academic thought. 2019; 3(8). Available from: https://cyberleninka.ru/article/n/osobennosti-sudebno-ekspertnogo-issledovaniya-tsifrovyh-sledov. (In Russ.)

12. Deryugin R.A., Fayskhanov I.F. On forensic examination of electronic information carriers and digital traces [Internet]. Vestnik Ural'skogo yuridicheskogo instituta MVD Rossii = Bulletin of the Ural Law Institute of the Ministry of Internal Affairs of Russia. 2019: 4. Available from: https://cyberleninka.ru/article/n/o-kriminalisticheskom-issledovanii-elektronnyh-nositeley-informatsii-i-tsifrovyh-sledov. (In Russ.)

13. Akhtyrskaya N.N. Procedural issues of recognizing a digital trace as proper evidence. Mezhdunarodnaya nauchno-prakticheskaya konferentsiya «Tsifrovoy sled kak ob»yekt sudebnoy ekspertizy» = International scientific and practical conference «Digital trace as an object of forensic examination» (Moscow, January 17, 2020). Moscow: RG-Press; 2020. 272 p. (In Russ.)

14. Karnaukhova O.G. Digital trace as a personality profile of the interrogated. Mezhdunarodnaya nauchno-prakticheskaya konferentsiya «Tsifrovoy sled kak ob»yekt sudebnoy ekspertizy» = International scientific and practical conference «Digital trace as an object of forensic examination» (Moscow, January 17, 2020). Moscow: RG-Press; 2020. 272 p. (In Russ.)

15. Kazarinova N.L., Kudrevatykh V.A., Mishakina M.G. Assessment of the actual result of the formation of a digital profile of a student of a pedagogical university based on a digital footprint. Pedagogicheskiy zhurnal = Pedagogical journal. 2022; 12; 6-2: 1140-1154. (In Russ.)

16. Loginova A.A., Denisov A.R. Application of digital footprint analysis technology to create a system that forms an individual digital profile of a student. VIII Mezhdunarodnaya nauchno-prakticheskaya konferentsiya «BIG DATA i analiz vysokogo urovnya» = VIII International Scientific and Practical Conference «BIG DATA and High-Level Analysis» (Minsk, May 11-12, 2022). Minsk: Bestprint UP; 2022: 411-420.

17. Nesterov S.A., Smolina Ye.M. The concept of a digital footprint and the analysis of a digital footprint in education. XXVI Mezhdunarodnaya nauchno-prakticheskaya konferentsiya «Sistemnyy analiz v proyektirovanii i upravlenii» = XXVI International scientific and practical conference «System analysis in design and management» (St. Petersburg, October 13-14, 2022). Saint Petersburg: Peter the Great St. Petersburg Polytechnic University; 2023: 309-314. (In Russ.)

18. Osanova A.S. Advantages and disadvantages of digital footprints: results of the analysis of a student's digital footprints. Mezhdunarodnaya nauchno-prakticheskaya konferentsiya, posvyashchennaya Dnyu kosmonavtiki «Aktual'nyye problemy aviatsii i kosmonavtiki» = International scientific and practical conference dedicated to Cosmonautics Day «Actual problems of aviation and cosmonautics» (Krasnoyarsk, April 11-15, 2022). Krasnoyarsk: Siberian State University of Science and Technology

named after Academician M.F. Reshetnev; 2022: 287-289. (In Russ.)

19. Cheremisina Ye.N., Kirpicheva Ye.Yu., Tokareva N.A. et al Digital traces: impact on IT education and digital transformation. Fizika elementarnykh chastits i atomnogo yadra = Physics of elementary particles and atomic nuclei. 2024; 55; 3: 519. (In Russ.)

20. Goncharova Ye.V. Digital trace as a qualitative characteristic of the volume of data in the digital economy. Mezhdunarodnaya nauchno-prakticheskaya konferentsiya «Tsifrovoy sled kak ob»yekt sudebnoy ekspertizy» = International scientific and practical conference «Digital trace as an object of forensic examination» (Moscow, January 17, 2020). Moscow: RG-Press; 2023: 55-56. (In Russ.)

21. Nikishin V.D. Digital and speech traces in the aspect of ensuring information (ideological) security in the Internet environment. Sudebnaya ekspertiza = Forensic examination. 2020; 1(61): 131-139. (In Russ.)

22. Baranova Ye.V., Shvetsov G.V. Methods and tools for analyzing a student's digital footprint when mastering an educational route [Internet]. PNiO. 2021: 2(50). Available from: https://cyberleninka.ru/article/n/metody-i-instrumenty-dlya-analiza-tsifrovogo-sleda-studenta-pri-osvoenii-obrazovatelnogo-marshruta. (In Russ.)

23. Pokolodina Ye.V. Obrabotka tsifrovogo sleda v sootvetstvii s model'yu deyatel'nosti cheloveka (gruppy lyudey) i informatsionno-kommunikatsionnykh sistem = Processing a digital footprint in accordance with the model of human activity (group of people) and information and communication systems [Internet]. Available from: https://academia-moscow.ru/catalogue/5532/681904/. (In Russ.)

24. Pokolodina Ye.V. Volodin S.M., Naboyshchikov F.P Development of recommendations for protecting the digital footprint of Internet users. IV Mezhdunarodnaya nauchnaya konferentsiya, posvyashchonnoy pamyati doktora tekhnicheskikh nauk, professora A.A. Tarasova i doktora tekhnicheskikh nauk, starshego nauchnogo sotrudnika O.V. Kazarina «Vzaimodeystviye vuzov, nauchnykh organizatsiy i uchrezhdeniy kul'tury v sfere zashchity informatsii i tekhnologiy bezopasnosti» = IV International Scientific Conference dedicated to the memory of Doctor of Technical Sciences, Professor A.A. Tarasov and Doctor of Technical Sciences, Senior Researcher O.V. Kazarin «Interaction of Universities, Scientific Organizations and Cultural Institutions in the Field of Information Protection and Security Technologies» (Moscow, April 19-21, 2023). Moscow: Russian State University for the Humanities; 2023: 218-224. (In Russ.)

Сведения об авторах

Елена Анатольевна Останина
Московский авиационный институт,
Москва, Россия,
Эл. почта: neka1818@mail.ru

Елена Владиславна Поколодина
Российский экономический университет им. Г.В. Плеханова, Москва, Россия
Эл. почта: Pokolodina.EV@rea.ru

Information about the authors

Elena A. Ostanina
Moscow Aviation Institute,
Moscow, Russia,
E-mail: neka1818@mail.ru

Elena V. Pokolodina
Plekhanov Russian University of Economics,
Moscow, Russia
E-mail: Pokolodina.EV@rea.ru