В. Д. ОЛИСЕЕНКО, М. В. АБРАМОВ
Санкт-Петербургский Федеральный исследовательский центр РАН (СПб ФИЦ РАН)
Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН),
Санкт-Петербургский государственный университет, Санкт-Петербург

ПРИМЕНЕНИЕ МЕТОДОВ РАСПОЗНОВАНИЯ ЛИЦ В ЗАДАЧЕ ИДЕНТИФИКАЦИИ АККАУНТОВ ПОЛЬЗОВАТЕЛЕЙ В РАЗЛИЧНЫХ СОЦИАЛЬНЫХ СЕТЯХ

В представленной работе поднимаются вопросы защиты сотрудников предприятий от социоинженерных атак при помощи оценки выраженности личностных особенностей, проводимой по профилю в социальных сетях. Для улучшения данной оценки необходимо найти профили одного пользователя в различных социальных сетях. Для более точного определения профилей предлагается использовать результат распознавания лиц в качестве атрибута в существующем методе. Данная работа посвящена возможности использования методов распознавания лиц для улучшения результатов идентификации аккаунтов пользователей в различных социальных сетях.

Введение. Киберпреступления становятся всё большей проблемой для предприятий, компаний и пользователей информационных систем. Согласно отчёту компании Verizon [1], ежегодно количество атак на информационные системы стремительно увеличивается. Вместе с этим также увеличивается процент атак с использованием социальной инженерии. Согласно исследованию [2], социальная инженерия проводилась в 84 % атак на предприятия с конечной целью установки вредоносного программного обеспечения. С программно-технической стороны вопросы анализа защищенности и защиты от кибератак достаточно хорошо изучены, активно исследуются различными коллективами [3, 4, 5]. В то же время вопросы анализа защищенности и защиты пользователей информационных систем от социоинженерных атак изучены в меньшей степени. Прежде всего, это обусловлено сложностью исследования данной области, так как объектом исследования является человек, а не программа. Сложность заключается в недетерминированности реакций пользователя на различные воздействия. В одной и той же ситуации на одно и то же социоинженерное атакующее воздействие пользователь в разное время может отреагировать по-разному. Ответные реакции пользователя могут зависеть от его личностных особенностей, психологического состояния, иных контекстов.

Оценка выраженности личностных особенностей пользователя может производиться посредством анализа публикуемого им в социальных сетях контента [6, 7]. При этом, чем больше такого контента удается извлечь, тем проще строить соответствующие оценки. Как правило, пользователи имеют несколько аккаунтов в разных социальных сетях. Извлечение информации из каждого позволяет агрегировать большее число сведений и, соответственно, упрощает построение оценок выраженности личностных особенностей пользователя. Таким образом, актуальной видится задача сопоставления профилей пользователей в разных социальных сетях с целью выявления принадлежащих одному человеку.

Задача сопоставления аккаунтов в различных социальных сетях не является новой, существуют подходы к её решению для различных социальных сетей. В [8] представлен обзор различных подходов и методов для определения принадлежности профилей в разных социальных сетях одному пользователю. Однако, представленные подходы применяются к социальным сетям «Facebook», «Twitter», «Foursquare» и некоторым другим, которые хоть и популярны в России, но не входят в топ-3 наиболее популярных [9]. Не все из предложенных в [8] подходов применимы к наиболее популярным в России социальным сетям «ВКонтакте» и «Одноклассники» [9]. Речь о том, что, например, социальная сеть «Twitter» не имеет многих атрибутов, которые обязательно присутствуют в социальной сети «ВКонтакте» или «Одноклассниках», таких как имя, фамилия, город, возраст и т. д. Кроме того, стоит отметить, что решение задачи сопоставления профилей в разных социальных сетях для разных ресурсов будет иметь свою специфику. Вместе с тем еще не существует канонического метода, позволяющего решать данную задачу без ошибок и с широкой применимостью.

Существуют подходы, ориентированные на социальные сети «ВКонтакте» и «Одноклассники», основанные на сравнении значений атрибутов профилей для выявления принадлежности их одному человеку [10, 11]. Но эти подходы не достигают стопроцентной точности и приме-

нимости, то есть актуальной остается задача автоматизации идентификации профилей пользователей в разных социальных сетях. Одним из вариантов улучшения данных подходов может быть учет такой важной информации из профиля пользователя как фотографии. Сравнивая лица на фотографиях профилей в разных социальных сетях и добавляя результаты сравнения в качестве признака, например, в алгоритм, предложенный в [11], ожидается, что можно будет улучшить оценку идентификации и сделать её более устойчивой.

Предлагаемый доклад посвящён возможности использования методов распознавания лиц на фотографиях пользователей в социальных сетях, в контексте использования их в подходе [11] определения профилей пользователя в социальных сетях «ВКонтакте» и «Одноклассники»

Предложенный подход.

Согласно подходу, предложенному в [11], решается задача бинарной классификации, где X – множество сопоставленных пар профилей пользователей социальных сетей «ВКонтакте» и «Одноклассники», а Y – множество классов {0; 1}, где 0 означает, что пара профилей не принадлежит одному пользователю, а 1 – что принадлежит. Признаками классификации выступают числовые результаты сопоставления значений соответствующих атрибутов из профилей пользователей «фамилия», «имя», «город», «возраст», «список друзей».

В рамках данного доклада предлагается рассмотреть возможность включения результатов распознавания лиц как признака бинарной классификации. Целесообразность данного включения обосновывается тем, что оно предположительно поможет избежать ложноположительного решения классифицирующей модели, когда владельцы профилей являются полными тезками, проживают в одном городе или имеют одинаковый возраст. Это может быть актуально для людей с распространёнными фамилиями и одинаковыми именами. Так только в социальной сети «ВКонтакте» людей с именем и фамилией «Иван Иванов», проживающих в Санкт-Петербурге 10 306 человек.

Существующие подходы. Распознавание лиц на фотографии состоит из четырёх этапов [12]:

1. Найти лица на фотографии;
2. Распознать каждое лицо, несмотря на разное освещение и угол фотографии;
3. Определить каждую особую черту лица;
4. Сравнить лица разных людей.

Для каждого из этапов существует множество разных методов их решения. Однако мы рассмотрим существующие комплексные программные решения. Так в статье [13] проводится обзор популярных библиотек компьютерного зрения, позволяющих решать каждый из этапов распознавания лиц. Авторы представленной работы рассматривают следующие библиотеки: OpenCV, OpenCV с IPP, LTI и VXL. Проанализировав данные библиотеки, авторы пришли к выводу, что OpenCV является лучшей библиотекой для распознавания лиц так как имеет следующие преимущества: быстрота работы, гибкость в выборе методов для каждого этапа распознавания лиц, открытый исходный код, также кроссплатформенная и кросс-языковая реализация библиотеки. Используя данную библиотеку, планируются получать оценки вероятности принадлежности лица на фотографии одному человеку. Таким образом, новый атрибут в модели [11] будет представлять собой вероятность принадлежности лица одному человеку, лежащего в диапазоне [0,1].

Заключение. В докладе представлен подход к использованию методов распознавания лиц на фотографиях пользователей в социальных сетях, в контексте использования их в подходе [11] определения профилей одних и тех же пользователей в социальных сетях «ВКонтакте» и «Одноклассники». Агрегирование большего количества сведений упрощает построение оценок выраженности личностных особенностей пользователя, и опосредовано, помогает в предотвращении социоинженерных атак на пользователей, компаний и предприятий.

Работа выполнена в рамках проекта по государственному заданию СПИИРАН № 0073-2019-0003, при финансовой поддержке РФФИ, проект №20-07-00839, №18-01-00626.

ЛИТЕРАТУРА

1. Verizon Data Breach Investigations Report 2018. 2019. URL: https://www.researchgate.net/profile/Suzanne_Widup/publication/324455350_2018_Verizon_Data_Breach_Investigations_Report/links/5ace9f0b0f7e9b18965a5fe5/2018-Verizon-Data-Breach-InvestigationsReport.pdf?origin=publication_detail
2. Ptsecurity – Актуальные киберугрозы: итоги 2019 года. 2020. URL: https://www.ptsecurity.com/ru-ru/research/analytics/cybersecurity-threatscape-2019/
3. **Браницкий А.А., Котенко И.В.** Анализ и классификация методов обнаружения сетевых атак // Труды СПИИРАН. 2016. № 45 (2). С. 207-244.
4. **Котенко И.В., Паращук И.Б.** Верификация недостоверных параметров модели обнаружения вредоносной информации // Вестник АГТУ. Серия: Управление, вычислительная техника и информатика. 2019. №2. С. 7–18.
5. **Zhang Y., Zhang L. Y., Zhou J., Liu L., Chen F**. A Review of Compressive Sensing in Information Security Field // IEEE Access. 2016. vol. 4. pp. 2507–2519.
6. **Абрамов М. В., Тулупьева Т. В., Тулупьев А. Л.** Социоинженерные атаки: социальные сети и оценки защищенности пользователей // СПб. ГУАП, 2018. 266 с. ISBN 978-5-8088-1377-5.
7. **Kharitonov N.A., Maximov A.G., Tulupyev A.L.** Algebraic Bayesian Networks: Naïve Frequentist Approach to Local Machine Learning Based on Imperfect Information from Social Media and Expert Estimates // Communications in Computer and Information Science. 2019.
8. **Hazimeh H., Mugellini E., Khaled O. A., CudréMauroux P.** SocialMatching++: A Novel Approach for Interlinking User Profiles on Social Networks // In Proceedings of PROFILES'17@ ISWC. 2017.
9. Статистика социальных сетей в России на 2018 год. URL: https://hiconversion.ru/blog/statistika-socialnyh-setej-v-rossii-na-2018-god/
10. Патент РФ № 2011145077 / Бартунов С.О., Коршунов А.В., Турдаков Д.Ю. и др. Способ интеграции профилей пользователей онлайновых социальных сетей. Опубл. 08.11.2011. Бюл. № 8.
11. **Корепанова А.А., Олисеенко В.Д., Абрамов М.В., Тулупьев А.Л**. Применение методов машинного обучения в задаче идентификации аккаунтов пользователя в двух социальных сетях // Компьютерные инструменты в образовании. 2019. № 3. С. 29–43. doi:10.32603/2071-2340-2019-3-29-4
12. **Al-Mukhtar F., Al-Dabagh M.** Real-Time Face Recognition System Using KPCA, LBP and Support Vector Machine. International Journal of Advanced Engineering Research and Science. 2017. 4. 184-189. 10.22161/ijaers.4.2.36.
13. **Булатников Е.В., Гоева А.А.** Сравнение библиотек компьютерного зрения для применения в приложении, использующем технологию распознавания плоских изображений // Вестник МГУП. 2015. №6.

V.D. Oliseenko, M.V. Abramov (St. Petersburg Federal Research Center of the RAS; St. Petersburg State University, St. Petersburg)

Applying Methods of Face Recognition in the Task of Identifying User Accounts in Various Social Networks

In the presented paper questions of protection of employees of the enterprises from social engineering attacks are raised by means of an estimation of expression of the personal features spent on a profile in social networks. To improve this assessment, it is necessary to find profiles of one person in different social networks. To define profiles more accurately, it is suggested to use the result of face recognition as an attribute in the existing method. This work is devoted to the possibility of using facial recognition methods to improve the results of identifying user accounts in various social networks.