DOI 10.37882/2223–2966.2022.11–2.18

ПРИМЕНЕНИЕ МЕТОДОВ ОБУЧЕНИЯ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ РАСПОЗНАВАНИЯ ЭМОЦИЙ РАБОТНИКА В ЦЕЛЯХ ПОВЕДЕНЧЕСКОЙ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

APPLICATION OF METHODS FOR STUDYING A CONVOLUTIONAL NEURAL NETWORK FOR PEOPLE WITH LIMITED EMOTIONS FOR THE PURPOSE OF BEHAVIORAL INFORMATION SECURITY

I. Mandritsa
V. Kopytov
A. Chernyshev
A. Makarov
D. Reznikov

Summary. The behavior of an employee of an organization in the field of information security must comply with the adopted information security policy in it. A stable working «state» changes to a problematic «unstable» state, and with a high probability leads to information threats to the organization, in the form of information leaks, or irreversible information threats. The emotions of a person (employee) always have a digital «footprint» and are present on the face. This "trace" fills in the metadata for the mathematical model of the organization's information security, where the excess of the calculated probability level determines the «access» of this employee to the organization's business process. For recognition emotion, a neural network consisting of 152 layers is used, with an output layer of 7 neurons, one for each emotion. During the training, a dataset of 28,000 full-face images of people's faces was used, expressing 7 emotions: anger. disgust, fear, happiness, sadness, surprise, calmness. The neural network is written in Python using the PyTorch library.

Keywords: neural network training, emotion recognition, behavioral information security.

Мандрица Игорь Владимирович
Доктор экономических наук, доцент
Северо-Кавказский федеральный университет
d_artman@mail.ru
Копытов Владимир Вячеславович
Доктор технических наук, профессор
Северо-Кавказский федеральный университет
v.kopytov@infocom-s.ru
Чернышев Александр Борисович
Доктор технических наук, доцент
Пятигорский институт (филиал) Северо-Кавказского федерального университета
achernyshev@ncfu.ru
Макаров Анатолий Михайлович
Доктор технических наук, профессор
Пятигорский государственный университет
mellin_22@mail.ru
Резников Дмитрий Константинович
Северо-Кавказский федеральный университет

Аннотация. Поведение работника организации в области информационной безопасности должно соответствовать принятой политике информационной безопасности в ней. Стабильное рабочее «состояние» меняется на проблемное «нестабильное» состояние, и с большой вероятностью ведет к информационным угрозам для организации, в виде утечек информации, или необратимых информационных угрозам. Эмоции человека (работника) всегда имеют цифровой «след», и присутствуют на лице. Этот «след» наполняет метаданные для математической модели информационной безопасности организации, где превышение уровня расчетной вероятности определяет «доступ» данного работника к бизнес-процессу организации. Для распознавания эмоций применяется нейронная сеть, состоящая из 152 слоев, с выходным слоем из 7 нейронов, по одному на каждую эмоцию. В ходе обучения применялся датасет из 28000 изображений лиц людей анфас, выражающих 7 эмоций: гнев. отвращение, страх, счастье, грусть, удивление, спокойствие. Нейронная сеть написана на языке Python, с использованием библиотеки PyTorch.

Ключевые слова: обучение нейронной сети, распознавание эмоций, поведенческая информационная безопасность.

Поведенческая информационная безопасность позволяет выявлять случаи непреднамеренного совершения ошибок работниками организации в обращении с информационными ресурсами фирмы [1]. Одной из причин информационной угрозы для организации является негативные действия сотрудника фирмы как результат воздействия на него различных эмоциональных отклонений, вызванные реакцией на внешние факторы. Зная набор этих факторов и степень их воздействия, можно проводить предположительный анализ психоэмоционального состояния и как следствие возможного типа поведения информацион-

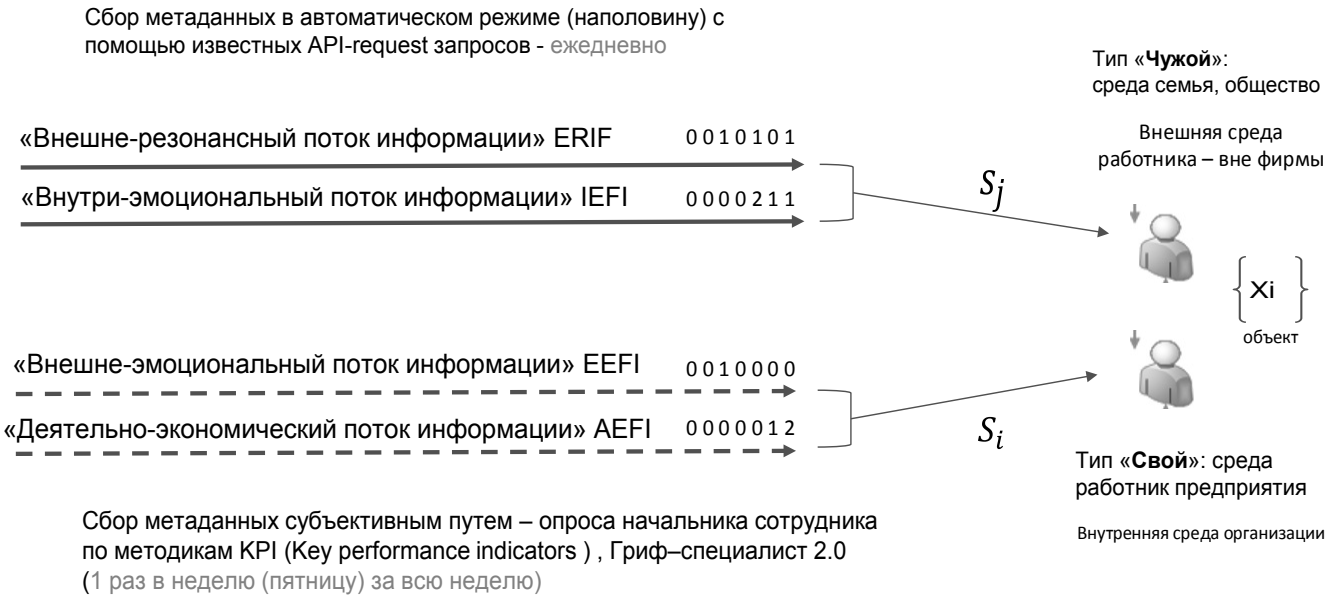

Рис. 1. Сбор метаданных о работнике организации из открытых и закрытых источников информационного пространства

ной безопасности работника в обращении с информацией организации, для этого и составляется следующая математическая модель внутреннего нарушителя, отражающая спектр негативных эмоциональных состояний человека и расчет вероятности что «субъект-работник» станет утечкой (инсайдером) (рисунок 1).

Метаданные раскрывают сведения о признаках и свойствах, характеризующих какие-либо сущности, позволяющие автоматически искать и управлять ими в больших информационных потоках. Для целей поставленной задачи — контроля психоэмоционального состояния работника в виде эмоций на его лице в компиляции с подтвержденными фактами его внешней среды проживания и внутри организации создание и накопление метаданных о работнике может проходить как в ручном, так и в автоматическом режиме. Из разного рода источников как отрытых, так и «сливов» информации из закрытых источников (баз данных должников, баз данных о правонарушениях на дорогах и т.п.) формируются потоки данных о человеке, его материальном обеспечении, возможных проблемах с законом. Метаданные для математической модели по нижеперечисленным факторам (таблица 1) берутся из открытых источников с помощью API request запросов и затем они компилируются с обработкой видеопотока эмоций данного работника внутри организации (фирмы) [2].

Одно подтвержденное событие из внешней среды возвращает значение 0 в 1 (таблица 1), что важно для расчета вероятности, значит — есть факт некой доли (веса) информационно-психологического воздействия данного потока информации на Xi — работника до последующего подтверждения факта.

После всех этапов модели, в конце алгоритма необходимо будет рассчитать вероятностные состояния «Свой» — S_i или «Чужой» — S_j от количества инцидентов $Q(Xi)$ воздействующих на него «извне», как показано на рисунке 1 – декомпозиция этапов модели в виде алгоритма расчета вероятности состояния работника «Чужой».

Соответственно на данном рисунке отражена гипотеза расчета полной вероятности ρ_j от наблюдаемого воздействия инцидентов на состояние $S_j(X_i)_m^{ext}$ работника X_i — «Чужой» по обобщенной формуле Байеса [6, 7]

$$P\left(\frac{D_{Xi}}{K^{Sj}}\right) = P(D_{Xi}) * P(\frac{K^{Sj}}{D_{Xi}})/P(K^{Sj}) \quad (1)$$

где $P\left(\frac{D_{Xi}}{K^{Sj}}\right)$ —

полученная полная вероятность по состоянию S_j — «Чужой»; $P(D_{Xi})$ — предварительная предельная вероятность (гипотеза порога) неудовлетворенности работника X_i.

Для целей задачи информационной безопасности организации для контроля психоэмоционального состояния авторы предлагается использование нейро-сетевых технологий распознавания эмоций работника

Таблица 1. Метаданные из «внешней среды жизни» работника с помощью API request запросов для последующего расчета вероятности резонансного психоэмоционального состояния работника [3]

№ пп	Для каждого Xi работника из состава персонала имеющих доступ к коммерческой информации	Одна неделя жизни работника Xi (частота наблюдений 2 раза в день)						
	ВРЕМЯ СБОРА ДАННЫХ	Понед.	Втор.	Среда	Четв.	Пятн.	Субб.	Воскр.
1	2	3	4	5	6	7	8	9
1	ИТОГ Машинный сбор информации API-запросы — ежедневные= [СУММ(ERIF)+-СУММ(IEFI)] (стр.2+стр.10)	0	1	0	0	0	3	1
2	Внешне-резонансный поток информации СУММ(ERIF) (стр.3: стр.9)	0	0	0	0	0	1	1
3	Чужие штрафы	0	0	0	0	0	0	0
4	Чужие долги	0	0	0	0	0	0	0
5	Чужие кредиты	0	0	0	0	0	0	0
6	Чужие болезни	0	0	0	0	0	1	1
7	Чужие угрозы	0	0	0	0	0	0	0
8	Чужие аварии и ущербы	0	0	0	0	0	0	0
9	Чужие проблемы от связей	0	0	0	0	0	0	0
10	Внутри-эмоциональный поток информации СУММ(IEFI) (стр.11: стр.17)	0	1	0	0	0	2	0
11	Свои штрафы	0	0	0	0	0	0	0
12	Свои долги	0	0	0	0	0	0	0
13	Свои кредиты	0	0	0	0	0	0	0
14	Свои болезни	0	0	0	0	0	0	0
15	Свои ссоры в семье	0	1	0	0	0	1	0
16	Свои аварии и ущербы	0	0	0	0	0	0	0
17	Свои депрессии	0	0	0	0	0	1	0
1	ИТОГ Ручной сбор информации ЭКСПЕРТНЫЕ-запросы — раз неделю= [СУММ(ERIF)+СУММ(EEFI)] (стр.2+стр.10)	0	3	2	0	3	0	0
2	Внешне-эмоциональный поток информации СУММ(EEFI) (стр.3: стр.8)	0	2	1	0	2	0	0
3	Факты конфликтов с коллегами	0	1	1	0	0	0	0
4	Факты ссор и агрессии	0	0	0	0	0	0	0
5	Факты выгорания и слабой мотивации	0	0	0	0	1	0	0
6	Факты ругательств в адрес руководства	0	1	0	0	0	0	0
7	Факты избыточного любопытства, занудства	0	0	0	0	0	0	0
8	факты эмпатии и ненависти	0	0	0	0	1	0	0
9	Внутри-эмоциональный поток информации СУММ(AEFI) (стр.10: стр.14)	0	1	1	0	1	0	0
10	Факты прогулов и лености	0	0	0	0	0	0	0
11	Факты снижения качества работы	0	0	1	0	1	0	0
12	Факты халатности и ошибок и вранья	0	0	0	0	0	0	0
13	Факты снижения производительности труда	0	0	0	0	0	0	0
14	Факты хищений и ущербов	0	1	0	0	0	0	0

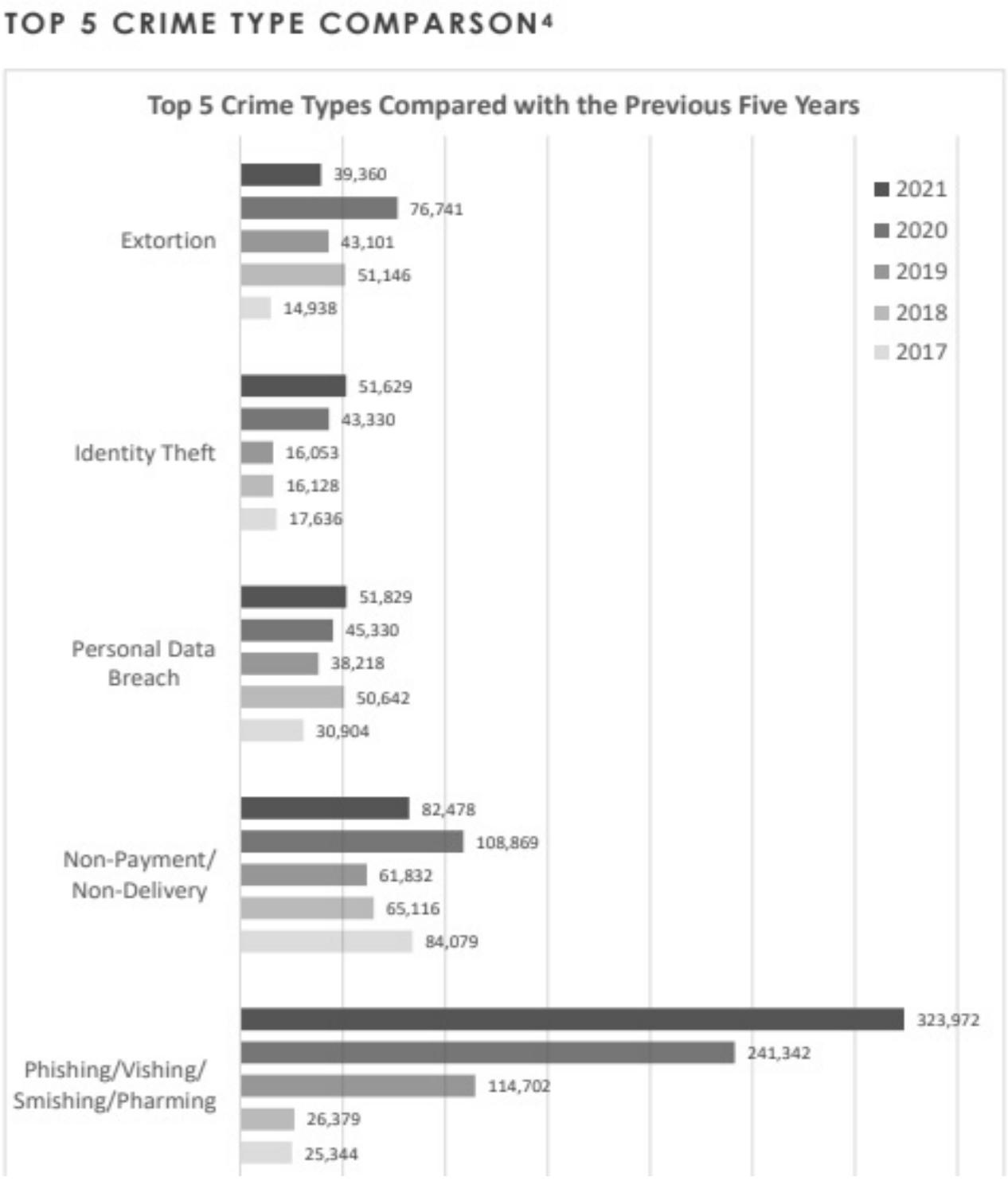

Рис. 2. Число самых распространенных киберпреступлений в динамике за 5 лет [9]

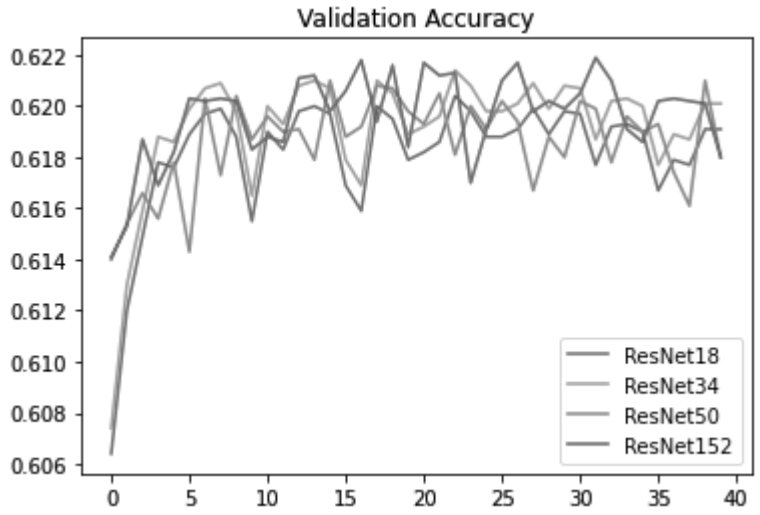

Рис. 3. Точность моделей на валидации в ходе тренировки нейросети по эпохам

внутри организации. Для чего можно использовать данные СКУД организации и ее системы внутреннего видеонаблюдения, свидетельствующие о перемещениях человека внутри объекта, его эмоциях на различных стадиях бизнес-процесса фирмы.

Сбор и анализ метаданных проходит в режиме реального времени. Используя API запросы с различных веб-сервисов, геолокации, видеопотока с камер видеонаблюдения, будет осуществляться в автоматическом режиме. Данные поступают на вход различных алгоритмов машинного обучения, которое определяет принадлежность и вес данных, исходя из присвоенной им важности, метаданным присваивается численный эквивалент, в последствии используемый в вычислении контрольной суммы вероятностей.

С вероятностью 1 не могут совпасть два не идентичных кадра (поток разбивается на отдельные кадры, для распознавания выбираются кадры с определенной периодичностью, для снижения нагрузки), поэтому экспертным путем определяется допустимое значение совпадения. Так же существует возможность совместно с «отпечатком» и временем его обнаружения сохранять ряд кадров не опознанных лиц, для того, чтоб в будущем, при увеличении качественных показателей работы алгоритма по распознаванию, сохранялась возможность повторного распознавания лица (или при переходе на новый алгоритм).

В результате в базу данных о каждом работнике вносятся появления на камере определенного работника и время появления, так же можно отслеживать перемещения этого работника по территории предприятия, тем самым отслеживая возможность длительного нахождения в отделах, в которых у работника нет задач к выполнению.

Распознавание эмоционально-психологического состояния в перспективе может применяться в различных сферах жизни человеков, в медицине, отрасли услуг, маркетинге и рекламе, а в рамках информационной безопасности организации при получении допусков ее персонала к ответственным этапам бизнес-процесса или функциональным обязанностям в бюджетной организации, при допуске к коммерческой или государственной тайне. На сегодня вообще распознавание эмоций как информационная задача уже достаточно разработано на программном уровне, и известны Пермская программа IT-компании New Vizion [4], программа которой называется «Гибридная нейро-экспертная система экспресс-диагностики характера и эмоционального состояния человека в видеопотоке». Ее принято считать особой видео-аналитикой, в которой не просто происходит распознавание человека в видеопотоке и идентификация личности, а происходит распознавание его психоэмоционального состояния. Если отдел безопасности в отчетах видеоаналитики видит, что каждый день в течение недели у сотрудника систематически наблюдается результирующая эмоция «агрессия», или «депрессия», то аналитик понимает, что работник находится в некой группе информационного риска.

Рассматривая ежегодную статистику совершенных киберпреступлений (рисунок 2), можно увидеть, что большую часть из них составляет, так называемый «фишинг». Внешние информационные атаки проводятся персонально на работников фирм (организаций), имеющих доступ к ценной информации (коммерческой,

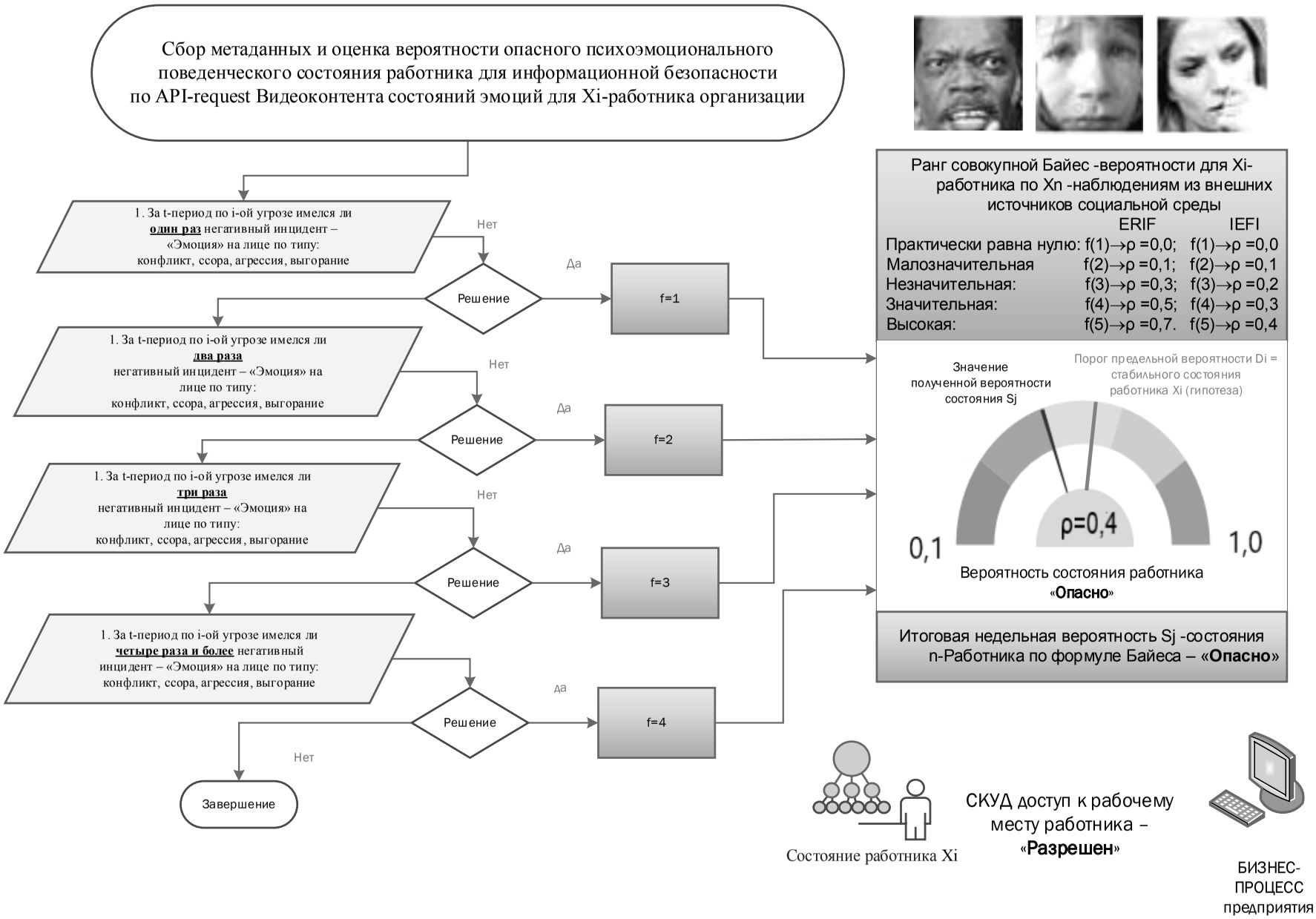

Рис. 4. Алгоритм определения доступа работника к бизнес-процессу фирмы через эмоцию как подтверждения воздействия на него психоэмоциональных факторов внешней среды

торговой, бизнес-информации или государственной тайне).

Осуществляться они могут по средствам спам-рассылки писем с несуществующими ссылками, мошенничество с использованием известных брендов, ложные антивирусы и программы обеспечения безопасности, телефонный фишинг. Такие атаки не требуют серьезных технических знаний уязвимостей различных информационных частей, а только базовые знания работы компьютерных сетей и веб-сервисов. Именно число фишинговых атак растет наибольшими темпами: от 25 тыс. в 2017 году возросло до 323 тыс. в 2021, то есть более чем в 13 раз.

Сотрудник коммерческой фирмы, имеющий средний доступ к информации фирмы, как правило неограничен на просмотр, копирование, перемещение и изменение данных, находящихся в 10 миллионах файлов, в это же время около 1000 конфиденциальных файлов открыты для доступа любого сотрудника фирмы. Такие выводы содержатся в отчете составленном Varonis [4]. Отчет содержит сведения и о использовании паролей с истекшими сроками действия, обнаруженные в 59% организаций, подвергшихся анализу.

Распознавание эмоций для информационной безопасности является только лишь частью комплексной задачи по обеспечению внутренней службой безопасности, так называемой превентивной мерой и вытекает из работ [3, 4], как совокупный механизм обнаружения внутреннего потенциального злоумышленника организации посредством математической модели обработки 23 факторов в виде ежедневного потока метаданных собираемых о каждом работнике организации. Распознавание эмоций по средствам нейронной сети в данной работе полагается на мимические проявления эмоций. Стоит отметить, что в силу физиологических особенностей человеческих организмов, не все эмоции поддаются однозначному распознаванию, при этом одинаковые эмоции могут проявляться у разных людей в разной степени искажения лица.

Распознавание состоит из пропускания через нейронную сеть изображения лица, с вынесением результата в виде вероятности. Эмоция определяется наибольшим значением вероятности. Предварительно датасет был подготовлен для обучения: изображения приведены в черно-белый формат, размер изображения приведен к 224*224 пикселей. В эксперименте на примере лиц известных личностей для распознавания эмоций была выбрана сверточная нейронная сеть архитектуры ResNet, с выходным слоем, состоящим из 7 нейронов, для каждой эмоции, с функцией потерь — перекрестной энтропией и алгоритмом оптимизации Адама [5, 13, 14].

В ходе экспериментальных тренировок на батчах различного размера, наилучшие результаты сетями из 18 и 34 слоев, были установлены на батчах, состоящих из 18 изображений [16]. Для сети из 50 слоев технически отсутствовала возможность использовать батч размером 18 изображений, поэтому число было снижено до 12, идентично для сети из 152 слоев, батч состоял из 9 изображений. Размер батча не стал фактором, снизившим точность классификации сети 50 и 152 слойных сетей. Разница в точности после 10 эпох обучения между ResNet150 и ResNet34 составляла 5–7%, между ResNet150 и ResNet152–4–7%. Сети ResNet18 и ResNet34 показывали соизмеримую точность +-0.02. Точность классификации сетью ResNet150 составляла 0.45. Для дальнейшего обучения была выбрана сеть с 150 слоями, т.к. показала наибольшую точность классификации. Тренировочный датасет был разбит на 4 равных части, после чего тренировка проходила по 10 эпох на каждой части последовательно. В случае начала перетренировки модель сбрасывалась до более оптимального состояния, со сменой на новую часть датасета, в результате точность после тренировки составила 0.51. С целью повышения точности, изображения пришлось предобработать снова: обрезан фон, частично волосы головы, в автоматическом режиме, с помощью встроенной функции библиотеки Matplotlib. Датасет был разбит на равные части.

Для тренировки на полученном датасете была взята не обученная сеть ResNet150, обучение было как описано выше. В добавок скорость обучения занижена с 1e-3 до 1e-4. В результате после тренировки на всем датасете прирост точности 10–15%, т.е. результирующая точность составила 0.62 (точность на тестовом датасете), что отражено на рисунке 3 [15, 17].

В результате сбора, анализа полученных метаданных с помощью математической модели (1) и компиляции совокупных вероятностей психоэмоционального состояния от двух потоков метаданных система информационной безопасности организации в полуавтоматическом режиме будет определять доступ каждого работника к бизнес-процессу организации через совокупную вероятность «поведенческой» информационной безопасности работника будет определять математически уровень доступа работника по типу «Разрешен» или «В доступе отказано», как показано на рисунке 4.

В результате компилируя двумя вероятностями, но в рамках совокупной вероятности по Байесу равной 1, система доступа работника к бизнес-процессу (к рабочему месту) станет триггером — регулятором доступа и соответственно повысит информационную безопасность организации от возможных угроз и ущербов от работника.

ЛИТЕРАТУРА

1. (Merrill Warkentin & Robert Willison (2009) Behavioral and policy issues in information systems security: the insider threat, European Journal of Information Systems, 18:2, 101–105, DOI: 10.1057/ejis.2009.12),
2. Стратегическое управление развитием информационной безопасности социально-экономических систем на основе умных технологий: Монография / Л.М. Борщ, С.В. Герасимова, А.Р. Жарова [и др.]. — Симферополь: Крымский федеральный университет им. В.И. Вернадского, 2022. — 392 с. — ISBN978–5–6047624–5–5. — EDN ZPEZJF.
3. Концепция динамической модели обнаружения внутреннего нарушителя (инсайдера) коммерческой организации / И.В. Мандрица, А.П. Жук, В.И. Петренко [и др.] // Проблемы информационной безопасности социально-экономических систем: VIII Всероссийская с международным участием научно-практическая конференция, Симферополь — Гурзуф, 17–19 февраля 2022 года. — Симферополь: Крымский федеральный университет им. В.И. Вернадского, 2022. — С. 19–23. — EDN ZEBZYO.
4. Звонарев, С.В. «Основы математического моделирования: учебное пособие / С.В. Звонарев. — Екатеринбург: Изд-во Урал. ун-та, 2019. —112 с. — режим доступа https://elar.urfu.ru/bitstream/10995/68494/1/ 978–5–7996–2576–4_2019.pdf
5. Мартьянов Е.А. Возможность выявления инсайдера статистическими методами // Системы и средства информатики. — 2017. — Т. 27, № 2. —С. 41–47;
6. Системно-динамическое моделирование информационных воздействий на социум / В.А. Минаев [и др.] // Вопросы радиоэлектроники. 2017. № 11. С. 35–43.
7. Медведев В.И., Ларина Е.А. Борьба с внутренними угрозами. Выявляем инсайдера — «Актуальная бухгалтерия», февраль 2014 — режим доступа https://kvftroubleshooterblog.wordpress.com/борьба-с-внутренними-угрозами-выявля/
8. Щеглов А.Ю., Щеглов К.А. Математические модели и методы формального проектирования систем защиты информационных систем. Учебное пособие. — СПб: Университет ИТМО, 2015. — 93 с.
9. Строгалев В.П., Толкачева И.О. Имитационное моделирование. — М.: Изд-во МГТУ им. Н.Э. Баумана, 2008;
10. Deep Residual Learning for Image Recognition [Electronic resource]. — https://arxiv.org/pdf/1512.03385.pdf
11. Neural Networks and Deep Learning [Electronic resource]. — http://neuralnetworksanddeeplearning.com/
12. FER-2013 dataset with 7 emotion types [Electronic resource]. — https://www.kaggle.com/datasets/ananthu017/emotion-detection-fer
13. Вероятностно-статистические методы принятия решений: Теория, примеры, задачи: учебное пособие /А.П. Науменко, И.С. Кудрявцева, А.И. Одинец; Минобрнауки России, ОмГТУ. — Омск: Изд-во ОмГТУ, 2018, ISBN978–5
14. Ивановский, Л.И. Использование глубокого обучения и сверточных нейронных сетей для анализа выражения лица / Л.И. Ивановский, О.А. Степанова, В.В. Хрящев // DSPA: Вопросы применения цифровой обработки сигналов. — 2018. — Т. 8. — № 4. — С. 170–173. — EDN YPVPVZ.
15. Estimation of Light Commercial Vehicles Dynamics by Results of Road Tests and Simulation / A.V. Tumasov, A.M. Groshev, R.A. Musarsky [et al.] // Journal of Communication and Computer. — 2014. — Vol. 11. — No 5. — P. 412–422. — DOI 10.17265/1548–7709/2014.05002. — EDN QCRJWD.
16. Ахметшин, Р.И. Распознавание эмоций человека на изображениях / Р.И. Ахметшин, А.П. Кирпичников, М.П. Шлеймович // Вестник Технологического университета. — 2015. — Т. 18. — № 11. — С. 160–163. — EDN UBLMEV.
17. Кирпичников, А.П. Автоматизированная система моделирования параметров быстропротекающих процессов / А.П. Кирпичников, С.А. Ляшева, О.Т. Шипина // Вестник Казанского технологического университета. — 2014. — Т. 17. — № 13. — С. 349–351. — EDN SNWZCJ.

© Мандрица Игорь Владимирович (d_artman@mail.ru), Копытов Владимир Вячеславович (v.kopytov@infocom-s.ru),
Чернышев Александр Борисович (achernyshev@ncfu.ru), Макаров Анатолий Михайлович (mellin_22@mail.ru),
Резников Дмитрий Константинович (mellin_22@mail.ru).
Журнал «Современная наука: актуальные проблемы теории и практики»