УНИВЕРСИТЕТ ИТМО

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
июль–август 2021 Том 21 № 4 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
July–August 2021 Vol. 21 No 4 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК УНИВЕРСИТЕТ ИТМО
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2021-21-4-592-598
УДК 004.056.053

Моделирование нарушений безопасности в системах машинного обучения

Максим Алексеевич Чекмарев[1]✉, Станислав Геннадьевич Клюев[2], Виктор Викторович Шадский[3]

[1,2,3] Краснодарское высшее военное училище им. С.М. Штеменко, Краснодар, 350063, Российская Федерация

[1] max.chek13@gmail.com✉, http://orcid.org/0000-0002-6832-9991
[2] s.g.klyuev@mail.ru, http://orcid.org/0000-0002-0534-9143
[3] vdvryazan57@yandex.ru, http://orcid.org/0000-0002-9221-2283

Аннотация

Предмет исследования. Широкое распространение машинного обучения, включая объекты критической информационной инфраструктуры, влечет за собой риски угроз безопасности при отсутствии надежных средств защиты. В работе рассмотрены процессы в системах машинного обучения по аналогии с информационными системами, которые подвержены вредоносным воздействиям. Представлены результаты моделирования событий, приводящих к нарушению безопасности в системах машинного обучения объектов критической информационной инфраструктуры. **Метод.** При моделировании применена технология создания функциональных моделей SADT (Structured Analysis and Design Technique) и методология IDEF0 (Integration definition for function modeling) как инструмент перехода от вербально-функционального описания исследуемого процесса к описанию в терминах математического представления. Для исследования сценариев перехода систем машинного обучения в опасное состояние и численной оценки вероятности нарушения безопасности выполнено математическое моделирование угроз с использованием логико-вероятностного метода. **Основные результаты.** Получена наглядная функциональная модель нарушения безопасности системы в виде контекстной диаграммы системы и двух уровней декомпозиции. Определена функция опасности системы и выведен арифметический полином вероятностной функции. **Практическая значимость.** Представленные модели позволят разработать методы и алгоритмы защиты систем машинного обучения от вредоносных воздействий и применять их в ходе оценки уровня защищенности.

Ключевые слова

машинное обучение, нарушение безопасности, целостность, конфиденциальность, функциональное моделирование, логико-вероятностное моделирование

Благодарности

Работа выполнена в Краснодарском высшем военном училище им. С.М. Штеменко в рамках диссертационного исследования в области обеспечения безопасности систем машинного обучения.

Ссылка для цитирования: Чекмарев М.А., Клюев С.Г., Шадский В.В. Моделирование нарушений безопасности в системах машинного обучения // Научно-технический вестник информационных технологий, механики и оптики. 2021. Т. 21, № 4. С. 592–598. doi: 10.17586/2226-1494-2021-21-4-592-598.

Modeling security violation processes in machine learning systems

Maxim A. Chekmarev[1]✉, Stanislav G. Klyuev[2], Viktor V. Shadskiy[3]

[1,2,3] Krasnodar Higher Military School, Krasnodar, 350063, Russian Federation

[1] max.chek13@gmail.com✉, http://orcid.org/0000-0002-6832-9991
[2] s.g.klyuev@mail.ru, http://orcid.org/0000-0002-0534-9143
[3] vdvryazan57@yandex.ru, http://orcid.org/0000-0002-9221-2283

Abstract

The widespread use of machine learning, including at critical information infrastructure facilities, entails risks of security threats in the absence of reliable means of protection. The article views the processes in machine learning systems as the ones occurring in information systems susceptible to malicious influences. The results of modeling events leading

to a security breach in machine learning systems operating at critical information infrastructure facilities are presented. For modeling, the technology of creating functional models SADT (Structured Analysis and Design Technique) and the IDEF0 (Integration definition for function modeling) methodology were used as a tool for transition from a verbal functional description of the process under study to a description in terms of mathematical representation. In order to study the scenarios of the transition of machine learning systems to a dangerous state and the numerical assessment of the probability of security violation, mathematical modeling of threats was carried out using the logical-probabilistic method. The authors obtained a visual functional model of system security violation in the form of a context diagram of the system and two levels of decomposition. The hazard function of the system is determined and the arithmetic polynomial of the probability function is derived. In further work the described models will allow researchers to develop methods and algorithms for protecting machine learning systems from malicious influences, as well as to apply them in assessing the level of security.

Keywords
machine learning, security breach, integrity, confidentiality, functional modeling, logical probabilistic modeling

Acknowledgements
The work was carried out at the Krasnodar Higher Military School as part of a dissertation research in the field of ensuring the security of machine learning systems.

For citation: Chekmarev M.A., Klyuev S.G., Shadskiy V.V. Modeling security violation processes in machine learning systems. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2021, vol. 21, no. 4, pp. 592–598 (in Russian). doi: 10.17586/2226-1494-2021-21-4-592-598.

Введение

Машинное обучение в современных условиях оказывается наиболее распространенным способом решения сложных задач. Разнообразие алгоритмов, высокая эффективность их применения и возможность работы с большими объемами данных позволяют внедрять системы машинного обучения не только в развлекательные сегменты или сферы бизнеса, но и на объекты критической информационной инфраструктуры.

Недавно были обнаружены проблемы уязвимости систем машинного обучения перед злонамеренными манипуляциями как на этапе обучения алгоритма, так и на этапе функционирования модели. Так, в настоящее время рядом авторов выявлены и описаны следующие проблемы.

Нарушение целостности системы машинного обучения. Происходит в результате вредоносного воздействия на набор обучающих данных («poisoning attack») либо на входные данные в процессе функционирования системы («evasion attack»). Снижение доступности в байесовских сетях и моделях с использованием метода опорных векторов в результате внедрения «плохих» данных описано в [1, 2]. В [3] отмечается, что экспериментальное воздействие на обучающую выборку в нейронных сетях даже при сильной защите приводит к падению точности модели на 11 % при введении 3 % неверных обучающих данных.

Целью вредоносного воздействия на входные данные является попытка обмануть обученную и протестированную модель машинного обучения, заставить выдавать нужный прогноз или результат. Так в [4] «подмешивание» невидимого человеческому глазу шума к исходному изображению панды (вероятность распознавания 57,7 %) привело к распознаванию изображения как гиббона с вероятностью 99,3 %. В [5] осуществлен обман системы распознавания лиц в результате добавления к изображению оправы для очков.

Нарушение конфиденциальности — возможность злоумышленника извлекать сведения о модели машинного обучения (структура, связи, весовые коэффициенты и др.) которые в дальнейшем могут быть применены для решения задач нарушения целостности [6]. Кроме того, могут быть извлечены и ценные для собственника и злоумышленника данные, на которых алгоритм обучался. В лабораторных атаках проведено успешное получение номеров кредитных карт и договоров социального страхования [7], статистической информации о базовых данных обучения [8], изображений лиц [9], геномной информации о пациентах [10].

При разработке программного обеспечения и систем любого рода существуют риски безопасности. Для создания безопасной системы машинного обучения эти риски необходимо учитывать на каждом этапе ее жизненного цикла. Стоит отметить, что в настоящее время существуют отдельные способы обеспечения безопасности, однако ни один из них до готового программного решения не доведен [11].

В настоящей работе в первую очередь изучена проблема построения моделей нарушения безопасности и исследование противоправных действий в отношении систем машинного обучения со стороны злоумышленника.

Выполнена первичная формализация исследуемого процесса путем его функциональной структуризации. Проведена декомпозиция целевой функции «Нарушение безопасности системы машинного обучения», выделение функциональных элементов и взаимосвязи между ними. Представлено описание функциональной модели средствами математической логики и выведение формулы вычисления вероятности нарушения безопасности системы машинного обучения.

Результаты данной работы позволят специалистам по информационной безопасности и разработчикам получить численную оценку уровня защищенности систем машинного обучения.

Функциональная модель нарушения безопасности

Осуществим моделирование функций, выполняемых рассматриваемой информационной системой, путем создания структурированного графического

изображения с использованием технологии создания функциональных моделей SADT (Structured Analysis a nd Design Technique) и методологии IDEF0[1] (Integration definition for function modeling) в соответствии со сценариями угроз безопасности информации УБИ.218-УБИ.222[2].

Построим контекстную диаграмму с определяющей функцией «Нарушение безопасности» (S0) (рис. 1).

Определим в качестве входных данных модели – обучающие данные алгоритма, входные данные модели и данные о модели машинного обучения (структура нейронной сети, связи между узлами, весовые значения и пр.). Выходными данными модели будут служить конфиденциальные данные, нерабочая модель, измененные характеристики модели, неверный результат/прогноз.

В качестве управляющих компонентов примем порядок обучения алгоритма и порядок функционирования модели машинного обучения. В качестве воздействующего механизма используем действия злоумышленника.

Результаты декомпозиции контекстной диаграммы, содержащей описание совокупности нарушений безопасности системы машинного обучения приведены на рис. 2, где S1 — функция «Нарушение целостности», а S2 — функция «Нарушение конфиденциальности».

Результаты декомпозиции функции «Нарушение целостности» представлены на рис. 3 в виде функций «Изменение параметров обучения» (S3) и «Обход параметров обученной модели» (S4).

Декомпозиция функции «Нарушение конфиденциальности» показана на рис. 4 в виде функций «Раскрытие информации о модели» (S5) и «Подмена модели» (S6).

На данном этапе достигнут максимальный уровень декомпозиции, детализация целевой функции завершена.

[1] ГОСТ Р 50.1.028-2001. Информационные технологии поддержки жизненного цикла продукции. Методология функционального моделирования. Введ. 01.07.2002. М.: ИПК Издво стандартов, 2001. 54 с.

[2] Федеральная служба по таможенному и экспортному контролю (ФСТЭК России). Банк данных угроз безопасности информации [Электронный ресурс]. Режим доступа: http://bdu.fstec.ru/threat/, свободный. Яз. рус. (дата обращения: 22.05.2021).

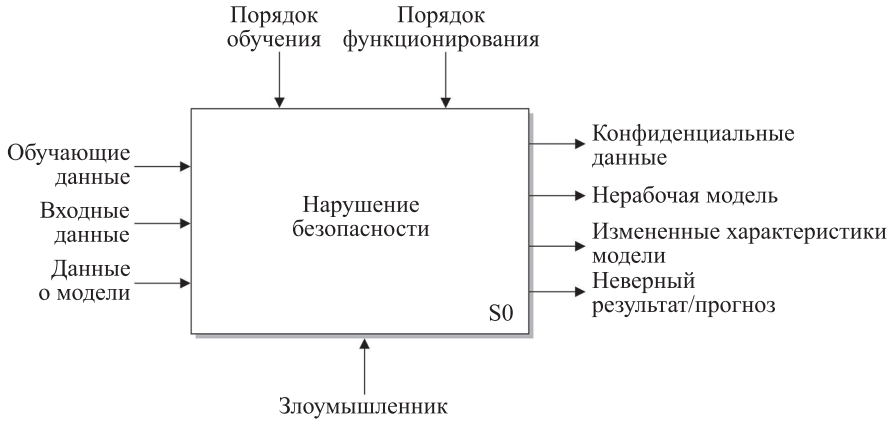

Рис. 1. Контекстная диаграмма системы

Fig. 1. System context diagram

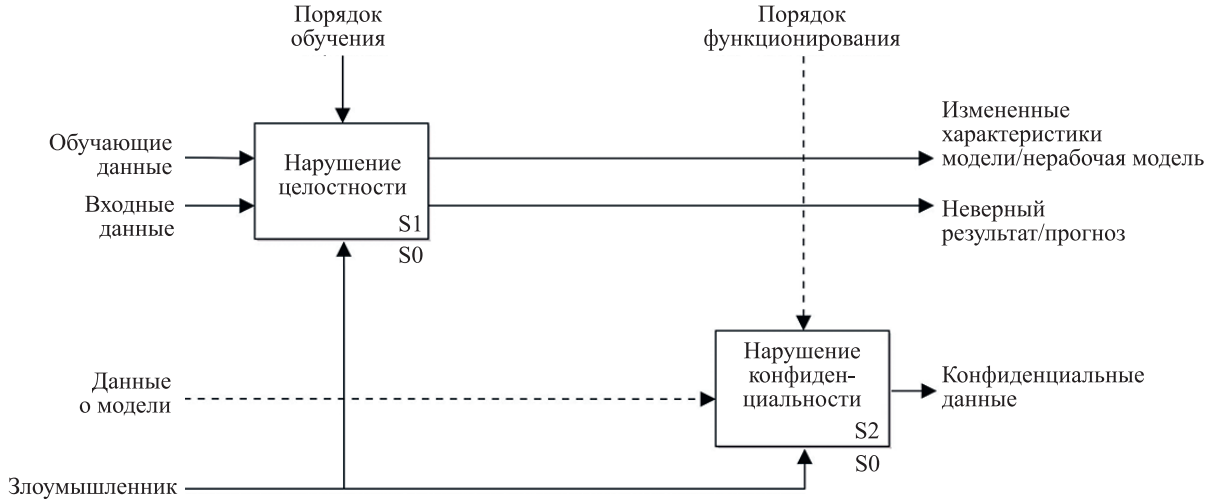

Рис. 2. Декомпозиция функции «Нарушение безопасности»

Fig. 2. Decomposition of the “Security Breach” function

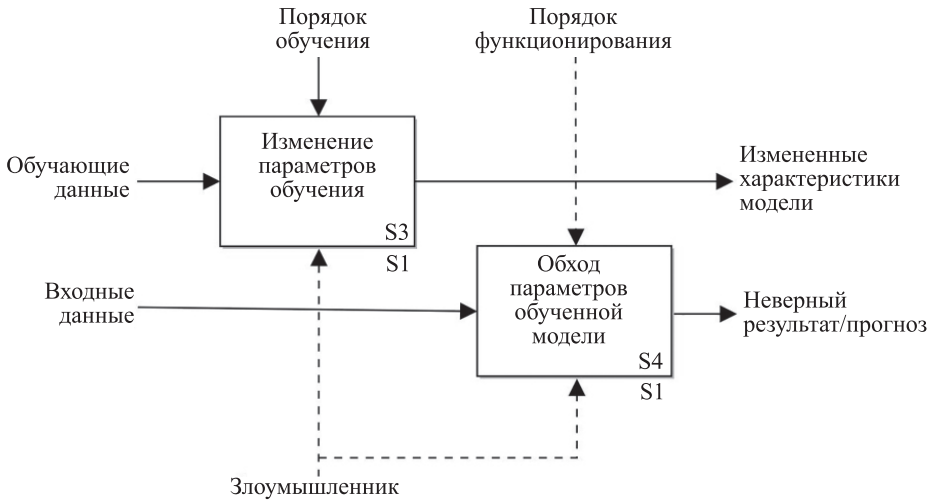

Рис. 3. Декомпозиция функции «Нарушение целостности»
Fig. 3. Decomposition of the “Integrity Violation” function

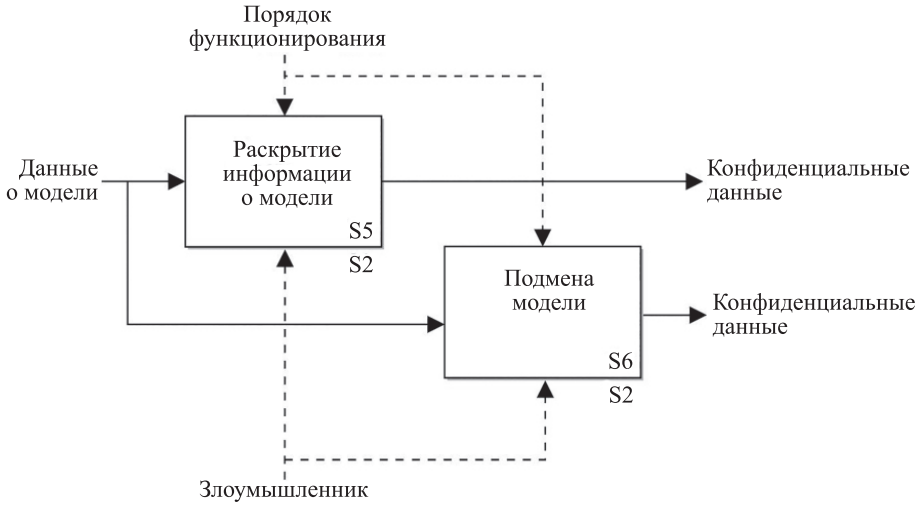

Рис. 4. Декомпозиция функции «Нарушение конфиденциальности»
Fig. 4. Decomposition of the “Privacy Breach” function

Построенная функциональная модель нарушения безопасности информации в результате вредоносного воздействия на систему машинного обучения является инструментом перехода от вербально-функционального описания исследуемого процесса к описанию в терминах математического представления и дальнейшему созданию логических и вероятностных моделей надежности и безопасности.

Логико-вероятностная модель нарушения безопасности

Осуществим логико-вероятностное моделирование нарушения безопасности информации системы машинного обучения в соответствии с алгоритмом, описанном в [12].

Сценарий перехода системы машинного обучения в опасное состояние показан на рис. 5. На схеме: z_i — булева переменная, обозначающая i-ое инициирующее событие; f_j — булева переменная, обозначающая j-ое опасное состояние; z_1 — вредоносное воздействие на обучающую выборку; z_2 — вредоносное воздействие на входные данные; z_3 — извлечение данных; z_4 — функционирование модели; z_5 — обучение алгоритма; f_{11} — изменение параметров обучения; f_{12} — обход параметров обученной модели; f_{13} — раскрытие информации о модели; f_{14} — подмена модели; f_{21} — нарушение целостности; f_{22} — нарушение конфиденциальности; f_{31} — нарушение безопасности.

В логико-вероятностной теории безопасности аналитическое описание опасного состояния системы осуществляется логической *функцией опасности системы* с помощью *кратчайших путей опасного функционирования*, представляющих собой такую конъюнкцию инициирующих событий z_i, ни одну из компонент которой нельзя изъять, не нарушив опасного функционирования системы. При этом $z_i = 0$, если инициирующее событие не наступило, и $z_i = 1$ — в обратном случае [13].

Опишем функцию опасности системы применительно к сценарию (рис. 5):

$$F = f_{31} = f_{21} \vee f_{22} = (f_{11} \vee f_{12}) \vee (f_{13} \vee f_{14}) = \\ = ((z_1 z_5) \vee (z_2 z_4) \vee ((z_3 z_4)) \vee (z_3 z_4)) = \\ = z_1 z_5 \vee z_2 z_4 \vee z_3 z_4 \vee z_3 z_4 = z_1 z_5 \vee z_2 z_4 \vee z_3 z_4. \quad (1)$$

Составим таблицу истинности логической функции опасности системы (1) (таблица), отражающую все значения функции при всех возможных значениях ее аргументов. При этом количество возможных значений функции (количество наборов) для n-переменных составляет 2^n [14].

На наборах 3, 7, 11, 15, 19, 23, 27, 31, когда аргументы z_4 и z_5 принимают значения равные 1, функция опасности системы (1) имеет неопределенное значение, так как события z_4 и z_5 являются несовместными — появление одного из них исключает появление другого [15]. Применительно к системе машинного обучения это объясняется тем, что она может находиться одновременно только в одном состоянии — обучаться или функционировать.

В соответствии с [12] представим неполностью определенную функцию опасности системы с несовместными событиями z_4 и z_5 арифметическим полиномом вероятностной функции с учетом вхождения их в группы совместных событий (ГСС), и используя известные соотношения [16]:

$$Q_1 \wedge Q_2 = Q_1 Q_2, \quad Q_1 \vee Q_2 = Q_1 + Q_2 - Q_1 Q_2, \quad (2)$$

где Q_i — вероятность наступления события.

В соответствии с (2) построим арифметические полиномы вероятностной функции для ГСС_1, состоящей из событий z_1 и z_5 и ГСС_2, состоящей из событий z_2, z_3 и z_4:

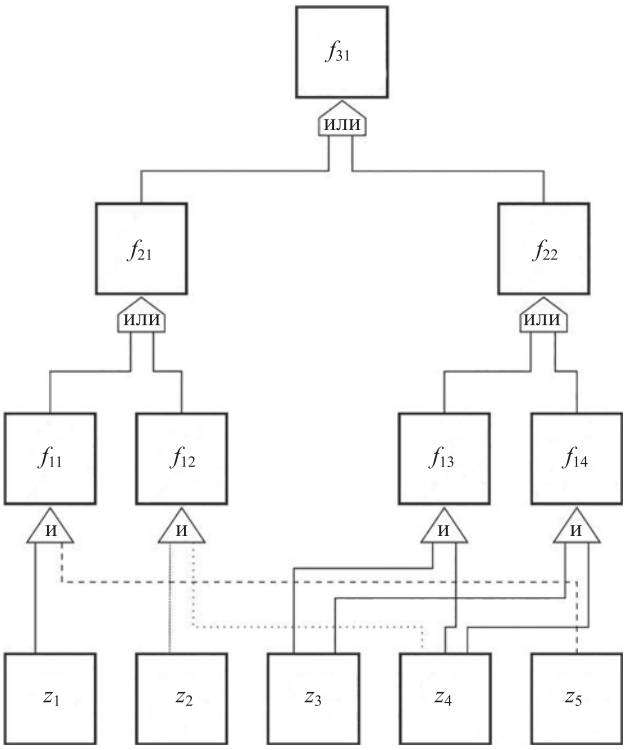

Рис. 5. Сценарий перехода системы машинного обучения в опасное состояние

Fig. 5. Scenario of the transition of a machine learning system to a dangerous state

$$P_1 = \{f(z_1 \wedge z_5)\} = Q_1 Q_5, \\ P_2 = \{f(z_2 z_4 \vee z_3 z_4)\} = Q_2 Q_4 + Q_3 Q_4 - Q_2 Q_3 Q_4, \quad (3)$$

где Q_i — вероятность перехода в опасное состояние i-го инициирующего события.

Таблица. Таблица истинности функции опасности системы

Table. Truth table for the hazard function of the system

Номер набора	z_1	z_2	z_3	z_4	z_5	$f(z_1 \ldots z_5)$	Номер набора	z_1	z_2	z_3	z_4	z_5	$f(z_1 \ldots z_5)$
0	0	0	0	0	0	0	16	1	0	0	0	0	0
1	0	0	0	0	1	0	17	1	0	0	0	1	1
2	0	0	0	1	0	0	18	1	0	0	1	0	0
3	0	0	0	1	1	–	19	1	0	0	1	1	–
4	0	0	1	0	0	0	20	1	0	1	0	0	0
5	0	0	1	0	1	0	21	1	0	1	0	1	1
6	0	0	1	1	0	1	22	1	0	1	1	0	1
7	0	0	1	1	1	–	23	1	0	1	1	1	–
8	0	1	0	0	0	0	24	1	1	0	0	0	0
9	0	1	0	0	1	0	25	1	1	0	0	1	1
10	0	1	0	1	0	1	26	1	1	0	1	0	1
11	0	1	0	1	1	–	27	1	1	0	1	1	–
12	0	1	1	0	0	0	28	1	1	1	0	0	0
13	0	1	1	0	1	0	29	1	1	1	0	1	1
14	0	1	1	1	0	1	30	1	1	1	1	0	1
15	0	1	1	1	1	–	31	1	1	1	1	1	–

Из (3) выводим арифметический полином вероятностной функции нарушения безопасности системы машинного обучения:

$$P = P_1 + P_2 = Q_1Q_5 + Q_2Q_4 + Q_3Q_4 - Q_2Q_3Q_4, \quad (4)$$

где P — вероятность нарушения безопасности системы.

С учетом несовместности z_4 и z_5 для (4) должно выполняться условие $Q_4 + Q_5 \le 1$ [15], а применительно к системе машинного обучения $Q_4 = 0$, $Q_5 = 1$ на этапе обучения алгоритма и $Q_4 = 1$, $Q_5 = 0$ на этапе функционирования модели.

Примеры реализации модели

Определим достоверность полученных результатов, рассмотрев два противоположных случая.

1. Защита системы машинного обучения от угроз злоумышленника полностью отсутствует, т. е. $Q_1 = Q_2 = Q_3 = 1$. В таком случае результат вычисления формулы (4) должен быть равен 1. Подставляя указанные значения в (4) получим:
   — для этапа обучения алгоритма системы:
   $$P = 1\cdot1 + 1\cdot0 + 1\cdot0 - 1\cdot1\cdot0 = 1;$$
   — для этапа функционирования модели системы:
   $$P = 1\cdot0 + 1\cdot1 + 1\cdot1 - 1\cdot1\cdot1 = 2 - 1 = 1.$$
2. Система машинного обучения защищена полностью, т. е. $Q_1 = Q_2 = Q_3 = 0$. В таком случае результат вычисления формулы (4) должен быть равен 0. Подставляя указанные значения в (4) получим:
   — для этапа обучения алгоритма системы:
   $$P = 0\cdot1 + 0\cdot0 + 0\cdot0 - 0\cdot0\cdot0 = 0;$$
   — для этапа функционирования модели системы:
   $$P = 0\cdot0 + 0\cdot1 + 0\cdot1 - 0\cdot0\cdot1 = 0,$$

что подтверждает валидность предложенных моделей.

Очевидно, что при разработке систем машинного обучения необходимо внедрение таких алгоритмов и моделей, которые позволяют минимизировать значение P, а в идеале свести его к 0.

Заключение

Получена наглядная функциональная модель нарушения безопасности системы в виде контекстной диаграммы системы и двух уровней декомпозиции. Определена функция опасности системы и выведен арифметический полином вероятностной функции нарушения безопасности системы машинного обучения, позволяющий на практике получать численное значение вероятности перехода ее в опасное состояние.

Следуют отметить, что моделирования подобного рода в отношении систем машинного обучения до настоящего времени не проводились. Это можно обусловить тем, что проблема обеспечения безопасности таких систем сравнительно нова, а угрозы сформулированы и внесены в Банк данных ФСТЭК России только в декабре 2020 года.

Таким образом, особенность и преимущества описанных моделей состоит в том, что в настоящее время они являются единственными решениями, позволяющими в дальнейшем разработать методы и алгоритмы защиты систем машинного обучения от вредоносных воздействий, а также применять их в ходе оценки уровня защищенности.

Литература

1. Nelson B., Barreno M., Chi F.J., Joseph A.D., Rubinstein B.I.P., Saini U., Sutton C., Tygar J.D., Xia K. Exploiting machine learning to subvert your spam filter // Proc. of First USENIX Workshop on Large-Scale Exploits and Emergent Threats. 2008 [Электронный ресурс]. URL: https://people.eecs.berkeley.edu/~tygar/papers/SML/Spam_filter.pdf (дата обращения: 25.03.2021).
2. Biggio B., Nelson B., Laskov P. Poisoning attacks against support vector machines // Proc. of the 29th International Conference on Machine Learning (ICML 2012). 2012. P. 1807–1814 [Электронный ресурс]. URL: https://icml.cc/2012/papers/880.pdf (дата обращения: 25.03.2021).
3. Steinhardt J., Koh P.W., Liang P. Certified defenses for data poisoning attacks // Advances in Neural Information Processing Systems. 2017. V. 30. P. 3518–3530.
4. Goodfellow I.J., Shlens J., Szegedy C. Explaining and harnessing adversarial examples // Proc. of the 3rd International Conference on Learning Representations. 2015.
5. Sharif M., Bhagavatula S., Bauer L., Reiter M.K. Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition // Proc. of the 23th ACM Conference on Computer and Communications Security (CCS). 2016 [Электронный ресурс]. URL: https://www.cs.cmu.edu/~sbhagava/papers/face-rec-ccs16.pdf (дата обращения: 11.04.2021). https://doi.org/10.1145/2976749.2978392
6. Tramèr F., Zhang F., Juels A., Reiter M.K., Ristenpart T. Stealing machine learning models via prediction APIs // Proc. of the 25th USENIX Conference on Security Symposium. 2016. P. 601–608.
7. Carlini N., Liu C., Erlingsson Ú., Kos J., Song D. The secret sharer: Evaluating and testing unintended memorization in neural networks // Proc. of the 28th USENIX Security Symposium. 2019. P. 267–284.

References

1. Nelson B., Barreno M., Chi F.J., Joseph A.D., Rubinstein B.I.P., Saini U., Sutton C., Tygar J.D., Xia K. Exploiting machine learning to subvert your spam filter. *Proc. of First USENIX Workshop on Large-Scale Exploits and Emergent Threats*, 2008. Available at: https://people.eecs.berkeley.edu/~tygar/papers/SML/Spam_filter.pdf (accessed: 25.03.2021).
2. Biggio B., Nelson B., Laskov P. Poisoning attacks against support vector machines. *Proc. of the 29th International Conference on Machine Learning (ICML 2012)*, 2012, pp. 1807–1814. Available at: https://icml.cc/2012/papers/880.pdf (accessed: 25.03.2021).
3. Steinhardt J., Koh P.W., Liang P. Certified defenses for data poisoning attacks. *Advances in Neural Information Processing Systems*, 2017, vol. 30, pp. 3518–3530.
4. Goodfellow I.J., Shlens J., Szegedy C. Explaining and harnessing adversarial examples. *Proc. of the 3rd International Conference on Learning Representations*, 2015.
5. Sharif M., Bhagavatula S., Bauer L., Reiter M.K. Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. *Proc. of the 23th ACM Conference on Computer and Communications Security (CCS)*, 2016. Available at: https://www.cs.cmu.edu/~sbhagava/papers/face-rec-ccs16.pdf (accessed: 11.04.2021). https://doi.org/10.1145/2976749.2978392
6. Tramèr F., Zhang F., Juels A., Reiter M.K., Ristenpart T. Stealing machine learning models via prediction APIs. *Proc. of the 25th USENIX Conference on Security Symposium*, 2016, pp. 601–608.
7. Carlini N., Liu C., Erlingsson Ú., Kos J., Song D. The secret sharer: Evaluating and testing unintended memorization in neural networks. *Proc. of the 28th USENIX Security Symposium*, 2019, pp. 267–284.

8. Ateniese G., Mancini L.V., Spognardi A., Villani A., Vitali D., Felici G. Hacking smart machines with smarter ones: How to extract meaningful data from machine learning classifiers // International Journal of Security and Networks. 2015. V. 10. N 3. P. 137–150. https://doi.org/10.1504/IJSN.2015.071829
9. Fredrikson M., Jha S., Ristenpart T. Model inversion attacks that exploit confidence information and basic countermeasures // Proc. of the 22nd ACM SIGSAC Conference on Computer and Communications Security. 2015. P. 1322–1333 [Электронный ресурс]. URL: https://rist.tech.cornell.edu/papers/mi-ccs.pdf (дата обращения: 25.03.2021). https://doi.org/10.1145/2810103.2813677
10. Fredrikson M., Lantz E., Jha S., Lin S., Page D., Ristenpart T. Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing // Proc. of the 23rd USENIX Security Symposium. 2014. P. 17–32 [Электронный ресурс]. URL: https://www.usenix.org/system/files/conference/usenixsecurity14/sec14-paper-fredrikson-privacy.pdf (дата обращения: 11.04.2021).
11. Запечников С.В. Модели и алгоритмы конфиденциального машинного обучения // Безопасность информационных технологий. 2020. Т. 27. № 1. С. 51–67. https://doi.org/10.26583/bit.2020.1.05
12. Финько О.А., Соколовский Е.П. Алгоритм оценки информационной безопасности в системах защиты информации на основе логико-вероятностного метода И.А. Рябинина // Известия ЮФУ. Технические науки. 2013. № 12(149). С. 172–178.
13. Рябинин И.А. Надежность и безопасность структурно-сложных систем: монография. СПб.: Политехника, 2000. 248 с.
14. Колмогоров А.Н., Драгалин А.Г. Математическая логика: Введение в математическую логику: учебное пособие. М.: ЛЕНАНД, 2017. 240 с.
15. Мхитарян В.С., Астафьева Е.В., Миронкина Ю.Н., Трошин Л.И. Теория вероятностей и математическая статистика: учеб. пособие / под ред. В.С. Мхитаряна. 2-е изд., перераб. и доп. М.: Московский финансово-промышленный университет «Синергия», 2013. 336 с. (Университетская серия).
16. Финько О.А. Модулярная арифметика параллельных логических вычислений: монография / под ред. В.Д. Малюгина. М.: Институт проблем управления им. В.А. Трапезникова РАН, 2003. 224 с.

8. Ateniese G., Mancini L.V., Spognardi A., Villani A., Vitali D., Felici G. Hacking smart machines with smarter ones: How to extract meaningful data from machine learning classifiers. *International Journal of Security and Networks*, 2015, vol. 10, no. 3, pp. 137–150. https://doi.org/10.1504/IJSN.2015.071829
9. Fredrikson M., Jha S., Ristenpart T. Model inversion attacks that exploit confidence information and basic countermeasures. *Proc. of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 2015, pp. 1322–1333 Available at: https://rist.tech.cornell.edu/papers/mi-ccs.pdf (accessed: 25.03.2021). https://doi.org/10.1145/2810103.2813677
10. Fredrikson M., Lantz E., Jha S., Lin S., Page D., Ristenpart T. Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing. *Proc. of the 23rd USENIX Security Symposium*, 2014, pp. 17–32. Available at: https://www.usenix.org/system/files/conference/usenixsecurity14/sec14-paper-fredrikson-privacy.pdf (accessed: 11.04.2021).
11. Zapechnikov S.V. Models and algorithms of privacy-preserving machine learning. *IT Security (Russia)*, 2020, vol. 27, no. 1, pp. 51–67. (in Russian). https://doi.org/10.26583/bit.2020.1.05
12. Finko O.A., Sokolovsky E.P. Risk assessment information security algorithm is based on the I. Ryabinin logical-probabilistic method. *Izvestiya SFedU. Engineering Sciences*, 2013, no. 12(149), pp. 172–178. (in Russian)
13. Riabinin I.A. *Reliability and Safety of Structural Complex Systems*. St. Petersburg, Politehnika Publ., 2000, 248 p. (in Russian)
14. Kolmogorov A.N., Dragalin A.G. *Mathematical logic: Introduction to the Mathematical Logic*. Moscow, LENAND Publ., 2017, 240 p. (in Russian)
15. Mkhitarian V.S., Astafeva E.V., Mironkina Iu.N., Troshin L.I. *Probability Theory and Mathematical Statistics*. Moscow, Synergy University Publ., 2013, 336 p. (in Russian)
16. Finko O. *Modular Arithmetic of Parallel Logic Computing*. Ed by V.D. Maliugin. Moscow, Institute of Control Sciences of the Russian Academy of Sciences Publ., 2003, 224 p. (in Russian)

Авторы

Чекмарев Максим Алексеевич — адъюнкт, Краснодарское высшее военное училище им. С.М. Штеменко, Краснодар, 350063, Российская Федерация, http://orcid.org/0000-0002-6832-9991, max.chek13@gmail.com

Клюев Станислав Геннадьевич — кандидат технических наук, доцент, Краснодарское высшее военное училище им. С.М. Штеменко, Краснодар, 350063, Российская Федерация, http://orcid.org/0000-0002-0534-9143, s.g.klyuev@mail.ru

Шадский Виктор Викторович — адъюнкт, Краснодарское высшее военное училище им. С.М. Штеменко, Краснодар, 350063, Российская Федерация, http://orcid.org/0000-0002-9221-2283, vdvryazan57@yandex.ru

Authors

Maxim A. Chekmarev — PhD Student, Krasnodar Higher Military School, Krasnodar, 350063, Russian Federation, http://orcid.org/0000-0002-6832-9991, max.chek13@gmail.com

Stanislav G. Klyuev — PhD, Associate Professor, Krasnodar Higher Military School, Krasnodar, 350063, Russian Federation, http://orcid.org/0000-0002-0534-9143, s.g.klyuev@mail.ru

Viktor V. Shadskiy — PhD Student, Krasnodar Higher Military School, Krasnodar, 350063, Russian Federation, http://orcid.org/0000-0002-9221-2283, vdvryazan57@yandex.ru

Статья поступила в редакцию 27.04.2021
Одобрена после рецензирования 30.05.2021
Принята к печати 16.07.2021

Received 27.04.2021
Approved after reviewing 30.05.2021
Accepted 16.07.2021