УДК 004.852 EDN SBLLSK

Машинное обучение для защиты от аномалий в программном обеспечении

Сергей Викторович Зыков[1], Мария Александровна Золотухина[2*], Святослав Александрович Золотухин[2]

[1]МИРЭА – Российский технологический университет, 119454, г. Москва, Проспект Вернадского, д. 78
Национальный исследовательский университет «Высшая школа экономики», 109028, г. Москва, Покровский бульвар, д. 11
[2]МИРЭА – Российский технологический университет, 119454, г. Москва, Проспект Вернадского, д. 78

*E-mail: rtu_mary@mail.ru

Аннотация. В статье приводятся примеры неисправностей в программном обеспечении и установление лучшего метода по производительности для анализа таких событий, как аномалии в программных средах, путей передачи информации и пост-фактор благоприятных и неблагоприятных событий. Также описывается подготовка к исследованиям применения искусственного интеллекта на практике, а именно использование логистической регрессии. Таким образом являясь наилучшим методом анализа и управления в системе предприятия, он позволяет ограничить или вовсе уменьшить аномалии. Рассмотрены вопросы устойчивости предприятия перед нынешними недостатками в системах, а также осуществление мониторинга и защиты информации, которые способствуют эффективному анализу данных.

Ключевые слова: машинное обучение, анализ данных, программная инженерия, защита данных.

Machine learning to protect against anomalies in software

Sergey Viktorovich Zykov[1], Maria Alexandrovna Zolotukhina[2*], Svyatoslav Alexandrovich Zolotukhin[2]

[1]MIREA – Russian Technological University, 119454, Moscow, Prospekt Vernadskogo, 78
National Research University Higher School of Economics, 109028, Moscow, Pokrovsky Boulevard, 11
[2]МИРЭА – Российский технологический университет, 119454, г. Москва, Проспект Вернадского, д. 78

*E-mail: rtu_mary@mail.ru

Abstract. The article provides examples of software malfunctions and the establishment of the best performance method for analyzing events such as anomalies in software environments, information transmission paths and the post-factor of favorable and unfavorable events. It also describes the preparation for research on the use of artificial intelligence in practice, namely the use of logistic regression. Thus, being the best method of analysis and management in the enterprise system, it allows you to limit or even reduce anomalies. The issues of sustainability of the enterprise before the current shortcomings in the systems, as well as the monitoring and protection of information, which contribute to the measured data analysis, are considered.

Keywords: machine learning, data analysis, software engineering, data protection.

1. Введение

Управление данными, а также управление бизнес – процессами и техническим оборудованием на предприятии должны быть защищенными на уровне всех структур, а именно, с помощью методики автоматизированных систем. Передача информации – это составляющая всей структуры предприятия [1]. За каждой командой стоит запрос в систему и ответ на нее, и если возникнет неисправность, то по цепочке будут распадаться все системы предприятия в условиях, несущих потери как с производственной стороны, так и с информационной. Важным аспектом структуры защиты программного обеспечения является обработка и анализ благоприятных и неблагоприятных событий, происходящих во внешних системах и во внутренних процессах обработки. Подача заявки пользователем или отправка любого документа являются внешними составляющими в организации, все программы, работающие в пределах одного предприятия, будут считаться внутренними [2].

2. Постановка задачи

Проблему можно обосновать, используя статистику прошлых лет по обнаружению аномалий в средах технического использования и производительности, а именно сбор статистических данных [3, 4]. На таком подходе строится датасет, далее используются методы искусственного интеллекта. Проблема заключается в том, что эффективности интеллектуального анализа данных часто не хватает для обработки и определения событий на основе машинного обучения, которое однообразно идентифицирует данные, следовательно, развития методики нет [5, 6]. Информация однообразна только если это диагностические величины, собранные с датчиков для определения параметров оборудования. В таком случае все категории данных описаны в документации и интеллектуальный анализ только способствует прогнозированию штатных случаев дисбалансов в технических установках [7, 8]. Следовательно, требуется присутствие человека для перенастройки параметров системы или оборудования. Это касается и анализа критериев аномалий в программном обеспечении. Для того, чтобы были изменения, т.е. развитие в решении проблемы идентификации аномалий со свойственной им структурой изменчивости, требуется нестандартный, инновационный подход к решению данного вопроса [9].

3. Методы и материалы исследования

Применение методов искусственного интеллекта, а именно, алгоритмов машинного обучения, интеллектуального анализа данных и обработки данных, позволяет определить зависимости параметров присутствующих неисправностей, аномалий и дефектов в программном обеспечении. Так для качественного результата нужно создать датасет с количественными характеристиками данных параметров ПО. Для высокой производительности алгоритма использовалась аналитическая платформа Loginom в которой посредством блочной структуры строилась модель анализа параметров. Модель организована с помощью блоков:

1. Блок датасета.
2. Блок обработки данных.
3. Блок алгоритма.
4. Блок результатов.

Для определения аномалий проводится анализ зависимостей и детализирование совокупности критериев диагностических данных.

4. Полученные результаты

Данный подход обычно использует интеллектуальные инструменты и платформы обработки данных [10]. Анализ базы знаний программного обеспечения дает определенные преимущества при разъяснении неблагоприятных событий, а именно:

1. Когда в последний раз было обновление ПО.
2. Подсчет статистики использования на рабочем месте.
3. Зафиксированные дефекты и даты их устранения.
4. Способы устранения.

Создание датасета для обучения алгоритма состоит из аномалий, предшествующих событий неисправностей и благоприятных событий [11]. Также многие неисправности связаны с факторами процесса передачи данных, а именно проводные или беспроводные системы. Модемы, USB – флеш накопители и т.д. являются переносчиками вредоносных программ, и соответственно, составляют часть неблагоприятных событий, ведущих к аномалиям. Такая зависимость требует определенного подхода, а именно применения логистической регрессии [12]. Чтобы не было переобучения, выборка делится на тестовую и обучающую в классическом стиле,

а именно 20/80 или 30/70. В данной работе использовалось деление 30/70, следовательно, нейросеть показывает высокую точность обучения [13].

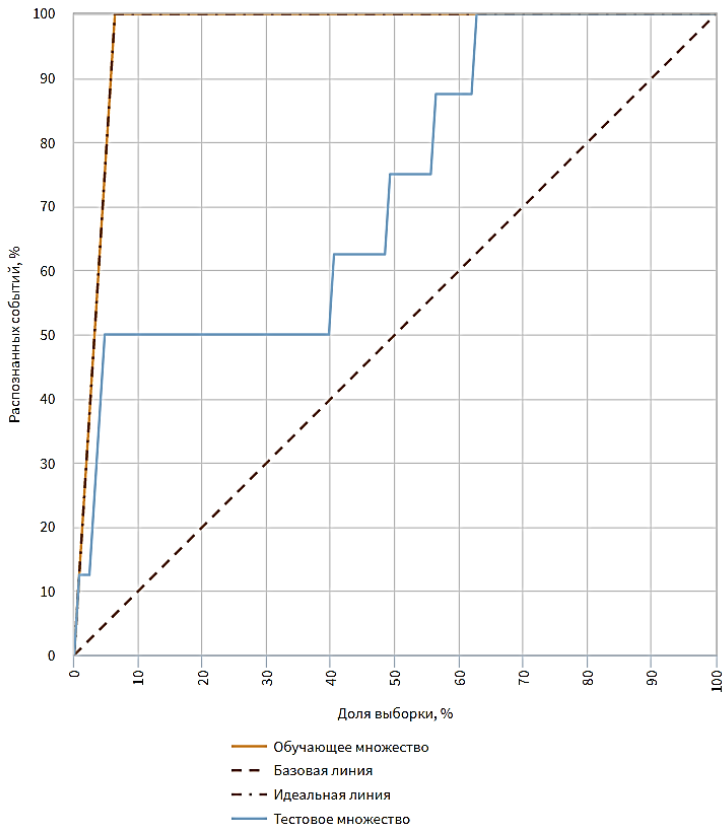

Рисунок 1. Процент распознанных событий.

На (рисунке 1) показаны обучающее и тестовое множество, при сортировке количество выбросов в тестовой выборке увеличилось до 7%. Модель установила 23 неблагоприятных события из 25 в структуре ПО – это также является допустимыми значениями [14]. Неблагоприятными событиями можно считать:

- ошибки в самой программе;
- изменения входной информации;
- действия пользователя;
- технические неисправности ПК.

Устойчивость программного обеспечения определяется быстрым откликом на запросы для одних и тех же задач, прописанных в правилах, осуществляющих контроль над процессом исполнения программ и перенаправление на параллельные модули [15]. Установленные критерии помогут в дальнейшем перенастраивать модули ПО под выданные показатели, тем самым сохраняя ПО в рабочем состоянии.

5. Выводы

Для определения лучшего метода по производительности обработки и выдачи результатов был проведен статистический анализ датасета и выбран метод логистической регрессии. Нейросеть показала высокую эффективность и устойчивость к выдвинутым параметрам. Для защиты ПО от возможных аномалий был применен искусственный интеллект. Определение неблагоприятных событий осуществлялось с помощью логистической регрессии. Таким образом, найденные параметры аномалий и быстрое их устранение дают возможность программному обеспечению проработать дольше с заданной надежностью.

Благодарности

Выражаем благодарность «Российскому технологическому университету – МИРЭА», за предоставленную возможность проводить исследования в области искусственных нейронных сетей, защиты данных и программной инженерии.

Список литературы

1. Луизи, Дж.В. Прагматичная архитектура предприятия: стратегии преобразования информационных систем в эпоху больших данных. Morgan Kaufmann, 2014. – 372 с. – ISBN: 9780128005026
2. Захарова, А. Визуальное обнаружение внутренних закономерностей в эмпирических данных / Алена Захарова, Евгения Вехтер, Алексей Шкляр и др. // Коммуникации в информатике и вычислительной технике; под редакцией Кравца А., Щербакова М., Кульцовой М., Грумпоса П. – Выпуск 754. – Cham: Springer, 2017. – 16 p.
3. Тревор Хасти. Основы статистического обучения: интеллектуальный анализ данных, логический вывод и прогнозирование / Хасти Тревор, Роберт Тибширани, Джером Фридман. – СПб: СПб. : ООО "Диалектика", 2020. – 768 с.
4. Хасти, Т., Тибширани, Р., Фридман Дж. Элементы статистического обучения. Интеллектуальный анализ данных, логический вывод и прогнозирование. 2-е изд. – Springer, 2009. – 745 с.
5. Виттен, И.Х. Data Mining Practical Machine Learning Tools and Techniques / И.Х. Виттен, Фрэнк Эйбе, М.А. Холл et al. // Издательство Моргана Кауфмана; – Netherlands: elsevier, 2017. – 654 p.

6. Бринк Х. Ричардс Дж. Феверолф М. Машинное обучение в реальном мире. – Санкт-Петербург: Питер, 2017. – 336 с. – ISBN: 978-5-496-02989-6
7. Зыков, С.В. Семантическая интеграция данных для безопасности и целостности корпоративных систем // Безопасность информационных технологий. – 2009. – № 3. – С.16-19
8. Бачотти А. Стабильность и управление линейными системами. Cham: Springer, 2019. — 200p. ISBN 978-3-030-02405-5
9. Лин Чжан, Бернард П. Зиглер, Юаньцзюнь Лайли. Разработка моделей для моделирования / Elsevier; 1-е издание, 2019 г. – 453 с.
10. Хинкель, Г. NMF: мультиплатформенный фреймворк моделирования // Международная конференция по теории и практике преобразований моделей. – Springer, Cham, 2018. – 184-194 с. – DOI:10.1007/978-3-319-93317-7_10
11. Бутакова, М.А., Чернов, А.В., Говда, А.Н., Верескун, В.Д., Карташов, О.О. Метод представления знаний для проектирования интеллектуальной системы ситуационного информирования. В: Абрахам А., Ковалев С., Тарасов В., Снасель В., Суханов А. (ред.) Материалы Третьей Международной научной конференции "Интеллектуальные информационные технологии для промышленности" (IITI'18). 2018. Достижения в области интеллектуальных систем и вычислений, том 875. – Springer, Cham. – 225-235 с. doi: 10.1007/978-3-030-01821-4_24.
12. Гудфеллоу, Я., Бенджио, И., Курвилль, А. Глубокое обучение / пер. с анг. А. А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с.
13. Дей Р., Рэй Г., Балас В.Е. Устойчивость и стабилизация линейных и нечетких систем с временной задержкой. Подход с Линейным Матричным Неравенством. Нью–Йорк: Спрингер, 2018. – 274 с. – DOI:10.1007/978-3-319-70149-3 ISBN: 978-3-319-70147-
14. Бурнашев, Р. А. и др. Исследования по разработке экспертных систем с использованием искусственного интеллекта // Международная конференция по архитектуре и технологиям информационных систем. – Springer, Cham, 2019. – 233-242 с.
15. Шолле, Ф. Глубокое обучение на Python. – СПб.: Питер, 2018. – 400 с.