Поиск аномалий с помощью автоэнкодеров

Д.А. Сафронов, Ю.А. Кацер, К.С. Зайцев

***Аннотация*. Целью настоящей работы является сокращение затрат на поиск неисправностей цифрового оборудования путем совершенствования методов распознавания аномалий на основе использования автоэнкодеров. Для этого авторами предлагается использовать два вида автоэнкодеров: глубокий feed-forward автоэнкодер и глубокий сверточный автоэнкодер. В качестве сравнения эффективности предлагается использовать два метода управляемого машинного обучения: логистической регрессии и опорных векторов. Сравнение подтвердило эффективность предложенных автоэнкодеров. Для проведения экспериментов с алгоритмами был выбран массив данных NSL-KDD, характеризующий сетевой поток и включающий более 10 тысяч значений по 41 метрике. Этот датасет содержит данные как нормального, так и аномального сетевого потока. Перед обучением алгоритмов текущий набор данных был обработан, путем исключения корреляций, бинаризации категориальных данных, группировки данных по категориям атак. Результаты применения предложенных авторами автоэнкодеров для поиска аномалий показали свою эффективность.**

***Ключевые слова* – поиск аномалий, машинное обучение, нейронные сети, автоэнкодеры, классификация, сетевые данные.**

1. ВВЕДЕНИЕ

В настоящее время информационные технологии, и, в частности, машинное обучение, используются в различных областях производства и социальной среды, таких как создание приложений для гаджетов, разработка систем анализа и классификации и т.д. Например, в приложениях, связанных с музыкальной сферой, алгоритмы подбирают музыку на основе ранее прослушанных музыкальных произведений.

Статья получена 06 июля 2022.
Сафронов Денис Алексеевич, Национальный Исследовательский Ядерный Университет МИФИ, магистрант, Safronov.thief@yandex.ru
Кацер Юрий Дмитриевич, компания «Цифрум» (Росатом), эксперт отдела ИИ, аспирант Сколтеха, YDKatser@rosatom.ru
Зайцев Константин Сергеевич, Национальный Исследовательский Ядерный Университет МИФИ, профессор, KSZajtsev@mephi.ru

Аналогично выглядит подбор фильмов по указанным предпочтениям в разных жанрах и так далее.

Также существует множество методов машинного обучения (в том числе и глубокого обучения), которые применяются для помощи человеку в повседневной жизни. Например, обработка видеопотоков для дальнейшей рекомендации пользователю [1], распознавание лиц или голосов [2]. Одним из таких применений является сохранение частных данных от воздействия злоумышленников во всемирной паутине. С помощью алгоритмов машинного обучения можно с большой точностью обнаружить аномалии в синхронизированных показаниях многих датчиков [3] или, в сетевой информации [4], что впоследствии поможет обнаруживать и предотвращать поломки в различных системах.

В дополнение к сказанному, можно видеть, что с растущим объемом данных и потребностью в их обработке в областях, допускающих обучение, лучше справляются методы глубокого обучения. Несмотря на все свои недостатки, нейронные сети, и автоэнкодеры в том числе, применяются и для поиска аномалий.

В настоящей статье предлагаются решения задачи детекции аномалий в сетевых данных. Предлагаемые решения представляют собой нейронные сети прямого распространения, т.е. автоэнкодеры.

2. СОПУТСТВУЮЩИЕ РАБОТЫ

Несмотря на широкое применение методов машинного обучения в областях по обнаружению аномалий, в настоящее время все популярнее становится использование методов глубокого обучения для обнаружения аномалий различного рода.

Например, в статье [3] применялся метод прогнозирующего обнаружения аномалий на судовом дизельном двигателе на основе эхо-состояний сети (ESN) и глубокого автоэнкодера. Использование этого метода обусловлено тем, что судовой дизельный двигатель представляет собой

очень сложную машину, и, поскольку он работает в условиях высокой температуры и высокого давления в течение длительного времени, вероятность возникновения нештатных ситуаций относительно велика. В то же время судовые дизели зачастую дороги и требуют высокой надежности при выполнении поставленной задачи, поэтому технологии обнаружения аномалий судовых дизелей уделяется большое внимание.

В статье [5], также рассмотрено применение автоэнкодеров при обслуживании гидроагрегатов ГЭС. Статья поднимает проблему нерационального применения технического обслуживания гидроагрегатов ГЭС, которое приводит к простою оборудования и уменьшению экономического эффекта. С одной стороны, данный подход правильный, так как обеспечивает безопасную эксплуатацию гидроагрегата, но с другой стороны можно реализовать более эффективный способ обслуживания данных гидроагрегатов ГЭС, путем технического обслуживания по состоянию. Он может не только удовлетворить требования своевременного обнаружения потенциальных угроз безопасности без остановки, но и обоснованно предсказать будущую тенденцию агрегата, поэтому обслуживание гидроагрегата будет более целенаправленным и точным [6, 7]. Для такого обслуживания гидроагрегатов применяется метод неконтролируемого обнаружения аномалий с использованием вариационной модальной декомпозиции (VMD) и глубокого автоэнкодера. Автоэнкодер на основе сверточной нейронной сети используется для завершения обучения без учителя, а остаток реконструкции автоэнкодера используется для обнаружения аномалий.

На основании проведенных экспериментов, авторы отмечают, что глубокий автоэнкодер может увеличить интервал между аномальным и нормальным распределением данных, а VMD может эффективно уменьшить количество выборок в области перекрытия. По сравнению с традиционным методом использования автоэнкодера предлагаемый метод улучшает полноту, точность и метрику F1.

Основываясь на проведенном анализе публикаций, можно сделать вывод, что интерес к применению автоэнкодеров в области обнаружения аномалий возрастает. Хотя, можно заметить, что методы построения автоэнкодеров применяется очень избирательно на данных, получаемых с датчиков промышленного оборудования (ГЭС, дизели кораблей и т.п.).

Однако методологии построения автоэнкодеров, часто показывают преимущества перед классическими методами машинного обучения.

3. АЛГОРИТМЫ ВЫЯВЛЕНИЯ АНОМАЛИЙ

В области искусственных нейронных сетей присутствует вид, называемый автоэнкодерами, для которых выходными данными являются восстановленные внутри сети входящие сигналы. То есть, путем сжатия размерности входных данных и построения связей между входными и выходными данными, нейронная сеть может формировать желаемый выходной сигнал, который является менее искаженным, чем входной.

Общую структуру (рис.1) автоэнкодеров можно представить несколькими частями:

- входной слой, содержащий начальные данные;
- энкодер – слой, переводящий входной сигнал (данные) в его представление;
- декодер – слой, восстанавливающий входящие данные по полученному представлению;
- выходной слой, содержащий восстановленные данные после обработки автоэнкодера.

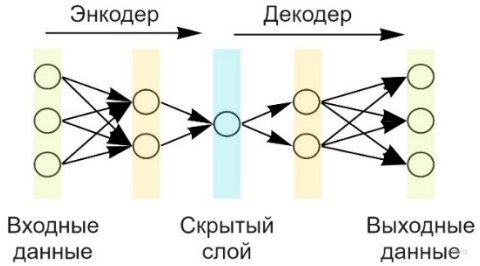

Рис. 1. Схема построения автоэнкодера.

Формально задача, которую выполняет автоэнкодер, поставлена следующим образом:

Пусть x – входной сигнал для автоэнкодера, h(x) – представление входного сигнала, f(h) – выходной сигнал.

Тогда необходимо восстановить сигнал на выходе таким образом, чтобы

$$x' = f(h), \text{ где } h = h(x) \qquad (1)$$

4. ТЕКУЩАЯ РАБОТА

Рассмотрим набор данных и виды автоэнкодеров, которые будем применять в настоящей работе.

4.1. Описание набора данных и среды

Используемый в этой статье набор данных был подготовлен и распространён лабораторий Линкольна Массачусетского технологического института (MIT). Текущий датасет называется NSL-KDD [8] и является наиболее известным и широко используемым набором данных для экспериментов по обнаружению аномалий в компьютерных сетях. Текущий датасет представляет собой подмножество данных DARPA 1998 года, которое было собрано путем моделирования работы типичной локальной сети ВВС США с многочисленными атаками.

Последовательность действий по обработке данных показана на рис. 2.

Рис. 2. Функциональная схема обработки данных.

Используемое программное обеспечение: язык программирования Python, фреймворк Keras с Tensorflow, библиотеки Scikit-learn, Numpy, Pandas.

4.2. Предварительная обработка данных

Реальные данные представляют набор данных, содержащий ошибки, выбросы и по этой причине являются неполными и непоследовательными, поэтому самый важный этап – предварительная обработка данных, реализуется перед началом обучения моделей.

Набор данных, описанный выше имитирует сетевой поток, к которому была применена предварительная обработка данных, включающая следующие этапы.

1. Группировка данных по типу сетевой атаки

При анализе представленного датасета видно, что таргетированный столбец, показывающий нормальность сетевого потока, разделен на множество сетевых атак, которые можно разделить на 4 большие группы:

- DoS,
- Probe,
- R2L,
- U2R.

Так как в данной работе необходимо проверить способность автоэнкодеров классифицировать аномалии по наборам данных (например, обнаруживать DoS атаку), то была реализована бинарная группировка данных, помечаемая «1», в случае сетевой DoS атаки и «0», в случае нормального сетевого потока.

2. Бинаризация категориальных данных

При анализе набора данных NSL-KDD было обнаружено, что в этом датасете присутствуют категориальные данные, которые для глубокого обучения не подходят – это параметры «protocol_type», «service», «flag».

Таким образом, была произведена бинаризация категориальных признаков с целью добавления новых параметров для обучения, которые в дальнейшем помогут повлиять на обучение в лучшую сторону.

3. Анализ корреляции данных

Поскольку в реализации моделей автоэнкодеров будут участвовать линейные функции активации нейрона, а прямая зависимость параметров может повлиять на обучение автоэнкодеров в худшую сторону, то возникла необходимость в построении тепловой карты, с помощью которой был проведен анализ корреляции параметров. По карте было обнаружено, что некоторые параметры имеют высокую зависимость как между собой, так и с таргетированным параметром, характеризующим сетевой поток («норма»/«аномалия»), поэтому такие параметры были исключены из датасета, что в дальнейшем позволит повысить эффективность обучения автоэнкодеров.

В работе были реализованы и исследовались два автоэнкодера - глубокий Feed-Forward и сверточный.

4.4. Глубокий Feed-Forward автоэнкодер

В этом автоэнкодере (рис. 3) реализованы три слоя: энкодера, декодера и один скрытый слой. По мере приближения к скрытому слою количество нейронов в слоях энкодера уменьшается до трех. Обратная ситуация

наблюдается в декодере – количество нейронов увеличивается до размерности входного слоя. Такие параметры предложенной архитектуры автоэнкодера, как количество слоев в энкодере и декодере, количество нейронов в каждом слое показали наименьшую ошибку MAE и оптимальную скорость обучения на «нормальных» данных. Поэтому они были выбраны для дальнейшего обучения глубокого Feed-Forward автоэнкодера на реальных данных, содержащих как нормальные, так и аномальные данные.

4.5. Сверточный автоэнкодер

Реализованный сверточный автоэнкодер – это нейронная сеть прямого распространения, способная заменять полносвязные слои в сверточные. Они, наряду с объединением слоев, преобразуют входные данные из широких и тонких (например, 100x100 пикселей с 3 каналами – RGB) в узкие и тонкие. Этот процесс помогает сети извлекать визуальные особенности из изображений и, следовательно, получать гораздо более точное представление скрытого пространства.

В этом автоэнкодере были реализованы два сверточных слоя в энкодере и декодере (рис. 4). При этом отметим, что минимальная размерность данных реализована в энкодере. Предложенная архитектура сверточного автоэнкодера показала наименьшую ошибку MAE и наименьшую скорость обучения на «нормальных» данных. При увеличении числа слоев в энкодере и декодере в несколько раз увеличивается время обучения сверточного автоэнкодера, что не подходит, например, для приложений ready-production областей. Поэтому именно эта архитектура была выбрана для дальнейшего обучения сверточного автоэнкодера на реальных данных, содержащих нормальные и аномальные данные.

4.6. Метрики оценки

Результаты работы моделей оценивались с помощью следующих метрик:

- точность (precision);
- полнота (recall);
- F-мера;
- AUC (Area Under Curve - площадь под ROC кривой).

Точность – это отношение количества правильно классифицированных параметров к общему числу истинных (отрицательных и положительных) параметров:

$$precision = \frac{TP}{TP + FP}, \qquad (2)$$

где TP (true positive) – истинно положительные результаты, FP (false positive) – ложноположительные результаты.

Полнота – это отношение количества правильно классифицированных параметров к общему числу параметров, которые реально относятся к истинному классу:

$$recall = \frac{TP}{TP + FN}, \qquad (3)$$

где TP (true positive) – истинно положительные результаты, FN (false negative) - ложноотрицательные результаты.

После вычисления значений метрик «точность» и «полнота», может случиться ситуация, при которой полнота классификатора низкая, а точность - высокая. Поэтому необходимо ввести дополнительную метрику, которая является средним гармоническим точности и полноты, т.н. F-мера. Эта метрика вычисляется следующим образом:

$$F_{measure} = \frac{2Precision \times Recall}{Precision + Recall} \qquad (4)$$

Кривая ROC (Receiver Operating Characteristic - рабочая характеристика приёмника) показывает компромисс между такими характеристиками как TPR (процент реальных аномалий, которые были правильно предсказаны классификатором) и FPR (процент реальных нормальных данных, которые были предсказаны классификатором как аномальные). Они вычисляются следующим образом:

$$TPR = \frac{TP}{TP + FN} \times 100\%, \qquad (5)$$

$$FPR = \frac{FP}{TN + FP} \times 100\%, \qquad (6)$$

где TN (true negative) - истинно отрицательные результаты.

Площадь под ROC кривой представлена метрикой AUC, значения которой варьируются от 0.5 до 1. AUC, равная 1, указывает на то, что модель идеальна, а AUC, равная 0,5, - на то, что модель выполняет случайное угадывание.

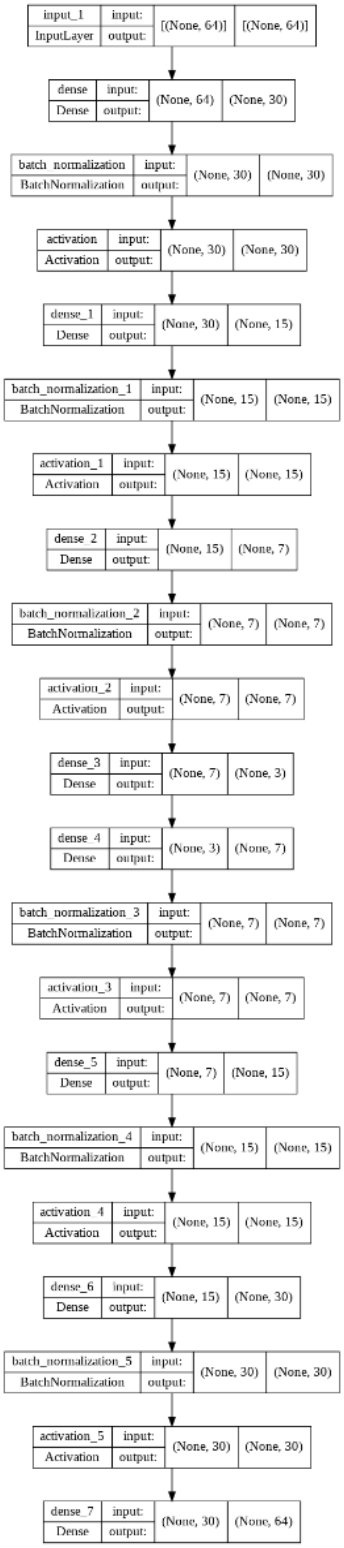

Рис. 3. Архитектура глубокого feed forward автоэнкодера.

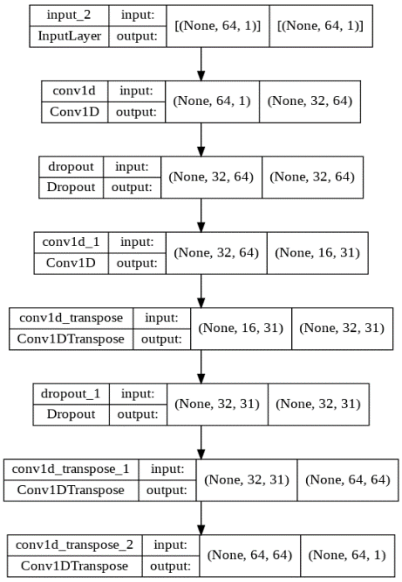

Рис. 4. Архитектура сверточного автоэнкодера.

5. РЕЗУЛЬТАТЫ СРАВНЕНИЯ МОДЕЛЕЙ И ОБСУЖДЕНИЕ

Таблица 1. Оценки алгоритмов.

Алгоритм поиска аномалий	**Виды данных**	**Метрики оценки методов**				
		Precision	Recall	F-мера	AUC	Время обучения
Глубокий Feed Forward автоэнкодер	Норма	0.94	0.93	0.94	0.957	6 мин 28 сек
	Аномалия	0.89	0.89	0.89		
Сверточный автоэнкодер	Норма	0.96	0.94	0.95	0.965	45 мин 47 сек
	Аномалия	0.90	0.93	0.92		
Линейная регрессия	Норма	1.00	0.40	0.57	0.698	1сек
	Аномалия	0.49	1.00	0.66		
Метод опорных векторов	Норма	0.90	0.93	0.92	0.878	3 мин 38 сек
	Аномалия	0.88	0.82	0.85		
Метод опорных векторов (оптимизи-рованный)	Норма	0.99	0.93	0.95	0.952	17 мин 45 сек
	Аномалия	0.88	0.98	0.93		

Было проведено сравнение эффективности детекции аномалий по предложенному датасету различными моделями автоэнкодеров и методами классического машинного обучения. Полученные оценки представлены в Таблице 1. Из анализа таблицы 1 следует.

1. Метрики Precision и Recall автоэнкодеров, с учетом небольшой несбалансированности данных, являются практически одинаковыми:

Несмотря на небольшую разбалансированность, метод опорных векторов смог найти баланс между метриками точности и полноты, но все еще уступает реализованным автоэнкодерам.

2. Для метрики F-мера складывается похожая ситуация, так как F-мера является средним гармоническим между полнотой и точностью.

Можно сказать, что эта метрика показывает, что метод линейной регрессии выделяется на фоне других методов тем, что он не смог справиться с качественной классификацией как нормальных, так и аномальных данных.

3. Анализируя величину AUC-ROC, можно заметить, что лучший показатель показал сверточный автоэнкодер, а далее идет глубокий feed-forward автоэнкодер. Метод опорных векторов уступает всего 10% сверточному автоэнкодеру за счёт своей сложной математической реализации и большим штрафам при классификации.

4. Несмотря на превосходство сверточного автоэнкодера, этот вид нейронной сети сильно уступает остальным алгоритмам по времени обучения. При больших данных и большем количестве слоев в энкодере и декодере этот показатель значительно возрастет, поэтому применимость данного автоэнкодера может быть снижена.

5. При обнаружении аномалий для сравнения с автоэнкодерами был оптимизирован метод опорных векторов. Из полученных результатов видно, что после оптимизации метод опорных векторов его характеристики качества улучшились, но время обучения выросло в 5 раз.

По факту проведенных исследований оказывается, что, несмотря на высокие показатели качества оптимизированного метода опорных векторов, можно получить схожие показатели качества у глубокого feed-forward автоэнкодера с более быстрым обучением.

6. ЗАКЛЮЧЕНИЕ

В работе исследовались подходы к решению задачи обнаружения (детекции) аномалий при обработке сетевого потока данных с помощью двух видов автоэнкодеров: глубокого feed-forward автоэнкодера и сверточного автоэнкодера.

В работе проведен сравнительный анализ эффективности выявления аномалий этими видами автоэнкодеров в сравнении с двумя классическими методами машинного обучения (методом логистической регрессии и методом опорных векторов). Сравнение подтвердило эффективность предложенных автоэнкодеров.

Для проведения испытаний предложенных алгоритмов был выбран NSL-KDD датасет, состоящий из метрик, характеризующих сетевой поток. Этот датасет состоит из нормальных и аномальных значений метрик потоков сетевых данных. Значения датасета были предварительно обработаны для выравнивания влияния метрик и исключения прямых корреляций.

Подводя итог, можно сказать, что разработанные автоэнкодеры показали хорошие результаты и могут являться основой для модификации других видов автоэнкодеров при выявлении аномалий.

БЛАГОДАРНОСТИ

Авторы выражают благодарность Высшей инжиниринговой школе НИЯУ МИФИ за помощь в возможности опубликовать результаты выполненной работы.

БИБЛИОГРАФИЯ

1. Almeida A. et al. The complementarity of a diverse range of deep learning features extracted from video content for video recommendation // Expert Systems with Applications. Pergamon, 2022. Vol. 192. P. 116335.
2. Hammouche R. et al. Gabor filter bank with deep autoencoder based face recognition system // Expert Systems with Applications. Pergamon, 2022. Vol. 197. P. 116743.
3. Qu C. et al. Predictive anomaly detection for marine diesel engine based on echo state network and autoencoder // Energy Reports. Elsevier Ltd, 2022. Vol. 8. P. 998–1003.
4. Ma Q. et al. A novel model for anomaly detection in network traffic based on kernel support vector machine // Computers and Security. Elsevier Ltd, 2021. Vol. 104.
5. Wang H. et al. Anomaly detection for hydropower turbine unit based on variational modal decomposition and deep autoencoder // Energy Reports. Elsevier Ltd, 2021. Vol. 7. P. 938–946.
6. Zhao W. et al. On the use of artificial neural networks for condition monitoring of pump-turbines with extended operation // Measurement: Journal of the International Measurement Confederation. Elsevier B.V., 2020. Vol. 163.
7. Egusquiza M. et al. Advanced condition monitoring of Pelton turbines // Measurement. Elsevier, 2018. Vol. 119. P. 46–55.
8. Protić D. Review of KDD Cup ’99, NSL-KDD and Kyoto 2006+ datasets // Vojnotehnicki glasnik. Centre for Evaluation in Education and Science (CEON/CEES), 2018. Vol. 66, № 3. P. 580–596.

Anomaly detection with autoencoders

D.A. Safronov, Y.D. Kazer, K.S. Zaytsev

Abstract — **The purpose of this work is to reduce the cost of troubleshooting digital equipment by improving anomaly recognition methods based on the use of autoencoders. To do this, the authors propose to use two types of autoencoders: a deep feed-forward autoencoder and a deep convolutional autoencoder, and as a comparison, it is proposed to use two supervised machine learning methods: the logistic regression method and the support vector machine. The comparison confirmed the effectiveness of the proposed autoencoders. The NSL-KDD dataset was chosen for experiments with the algorithms. It includes more than 10,000 measurements and 41 parameters characterizing the network flow. This dataset contains both normal and abnormal network stream data. Before training machine learning algorithms and autoencoders, the current data set was pre-processed: correlation elimination, categorical data binarization, data grouping into attack categories. The results of using autoencoders developed by the authors to search for anomalies have shown the effectiveness**

Keywords – **anomaly detection, machine learning, neural networks, autoencoders, classification, network data.**

REFERENCES

1. Almeida A. et al. The complementarity of a diverse range of deep learning features extracted from video content for video recommendation // Expert Systems with Applications. Pergamon, 2022. Vol. 192. P. 116335.
2. Hammouche R. et al. Gabor filter bank with deep autoencoder based face recognition system // Expert Systems with Applications. Pergamon, 2022. Vol. 197. P. 116743.
3. Qu C. et al. Predictive anomaly detection for marine diesel engine based on echo state network and autoencoder // Energy Reports. Elsevier Ltd, 2022. Vol. 8. P. 998–1003.
4. Ma Q. et al. A novel model for anomaly detection in network traffic based on kernel support vector machine // Computers and Security. Elsevier Ltd,
5. Wang H. et al. Anomaly detection for hydropower turbine unit based on variational modal decomposition and deep autoencoder // Energy Reports. Elsevier Ltd, 2021. Vol. 7. P. 938–946..
6. Zhao W. et al. On the use of artificial neural networks for condition monitoring of pump-turbines with extended operation // Measurement: Journal of the International Measurement Confederation. Elsevier B.V., 2020. Vol. 163.
7. Egusquiza M. et al. Advanced condition monitoring of Pelton turbines // Measurement. Elsevier, 2018. Vol. 119. P. 46–55.
8. Protić D. Review of KDD Cup '99, NSL-KDD and Kyoto 2006+ datasets // Vojnotehnicki glasnik. Centre for Evaluation in Education and Science (CEON/CEES), 2018. Vol. 66, № 3. P. 580–596.