



ИСПОЛЬЗОВАНИЕ ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНЫХ НЕЙРОННЫХ СЕТЕЙ ПРИ ВЫЯВЛЕНИИ АНОМАЛИЙ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА

Рассмотрены генеративно-сопоставительные нейронные сети, а также их применение в задаче выявления аномалий технологического процесса. Для проведения экспериментов использовался набор данных Gas Pipeline, описывающий работу системы бензопровода. В ходе экспериментальных исследований, на примерах, соответствующих нормальному состоянию технологического процесса, была обучена генеративно-сопоставительная модель BiGAN. Показаны преимущества применения генеративно-сопоставительных моделей в задаче выявления аномалий технологического процесса.

Ключевые слова: информационная безопасность, автоматизированные системы управления технологическими процессами, выявление аномалий, глубокое обучение, генеративно-сопоставительные сети.

Alabugin S.K., Sokolov A.N.

GENERATIVE ADVERSARIAL NETWORKS USAGE IN DETECTING ANOMALIES OF THE INDUSTRIAL PROCESS

The paper considers generative adversarial networks (GAN) and their application in industrial process anomaly detection. Experiments were conducted using the Gas Pipeline dataset. This dataset describes the operation of the gas pipeline system. In the course of experimental studies, the generative-adversarial model BiGAN was trained only on examples corresponding to the normal state of the industrial process. The advantages of the use of generative adversarial models in the problem of industrial process anomaly detection are shown.

Keywords: information security, industrial control systems, anomaly detection, deep learning, generative adversarial networks.

Автоматизированные системы управления технологическим процессом (АСУ ТП) часто размещаются на промышленных объектах, от работы которых зависит качество жизни значительного числа граждан и/или экономическое благополучие отдельного региона или страны. По этой причине промышленные объекты и размещенные на них АСУ ТП могут оказаться целью кибератак, приводящих, в том числе, к остановке производства, техногенным катастрофам, материальным и репутационным потерям [1-3].

Поскольку современная АСУ ТП является не только информационной, но и физической системой, один из способов обнаружения вторжений заключается в мониторинге непосредственно технологического процесса и выявление аномалий. Зачастую, аномалия указывает на неправильную работу системы, которая, возможно, вызвана кибератакой.

Среди нескольких имеющихся подходов к выявлению аномалии, основанных на использовании машинного обучения [4] выделяются генеративно-сопоставительные нейронные сети (Generative adversarial network, GAN) [5]. Генеративно-сопоставительные нейронные сети представляют собой специфический класс нейронных сетей, в архитектуре которых имеются две модели: генератор и дискриминатор. Архитектура такой сети представлена на рис. 1.



Рис.1 Схема генеративно-сопоставительной сети

Генератору G на вход подаётся вектор, состоящий из случайного шума – вектор из некоторого пространства скрытых переменных (latent space) Z , на котором задано априорное распределение $p_z(z)$. На выходе генератор выдаёт новый объект из пространства данных X . Формально, сеть-генератор можно описать в виде следующей функции:

$$G = G(z; \theta_g): Z \rightarrow X$$

где θ_g – параметры сети-генератора. В ходе обучения, генератор аппроксимирует распределение p_{data} выборки реальных данных

X . Следовательно, по окончании обучения, распределение порождаемых генератором объектов p_{gen} должно быть приближенно к распределению реальных объектов p_{data} .

Для того, чтобы генератор с каждой итерацией обучения обладал лучшей аппроксимацией распределения p_{data} , используется сеть дискриминатор. Дискриминатор – как правило, является обычным бинарным классификатором, на вход которому подаётся объект x из пространства данных X . На выходе, дискриминатор выдаёт вероятность принадлежности объекта к тому или иному классу: реальный объект или объект, порождённый дискриминатором. Формально, дискриминатор определяется как:

$$D = D(x; \theta_d): X \rightarrow [0, 1]$$

где θ_d – параметры сети-дискриминатора. Для того, чтобы аппроксимировать распределение p_{data} генератору нужно научиться обманывать дискриминатор: научиться порождать такие объекты, которые дискриминатор не в состоянии отличить от настоящих.

В настоящее время генеративно-сопоставительные сети применяются, в частности, для генерации синтетических изображений [6-7]. В ходе обучения, сеть учится аппроксимировать статистическое распределение, задаваемое реальными данными, или, говоря точнее, генератор учится отображать простое распределение (из которого берутся векторы

шума) в сложное, произвольное распределение реальных данных.

Было предложено несколько способов применения GAN в задаче выявления аномалий. Мы рассмотрим один из них, основанный на применении архитектуры Bidirectional Generative Adversarial Network (BiGAN, двунаправленная генеративно-сопоставительная сеть) [8]. Схема архитектуры BiGAN представлена на рис. 2.

В обычную схему GAN добавлена сеть-кодировщик (encoder, E), которая отображает

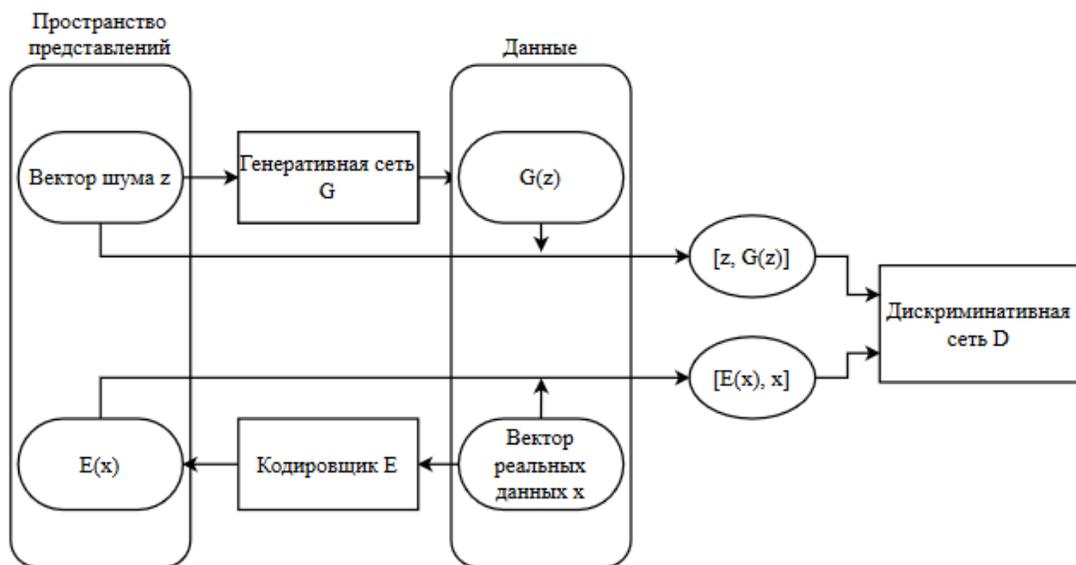


Рис.2 Схема сети BiGAN

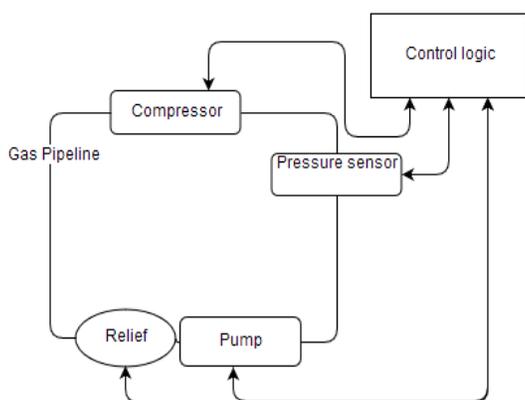


Рис.3 Схема установки Gas Pipeline

ся генератору. Таким образом, выход генератора отображается в пространство реальных данных. После этого мы можем считать метрикой аномальности расстояние до ближайшего примера из реальных данных.

Для проведения экспериментов, в исследовании использовался набор данных Gas Pipeline [9]. В нём представлены данные, соответствующие нормальной работе системы и данные, соответствующие различным атакам. Схема системы представлена на рис. 3.

Система состоит из бензопровода, помпы, компрессора, датчика давления и предохранительного клапана, управляемого соленоидом. Необходимый уровень давления в системе поддерживается с помощью пропорционально-интегрально-дифференцирующего (ПИД) регулятора. Для коммуникации в описанной системе используется протокол прикладного уровня Modbus. Сетевые пакеты

с метками времени после некоторой обработки составляют набор данных, таким образом запись набора данных соответствует сетевому пакету Modbus. Каждая запись содержит значения 16 признаков, некоторые из которых несут в себе информацию о сети (адрес назначения пакета, код функции Modbus и т.д.), а остальные - характеризуют состояние технологического процесса.

Экспериментальные исследования были проведены при помощи библиотеки машинного обучения Tensorflow. Для применения BiGAN были реализованы модели генератора, дискриминатора и кодировщика. Каждая из этих моделей является обычной полносвязной сетью.

Так как в наборе данных часто встречаются пропуски (в частности, среди значений признаков, описывающих состояние технологического процесса), а значения разных признаков имеют существенно различающиеся масштабы, данные были подвергнуты предобработке. Для заполнения пропусков использовались последние известные значения признака. После этого, данные были разделены на обучающую и тестовую выборки: в обучающую выборку было включено 95% всех записей, описывающих нормальную работу системы. Все признаки были нормализованы относительно обучающей выборки.

Обучение модели BiGAN осуществлялось в течении 50 эпох, лучшие результаты были получены на 39-й эпохе. Уже обученная модель тестировалась следующим образом: объект из тестовой выборки подавался на

вход кодировщику, а выхлоп кодировщика подавался генератору. В пространстве данных находилось расстояние между полученным объектом и ближайшим настоящим объектом, соответствующим нормальной работе системы. Это расстояние использовалось в качестве метрики аномальности, и, для получения лучших результатов, был подобрано пороговое значение. Лучшие, полученные после подбора порога, результаты, в сравнении с несколькими уже известными представлены в табл. 1.

Как видно из таблицы, полученные на наборе данных Gas Pipeline результаты, не являются лучшими. Однако, в силу своей специфики, используемый подход не требует для обучения примеров, соответствующих вторжению, и обучался лишь на данных, соответствующих нормальной работе системы. На практике, это является очень важным преимуществом, так как позволяет затратить меньше сил на подготовку обучающего набора данных.

Таблица 1

Результаты экспериментов

Алгоритм	Accuracy	Precision	Recall
Исследуемая модель	0.95	0.93	0.98
Сеть Байеса	0.87	0.97	0.59
Support Vector Data Description	0.76	0.95	0.21
Isolation Forest	0.70	0.51	0.13
K-means	0.57	0.83	0.57
Support Vector Machine (SVM)	0.94	0.94	0.94
Random Forest	0.99	0.99	0.99

Литература

1. Falliere, N., Murchu, L. O., & Chien, E. (2011). W32. Stuxnet Dossier Version 1.4. Symantec Security Response.
2. Lee, R., Assante, M., & Conway, T. (2014). ICS CP/PE (Cyber-to-Physical or Process Effects) case study paper—German steel mill cyber attack. Sans ICS.
3. Lee, R. M., Assante, M. J., & Conway, T. (2016). Analysis of the cyber attack on the Ukrainian power grid. SANS Industrial Control Systems, 23.
4. Асъяев Г. Д., Соколов А. Н. Обнаружение вторжений на основе анализа аномального поведения локальной сети с использованием алгоритмов машинного обучения с учителем // Вестник УрФО. Безопасность в информационной сфере. – 2020. – №. 1 (35). – С. 77-83.
5. Goodfellow I. et al. Generative adversarial nets // Advances in neural information processing systems. – 2014. – С. 2672-2680
6. Reed S. et al. Generative Adversarial Text to Image Synthesis // International Conference on Machine Learning. – 2016. – С. 1060-1069.
7. Karras T., Laine S., Aila T. A style-based generator architecture for generative adversarial networks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2019. – С. 4401-4410.
8. Donahue J., Krähenbühl P., Darrell T. Adversarial feature learning // arXiv preprint arXiv:1605.09782. – 2016.
9. Morris T. H., Thornton Z., Turnipseed I. Industrial control system simulation and data logging for intrusion detection system research // 7th annual southeastern cyber security summit. – 2015. – С. 3-4.

References

1. Falliere, N., Murchu, L. O., & Chien, E. (2011). W32. Stuxnet Dossier Version 1.4. Symantec Security Response.
2. Lee, R., Assante, M., & Conway, T. (2014). ICS CP/PE (Cyber-to-Physical or Process Effects) case study paper—German steel mill cyber attack. Sans ICS.
3. Lee, R. M., Assante, M. J., & Conway, T. (2016). Analysis of the cyber attack on the Ukrainian power grid. SANS Industrial Control Systems, 23.

4. Asyaev G. D., Sokolov A. N. Obnaruzhenie vtorzhenij na osnove analiza anomal'nogo povedenija lokal'noj seti s ispol'zovaniem algoritmov mashinnogo obuchenija s uchitelem //Vestnik UrFO. Bezopasnost' v informacionnoj sfere. – 2020. – no. 1(35). – pp. 77-83.
 5. Goodfellow I. et al. Generative adversarial nets //Advances in neural information processing systems. – 2014. – C. 2672-2680
 6. Reed S. et al. Generative Adversarial Text to Image Synthesis //International Conference on Machine Learning. – 2016. – C. 1060-1069.
 7. Karras T., Laine S., Aila T. A style-based generator architecture for generative adversarial networks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2019. – C. 4401-4410.
 8. Donahue J., Krähenbühl P., Darrell T. Adversarial feature learning //arXiv preprint arXiv:1605.09782. – 2016.
 9. Morris T. H., Thornton Z., Turnipseed I. Industrial control system simulation and data logging for intrusion detection system research //7th annual southeastern cyber security summit. – 2015. – C. 3-4.
-

АЛАБУГИН Сергей Константинович, аспирант кафедры защиты информации высшей школы электроники и компьютерных наук ФГАОУ ВО «Южно-Уральский государственный университет (национальный исследовательский университет)». Россия, 454080, г. Челябинск, проспект Ленина, д. 76. E-mail: sergei_alabugin@mail.ru

СОКОЛОВ Александр Николаевич, кандидат технических наук, доцент, заведующий кафедрой защиты информации высшей школы электроники и компьютерных наук ФГАОУ ВО «Южно-Уральский государственный университет (национальный исследовательский университет)». Россия, 454080, г. Челябинск, проспект Ленина, д. 76. E-mail: sokolovan@susu.ru

ALABUGIN Sergei, postgraduate student of the department of information security of the school of electrical engineering and computer science in FSAEI HE «South Ural State University (national research university)». 76, Lenin prospect, Chelyabinsk, Russia, 454080. E-mail: sergei_alabugin@mail.ru

SOKOLOV Alexander, Ph.D., Associate professor, Head of the department of information security of the school of electrical engineering and computer science in FSAEI HE «South Ural State University (national research university)». 76, Lenin prospect, Chelyabinsk, Russia, 454080. E-mail: sokolovan@susu.ru