УДК 004.056

РАЗРАБОТКА МЕТОДА ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ СИСТЕМ ОБНАРУЖЕНИЯ СЕТЕВЫХ ВТОРЖЕНИЙ С ПРИМЕНЕНИЕМ УСЛОВНЫХ ГЕНЕРАТИВНЫХ МОДЕЛЕЙ ДЛЯ ОБРАБОТКИ НЕСБАЛАНСИРОВАННЫХ ДАННЫХ

Нгуен Т.Ч. (ИТМО), **Фан Н.Т.** (ИТМО), **До Х.Н.Ч.** (ИТМО)
Научный руководитель – кандидат технических наук, доцент Левко И.В. (ИТМО)

Введение. В задачах обнаружения сетевых вторжений обучающие данные обычно имеют сильный дисбаланс: «нормальный» трафик и массовые атаки представлены десятками тысяч наблюдений, тогда как редкие классы атак встречаются крайне редко. В результате стандартные модели машинного обучения демонстрируют низкую полноту по миноритарным классам, а общая точность перестаёт отражать реальную полезность IDS. Классические методы пересэмплирования (например, SMOTE) ограничены при смешанных признаках и высокой кардинальности категорий, что особенно характерно для сетевых датасетов. Поэтому актуальна разработка метода, который повышает качество IDS на редких атаках за счёт контролируемой генерации синтетических примеров.

Основная часть. Предложен метод повышения эффективности IDS на несбалансированных данных с использованием условных генеративных моделей для табличных признаков. Метод включает: (1) предобработку признаков сетевого трафика (обработка пропусков, устойчивое масштабирование числовых признаков, кодирование категориальных; для признаков высокой кардинальности применяется ограничение one-hot и/или ordinal-кодирование во избежание взрыва размерности); (2) обучение условного генератора по метке класса. Рассмотрены две реализации: условная WGAN-GP (cWGAN-GP) как более стабильная модификация GAN [2] и условный вариационный автоэнкодер (CVAE) как устойчивый базовый генератор [3]; (3) контролируемое дополнение обучающей выборки. Для предотвращения ухудшения качества из-за чрезмерной генерации сверхредких классов вводятся ограничения: cap_per_class (верхняя граница целевого размера класса), augment_min_count (генерация только для классов с достаточным числом реальных примеров) и режим augment_only (селективная генерация для выбранных миноритарных классов).

Оценка проводилась в задаче многоклассовой классификации атак на наборах NSL-KDD и UNSW-NB15. Основные метрики — Macro-F1 и Balanced Accuracy, дополнительно анализировалась полнота по классам. Для контроля качества синтетики использовались: TSTR (обучение на синтетике, тестирование на реальных данных), меры близости распределений (KS для числовых и JS для категориальных признаков) и прокси-оценки приватности (отличимость синтетики от реальных записей). Эксперименты показали, что на NSL-KDD условная генерация повышает устойчивые к дисбалансу метрики: Macro-F1 увеличивается с 0,468 до 0,493, а полнота редкого класса r2l возрастает с 0,005 до 0,052. На UNSW-NB15 выявлено, что наивная балансировка всех классов может ухудшать качество из-за сдвига распределения, поэтому применена улучшенная стратегия cap+selective. В результате для CVAE при cap_per_class=5000 и селективном дополнении выбранных миноритарных классов получено устойчивое улучшение по трём запускам (seeds 42/43/44): среднее Macro-F1 возрастает с 0,493 до 0,501, а Balanced Accuracy — с 0,487 до 0,492.

Выводы. Разработан и экспериментально проверен метод повышения эффективности IDS на несбалансированных данных с использованием условных генеративных моделей и контролируемой стратегии дополнения обучающей выборки. Показано, что оптимальная политика генерации зависит от свойств датасета: на NSL-KDD полезна более сильная балансировка, а на UNSW-NB15 требуются ограничения объёма синтетики и селективная генерация для предотвращения деградации качества. Полученные результаты могут быть использованы при подготовке данных и обучении IDS для повышения полноты обнаружения редких типов атак.

Список использованных источников:

1. Mirza M., Osindero S. Conditional Generative Adversarial Nets. 2014.
2. Gulrajani I., Ahmed F., Arjovsky M., Dumoulin V., Courville A. Improved Training of Wasserstein GANs. 2017.
3. Kingma D.P., Welling M. Auto-Encoding Variational Bayes. 2013.
4. Tavallaee M., Bagheri E., Lu W., Ghorbani A.A. A detailed analysis of the KDD CUP 99 data set. 2009.
5. Moustafa N., Slay J. UNSW-NB15: a comprehensive data set for network intrusion detection systems. 2015.