https://doi.org/10.26160/2474-5901-2026-54-65-69

ПРИМЕНЕНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ВЫЯВЛЕНИЯ СЕТЕВЫХ АТАК И ПОВЫШЕНИЕ ОБНАРУЖЕНИЯ РЕДКИХ АТАК БАЛАНСИРОВКОЙ КЛАССОВ

Архипов М.О.
Санкт-Петербургский государственный лесотехнический университет имени С.М. Кирова, Санкт-Петербург, Россия

Ключевые слова: кибербезопасность, обнаружение вторжений, IDS, машинное обучение, сетевые атаки, NSL-KDD, балансировка классов, SMOTE, редкие атаки, защита информации.
Аннотация. В статье оценивается применимость машинного обучения к обнаружению сетевых атак. На наборе NSL-KDD обучены и сопоставлены пять классификаторов – дерево решений, случайный лес, градиентный бустинг, логистическая регрессия и метод k ближайших соседей; проверка выполнена на отдельной тестовой выборке с ранее не встречавшимися атаками. Лучший баланс точности и полноты показал градиентный бустинг (F1 = 0,801), наибольший ROC-AUC – случайный лес (0,960). Анализ по типам атак выявил резкий перекос: массовые атаки DoS и Probe обнаруживаются в 84% случаев, тогда как редкие R2L и U2R почти ускользают (5-15% в зависимости от модели). Балансировка обучающей выборки методом SMOTE повысила обнаружение U2R до 34%, R2L – до 13% при сохранении точности и росте F1.

APPLICATION OF MACHINE LEARNING ALGORITHMS FOR NETWORK ATTACK DETECTION

Arkhipov M.O.
Saint-Petersburg State Forest Technical University named after S.M. Kirov, Saint-Petersburg, Russia

Keywords: cybersecurity, intrusion detection, IDS, machine learning, network attacks, NSL-KDD, class balancing, SMOTE, rare attacks, information security.
Abstract. The paper evaluates the applicability of machine learning to network attack detection. Five classifiers – logistic regression, k-nearest neighbours, decision tree, random forest and gradient boosting – were trained and compared on the NSL-KDD dataset; evaluation was carried out on a separate test set containing previously unseen attacks. Gradient boosting achieved the best balance of precision and recall (F1 = 0.801), while random forest reached the highest ROC-AUC (0.960). Analysis by attack type revealed a sharp imbalance: high-volume DoS and Probe attacks are detected in 84% of cases, whereas rare R2L and U2R attacks almost slip through (5–15% depending on the model). Balancing the training set with SMOTE increased U2R detection to 34% and R2L to 13% while preserving precision and improving F1.

Введение

Объём и изощрённость сетевых атак растут быстрее, чем успевают обновляться сигнатурные средства защиты. Классические системы обнаружения вторжений (IDS) опираются на заранее описанные правила и потому слепы к угрозам, которых раньше не видели. Машинное обучение предлагает другой путь: вместо фиксированных сигнатур – модель, которая сама учится отличать нормальный трафик от вредоносного по статистике

соединений [1, 2]. Но в наивном исполнении подход обманчив: завышенные оценки качества, полученные на «удобных» выборках, в реальной сети рассыпаются [3].

Цель работы – проверить, насколько разные алгоритмы машинного обучения справляются с обнаружением сетевых атак в честных условиях, и нащупать границу их применимости. Мы сравнили пять классификаторов, а затем разобрали, какие типы атак модель ловит уверенно, а какие пропускает.

Данные и постановка задачи

Эксперимент построен на наборе NSL-KDD – переработанной версии классического KDD'99, из которой удалили дубликаты, искажавшие оценки [1]. Каждая запись описывает одно сетевое соединение 41 признаком: длительность, протокол, сервис, число переданных байт, частоты ошибочных флагов и десятки производных статистик. Обучающая часть содержит 125 973 соединения, тестовая – 22 544.

Один момент здесь принципиален. Модели учились на KDDTrain+, а проверялись на отдельном KDDTest+, где есть атаки, не встречавшиеся в обучении, — это ближе к боевым условиям и строже к моделям. Метку соединения свели к бинарной (0 – норма, 1 – атака), сохранив исходную категорию атаки (DoS, Probe, R2L, U2R) для разбора по типам. Категориальные признаки закодировали методом one-hot, размерность выросла до 122.

Методика исследования

В сравнение вошли пять моделей: дерево решений и два его ансамблевых развития – случайный лес и градиентный бустинг, – а также логистическая регрессия и метод k ближайших соседей [4]. Для линейной модели и kNN, опирающихся на расстояния, признаки предварительно стандартизировали. Качество измеряли пятью метриками – accuracy, precision, recall, F1 и ROC-AUC.

У этих метрик разный вес. Высокая precision означает мало ложных тревог, высокий recall – что мимо проходит мало атак; на практике между ними ищут компромисс, и его отражает F1. Эксперимент выполнен на Python с библиотекой scikit-learn.

Результаты и обсуждение

Итоги на тестовой выборке собраны в таблице 1; модели упорядочены по убыванию F1.

Первое, что бросается в глаза, – разрыв между precision и recall. Ложных тревог модели почти не дают: precision держится в диапазоне 0,92-0,97. Но каждая третья атака всё же проходит незамеченной – recall не поднимается выше 0,68. Лучший компромисс показал градиентный бустинг (F1 = 0,801), а по ROC-AUC вперёд вышел случайный лес (0,960). ROC-кривые на рисунке 1 наглядно разводят участников: ансамблевые модели уверенно отрываются от линейной регрессии и одиночного дерева.

Табл. 1. Качество обнаружения сетевых атак на тесте KDDTest+

Модель	Accuracy	Precision	Recall	F1	ROC-AUC
Градиентный бустинг	0,808	0,972	0,682	0,801	0,944
Дерево решений	0,779	0,967	0,633	0,765	0,763
k ближайших соседей	0,770	0,923	0,651	0,763	0,820
Случайный лес	0,765	0,967	0,607	0,746	0,960
Логистическая регрессия	0,753	0,917	0,623	0,742	0,779

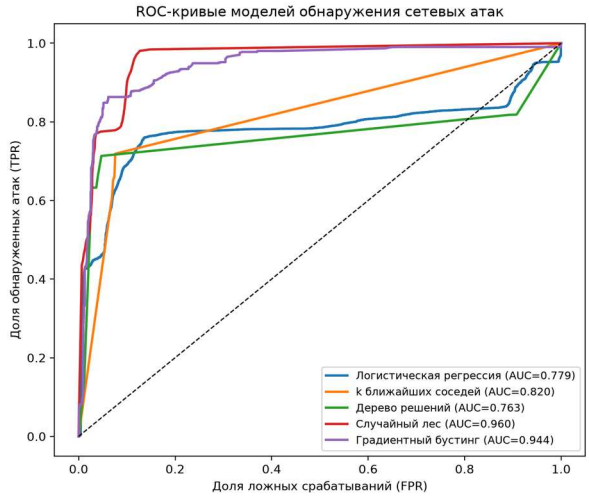

Рис. 1. ROC-кривые моделей обнаружения сетевых атак

Усреднённые метрики, однако, скрывают самое важное – стоит разложить обнаружение по типам атак, и картина становится тревожной. Массовые DoS и Probe модель ловит в 84% случаев, а редкие R2L и U2R почти ускользают – лишь 15%. Причина в том, что подбор пароля (R2L) или повышение привилегий (U2R) по статистике соединения почти неотличимы от обычной активности, да и в обучающих данных их единицы. Вывод для практики: ML-детектор закрывает массовые угрозы, а точечные вторжения требуют отдельных средств. Дополнительно отметим, что решение модели определяют прежде всего объёмные характеристики соединения (число переданных байт, тип сервиса), которые ярко выражены у DoS и сканирования, но не у R2L и U2R.

Повышение детекции редких атак балансировкой классов

Низкое обнаружение R2L и U2R – не приговор, а задача. Их главная беда в малочисленности: в обучающей выборке всего 995 примеров R2L и 52 примера U2R против десятков тысяч DoS. Распознаванию таких атак модель толком негде научиться. Напрашивающееся лекарство – балансировка. Мы взяли случайный лес как базовый детектор (на нём исходное обнаружение редких атак ещё ниже, чем у бустинга: R2L 5%, U2R 10%) и дополнили обучающую выборку синтетическими примерами редких классов методом

SMOTE [5], после чего переобучили модель. Сравнение до и после собрано в таблице 2 и на рисунке 2.

Табл. 2. Обнаружение по типам атак и общие метрики до и после балансировки

Показатель	До (базовая)	После (баланс.)	Δ
DoS	0,787	0,828	+0,041
Probe	0,730	0,772	+0,042
R2L	0,052	0,129	+0,077
U2R	0,104	0,343	+0,239
Precision	0,967	0,969	+0,002
Recall	0,607	0,658	+0,051
F1	0,746	0,784	+0,038

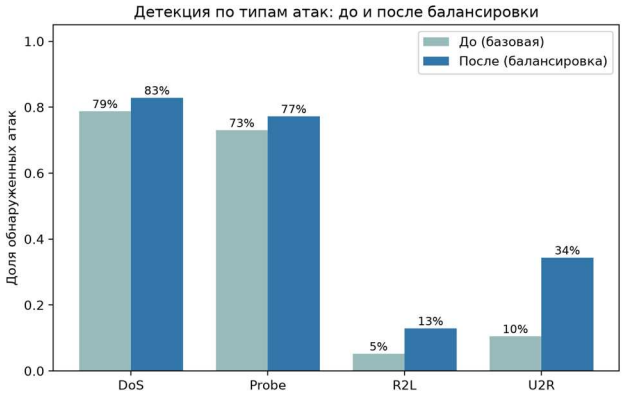

Рис. 2. Детекция по типам атак до и после балансировки (случайный лес)

Эффект заметный и, что особенно ценно, бесплатный по цене ложных тревог. Обнаружение U2R подскочило с 10 до 34%, R2L – с 5 до 13%, подтянулись и массовые классы. Точность почти не сдвинулась (0,967 против 0,969), а F1 вырос с 0,746 до 0,784. Балансировка добавила чувствительности к редким атакам, не завалив администратора ложными срабатываниями. Проблему это снимает не полностью, но даже простой SMOTE даёт ощутимый сдвиг, а специализированные методы способны пойти дальше.

Заключение

Сравнение пяти алгоритмов на наборе NSL-KDD показало, что машинное обучение обнаруживает сетевые атаки неравномерно. По F1 лучшим оказался градиентный бустинг (0,801), по ROC-AUC – случайный лес (0,960). Главное же касается не выбора модели, а её пределов: массовые DoS и Probe распознаются в 84% случаев, а редкие R2L и U2R — лишь в 5-15%. Этот перекос мы не только зафиксировали, но и частично устранили — балансировка методом SMOTE подняла обнаружение U2R с 10 до 34%, R2L – с 5 до 13% без потери точности. Научная ценность работы – в честной оценке

на раздельных выборках, количественной демонстрации перекоса детекции и измеримом его сокращении. Практический вывод: ML-модуль стоит встраивать в IDS как первый рубеж против массовых угроз, усиливая его балансировкой и специализированными методами под редкие классы.

Список литературы / References

1. Tavallaee M., Bagheri E., Lu W., Ghorbani A. A. Детальный анализ набора данных KDD CUP 99 // IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA). 2009, pp. 1-6. DOI: 10.1109/CISDA.2009.5356528.
2. Buczak A.L., Guven E. Обзор методов интеллектуального анализа данных и машинного обучения для обнаружения вторжений в кибербезопасности // IEEE Communications Surveys & Tutorials. 2016, vol. 18, no., pp. 1153-1176. DOI: 10.1109/COMST.2015.2494502.
3. Sommer R., Paxson V. За пределами замкнутого мира: о применении машинного обучения для обнаружения сетевых вторжений // IEEE Symposium on Security and Privacy. 2010, pp 305-316. DOI: 10.1109/SP.2010.25.
4. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E. Scikit-learn: машинное обучение на языке Python // Journal of Machine Learning Research. 2011, vol. 12, pp. 2825-2830.
5. Chawla N.V., Bowyer K. W., Hall L. O., Kegelmeyer W. P. SMOTE: метод синтетической генерации примеров миноритарного класса // Journal of Artificial Intelligence Research. 2002, vol. 16, pp. 321-357. DOI: 10.1613/jair.953.

Архипов Максим Олегович – студент	**Arkhipov Maksim Olegovich** – student
max.arkhipov.2004@gmail.com	

Received 25.06.2026