УДК 004.853

RLHF ИЛИ ЭФФЕКТИВНОЕ ИСПОЛЬЗОВАНИЕ ЭКСПЕРТНЫХ ОЦЕНОК

Муромцев Р.М.[1]
Научный руководитель – кандидат экономических наук, доцент Шелупанова П.А.[1]

1 – Томский государственный университет систем управления и радиоэлектроники

Введение. В последние годы мы наблюдаем значительное увеличение числа и сложности киберугроз. Эти угрозы становятся всё более изощрёнными, что делает их обнаружение сложной задачей. Злоумышленники постоянно разрабатывают новые методы атак, которые способны обходить традиционные системы безопасности. Это приводит к тому, что многие известные методы кибербезопасности, основанные на предварительно определенных правилах и подписях, оказываются неэффективными. Эти методы часто не могут своевременно обнаруживать новые, неизвестные ранее виды атак, особенно те, которые используют сложные стратегии для маскировки своего присутствия в сети.

В этом контексте развитие искусственного интеллекта (ИИ) представляет собой перспективное направление для усиления защитных механизмов в сфере кибербезопасности. ИИ способен анализировать большие объемы данных и выявлять сложные закономерности, что может помочь в распознавании и предотвращении кибератак. Одним из наиболее обещающих подходов в этой области является обучение с подкреплением от человека (Reinforcement Learning from Human Feedback, RLHF) [4, 7, 8]. Этот подход позволяет системам ИИ обучаться, используя не только структурированные данные, но и опыт и знания человеческих экспертов. В среде кибербезопасности, которая характеризуется быстрыми изменениями и появлением новых угроз, способность адаптироваться и обучаться на основе актуальной обратной связи от экспертов становится ключевым фактором в разработке эффективных защитных механизмов.

Основная часть. Система RLFH состоит из нескольких компонентов, включая модуль сбора данных, модуль обучения и модуль реагирования. Разберемся подробнее за что отвечает каждый отдельный модуль системы.

Модуль сбора данных отвечает за предварительную обработку и сбор данных, включая обратную связь от эксперта. Это необходимая часть для реализации RLHF подхода.

Модуль обучения состоит из environment и agent, где agent – модуль реагирования. Представим, что environment (S, A) это математическая модель нашей системы, в рамках который при определенном состоянии системы S agent выполняет действия A, направленные на получения максимального Reward (r). Также не стоит забывать о том, что экспертная оценка также играет немаловажную роль, а именно приоритет возможных действий составляется исходя из Human Feedback [9], то есть экспертной оценки. Таким образом, описанная реализация обучения представлена на рисунке 1.

Модуль реагирования использует anomaly transformer [1, 2, 3, 5, 6] для выявления аномалий во временных рядах событий, а также полносвязанную нейронную сеть для ранжирования возможных действий. Концепция системы представлена на рисунке 2.

Отличие anomaly transformer от остальных трансформеров заключается в слое anomaly-attention layer, который используется механизм внимания для выделения наиболее важных частей входных данных при обнаружении аномалий. Это особенно полезно для анализа аномалий в событиях системы, чтобы сосредоточиться на отклоняющихся или необычных паттернах в данных.

Также, этот подход использует онлайн-стратегию обучения, что позволяет сделать модель более устойчивой к выбросам ввиду вычисления оценки выравнивания (alignment scores), которая вычисляется на каждом шаге прогнозирования и дает оценку, которая указывает, насколько хорошо элементы входной последовательности соответствуют

текущему выходу на позиции.

В качестве эксперимента был проведен сравнительный анализ популярных моделей на 3 разных наборах данных, среди которых SMD (Server Machine Dataset), SML (этот набор данных собран с системы мониторинга, установленной в домашнем доме), а также SMAP (Soil Moisture Active Passive Data). В качестве рассматриваемых моделей были взяты LSTM, LSTM-VAE, THOC, InterFusion, OmniAnomaly. А в качестве метрики f1-score. Исходя из результатов можно сказать, что модель система с использованием RLHF показывает значительное превосходство над остальными. Результат сравнения представлен в таблице 1.

Выводы. В заключении, применение RLHF в кибербезопасности также подчеркивает важность междисциплинарного взаимодействия в современной науке и технологиях. Эффективное сочетание областей, таких как искусственный интеллект, машинное обучение и кибербезопасность, открывает новые горизонты для защиты цифровых систем и данных.

Тем не менее, важно признать и предстоящие вызовы, связанные с реализацией и внедрением таких систем на практике. К ним относятся вопросы конфиденциальности и безопасности данных, необходимость обеспечения непрерывного и качественного обучения моделей, а также постоянное обновление и поддержка системы для соответствия текущим угрозам.

Список использованных источников

1. Jiehui Xu , Haixu Wu, Jianmin Wang, Mingsheng Long “ANOMALY TRANSFORMER: TIME SERIES ANOMALY DETECTION WITH ASSOCIATION DISCREPANCY”// https://arxiv.org/pdf/2110.02642.pdf
2. Ahmed Abdulaal, Zhuanghua Liu, and Tomer Lancewicki //Practical approach to asynchronous multivariate time series anomaly detection and localization. KDD, 2021
3. Haibin Cheng, Pang-Ning Tan, Christopher Potter, and Steven A. Klooster. A robust graph-based algorithm for detection and characterization of anomalies in noisy multivariate time series //ICDM Workshops, 2008
4. Xingchen Wu, Qin Qiu, Jiaqi Li, and Yang Zhao "Intell-dragonfly: A Cybersecurity Attack Surface Generation Engine Based On Artificial Intelligence-generated Content Technology" //https://arxiv.org/pdf/2311.00240.pdf
5. Shohreh Deldari, Daniel V. Smith, Hao Xue, and Flora D. Salim. Time series change point detection with self-supervised contrastive predictive coding //In WWW, 2021
6. Zekai Chen, Dingshuo Chen, Zixuan Yuan, Xiuzhen Cheng, and Xiao Zhang. Learning graph structures with transformer for multivariate time series anomaly detection in iot //ArXiv, abs/2104.03466, 2021
7. Christian Wirth and Johannes Fürnkranz. Preference-based reinforcement learning: A preliminary survey //In ECML/PKDD Workshop on Reinforcement Learning from Generalized Feedback: Beyond Numeric Rewards, 2013
8. Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control //In International Conference on Intelligent Robots and Systems, pages 5026–5033, 2012
9. Patrikk D Sørensen, Jeppeh M Olsen, and Sebastian Risi. Breeding a diversity of super mario behaviors through interactive evolution //In Computational Intelligence and Games (CIG), 2016 IEEE Conference on, pages 1–7. IEEE, 2016