

УДК 004.89

РАЗРАБОТКА И ВАЛИДАЦИЯ ИНСТРУМЕНТОВ ДЛЯ СТРЕСС-ТЕСТИРОВАНИЯ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ

Медведев О.А., Старший оператор научной роты, III курс
Военная академия связи имени Маршала Советского Союза С. М. Буденного,
г. Санкт-Петербург

В последние годы машинное обучение стало неотъемлемой частью множества прикладных систем, используемых в самых различных областях – от медицины и финансов до промышленности и транспорта. Вместе с ростом масштабов и сложности моделей возникает необходимость не только создавать эффективные алгоритмы, но и обеспечивать их устойчивость к непредвиденным и экстремальным воздействиям, которые могут привести к сбоям или неправильным решениям. В этом контексте разработка и валидация инструментов для стресс-тестирования моделей машинного обучения приобретает особое значение как способ повышения надежности и безопасности ИИ-систем [1].

Стресс-тестирование представляет собой комплекс мероприятий, направленных на проверку поведения модели в условиях экстремальных, нетипичных или редких, но критичных ситуаций. Такие ситуации могут включать сильные шумы и искажения в данных, внезапные изменения распределения входных параметров, целенаправленные атаки, ограниченный доступ к информации, а также сбои в инфраструктуре и аппаратных средствах. Цель стресс-тестирования – выявить пределы работоспособности модели, определить сценарии, в которых она может проявлять нестабильность или неправильные предсказания, и на основе полученных данных улучшить архитектуру, алгоритмы и процессы обучения [2].

Процесс разработки инструментов для стресс-тестирования начинается с формализации требований и определения критериев, которым должны соответствовать создаваемые решения. Инструменты должны обеспечивать возможность генерации широкого спектра стрессовых сценариев с учётом специфики прикладной области и архитектуры модели. Это включает как случайные воздействия – например, шумы различной интенсивности и распределения, – так и продвинутое создание адверсариальных примеров, направленных на имитацию реальных угроз безопасности. Важно, чтобы инструменты были гибкими, масштабируемыми и совместимыми с основными ML-фреймворками и инфраструктурами.

Валидация инструментов стресс-тестирования является критическим этапом, гарантирующим их корректность и эффективность. Она предполагает проверку адекватности создаваемых воздействий, точности измерений и способности выявлять реальные уязвимости моделей.

Валидация проводится с использованием разнообразных эталонных наборов данных и моделей, отражающих широкий спектр задач и условий эксплуатации.

Кроме того, важным критерием является воспроизводимость тестов и стабильность результатов, что обеспечивает доверие к инструментам и возможность их использования в промышленной практике.

Техническая реализация инструментов часто базируется на современных программных технологиях, включая контейнеризацию, автоматизацию с использованием CI/CD-пайплайнов и интеграцию с MLOps процессами. Это позволяет обеспечить непрерывный мониторинг устойчивости моделей на всех этапах жизненного цикла, своевременное выявление деградации и автоматическую адаптацию стратегий защиты. Возможность масштабирования и параллельного выполнения тестов делает инструменты применимыми для моделей с большими объёмами данных и сложной архитектурой.

Особое внимание уделяется удобству использования и гибкости настройки инструментов. Понятные интерфейсы, возможность кастомизации сценариев стресс-тестирования, автоматический сбор и визуализация результатов позволяют разработчикам и аналитикам быстро интерпретировать данные и принимать обоснованные решения. Это способствует широкой адаптации технологий стресс-тестирования в различных индустриях и уменьшает порог входа для организаций с разным уровнем экспертизы.

Разработка и валидация таких инструментов стимулирует дальнейшие исследования в области повышения робастности моделей, позволяя выявлять новые типы уязвимостей и совершенствовать методы защиты. Накопленные данные и опыт применяются для создания более надёжных архитектур, методов аугментации данных, а также для улучшения процессов обучения и валидации моделей. Кроме того, внедрение стресс-тестирования способствует формированию культуры качества и ответственности в разработке ИИ-систем [3].

Нельзя не отметить и важность соответствия инструментов нормативным требованиям и стандартам безопасности, особенно в областях с высокими рисками, таких как медицина, авиация и финансовый сектор. Инструменты стресс-тестирования становятся неотъемлемой частью процедур сертификации и аудита, обеспечивая объективную и воспроизводимую оценку надёжности моделей. Это существенно повышает уровень доверия пользователей и регуляторов к интеллектуальным системам [4].

В результате разработка и валидация инструментов для стресс-тестирования моделей машинного обучения представляют собой сложный и многоступенчатый процесс, требующий объединения усилий исследователей, инженеров и практиков. Комплексный подход, включающий формализацию требований, техническую реализацию, проверку качества и интеграцию с современными процессами разработки, позволяет создавать эффективные решения, способные обеспечивать высокую устойчивость и безопасность ИИ-продуктов в реальных условиях эксплуатации. Продолжение работы в этом направлении будет способствовать развитию доверенного искусственного интеллекта и расширению сфер его успешного применения [5, 6].

Список литературы:

1. Mern, J. T. A Survey of Software Engineering for Machine Learning / J. T. Mern, A. B. Arrieta, X. Xie // ACM Computing Surveys. – 2023. – Vol. 55, № 11. – Article 235.
2. Zhang, X. A Survey of Deep Learning Testing / X. Zhang, H. Zhang, H. Sun [et al.] // ACM Computing Surveys. – 2022. – Vol. 55, № 13s. – Article 277.
3. Breck, E. The ML Test Score : A Rubric for ML Production Readiness and Technical Debt Reduction / E. Breck, S. Cai, E. Nielsen [et al.] // 2017 IEEE International Conference on Big Data. – Boston : IEEE, 2017. – P. 1123-1132. – DOI 10.1109/BigData.2017.8258038.
4. Renggli, C. Data Quality Validation for Machine Learning / C. Renggli, B. Karlaš, B. Ding [et al.] // Proceedings of Machine Learning and Systems. – 2021. – Vol. 3. – P. 802-815.
5. Uesato, J. Adversarial Risk and the Dangers of Evaluating against Weak Attacks / J. Uesato, B. O'Donoghue, A. van den Oord, P. Kohli // Proceedings of the 35th International Conference on Machine Learning. – 2018. – Vol. 80. – P. 5025-5034.
6. Vogelsang, A. Requirements Engineering for Machine Learning : Perspectives from Data Scientists / A. Vogelsang, M. Borg // 2019 IEEE 27th International Requirements Engineering Conference Workshops. – Jeju : IEEE, 2019. – P. 245-251. – DOI 10.1109/REW.2019.00050.