

УДК 004.89

АНАЛИЗ МЕТОДОВ УСИЛЕНИЯ УСТОЙЧИВОСТИ МОДЕЛЕЙ ЧЕРЕЗ АДАПТИВНОЕ ТЕСТИРОВАНИЕ

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

В современных условиях бурного развития искусственного интеллекта и машинного обучения перед исследователями и практиками стоит задача не только создания точных и эффективных моделей, но и обеспечения их устойчивости к разнообразным воздействиям и внешним факторам. Особенно остро эта проблема проявляется в прикладных областях, где модели подвергаются как случайным шумам, так и целенаправленным атакам, способным существенно снизить качество и надёжность работы систем. В этом контексте анализ методов усиления устойчивости моделей через адаптивное тестирование становится одним из ключевых направлений, направленных на выявление и устранение уязвимостей, повышение робастности и обеспечение долгосрочной стабильности функционирования [1].

Адаптивное тестирование представляет собой динамический и интерактивный процесс, в ходе которого сценарии проверки модели изменяются и уточняются на основе получаемых результатов и выявленных особенностей поведения. В отличие от традиционных методов, предполагающих заранее фиксированные и статичные тестовые наборы и параметры, адаптивный подход позволяет более эффективно исследовать пространство возможных искажений и атак, фокусируясь на наиболее критичных и значимых для конкретной модели аспектах. Такая стратегия обеспечивает более глубокое понимание уязвимостей и способствует развитию целевых методов защиты [2].

Основной механизм адаптивного тестирования заключается в циклическом взаимодействии между генератором тестовых примеров и анализатором результатов. После каждого этапа проверки система анализирует реакцию модели, выявляет зоны повышенной чувствительности или аномалии в поведении и формирует новые тестовые случаи, ориентированные на проверку и усиление обнаруженных слабых мест. Такой итеративный процесс продолжается до достижения заданных критериев устойчивости или исчерпания ресурсов тестирования. Важно, что адаптивный подход позволяет снизить объём необходимых тестов, концентрируя усилия на наиболее информативных и потенциально опасных сценариях.

В рамках адаптивного тестирования применяются разнообразные методы генерации тестов: от классических преобразований и шумов до сложных адверсариальных атак, которые модифицируются в соответствии с реакцией модели. Использование машинного обучения и оптимизационных алгоритмов для создания новых примеров позволяет обнаруживать ранее неизвестные уязвимости и повышать эффективность проверки. Кроме того, интеграция методов активного обучения и самообучения способствует автоматизации и интеллектуализации процесса тестирования, снижая необходимость участия человека и повышая масштабируемость.

Анализ результатов адаптивного тестирования требует применения комплексных метрик, способных учитывать не только качество предсказаний, но и изменения внутренней структуры модели, распределения выходных значений, а также динамику реакции на последовательные воздействия. Это позволяет формировать целостное представление об устойчивости и выявлять закономерности, которые сложно заметить при использовании статичных тестов. Визуализация и интерпретация таких данных являются важным компонентом, помогающим разработчикам быстро понимать проблемы и принимать обоснованные решения [3].

Практическая реализация методов адаптивного тестирования тесно связана с использованием современных вычислительных инфраструктур и программных платформ, поддерживающих автоматизацию и параллельное выполнение тестов. Внедрение таких систем в MLOps пайплайны обеспечивает непрерывный мониторинг устойчивости моделей на протяжении всего их жизненного цикла, позволяя своевременно обнаруживать деградацию и оперативно реагировать на новые угрозы. Это особенно важно для задач с высокими требованиями к безопасности и надёжности, где сбои могут привести к серьёзным последствиям [4].

Важным аспектом является также возможность кастомизации адаптивного тестирования под особенности конкретных прикладных областей, учитывая специфику данных, требований к точности, допустимым рискам и нормативным ограничениям. Такой подход обеспечивает баланс между глубиной проверки и затратами ресурсов, позволяя создавать эффективные и экономичные решения. Кроме того, адаптивное тестирование способствует развитию новых архитектур и алгоритмов, которые изначально проектируются с учётом требований к устойчивости и способности к саморегуляции [5].

Не менее значимой становится роль анализа эффективности методов усиления устойчивости, применяемых на основе результатов тестирования. Это включает оценку влияния различных техник регуляризации, методов аугментации, защитных слоёв и алгоритмов детекции атак на поведение модели. Полученные данные позволяют не только улучшать существующие модели,

но и создавать новые подходы, оптимизированные под конкретные угрозы и условия эксплуатации.

В конечном итоге, анализ методов усиления устойчивости моделей через адаптивное тестирование представляет собой динамичное и многоуровневое направление, сочетающее теоретические исследования и практические решения. Оно способствует формированию более надёжных, безопасных и доверенных систем машинного обучения, способных эффективно функционировать в условиях реального мира, где разнообразие и сложность воздействий постоянно растут. Развитие и внедрение таких подходов является залогом устойчивого прогресса в области искусственного интеллекта и повышения его вклада в общество и экономику.

Список литературы:

1. Segura, S. A Survey on Metamorphic Testing / S. Segura, G. Fraser, A. B. Sanchez, A. Ruiz-Cortes // IEEE Transactions on Software Engineering. – 2016. – Vol. 42, № 9. – P. 805-824. – DOI 10.1109/TSE.2016.2532875.
2. Chen, T. Y. Metamorphic Testing : A Review of Challenges and Opportunities / T. Y. Chen, S. C. Cheung, S. M. Yiu // ACM Computing Surveys. – 2018. – Vol. 51, № 1. – Article 4. – DOI 10.1145/3143561.
3. Harel-Canada, F. Is Neuron Coverage a Meaningful Measure for Testing Deep Neural Networks? / F. Harel-Canada, L. Wang, M. A. Gulzar, Q. Gu // Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. – New York : ACM, 2020. – P. 851-862. – DOI 10.1145/3368089.3409754.
4. Guo, J. DLFuzz : Differential Fuzzing Testing of Deep Learning Systems / J. Guo, Y. Jiang, Y. Zhao [et al.] // Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. – New York : ACM, 2018. – P. 739-743. – DOI 10.1145/3236024.3264835.
5. Wang, Z. DeepGlobal : A Framework for Global Robustness Evaluation of Deep Neural Networks / Z. Wang, M. Yan, J. Chen [et al.] // Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering. – New York : ACM, 2020. – P. 588-600. – DOI 10.1145/3324884.3416540.