

УДК 004.89

ИСПОЛЬЗОВАНИЕ СИМУЛЯЦИЙ ДЛЯ ВЫЯВЛЕНИЯ ЧУВСТВИТЕЛЬНОСТИ МОДЕЛЕЙ К ПРИЗНАКАМ РИСКА

Коваленко М.Г., студент гр. 024, I курс
Уральский федеральный университет имени первого Президента России
Б. Н. Ельцина, г. Екатеринбург

Использование симуляций для выявления чувствительности моделей к признакам риска представляет собой важный инструмент в области машинного обучения, особенно когда речь идет о создании и оценке моделей, применяемых в критически важных и высокорисковых областях, таких как финансовые услуги, здравоохранение, страхование, правосудие и безопасность. Модели машинного обучения, которые используются для принятия решений, влияющих на жизнь и благосостояние людей, должны обладать высокой степенью точности, справедливости и устойчивости к различным внешним воздействиям. Однако, несмотря на огромный потенциал и возможности, которые открывает использование искусственного интеллекта, существует опасность, что модели могут проявлять чрезмерную чувствительность к определенным признакам, особенно к таким, которые являются признаками риска, и, как следствие, принимать решения, искажающие справедливость или не соответствующие этическим стандартам [1].

Признаки риска в контексте машинного обучения могут быть связаны с различными характеристиками данных, такими как возраст, доход, профессия, состояние здоровья, местоположение или даже история кредитования. Например, в системах кредитного скоринга модели могут ошибочно отдать предпочтение или, наоборот, отклонить заявку на кредит, если модель слишком сильно зависит от одного-единственного признака, такого как возраст или место жительства, при этом игнорируя другие важные факторы. В таких случаях важно выявить, какие признаки, и в какой степени, влияют на решения модели, чтобы минимизировать риск ошибок и несправедливости. Одним из наиболее эффективных способов для этого является использование симуляций, которые позволяют воссоздавать возможные сценарии поведения модели в изменяющихся условиях.

Симуляции, в контексте машинного обучения, представляют собой метод, при котором моделируются различные варианты входных данных с целью тестирования чувствительности модели к изменениям в этих данных. Этот подход позволяет исследовать, как модель реагирует на изменения в значениях

признаков, которые могут быть связаны с рисками. Например, в финансовых системах можно провести симуляцию, где изменяются уровни доходов или кредитных историй клиентов, чтобы увидеть, как эти изменения влияют на решение модели о выдаче кредита. Такой подход позволяет точно определить, какие признаки (например, возраст, пол, семейное положение) наиболее влияют на решение модели и насколько эти признаки могут быть чувствительными в контексте предсказаний [2].

Одним из примеров применения симуляций является использование методов Монте-Карло, которые позволяют проводить множество случайных проб, имитируя различные вариации входных данных. Этот метод позволяет понять, как изменения в значениях признаков риска могут повлиять на итоговое решение модели, и поможет выявить потенциальные уязвимости или неожиданные зависимости. Например, если модель предсказывает вероятность заболевания на основе медицинских данных, с помощью симуляции можно изменить значения таких признаков, как возраст, пол, привычки пациента и другие, чтобы понять, насколько сильно эти изменения влияют на результаты предсказаний. В результате можно выявить, что модель может быть чрезмерно чувствительна к одному из признаков, что может привести к ошибочным и несправедливым прогнозам.

Также в симуляциях может быть использован метод вариации признаков, при котором происходит пошаговое изменение значений отдельных признаков для проверки реакции модели. Например, для системы, которая принимает решения по страхованию, можно варьировать такие признаки, как возраст, тип автомобиля или место проживания, чтобы понять, насколько изменения в этих характеристиках могут повлиять на итоговую оценку рисков и принятие решения. Такой подход позволяет анализировать степень чувствительности модели к рисковым признакам и выявить, в какой степени определенные переменные могут искажать результаты модели [3].

Симуляции также могут быть полезны для оценки устойчивости модели к внешним изменениям или колебаниям в данных. В условиях, когда данные могут быстро изменяться, например, из-за экономических или социальных факторов, важно понять, насколько устойчивой будет модель к этим изменениям. С помощью симуляций можно воссоздать различные сценарии изменений данных, такие как изменения в экономической ситуации, кризисы, изменения в законодательстве или политике, чтобы проверить, насколько гибко и корректно модель будет реагировать на новые, потенциально рискованные ситуации. Важно отметить, что эти тесты позволяют не только выявить уязвимости модели, но и подсказывают, какие методы и подходы могут быть использованы для корректировки модели в случае изменений в данных или среде.

Важной частью процесса симуляции является анализ чувствительности, который помогает выявить признаки, оказывающие наибольшее влияние на поведение модели. Такой анализ позволяет понять, какие данные являются ключевыми для принятия решения и какие из них могут вызывать чрезмерную чувствительность модели. Например, можно определить, что если модель чувствительна к изменениям в уровне дохода при определении кредитоспособности, это может привести к несправедливым результатам для людей с низким доходом, даже если они могут быть надёжными заемщиками. В этом случае можно рассмотреть возможность регуляризации модели или перераспределения весов для уменьшения влияния таких признаков.

Кроме того, симуляции дают возможность разрабатывать и тестировать сценарии стресс-тестирования, когда модель испытывается на экстремальных данных, которые не встречаются в реальной практике, но могут возникнуть в случае кризисов или неожиданных событий. Например, в случае резкого экономического спада симуляции могут помочь предсказать, как модель будет работать при изменении уровня безработицы или снижения потребительского спроса. Это может быть особенно важным для моделей, которые работают в нестабильных или высокорисковых сферах, таких как финансовые учреждения, страховые компании и медицинские системы, где принятие неправильных решений может повлечь серьёзные последствия для клиентов и для бизнеса [4].

В заключение, использование симуляций для выявления чувствительности моделей к признакам риска является мощным инструментом для обеспечения справедливости, устойчивости и этичности машинного обучения. Этот подход позволяет глубже понять, как изменения в признаках могут влиять на результаты предсказаний модели, а также выявить скрытые риски, которые могут быть потенциально опасными для пользователей.

Регулярное использование симуляций помогает моделям оставаться гибкими, адаптироваться к изменениям данных и улучшать их работу в условиях реальных рисков, обеспечивая при этом справедливость и безопасность принятия решений в различных отраслях [5].

Список литературы:

1. Saltelli, A. Global Sensitivity Analysis: The Primer / A. Saltelli, M. Ratto, T. Andres [et al.]. – Chichester : Wiley, 2008. – 292 p. – DOI 10.1002/9780470725184.
2. Morris, M. D. Factorial Sampling Plans for Preliminary Computational Experiments / M. D. Morris // Technometrics. – 1991. – Vol. 33, № 2. – P. 161–174. – DOI 10.1080/00401706.1991.10484804.
3. Sobol, I. M. Global Sensitivity Indices for Nonlinear Mathematical Models and Their Monte Carlo Estimates / I. M. Sobol // Mathematics and Computers in

Simulation. – 2001. – Vol. 55, № 1–3. – P. 271–280. – DOI 10.1016/S0378-4754(00)00270-6.

4. Borgonovo, E. A New Uncertainty Importance Measure / E. Borgonovo // Reliability Engineering & System Safety. – 2007. – Vol. 92, № 6. – P. 771–784. – DOI 10.1016/j.ress.2006.04.015.

5. Пылов, П. А. Экстракция признаков в моделях последовательного глубокого обучения / П. А. Пылов, А. В. Протодяконов // Россия молодая : сб. материалов XIV Всероссийской научно-практической конференции с международным участием, Кемерово, 19–21 апреля 2022 года. – Кемерово : Кузбасский государственный технический университет имени Т. Ф. Горбачева, 2022. – С. 31525.1–31525.3. – EDN MZEHGM.