

УДК 004.896

СОЗДАНИЕ НАБОРОВ ДАННЫХ ДЛЯ МАШИННОГО ОБУЧЕНИЯ В ОБЛАСТИ ПРОЕКТИРОВАНИЯ АППАРАТУРЫ

Морозов О.В. (Университет ИТМО)

Научный руководитель – к.т.н., доцент факультета программной инженерии и компьютерной техники Антонов А.А.
(Университет ИТМО)

Введение. Использование методов машинного обучения в средствах автоматизированного проектирования становится шире с каждым годом. Регулярно разрабатываются новые модели в области проектирования аппаратуры, применяемые на разных этапах проектирования от высокоуровневого синтеза до компоновки интегральных схем. В список применений машинного обучения входят: ускорение построения топологии на кристалле, синтез аппаратуры из HDL-спецификаций, валидация и верификация полученных результатов, проверка соответствия преобразуемых схем, поиск уязвимостей, анализ проектного пространства. Однако вопросы аугментации и предобработки датасетов остаются слабо систематизированными. Так, в исследованиях редко можно увидеть методику подбора данных, а количество типов генерируемых для выборки вычислителей редко превышает 1-3.

Основная часть. Задача формирования датасетов для построения моделей машинного обучения подразумевает сбор банка примеров, в котором есть заведомо известные недостатки микроархитектуры и способы их устранения. Таким образом, первым шагом для решения задачи является выделение подобных признаков в существующих примерах, либо же создание правила, по которому будут генерироваться примеры, заранее содержащие эти признаки.

Исследуемые методы получения примеров включает в себя: ручное проектирование, проектирование с параметризацией, проектирование при помощи высокоуровневого синтеза, проектирование средствами искусственного интеллекта, аугментация существующих датасетов.

Отдельной проблемой в этом процессе является верификация полученных спецификаций, поэтому важным этапом становится автоматизация тестирования. В отдельных случаях, целесообразно применение методов повышения скорости синтеза спецификаций. Для этого могут задействоваться такие техники, как скриптовые макросы, предварительная отбраковка, предобработка списков соединений.

Оцениваемые характеристики методов включают в себя быстродействие, точность полученных моделей, разнообразие получаемых примеров и влияние на переобучение модели машинного обучения, возможные параметры модели, трудозатратность применения.

Выводы. В ходе работы были исследованы различные методы формирования датасетов. При помощи описанных методов сгенерированы выборки и протестированы для модели линейной регрессии с целью оценки характеристик полученных выборок. В дальнейшем результаты данной работы будут использованы для разработки методов исследования проектного пространства в задачах архитектурного проектирования аппаратуры.

Список использованных источников:

1. Huang G. et al. Machine learning for electronic design automation: A survey //ACM Transactions on Design Automation of Electronic Systems (TODAES). – 2021. – Т. 26. – №. 5. – С. 1-46.
2. Rosa J. P. S. et al. Using artificial neural networks for analog integrated circuit design automation. – Berlin : Springer, 2020. – Т. 1.
3. Thakur S. et al. Benchmarking Large Language Models for Automated Verilog RTL

Code Generation //arXiv preprint arXiv:2212.11140. – 2022.

4. Zennaro E. et al. A machine learning approach for area prediction of hardware designs from abstract specifications //2018 21st Euromicro Conference on Digital System Design (DSD). – IEEE, 2018. – C. 413-420.

5. Varambally B. S., Sehgal N. Optimising Design Verification Using Machine Learning: An Open Source Solution //arXiv preprint arXiv:2012.02453. – 2020.

6. Boser B. E., Guyon I. M., Vapnik V. N. A training algorithm for optimal margin classifiers //Proceedings of the fifth annual workshop on Computational learning theory. – 1992. – C. 144-152.