

УДК 004.457

ИСПОЛЬЗОВАНИЕ РСЕМО-МОДЕЛЕЙ В ЗАДАЧАХ МОДЕЛИРОВАНИЯ АППАРАТНЫХ НЕЙРОННЫХ СЕТЕЙ

Кормилицын К.А. (Университет ИТМО), Кормилицына Е.К. (Университет ИТМО)

Научный руководитель – к.т.н, доцент Кустарев П.В.
(Университет ИТМО)

С каждым годом все большую популярность набирают алгоритмы обработки информации на основе нейронных сетей. Однако существуют ограничения, которые мешают использовать данные алгоритмы для управления техническими системами и объектами. Одно из ограничений - отсутствие математического аппарата, который позволил бы формально оценить выполнение ограничений реального времени для искусственных нейронных сетей. В рамках данной проблемы в статье решается частная задача - предложен и апробирован метод валидации временных характеристик аппаратных искусственных нейронных сетей, реализованных на FPGA. Метод базируется на аппарате сетей массового обслуживания, имеет прикладную направленность и прошел экспериментальную апробацию.

Введение. С каждым годом все большую популярность начинают набирать приложения на основе искусственных нейронных сетей (ИНС). В том числе их пробуют использовать для управления техническими и киберфизическими системами. Для таких встроенных систем управления важны высокие производительность и энергоэффективность, которые достигаются аппаратной реализацией вычислителей. В последние несколько лет основными платформами для аппаратных нейронных сетей являются FPGA и графические процессоры. Критичным условием применения нейросетей во встроенных системах является предсказуемость их поведения как в функциональном, так и во временном аспектах. И первое, и второе – далеко нерешенная задача. Фокус статьи на второй задаче. На сегодняшний день нет устоявшихся методологий подтверждения выполнения требований реального времени. Автор исследует возможности моделирования аппаратных нейросетей на базе FPGA с помощью сетей массового обслуживания (СеМО). В статье показано, что различные структуры аппаратных нейросетей хорошо отображаются на модель СеМО и анализ функционально-динамических параметров при помощи математического аппарата и средств симуляции СеМО позволит оптимизировать архитектуру аппаратных нейросетей для достижения ограничений реального времени.

Основная часть. Искусственный нейрон, состоит из многовходового сумматора, блока функции активации и набора коэффициентов (синаптические веса, смещение, параметры функции активации), хранимых в памяти. Память используется также для аппаратной табличной реализации функций активации. Из современных технологий только FPGA может обеспечить множество независимых блоков памяти со сверхбыстрым доступом. Это причина популярности FPGA для реализации ИНС и одновременно причина того, что память будет «узким» местом при разработке ИНС на базе FPGA. Структуру нейронной сети можно прямо отобразить на структуру разомкнутой сети массового обслуживания (РСеМО). Каждый нейрон – узел РСеМО - можно представить в виде обслуживающего прибора, без очереди. Время обслуживания в нейроне определяется быстродействием FPGA. Несмотря на видимую «жесткую» синхронизацию цифровой схемы, время обслуживания не фиксировано и имеет Гауссово распределение. Это объясняется джиттером синхросигнала, накапливающимся при последовательном обходе слоев ИНС. Входной поток данных сети (заявок), случайный, распределенный по равномерному закону. «Узким местом» в FPGA-реализациях ИНС является встроенная память. Способом экономии памяти является использование одного блока функции активации несколькими нейронами, но это приводит к коллизиям и задержкам из-за параллельных обращений нейронов к блокам памяти и ставит задачу оптимизации топологии нейронной сети по критерию быстродействия. С помощью РСеМО-модели, можно

проводить симуляция и рассчитывать характеристики для FPGA-ИНС с несколькими способами объединения нейронов. Таким образом с помощью РСeМО-модели можно выбрать оптимальную модель с точки зрения производительности, загруженность элементов памяти и отказоустойчивость системы, как следствие, повышая надежность аппаратных реализаций нейросетей

Выводы. В результате исследования был сделан вывод, что использование модели РСeМО для моделирования аппаратных нейронных сетей реализованных на FPGA является целесообразным и с помощью данного математического аппарата можно оценить временные характеристики аппаратных нейронных сетей. Модель РСeМО позволяет имитировать систему с различными компоновками нейронов и дисциплинам их обслуживания. Таким образом существует возможность выбрать оптимальную компоновку сети с точки зрения производительности и вероятностей отказа системы. Модели РСeМО отлично подходят для описания как параллельных, так и последовательных процессов, проходящих в аппаратной реализации нейронных сетей.