

Реализация нейросетевых алгоритмов на вычислительной платформе с архитектурой SIMD

А. А. Емельянов, С. А. Золотов, Е. И. Цыкунов

1. Введение

Целью представляемой работы является интеграция бурно развивающихся нейросетевых алгоритмов с современными параллельными вычислителями. Актуальность данной темы обусловлена необходимостью иметь высокоуровневый инструмент разработки, моделирования и внедрения высокопроизводительных нейронных сетей. Потребность в таком инструменте возникает как на этапе изучения и разработки нейросетевых алгоритмов, так и на этапе их тестирования и внедрения.

2. Формулировка и анализ задачи

Инженерам ООО «Информационный Вычислительный Центр» была поставлена задача разработать комплекс программных средств, обеспечивающих высокопроизводительную обработку потоков данных (в том числе видео изображений) с использованием нейросетевых алгоритмов. В результате изучения и анализа задачи было принято решение создать ПО промежуточного слоя реализующего поведение нейронных слоев разных типов и их взаимодействие на аппаратной платформе nVidia GeForce/CUDA.

3. Текущие результаты

На сегодняшний день реализована значительная часть базовых математических алгоритмов, адаптированных на целевую архитектуру, и использующихся в широком круге нейронных сетей различных типов. Тесты реализованных математических алгоритмов на младшей модели графического процессора nVidia GeForce 8600 показали увеличение производительности, по сравнению с Intel®Core™2 Quad CPU 2.5 ГГц в 90-110 раз.

На основе базовых алгоритмов создаются программные модели общеизвестных типовых нейросетей. Из них реализованы: SOM (самоорганизующаяся карта Кохонена), алгоритм нейронного газа, многослойные нейронные сети с нелинейной сжимающей функцией. В разработке находятся: нейронная сеть типа Когнитрон и её расширение Неокогнитрон, а также нейросеть на основе адаптивной резонансной теории (АРТ архитектура). Каждый тип нейросети сопровождается соответствующим алгоритмом её обучения.

Разрабатываемое ПО представляет собой бинарные динамические библиотеки с реализацией низкоуровневых базовых математических функций и библиотеку пользователя с высокоуровневым программным интерфейсом. Базовые математические функции разработаны и реализованы на языке С, библиотека пользователя – на языке С++. Библиотека пользователя представляет собой высокоуровневый Объектно-Ориентированный интерфейс пользователя, позволяющий создавать, настраивать, обучать и моделировать нейросети различных типов, а также без изменения кода внедрять разработанные нейронные структуры в целевые приложения.

4. Планы на ближайшее будущее

В первую очередь планируется реализовать элементарные нейросети (нейронная сеть типа Когнитрон и её расширение Неокогнитрон, сети Хемминга и Хопфилда), а также сложно организованные нейросети (нейросеть на основе адаптивной резонансной теории «АРТ архитектура» и сеть ДАП). Также продолжают работы по оптимизации базовых математических алгоритмов для увеличения скорости вычислений, и созданию новых более совершенных алгоритмов обучения и поведения как реализованных нейросетей, так и планируемых к реализации.