

Лекция 19: стек технических средств для реализации методов ML

Евгений Борисов

Нейросети

история

[1950] математические модели биологического нейрона

McCulloch W.S., Pitts W. A logical Calculus of Ideas Immanent in Nervous Activity
– Bull. Mathematical Biophysics, 1943

[1960] модели нейронных сетей с одним обрабатывающим слоем

F.Rosenblatt Principles of Neurodynamics. - New York: Spartan Books, 1962.

[1970-80] метод обратного распространения, нейронные сети с несколькими обрабатывающими слоями

Галушкин А. И. Синтез многослойных систем распознавания образов.
– М.: «Энергия», 1974.

D.E.Rumelhart, G.E.Hinton, R.J.Williams Learning internal representations
by error propagation. // In Parallel distributed processing,
vol. 1, pp. 318-62. Cambridg, MA: MIT Press, 1986.

[2005] концепция Deep Learning

Нейросети

концепция Deep Learning

многоуровневые модели интеллектуальных систем
(т.е. представленные несколькими слоями модели)

искусственные нейронные сети

- много данных

- через большое количество слоёв

- с большим количеством нейронов

автоматическое извлечение признаков

улучшенная обобщающая способность

проблема "исчезающего" градиента (vanishing gradient)

предобучение

Нейросети

общее описание стека технологий

прикладные программные средства

вычислительные библиотеки

программный интерфейс с аппаратурой

аппаратные вычислительные средства

Нейросети

аппаратная часть : про модельные данные

набор с изображениями цифр MNIST - 70 000 примеров - 53MB

набор картинок CIFAR-10 - 60 000 примеров - 163MB

набор изображений PASCAL-2011 - 11 530 - 2GB

набор изображений ILSVRC-2015: CLS-LOC - 100 000 - 155GB

необходимо использовать средства соответствующей производительности

- высокопроизводительные параллельные вычисления

Нейросети

аппаратная часть :
параллельные вычисления

классификация по двум параметрам

- способ организации памяти
- способ организации управления

Нейросети

аппаратная часть :

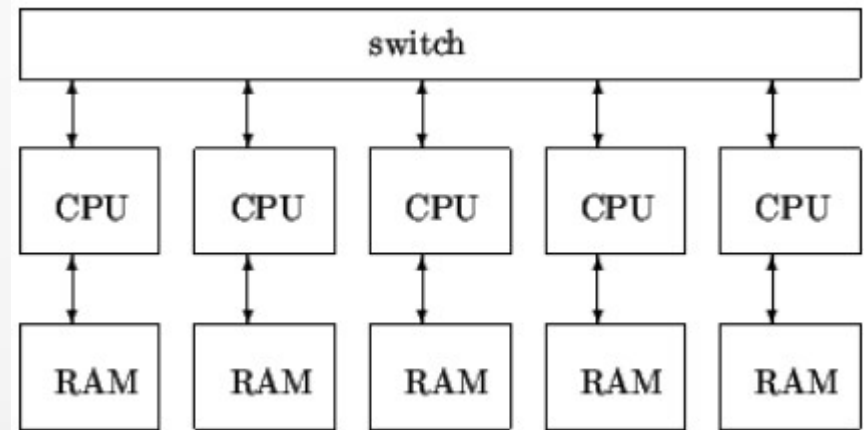
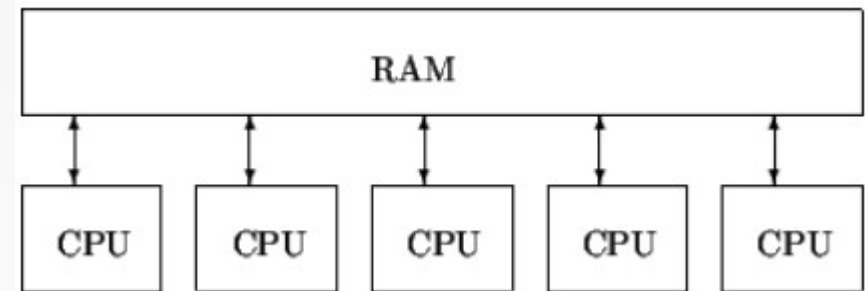
основные типы организации памяти параллельных вычислителей

общая память - все процессора работают в едином адресном пространстве с равноправным доступом к памяти

- + относительно просто программировать,
- ограниченная масштабируемость

распределенная память - каждый процессор имеет собственную локальную память, прямой доступ к этой памяти других процессоров невозможен.

- + хорошая масштабируемость
- сложно программировать
(больше вычислений, меньше обменов)



Нейросети

аппаратная часть :

способ организации управления параллельными вычислителями

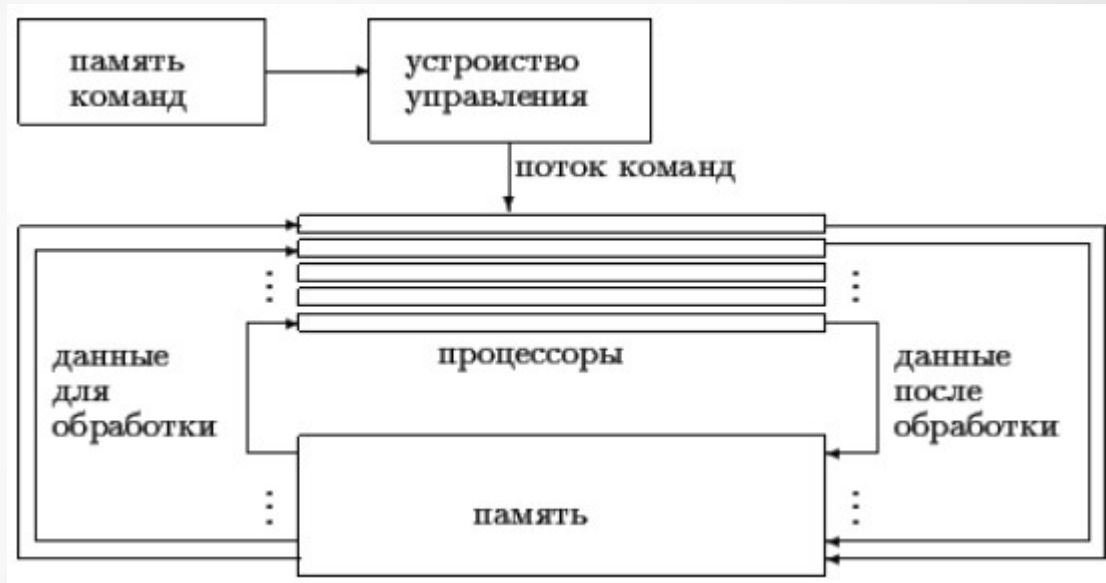
классификация Флинна

SIMD "single instruction multiple data" (один поток команд много потоков данных)

SISD

MISD

MIMD "multiple instruction multiple data" (много потоков команд много потоков данных)



Нейросети

аппаратная часть :

задачи при организации параллельных вычислений

автоматическое распараллеливание последовательных программ

автоматическая балансировка загрузки процессоров

Нейросети

аппаратная часть :

основные типы параллельных вычислителей

SMP - симметричные мультипроцессорные системы (общая память)

PVP - параллельные векторные системы (специальные векторно-конвейерные процессоры)

MPP - системы массового параллелизма. несколько однородных вычислительных узлов; связаны специальной высокоскоростной сетью. (распределенная память)

NUMA - системы с неоднородным доступом к памяти. память физически распределена, но логически общедоступна.

Cluster - «упрощённый» вариант MPP

Grid - «кластер из кластеров»

GPGPU - General-Purpose Graphics Processing Units

Нейросети

аппаратная часть

HPC технологии можно комбинировать

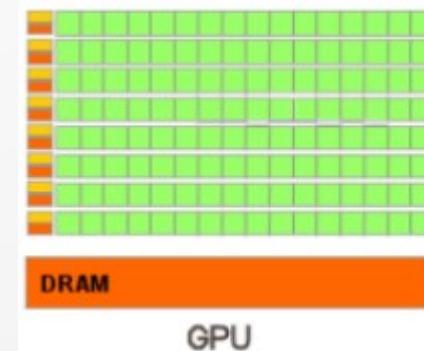
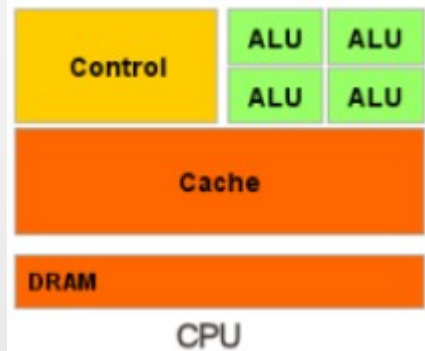
Пример: Грид состоящий из нескольких кластеров,
кластеры состоят из вычислительных узлов SMP
вычислительные узлы имеют GPU

Нейросети

аппаратная часть : подробней про GPGPU

параллельная программа GPGPU состоит из двух частей

1. основная - выполняется на CPU (host)
2. kernel - выполняется на GPU (device)



Нейросети

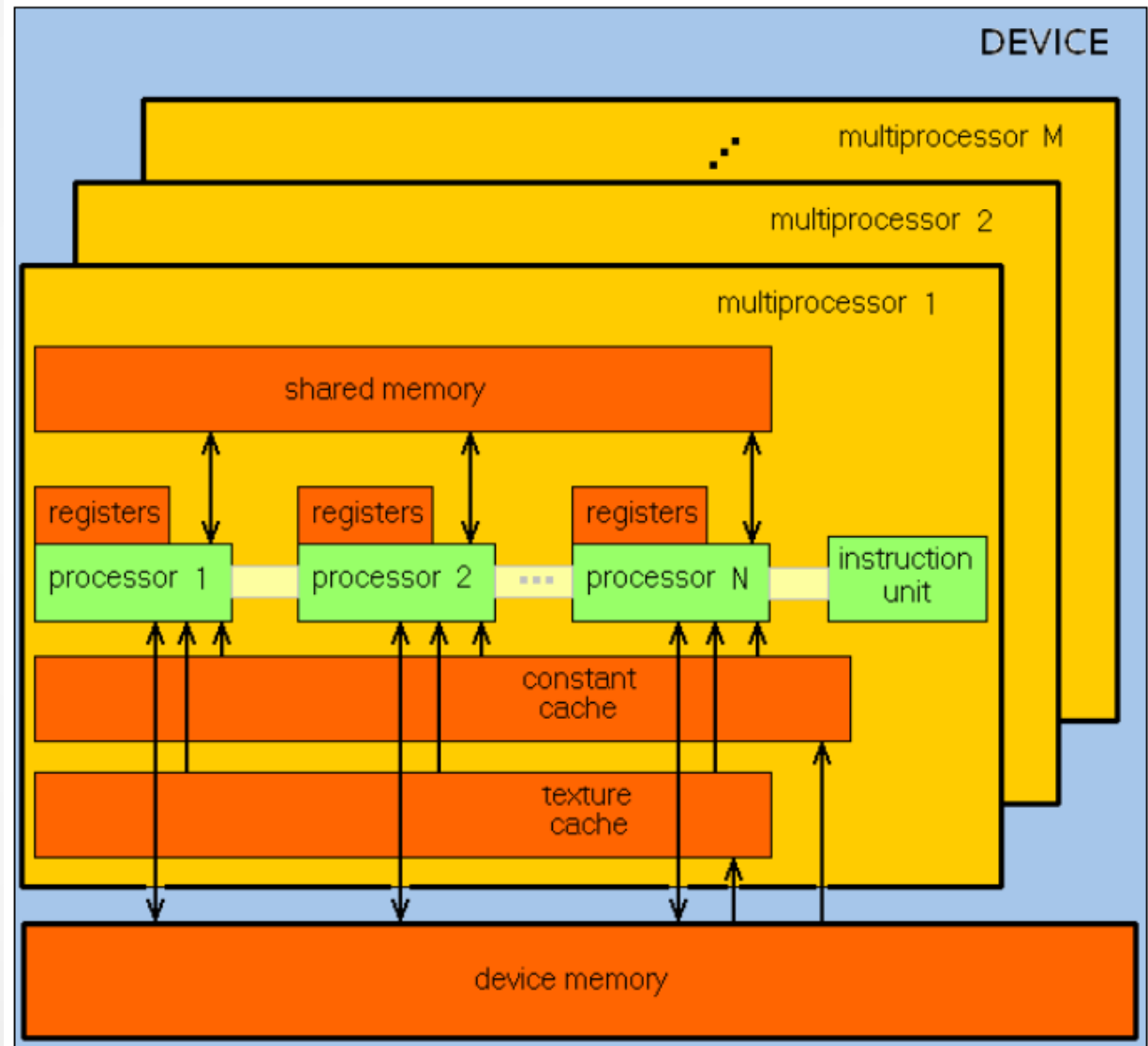
аппаратная часть : подробней про GPGPU

GPU - массив потоковых процессоров (Streaming Processor Array), состоит из Texture Processor Clusters (TPC)

TPC состоит из Streaming Multi-processor (SM)

SM содержит несколько Streaming Processors (SP) или ядер.

ядра мультипроцессора работают по схеме SIMD



Нейросети

программный интерфейс с аппаратурой

распределенная память: MPI, Spark

общая память: OpenMP, OpenCL

GPU: OpenCL, CUDA

прикладные программные средства

вычислительные библиотеки

программный интерфейс с аппаратурой

аппаратные вычислительные средства

Нейросети

программный интерфейс с аппаратурой :

вычислительные библиотеки для систем с распределенной памятью

MPI - стандарт параллельного программирования для распределённых систем

на каждом узле кластера запускаем вычислительный процесс

процессы могут обмениваться данными

есть механизм синхронизации процессов

Apache Spark - средство для обработки данных в распределённых хранилищах

Нейросети

программный интерфейс с аппаратурой :

вычислительные библиотеки для систем с общей памятью

OpenMP - стандарт параллельного программирования для SMP

OpenCL - стандарт параллельного программирования для SMP и GPU

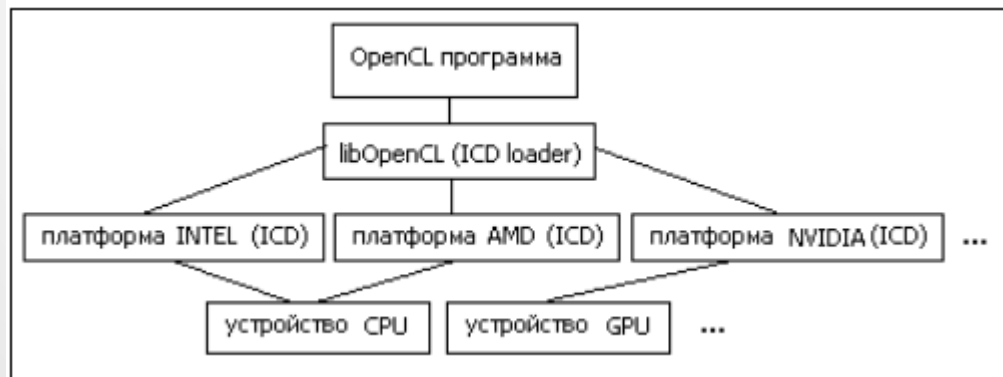
Нейросети

программный интерфейс с аппаратурой :

вычислительные библиотеки для GPGPU

CUDA - библиотека параллельного программирования для устройств nVidia
compute capability 7.5, cuBLAS, cuDNN

OpenCL - стандарт параллельного программирования для SMP и GPU



Нейросети

прикладные средства

Caffe и DIGITS - система моделирования свёрточных сетей

Theano - пакет ML от университета Монреаля

TensorFlow - пакет ML от Google

CNTK - пакет ML от Microsoft

Keras - библиотека-надстройка над Theano, TensorFlow, CNTK

Torch - пакет ML основанный на языке Lua

Нейросети

прикладные средства

Caffe - система моделирования свёрточных сетей

имеет интерфейсы для нескольких популярных языков

web интерфейс DIGITS

можно описывать модели в спецификациях на языке protobuf

поддерживает вычисления на GPU

Нейросети

прикладные средства

Theano - пакет ML от университета Монреаля

пакет языка Python предназначенный для символьных вычислений

численные расчёты работают непосредственно с данными в памяти

символьные вычисления (symbolic computation) работают с алгебраическими выражениями

может выполнять аналитическое дифференцирование и оптимизацию описанных пользователем функций

позволяет компилировать и выполнять заданные пользователем функций в машинный код для CPU и/или GPU

Нейросети

прикладные средства

TensorFlow - пакет ML от Google

пакет языка Python предназначенный для символьных вычислений

может выполнять аналитическое дифференцирование и оптимизацию описанных пользователем функций

позволяет компилировать и выполнять заданные пользователем функций в машинный код для CPU и/или GPU

содержит TensorBoard - инструмент для визуализации графа вычислений

Нейросети

прикладные средства

Keras - библиотека-надстройка над Theano, TensorFlow, CNTK

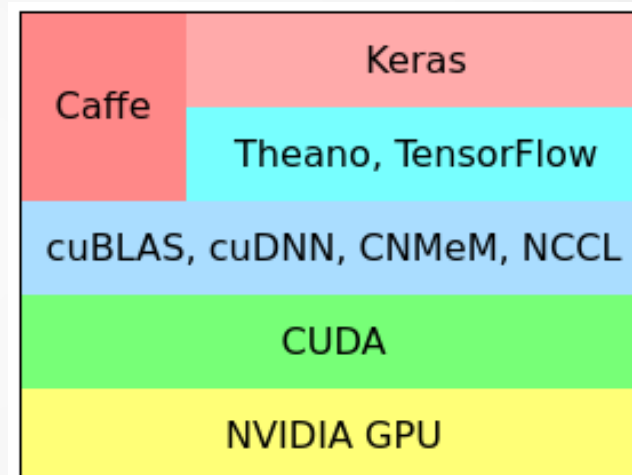
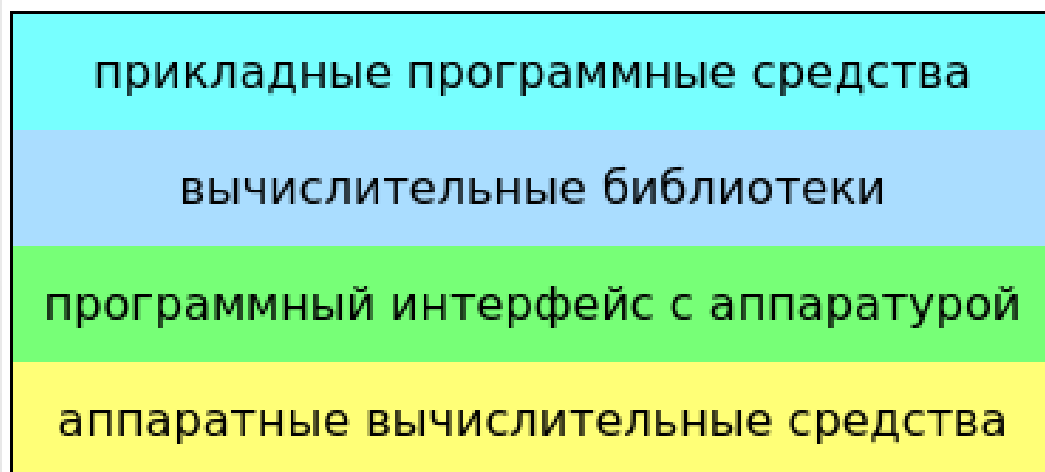
пакет языка Python

содержит примитивы для реализации моделей Deep Learning

использует Theano, TensorFlow или CNTK в качестве основы (backend)

Нейросети

описание стека технологий



Нейросети: литература

git clone https://github.com/mechanoid5/ml_lectorium.git

Евгений Борисов О технических средствах для реализации методов Deep Learning.

<http://mechanoid.kiev.ua/ml-deep-learning-tools.html>

Евгений Борисов Технология параллельного программирования CUDA.

<http://mechanoid.kiev.ua/parallel-cuda.html>

Нейросети



Вопросы ?

Нейросети: практика

источники данных для экспериментов



sklearn.datasets
UCI Repository
kaggle

