Machine Learning for Adaptive Multi-Core Machines

Noel Lopes
Supervisor: Prof. Dr. Bernardete Ribeiro

CISUC
University of Coimbra, Portugal

September 17, 2013
Outline

▶ Introduction
▶ Objectives
▶ Contributions
▶ High-performance Deep Learning
▶ Conclusions
Machine Learning
Need to Scale up

- Large datasets
- High-dimensional inputs
- Inference time constraints
- Algorithm complexity
- Adequate model selection
- Cascade predictors
Machine Learning
Big Data

Data sources
Real Data
Computer Simulation Models
Artificial Data

Data streams

Extract useful and relevant information

Large volumes of data

vastly exceeds our capacity to analyze it

Persistent repositories of (accumulated) Data

challenge

Noel Lopes
Machine Learning for Adaptive Multi-Core Machines
Machine Learning
Big Data

Data sources -> Real Data -> Computer Simulation Models -> Artificial Data

Data streams -> Large volumes of data

ML Algorithms

Extracted information

Persistent repositories of (accumulated) Data
Scientific Contributions

- Machine Learning
  - Supervised Learning
  - Semi-supervised Learning
  - Unsupervised Learning
- GPUMLib – GPU ML Library
Scientific Contributions
Supervised Learning

- Machine Learning
  - **Supervised Learning**
    - Autonomous Training System (ATS)
    - Neural Selective Input Model (NSIM)
    - Incremental Hypersphere Classifier (IHC)
  - Semi-supervised Learning
  - Unsupervised Learning
- GPUMLib – GPU ML Library

![Diagram](image-url)

Main Network with selective actuation neurons

Space Network
Scientific Contributions
Supervised Learning

- Machine Learning
  - **Supervised Learning**
    - Autonomous Training System (ATS)
    - Neural Selective Input Model (NSIM)
    - Incremental Hypersphere Classifier (IHC)
  - Semi-supervised Learning
  - Unsupervised Learning
- GPUMLib – GPU ML Library
Scientific Contributions
Supervised Learning

- Machine Learning
  - **Supervised Learning**
    - Autonomous Training System (ATS)
    - Neural Selective Input Model (NSIM)
    - **Incremental Hypersphere Classifier (IHC)**
  - Semi-supervised Learning
  - Unsupervised Learning
- GPUMLib – GPU ML Library
Scientific Contributions

Semi-Supervised Learning

▶ Machine Learning
  ▶ Supervised Learning
  ▶ Semi-supervised
    ▶ Semi-supervised Non-Negative Matrix Factorization
  ▶ Unsupervised Learning
▶ GPUMLib – GPU ML Library

\[
V^{1} \ V^{2} \ \cdots \ V^{C} \approx \ D^{N} \ r^{N} \ r^{1} \ r^{2} \ r^{C} \ H^{1} \ H^{2} \ \cdots \ H^{C}
\]
Scientific Contributions
Unsupervised Learning

- Machine Learning
  - Supervised Learning
  - Semi-supervised
  - **Unsupervised Learning**
    - Deep Belief Networks (Adaptive Step Size technique)
- GPUMLib – GPU ML Library
Scientific Contributions
Case studies and Benchmarks

- Case studies
  - biomedical
  - finance and business
  - bio-informatics

- Yale face database
- ORL face database
- MNIST hand-written digits
- HHreco multi-stroke images
Historical Single-/Double-Precision Peak Compute Rates

- **Precision**
  - **SP**
  - **DP**

- **Vendor**
  - **AMD (GPU)**
  - **NVIDIA (GPU)**
  - **Intel (CPU)**
  - **Intel Xeon Phi**

**GFLOPS**

**Date**

- 2002
- 2004
- 2006
- 2008
- 2010
- 2012
Scientific Contributions
GPUMLib – GPU ML Library

Host (CPU) and device (GPU) memory access framework
- HostArray
- HostMatrix
- CudaArray
- DeviceArray
- DeviceMatrix
- ...

C++ classes (algorithms)
- Back-Propagation
- Radial Basis Functions
- Deep Belief Networks
- Restricted Boltzmann Machines
- Multiple Back-Propagation
- Support Vector Machines
- Non-Negative Matrix Factorization
- ...

Common Host (CPU) Classes
Common CUDA Kernels
CUDA (GPU) Kernels
- Multiple Back-Propagation
- Support Vector Machines
- Non-Negative Matrix Factorization
- Nonlinear Dimension Reduction
- Radial Basis Functions
- Restricted Boltzmann Machines
- Self Organizing Maps
- ...

http://gpumlib.sourceforge.net/

Noel Lopes
Machine Learning for Adaptive Multi-Core Machines
Deep Belief Networks
Deep architecture

Object Recognition/
Invariant
Representation
("Dog")

Increasingly
Complex
Features

Simple Inputs

Feature Detection
(feedforward input)

Prediction
(feedback inference)
For the binary units \( h_j \in \{0, 1\} \) and \( v_i \in \{0, 1\} \) the energy function of the whole network is:

\[
E(v, h) = - \sum_{i,j} W_{ij} v_i h_j - \sum_i c_i v_i - \sum_j b_j h_j
\]  

(1)

where \( W \) is the matrix of weights, and \( b \) and \( c \) are the bias units w.r.t. hidden and visible layers, respectively.
Restricted Boltzmann Machines (RBMs)

Given a random training vector \( \mathbf{v} \), the state of a given hidden unit \( j \) is set to 1 with probability:

\[
p(h_j = 1|\mathbf{v}) = \sigma(b_j + \sum_i v_i W_{ij})
\]  

(2)

Similarly:

\[
p(v_i = 1|h) = \sigma(c_i + \sum_j h_j W_{ij})
\]  

(3)

where \( \sigma(x) \) is the sigmoid squashing function \( \frac{1}{1+e^{-x}} \).
Training an RBM

Alternating Gibbs Sampling

\[ p(h_j = 1 | v) = \sigma(b_j + \sum_{i=1}^{I} v_i W_{ji}) \]
Training an RBM

Alternating Gibbs Sampling

\[ p(v_i = 1|h) = \sigma(c_i + \sum_{j=1}^{J} h_j W_{ji}) \]
Training an RBM

Alternating Gibbs Sampling

\[ p(v_i = 1 | h) = \sigma(c_i + \sum_{j=1}^{J} h_j W_{ji}) \]
Training an RBM
Alternating Gibbs Sampling

\[ p(v_i = 1|h) = \sigma(c_i + \sum_{j=1}^{J} h_j W_{ji}) \]
Training an RBM
Alternating Gibbs Sampling

\[ v^{(0)} = x \]
\[ \langle v_i h_j \rangle_0 \]
Training an RBM
Contrastive Divergence (CD–$k$)

- To solve this problem, Hinton proposed the Contrastive Divergence algorithm.
- CD–$k$ replaces $\langle \cdot \rangle_\infty$ by $\langle \cdot \rangle_k$ for small values of $k$.

\[
\Delta W_{ji} = \gamma (\langle v_i h_j \rangle_0 - \langle v_i h_j \rangle_k)
\]  
(4)
Deep Belief Networks (DBNs)

Noel Lopes
Machine Learning for Adaptive Multi-Core Machines
Deep Belief Networks (DBNs)

GPU Implementation Results

\[ N = 60,000 \]

**MNIST average training time per epoch.**
Deep Belief Networks (DBNs)

Adaptive Step Size

Average reconstruction error (RMSE).
Deep Belief Networks (DBNs)

Demonstration
Conclusions
Future Work

- Big Data Problem:
  - Novel ML algorithms
  - Scale-up existing algorithms
    - High-performance (GPU) ML implementations

- Size matters:
  - Enhancing GPUMLib algorithms with Big Data in mind
Publications
First author

- **5** Journal Articles
- **15** Conference Articles
- **30+** Citations

Noel Lopes
Machine Learning for Adaptive Multi-Core Machines
Noel Lopes and Bernardete Ribeiro. 
Towards adaptive learning with improved convergence of deep belief networks on graphics processing units. 
*Pattern Recognition*, 2013.

Noel Lopes and Bernardete Ribeiro. 
Towards a hybrid NMF-based neural approach for face recognition on GPUs. 

Noel Lopes and Bernardete Ribeiro. 
Handling missing values via a neural selective input model. 

Noel Lopes and Bernardete Ribeiro. 
GPUMLib: An efficient open-source GPU machine learning library. 

Noel Lopes and Bernardete Ribeiro. 
An evaluation of multiple feed-forward networks on GPUs. 
Noel Lopes, Bernardete Ribeiro, and João Gonçalves.  
Restricted Boltzmann machines and deep belief networks on multi-core processors.  
In *The 2012 International Joint Conference on Neural Networks (IJCNN)*, 2012.

Noel Lopes and Bernardete Ribeiro.  
Improving convergence of restricted Boltzmann machines via a learning adaptive step size.  

Noel Lopes, Daniel Correia, Carlos Pereira, Bernardete Ribeiro, and António Dourado.  
An incremental hypersphere learning framework for protein membership prediction.  

Noel Lopes and Bernardete Ribeiro.  
A robust learning model for dealing with missing values in many-core architectures.  


Noel Lopes and Bernardete Ribeiro.
A strategy for dealing with missing values by using selective activation neurons in a multi-topology framework.
In *IEEE World Congress on Computational Intelligence (WCCI 2010)*, 2010.

Noel Lopes and Bernardete Ribeiro.
Stochastic GPU-based multithread implementation of multiple back-propagation.

Noel Lopes and Bernardete Ribeiro.
Non-negative matrix factorization implementation using graphic processing units.

Noel Lopes and Bernardete Ribeiro.
A hybrid face recognition approach using GPUMLib.
Noel Lopes and Bernardete Ribeiro.
Fast pattern classification of ventricular arrhythmias using graphics processing units.

Noel Lopes and Bernardete Ribeiro.
GPU implementation of the multiple back-propagation algorithm.

Noel Lopes and Bernardete Ribeiro.
MBPGPU: A supervised pattern classifier for graphical processing units.
Machine Learning for Adaptive Multi-Core Machines

Noel Lopes
Supervisor: Prof. Dr. Bernardete Ribeiro

CISUC
University of Coimbra, Portugal

September 17, 2013