Benefits of Scaled Machine Learning

Reza Zadeh

Twitter: @Reza_Zadeh | http://reza-zadeh.com
Machine Learning Pipeline

Data → Learning Algorithm → Trained Model → Replicate model → Serve Model

Repeat entire pipeline
Scaling Machine Learning

Datasets and models growing faster than processing speeds

Solution is to parallelize on clusters and GPUs
Scaled ML at Matroid

Object recognition in Princeton ModelNet

» First on leaderboard for 40-class dataset

Matrix Computations and Optimization in Apache Spark

» Won KDD Best Paper Award runner-up
From Image Recognition to Object Recognition
Object recognition

Given 3D model, figure out what it is

» bathtub

Try using image recognition on projections, but that only goes so far.
Convolutional Network

Slide a two-dimensional patch over pixels.

How to adapt to three dimensions?
Multi-View CNN

Rotate camera around object

@Reza_Zadeh
Volumetric (V-CNN)

Simple idea: slide a three-dimensional volume over voxels.

@Reza_Zadeh
Volumetric CNNs

Use two different Volumetric CNNs (VCNN-I and VCNN-II). Example of one:

@Reza_Zadeh
FusionNet

Fusion of two volumetric representation CNNs and one pixel representation CNN

http://arxiv.org/abs/1607.05695
Matrix Computations and Optimization in Apache Spark
Traditional Network Programming

Message-passing between nodes (e.g. MPI)

**Very difficult** to do at scale:
» How to split problem across nodes?
  • Must consider network & data locality
» How to deal with failures? (inevitable at scale)
» Even worse: stragglers (node not failed, but slow)
» Ethernet networking not fast
» Have to write programs for each machine

Rarely used in commodity datacenters
Data Flow Models

Restrict the programming interface so that the system can do more automatically

Express jobs as graphs of high-level operators

» System picks how to split each operator into tasks and where to run each task

» Run parts twice fault recovery

Biggest example: MapReduce

Nowadays: Spark, TensorFlow
Deep Learning Data Flow

TensorFlow, Theano, etc: Provide automatic differentiation

Model creator doesn’t have to do the algebra or optimization

Makes it easier to explore models
Spark Computing Engine

Extends a programming language with a distributed collection data-structure

- “Resilient distributed datasets” (RDD)
- DataFrames and Datasets (inspired from R)

Open source at Apache

- Most active community in big data, with 200+ companies contributing

Clean APIs in Java, Scala, Python, R

@Reza_Zadeh
Machine Learning Pipeline

- Data
- Learning Algorithm
- Trained Model
- Replicate model
- Serve Model

Repeat entire pipeline
MLlib: Available algorithms

classification: logistic regression, linear SVM, naïve Bayes, least squares, classification tree, neural networks

regression: generalized linear models (GLMs), regression tree

collaborative filtering: alternating least squares (ALS), non-negative matrix factorization (NMF)

clustering: k-means

decomposition: SVD, PCA

optimization: stochastic gradient descent, L-BFGS
Simple Observation

Matrices are often quadratically larger than vectors

\[
A: n \times n \quad \text{(matrix)} \quad O(n^2)
\]

\[
v: n \times 1 \quad \text{(vector)} \quad O(n)
\]

Even \( n = 1 \) million makes cluster useful
Spark TFOCS

Conic optimization program solver

Solve e.g. LASSO

\[ \frac{1}{2} \|Ax - b\|_2^2 + \lambda \|x\|_1 \]

General Linear Programs

\[ \text{minimize } c \cdot x + \frac{1}{2} \mu \|x - x_0\|_2^2 \text{ s.t. } Ax = b \text{ and } x \geq 0 \]
Spark TFOCS

The implementation of TFOCS for Spark closely follows that of the Matlab TFOCS package.

Matrix Computations shipped to cluster, vector operations on driver
Eigenvalue Decomposition

ARPACK written in Fortran, decades ago

With foresight!

Use Eigenvalue Decomposition for Singular Value Decomposition

@Reza_Zadeh
Singular Value Decomposition

ARPACK: Very mature Fortran77 package for computing eigenvalue decompositions

JNI interface available via netlib-java

Distributed using Spark

@Reza_Zadeh
Square SVD via ARPACK

Only interfaces with distributed matrix via matrix-vector multiplies

\[ K_n = \begin{bmatrix} b & Ab & A^2b & \cdots & A^{n-1}b \end{bmatrix} \]

The result of matrix-vector multiply is small.

The multiplication can be distributed.

@Reza_Zadeh
Example mllib.linalg algorithms

» Matrix Multiplication
» Singular Value Decomposition (SVD)
» QR decomposition
» Optimization primitives

» … yours?

Simple idea goes a long way

@Reza_Zadeh
DEEM workshop

Held in conjunction with SIGMOD/PODS

May 14th, 2017 – Submissions open!
Thank you!

Matrix Computations paper


FusionNet paper


@Reza_Zadeh