TOWARDS DEEP LEARNING TECHNIQUES ON DISTRIBUTED SYSTEM

Presented by: Xiu Zhang
20161025
Outlines

• Significance
• Focus
• Task
• Evaluation
SIGNIFICANCE
Big Data: “4V”
Deep Learning

BIG DATA & DEEP LEARNING

Performance

Amount of Data

Deep Learning

Most Learning Algorithms

Input nodes

Connections

Hidden nodes

Output nodes
Distributed Computation
FOCUS: LEARN DISTRIBUTED DL
Current Distributed System Framework

http://yahoohadoop.tumblr.com/post/139916563586/caffeonspark-open-sourced-for-distributed-deep
CaffeonSpark: Latency & Simplicity

Hadoop/Spark Cluster

Spark application program:
1. Prepare datasets
2. DL training & Test
3. Apply DL model

Hadoop Datasets

http://yahoohadoop.tumblr.com/post/139916563586/caffeonspark-open-sourced-for-distributed-deep
TASK:
IMAGE CLASSIFICATION
Image Classification Framework

Traditional recognition: “Shallow” architecture

Deep learning: “Deep” architecture
EVALUATION
Experiments

• Dataset: PASCAL VOC 2007 Dataset
  – 5011 Training/Validation
  – 4902 Testing
  – 20 object classes

• Classifier:
  – SVM : one-vs-all
Experiments

• Evaluation Protocol
  – Time & Precision Comparison
  – Precision
    • MAP (Mean Average Precision)
    • Confusion Matrix to show inter-class correlations
Steps to do

- Pre-processing for the image dataset
- Learn How to use Caffe & CaffeonSpark
- Learn DL Models
- Evaluate the performance
- Analysis the results
THANK YOU
Any QUESTIONS?