BIOPHOTONICS
With PredictionIO, Spark and Deep Learning

Prajod Vettiyattil, Architect, Wipro
@prajods
https://in.linkedin.com/in/prajod
ABOUT ME

• Architect at Wipro
  • Big Data division of Open Source Solutions team
• Machine Learning
• Video Analytics
• Platform design and implementation
• Domain solutions
• Spark, Java, Python, DL4J, Tensorflow
AGENDA

- Bio photonics
- Applications
- PredictionIO
- Apache Spark
- DeepLearning4J and Tensorflow
- Cell detection process
- Deep learning and CNN
- Solution Architecture
SESSION OVERVIEW

In 4 slides
APPLICATIONS

- Self driving cars
- Robots
- Drones
- Industrial automation
- Physical security
- Medical labs
- Wherever images or videos are used
SESSION OVERVIEW

• Need in the healthcare domain
  • Speed up and automate, cell detections, counting and analysis
    • Diagnosis
    • Medical research

• Solution
  • Train a Deep Learning Model using digital images of living cells
  • Recognize test images with high accuracy

• Technology used
• Training process
CLASSIFICATION NEED

Input from the microscope

Expected output

Biophotonics using Apache PredictionIO, Spark and Deep Learning
@prajods

#apacheconbigdata
Biophotonics using Apache PredictionIO, Spark and Deep Learning
@prajods
#apacheconbigdata
BACKGROUND

How its done
INTRODUCTION

- Photonics: study and harness light
- The World of Small Things
- Microscopic life
- High end microscopes
- Data set scarcity
- Accessibility

nigms.nih.gov
LIVE CELL IMAGING

© Northwestern University
CONFOCAL MICROSCOPE

- Very high resolution
- Spatial features
IMAGE COMPARISON

Widefield Image

Confocal Image

meyerinst.com
ELECTRON MICROSCOPE

Ref: emc.sc.edu
What to do with all these images of micro stuff?
Spend hours peering through the lens?
• Even then
  • How many cells can one count in a minute?
  • How accurate is our ability to visually differentiate between bacterium A vs bacterium B?
  • How many patient blood samples can one analyze in an hour?
  • Can a doc detect all abnormalities with his endoscope?
  • How accurate is human visual diagnosis?
AUTOMATED ANALYSIS OF CELLS

- Detection of cells
- Count cells
- Distinguish cell A vs cell B
- Detect physical abnormalities
- Cell lifecycle analysis
Biophotonics using Apache PredictionIO, Spark and Deep Learning
#apacheconbigdata @prajods

TECHNOLOGY
PREDICTION IO

• Simplifies Machine Learning projects
  • Data storage
  • Training
  • Evaluate models
  • Deploy models
  • Serving predictions
PREDICTION IO

- DASE architecture
  - Data
  - Algorithm
  - Serving
  - Evaluation
PREDICTION IO

• Readymade ML templates
  • Classification
  • Regression
  • Recommendation
  • NLP
  • Clustering
  • Similarity
Biophotonics using Apache PredictionIO, Spark and Deep Learning
@prajods

#apacheconbigdata
PREDICTIONIO: PRODUCT VIEW

Client application

Event Server (Spray+Storage)

Other components

Serving Engine (Spray+Spark)

PredictionIO

Storage (Hbase/Postgres/MySQL)

Training Engine (Spark)

Evaluator

Biophotonics using Apache PredictionIO, Spark and Deep Learning
@prajods

#apacheconbigdata
APACHE SPARK

- Fast in memory data processing
  - Real time and batch modes
- Complements Hadoop
  - Replaces Hadoop MR
  - Adds
    - In memory processing
    - Stream processing
    - Fast for interactive queries
- YARN or Mesos for clustering
- Java, Scala, Python, R

Biophotonics using Apache PredictionIO, Spark and Deep Learning
@prajods

#apacheconbigdata
<table>
<thead>
<tr>
<th>Spark SQL</th>
<th>Spark Streaming</th>
<th>SparkML</th>
<th>GraphX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Apache Spark Core</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SPARK: DEPLOYMENT VIEW

Master Node

Spark Driver

Spark’s Cluster Manager

Worker Node

Executor

Task

Cache

Worker Node

Executor

Task

Cache

SPARK Driver

Biophotonics using Apache PredictionIO, Spark and Deep Learning
@prajods
#apacheconbigdata
DEEPLEARNING4J (DL4J)

- Deep learning library
- Open source
- Apache 2.0 license
- Java based
- Distributed execution
- Runs on Spark and Hadoop
• Deep Learning framework
  • from the Google Brain Team
• Python and C++ SDKs
• Dataflow graph based processing
• Tensors and Operations
• Numerical operations
• Lazy evaluation
• Distributed and parallel
  • Training and inference
• Good documentation
• Useful examples

Ref: tensorflow.org
• CPU, GPU
• Mobile: IOS and Android
• Core API in C
• Compiled models
• Visualization using TensorBoard
• Tensorflow Serving
WHAT DOES IT INVOLVE?
THE CELL DETECTION PROCESS

• Data gathering
• Data preparation
• Data extraction
• Model training
• Evaluation
DATA GATHERING

- “Google” it?
- Cell image data sets are not common
- Very few youtube videos
- Get the data set from the labs
- Caveat: Competitive information
DATA EXTRACTION

- Extract your own data sets from videos
- Different angles, lighting, perspective
- Multiple cells
- Image processing techniques
  - Edge detection
  - Segmentation
  - Background subtraction
  - Otsu
  - Watershed
MODEL TRAINING

• Custom models
  • Build your own
  • High difficulty in hyper parameter tuning
  • Very high training effort
  • Small sizes
  • Poor accuracy

• Transfer learning
  • Reuse an existing image detection model
    • Tensorflow’s inception
  • Replace its final layer/s
  • Very little hyper parameter tuning
  • Involves lower training time
EVALUATION

• Test, test, test
• Primary tests
  • Accuracy
  • Precision
  • Recall
  • F1 score
• Cross validate
  • With data sets
  • With different algorithms
DEEP LEARNING AND CNN
DEEP LEARNING: WHAT

- Neural Network based Machine Learning
- Neural Network
  - 1 Input layer
  - 1 or more hidden layers
  - 1 Output layer
- Basic unit of NN
  - Neuron
- Combine neurons in many ways using multiple parameters
A TWO LAYER NEURAL NETWORK
DEEP LEARNING: WHY

- For complex input patterns
- Higher accuracy
  - for image analysis
- For spatial data analysis
- Higher training time
- Parallel execution
- Higher level architectures for evolving needs
  - CNN, RNN, LSTM
CONVOLUTIONAL NEURAL NETWORKS
CONVOLUTIONAL NEURAL NETWORK

• A type of Neural Network
• Pass many filters(kernels) over an image
  • Capture its features
CNN: APPLICATIONS

- Self driving cars
- Robotics
- Drones
- Industrial automation
- Physical security
- Medical labs
- Wherever images or videos are used
CNN ARCHITECTURE

1st Convolution → Non linear transform (ReLU, tanh etc) → Pooling → ... → Nth Convolution → Non linear transform (ReLU, tanh etc) → Pooling → Fully connected layers → Output probabilities → Class probabilities

Feature extraction + dimension reduction

Classification
SOLUTION ARCHITECTURE

For cell detection
SOLUTION ARCHITECTURE: TRAINING

OpenCV

- Segment images
- Extract approx. images

Tensorflow

- Retrain inception
- Evaluate accuracy, precision, recall, F1 score

Biophotonics using Apache PredictionIO, Spark and Deep Learning
@prajods

#apacheconbigdata
SOLUTION ARCHITECTURE: PREDICTION

Client App

Spray

PredictionIO

Spark

Tensorflow + Cell Model

PySpark
OTHER PROJECTS

- Diabetic retinopathy
  - Google, IBM
- Cancer cell detection
  - Many institutions
- Startups
- Universities
- CT, MRI scans
SUMMARY

- Photonics
- Analyzing the small stuff
- Automating using Deep Learning
- PredictionIO + Spark + Tensorflow
- Distributed training and evaluation
REFERENCES

• Analysis of GPU tech used for biological research
  • https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3496509/

• Rethinking biology, by an MIT startup co-founder
  • http://nikhilbuduma.com/2014/12/27/the-cell-reimagined/

• Live cell imaging
  • https://www.leica-microsystems.com/science-lab/topics/what-is-live-cell-imaging/topic/Topic/////page_t/8/

• Good explanation of neural nets
  • http://karpathy.github.io/neuralnets/

• Stanford course material on CNN
  • http://cs231n.github.io/convolutional-networks/

• Intuitive explanation of CNN
  • https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

• The current generation of automation for cell analysis
  • https://en.wikipedia.org/wiki/Automated_analyser
Prajod Vettiyattil, Architect, Wipro
@prajods
https://in.linkedin.com/in/prajod