Beyond Training –
The next steps of Machine Learning

Chris Parsons
chrisparsons@uk.ibm.com
@chrisparsonsdev
/in/chrisparsonsdev
What is this talk?

Part 1 –

• What is Machine Learning?
• AI Infrastructure
• PowerAI
• Use Cases

Part 2 –

• Why?
• AI Vision
• iOS 101
• Considerations
• Demo
What is machine learning?
5041
Integrated & Supported AI Platform
Higher Productivity for Data Scientists
Enable non-Data Scientists to use AI

Enable non-data scientists to use AI
(Tools for ease of use)

Open Source: Supported Distribution
- TensorFlow
- Caffe
- PyTorch
- Keras
- Chainer

Faster Training for Data Scientists
(SW & HW Optimizations)

GPU-Accelerated Power Servers
Storage
Distributed Deep Learning
Distributed Deep Learning (DDL)

Deep learning training takes days to weeks

Limited scaling to multiple x86 servers

PowerAI with DDL enables scaling to 100s of servers

16 Days Down to 7 Hours
58x Faster

Near Ideal Scaling to 256 GPUs

95% Scaling with 256 GPUs

Caffe with PowerAI DDL, Running on Minsky (S822Lc) Power System
Large AI Models Train ~4 Times Faster

POWER9 Servers with NVLink to GPUs vs x86 Servers with PCIe to GPUs

Caffe with LMS (Large Model Support) Runtime of 1000 Iterations

GoogleNet model on Enlarged ImageNet Dataset (2240x2240)
Snap ML
Distributed GPU-Accelerated Machine Learning Library

Snap Machine Learning (ML) Library

- Logistic Regression
- Support Vector Machines (SVM)
- libGLM (C++ / CUDA Optimized Primitive Lib)
- Distributed Training

- Linear Regression
- More Coming Soon
- Distributed Hyper-Parameter Optimization

APIs for Popular ML Frameworks

- python
- scikit-learn
- (coming soon)
- Spark
Snap ML: Training Time Goes From An Hour to Minutes

46x faster than previous record set by Google

Workload: Click-through rate prediction for advertising

Logistic Regression Classifier in Snap ML using GPUs vs TensorFlow using CPU-only

Dataset: Criteo Terabyte Click Logs (http://labs.criteo.com/2013/12/download-terabyte-click-logs/)
4 billion training examples, 1 million features

Model: Logistic Regression: TensorFlow vs Snap ML

Test LogLoss: 0.1293 (Google using TensorFlow), 0.1292 (Snap ML)

Platform: 89 CPU-only machines in Google using TensorFlow versus 4 AC922 servers (each 2 Power9 CPUs + 4 V100 GPUs) for Snap ML

Google data from this Google blog

Logistic Regression in Snap ML (with GPUs) vs TensorFlow (CPU-only)

Runtime (Minutes)

1.1 Hours

1.53 Minutes

90 x86 Servers (CPU-only)

Google CPU-only

Snap ML Power + GPU

4 Power9 Servers With GPUs

46x Faster
Retail Bank – Masked Face Detection
More Use Cases
What is this talk?

Part 1 –

• What is Machine Learning?
• AI Infrastructure
• PowerAI
• Use Cases

Part 2 –

• Why?
• AI Vision
• iOS 101
• Considerations
• Demo
Why?
PowerAI Vision

- Automatically build Image Classification and Object Detection models
- Makes the data science easy!
- Curate datasets and train on one platform
- Easily deploy models as a web API
- Handles video and image workloads
XHR

XMLHttpRequest object calls Web Service

Web service

/Flights?from=London&to=Tokyo

XMLHttpRequest internal response fills up with Web Service response...
Who is Jason?

```json
{
    "classified":{
        "Larus": "1.0000"
    },
    "imageUrl": "https://aivisioninstance.com/api/tmp/img.png",
    "result": "success"
}
```
iOS Application Development 101

Swift
UI Kit
iOS Network Stack
Getting Images

```swift
@IBAction func chooseImageFromAlbum(_ sender: Any) {
    let pickerView = UIImagePickerController()
    pickerView.delegate = self
    pickerView.sourceType = .photoLibrary
    self.present(pickerView, animated: true, completion: nil)
}

@IBAction func chooseImageFromCamera(_ sender: Any) {
    let pickerView = UIImagePickerController()
    pickerView.delegate = self
    pickerView.sourceType = .camera
    self.present(pickerView, animated: true, completion: nil)
}
```
//URL for your AI Vision instance and model
let urlString = "API_URL_HERE"

//Set up HTTP Request Object
var request = URLRequest(url: URL(string: urlString)!
request.httpMethod = "POST"
let boundary = "Boundary-\(UUID().uuidString)"
request.setValue("multipart/form-data; boundary=\(boundary)", forHTTPHeaderField: "Content-Type")
request.setValue("gzip, deflate", forHTTPHeaderField: "Accept-Encoding")

let imageData = UIImagePNGRepresentation(image)!
let fileName = "upload.png"
let fullData = photoDataToFormData(data: imageData, boundary: boundary, fileName: fileName)

request.setValue(String(fullData.count), forHTTPHeaderField: "Content-Length")

request.httpBody = fullData
request.httpShouldHandleCookies = false
Demo

• Build dataset in AI Vision
• Train Model in AI Vision
• Deploy AI Vision Web API
• Build iOS App
• Connect to API
• Test iOS App
Thank You

Chris Parsons
chrisparsons@uk.ibm.com
@chrisparsonsdev
/in/chrisparsonsdev

Code and examples:
https://developer.ibm.com/code
https://github.com/chrisparsonsdev/vision_ios
https://ibm.biz/powerai_counting_cars