Hello, PowerAI

IBM PowerAI Platform

PowerAI RPM
Has all the DL frameworks and supporting Libraries

PowerAI Software Distribution

- Caffe
- NVIDIA Caffe
- IBM Caffe
- torch
- TensorFlow
- theano
- Chainer

Supporting Libraries
- DIGITS
- OpenBLAS
- Distributed Frameworks
- Bazel
- NCCL

IBM Power System for HPC, with NVLink
Breakthrough performance for GPU accelerated applications, including Deep Learning and Machine Learning.

Hardware
Minsky
P100 GPUs
OS – Ubuntu 16.04
Agenda

- What is AI?
- Why “Neural” Networks?
- What is Deep Learning?
- Distributed Deep Learning
- Large Model Support
- Identifying Use Cases
- PowerAI Installation Process
- More information
- Join in!
What is AI?

Artificial Intelligence
The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision making, and translation between languages.

Machine Learning
The capacity of a computer to learn from experience, i.e. to modify its processing on the basis of newly acquired information.

Deep Learning
The study of artificial neural networks and related Machine Learning Algorithms that include more than one hidden layer.
What is Machine Learning?
What is Machine Learning?
What is Machine Learning?
What is Machine Learning?
Why "Neural" Networks?

Remember this?

Bias +1

Point X position

Point Y position

Sigmoid activation function

Output
Red if OUT < 0.5
Blue if OUT >= 0.5

Input weights

w0 w1 w2
Why “Neural” Networks?

- Dendrite
- Cell body
- Axon
- Axon terminal
- Node of Ranvier
- Schwann cell
- Myelin sheath
- Nucleus
What is Deep Learning?

Simple Neural Network

Deep Learning Neural Network

- Input Layer
- Hidden Layer
- Output Layer

Credit: Hacker Noon
Distributed Deep Learning

Server
Distributed Deep Learning

Server

DDL over IB/FC

Server
Large Model Support
Why does LMS matter?

Tumor Proliferation Assessment – mitosis detection
Images from electron-microscope
Size of image - 70K * 60K

Data Transformation

<table>
<thead>
<tr>
<th>Framework</th>
<th>Format</th>
<th>Input Size (Faster R-CNN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caffe</td>
<td>LMDB</td>
<td>1K*1K</td>
</tr>
<tr>
<td>TensorFlow</td>
<td>TensorRecord</td>
<td>1K*1K</td>
</tr>
</tbody>
</table>

Data Distribution among training, validation and testing

Data Shuffle
<table>
<thead>
<tr>
<th>Manufacturing</th>
<th>Financial Services</th>
<th>Healthcare &amp; Life Sciences</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Predictive Maintenance</td>
<td>• Risk Analysis</td>
<td>• Patient Triage</td>
</tr>
<tr>
<td>• Process Optimisation</td>
<td>• Cross/Up Selling</td>
<td>• Proactive Health Management</td>
</tr>
<tr>
<td>• Demand Forecasting</td>
<td>• Credit Checks</td>
<td>• Real Time Alerts and Diagnostics</td>
</tr>
<tr>
<td></td>
<td>• Customer Segmentation</td>
<td>• Disease Identification</td>
</tr>
<tr>
<td>Retail</td>
<td>Utilities</td>
<td></td>
</tr>
<tr>
<td>• Inventory Planning</td>
<td>• Smart Grid Management</td>
<td></td>
</tr>
<tr>
<td>• Cross-Channel marketing</td>
<td>• Carbon Emissions</td>
<td></td>
</tr>
<tr>
<td>• Customer ROI and Lifetime</td>
<td>• Customer Specific Pricing</td>
<td></td>
</tr>
<tr>
<td>Value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Scheduling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Pricing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Social Media Analytics and</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer Sentiment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictive Maintenance</td>
<td>Risk Analysis</td>
<td>Patient Triage</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>Process Optimisation</td>
<td>Cross/Up Selling</td>
<td>Proactive Health Management</td>
</tr>
<tr>
<td>Demand Forecasting</td>
<td>Credit Checks</td>
<td>Real Time Alerts and Diagnostics</td>
</tr>
<tr>
<td></td>
<td>Customer Segmentation</td>
<td>Disease Identification</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inventory Planning</td>
<td>Smart Grid Management</td>
<td>Scheduling</td>
</tr>
<tr>
<td>Cross-Channel marketing</td>
<td>Carbon Emissions</td>
<td>Pricing</td>
</tr>
<tr>
<td>Customer ROI and Lifetime Value</td>
<td>Customer Specific Pricing</td>
<td>Social Media Analytics and Customer Sentiment</td>
</tr>
</tbody>
</table>
POWERAI INSTALLATION
redbooks@pts153:~/downloads$ ls
7fa2af80.pub
cuda-repo-ubuntu1604-8-0-local-ga2v2_8.0.61-1_ppc64el.deb
libcudnn6_6.0.21-1+cuda8.0_ppc64el.deb
mldl-repo-local_4.0.0_ppc64el.deb
nvidia-driver-local-repo-ubuntu1604-384.81.0-1_ppc64el.deb
redbooks@pts153:~/downloads$ sudo apt-key add 7fa2af80.pub
[sudo] password for redbooks:
OK
redbooks@pts153:~/downloads$ sudo dpkg -i cuda-repo-ubuntu1604-8-0-local-ga2v2_8.0.61-1_ppc64el.deb
redbooks@pts153:~/downloads$ sudo dpkg -i cuda-repo-ubuntu1604-8-0-local-ga2v2_8.0.61-1_ppc64el.deb
Selecting previously unselected package cuda-repo-ubuntu1604-8-0-local-ga2v2.
(Reading database ... 57530 files and directories currently installed.)
Preparing to unpack cuda-repo-ubuntu1604-8-0-local-ga2v2_8.0.61-1_ppc64el.deb ...
Unpacking cuda-repo-ubuntu1604-8-0-local-ga2v2 (8.0.61-1) ...
Setting up cuda-repo-ubuntu1604-8-0-local-ga2v2 (8.0.61-1) ...
OK
redbooks@pts153:~/downloads$
redbooks@pts153:~/$ downloads$ sudo apt-get update
Get:1 file:/var/cuda-repo-8-0-local-ga2v2 InRelease
Ign:1 file:/var/cuda-repo-8-0-local-ga2v2 InRelease
Get:2 file:/var/cuda-repo-8-0-local-ga2v2 Release [574 B]
Get:2 file:/var/cuda-repo-8-0-local-ga2v2 Release [574 B]
Get:3 file:/var/cuda-repo-8-0-local-ga2v2 Release.gpg [819 B]
Get:3 file:/var/cuda-repo-8-0-local-ga2v2 Release.gpg [819 B]
Get:5 file:/var/cuda-repo-8-0-local-ga2v2 Packages [10.8 kB]
Hit:6 http://us.ports.ubuntu.com/ubuntu-ports xenial InRelease
Get:7 http://ports.ubuntu.com/ubuntu-ports xenial-security/universe ppc64el Packages [146 kB]
Fetched 452 kB in 0s (549 kB/s)
Reading package lists... Done
redbooks@pts153:~/downloads$ export PATH=/usr/local/cuda-8.0/bin:${PATH}:
redbooks@pts153:~/downloads$ export LD_LIBRARY_PATH=/usr/local/cuda-8.0/lib64:${LD_LIBRARY_PATH}:
redbooks@pts153:~/downloads$
redbooks@pts153:$ sudo dpkg -i libcudnn*
Selecting previously unselected package libcudnn6.
(Reading database ... 73296 files and directories currently installed.)
Preparing to unpack libcudnn6_6.0.21-1+cuda8.0_ppc64el.deb ... 
Unpacking libcudnn6 (6.0.21-1+cuda8.0) ...
Selecting previously unselected package libcudnn6-dev.
Preparing to unpack libcudnn6-dev_6.0.21-1+cuda8.0_ppc64el.deb ... 
Unpacking libcudnn6-dev (6.0.21-1+cuda8.0) ...
Setting up libcudnn6 (6.0.21-1+cuda8.0) ...
Setting up libcudnn6-dev (6.0.21-1+cuda8.0) ...
update-alternatives: using /usr/include/powerpc64le-linux-gnu/cudnn_v6.h to provide /usr/include/cudnn.h (libcudnn) in auto mode
Processing triggers for libc-bin (2.23-0ubuntu5) ...
sudo dpkg -i nvidia-driver-local-repo-ubuntu1604-384.81_1.0-1_ppc64el.deb
Selecting previously unselected package nvidia-driver-local-repo-ubuntu1604-384.81.
(Reading database ... 73308 files and directories currently installed.)
Preparing to unpack nvidia-driver-local-repo-ubuntu1604-384.81_1.0-1_ppc64el.deb ...
Unpacking nvidia-driver-local-repo-ubuntu1604-384.81 (1.0-1) ...
Setting up nvidia-driver-local-repo-ubuntu1604-384.81 (1.0-1) ...
redbooks@pts153:~/downloads$ sudo apt-get update
redbooks@pts153:~/downloads$ sudo apt-get install cuda-drivers
**redbooks@pts153:**

```
sudo reboot
```

Connection to closed by remote host.
Connection to closed.

```
chrisparsons@Chriss-MBP ~ $
```
```
redbooks@pts153:~$ downloads$ nvidia-smi
Mon Oct 16 15:19:43 2017

+--------------------------------------------------------------------------
<table>
<thead>
<tr>
<th>NVIDIA-SMI 384.81  Driver Version: 384.81</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPU</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>N/A</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>N/A</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>N/A</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>N/A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Processes:</th>
<th>GPU Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPU</td>
<td>PID</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>No running processes found</td>
<td></td>
</tr>
</tbody>
</table>
```

redbooks@pts153:~/downloads$ sudo dpkg -i mldl-repo-local_4.0.0_ppc64el.deb
[sudo] password for redbooks:
(Reading database ... 87938 files and directories currently installed.)
Preparing to unpack mldl-repo-local_4.0.0_ppc64el.deb ...
Unpacking mldl-repo-local (4.0.0) ...
Setting up mldl-repo-local (4.0.0) ...
OK
redbooks@pts153:~$ sudo apt-get update
Get:1 file:/var/cuda-repo-8-0-local-ga2v2 InRelease
Ign:1 file:/var/cuda-repo-8-0-local-ga2v2 InRelease
Get:2 file:/opt/DL/repo xenial InRelease [1,830 B]
Get:3 file:/var/nvidia-driver-local-repo-384.81 InRelease
Ign:3 file:/var/nvidia-driver-local-repo-384.81 InRelease
Get:4 file:/var/cuda-repo-8-0-local-ga2v2 Release [574 B]
Get:2 file:/opt/DL/repo xenial InRelease [1,830 B]
Get:5 file:/var/nvidia-driver-local-repo-384.81 Release [574 B]
Get:4 file:/var/cuda-repo-8-0-local-ga2v2 Release [574 B]
Get:5 file:/var/nvidia-driver-local-repo-384.81 Release [574 B]
Hit:6 http://ports.ubuntu.com/ubuntu-ports xenial-security InRelease
Get:7 file:/opt/DL/repo xenial/main ppc64el Packages [38.7 kB]
Hit:9 http://us.ports.ubuntu.com/ubuntu-ports xenial InRelease
Get:13 http://us.ports.ubuntu.com/ubuntu-ports xenial-updates/universe ppc64el Packages [460 kB]
Fetched 665 kB in 1s (503 kB/s)
Reading package lists... Done
redbooks@pts153:~$
redbooks@pts153:/downloads$ sudo apt-get install power-mldl
redbooks@pts153:~$ downloads$ sudo apt-get install linux-tools-common
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
  linux-tools-common
0 upgraded, 1 newly installed, 0 to remove and 143 not upgraded.
Need to get 114 kB of archives.
After this operation, 433 kB of additional disk space will be used.
Get:1 http://us.ports.ubuntu.com/ubuntu-ports xenial-updates/main ppc64el linux-tools-common all 4.4.0-97.120 [114 kB]
Fetched 114 kB in 0s (260 kB/s)
Selecting previously unselected package linux-tools-common.
Reading database ... 124721 files and directories currently installed.
Preparing to unpack .../linux-tools-common_4.4.0-97.120_all.deb ...
Unpacking linux-tools-common (4.4.0-97.120) ...
Processing triggers for man-db (2.7.5-1) ...
Setting up linux-tools-common (4.4.0-97.120) ...
redbooks@pts153:~$ downloads$
```
redbooks@pts153:~/downloads$ ls /opt/DL/
bazel  caffe  caffe-bvlc  caffe-ibm  caffe-nv  chainer  ddl  ddl-tensorflow  digits  nccl  open
```
PERFORMANCE OPTIMISATIONS
redbooks@pts153:~/downloads$ sudo ppc64_cpu --smt=2
redbooks@pts153:~/downloads$ sudo nvidia-smi -pm ENABLED
Enabled persistence mode for GPU 00000002:01:00.0.
Enabled persistence mode for GPU 00000003:01:00.0.
Enabled persistence mode for GPU 0000000A:01:00.0.
Enabled persistence mode for GPU 0000000B:01:00.0.
All done.
redbooks@pts153:~/downloads$ sudo nvidia-smi -ac 715,1480
Applications clocks set to "(MEM 715, SM 1480)" for GPU 00000002:01:00.0
Applications clocks set to "(MEM 715, SM 1480)" for GPU 00000003:01:00.0
Applications clocks set to "(MEM 715, SM 1480)" for GPU 0000000A:01:00.0
Applications clocks set to "(MEM 715, SM 1480)" for GPU 0000000B:01:00.0
All done.
```
redbooks@pts153:~/downloads$ source /opt/DL/tensorflow/bin/tensorflow-activate
redbooks@pts153:~/downloads$
```
redbooks@pts153:~/downloads$ tensorflow-test
More Information/Join In!

- PowerAI Installation Tutorial: http://ibm.biz/poweraivideo
- PowerAI Developer Content: https://ibm.biz/Bdj3hi
- Meetup (21/12/17): https://www.meetup.com/powerailondon/

- chrisparsons@uk.ibm.com
- Twitter: @chrisparsonsdev
Thank You