LiCO: Simplifying and accelerating deep learning application development

David Ellison, PhD 2018-05-24

NOVO

18 Lenovo Internal. All rights reservec

Lessons From Medicine: 'Modernizing' Deep Learning Model Development



Experimental Process for Deep Learning



AI Use Cases that Lenovo is Exploring Today

For better quality control on

Manufacturing Lines

Mark III

Lenovo

For proactive actions improving

Water & Energy Conservation



≜UCL

To advance particle

physics at

CERN's Large

Hadron Collider

Lenovo

A more efficient and effective

Radiology Practice

CodaLab

Current Models of Development (CRISP-DM)





Progression vs. Safety



Leonard McCombe/Time Life Pictures/Getty Images





CNBC & Guyana News

Experimental Process for Deep Learning



How Will We Measure Success?



1. Choose metrics first

- 2. Get data and access early
- 3. Start collecting historical

data ASAP

4. Plan for data restructuring

Step 1: Create a list of use cases. Sample list for consumer-packagedgoods company

Sales/customer relationship management (CRM)

- 1. Overall brand management
- 2. Overall campaign management
- 3. 360° view of shopper
 - 4. Targeted acquisition campaigns
 - 5. Real-time image advertising (awareness)
- 6. Retargeting campaign

Marketing

- 7. Optimization of spend across media
- 8. Optimization of spend within digital media
- 9. Digital attribution modeling
- 10. Performance advertising (sales)

Innovation

- 11. Consumer insights (social listening/sentiment analysis)
- 12. New product success (predictive behavior model)
- 13. Product customization at scale
- Open innovation on promotion mechanisms
 New digital sales models

Sample impact vs feasibility matrix

Step 2: Prioritize them.



McKinsey&Company

McKinsey Co 2018, Ten Red Flags Signaling Your Analytics Program Will Fail

Trade-off Between Metrics





Canzaini et al 2017 An Analysis of Deep Neural Network Models for Practical Applications

Deploy Simplest Model Possible



1. Statistical learning

2. Machine learning

3. Deep learning



David Ellison, unpublished data

How Will the Deep Learning Model be Used?

- Know how to:
- 1. Get data to your model
- 2. Evaluate your model
- 3. Use results of your model



Image from Intel Customer Engagements

Metrics and Decision-Making



- What metrics will be used to evaluate ethics?
 - Excluding obviously unethical features is not enough
 - What is the balance of fairness and safety?

- 2. Who decides what is ethical?
 - Programmers/Management
 - External board
- 3. At what point is it unethical NOT to use an algorithm?



Top Image: Ohio State Bar, individuals pictured are actors/models Bottom Image: Uber dash cam of Tempe, AZ accident

Safety in AI



- Avoid Negative Side Effects
 - Cleaning robot should not knock things over to clean faster. Should not manually specify rules
- Avoiding Reward Hacking
 - Ensure the cleaning robot won't game its own reward function. Might cover over messes, or hide not to tell humans about the mess
- Scalable Oversight
 - How to control for big, infrequent mistakes that require human intervention. How does the cleaning robot it treat candy wrappers vs. lost cellphones
- Safe Exploration
 - Explore new space without obvious mistakes. Cleaning robot attempting to wet mop electric outlets
- Robustness to Distributional Shifts
 - Adapting to a change in environment, e.g. office cleaning vs factory cleaning

Amodei et al, 2016 Concrete Problems in Al Safety, p3

Experimental Process for Deep Learning



LiCO Stack for Artificial Intelligence



Web Portal
Pre-trained Models
Workflow Templates

Distributed AI TrainingAI Frameworks



Open-Source Cluster Management
 Optimized HW Libraries



Shared under NDA with IBM May 22 2018

LiCO for Cluster Deployment

Easily manage AI & HPC clusters, workloads & users.



Efficiently manages **workflows** for AI training tasks

- Easy-to-use GUI & templates to simplify job management
- Leverage pre-trained models, monitor training & manage job history for rapid deployment.
- Multi-user support & intuitive admin controls.



Simplifies open source deployment in enterprises

- Validated stacks reduce time-to-development
- Choice of multiple AI frameworks through containers



Offers **flexibility** with hardware infrastructure

 Supports both NVIDIA GPUs and Intel processors to suit varying workloads



Shared under NDA with IBM May 22 2018

Optimally uses hardware resources for better TCO

- Improve scaling efficiency for distributed model training
- Manages cluster resources among multiple jobs



enovo-

Track Data Sets, Topologies, Models



😭 LiCO	≡				Welcome, demoadmin	anny ka
☆ Home ★ Submit job	Home > Train model > Datasets > mnist mnist Completed		Size 1.4 MB	Create time End time 2018-03-23 11:41 2018-03-23	Back accorney (A	counsy) kos (5:dtmarWBBLoss)
 Dobs Train model Datasets Topologies Models Pretrained Models Expert mode Admin 	 Control Control Cont	ome > Train model > Topologies Create Creat	■ Home > Train model > Models > mnist_de2_20180510103649 mnist_de2_20180510103649 ① Completed 2018-05-10 13:37 - 2018-05-10 13:38	Epoch	Welco Training Accuracy 96.88% 93.36%	me, demoadmin
	 <<p>< Topologies</p> S Models 	 Train model Datasets Topologies Models Pretrained Models Expert mode Admin 	Monitor Log Data analysis Test Export	Information Training loss 2-5 - - - - - - - - - - - - -	Processing speed	favil kernel söte: 5 sinkle: 1 poli, 0

Complete Model History





2018 Lenovo Internal. All rights reserved.

Consistent Visual Inspections







Flexibility Across Frameworks



Lenovo

Others Can Validate & Adjust Model

😭 LiCO	Ξ						Welcome, demoadmi	in Q
命 Home	Home \rightarrow Train model \rightarrow Models \rightarrow	mnist_exp12						Back
	mnist_exp12 🕀 📿 🗖				Epoch Training	Accuracy Test Accuracy	Validation Acc	curacy
🛨 Submit job	Rerun							
🗐 Jobs	LiCO	Copy model					×	Welcome, demoadmin
🕑 Train model 🔷	ழி Home	1 * Name	mnist_exp12_20180520183535		* Training epochs	64		Back couracy Validation Accuracy
🖨 Datasets	🛨 Submit job	* Topology	LeNet	Choose	* Train batch size	64		.8% 91.67%
🛠 Topologies	🗐 Jobs	Python Laver File		Browse	* Test batch size	64		
😂 Models	🕑 Train model 🔷							
😚 Pretrained Models	🖹 Datasets	Initial weight		Choose	* Validation batch size	64		ssing speed
♡ Expert mode	≪° Topologies	* Datasets	Training Dataset:mnist	Choose	* Snapshot Interval	1	Epochs	
招 Admin ~	😂 Models		Validation		* Weight Decay	0.0005		
	😚 Pretrained Models	* Queue	stark	~				
					* Regularization Type	L2	~	
	♥ Expert mode	* Number of nodes	2		* Image normalization	Yes	~	
	🕆 Admin	* Cores per node	72					
					Crop size			28 31 34 37 40 43 46 49 52 55 58 61 64
		* Container Images	intel-caffe-system	~	* Base learning rate	0.0001		
		Description						rocessing speed
				11	* Learning Rate Policy	Constant	~	

Lessons From Medicine: 'Modernizing' Deep Learning Model Development



Experimental Process for Deep Learning



AI Use Cases that Lenovo is Exploring Today

For better quality control on

Manufacturing Lines

Mark III

Lenovo

For proactive actions improving

Water & Energy Conservation



≜UCL

To advance particle

physics at

CERN's Large

Hadron Collider

Lenovo

A more efficient and effective

Radiology Practice

CodaLab

AI: Better quality control for Manufacturing

Breakthrough innovation in Manufacturing with Lenovo AI

Current State of Manufacturing:

- Better Quality Control directly related to:
 - More yield at higher speed
 - Lower production costs by faster adjustments to process

Where AI can help:

- Leverage cameras and sensors through-out the production lifecycle
- Better manage quality through product age of customizations

Al Inference and Training:

- Image / Pattern recognition and analysis
- ML training of products at different stages of production

See demo in action at SC17, Booth 1353

2017 Lenovo Internal. All rights reserved.



AI: For Proactive Actions Improving Drought Management

Lenovo AI helping conserve water and energy for food security

Current State and Challenges :

- Agriculture accounts for 70% of the total global freshwater withdrawals
- As water availability for agriculture becomes scarce and adds cost, it is critical to identify areas that will have impacts on certain crops to manage drought

Where AI can help:

 <u>Predict drought prone areas</u>: AI and deep learning can help identifying patterns from large amounts of geospatial data that will most likely face water challenges

Al Inference and Training:

 Deep learning training with satellite images for feature extraction of agriculture land, crops, and water resources to access drought conditions



AI: To Advance Particle Physics in the Large Hadron Collider

Lenovo Al progressing advanced Particle Physics

Current State and Challenges :

- Reconstructing particle trajectories from thousands of sensor measurements in the detector is an important data analytics task
- Traditional computational methods demand high amounts of resources and don't scale well as the LHC pushes to higher collision frequencies

Where AI can help:

• <u>Pattern recognition</u>: Reconstructing particle trajectories using imaging data from the collider much more efficiently than traditional methods

Al Inference and Training:

 Machine learning methods using binary image data from experiments combined with integral transforms for pattern recognition



AI: A More Efficient and Effective Radiology Practice

Lenovo Helping advance Cancer Research

Current State of Radiology:

- Central to patient care
- Top 4 area of expense for hospitals
- 93% of US Radiology groups under staffed

Where AI can help:

- <u>Better Efficiency</u>: AI helps prioritize highest risk patients, optimizes radiologist time and lead to better patient outcome
- <u>Improve Effectiveness</u>: AI can offer a second screening to physician finding to reduce avoidable errors

Al Inference and Training:

- Image / Pattern recognition and analysis
- ML training with millions of patient images and outcomes

See demo in action at SC17, Booth 1353







enovo.

Different is better

Appendix

Dynamic Analysis of Cell Secretions for Cardiac Repair



Ellison, Gupta, Suhail, et al. microFLISA: A New Experimental and Computational Platform For Analysis of Dynamic Secretomes Uncovers a Cardioprotective Secretion Signature. Under Review with Nature BME.

Controlled: Track Data Sets



Principle: Keep Snap Shots of Your Data

- Minimally ensuring expected class balance and distributions ullet
- Extremely large or streaming data poses more complex problems •

Controlled: Track Topologies

😭 LiCO	≡				
合 Home	Home > Train mo	del > Topologies			
🛨 Submit job	Create				
卣 Jobs	ID 11 D	Name 11	Framework 11	Type ↑↓	Description 11
P Train model ^	5 I	LeNet_Sigmoid	Intel-Caffe	Private	Ellison - LeNet with modified activation layer (ReLU layer changed to Sig
Datasets	4 1	MiroNet	Intel-Caffe	Private	Modified LeNet
< Topologies	3 (GoogLeNet	Intel-Caffe	System	
😂 Models	2 4	AlexNet	Intel-Caffe	System	
😚 Pretrained Models	1 I	LeNet	Intel-Caffe	System	
💝 Expert mode	Total 5 10/j	page 🗸			
😫 Admin 🗸 🗸					

Principle: Record and Check Neural Network Topologies

- Changes in activation functions can have unexpected results •
- Quick visualization for sanity checks •
- Enable faster transfer learning ۲

melane Sarah In

Bath 64 /Bath 64

Controlled: Consistent Job Runs/Models



Principle: Robustly Record Successful Models

- Establish metrics and consistently measure the across models
- Look for and record anomalies
- Simple to systematically do hyper parameter tuning

Controlled: Remember Poor Models



Principle: Record and Learn from Poor Models

- Mitigates survival bias
- Data needed for later robustness analysis

Controlled: Flexibility Across Frameworks



Total 12 < 1 > Go to 1

Controlled: Same Metrics Across Frameworks



Reproducible: Others Can Validate Your Model



Principle: If Others Cannot Easily Reproduce Your Model, No One Will

Reproducible: Others Can Make Adjustments

😭 Lico	Copy model					×	Welcome, demoadmin
ப் Home	I * Name	mnist_exp12_20180520183535		* Training epochs	64]	Back Accuracy Validation Accuracy
🛨 Submit job	* Topology	LeNet	Choose	* Train batch size	64		.8% 91.67%
🗐 Jobs	Python Layer File		Browse	* Test batch size	64		
⑦ Train model ^	Initial weight		Choose	* Validation batch size	64		
🖹 Datasets	initial weight			validation oaten size			ssing speed
ổ Topologies	* Datasets	Training Dataset:mnist Validation	Choose	* Snapshot Interval	1 Epochs		
Models Models		Vallaution		* Weight Decay	0.0005		
😚 Pretrained Models	* Queue	stark	~	* Regularization Type	L2 ~		
💱 Expert mode	* Number of nodes	2					
🕆 Admin	* Cores per node	72		* Image normalization	Yes 🗸		
				Crop size			28 31 34 37 40 43 46 49 52 55 58 61 64
	* Container Images	intel-caffe-system	~	* Base learning rate	0.0001	ן	
	Description					J	rocessing speed
			1	* Learning Rate Policy	Constant ~		

Principle: Others Should be Able to Test the Boundaries/Robustness of Your Model