Machine Learning Fast & Slow

Suman Deb Roy

Lead Data Scientist @ betaworks







D**S**TS BLOGLOVIN' **GIPHY**





Digg on Messaging

www.digg.com/messaging





www.poncho.is

www.rundexter.com







1: Poncho

A weather cat that sends you personalized weather messages.

• Algorithms + Humans

 Not every feature in weather data has equal importance – what's actionable?





2: Digg Trending

- Ranked each day:
 - 10 million RSS feeds, 200 million tweets, 7.5
 million new articles ranked each day



Clinton Wins DC, Final Primary Of The Year, And Meets With Sanders

cnn.com · The contest doesn't change the general election match-up. Clinton clin... on.digg.com



Authorities Looking Into Orlando Shooter's

theguardian.com · The wife of the Orlando gun charges if ... on.digg.com

Share

m.me/digg

Share

Up next, the Digg editors' most interesting stories from today.



Type a message...

Ŧ



3: Digg Deeper



Instapaper



4: Instapaper's InstaRank

••••• 02-UK 🗢 🐇	17:03	o 43% 📭	••••• 02-UK 🗢 🕆	17:03	o 43% 💶	••••• 02-UK 🗢	* 17:07	0 43% • +
≡ Re	ead Later	Edit	≡ R	ead Later	Edit	≡	Read Late	r Edit
Popular articles	first	×	Longest articles	first	×	Sort		Filter
From AAA to do	ing it our way		iOS 7: The Pixel	Envy Review	– Pixel	Biscuit for iP appfreakblog.com	Pad is a class • by AppFreak	sy Reddit cli
By Julia Alexander on O	ctober 11, 2013 at 12	:oop•F•Al	A year ago — to the day	y — I closed my review	v of iOS 6	Biscuit by Matthia	s Tretter is a Reddi	t client focused o
Tesco Hudl tabl theguardian.com • by S	et review: a lo Samuel Gibbs	t of tab	iOS: A visual his The Verge • by Verge St	story aff			Sort By	
Sinéad O'Connor's open letter to Miley Cyrus			In what is widely regarded as his greatest presentatio			Newest Saved		
						0	Idest Sav	ed
Countdown To T Forbes • by Dina Gachn	The Emmys: Us	sing So	Lanyrd: from id	ea to exit - the	story o	Lo	ngest Arti	cles
Why the military is ditching double-A batteries http://		RT @Natbat: I've written the full story of our last 3 yea						
		$\tilde{a} \in a^*$				Sh	ortest Arti	cles
Amount Conver MacStories • by Federic	sions From La	unch C EDT	And Then Steve	Said, 'Let The VOGELSTEIN	ere Be a		Popularity	/
In September, I reviewe	ed Amount, an elegar	nt and eas	The 55 miles from Cam	pbell to San Francisc	o make for		Shuffle	
Opinion: To riva	al the film indu	ıstry, g	Apple Re-Invent	ts Its Wheel W	ith iOS		Cancel	



5: Scale Model

Communities Not Keywords



5/21150 (e	poch 0.674),	train_loss =	2.35138823,	grad/param
6/21150 (e	poch 0.676),	train_loss =	2.32717855,	grad/param
7/21150 (e	poch 0.678),	train_loss =	2.28651274.	grad/param
8/21150 (e	poch 0.681).	train loss =	2.34484470.	arad/param
9/21150 (e	poch (0.683).	train loss -	2.3554824	arad/param
0/21150 (0	Training	Neural Net	tworks 😶	anad/param
1/21150 (4	anoch @ 688)	train loss -	2 30763116	grad/param
2/21150 (4	apoch (0.600),	train loss -	2 35106219	grad/param
3/21150 (apoch (0.603)	train_loss =	2 20260427	grou/param
4/21150 (apoch 0.695),	train_loss =	2.29209127,	grad/param
5/21150 (epoch 0.095),	train_loss =	2.32/431/3,	graa/param
	epoch 0.097),	train_loss =	2.35125453,	graa/param
07/21150 (0	epoch 0.700),	train_loss =	2.35151182,	graa/param
9//21150 (0	epoch 0.702),	train_loss =	2.42091541,	graa/param
98/21150 (0	epoch 0.704),	train_loss =	2.32718820,	graa/param
99/21150 (epoch 0.707),	train_loss =	2.35932102,	grad/param
00/21150 (epoch 0.709),	train_loss =	2.24577900,	grad/param
01/21150 (epoch 0.712),	train_loss =	2.27098633,	grad/param
02/21150 (epoch 0.714),	train_loss =	2.31136981,	grad/param
03/21150 (epoch 0.716),	train_loss =	2.23584678,	grad/param
04/21150 (epoch 0.719),	train_loss =	2.23303004,	grad/param
05/21150 (epoch 0.721),	train_loss =	2.30332945,	grad/param
06/21150 (epoch 0.723),	train_loss =	2.2898/039,	grad/param
Ø7/21150 (epoch 0.726),	train_loss =	2.22005898,	grad/param
908/21150 (epoch 0.728),	train_loss =	2.29200139,	grad/param
909/21150 (epoch 0.730),	train_loss =	2.33206190,	grad/param r
510/21150 (epocn 0.735),	train_loss =	2.23010755,	grad/param r
511/21150 ((epoch 0.735),	train_loss =	2 35167313	grad/param r
312/21150 (epoch 0.750),	train loss =	2.33170996	arad/param r
313/21150 ((epoch 0 742)	train loss =	2.32389178.	arad/param r
315/21150 ((enoch 0, 745)	train loss =	2.27157382.	grad/param r
316/21150 (epoch 0.747).	train_loss =	2.28681223.	grad/param n
317/21150	(epoch 0.749).	train_loss =	2.30380847,	grad/param n
318/21150	(epoch 0.752),	train_loss =	2.34555733,	grad/param n
319/21150	(epoch 0.754),	train_loss =	2.29328290,	grad/param n
320/21150 ((epoch 0.757),	train_loss =	2.26914802,	grad/param n
321/21150	(epoch 0.759),	train_loss =	2.23794461,	grad/param n
322/21150	(epoch 0.761),	train_loss =	2.23613483,	grad/param n
323/21150	(epoch 0.764),	train_loss =	2.26098966,	grad/param n
324/21150	(epoch 0.766),	train_loss =	2.27066436,	grad/param n

$\begin{array}{c} \text{MACHINE LEARNING} \\ \text{WAS} \, HARD \end{array}$

Don't get technical with me.

$\mathsf{ITS}\,STILL\,\,\mathsf{HARD}$



Moving fast and slow

- Fast:
 - Experience, Similar Problems, Pre-existing pipelines
- Slow:

- New type of data, Bootstrap, Scaling

- Main challenge:
 - how to jump between states, when to change gears.



Effects of moving Fast

- Technical debt?
 - -Refactoring code
 - improving unit tests
 - -delete dead code
 - reducing dependencies
 - -tightening APIs
 - improving documentation

Effects of moving Slow

- Growth debt?
 - Waiting team mates
 - Uncertain quality assurance
 - Piling up further requests
 - Hypothesis might not be feedback driven
 - Overthinking the solution

Maintenance

Code Level

- How researchable, reusable, deployable

- System Level
 - Eroding abstraction boundaries

- Data Level
 - Data influences ML behavior.

Data vs. Code Organization

• Snapshotting .. Detects bias

Interface at the method , be procedural
 – Easy to execute portions of the code.

- Separate hyper-arguments from parameters
 - Parameter: How your model is specified
 - Hyper-Arguments: How your algorithm should run

Unstable APIs

- Who owns the data stream?
- Who owns the model ?
- Ownership by
 - entire solution
 - Expertise? DB ? Pipelines? Algorithms? Stats
- Debug?

- Frozen versioning instead of continual

Feature Erosion

• User behavior with new model could make features of current model unimportant

• How can we detect this?

• How can we prevent this?

Predictor Variables

• Myth: If you add a few more variables, the predictor will be better.

 If the predictors have realistic priors, their coefficients could be appropriately pulled down (in expectation) and over fitting shouldn't be such a problem

Visualizations

Any ML algorithm must be seen to believe it.



Visualizations



Research vs. Production

 Collaboration looks very different based on the end goals

• Do you need to master git or just get by

• How quickly can you move something from iPython to production grade?

Even the best tools.

- Lets talk about iPython notebooks:
 - Version Control
 - Fragmented Code is deadly for production grade.
 - Security issue : all those open ports
 - Code Reviews and Pull Requests.

Heuristic Escape

"Heuristic is an algorithm in a clown suit. It's less predictable, it's more fun, and it comes without a 30day, money-back guarantee."

- Steve McConnell, Code Complete

Domain of Impact

 Most engineers and computers scientists will conceptualize domains as primarily a rational, evidence-based, problem-solving enterprise focused on well-defined conditions.

• But the real world is more complex!

• e.g.,: Trending News Algorithms

Invention vs. Innovation

- What is ML good at? Both ?
- Not outside the box, instead connect them.
- innovation = improve significantly by adjusting ML method
- invention = totally new ML method.

Fitting ML into the betaworks model



Code & Data Residence

- ML module transfer
 - Code transfer
 - Core module
 - Model updating component
 - Analysis component
 - Data transfer
 - Infrastructure rebuild?
 - Performance
 - maintenance

Powered by deepNews

Research ready pipelines



Powered by deepNews + Scale Model

Second order Analysis

Reach in Conservatives (X), Progressives (Y) and Media (bubble size) communities



Conservatives Reach

Progressives Reach



Conversational Software

whats going on with benzema

0 (5)

-

00

 $(\cdot \cdot)$

Here ya go:





HUMAN BOT INVOLVEMENTA HUMAN INTER > Personality CONNECTION INCREASES 1 Episodic Memory Emotion ALGORITHMS PERFORM Social BETTER Learning Topics / Contextual Facts / Transactional

ZERO automated solutions Affective Computing trending dígg deeper topics Topic Modeling DBpedía Freebase Apps for transactional tasks APIS

MANY automated solutions



LOW VALUE of historical data

Data Types by Company

- Digg has topic modeling/ news data
- Scale model has social graph data
- Poncho has weather data/editorialized personality
- Giphy has gifs (emotion++)
- Instapaper has reading data
- Dexter has hooks to APIs

Transfer Learning



Yosinski et. al. How transferrable are deep learning features?, in NIPS 2014

To Sum up

- Constraints to ML solutions occur at three levels:
 - Algorithmic
 - Data
 - Humans
- These parameters lead to several oscillating cycles of fast and slow impact of ML
- Whats good for you?

ML 2016

- Understood by few, hyped by some, revered by most.
- Can be the difference between a company scaling vs. close shop.
- Almost every company can have at least 1 product feature powered by ML.
- Be careful about bias in data.

Suman Deb Roy

suman@betaworks.com | @_roysd



data.betaworks.com