# Better Machine Learning Through Data

Saleema Amershi Machine Teaching Group Microsoft Research



August 14, 2016

#### Making better sense of data.

Better data makes better machine learning.





# Machine learning research often takes the data as given.

When Algorithms Discriminate – *The New York Times*, 2015

Big Data's all-too-human failings – *Reuters*, 2016

# Artificial Intelligence's White Guy Problem – *The New York Times, 2016*

#### Mapping Crime – Or Stirring Hate?– Financial Times, 2014

#### Making better sense of data.

Better data makes better machine learning.

Most influence practitioners have on machine learning is through data.



In research, data is often taken as given.



In practice, the In research, algorithm is often data is often taken as given. taken as given.



In practice, the algorithm is often taken as given. "Data scientists, according to interviews and expert estimates, spend **50 percent to 80 percent of their time** mired in this more mundane labor of collecting and preparing unruly digital data." – New York Times, 2014





[Patel et al., CHI 2008]





Iterations are driven by evaluating models on data.



In practice, most effort is spent crafting input data.

Iterations are driven by evaluating models on data.



#### Machine learning in theory



# Machine learning in practice









































Pre-defined high-level categories.

Does not support concept evolution (refining the target concept as data is observed).
#### How common is concept evolution?

Nine machine learning experts labeled the same 200 pages in two sessions (4 weeks apart).



Average consistency 81.7% (SD=6.8%)

6 out of 9 people's labels changed significantly (via Chi Square test of symmetry)

#### Proposed Solution – Structured Labeling

Enable people to **explicitly organize their concept** via *grouping* and *tagging* within a traditional labeling scheme.

# Traditional Labeling



Pre-defined high-level categories.







ls this a Cat?	Cat Definitely Cat
WANTED * DEAD AND ALIVE * SCHRÖDINGER'S CAT * \$10,000,000 CASH REWARD	Cat Poster







Grouping within high-level categories. User provided tags on groups aid recall. Can move, merge and split groups as desired.

# Assisted Structured Labeling



Grouping recommendations to improve label consistency.

# Assisted Structured Labeling



Grouping recommendations to improve label consistency.

Similar items to help users make decisions.

# Findings

People revised labels significantly more with structured labeling People labeled more consistently

People preferred it over traditional labeling



# Structured Labeling Summary

Current tools do not support **concept evolution**.

**Structured labeling** helps people refine their concepts by surfacing labeling decisions and aiding recall.

People used structured labeling when it was available and **labeled more consistently**.

**Structure contains additional information** (e.g., group related features, group related accuracy, decisions made...)



Structured labeling improves consistency [CHI 2014] eature Insight [VAST 2015] ModelTracker [CHI 2015, VAST 2016] "At the end of the day, some machine learning projects succeed and some fail. What makes the difference? **Easily the most important factor is the features used**."

...yet, little guidance or best practices exist.

#### How do people come up with features?

Look for features used in related domains.

Use intuition or domain knowledge.

Apply automated techniques

**Feature ideation** – Think of and experiment with custom features (a "black art").

### Proposed Solution – Feature Insight

Support compare and contrast of data.

#### What makes a cat a cat?





#### What makes a cat a cat?





### Proposed Solution – Feature Insight

Support **compare and contrast** of data. Comparing pairs vs sets?

### Comparing Pairs vs Sets

Sets may help people think of generalizable features.

#### Positive



#### Negative



#### Positives



VS



### Proposed Solution – Feature Insight

Support **compare and contrast** of data. Comparing pairs vs sets? Raw data vs visual summaries?

### Looking at Raw Data vs. Visual Summaries

Visual summaries may reveal relevant characteristics and hide irrelevant noise.

VS

#### Raw Data



#### **Visual Summary**

Positives Predicted as Negative		Tern	ns related	to: kitter	Negatives Predicted as Positive			
kittens	_	kittens	breed	and	cats	<b>_</b> c	artoon	
kitten	for	sale	are	have	cattery	📕 fi	ind	
breed	healthy	kitte	en in	formation	they	<b></b> o	n	
cattery	male	to	pedigree	e a	female	∎ b	,	
society	of	pet	the	cat	in	is	5	
persian	only	click	page	bree	ders	a	ny	
healthy	us	which	check	out		<b></b> h	iow	
male	quality	our	like	site	this	• ty	ypes	
🛓 under	their	about	regis	tered	very	∎ d	lesign	
<ul> <li>american</li> </ul>	11		L	hd	:.	<b>_</b> w	vhen	





Visual summaries led to *better* features Visual summaries preferred over looking at raw data Sets useful only in combination with visuals

# Feature Insight Summary

Featuring is arguably the most important step in machine learning, but there is little guidance on feature ideation.

Feature Insight supports error comparison, examination of sets, and visual summaries.

Visual summaries help people create better quality features.







Summary statistics hide important information about model behavior.



???

#### How do people evaluate performance? Collect & **Evaluate** Create Label Results Features \_ 5 Samples = M 🕷 🖾 📰 🖓 🔴 🦫 Visible: 50 of 50 Vi

Summary statistics hide important information about model behavior.



0.70

Predicted

Positive

143

35

0.71

???

Negative

72

190

# How do people evaluate performance?

Create

Features

Collect & Label – Samples

Summary statistics hide important information about model behavior.

Switching tools to examine data is disruptive and leads to a trial-anderror approach [Patel *et al.,* AAAI 2008].



# Example: Predicting Income Levels

x∎		o- @- <b>⊈</b> -	Ŧ				UCI_adı	ult_income - Exce	I						? 🖬 — 🛛	□ ×
F	LE	HOME INSER	T PAGE	E LAYOUT	FORMULAS	DATA RE	VIEW VIEW	LOAD TEST	TEAM					Sa	leema Amershi 👻 📿	
A1	A1 • $f_x$ age													~		
	Α	В	с	D	E	F	G	н		J	К	L	М	N	0	P 🔺
1	age	workclass	fnlwgt	education	educatior	marital-status	occupation	relationship	race	sex	capital-ga	capital-lo	hours-per	native-co	abel(IsOver50K)	
2	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	. 0	40	United-S	t O	
3	50	Self-emp-not-	83311	Bachelors	13	Married-civ-sp	Exec-manageri	Husband	White	Male	0	0	13	United-S	t O	
4	38	Private	215646	HS-grad	9	Divorced	Handlers-clean	Not-in-family	White	Male	0	0	40	United-S	t O	
5	53	Private	234721	11th	7	Married-civ-sp	Handlers-clean	Husband	Black	Male	0	0	40	United-S	t O	
6	28	Private	338409	Bachelors	13	Married-civ-sp	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	0	
7	37	Private	284582	Masters	14	Married-civ-sp	Exec-manageri	Wife	White	Female	0	0	40	United-S	t O	
8	49	Private	160187	9th	5	Married-spous	Other-service	Not-in-family	Black	Female	0	0	16	Jamaica	0	
9	52	Self-emp-not-	209642	HS-grad	9	Married-civ-sp	Exec-manageri	Husband	White	Male	0	0	45	United-S	t 1	
10	31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family	White	Female	14084	0	50	United-S	t 1	
11	42	Private	159449	Bachelors	13	Married-civ-sp	Exec-manageri	Husband	White	Male	5178	0	40	United-S	t 1	
12	37	Private	280464	Some-colleg	10	Married-civ-sp	Exec-manageri	Husband	Black	Male	0	0	80	United-S	t 1	
13	30	State-gov	141297	Bachelors	13	Married-civ-sp	Prof-specialty	Husband	Asian-Pac	Male	0	0	40	India	1	
14	23	Private	122272	Bachelors	13	Never-married	Adm-clerical	Own-child	White	Female	0	0	30	United-S	t 0	
15	32	Private	205019	Assoc-acdm	12	Never-married	Sales	Not-in-family	Black	Male	0	0	50	United-S	t 0	
16	40	Private	121772	Assoc-voc	11	Married-civ-sp	Craft-repair	Husband	Asian-Pac	Male	0	0	40	?	1	
17	34	Private	245487	7th-8th	4	Married-civ-sp	Transport-mov	Husband	Amer-Ind	Male	0	0	45	Mexico	0	
18	25	Self-emp-not-	176756	HS-grad	9	Never-married	Farming-fishin	Own-child	White	Male	0	0	35	United-S	t O	
19	32	Private	186824	HS-grad	9	Never-married	Machine-op-in	Unmarried	White	Male	0	0	40	United-S	t O	
20	38	Private	28887	11th	7	Married-civ-sp	Sales	Husband	White	Male	0	0	50	United-S	t O	
21	43	Self-emp-not-	292175	Masters	14	Divorced	Exec-manageri	Unmarried	White	Female	0	0	45	United-S	t 1	
22	40	Private	193524	Doctorate	16	Married-civ-sp	Prof-specialty	Husband	White	Male	0	0	60	United-S	t 1	
22	54	Drivato	202146	US arad	٩	Soparated	Other convice	Unmarried	Plack	Fomalo	n	0	20	United S	+ 0	
		UCI_adult	Income	(+)						•						

READY

COUNT: 15 🏢 🗉 🗕 — — 🕂 100%

#### Decision Tree 86% Accuracy



#### Support Vector Machine 85% Accuracy



#### Decision Tree 86% Accuracy

#### Support Vector Machine 85% Accuracy

☆ 〓

Help

fr

2

8

2

1


### ModelTracker Demo

# Significantly faster and more accurate performance analysis

ModelTracker												Ch ModelTracker x										🔺 You	×
[ ←	→ C 🔒	https://modeltracker	.azurewebsites.ne	et/modeltracker.	html						☆ =	< →	← → C A https://modeltracker.azurewebsites.net/modeltracker.html										<u>ح</u>
	File View	/									Help	File	File View										
														Predicted									
1	1.0 -		A	- <u></u>	_ <u>_</u>				A				C0	C1		C3	C4	C5		C7	C8	C9	
												CO	0	0	1	5	1	5	9	1	16	0	0.00
0	0.9 -											C1	0	0	1	4	0	5	1	3	4	0	0.00
												C2	4	12	0	11	5	0	16	12	11	3	0.00
	.8 -											C3	1	14	16	0	0	15	3	6	22	8	0.00
	.7-											<u>문</u> C4	1	2	1	2	0	1	7	2	2	43	0.00 8
												ਵੀ C5	5	17	1	34	12	0	11	2	8	10	0.00 🛱
0	0.6 -											C6	7	2	2	1	1	5	0	1	5	0	
Score												C7	6	14	9	1	13	0	0	0	3	18	0.00
ction o	).5 -											C8	1	34	5	17	5	8	14	4	0	18	0.00
Predi												C9	0	7	3	7	12	4	6	19	12	0	0.00
C	0.4 -												0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
													Precision										
												Dataset: mr	ist.randomfore	st.csv 10 Classes, 50	v 10 Classes, 5000 Instances, 640 Shown Acc.: 0.00, Prec.: 0.00, Recall: 0.00, TP: 0.0, FP: 64.0, TN: 9				2.0, FN: 64.0 0 Hovered			0 Hovered, 0 Se	elected init 🖽
0	0.2 -									2		*	Imag	je TRUE	Assigned	Correct 🕴	Prediction Score	C0	C1	C2	C3	C4	C5
												<b>Ω</b>	8	C8	C1	0	0.4238	0.0059	0.4238	0.1503	0.0533	0.0191	0.0949
0	0.1 -						1				El .	☆	2	C2	C1	0	0.3203	0.0067	0.3203	0.1903	0.2118	0.0199	0.0497
												11	9	C1	C7	0	0.4218	0.0433	0.0287	0.1562	0.1032	0.0510	0.0702
0	0.0 - R=0.92 P=	0.95 R=0.97 P=0.85	R=0.85 P=0.91	R=0.84 P=0.84	R=0.87 P=0.90	R=0.78 P=0.89	R=0.95 P=0.88	R=0.88 P=0.	90 R=0.77	P=0.81	R=0.86 P=0.81	й 2	L V	C2	C1	0	0.3120	0.0256	0.0538	0.1101	0.0267	0.1049	0.0655
	TN=4497 TP EN=38 EP	=440 TN=4321 TP=559 =25 FN=18 FP=102	TN=4473 TP=414 FN=74 FP=39	TN=4394 TP=439 EN=85 EP=82	TN=4467 TP=423 EN=61 EP=49	TN=4498 TP=359 EN=100 EP=43	TN=4425 TP=484 FN=24 FP=67	TN=4432 TP=4 EN=64 EP=4	54 TN=4454	TP=357 TI FP=83	N=4399 TP=431 EN=70 EP=100	☆	9	C9	C2	0	0.1860	0.1231	0.0396	0.1860	0.0245	0.0945	0.0658
	C0	C1	C2	C3	C4	C5	C6	C7	C	3	C9	\$	5	C5	C1	0	0.4871	0.0068	0.4871	0.0669	0.1239	0.0061	0.1210
Dataset: mnist.randomforest.csv 10 Classes, 5000 Instances, 5000 Shown Acc.: 0.87, Prc.: 0.87, Recal: 0.87, TP: 436.0, FP: 64.0, TN: 4436.0, FN: 64.0 0 Hovered, 0 Selected 🕍 🖽 🏢 🔳 🍭 🔍											☆	2	C2	C1	0	0.3875	0.0036	0.3875	0.3495	0.0322	0.0141	0.0297	
	•	Image TRUE	Assigned	Correct	Prediction Score	0	C1	C2	C3	C4	C5	\$	5	C5	C0	0	0.4213	0.4213	0.0078	0.0246	0.2064	0.0337	0.1534
	7	7 C7	C7	1	0.5849	0.0103	0.0226	0.0646	0.0330	0.0621	0.0279	<u>Å</u>	)	C7	C2	0	0.2413	0.0120	0.1797	0.2413	0.1168	0.0443	0.0381
The second se	ų.	/ C1	C1	1	0.3576	0.0111	0.3576	0.1180	0.0431	0.0450	0.0447	\$	1	C5	C1	0	0.5081	0.0239	0.5081	0.1009	0.0923	0.0069	0.1085
1	ł	<b>7</b> C7	C7	1	0.6777	0.0185	0.0310	0.0485	0.0324	0.0328	0.0491	<u>1</u>	8	C8	C1	0	0.3517	0.0070	0.3517	0.1076	0.0450	0.0449	0.0894
12	ž	1 C9	C9	1	0.2158	0.0146	0.0634	0.0233	0.1524	0.1435	0.1895	1	3	C3	C9	0	0.2530	0.0050	0.0582	0.0552	0.1815	0.1196	0.0768
7.	ł	5 C5	C5	1	0.2964	0.0346	0.2051	0.1039	0.1367	0.0286	0.2964	17 	5	C5	C9	0	0.3221	0.0345	0.0214	0.0421	0.0438	0.1390	0.2261
T.	ż	3 C3	C3	1	0.2365	0.0121	0.0946	0.0367	0.2365	0.0617	0.1952	<b>X</b>	4	C4	C9	0	0.2202	0.0179	0.0240	0.0720	0.0738	0.1659	0.1141
1	ł	Z C2	C2	1	0.3628	0.0069	0.1414	0.3628	0.1552	0.0153	0.0526	и "	3	03	07	0	0.2182	0.0073	0.0931	0.1455	0.0528	0.0809	0.0473
T.	7	/ C1	C1	1	0.8319	0.0017	0.8319	0.0342	0.0166	0.0087	0.0201	и "	3	03	09	0	0.2246	0.0072	0.0440	0.0875	0.1344	0.0392	0.0049
TA I	7	<b>2</b> C2	C2	1	0.7116	0.0192	0.0188	0.7116	0.0230	0.0390	0.0157	ж 	5	07	0	0	0.2348	0.2348	0.0276	0.0808	0.1714	0.0353	0.1034
T.	2	/ C1	C1	1	0.7070	0.0014	0.7070	0.0384	0.0223	0.0234	0.0203	ж Х	t	C7	03	0	0.1905	0.0033	0.1350	0.0272	0.1905	0.00/6	0.0715
n n	7	C7 C7	C7	1	0.7910	0.0081	0.0135	0.0240	0.0252	0.0417	0.0272	×	2	03	0	0	0.1047	0.0066	0.0427	0.1199	0.1000	0.0700	0.0555
•											•	4											•

**Common Confusion Matrix** 

#### ModelTracker

ModelTracker Summary

Current tools for performance analysis and debugging hide a lot of important information about model behavior.

ModelTracker supports estimating performance at multiple levels of granularity while enabling direct access to data.

People are significantly **faster and more accurate** at performance analysis with ModelTracker.





#### Many more opportunities to better support machine learning in practice.



Many more opportunities to better support machine learning in practice.



Many more opportunities to better support machine learning in practice and theory.

#### Making better sense of data.

Better data means better machine learning.

Most influence practitioners have on machine learning is through data.

Many more opportunities!

Thanks! Questions?

## Better Machine Learning Through Data

Saleema Amershi, <u>samershi@microsoft.com</u> Machine Teaching Group Microsoft Research



August 14, 2016