Machine Learning for Modern Artificial Intelligence

Hsuan-Tien Lin

National Taiwan University

December 19, 2019
invited talk for Wistron NeWeb Corporation

Outline

ML for (Modern) AI

ML Research for Modern AI

ML for Future AI

From Intelligence to Artificial Intelligence

intelligence: thinking and acting smartly

- humanly
- rationally

artificial intelligence: **computers** thinking and acting smartly

- humanly
- rationally

humanly ≈ **smartly** ≈ rationally

—are humans rational? :-)

Humanly versus Rationally

What if your self-driving car decides one death is better than two—and that one is you? (The Washington Post `http://wpo.st/ZK-51`)

```
You're humming along in your self-driving
car, chatting on your iPhone 37 while the
machine navigates on its own.  Then a swarm
of people appears in the street, right in
the path of the oncoming vehicle.
```

Car Acting Humanly

to save my (and passengers') life, stay on track

Car Acting Rationally

avoid the crowd and crash the owner for minimum total loss

which is smarter?

—depending on where I am, maybe? :-)

(Traditional) Artificial Intelligence

Thinking Humanly

- cognitive modeling —now closer to Psychology than AI

Thinking Rationally

- formal logic—now closer to Theoreticians than AI practitioners

Acting Humanly

- dialog systems
- humanoid robots
- computer vision

Acting Rationally

- recommendation systems
- cleaning robots
- cross-device ad placement

acting humanly or rationally:
more academia/industry attentions nowadays

Traditional vs. Modern [My] Definition of AI

Traditional Definition

humanly ≈ intelligently ≈ rationally

My Definition

intelligently ≈ easily

is your smart phone 'smart'? :-)

modern artificial intelligence
= **application** intelligence

Examples of Application Intelligence

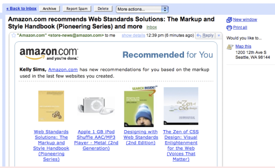
Siri

By Bernard Goldbach [CC BY 2.0]

iRobot

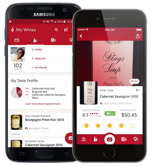
By Yuan-Chou Lo [CC BY-NC-ND 2.0]

Amazon Recommendations



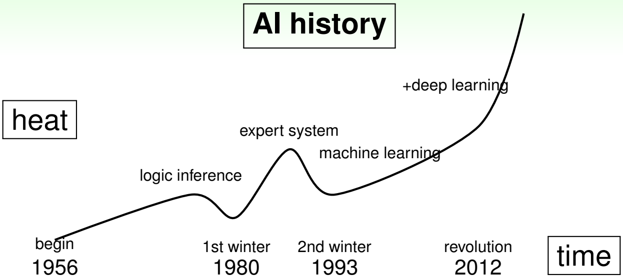
By Kelly Sims [CC BY 2.0]

Vivino



from `nordic.businessinsider.com`

AI Milestones



- first AI winter: AI cannot solve 'combinatorial explosion' problems
- second AI winter: expert system failed to scale

reason of winters: **expectation mismatch**

What's Different Now?

More Data
- cheaper storage
- Internet companies

Better Algorithms
- decades of research
- e.g. deep learning

Faster Computation
- cloud computing
- GPU computing

Healthier Mindset
- reasonable wishes
- key breakthroughs

data-enabled AI: mainstream nowadays

Machine Learning and AI

Easy-to-Use

Acting Humanly

Acting Rationally

Machine Learning

machine learning: core behind modern (data-driven) AI

ML Connects Big Data and AI

From Big Data to Artificial Intelligence

big data → ML → artificial intelligence

ingredient

→ tools/steps

→ dish

"cooking" needs many possible
tools & procedures

Bigger Data Towards Better AI

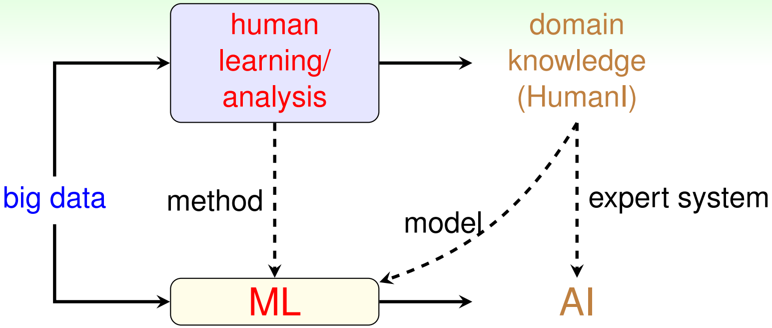
best route by shortest path

best route by current traffic

best route by predicted travel time

big data **can** make machine look smarter

ML for Modern AI



- human sometimes **faster learner** on **initial (smaller) data**
- industry: **black plum is as sweet as white**

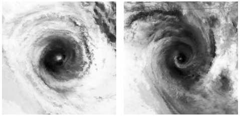
often important to leverage human learning, especially **in the beginning**

Application: Tropical Cyclone Intensity Estimation

meteorologists can ‘feel’ & estimate TC intensity from image

human learning/ analysis

domain knowledge (HumanI)



TC images

CNN

polar rotation invariance

current weather system

ML

intensity estimation

better than current system & ‘**trial-ready**’

(Chen et al., KDD 2018)

(Chen et al., Weather & Forecasting 2019)

Outline

ML for (Modern) AI

ML Research for Modern AI

ML for Future AI

Cost-Sensitive Multiclass Classification

What is the Status of the Patient?

?

H7N9-infected

cold-infected

healthy

- a **classification** problem
  —grouping 'patients' into different 'status'

are all mis-prediction costs equal?

Patient Status Prediction

error measure = society cost

actual \ predicted	H7N9	cold	healthy
H7N9	0	1000	**100000**
cold	100	0	3000
healthy	100	30	0

- H7N9 mis-predicted as healthy: **very high cost**
- cold mis-predicted as healthy: high cost
- cold correctly predicted as cold: no cost

human doctors consider costs of decision;
how about computer-aided diagnosis?

Our Works

	binary	multiclass
regular	well-studied	well-studied
cost-sensitive	known (Zadrozny et al., 2003)	ongoing (our works, among others)

selected works of ours

- cost-sensitive SVM (Tu and Lin, ICML 2010)
- cost-sensitive one-versus-one (Lin, ACML 2014)
- cost-sensitive deep learning (Chung et al., IJCAI 2016)

why are people **not**
using those **cool ML works for their AI? :-)**

Issue 1: Where Do Costs Come From?

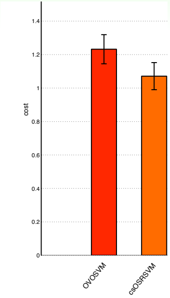
A Real Medical Application: Classifying Bacteria

- by human doctors: **different treatments** $\Longleftrightarrow$ serious costs
- cost matrix averaged from two doctors:

	Ab	Ecoli	HI	KP	LM	Nm	Psa	Spn	Sa	GBS
Ab	0	1	10	7	9	9	5	8	9	1
Ecoli	3	0	10	8	10	10	5	10	10	2
HI	10	10	0	3	2	2	10	1	2	10
KP	7	7	3	0	4	4	6	3	3	8
LM	8	8	2	4	0	5	8	2	1	8
Nm	3	10	9	8	6	0	8	3	6	7
Psa	7	8	10	9	9	7	0	8	9	5
Spn	6	10	7	7	4	4	9	0	4	7
Sa	7	10	6	5	1	3	9	2	0	7
GBS	2	5	10	9	8	6	5	6	8	0

issue 2: is cost-sensitive classification
really useful?

Cost-Sensitive vs. Traditional on Bacteria Data

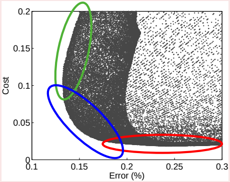


(Jan et al., BIBM 2011)

cost-sensitive better than **traditional**;
but why are people **still not**
using those cool ML works for their AI? :-)

Issue 3: Error Rate of Cost-Sensitive Classifiers

The Problem



- cost-sensitive classifier: low cost but high error rate
- traditional classifier: low error rate but high cost
- how can we get the blue classifiers?: low error rate and low cost

cost-and-error-sensitive:
more suitable for **real-world medical needs**

Improved Classifier for Both Cost and Error

(Jan et al., KDD 2012)

Cost

iris	$\approx$
wine	$\approx$
glass	$\approx$
vehicle	$\approx$
vowel	◯
segment	◯
dna	◯
satimage	$\approx$
usps	◯
zoo	◯
splice	$\approx$
ecoli	$\approx$
soybean	$\approx$

Error

iris	◯
wine	◯
glass	◯
vehicle	◯
vowel	◯
segment	◯
dna	◯
satimage	◯
usps	◯
zoo	◯
splice	◯
ecoli	◯
soybean	◯

now, **are people using those cool ML works for their AI? :-)**

Lessons Learned from Research on Cost-Sensitive Multiclass Classification

?

H7N9-infected

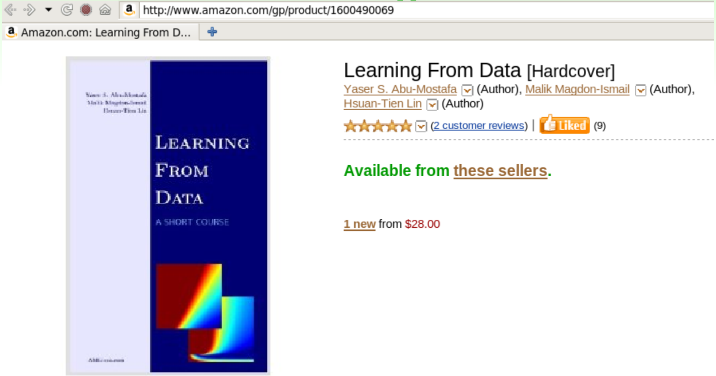
cold-infected

healthy

1. more realistic (generic) in academia
   $\neq$ **more realistic (feasible) in application**
   e.g. the 'cost' of **inputting a cost matrix? :-)**
2. **cross-domain collaboration** important
   e.g. getting the 'cost matrix' from **domain experts**
3. not easy to win **human trust**
   —humans are somewhat **multi-objective**

Label Space Coding for Multilabel Classification

What **Tags**?



?: {machine learning, ~~data structure~~, data mining, ~~object oriented programming~~, artificial intelligence, ~~compiler~~, ~~architecture~~, ~~chemistry~~, textbook, ~~children book~~, ~~... etc.~~ }

a **multilabel** classification problem:
tagging input to multiple categories

Binary Relevance: Multilabel Classification via Yes/No

Binary Classification

{yes, no}

multilabel w/ L classes: L Y/N questions

machine learning (Y), data structure (N), data mining (Y), OOP (N), AI (Y), compiler (N), architecture (N), chemistry (N), textbook (Y), children book (N), *etc.*

- **Binary Relevance** approach:
  transformation to **multiple isolated binary classification**
- disadvantages:
  - **isolation**—hidden relations not exploited (e.g. ML and DM highly correlated, ML subset of AI, textbook & children book disjoint)
  - **unbalanced**—few yes, many no

Binary Relevance: simple (& good) benchmark with known disadvantages

From Label-set to Coding View

	label set	apple	orange	strawberry	**binary code**
	$\{o\}$	0 (N)	1 (Y)	0 (N)	$[0, 1, 0]$
	$\{a, o\}$	1 (Y)	1 (Y)	0 (N)	$[1, 1, 0]$
	$\{a, s\}$	1 (Y)	0 (N)	1 (Y)	$[1, 0, 1]$
	$\{o\}$	0 (N)	1 (Y)	0 (N)	$[0, 1, 0]$
	$\{\}$	0 (N)	0 (N)	0 (N)	$[0, 0, 0]$

subset of $2^{\{1,2,\cdots,L\}}$ $\Leftrightarrow$ **length-L binary code**

A NeurIPS 2009 Approach: Compressive Sensing

General Compressive Sensing

sparse (many 0) binary vectors $\mathbf{y} \in \{0, 1\}^L$ can be **robustly compressed** by projecting to $M \ll L$ basis vectors $\{\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_M\}$

Comp. Sensing for Multilabel Classification (Hsu et al., NeurIPS 2009)

1. **compress**: encode original data by **compressive sensing**
2. **learn**: get regression function from compressed data
3. **decode**: decode regression predictions to sparse vector by **compressive sensing**

Compressive Sensing:
seemly strong competitor
from related theoretical analysis

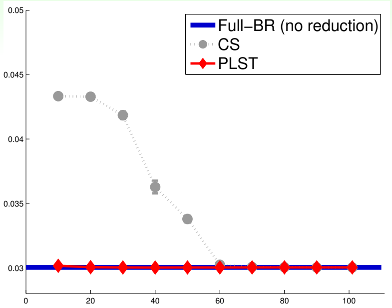
Our Proposed Approach: Compressive Sensing ⇒ PCA

Principal Label Space Transformation (PLST), i.e. PCA for Multilabel Classification (Tai and Lin, NC Journal 2012)

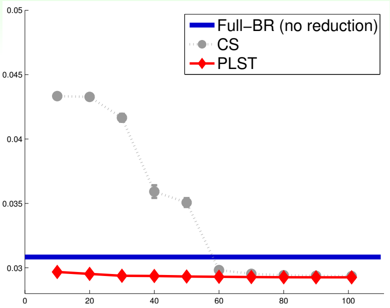
1. **compress**: encode original data by **PCA**
2. **learn**: get regression function from compressed data
3. **decode**: decode regression predictions to label vector by **reverse PCA + quantization**

does PLST perform better than CS?

Hamming Loss Comparison: PLST vs. CS



mediamill (Linear Regression)



mediamill (Decision Tree)

- **PLST** better than CS: faster, **better** performance
- similar findings across **data sets** and **regression algorithms**

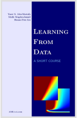
Why? CS creates
harder-to-learn regression tasks

Our Works Continued from PLST

1. **Compression** Coding (Tai & Lin, NC Journal 2012 with 249 citations)
   —**condense** for efficiency: better (than CS) approach PLST
   — key tool: PCA from Statistics/Signal Processing
2. **Learnable-Compression** Coding (Chen & Lin, NeuIPS 2012 with 186 citations)
   —**condense learnably** for **better** efficiency: better (than PLST) approach CPLST
   — key tool: Ridge Regression from Statistics (+ PCA)
3. **Cost-Sensitive** Coding (Huang & Lin, ECML Journal Track 2017)
   —**condense cost-sensitively** towards application needs: better (than CPLST) approach CLEMS
   — key tool: Multidimensional Scaling from Statistics

cannot thank **statisticans**
enough for those tools!

Lessons Learned from Label Space Coding for Multilabel Classification



?: {machine learning, ~~data structure~~, data mining, ~~object oriented programming~~, artificial intelligence, ~~compiler~~, ~~architecture~~, ~~chemistry~~, textbook, ~~children book~~, ~~. . . etc.~~ }

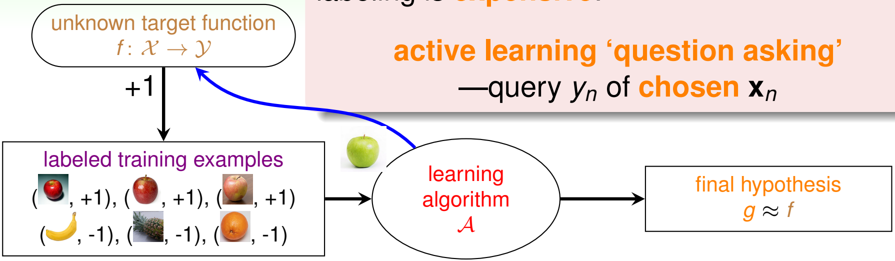
1. Is Statistics the same as ML? Is Statistics the same as AI?
   - **does it really matter?**
   - Modern AI should embrace **every useful tool from other fields**.
2. good tools **not necessarily most sophisticated tools**
   e.g. PCA possibly more useful than CS
3. more-cited paper $\neq$ more-useful AI solution
   —citation count **not the only impact measure**

Active Learning by Learning

Active Learning: Learning by ‘Asking’

labeling is **expensive**:

active learning ‘question asking’
—query y_n of **chosen** $\mathbf{x}_n$



active: improve hypothesis with fewer labels
(hopefully) by asking questions **strategically**

Pool-Based Active Learning Problem

Given

- labeled pool $\mathcal{D}_l = \left\{ (\text{feature } \mathbf{x}_n \right.$  $\left., \textbf{label } y_n \text{ (e.g. IsApple?)}) \right\}_{n=1}^{N}$
- unlabeled pool $\mathcal{D}_u = \left\{ \tilde{\mathbf{x}}_s \right\}_{s=1}^{S}$

Goal

design an algorithm that iteratively

1. **strategically query** some $\tilde{\mathbf{x}}_s$  to get associated $\tilde{y}_s$
2. move $(\tilde{\mathbf{x}}_s, \tilde{y}_s)$ from $\mathcal{D}_u$ to $\mathcal{D}_l$
3. learn **classifier** $g^{(t)}$ from $\mathcal{D}_l$

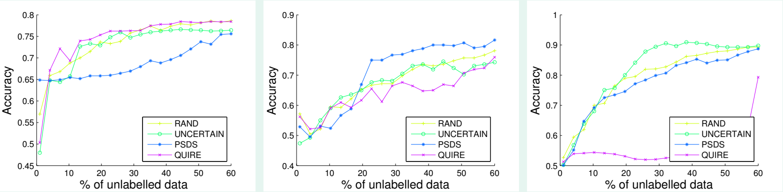
and improve **test accuracy of** $g^{(t)}$ w.r.t **#queries**

how to **query strategically**?

How to Query Strategically?

Strategy 1	Strategy 2	Strategy 3
ask **most confused** question	ask **most frequent** question	ask **most debateful** question

- **choosing** one single strategy is **non-trivial**:



application intelligence: how to **choose strategy smartly**?

Idea: Trial-and-Reward Like Human

when do humans **trial**-and-**reward**?
gambling

K strategies: $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_K$

try one strategy

"goodness" of strategy as **reward**

K bandit machines: $\mathcal{B}_1, \mathcal{B}_2, \cdots, \mathcal{B}_K$

try one bandit machine

"luckiness" of machine as **reward**

intelligent choice of strategy
$\Longrightarrow$ intelligent choice of **bandit machine**

Active Learning by Learning (Hsu and Lin, AAAI 2015)

K strategies:
$\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_K$

try one strategy

"goodness" of strategy as **reward**

Given: K existing active learning strategies

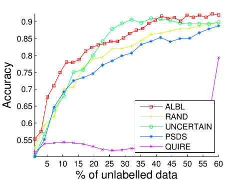
for $t = 1, 2, \ldots, T$

1. let some bandit model **decide strategy $\mathcal{A}_k$ to try**
2. **query the $\tilde{\mathbf{x}}_s$** suggested by $\mathcal{A}_k$, and compute $g^{(t)}$
3. evaluate **goodness of $g^{(t)}$** as **reward** of **trial** to update model

proposed Active Learning by Learning (ALBL):
motivated but unrigorous reward design

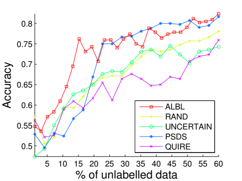
Comparison with Single Strategies

UNCERTAIN Best



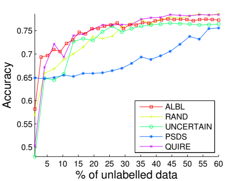
vehicle

PSDS Best



sonar

QUIRE Best



diabetes

- **no single best strategy** for every data set
  —choosing needed
- proposed **ALBL** consistently **matches the best**
  —similar findings across other data sets

'application intelligence' outcome:
open-source tool released
(`https://github.com/ntucllab/libact`)

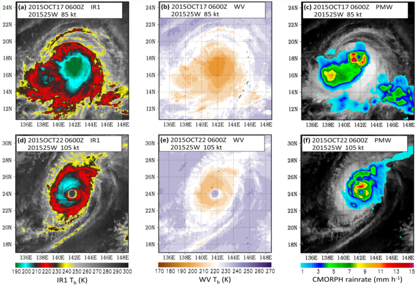
Lessons Learned from Research on Active Learning by Learning

by DFID - UK Department for International Development; licensed under CC BY-SA 2.0 via Wikimedia Commons

1. **scalability bottleneck** of ‘application intelligence’: **choice** of methods/models/parameter/. . .
2. think outside of the **math** box: ‘unrigorous’ usage may be **good enough**
3. important to be **brave** yet **patient**
   - **idea: 2012**
   - paper: (Hsu and Lin, AAAI 2015); software: (Yang et al., 2017)

Tropical Cyclone Intensity Estimation

Experienced Meteorologists Can ‘Feel’ and Estimate Tropical Cyclone Intensity from Image



Can ML do the same/better?

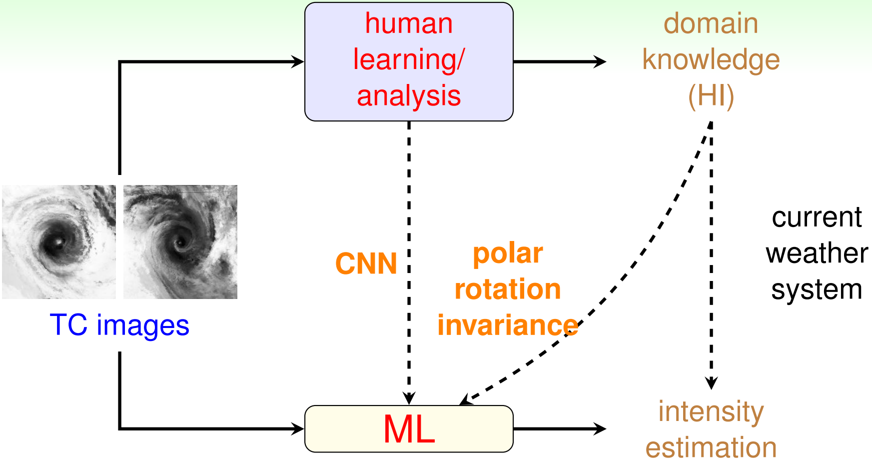
- lack of **ML-ready datasets**
- lack of **model that properly utilizes domain knowledge**

issues addressed in our latest works

(Chen et al., KDD 2018)

(Chen et al., Weather & Forecasting 2019)

Recall: Flow behind Our Proposed Model



is proposed **CNN-TC** better than current weather system?

Results

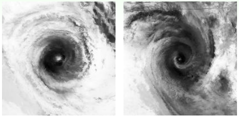
RMS Error

ADT	11.75
AMSU	14.40
SATCON	9.66
CNN-TC	**9.03**

CNN-TC much better than current weather system (SATCON)

why are people **not**
using this **cool ML model? :-)**

Lessons Learned from Research on Tropical Cyclone Intensity Estimation



1. again, **cross-domain collaboration** important
   e.g. even from 'organizing data' to be ML-ready
2. not easy to claim **production ready**
   —can ML be used for '**unseenly-strong** TC'?
3. good AI system requires **both human and machine learning**
   —still an 'art' to blend the two

Outline

ML for (Modern) AI

ML Research for Modern AI

ML for Future AI

AI: Now and Next

2010–2015

AI becomes **promising**, e.g.

- initial success of **deep learning** on ImageNet
- mature tools for SVM (**LIBSVM**) and others

2016–2020

AI becomes **competitive**, e.g.

- super-human performance of **alphaGo** and others
- all big technology companies become **AI-first**

2021–

AI becomes **necessary**

- "You'll not be replaced by AI, but **by humans who know how to use AI**"

  (Sun, Chief AI Scientist of Appier, 2018)

Needs of ML for Future AI

more creative

win human **respect**

e.g. Appier's 2018 work on **design matching clothes**

(Shih et al., AAAI 2018)

more explainable

win human **trust**

e.g. my students' work on **automatic bridge bidding**

(Yeh et al., IEE ToG 2018)

more interactive

win human **heart**

e.g. my student's work (w/ DeepQ) on **efficient disease diagonsis**

(Peng et al., NeurIPS 2018)

Summary

- ML for (Modern) AI:
  tools + human knowledge $\Rightarrow$ **easy-to-use application**
- ML Research for Modern AI:
  need to be **more open-minded**
  —in methodology, in collaboration, in KPI
- ML for Future AI:
  crucial to be '**human-centric**'

Thank you! Questions?