Machine Learning for Modern Artificial Intelligence

Hsuan-Tien Lin
林軒田

Dept. of Computer Science and Information Enginnering,
National Taiwan University
國立臺灣大學資訊工程學系

December 17, 2020
Keynote talk for International Computer Symposium 2020 &
教育部人工智慧技術及應用人才培育計畫成果發表會

About Me

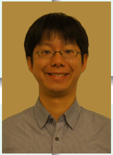
Professor

National Taiwan University

Chief Data Science Consultant

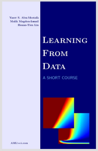
(former Chief Data Scientist)

Appier Inc.



Co-author

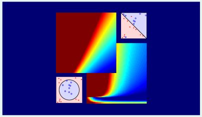
Learning from Data

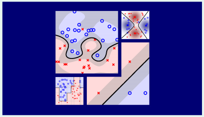


Instructor

NTU-Coursera MOOCs

ML Foundations/Techniques





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 $\approx$ **smartly** $\approx$ 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 $\approx$ intelligently $\approx$ rationally

My Definition

intelligently $\approx$ easily

is your smart phone 'smart'? :-)

modern artificial intelligence
$=$ **application** intelligence

Examples of Application Intelligence

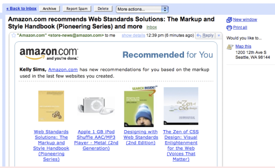
Siri

By Bernard Goldbach [CC BY 2.0]

iRobot

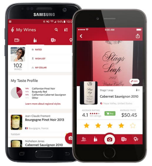
By Yuan-Chou Lo [CC BY-NC-ND 2.0]

Amazon Recommendations



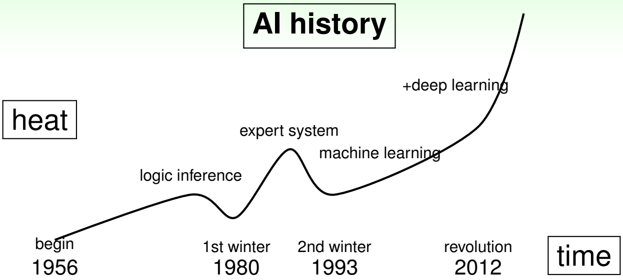
By Kelly Sims [CC BY 2.0]

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

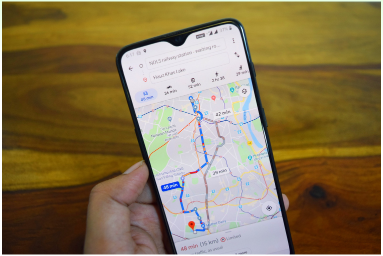
Better Algorithms
- decades of research
- e.g. deep learning

Faster Computation
- cloud computing
- GPU computing

Healthier Mindset
- reasonable wishes
- key breakthroughs

data-enabled AI: mainstream nowadays

Bigger Data Towards Easier-to-use AI



By deepanker70 on `https://pixabay.com/`

past

best route by shortest path

present

best route by current traffic

future

best route by predicted travel time

big data **can** make machine look smarter

Machine Learning Connects Big Data and AI

From Big Data to Artificial Intelligence

big data → **ML** → artificial intelligence

ingredient

→

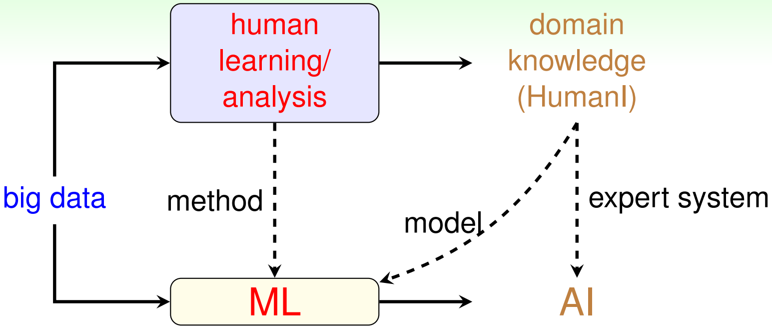
tools/steps

→

dish

"cooking" needs many possible
tools & procedures

ML for Modern AI



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

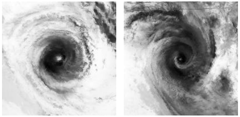
often important to leverage human learning, especially **in the beginning**

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 '18; Chen et al., Weather & Forecasting '19)

Outline

Cost-Sensitive Multiclass Classification

What is the Status of the Patient?

By DataBase Center for Life Science;
licensed under CC BY 4.0 via Wikimedia Commons

?

COVID19

cold

healthy

Pictures Licensed under CC BY-SA 3.0 from 1RadicalOne on Wikimedia Commons

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

are all mis-prediction costs equal?

Patient Status Prediction

error measure = society cost

actual \ predicted	COVID19	cold	healthy
COVID19	0	1000	**100000**
cold	100	0	3000
healthy	100	30	0

- COVID19 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?

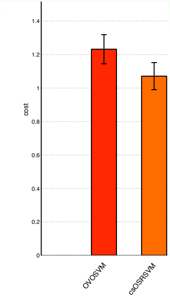
A Real Medical Application: Classifying Bacteria

- 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

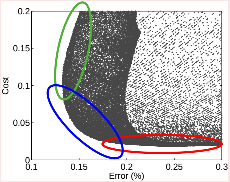


(Jan et al., BIBM 2011)

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

See Page 16 of the Slides for Sources of the Pictures

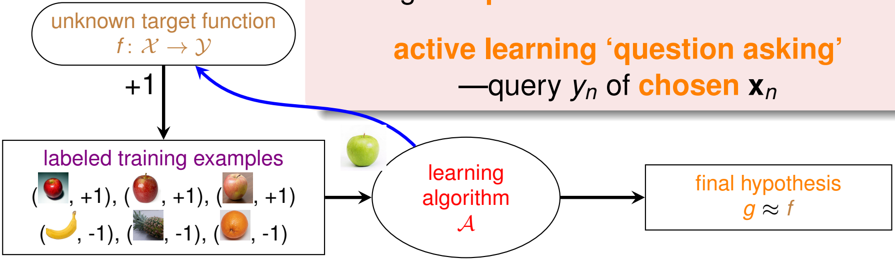
1. more realistic (generic) in academia
   $\neq$ **more realistic (feasible) in application**
   e.g. the 'cost' of **inputting a cost matrix? :-)**
2. **cross-domain collaboration** important
   e.g. getting the 'cost matrix' from **domain experts**
3. not easy to win **human trust**
   —humans are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

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

ask **most confused** question

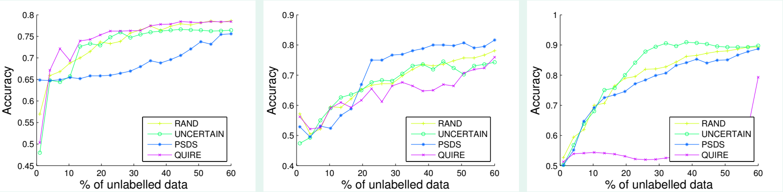
Strategy 2

ask **most frequent** question

Strategy 3

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

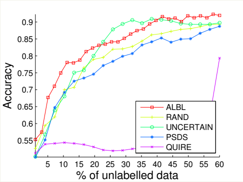
for $t = 1, 2, \ldots, T$

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

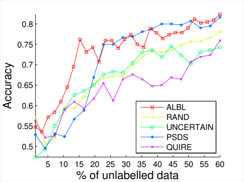
Comparison with Single Strategies

UNCERTAIN Best



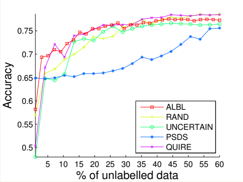
vehicle

PSDS Best



sonar

QUIRE Best

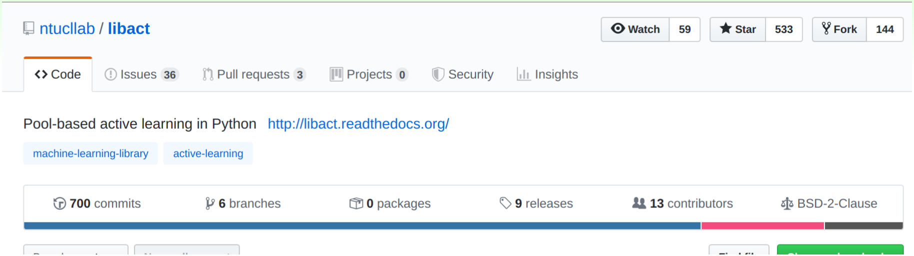


diabetes

- **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`)

Have We Made Active Learning More Realistic? (1/2)



Yes!

open-source tool `libact` developed (Yang, 2017)

`https://github.com/ntucllab/libact`

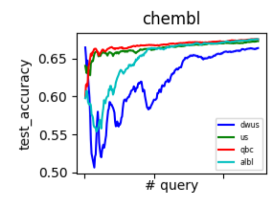
- including uncertainty, QUIRE, PSDS, . . ., **and ALBL**
- received > 500 **stars** and continuous **issues**

"`libact` is a Python package designed to **make active learning easier** for real-world users"

Have We Made Active Learning More Realistic? (2/2)

No!

- single-most raised **issue**: hard to install on Windows/Mac —because several strategies requires some C packages
- performance in a recent industry project:



  - **uncertainty** sampling **often suffices**
  - ALBL **dragged down by bad strategy**

"`libact` is a Python package **designed to make active learning easier** for real-world users"

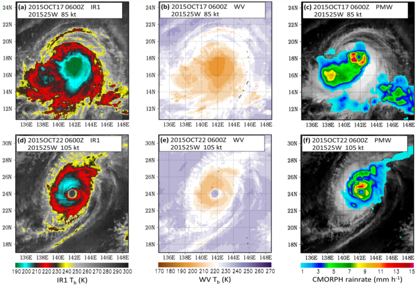
Lessons Learned from Research on Active Learning by Learning

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)
4. easy-to-use in design $\neq$ **easy-to-use in reality**

Tropical Cyclone Intensity Estimation

Experienced Meteorologists Can 'Feel' and Estimate Tropical Cyclone Intensity from Image



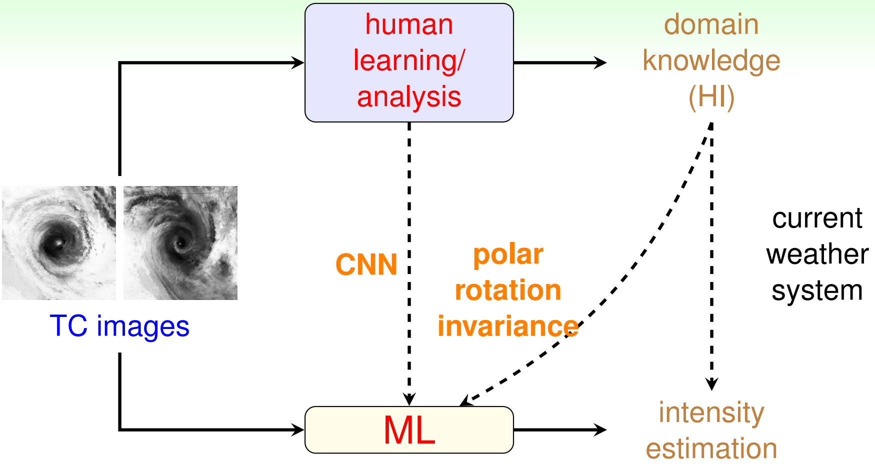
Can ML do the same/better?

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

issues addressed in our latest works

(Chen et al., KDD '18; Chen et al., Weather & Forecasting '19)

Recall: Flow behind Our Proposed Model



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

Results

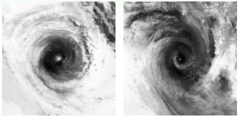
RMS Error

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 $|$

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

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

2016–2020: AI $+$

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

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

2021–: AI $\times$

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?