Machine Learning for Modern Artificial Intelligence

Hsuan-Tien Lin

National Taiwan University

June 25, 2019
other versions presented in Academia Sinica
& TWSIAM Annual Meeting

Outline

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? :-)

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

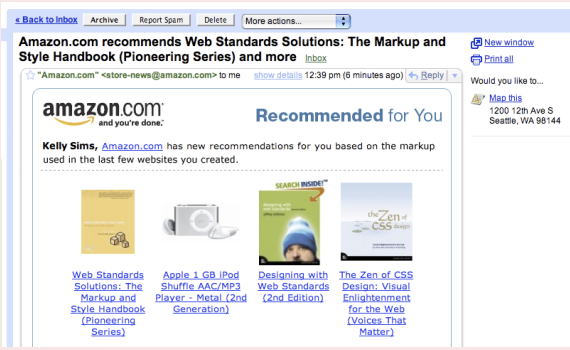
Siri

By Bernard Goldbach [CC BY 2.0]

iRobot

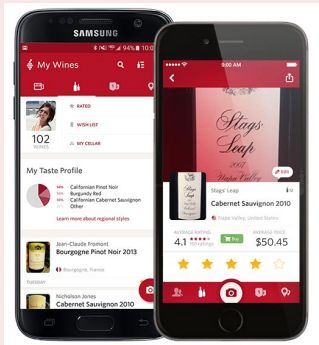
By Yuan-Chou Lo [CC BY-NC-ND 2.0]

Amazon Recommendations



By Kelly Sims [CC BY 2.0]

Vivino



from `nordic.businessinsider.com`

Machine Learning and AI

Easy-to-Use

Acting Humanly

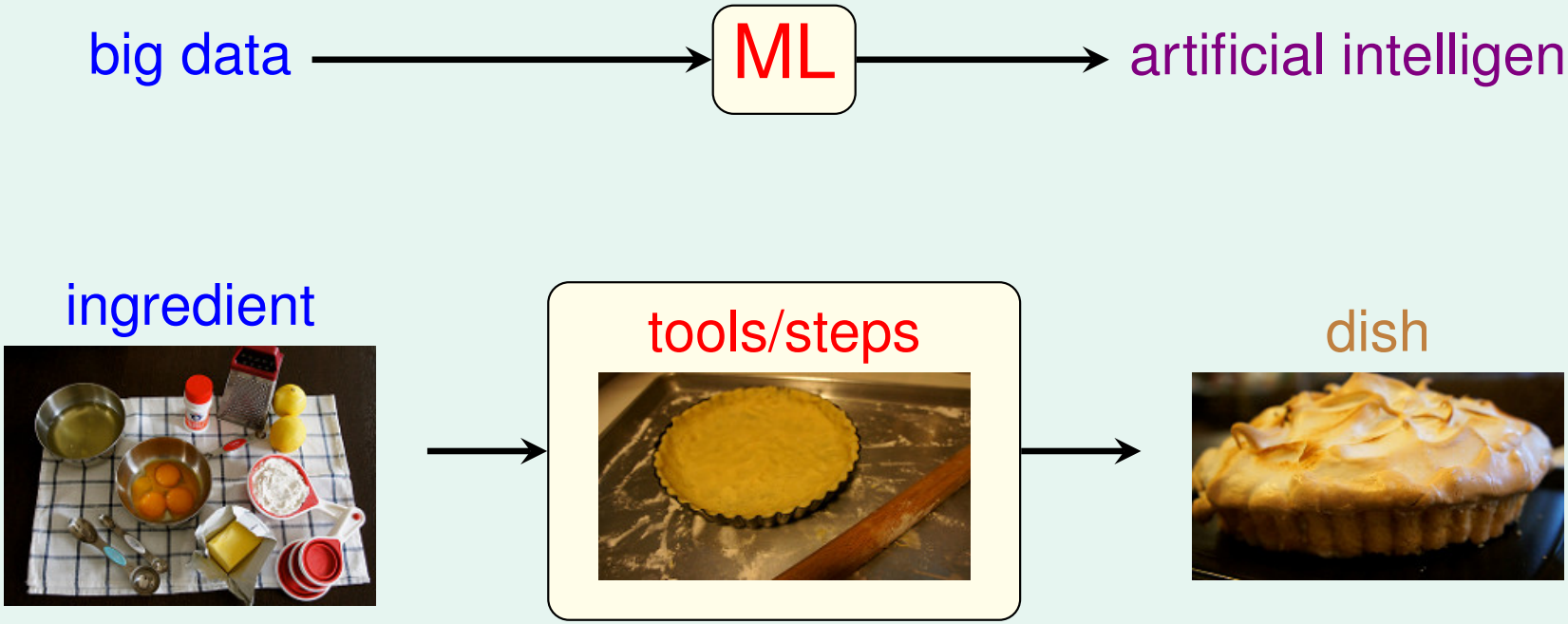
Acting Rationally

Machine Learning

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

ML Connects Big Data and AI

From Big Data to Artificial Intelligence



(Photos Licensed under CC BY 2.0 from Andrea Goh on Flickr)

"cooking" needs many possible
tools & procedures

Bigger Data Towards Better AI

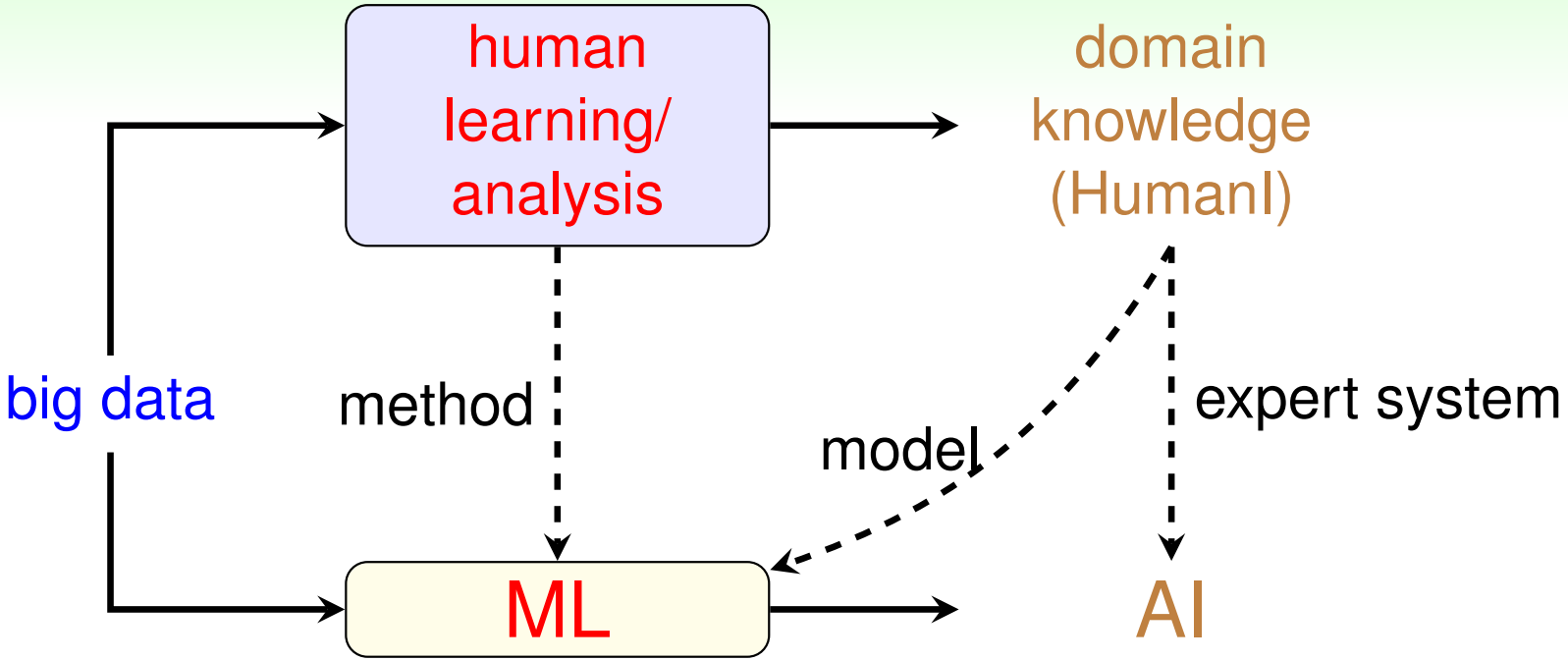
best route by shortest path

best route by current traffic

best route by predicted travel time

big data **can** make machine look smarter

ML for Modern AI

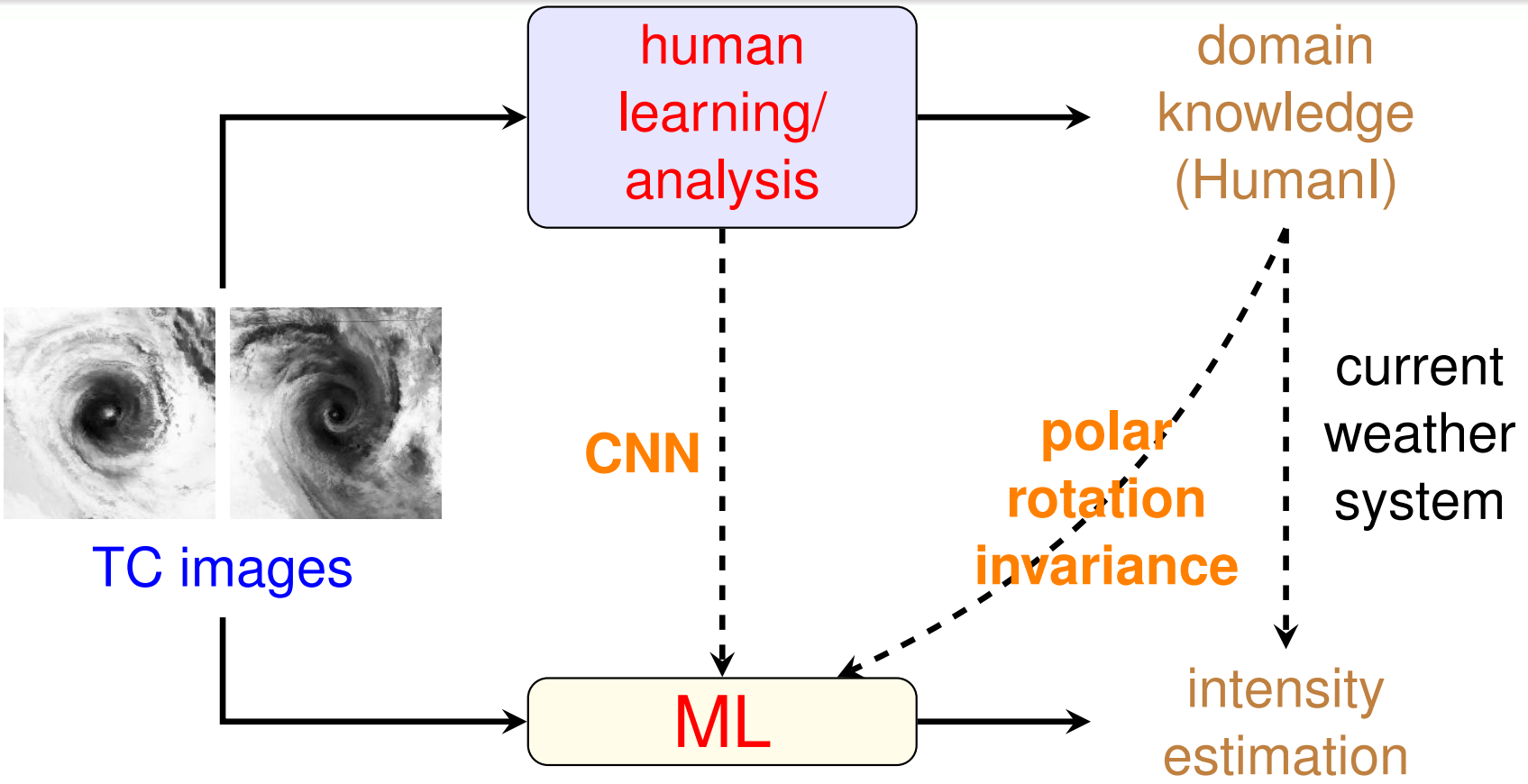


- human sometimes **faster learner** on **initial (smaller) data**
- industry: **black plum is as sweet as white**

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

Application: Tropical Cyclone Intensity Estimation

meteorologists can 'feel' & estimate TC intensity from image



better than current system & '**trial-ready**'

(Chen et al., KDD 2018)

(Chen et al., Weather & Forecasting 2019)

Outline

ML for (Modern) AI

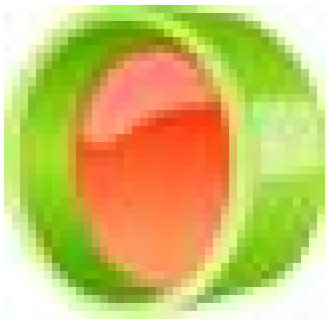
ML Research for Modern AI

ML for Future AI

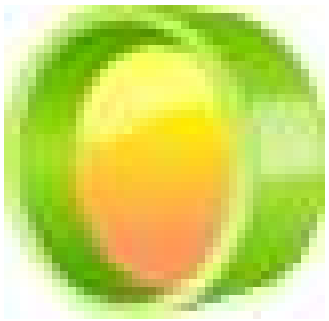
Cost-Sensitive Multiclass Classification

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?

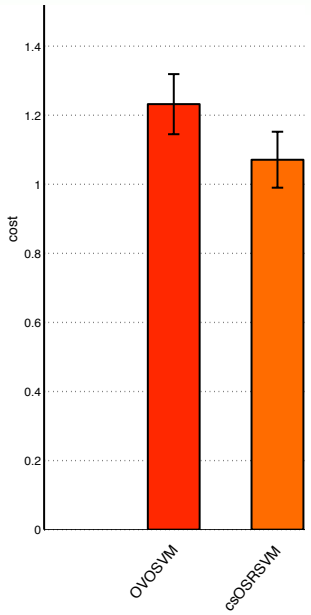
A Real Medical Application: Classifying Bacteria

- 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

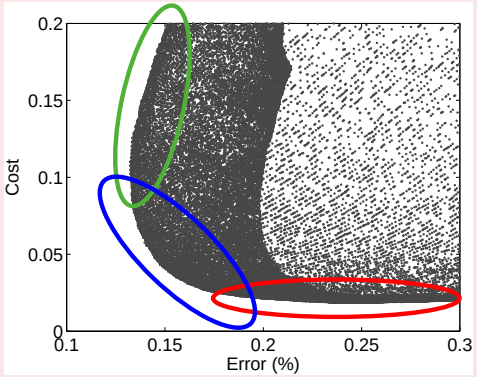


(Jan et al., BIBM 2011)

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

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

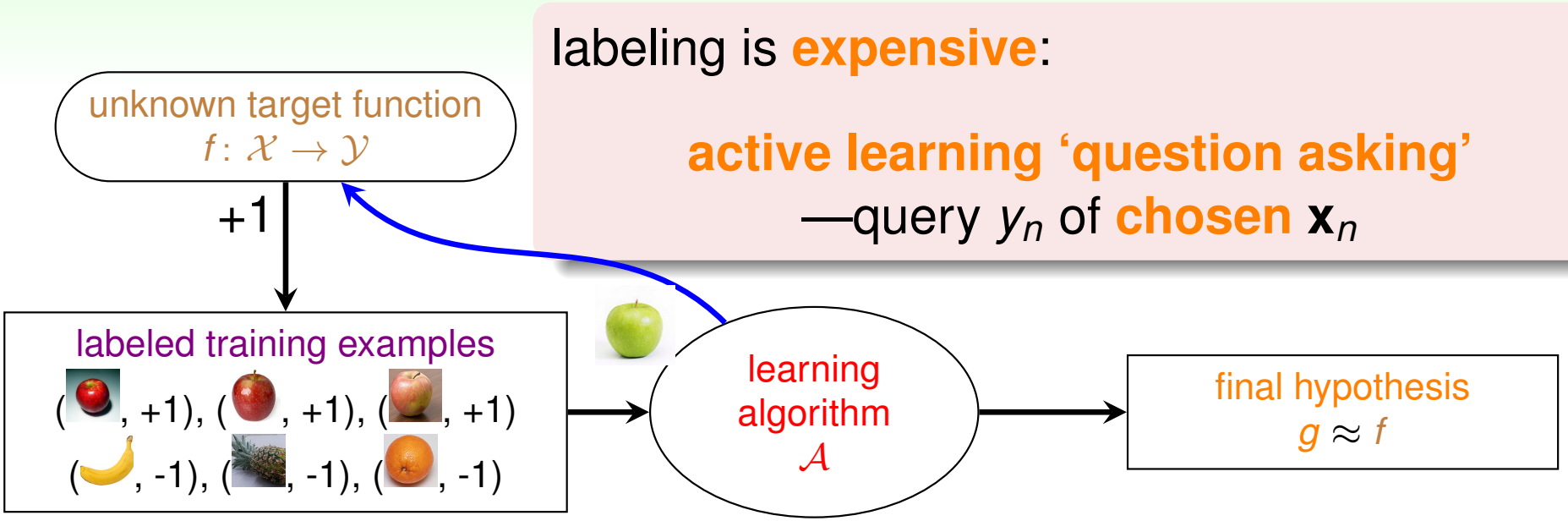
cold-infected

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**

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

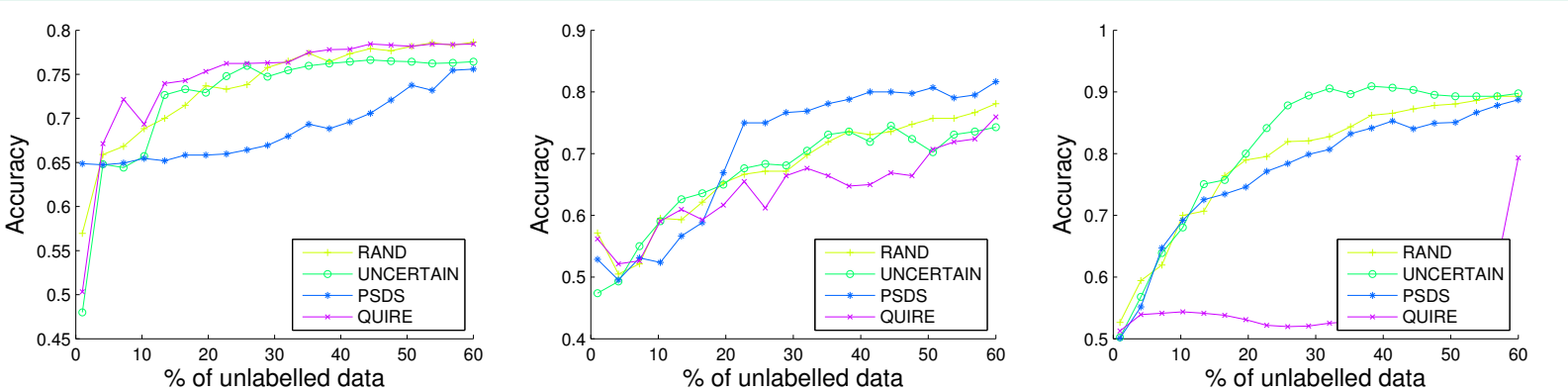
and improve **test accuracy of** $g^{(t)}$ w.r.t **#queries**

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

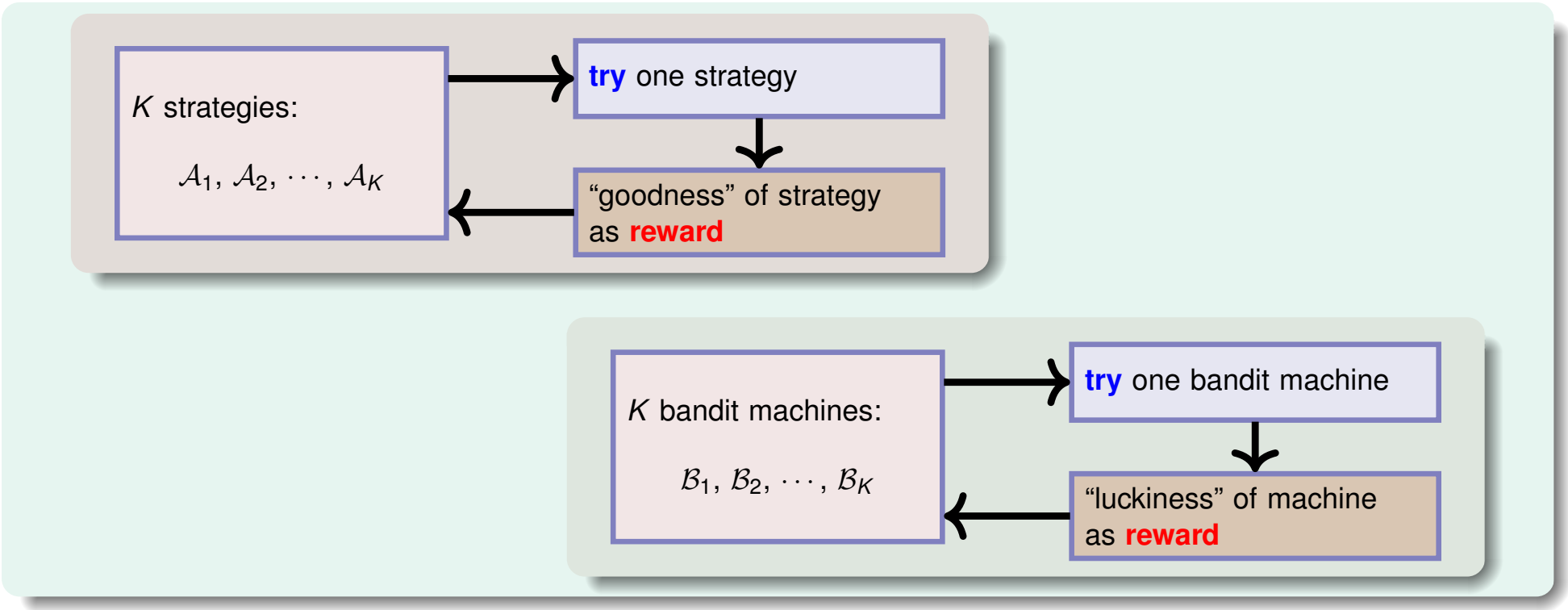
- **choosing** one single strategy is **non-trivial**:



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

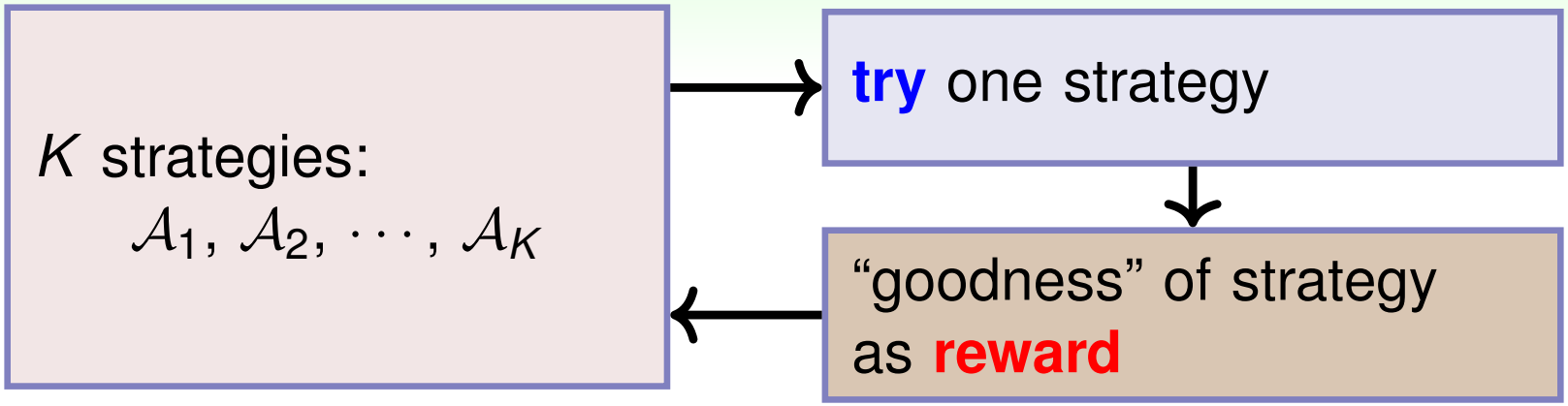
Idea: Trial-and-Reward Like Human

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



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

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



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

only remaining problem: **what reward**?

Design of Reward

ideal reward after updating classifier $g^{(t)}$ by the query $(\mathbf{x}_{n_t}, y_{n_t})$:

accuracy of $g^{(t)}$ on **test set** $\{(\mathbf{x}'_m, y'_m)\}_{m=1}^{M}$

—test accuracy **infeasible** in practice because labeling **expensive**

more feasible reward: training accuracy on the fly

accuracy of $g^{(t)}$ on **labeled pool** $\{(\mathbf{x}_{n_\tau}, y_{n_\tau})\}_{\tau=1}^{t}$

—but **biased** towards **easier** queries

weighted training accuracy as a better reward:

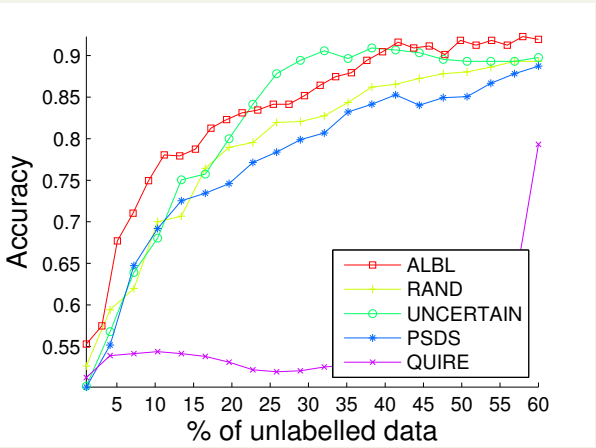
acc. of $g^{(t)}$ on inv.-prob. weighted **labeled pool** $\left\{(\mathbf{x}_{n_\tau}, y_{n_\tau}, \frac{1}{p_\tau})\right\}_{\tau=1}^{t}$

—'bias correction' from querying probability within bandit model

Active Learning by Learning (ALBL):
bandit + **weighted training acc. as reward**

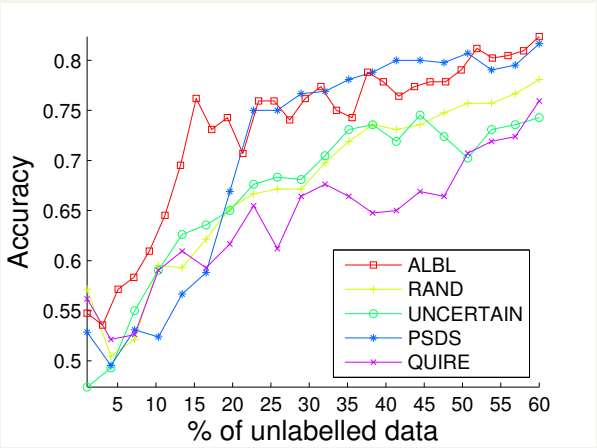
Comparison with Single Strategies

UNCERTAIN Best



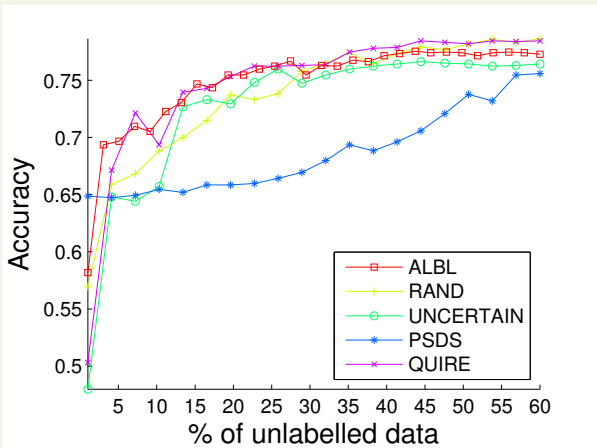
`vehicle`

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

ALBL: effective in **making intelligent choices**

Discussion for Statisticians

weighted training accuracy $\frac{1}{t}\sum_{\tau=1}^{t}\frac{1}{p_\tau}\left[\!\left[y_{n_\tau} = g^{(t)}(\mathbf{x}_{n_\tau})\right]\!\right]$ as reward

- is reward unbiased estimator of test performance?
  no for learned $g^{(t)}$ (yes for fixed g)
- is reward fixed before playing?
  no because $g^{(t)}$ learned from $(\mathbf{x}_{n_t}, y_{n_t})$
- is reward independent of each other?
  no because past history all in reward

—ALBL: tools from statistics + **wild/unintended usage**

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

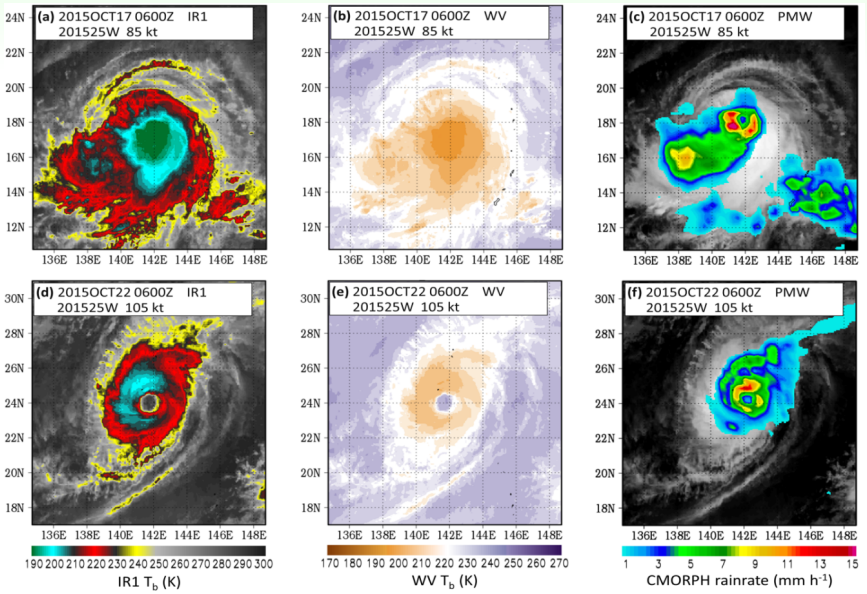
Lessons Learned from Research on Active Learning by Learning

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

1. Is Statistics the same as ML or AI?
   - **does it really matter?**
   - Modern AI should embrace **every useful tool from other fields**.
2. **scalability bottleneck** of 'application intelligence': **choice** of methods/models/parameter/. . .
3. think outside of the **math** box: 'unintended' usage may be **good enough**
4. 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?

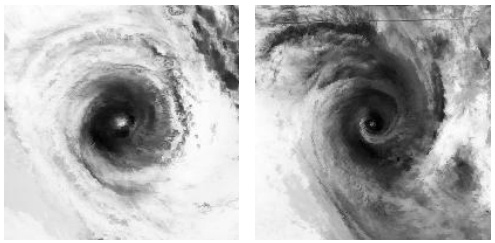
- lack of **ML-ready datasets**
- lack of **model that properly utilizes domain knowledge**

issues addressed in our latest works

(Chen et al., KDD 2018)

(Chen et al., Weather & Forecasting 2019)

Flow behind Our Proposed Model



TC images

CNN

polar rotation invariance

human learning/ analysis

domain knowledge (HI)

current weather system

ML

intensity estimation

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

Results

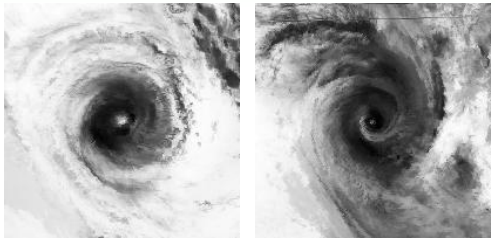
RMS Error

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?