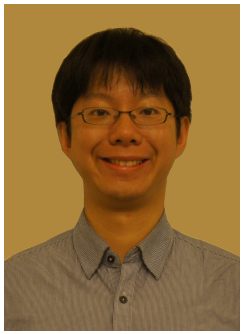
Machine Learning for Artificial Intelligence in Medicine Applications

林軒田
Hsuan-Tien Lin
htlin@csie.ntu.edu.tw

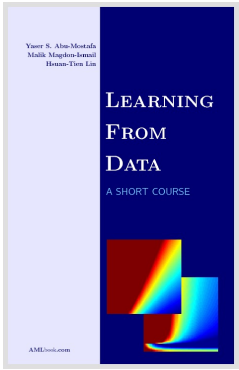
國立台灣大學
National Taiwan University

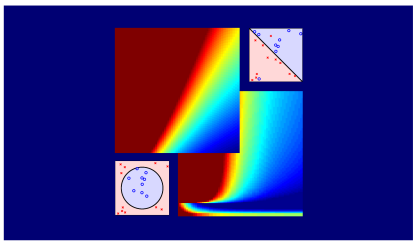
Chang Gung Memorial Hospital, 2019/03/05

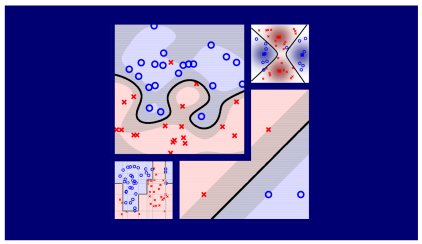
About Me
Hsuan-Tien Lin



- Professor, Dept. of CSIE, National Taiwan University
- Chief Data Science Consultant, Appier
- Co-author of textbook "*Learning from Data: A Short Course*"
- Instructor of the NTU-Coursera Mandarin-teaching ML Massive Open Online Courses
  - "*Machine Learning Foundations*": `www.coursera.org/course/ntumlone`
  - "*Machine Learning Techniques*": `www.coursera.org/course/ntumltwo`







Disclaimer

researched on quite a few ML-related topics, but . . .

limited first-hand experience in ML for AI in Medicine Applications

- Peng et al., . . . for fast disease diagnosis, NeurIPS 2018:
  building family-medicine doctor-bot
- Chou and Lin, ML for interactive verification, PAKDD 2014:
  effective use of doctor's time on screening X-ray scans
- Jan et al., Cost-sensitive classification on pathogen species of bacterial meningitis . . ., BIBM 2011:
  leveraging doctor's domain knowledge (to be introduced)
- Lin and Li. Analysis of SAGE results with combined learning techniques. In ECML/PKDD Discovery Challenge 2015:
  using machine learning properly on small medical data

will talk more about
general wisdom (hopefully),
less about specific techniques

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? :-)

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

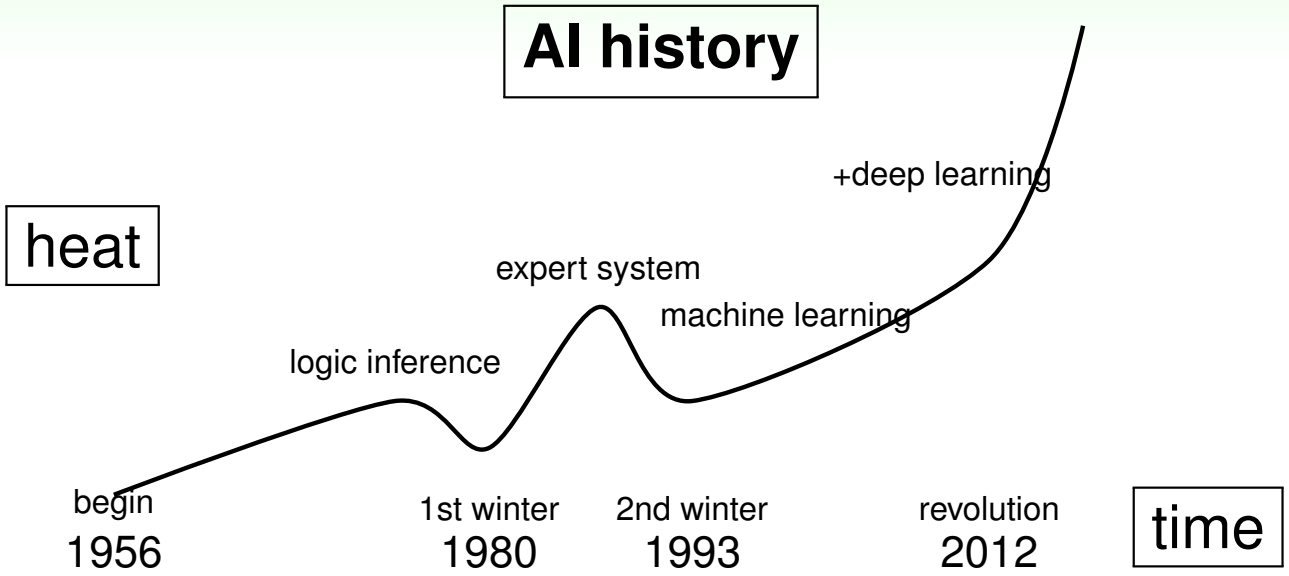
My Definition

intelligently $\approx$ easily

is your smart phone 'smart'? :-)

user-needs-driven AI is important

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

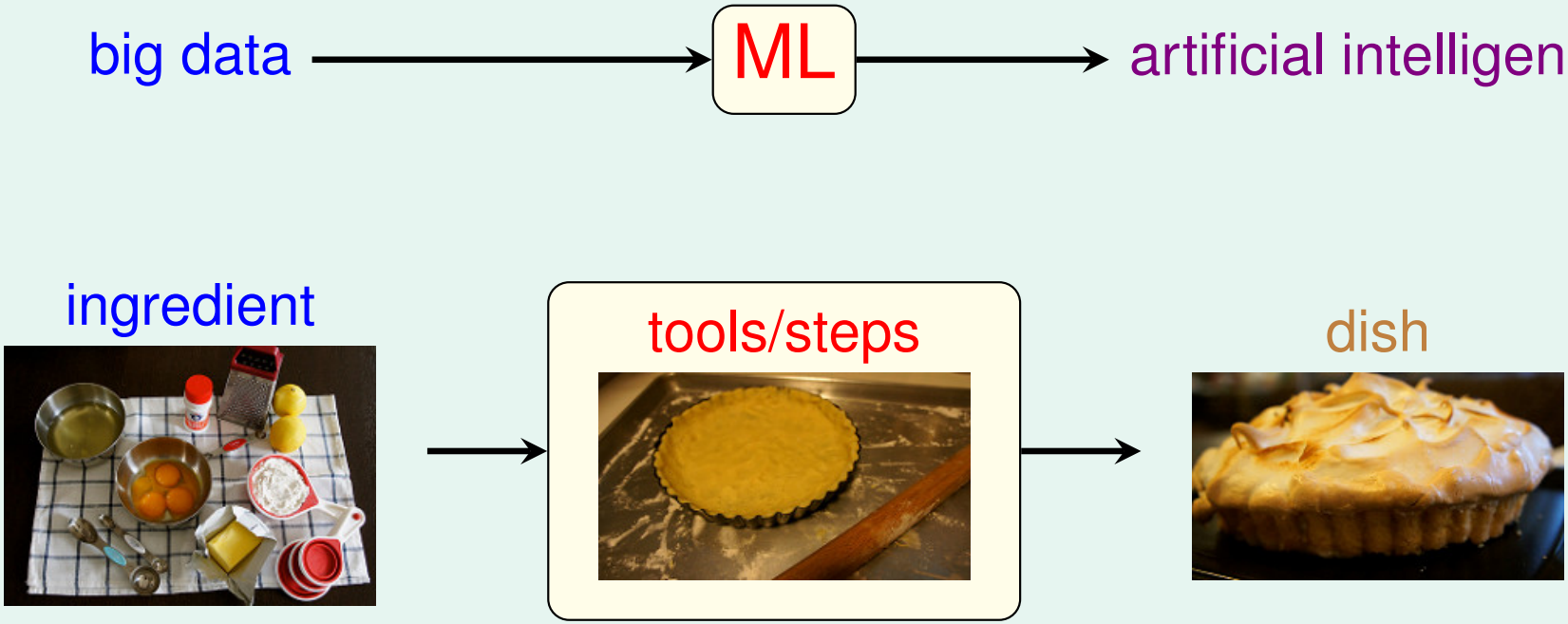
Acting Rationally

Machine Learning

machine learning: core behind modern (data-enabled) 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)

ML Scientist
≡ restaurant **chef**

Bigger Data Towards Better AI

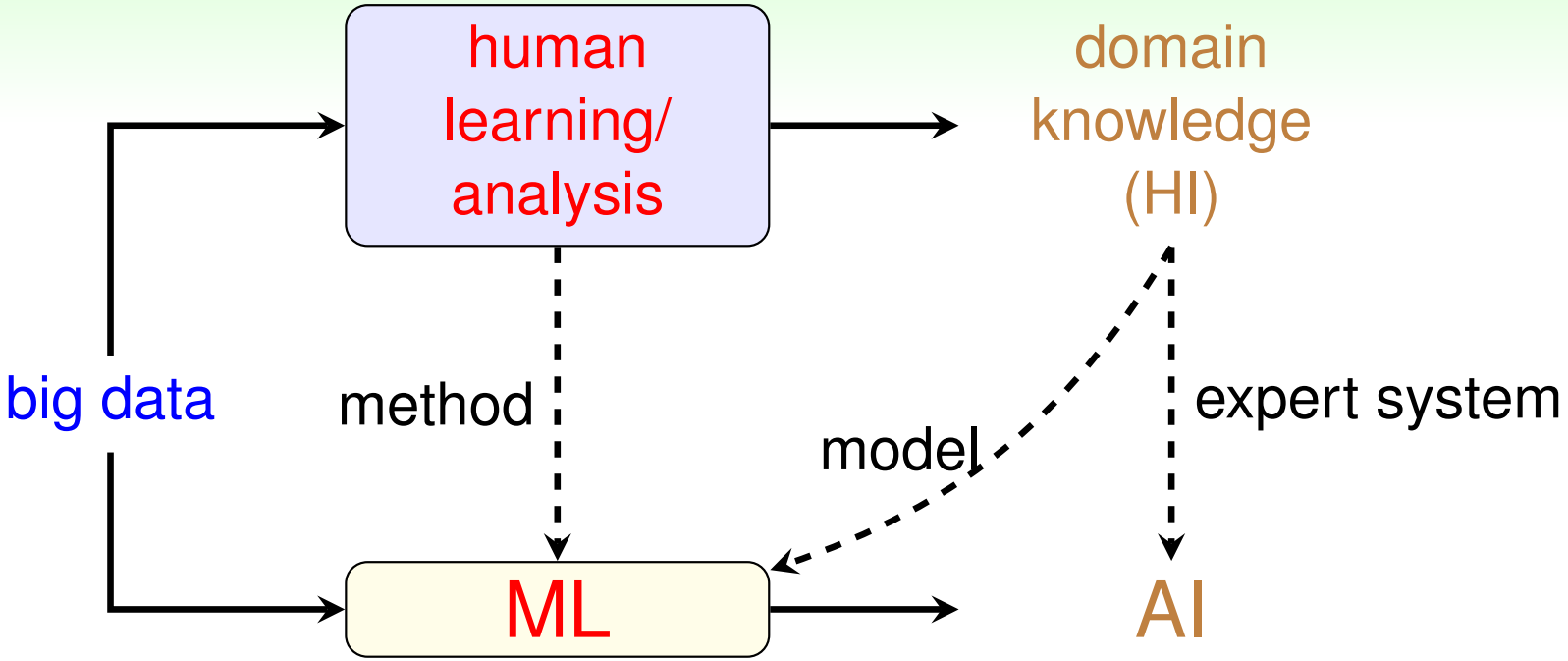
best route by shortest path

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**

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**

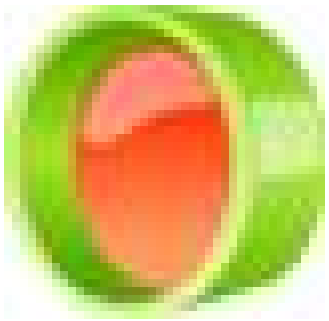
- "You'll not be replaced by AI, but **by humans who know how to use AI**"

  (Sun, Chief AI Scientist of Appier, 2018)

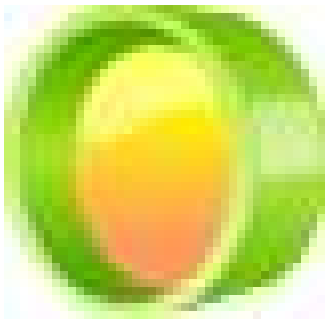
Outline

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?

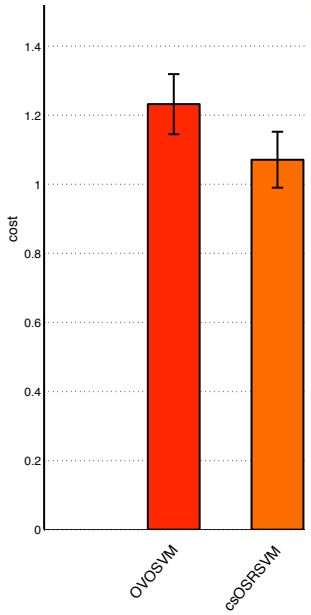
A Real Medical Application: Classifying Bacteria

- 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

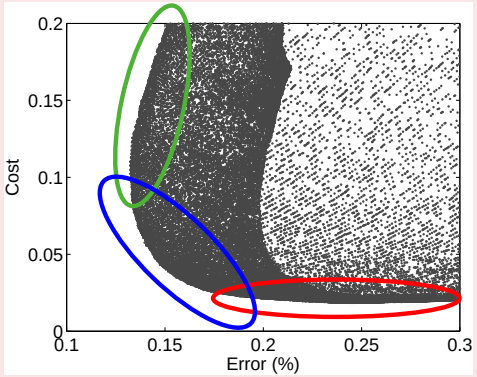


(Jan et al., BIBM 2011)

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 **inputing 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**

Outline

Is Logistic Regression Part of ML?

No

- developed in 1958, **even before "ML" named**
- applied on medicine research **long before "ML" popularized** (e.g. `https://www.ncbi.nlm.nih.gov/pubmed/11576808`)

Yes

- wikipedia: "*Logistic regression is an **important ML algorithm**.*"
- **special case of** modern **deep learning** approaches
- widely included **in ML tool boxes**

my biased opinion:

LogReg analysis: not (typical) ML;
LogReg algorithm: (typical) ML

but **both important for modern AI**

Shall We Replace Our Logistic Regression Model with Fancy ML Models?

Yes

- ML may provide **more opportunities** for better solving your problem
  - consider **more factors**
  - leverage **non-linear relationship**
  - **learn** $\rightarrow$ **analyze** (ML) v.s. analyze $\rightarrow$ regress

No

- LogReg: **safe first-hand choice** in ML anyway —philosophy of **linear first**
- not really **replacing**, but worth **comparing**
- super big ML jungle: **risky if lost**

concrete suggestions:

- compare with ("try") **some mature ML models**
- consult/collaborate with **ML specialist** if using advanced ML models

Some Mature ML Models Recommended

Random Forest

- voting of many (random) decision trees
- analysis: **feature importance**
- benefit: **robust** and **efficient** in general

Gradient Boosted Decision Tree

- optimized combination of decision trees
- analysis: **feature importance**
- benefit: **accurate** for many applications

(RBF-) Support Vector Machine

- optimized combination of key examples
- analysis: **key examples** (support vectors)
- benefit: **robust** for mid-sized data

suggested reading:

A Practical Guide to Support Vector Classification

`https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf`

Can We Explain ML Predictions?

courtesy of my Appier colleague
Jen-Yee Hong, M.D.

Yes

- **for simple models** like LogReg using statistics tools **or feature importance**
- ongoing research to explain complex ML models **with some initial success on visual data**

No

- **not generally applicable** to every ML model nowadays

explainable ML is **getting more important**

Can We Trust ML Predictions?

courtesy of my Appier colleague
Jen-Yee Hong, M.D.

Yes

- ML can be more accurate if **properly used**

No

- non-ML-specialists may not **properly use** ML tools
- need more **honest success stories** before winning human trust

trust **needs accumulation**

Summary

- ML for (Modern) AI:
  tools + human knowledge $\Rightarrow$ **easy-to-use application**
- ML Research for AI in Medicine Applications:
  collaborative to keep discovering new research directions
- Suggestions to Medicine Researchers on Using ML-driven AI:
  ML provides more **opportunities** but needs **care**

Thank you! Questions?