Machine Learning for Modern Artificial Intelligence

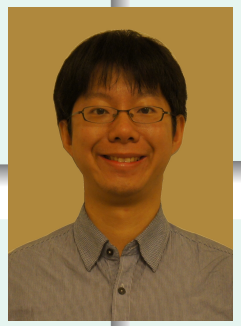
Hsuan-Tien Lin
林軒田

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

January 25, 2022
AI & Data Science Workshop

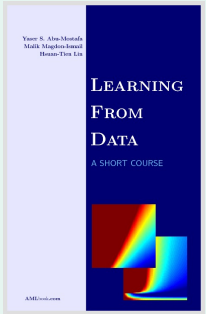
About Me

Professor
National Taiwan University

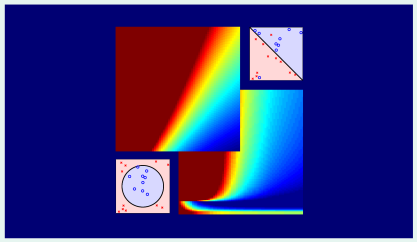


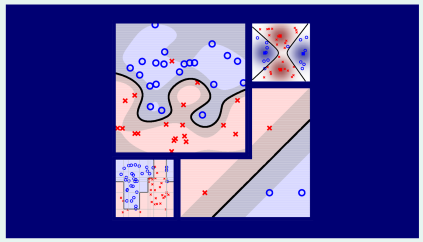
Chief Data Science Consultant
(former Chief Data Scientist)
Appier Inc.

Co-author
Learning from Data



Instructor
NTU-Coursera MOOCs
ML Foundations/Techniques





Outline

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
- character recognition

acting humanly or rationally:
more academia/industry attention 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

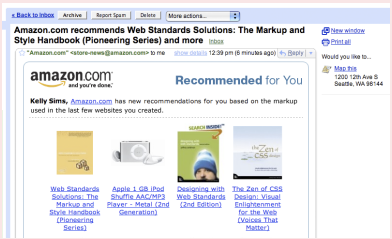
Siri

By Bernard Goldbach [CC BY 2.0]

iRobot

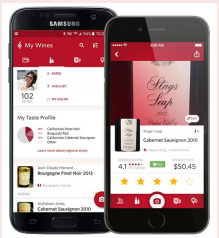
By Yuan-Chou Lo [CC BY-NC-ND 2.0]

Amazon Recommendations



By Kelly Sims [CC BY 2.0]

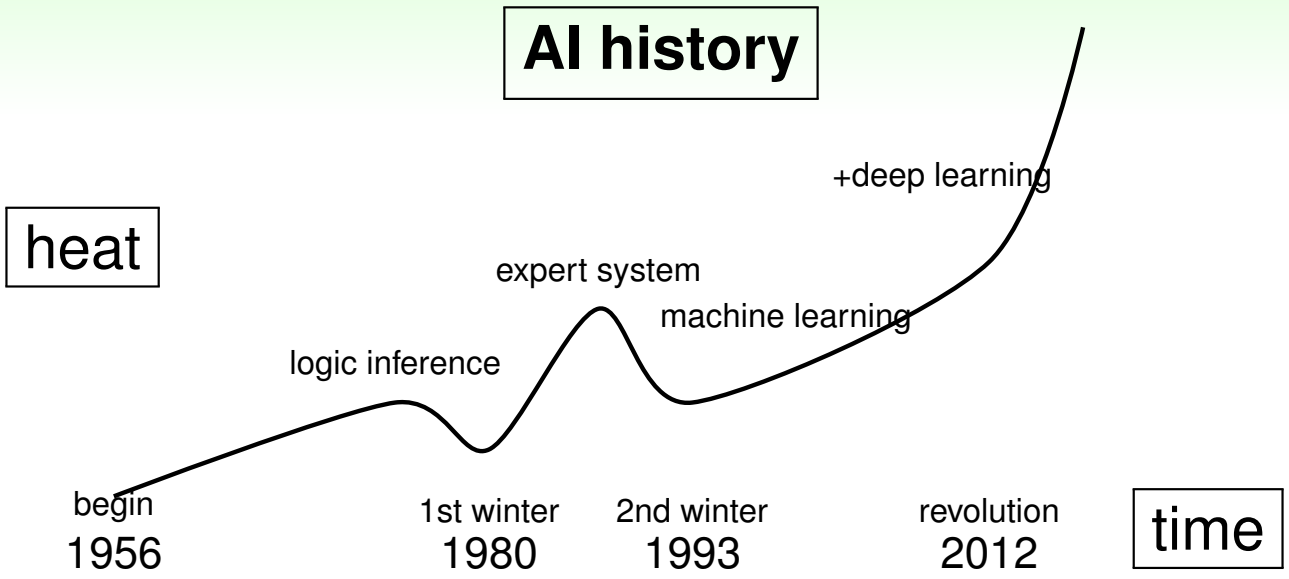
Vivino



From `nordic.businessinsider.com`

application intelligence is everywhere!

AI Milestones



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

reason of winters: **expectation mismatch**

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 ×

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)

what is **differeent** now?

What is Different Now?

More Data

- cheaper storage
- Internet companies

Better Algorithms

- decades of research
- e.g. deep learning

Faster Computation

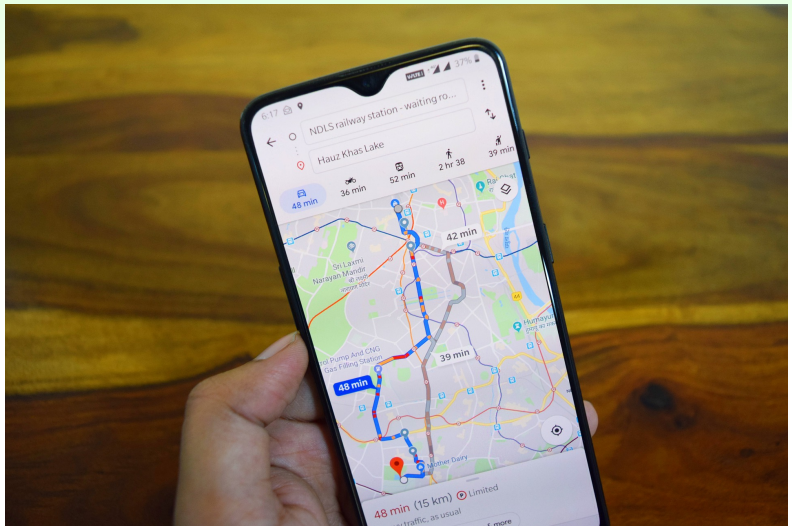
- cloud computing
- GPU computing

Healthier Mindset

- reasonable wishes
- key breakthroughs

data-enabled AI: mainstream nowadays

Bigger Data Enable 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

many possibilities when
using the right tools

ML-based AI Applications (1/4): Medicine

data → ML → AI

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

for computer-assisted diagnosis

- data:
  - **patient status**
  - **past diagnosis from doctors**
- AI: dialogue system that **efficiently identifies disease of patient**

my student's earlier work
as intern @ HTC DeepQ

ML-based AI Applications (2/4): Communication

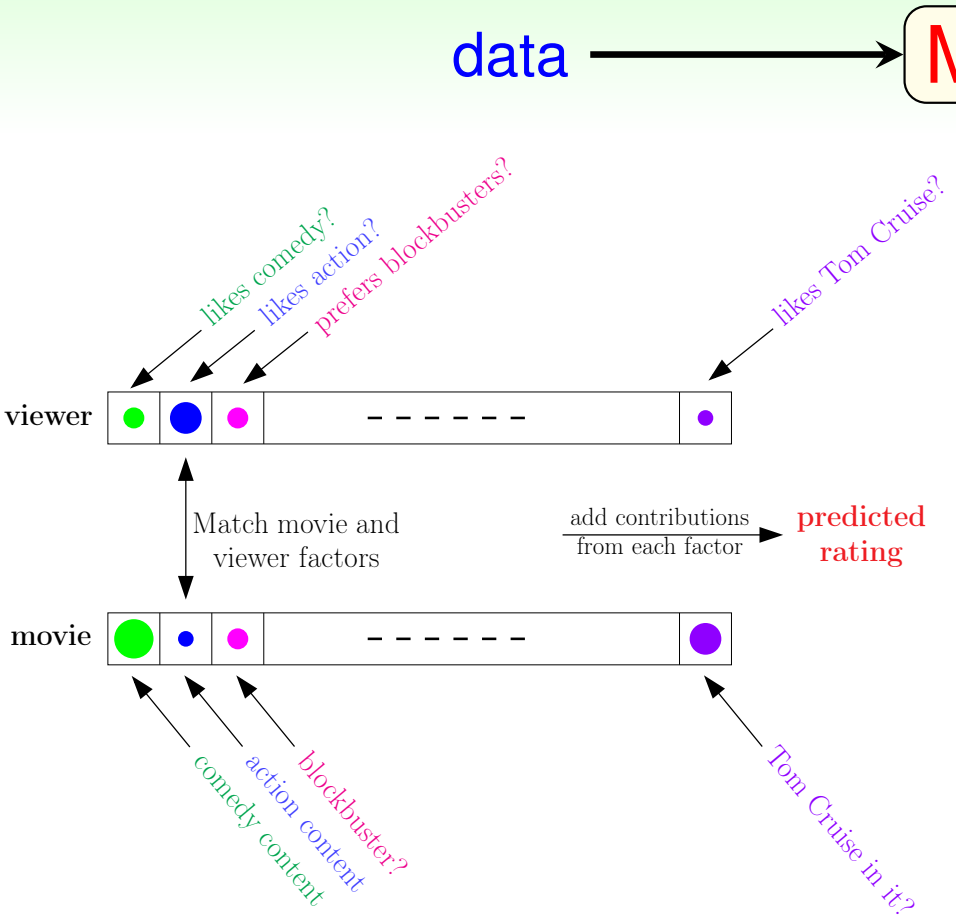
data → ML → AI

for 4G LTE communication

- data:
  - **channel information** (the channel matrix representing mutual information)
  - **configuration** (precoding, modulation, etc.) that reaches the highest throughput
- AI: predict **best configuration to the base station** in a new environment

my student's earlier work as intern @ MTK

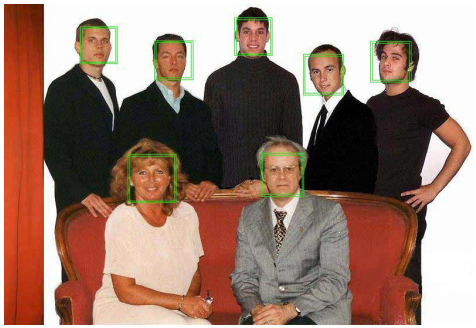
ML-based AI Applications (3/4): Entertainment



- data: how many users have rated some movies
- AI: predict how a user would rate an unrated movie

world-champion from
National Taiwan Univ. in KDDCup 2011

ML-based AI Applications (4/4): Security



original picture by F.U.S.I.A. assistant and derivative work by Sylenius via Wikimedia Commons

face recognition

- data: **faces and non-faces**
- AI: predict **which boxes contain faces**

mature ML technique, but often need **tuning** for different application intelligence needs

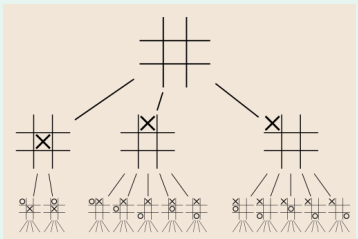
Good AI Needs Both ML and Non-ML Techniques

AlphaGo

(Public Domain, from Wikipedia; used here for education purpose; all other rights still belong to Google DeepMind)

Non-ML Techniques

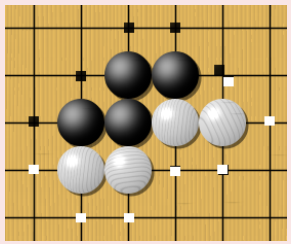
Monte C. Tree Search
≈ **move simulation** in brain



(CC-BY-SA 3.0 by Stannered on Wikipedia)

ML Techniques

Deep Learning
≈ **board analysis** in human brain



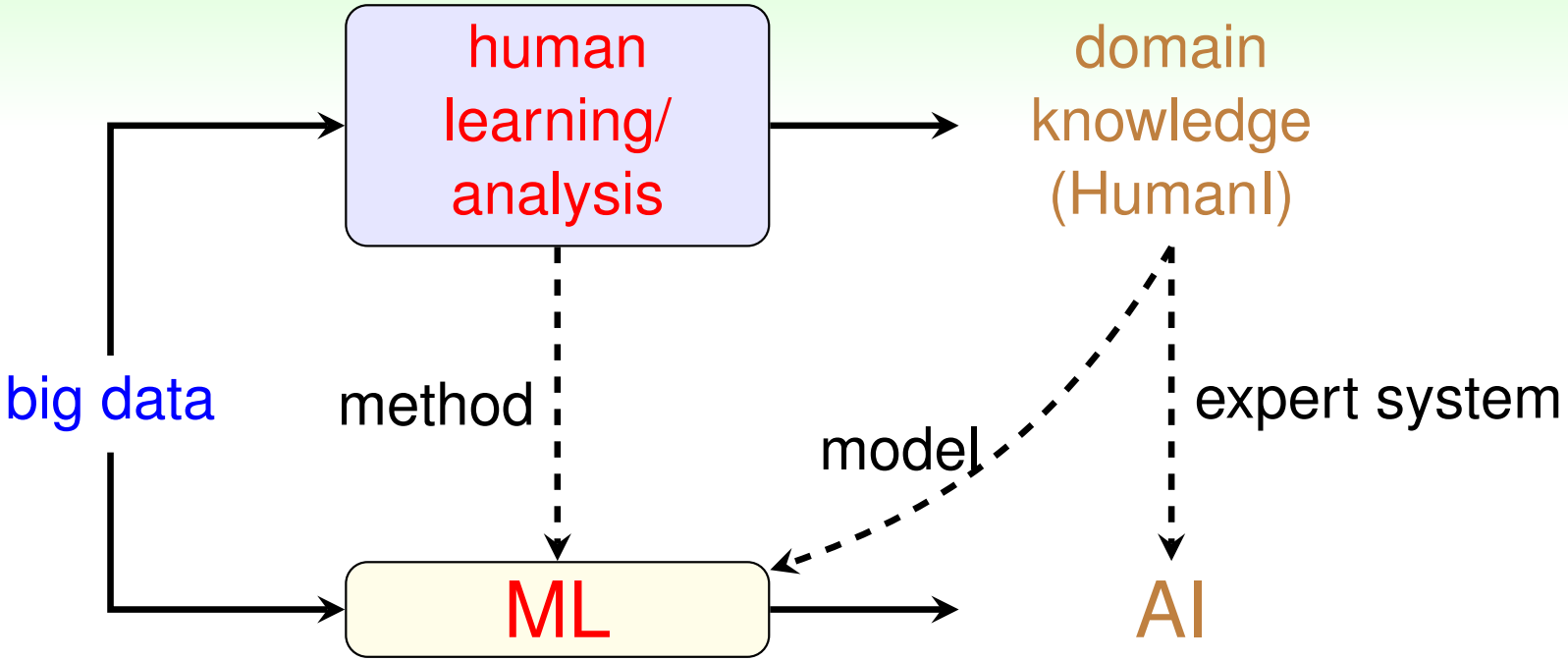
(CC-BY-SA 2.0 by Frej Bjon on Wikipedia)

Reinforcement Learn.
≈ **(self)-practice** in human training

(Public Domain, from Wikipedia)

good AI: important to use the **right** techniques—ML **& others, including human**

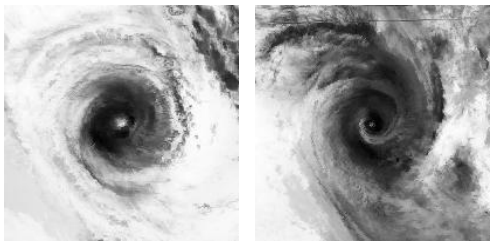
Full Picture of ML for Modern AI



industry: **black plum is as sweet as white**

Example: Tropical Cyclone Intensity Estimation

meteorologists can 'feel' & estimate TC intensity from image



TC images

human learning/ analysis

domain knowledge (HumanI)

CNN

polar rotation invariance

current weather system

ML

intensity estimation

better than current system

(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?

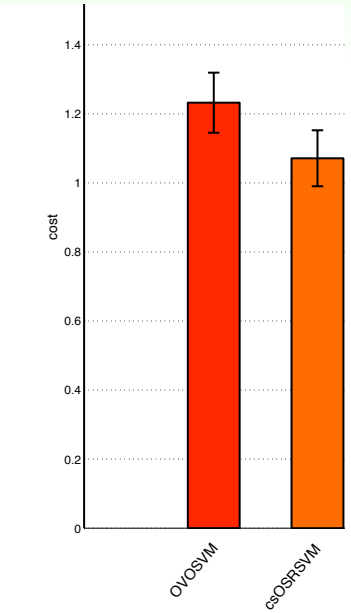
A Real Medical Application: Classifying Bacteria

- 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

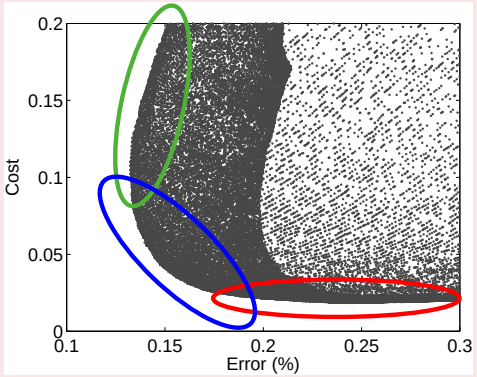


(Jan et al., BIBM 2011)

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	≈
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

?

COVID19

cold

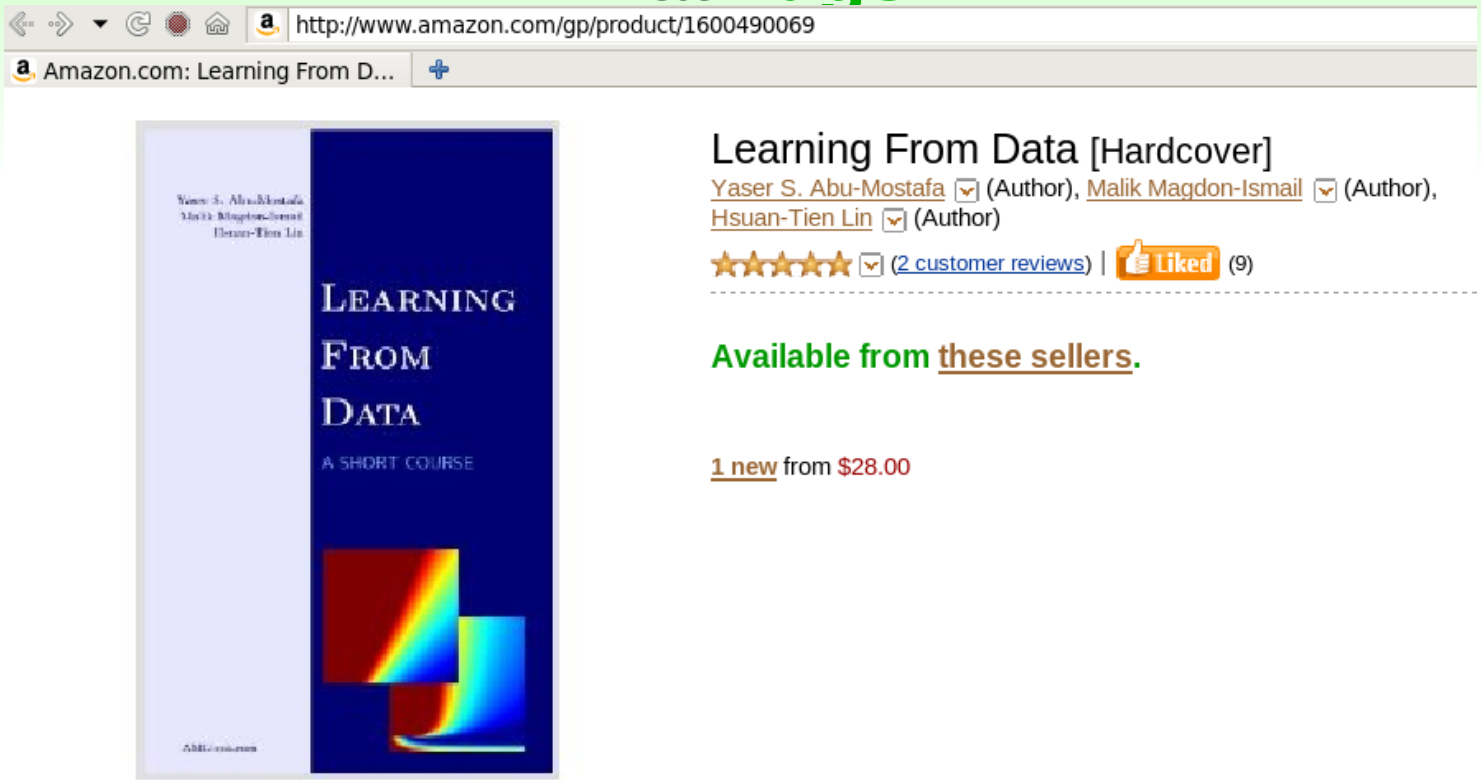
healthy

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**

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

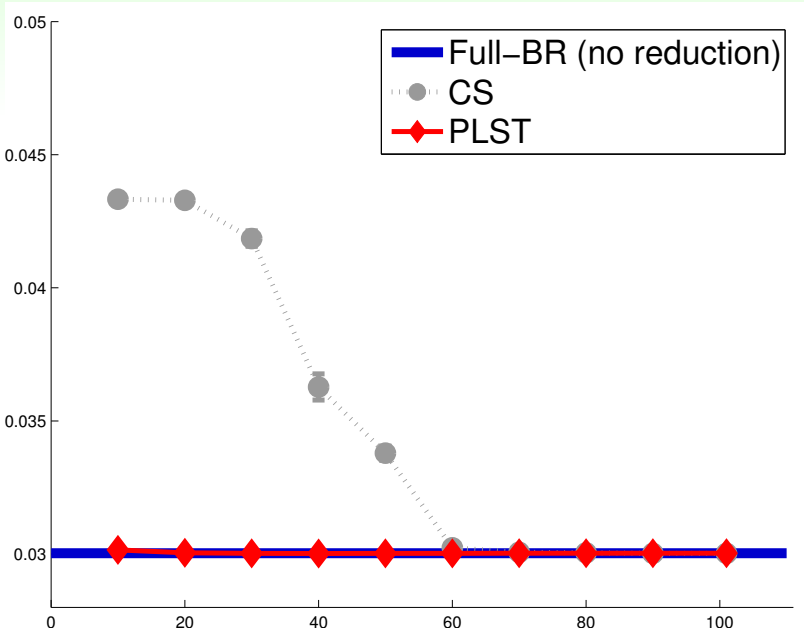
Our Proposed Approach: Compressive Sensing $\Rightarrow$ PCA

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

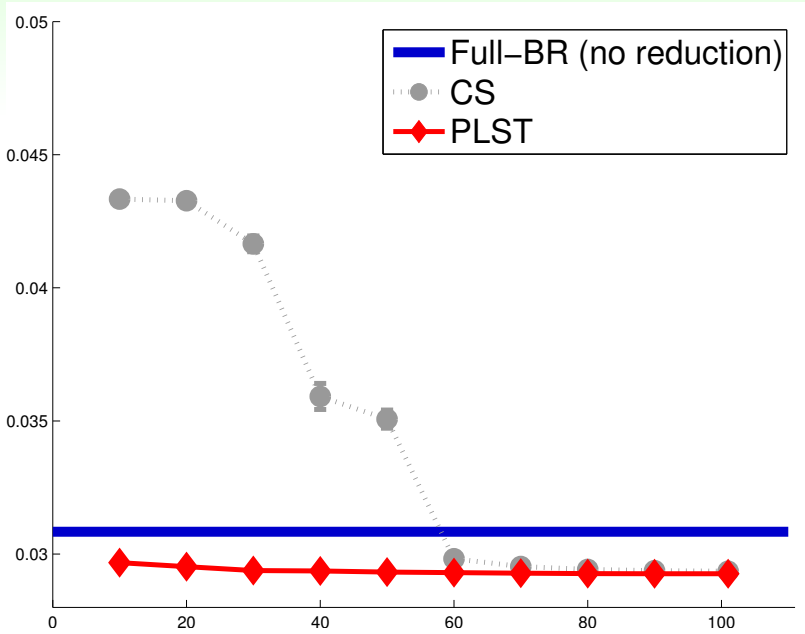
1. **compress**: encode original data by **PCA**
2. **learn**: get regression function from compressed data
3. **decode**: decode regression predictions to label vector by **reverse PCA + quantization**

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**

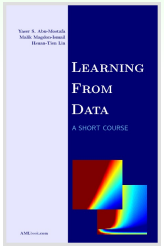
Why? CS creates
harder-to-learn regression tasks

Our Works Continued from PLST

1. **Compression** Coding (Tai & Lin, NC Journal 2012 with 342 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 262 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 with 48 citations)
   —**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

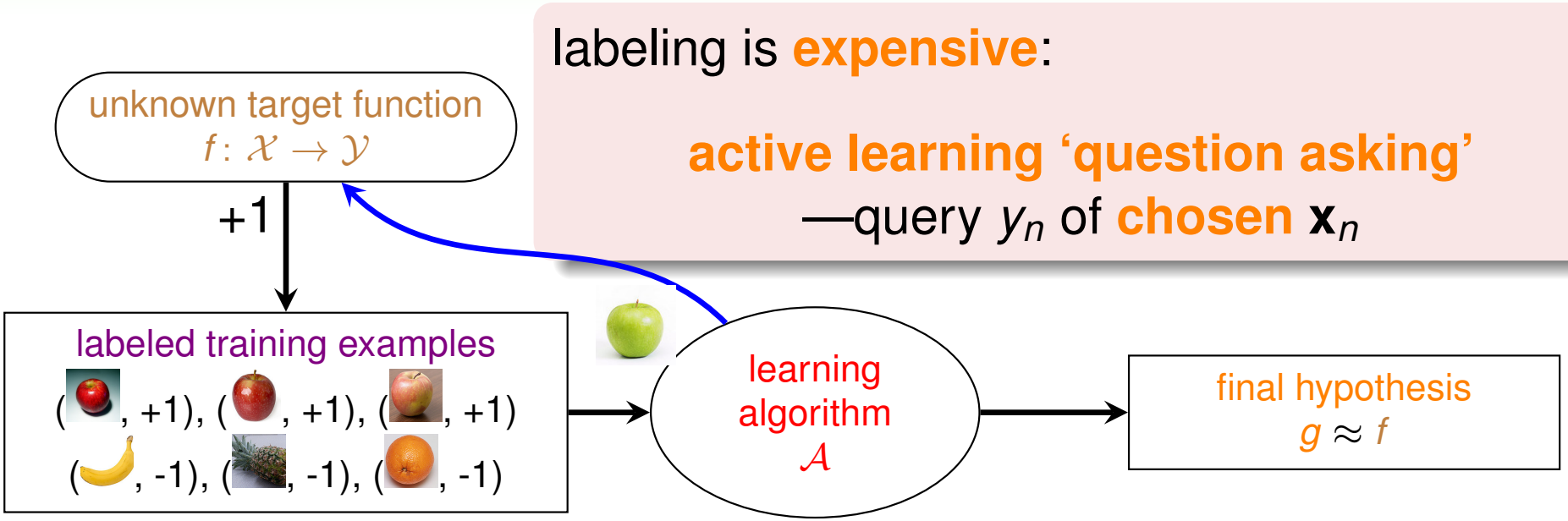


?: {machine learning, ~~data structure~~, data mining, ~~object oriented programming~~, artificial intelligence, ~~compiler~~, ~~architecture~~, ~~chemistry~~, textbook, ~~children book~~, ~~. . . etc.~~ }

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**
   - all fields could find their **concrete roles** in AI
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'



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

ask **most confused** question

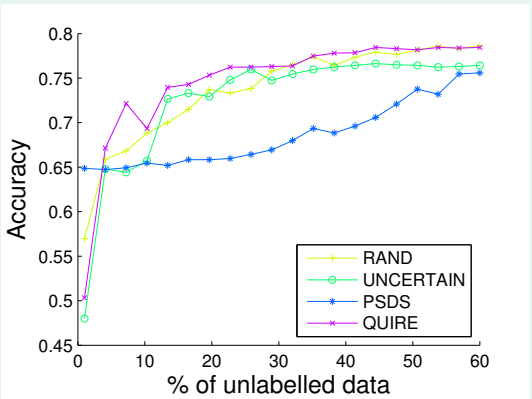
Strategy 2

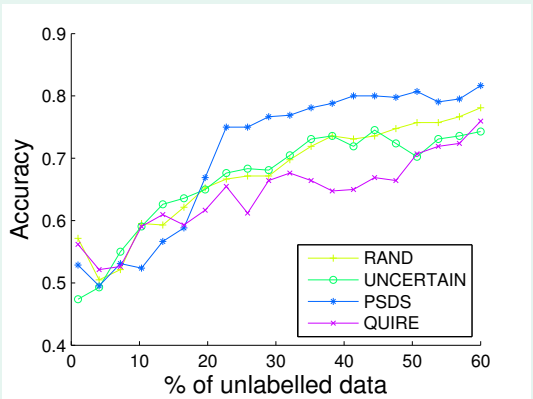
ask **most frequent** question

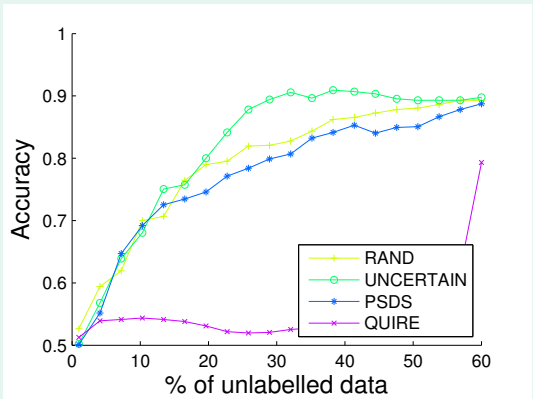
Strategy 3

ask **most debateful** question

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



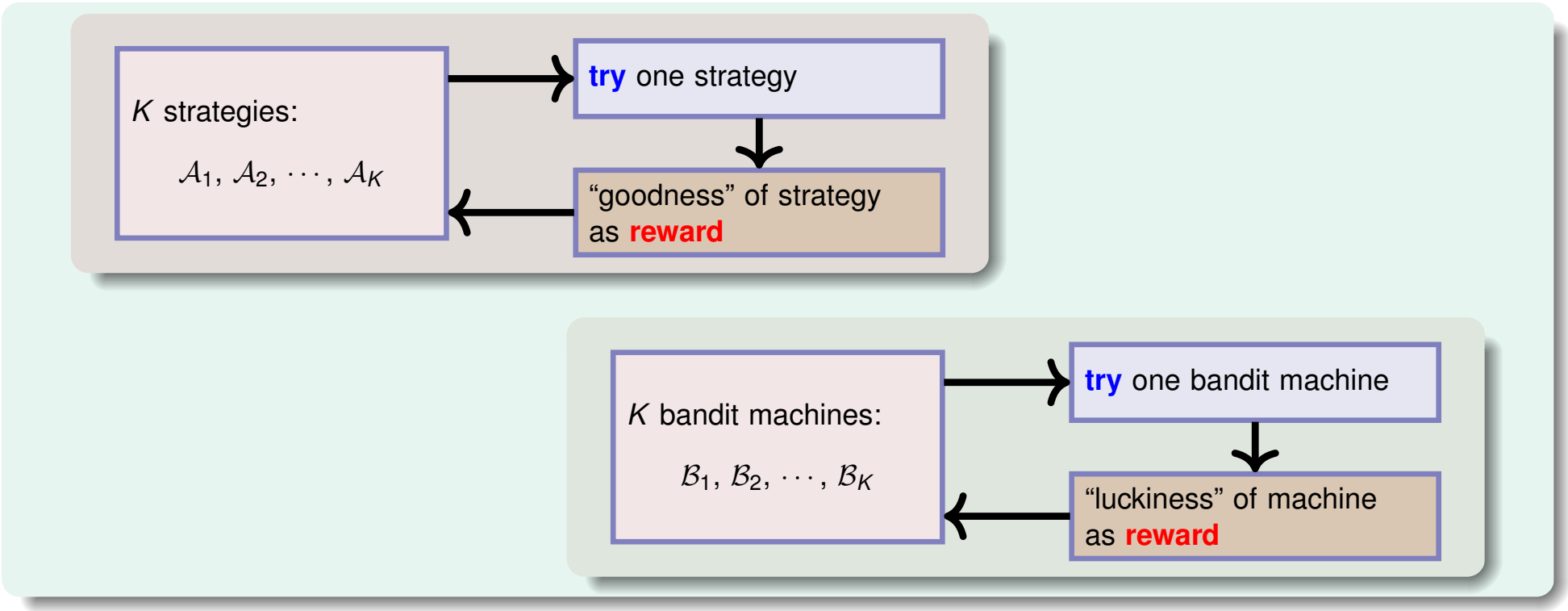




application intelligence: how to **choose strategy smartly**?

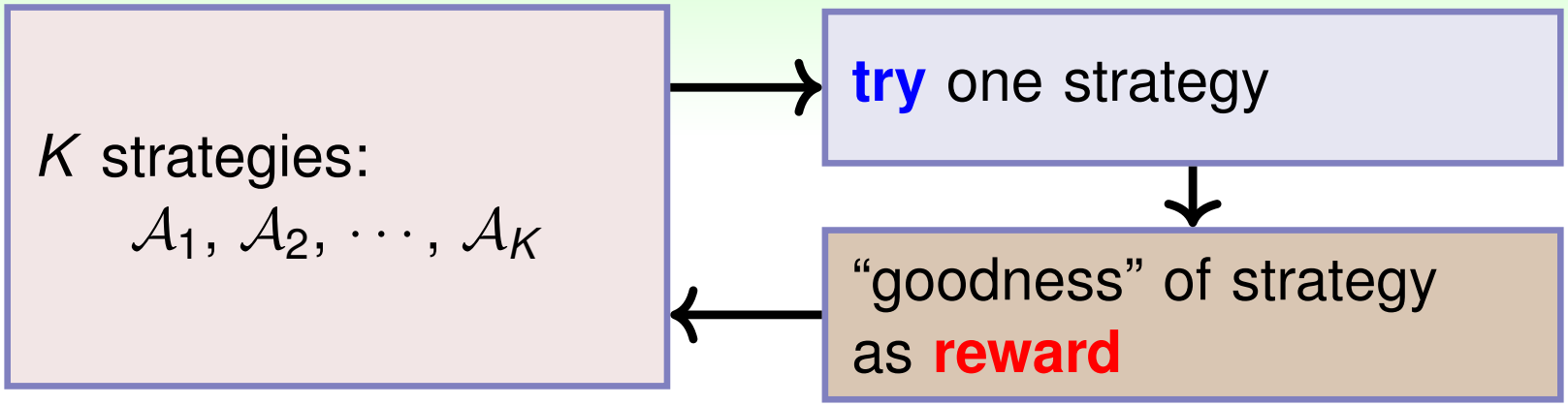
Idea: Trial-and-Reward Like Human

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



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

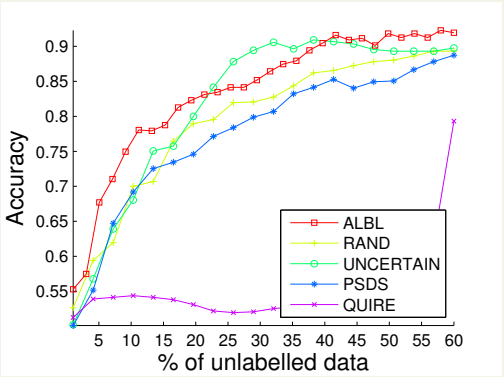
for $t = 1, 2, \ldots, T$

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

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

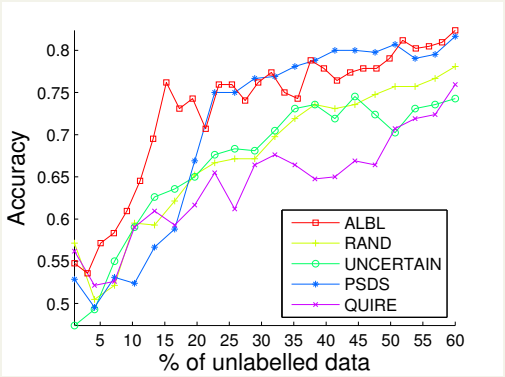
Comparison with Single Strategies

UNCERTAIN Best



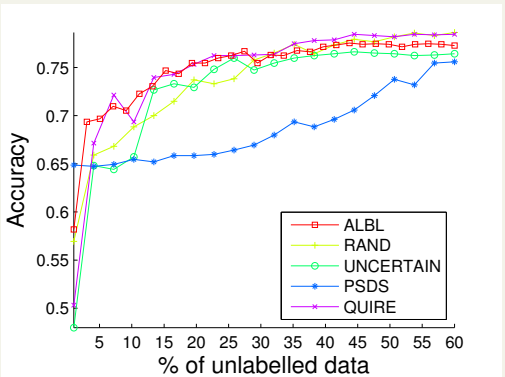
vehicle

PSDS Best



sonar

QUIRE Best



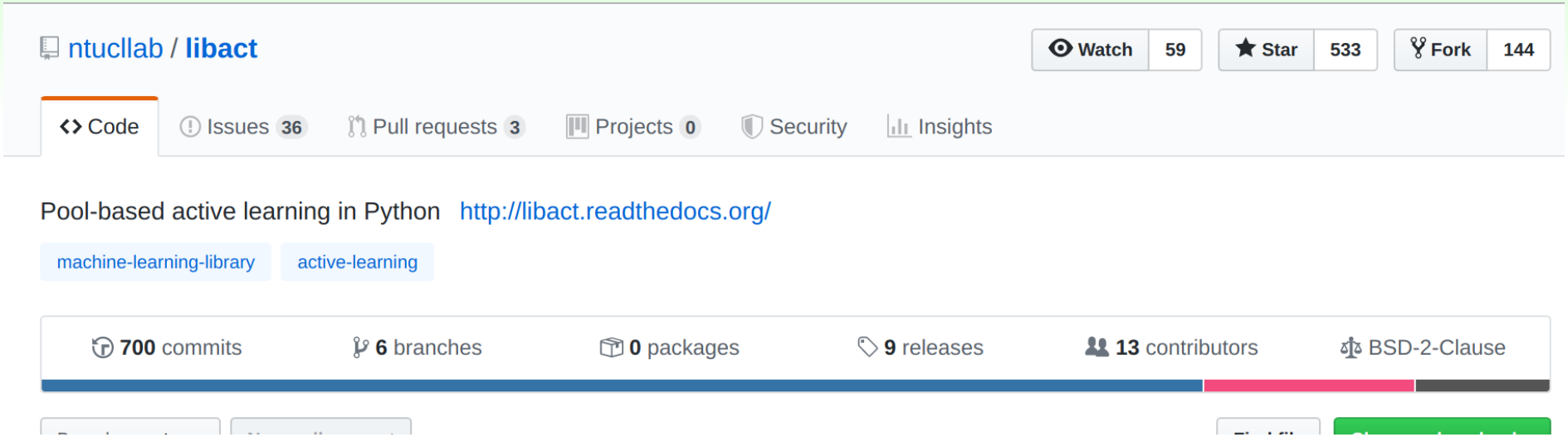
diabetes

- **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/ntuclab/libact`)

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



Yes!

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

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

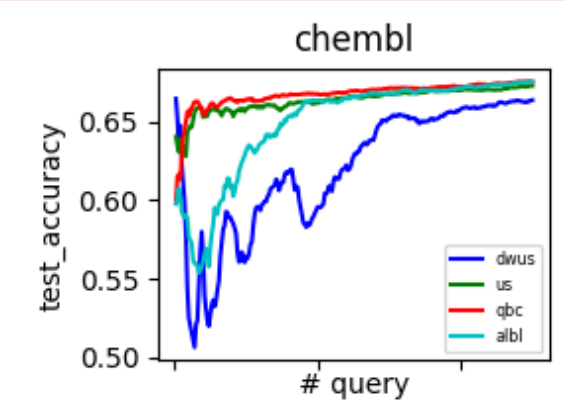
- including uncertainty, QUIRE, PSDS, . . ., **and ALBL**
- received $>$ 700 **stars**, $>$ 40 **citations**, and continuous **issues**

"`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 an 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"

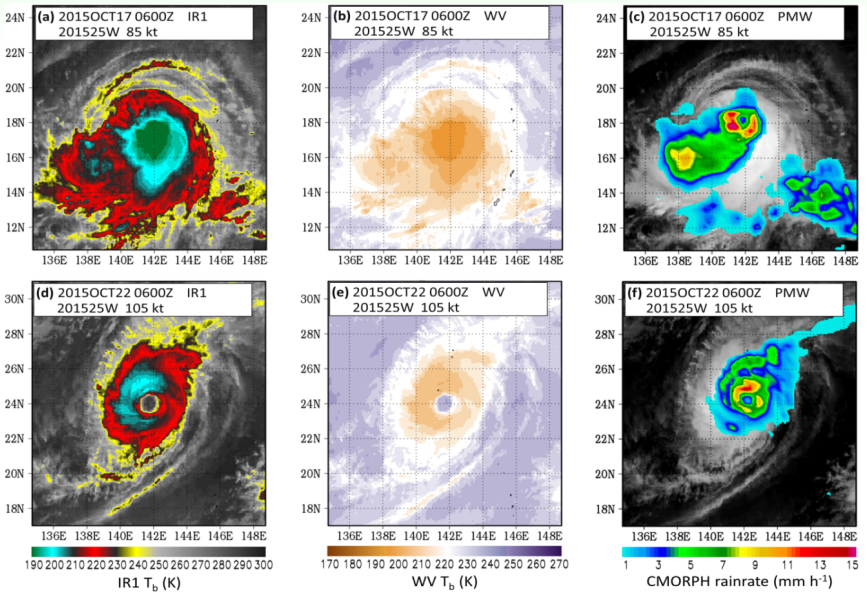
Lessons Learned from Research on Active Learning by Learning

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



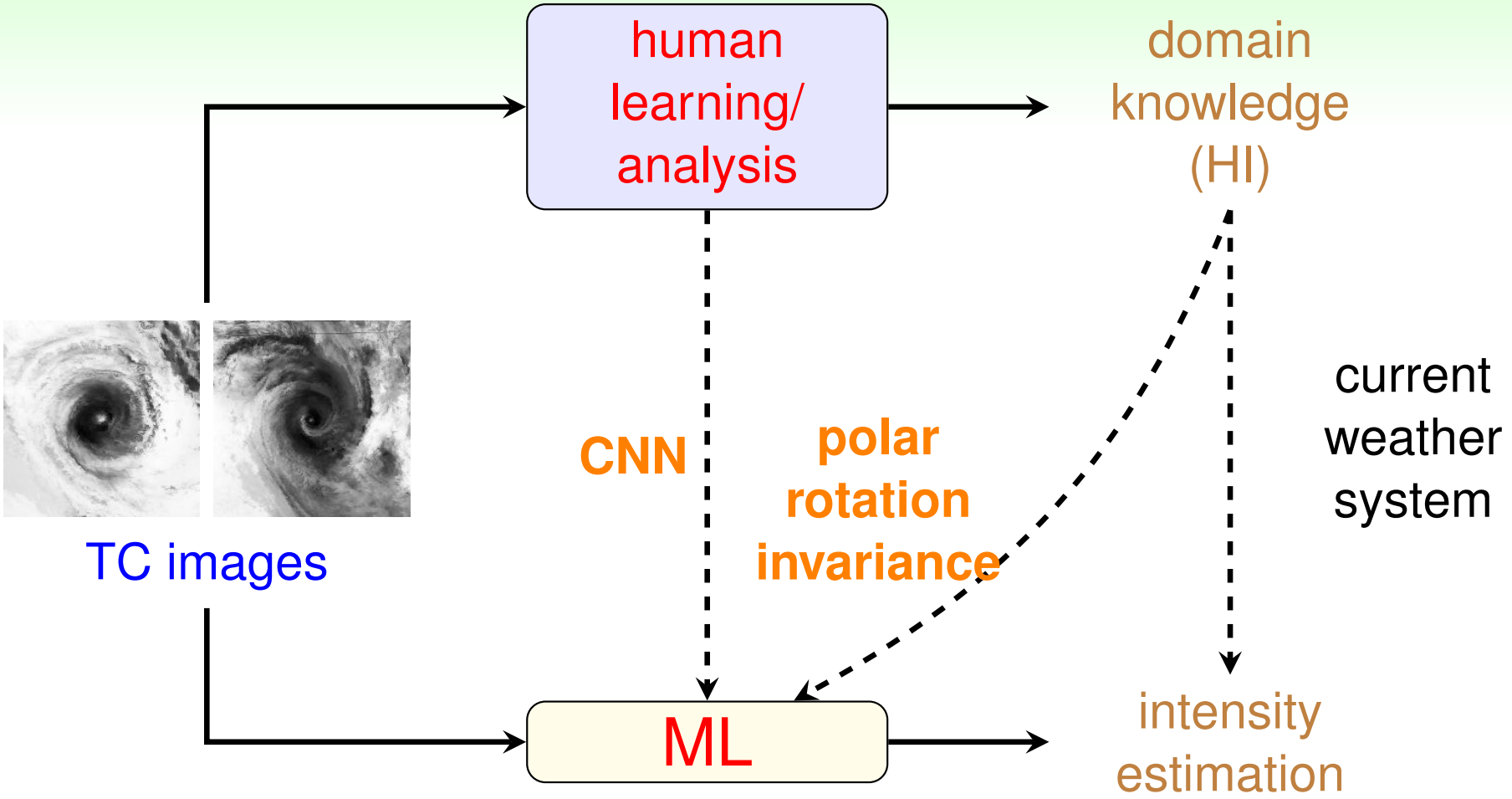
Can ML do the same/better?

- 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

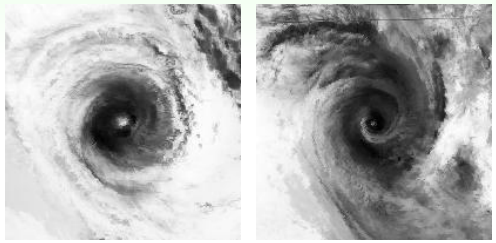
RMS Error

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

Frequently Asked Questions of ML for AI (1/4)

What is the best AI project for (my precious big) data?

My Polite Answer

good start already ☺, any more thoughts that you have in mind?

My Honest Answer

I don't know.

or a slightly longer answer:
if you don't know, I don't know.

A Similar Scenario

What is the best AI project for
(my precious big) data?

how to find a research topic for my thesis?

My Polite Answer

good start already ☺, any more thoughts that you have in mind?

My Honest Answer

I don't know.

or a slightly longer answer:
I don't know, but perhaps you can **start** by thinking about **motivation** and **feasibility**.

Finding AI Projects ≈ Finding Research Topics

- motivation: what are you interested in?
- feasibility: what can or cannot be done?

motivation

- something publishable? oh, possibly **just for people in academia** ☺
- something that **improves xyz performance**
- something that inspires deeper study

—helps generate questions

feasibility

- **modeling**
- **computational**
- budget
- timeline
- . . .

—helps filter questions

tip: important for **first AI project** to be **of high success possibility**

Frequently Asked Questions of ML for AI (2/4)

Should I use ML (or my precious Deep Learning) for my AI project?

My Polite Answer

let's understand more about the constraints of your project, shall we ☺?

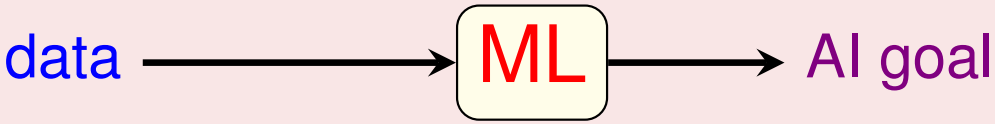
My Honest Answer

I don't know.

or a slightly longer answer:
if you don't know, I don't know.

Necessary Conditions for Using ML

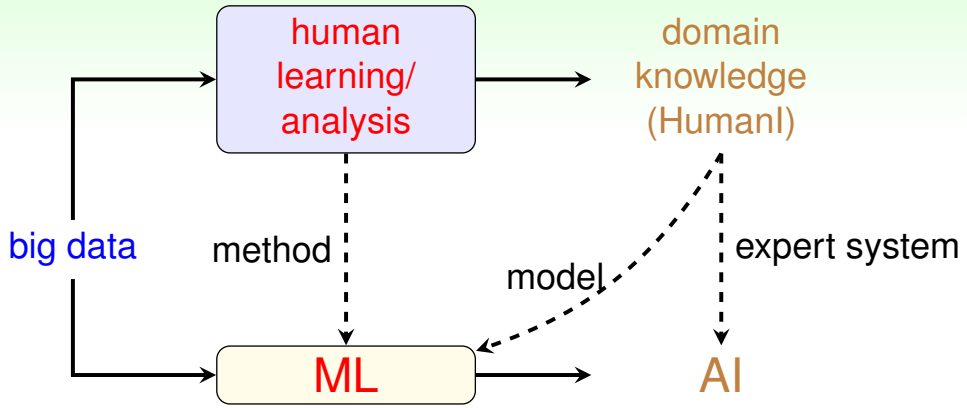
machine learning: improving some AI goal
with experience **accumulated** from data



1. exists some "underlying pattern" to be learned
   —so "AI goal" possible
2. but no programmable (easy) definition
   —so "ML" is needed
3. somehow there is enough data about the pattern
   —so ML has some "inputs" to learn from

necessary, but **not sufficient**, for using ML

Human Learning versus Machine Learning



Human Learning

- subjective
- produce domain knowledge
- fast basic solution

Machine Learning

- objective
- leverage computing power
- continuous improvement

tip: **use humans as much as possible first** before going to machines

Frequently Asked Questions of ML for AI (3/4)

What is the best machine learning model for (my precious big) data and AI?

My Polite Answer

the best model is **data-dependent**, let's chat about your data first

My Honest Answer

I don't know.

or a slightly longer answer:
I don't know about **best**, but perhaps you can **start** by thinking about **simple models**.

Sophisticated Model for AI

What is the best machine learning model for (my precious big) data and AI?

What is the most sophisticated machine learning model for (my precious big) data and AI?

- myth: my AI works best with most sophisticated model
- sophisticated model:
  - time-consuming to train and predict
  - difficult to tune or modify
  - hard to "simplify" nor "analyze"

sophisticated model shouldn't be **first choice**

Simple First

What is the first machine learning model for (my precious big) data and AI?

Taught in ML Foundations on NTU@Coursera

simple model first:

- efficient to train and predict
- easy to tune or modify
- somewhat "analyzable"
- little risk

tip: KISS Principle
—*Keep It Simple, ~~Stupid~~* **Safe**

Frequently Asked Questions of ML for AI (4/4)

How to Get my AI Project Started?

Old Me

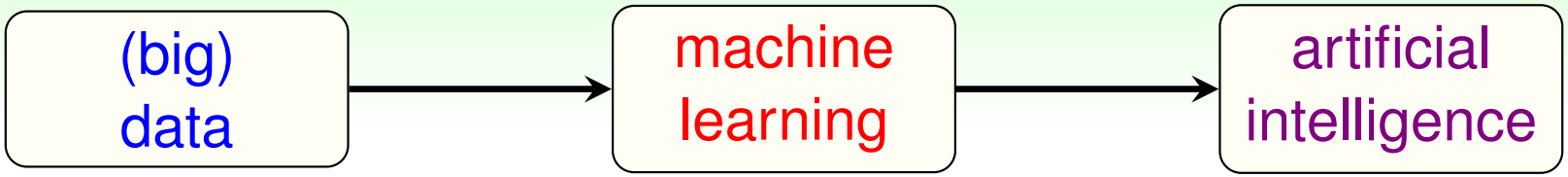
I don't know. ☺

New Me

I know one key factor!

let's see what the key factor is

Todos in AI Project



data

- gathering
- cleaning
- storing
- . . .

techniques

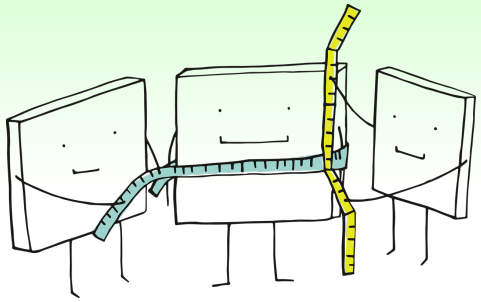
- modeling
- computation
- non-ML tech.
- . . .

usage

- **evaluation**
- deployment
- scalability
- . . .

key first step: set up **evaluation criteria**

Evaluation Criteria Guide AI Project Planning



(free image by Manfred Steger from Pixabay)

suggest improvement opportunities

⇕

data

hint
preparation steps

⇐

techniques

assist
model/tech. choices

⇐

usage

define
acceptance goals

tip: always start with
reasonable & measurable criteria
to describe prioritized AI goal

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 AI in Reality:
  - motivated/feasible project with **measurable criteria**
  - human and/or **simple** model first

Thank you! Questions?