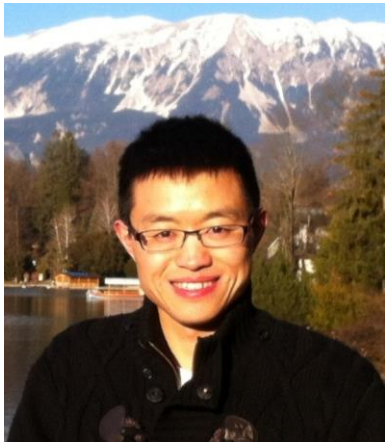
Advanced Machine Learning

Jie Tang
Department of Computer Science & Technology
Tsinghua University

A bit about Jie…



- Jie Tang, Full Professor, Vice Chair of Dept. of Computer Science of Tsinghua University. Interests include artificial intelligence, data mining, machine learning, knowledge graph.
- I have been visiting scholar at Cornell U. (working with John Hopcroft, Jon Kleinberg), UIUC (working with Jiawei Han), CUHK (with Jeffrey Yu), and HKUST (with Qiong Luo).
- I was awarded with the NSFC for Distinguished Young Scholars, NSFC for Excellent Young Scholars, Newton Advanced Fellowships Award
- Have published more than 200 paper on international conf/journals, including KDD (25), IJCAI/AAAI (20+), NIPS/ICML, ACM/IEEE Trans. (30)
- #Citation: 14,591 and *h*-index: 61
- Have a notable system, AMiner.org for academic researcher network analysis. The system has attracted 10 million users from 220 countries/regions.
- **Homepage:** http://keg.cs.tsinghua.edu.cn/jietang/

Contact Information

- Jie Tang
  - Software Division, Department of Computer Science, Tsinghua U.
  - Office: Rm 1-308, FIT Building
  - E-mail: jietang@tsinghua.edu.cn
  - Phone: 62788788-20, 13911215746
  - Open hours: Tuesday afternoon 2:00pm-5:00pm

联系信息

- Head TA
  - 张帆进、孙梦阳
  - FIT楼1-308
  - E-mail: zfj17@mails.tsinghua.edu.com
    sunmy19@mails.tsinghua.edu.cn
  - 电话：18811512596, 18641153593
  - Open hours: 14:00pm-18:00pm from M to F
- Associate TA
  - 岑宇阔，cyk18@mails.tsinghua.edu.cn

Resources

- Conferences:
  - Theory: NIPS, COLT, STOC/FOCS
  - Algorithm: ICML, KDD, UAI, IJCAI/AAAI
  - App: SIGIR, WWW, ACL
- Journals:
  - JMLR, JAIR, MLJ, ACM TKDD, IEEE TKDE
- Handbooks
  - Yoshua Bengio, Ian J. Goodfellow, Aaron Courville. Deep Learning. 2016.
  - Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2007.
  - Daphne Koller, Nir Friedman. Probabilistic Graphical Models. MIT Press, 2009

Artificial Intelligence

Artificial Intelligence



AlphaGo



Self-



- dog + cat =



Image

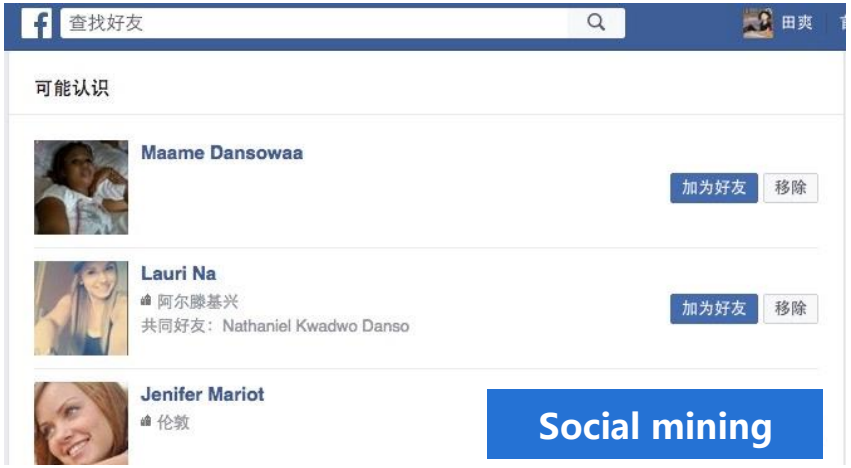
Look broader



QA



Recommendatio

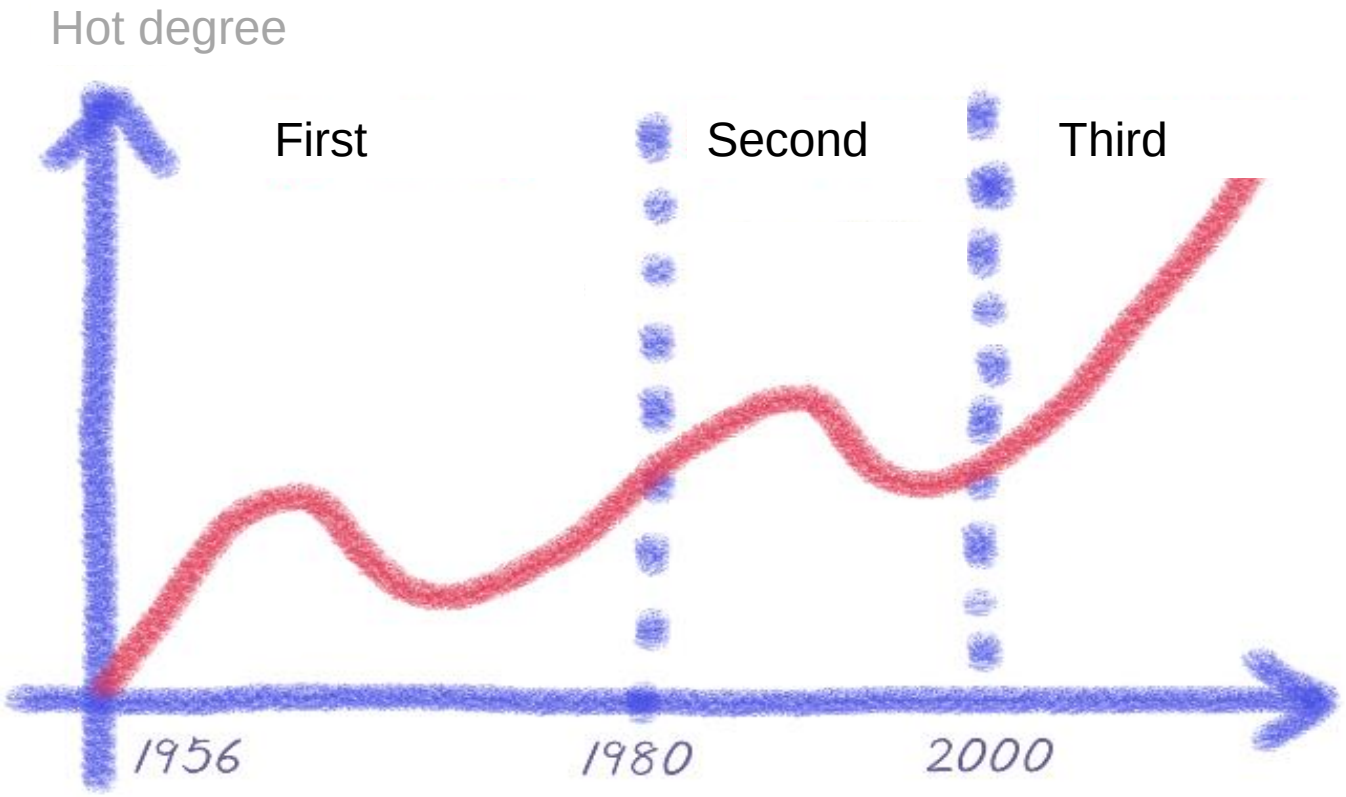


Social mining



Search engine

AI History

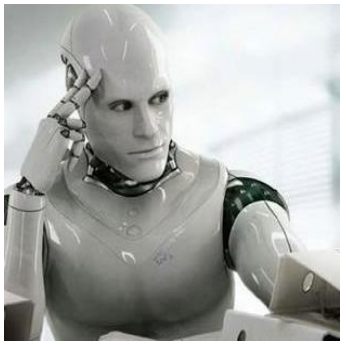


First	Second	Third
• Logic Theory Machine • Perceptron	• Expert System • Decision Tree	• Big Data • Deep Learning

National AI Strategy

国务院印发《新一代人工智能发展规划》

“要把人工智能发展放 在国家战略层面系统布局、主动谋划，牢牢把握人工智能发展新 阶段国际竞争的战略主动，打造竞争新优势、开拓发展新空间， 有效保障国家安全。” 2017, July



Intelligent



AI Education

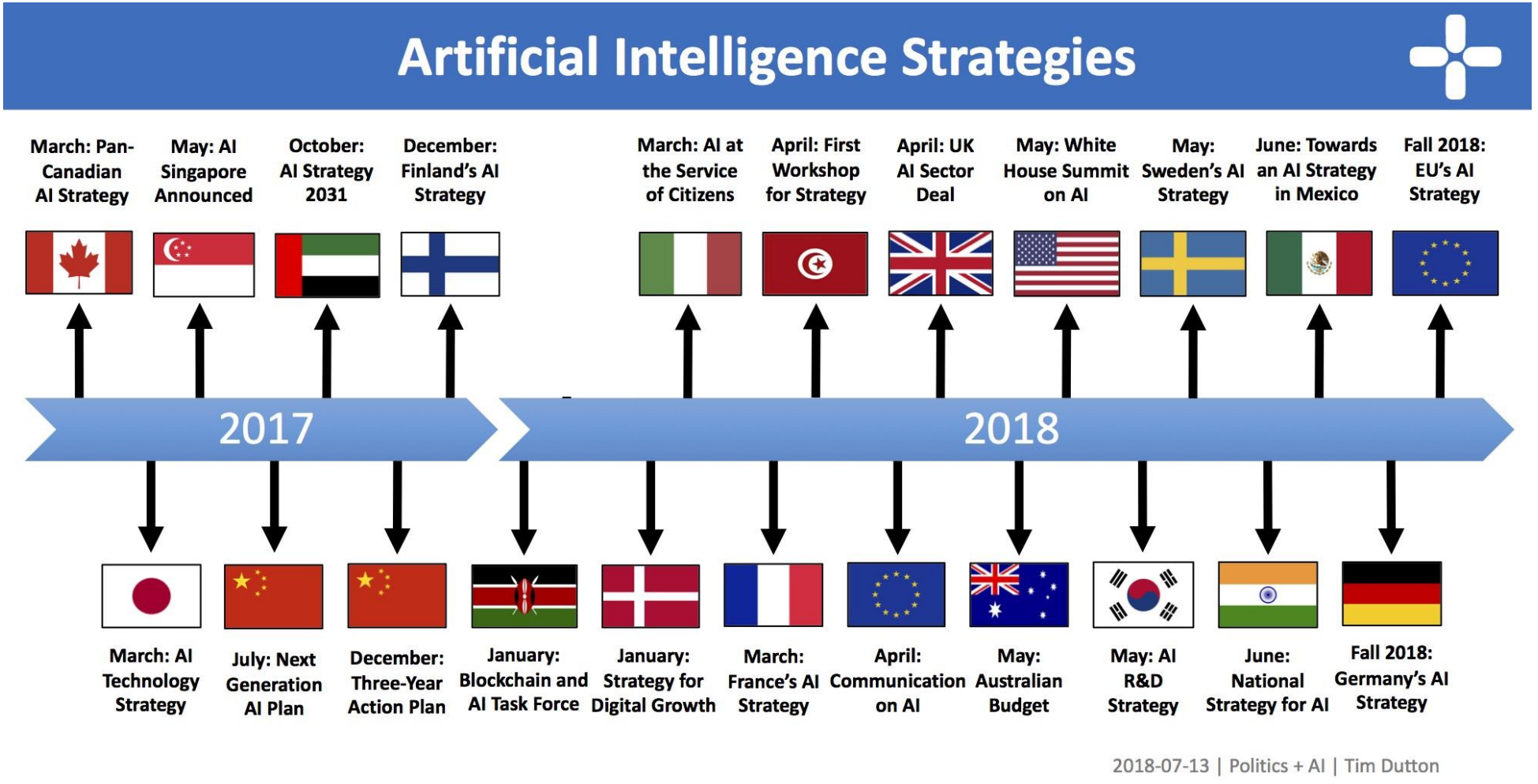


AI Agriculture

AI
Everything…

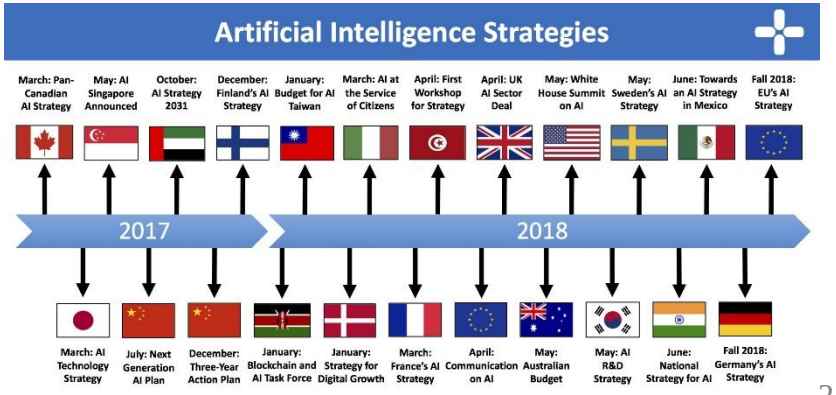
….

AI Strategies



National AI Strategies

- Example Countries
  - In 2017, Canada launched Pan-Canadian Artificial Intelligence Strategy
    - AMII, Vector, Mila
  - This year, UK is putting AI and Data at the heart of its Industrial Strategy
  - In March 2018, France announces his AI strategy with investing €1.5 billion over 5 yrs on AI research and innovation
  - US recently also release a national strategy for artificial intelligence. Moreover,
    - OpenAI
    - TensorFlow



AI开放平台

- 部分国际AI开放平台
  - OpenAI (Sutskever)
  - Mila (Bengio)
  - Vector Institute (Hinton)







OpenAI

- Research Director: Ilya Sutskever
- Sponsored by
  - 马斯克
  - Y Combinator总裁阿尔特曼
  - 天使投资人彼得•泰尔等
- 目标：制造通用机器人和使用自然语言的聊天机器人
- 2年来重要成果
  - 创建了一个名为 Universe 的软件平台
  - 他们创造了AI游戏玩家，在Dota2中击败 99.95%玩家
  - 建立了能完成日本相扑摔跤或者踢足球的虚拟机器人

Mila (Bengio)

- Founder: Yoshua Bengio
- 以人为本的AI平台
  - 培养年轻人
  - 专注Deep Learning
- 创新创业



Vector Institute (Hinton)

- Founder: Geoffrey Hinton
- 以研究为主
  - 专注Deep Learning
- 创新创业

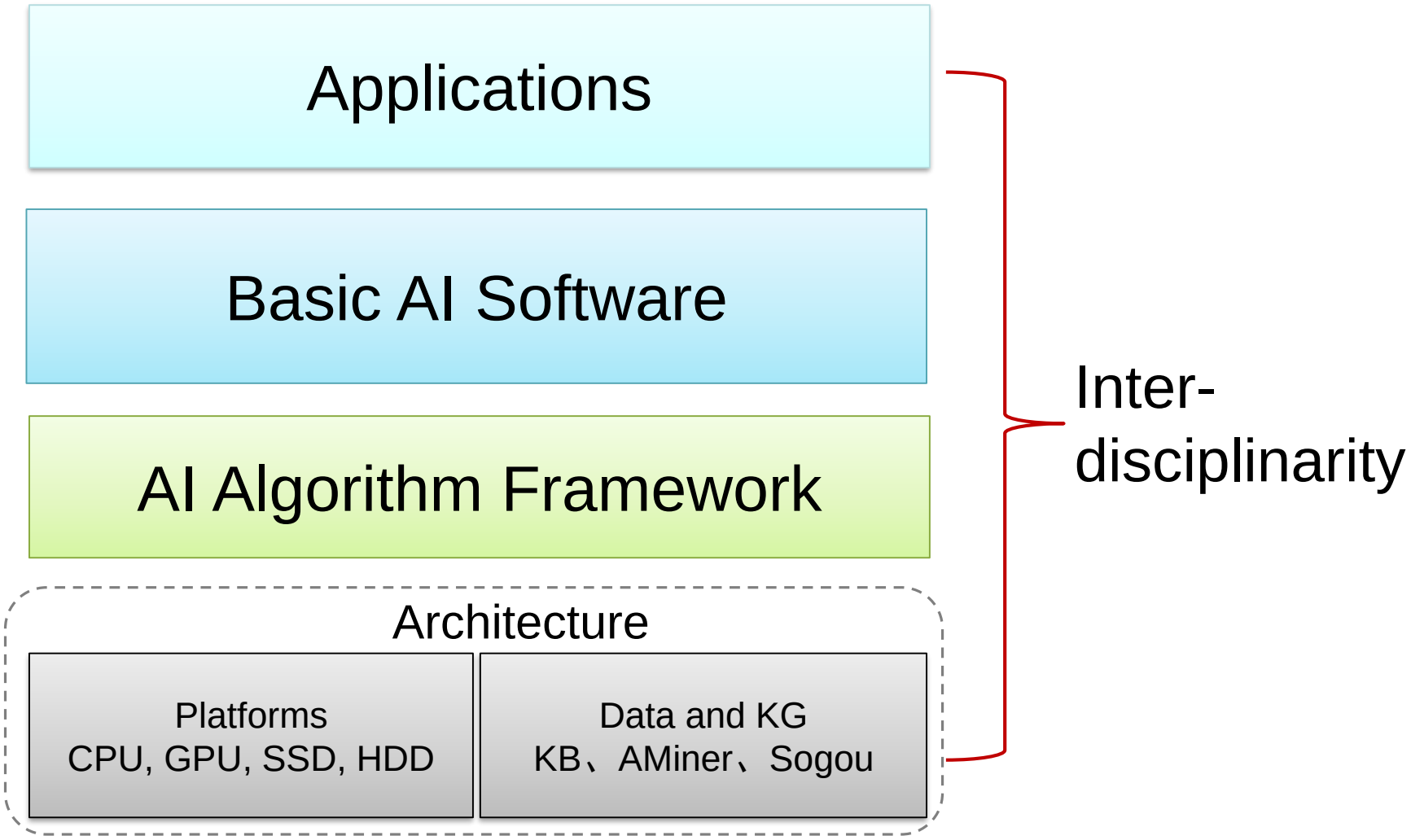


AI in Tsinghua

- In September, we established Tsinghua AI Institute
  - By Prof. Bo Zhang
- Focus
  - Fundamental theories and methodologies
  - Interdisciplinarity

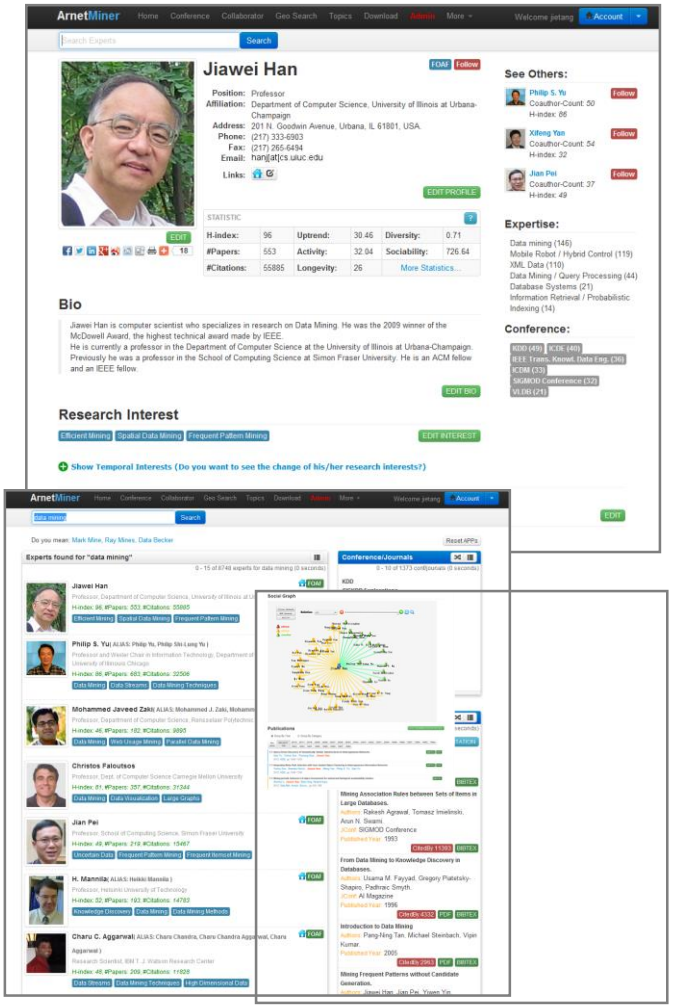


THU Algorithm and BigData Platform



THU Algorithm and BigData Platform

- Academic search and mining system —http://aminer.org
  - Online since 2006
  - >136 million researcher profiles
  - >201 million publication papers
  - >341 million requests
  - 100K IP access from 170 countries per month
- Deep analysis, mining, and search

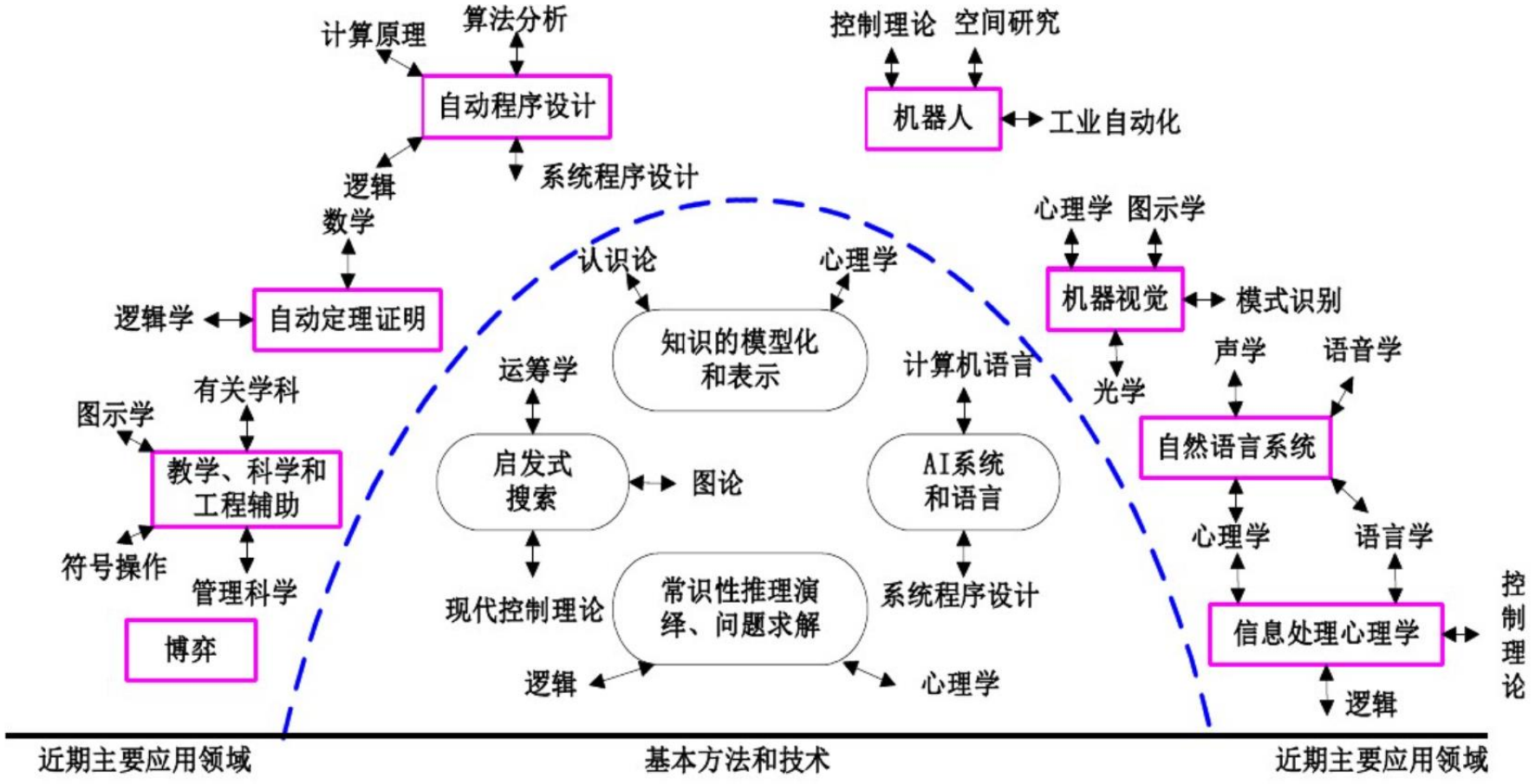


Beijing AI Institute

- In November, Beijing launched Beijing AI Institute
  - By Dr. Hongjiang Zhang
- Focus
  - ground-breaking research

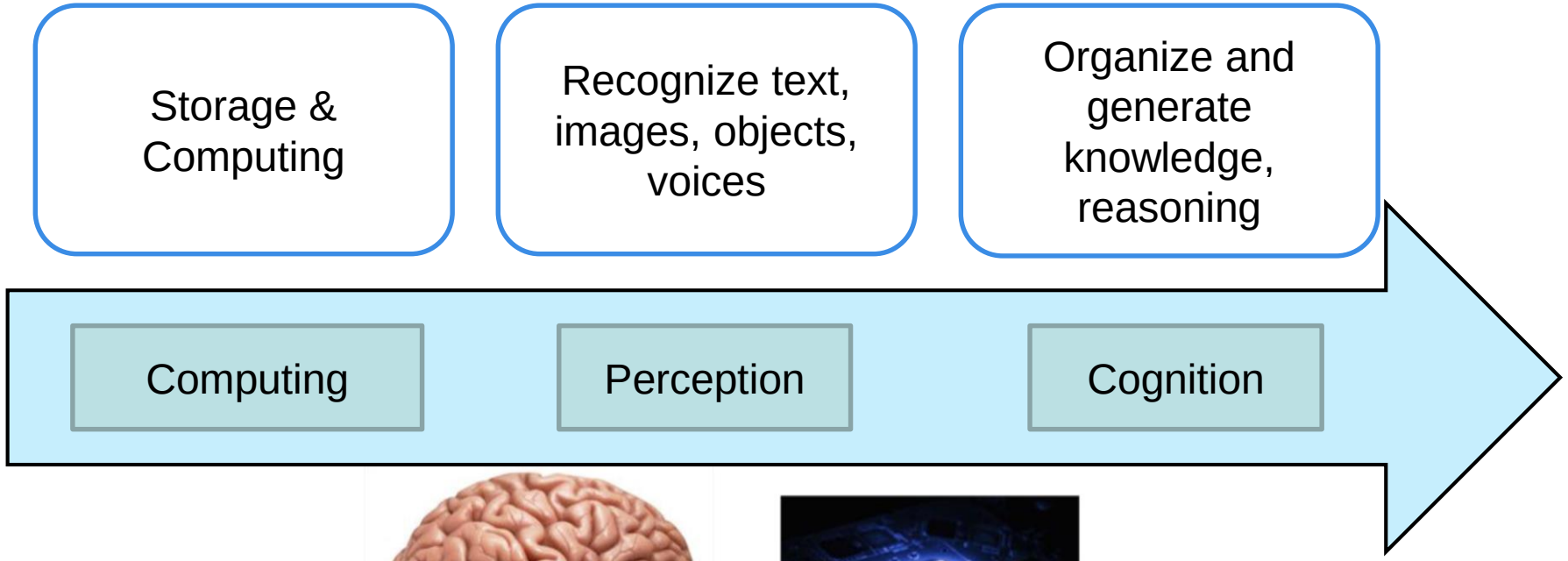


学科领域交叉与渗透

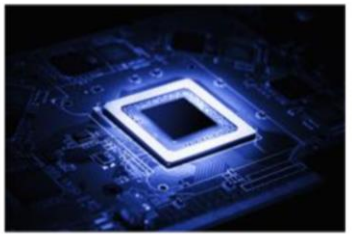

计算原理
算法分析
自动程序设计
逻辑
数学
系统程序设计
逻辑学
自动定理证明
有关学科
图示学
教学、科学和
工程辅助
符号操作
管理科学
博弈
控制理论
空间研究
机器人
工业自动化
心理学
图示学
机器视觉
模式识别
光学
声学
语音学
自然语言系统
心理学
语言学
信息处理心理学
逻辑
控制理论
认识论
心理学
知识的模型化
和表示
运筹学
启发式
搜索
图论
现代控制理论
常识性推理演
绎、问题求解
逻辑
心理学
计算机语言
AI系统
和语言
系统程序设计
近期主要应用领域
基本方法和技术
近期主要应用领域

AI Trend

- From perceptron to cognition







Stochastic vs Deterministic

Uncertainty!

Machine Learning Trends

Big Data

Deep Learning

Collective Learning

ML is hot & hot …Bengio@THU



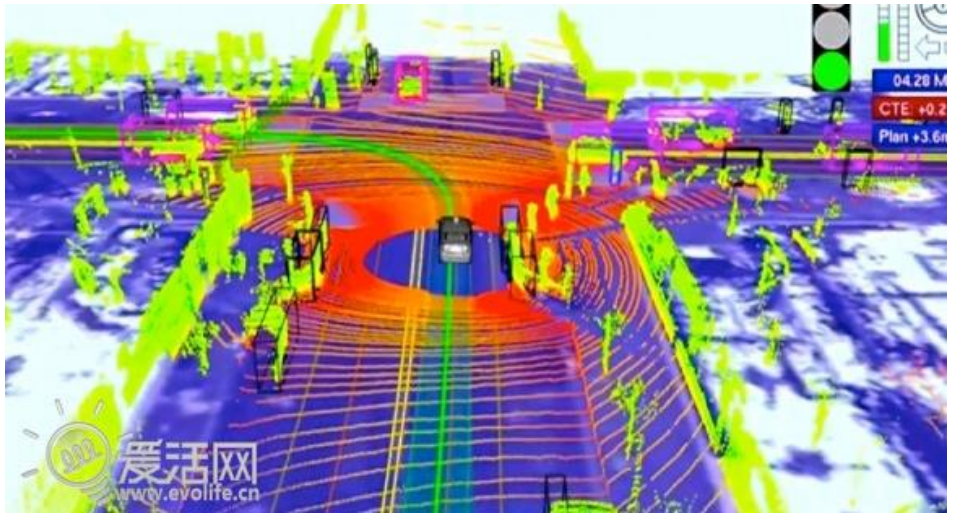


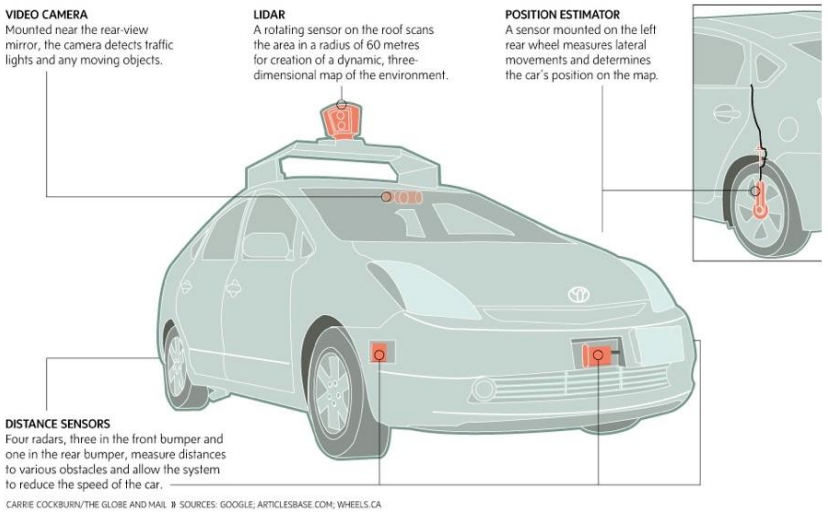
ML is hot & hot…@NIPS



BIG DATA—Google driverless car

Waymo—Head for market





Google's self-driving cars last month hit a 1,725,911-miles-driven milestone.

--2016/06

The system combines information gathered from Google Street View with artificial intelligence software that combines input from video cameras inside the car, a LIDAR sensor on top of the vehicle, radar sensors on the front of the vehicle and a position sensor attached to one of the rear wheels that helps locate the car's position on the map

Self-driving cars to create 1GB of data per second

http://www.fiercebigdata.com/story/self-driving-cars-create-1gb-data-second/2013-07-22

Deep Learning for Self-Driving Cars

6.S094: Deep Learning for Self-Driving Cars

Home DeepTesla DeepTraffic Leaderboard Resources View Profile Register Login

This class is an introduction to the practice of deep learning through the applied theme of building a self-driving car. It is open to beginners and is designed for those who are new to machine learning, but it can also benefit advanced researchers in the field looking for a practical overview of deep learning methods and their application.

Interested in the class? Here are some things you could do:

1. Register an account on the site to stay up-to-date. The material for the course is free and open to the public.
2. Follow the DeepTraffic and the DeepTesla tutorials, checkout the DeepTraffic leaderboard.
3. Design and evaluate neural networks for the DeepTraffic Game and the DeepTesla Simulation simulations all in the browser.
4. Watch the lectures and guest talks below.

Course Information:

- **Initial Offering:** Winter 2017
- **Instructor:** Lex Fridman
- **Contact:** deepcars@mit.edu



Lex Fridman
Instructor



Benedikt Jenik
TA



William Angell
TA



Spencer Dodd
TA



Dan Brown
TA

http://selfdrivingcars.mit.edu/

End-to-End Deep Learning for Self-Driving Cars

← Previous Next →



End-to-End Deep Learning for Self-Driving Cars

Share:

Posted on August 17, 2016 by Ben Firner, Beat Flepp, Karol Zieba, Larry Jackel, Mariusz Bojarski and Urs Muller | 7 Comments
Tagged Autonomous Vehicles, Deep Learning, DRIVE PX, Torch

In a new automotive application, we have used convolutional neural networks (CNNs) to map the raw pixels from a front-facing camera to the steering commands for a self-driving car. This powerful end-to-end approach means that with minimum training data from humans, the system learns to steer, with or without lane markings, on both local roads and highways. The system can also operate in areas with unclear visual guidance such as parking lots or unpaved roads.



SUBSCRIBE: RSS EMAIL
CONNECT:
FOLLOW @GPUCOMPUTING



RESOURCES

- About Parallel Forall
- NVIDIA Developer Forums
- Accelerated Computing Newsletter

RECENT POSTS

- An Even Easier Introduction to CUDA
- NVIDIA DIGITS Assists Alzheimer's Disease Prediction
- Beyond GPU Memory Limits with Unified

https://devblogs.nvidia.com/parallelforall/deep-learning-self-driving-cars/

Many companies

- Ford just invested $1 billion in self-driving cars
  - http://money.cnn.com/2017/02/10/technology/ford-argo-self-driving-cars/
- Uber joins race for driverless cars
  - Collective learning



Also Tesla, though it is still a bit mysterious technically.



Startup: Drive.ai, 地平线, pony.ai, 驭势科技…

Drive.ai uses deep learning to teach self-driving cars – and to give them a voice

Posted Aug 30, 2016 by Darrell Etherington (@etherington)

Next Story





DEEP LEARNING—IBM Watson

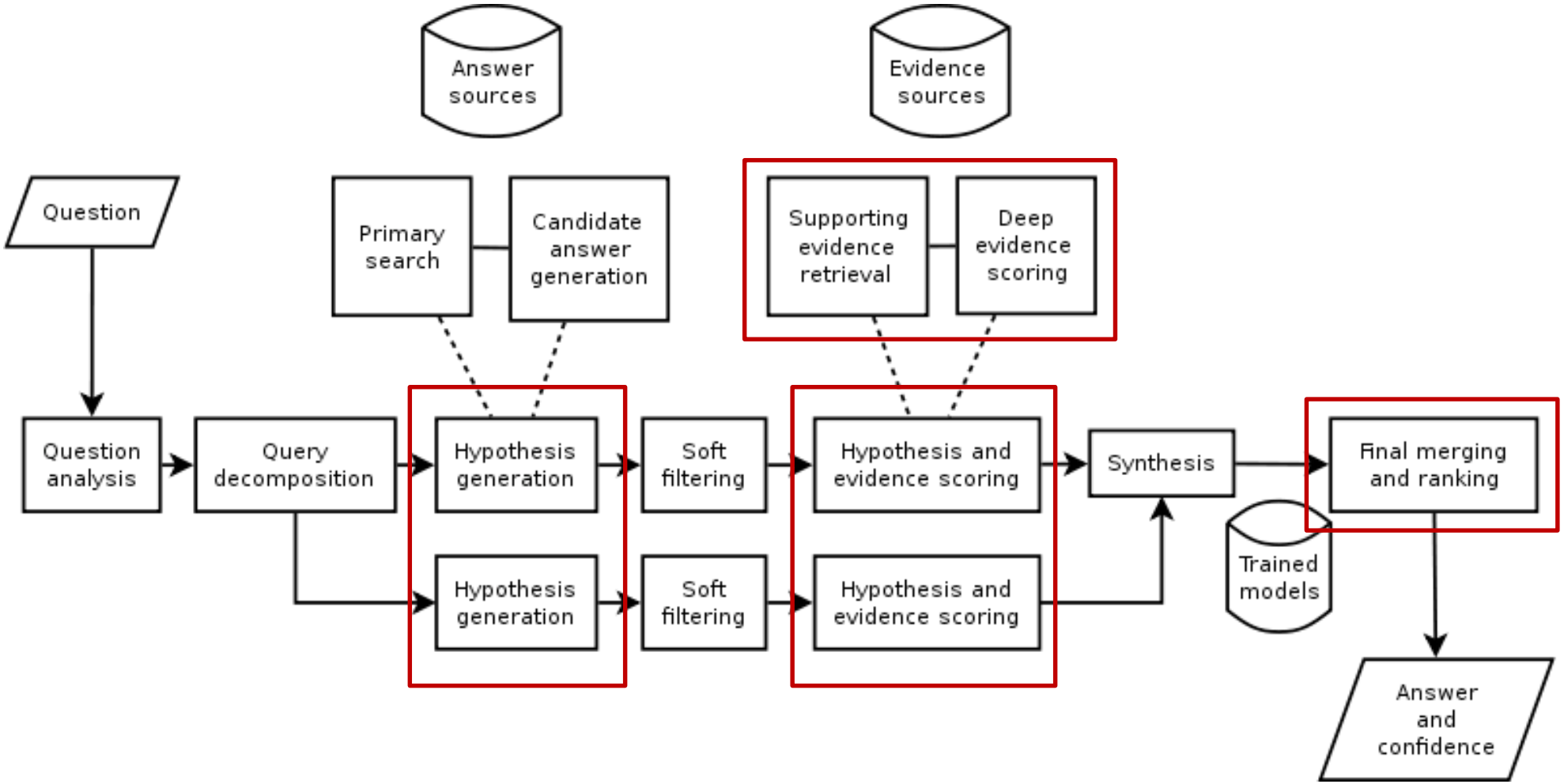


In 2011, as a test of its abilities, Watson competed on the quiz show Jeopardy!, in the show's only human-versus-machine match-up to date

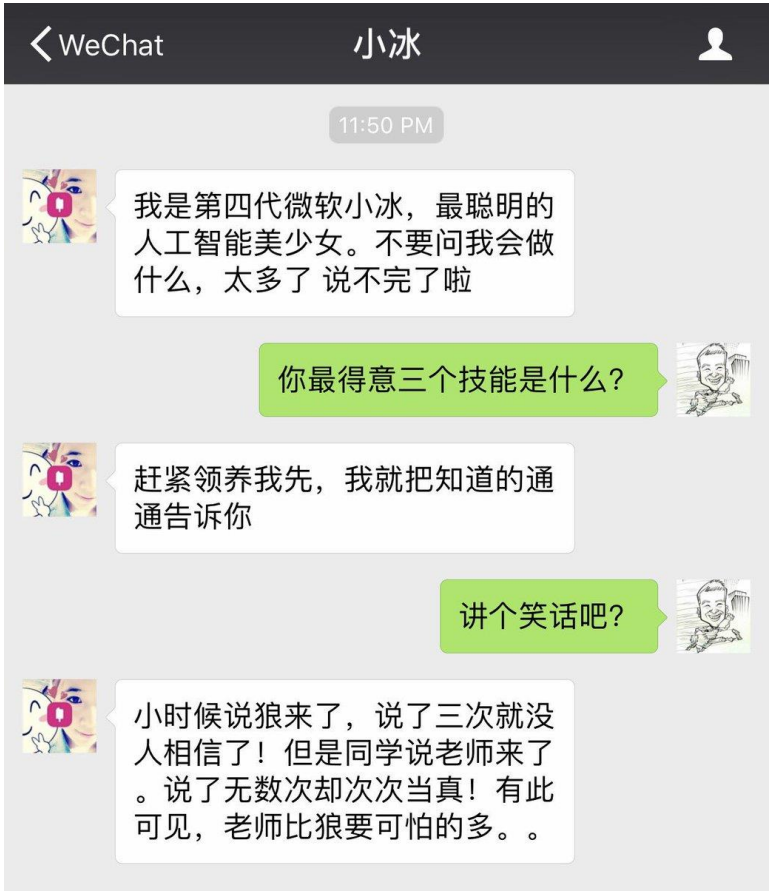
Watson is a Question answering (QA) computing system built by IBM. IBM describes it as "an application of advanced Natural Language Processing, Information Retrieval, Knowledge Representation and Reasoning, and Machine Learning technologies to the field of open domain question answering" which is "built on IBM's *DeepQA* technology for hypothesis generation, massive evidence gathering, analysis, and scoring

Watson had access to 200 million pages of structured and unstructured content consuming four terabytes of disk storage including the full text of Wikipedia, but was not connected to the Internet during the game. For each clue, Watson's three most probable responses were displayed on the television screen. Watson consistently outperformed its human opponents on the game's signaling device.

IBM Wastson (cont.)



XiaoIce


WeChat
小冰
11:50 PM
我是第四代微软小冰，最聪明的人工智能美少女。不要问我会做什么，太多了 说不完了啦
你最得意三个技能是什么?
赶紧领养我先，我就把知道的通通告诉你
讲个笑话吧?
小时候说狼来了，说了三次就没人相信了！但是同学说老师来了。说了无数次却次次当真！有此可见，老师比狼要可怕的多。。


你喜欢奥巴马还是特朗普？
。。。你丑你先说呀！
。。。You are ugly you say it!
Microsoft Translator

SQuAD

What is SQuAD?

Stanford **Q**uestion **A**nswering **D**ataset (SQuAD) is a new reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or *span*, from the corresponding reading passage. With 100,000+ question-answer pairs on 500+ articles, SQuAD is significantly larger than previous reading comprehension datasets.

Explore SQuAD and model predictions

Read the paper (Rajpurkar et al. '16)

Getting Started

We've built a few resources to help you get started with the dataset.
Download a copy of the dataset (distributed under the CC BY-SA 4.0 license):

Training Set v1.1 (30 MB)

Dev Set v1.1 (5 MB)

To evaluate your models, we have also made available the evaluation script we will use for official evaluation, along with a sample prediction file that the script will take as input. To run the evaluation, use `python evaluate-v1.1.py <path_to_dev-v1.1> <path_to_predictions>`.

Evaluation Script v1.1

Sample Prediction File (on Dev v1.1)

Once you have a built a model that works to your expectations on the dev set, you submit it to get official scores on the dev and a hidden test set. To preserve the

Leaderboard

Since the release of our dataset, the community has made rapid progress! Here are the ExactMatch (EM) and F1 scores of the best models evaluated on the test set of v1.1. Will your model outperform humans on the QA task?

Rank	Model	EM	F1
	Human Performance *Stanford University* (Rajpurkar et al. '16)	82.304	91.221
1 Jan 22, 2018	Hybrid AoA Reader (ensemble) *Joint Laboratory of HIT and iFLYTEK Research*	82.482	**89.281**
1 Feb 19, 2018	Reinforced Mnemonic Reader + A2D (ensemble model) *Microsoft Research Asia & NUDT*	**82.849**	88.764
2 Feb 02, 2018	Reinforced Mnemonic Reader (ensemble model) *NUDT and Fudan University* https://arxiv.org/abs/1705.02798	82.283	88.533
2 Jan 03, 2018	r-net+ (ensemble) *Microsoft Research Asia*	82.650	88.493
2 Jan 05, 2018	SLQA+ (ensemble) *Alibaba iDST NLP*	82.440	88.607
3 Dec 17, 2017	r-net (ensemble) *Microsoft Research Asia* http://aka.ms/rnet	82.136	88.126
3 Dec 22, 2017	AttentionReader+ (ensemble) *Tencent DPDAC NLP*	81.790	88.163
4 Feb 19, 2018	Reinforced Mnemonic Reader + A2D (single model) *Microsoft Research Asia & NUDT*	80.919	87.492
4 Nov 17, 2017	BiDAF + Self Attention + ELMo (ensemble) *Allen Institute for Artificial Intelligence*	81.003	87.432

LittleMU (小木)


学堂小木
如何注册学堂
1、pc端打开http://www.xuetangx.com，右上角点击注册，您可以使用邮箱或手机号进行注册。或者使用第三方账户，比如新浪微 …
来自问答库
二叉树是什么
二叉树：
在计算机科学中，二叉树是每个节点最多有两个子树的树结构。通常子树被称作"左子树"（left subtree）和"右子树"（right subtree）。二叉树常被用于实现二叉查找树和二叉堆。二叉树的每个结点至多只有二棵子树(不存在度大于2的结点)，二叉树的子树有左右之分，次序不能颠倒。二叉树的第i层至多有2^{i-1}个结点；深度为k的二叉树至多有2^k-1个结点；对任何一棵二叉树T，如果其终端结点数为n_0，度为2的结点数为n_2，则n_0=n_2+1。一棵深度为k，且有2^k-1个节点称之为满二叉树；深度为k，有n个节点的二叉树，当且仅当其每一个节点都与深度为k的满二叉树中，序号为1至n的节点对应时，称之为完全二叉树。
另见：
满二叉树
…同学们好，我们这一节介绍树的一种特殊，但又不失代表性的特例。
在这里提问，按enter(回车键)发送
30240184X 数据结构(上)(自主模式)
讨论区
Wiki
课程进度
常见问题
教材
习题解析
05E4-1 次序
小木提问：什么是层次遍历？ (2个同学已问过类似问题)
5.二叉树
(e4) 层次遍历
0:16 / 3:34
字幕
高清
1.00x
下载字幕
.txt

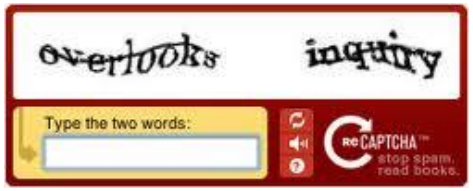


launched in 2013

MECHANISM—Collective Learning

- Learning in Social Networks
  - CAPTCHA + OCR

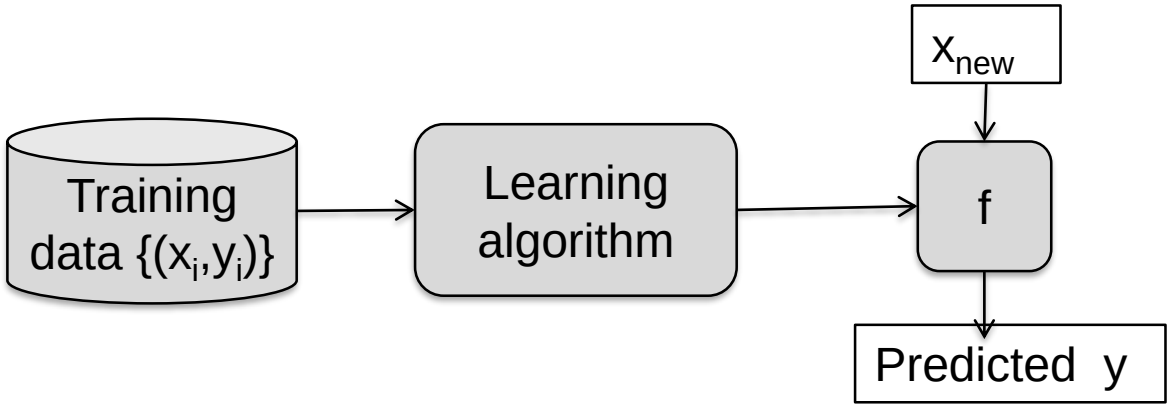
Lius von Ahn
--AP@CMU

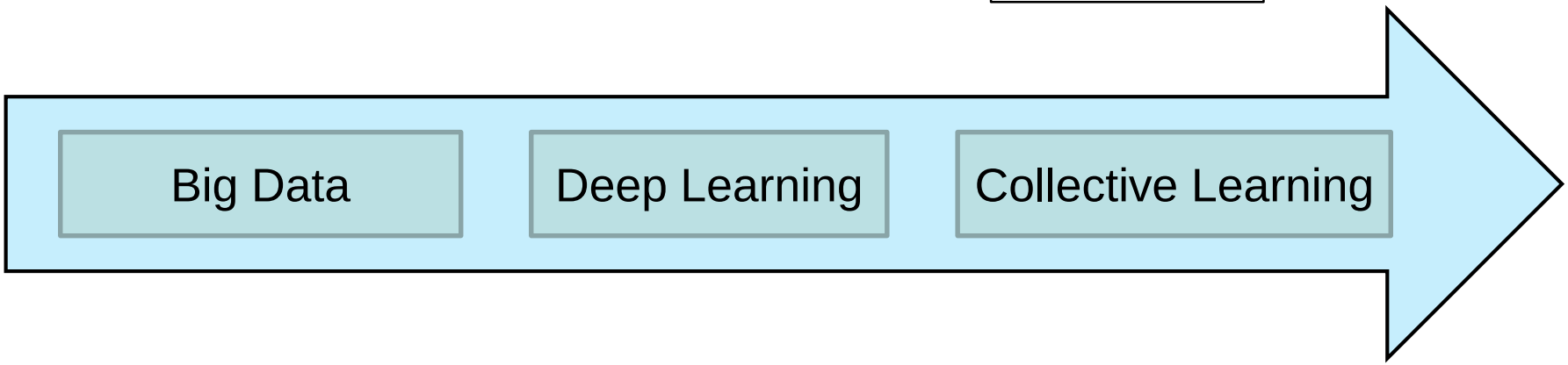


  - DUOLINGO

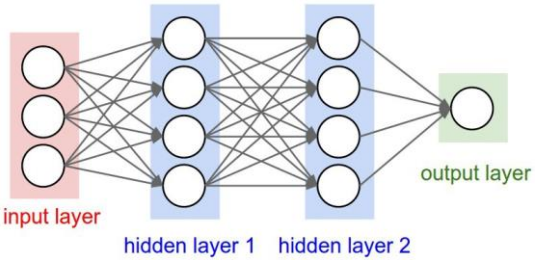


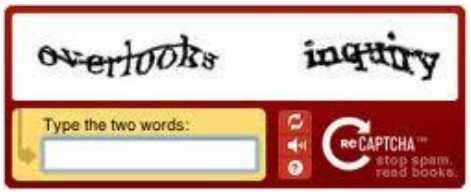
Machine Learning Trend





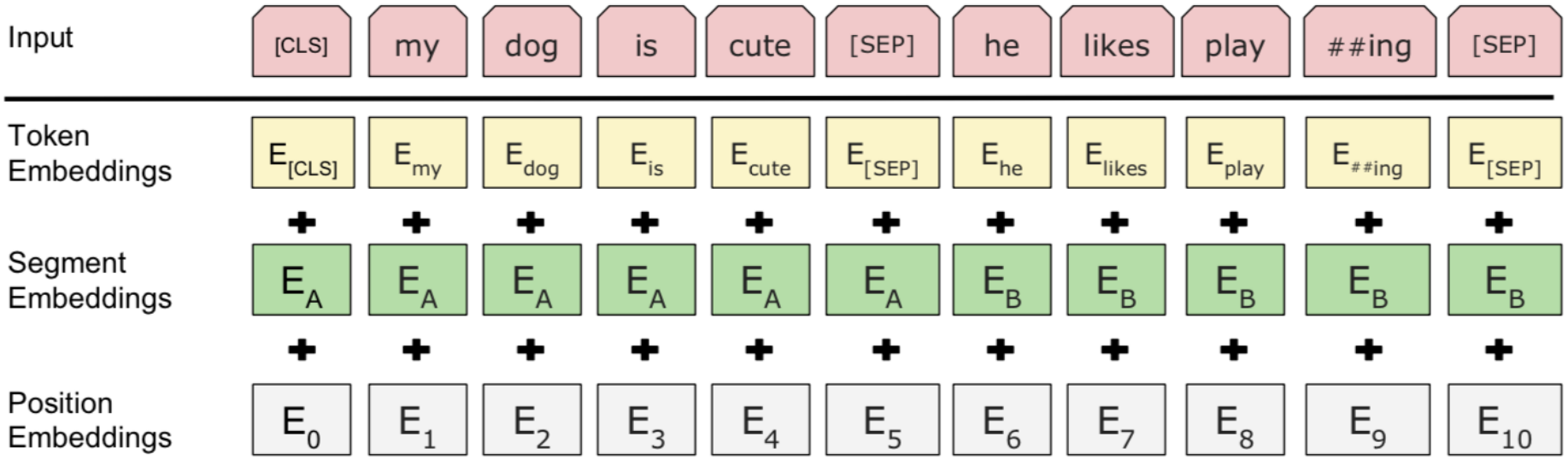






Recent milestones (2018/2019)

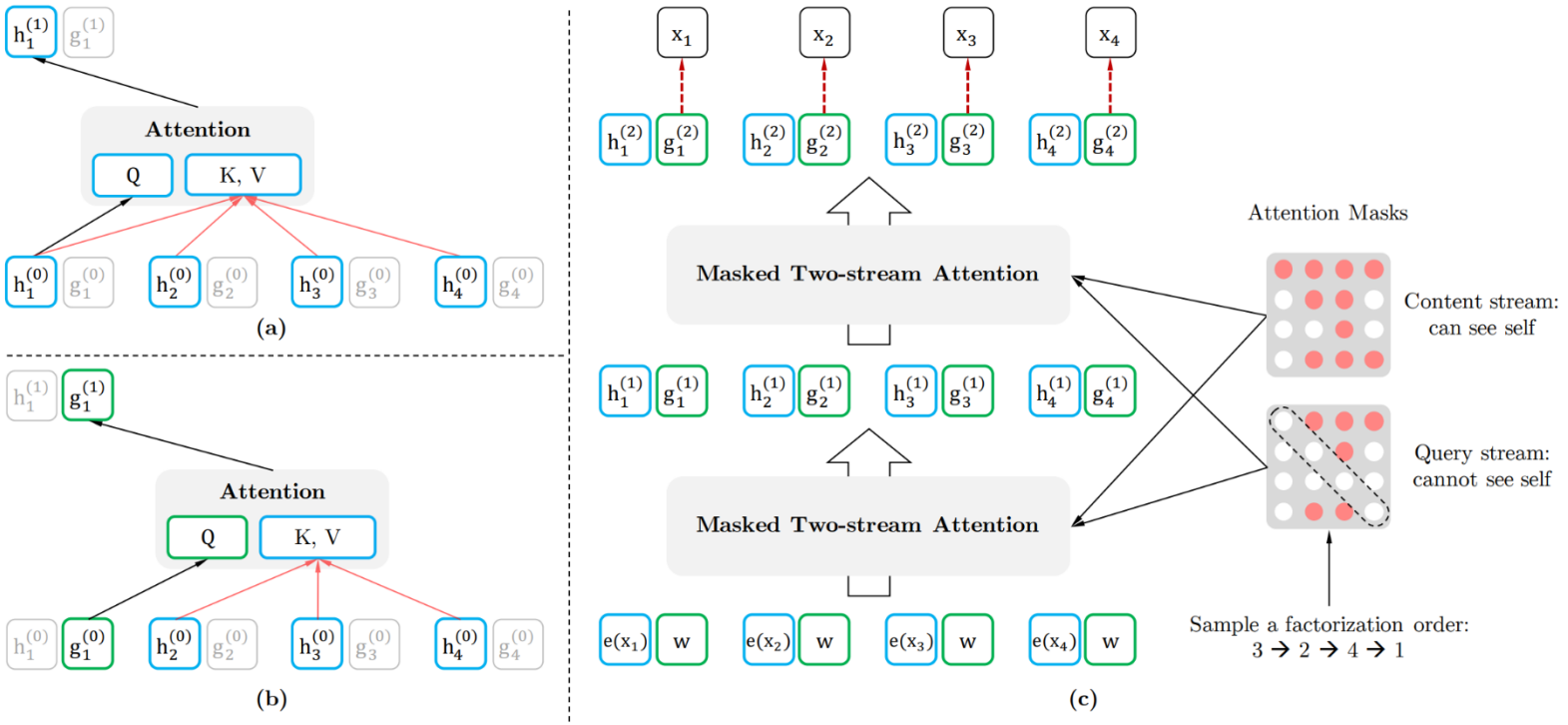
BERT



- Pre-train
- Fine tune
- Beat all state-of-the-arts on 11 NLP tasks in 2018

https://arxiv.org/pdf/1810.04805.pdf

XLNet



- Autoregressive Model
- Beat BERT in 2019

https://arxiv.org/pdf/1906.08237.pdf

ALBERT

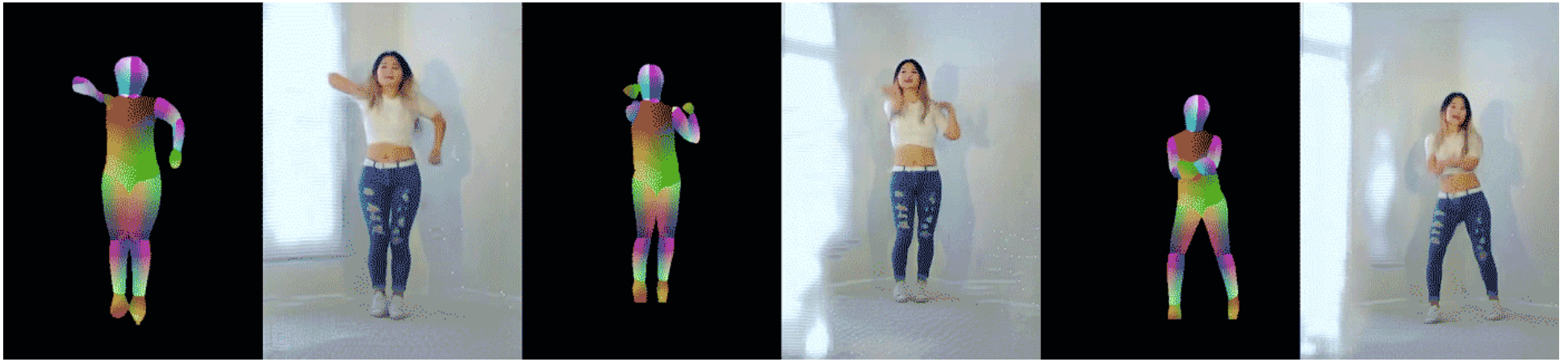
Model		Parameters	SQuAD1.1	SQuAD2.0	MNLI	SST-2	RACE	Avg	Speedup
BERT	base	108M	90.5/83.3	80.3/77.3	84.1	91.7	68.3	82.1	17.7x
	large	334M	92.4/85.8	83.9/80.8	85.8	92.2	73.8	85.1	3.8x
	xlarge	1270M	86.3/77.9	73.8/70.5	80.5	87.8	39.7	76.7	1.0
ALBERT	base	12M	89.3/82.1	79.1/76.1	81.9	89.4	63.5	80.1	21.1x
	large	18M	90.9/84.1	82.1/79.0	83.8	90.6	68.4	82.4	6.5x
	xlarge	59M	93.0/86.5	85.9/83.1	85.4	91.9	73.9	85.5	2.4x
	xxlarge	233M	**94.1/88.3**	**88.1/85.1**	**88.0**	**95.2**	**82.3**	**88.7**	1.2x

- A Lite BERT
- Parameter-reduction techniques
- Beat XLNet and all the others

https://arxiv.org/pdf/1909.11942.pdf

Video-to-Video Synthesis

- The best video synthesis performance

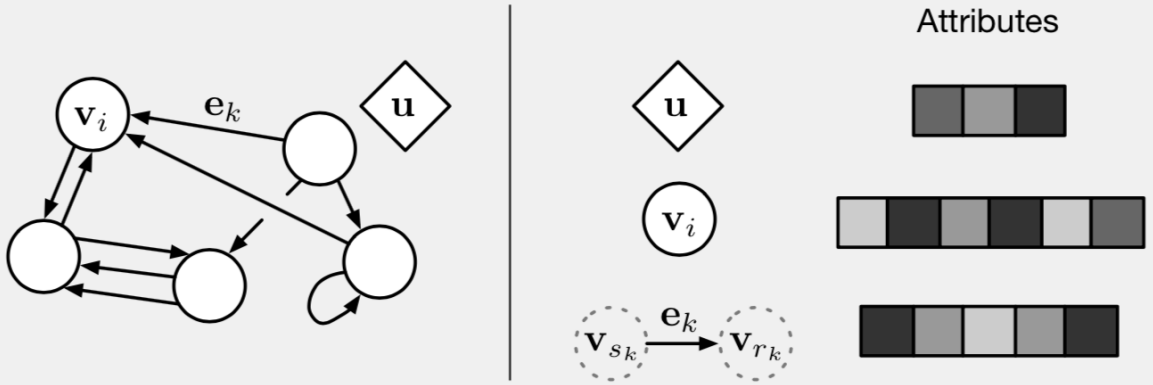


https://arxiv.org/abs/1808.06601

graph_net

- By DeepMind

Box 3: Our definition of "graph"



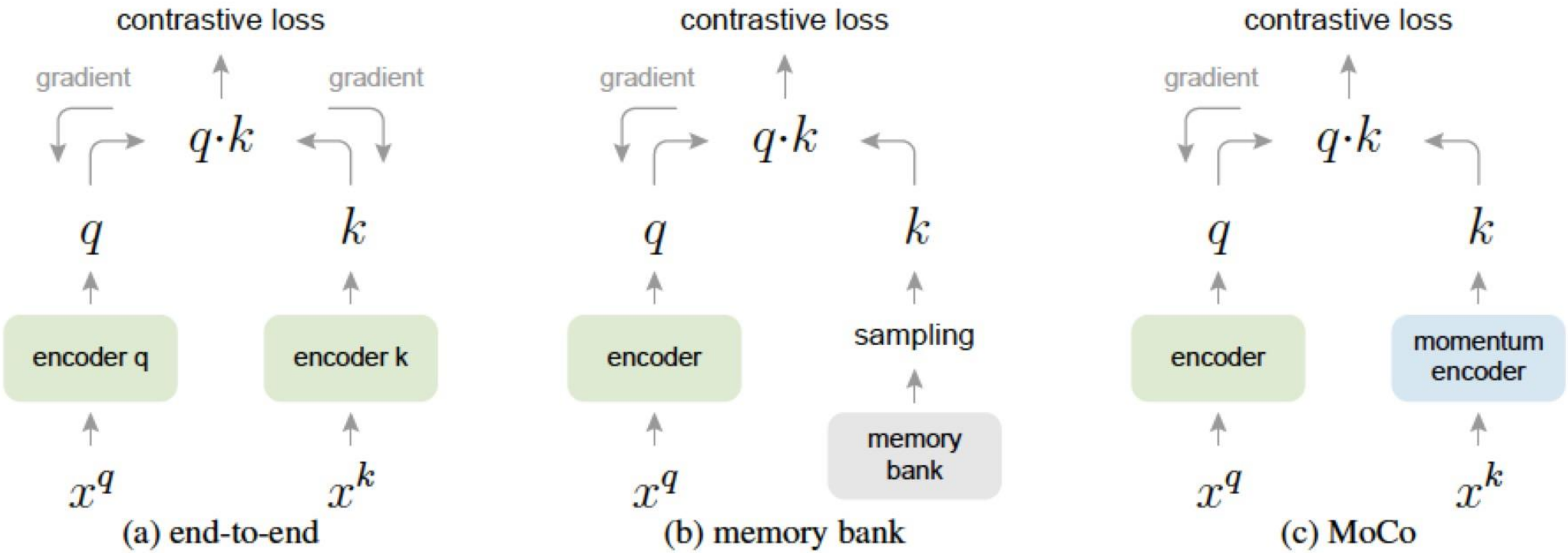
Here we use "graph" to mean a directed, attributed multi-graph with a global attribute. In our terminology, a node is denoted as $\mathbf{v}_i$, an edge as $\mathbf{e}_k$, and the global attributes as $\mathbf{u}$. We also use s_k and r_k to indicate the indices of the sender and receiver nodes (see below), respectively, for edge k. To be more precise, we define these terms as:

Directed : one-way edges, from a "sender" node to a "receiver" node.
Attribute : properties that can be encoded as a vector, set, or even another graph.
Attributed : edges and vertices have attributes associated with them.
Global attribute : a graph-level attribute.
Multi-graph : there can be more than one edge between vertices, including self-edges.

Figure 2 shows a variety of different types of graphs corresponding to real data that we may be interested in modeling, including physical systems, molecules, images, and text.

https://arxiv.org/abs/1806.01261

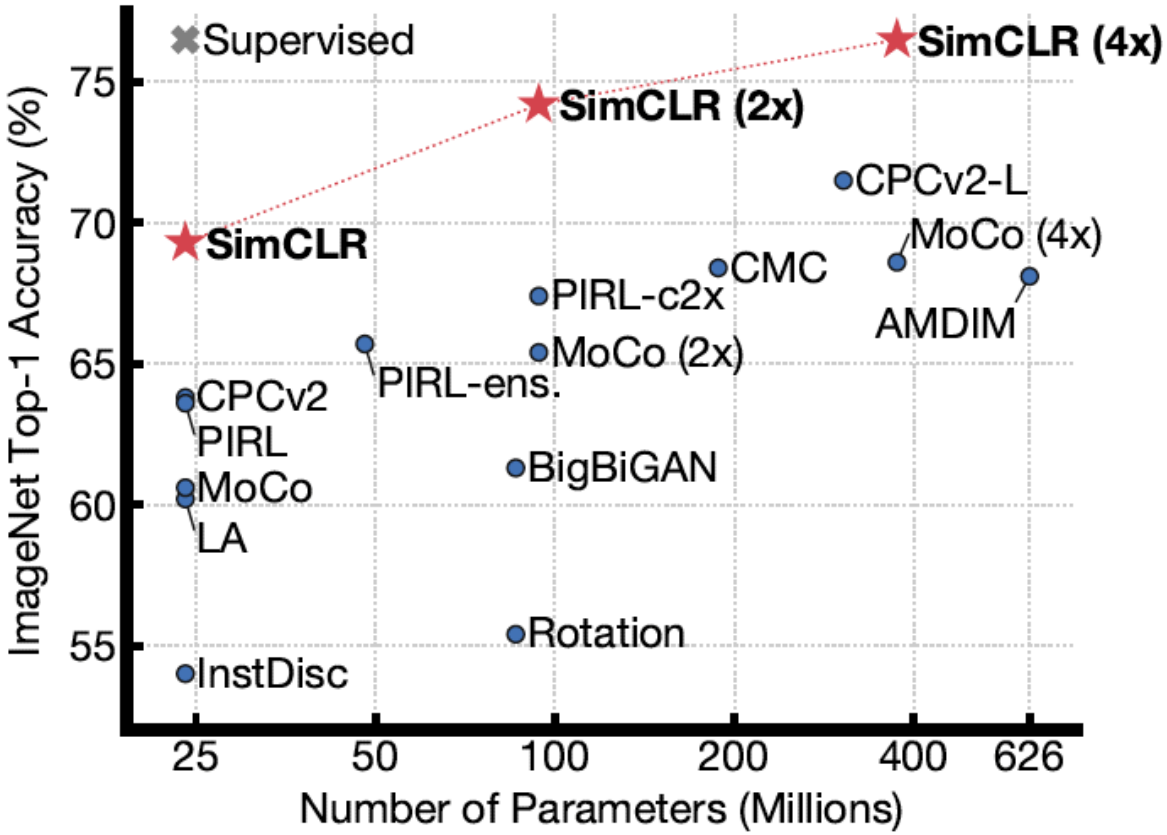
MoCo



- Unsupervised visual representation learning
- Momentum contrastive learning
- Outperform its supervised pre-training counterparts

https://arxiv.org/abs/1911.05722

SimCLR



- Simplified contrastive learning framework
- Outperform previous self-supervised and semi-supervised methods on ImageNet

https://arxiv.org/abs/2002.05709



Overview of Class

投票：你了解/熟悉的先修知识

A 概率统计

B 线性代数

C 最优化/凸优化

D 机器学习

E 会一种编程语言（C/C++/Python/Java）

F 没有接触过机器学习

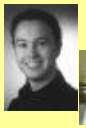
The State of Machine Learning

Classification models

Decision tree 
Bayesian classifier
Perceptron
Neural networks

Maximal Margin

SVM (Vapnik)



Maximal margin sequential learning

M3N network
(Ben Taskar)


Deep learning

Deep belief networks




Topic models

PLSI, LDA, etc.





Sequential learning

HMM

Sequential learning

MEMM, CRF,
voted perceptron




Graphical models

Factor graph,
Exponential model






For more: http://arnetminer.org/expertisesearch.do?keyword=Machine%20Learning

Agenda

- Overview
  - Course overview
  - ML Introduction

Agenda

- Machine Learning Basics
  - Data representation and machine learning
    - Evaluation
    - Perceptron, Bayesian classification, Decision tree, etc.
  - Unsupervised learning
    - Evaluation
    - K-means
  - Linear Algebra
  - Probability and Information Theory
  - Numerical Computations

Agenda

- Machine Learning Basics
  - Data representation and machine learning
    - Evaluation
    - Perceptron, Bayesian classification, Decision tree, etc.
  - Unsupervised learning
    - Evaluation
    - K-means
  - Linear Algebra
  - Probability and Information Theory
  - Numerical Computations

Agenda (6 hours)

- Deep Forward Networks
  - Gradient-based learning
  - Architecture design
  - Back propagation

Agenda

- Regularization for Deep Learning
  - Parameter norm penalties
  - Semi-supervised learning
  - Multitask learning
  - Bagging and other algorithms
  - Dropout
  - Adversarial training

Agenda

- Training Deep Models
  - Basic algorithms
  - Parameter initialization strategies
  - Approximate second-order methods
  - Optimization strategies and meta-algorithms

Agenda

- Convolutional Networks
  - Convolution operation
  - Pooling
  - Variants
  - Learning algorithms

Agenda

- Sequential Neural Networks
  - Recurrent neural networks
  - Bidirectional RNNs
  - Sequence-to-sequence
  - Deep recurrent networks
  - Recursive neural networks
  - Long short-term memory (LSTM)

Agenda

- Sequential Neural Networks
  - Recurrent neural networks
  - Bidirectional RNNs
  - Sequence-to-sequence
  - Deep recurrent networks
  - Recursive neural networks
  - Long short-term memory (LSTM)

Agenda

- Hands-on Tutorials
  - Tensorflow: https://www.tensorflow.org/
  - Keras: https://keras.io/
  - CNTK: https://www.microsoft.com/en-us/research/product/cognitive-toolkit/
  - MXNet: http://mxnet.io/
  - Torch: http://torch.ch/
  - Theano: http://deeplearning.net/software/theano/

Agenda

- Linear Factor Models
  - Probabilistic PCA and Factor Analysis
  - Independent Component Analysis
  - Sparse Coding
  - Manifold Interpretation

Agenda

- Autoencoders
  - Undercomplete autoencoders
  - Regularized autoencoders
  - Representational power
  - Stochastic encoders and decoders
  - Learning manifolds with autoencoders
  - Contrastive autoencoders
  - Predictive sparse decomposition

Agenda

- Representation Learning
  - Greedy Layer-wise unsupervised pretraining
  - Transfer learning with domain adaptation
  - Distributed representation
  - Exponential gains from depth

Agenda

- Structured probabilistic models
  - Graphical models
  - Sampling from graphical models
  - Learning algorithms
  - The deep learning approach

Agenda

- MCMC and Approximate Inference
  - Stochastic maximum likelihood
  - Contrastive divergence
  - Score matching and ratio matching
  - Expectation maximization
  - MAP inference
  - Variational inference

Agenda

- Reinforcement Learning
  - Multi-armed Bandits
  - Finite Markov Decision Processes
  - On-policy Prediction
  - Off-policy Prediction

Agenda

- Deep Generative Models
  - (Restricted) Boltzmann Machines
  - Deep Belief Networks
  - Deep Boltzmann Machines
  - Convolutional Boltzmann Machines
  - Directed generative models
  - Generative stochastic networks

Prerequisites

- Programming and other basic CS skills
  - C, C++
  - Java, C#, .NET
  - Perl, Python
- Theories
  - Data representation (e.g., vector space model, language model)
  - Basic probability theory (e.g., likelihood, conditional probability, posterior probability, Bayes)
  - Basic linear algebra (e.g., linear transformations, eigenvalues, least-squares best fit)

投票：你想了解哪些机器学习知识？

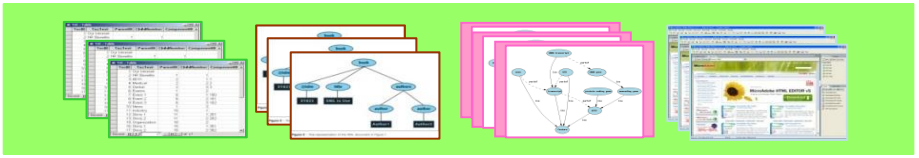
- A 监督学习
- B 无监督学习
- C 学习理论
- D 概率图模型
- E 深度学习
- F 强化学习
- G 集成学习
- F 其他 ____

Potential achievements

- Able to understand the underlying principles of classical ML algorithms
- Able to apply right ML algorithms to the applications at you hand
- Able to design effective ML algorithms to solve new problems

- This year, we will provide MS Azure platform for practicing large-scale deep learning…
  - Please register a new outlook account first…

Overview of Class (cont.)



Supervised learning

Unsupervised learning

Machine Learning

Lectures on ML

Generative learning

Topic model
Non-parametric
Representation learning
Convolutional networks

Discriminative learning

SVM
CRF
Max-margin
Sparse coding
Adversial learning

Seminar-style class

Presentation

Algorithm design

Output

New algorithms
Paper
Fun!

Evaluation

Homework
(50 points)

Correctness
efficiency

Projects
(50 points)

Novelty (20)
Completeness (30)

In this course, we try to learn something and try to achieve something…

Homework

- Homework (40’)
  - Hw1: one math
  - Hw2: one algorithm implementation
    - Bonus: LDA by Gibbs sampling
  - Hw3: one application
  - Hw4: one challenge (Coronavirus infection prediction)
- Project (50’)
  - Prj1: To design new ML algs. to real problems
  - Prj2: To solve novel problems using ML algs.
- Topic Survey (10’)
  - One-week workload, publishable
- Voluntary weekly share for novel papers/techniques (extra points)

4+1+1

For project and survey, you can choose to team up (<=2 members)

Final Presentation



References

- Handbooks:
  - Yoshua Bengio, Ian J. Goodfellow, and Aaron Courville. Deep Learning. 2016.
  - Christopher M. Bishop. *Pattern Recognition and Machine Learning*, Springer, 2007.
- Additional references
  - John Hopcroft. Computer Science Theory for the Information Age. 2011.
  - Daphne Koller and Nir Friedman. *Probabilistic Graphical Models.* MIT Press, 2009
  - Michael I. Jordan. *An Introduction to Probabilistic Graphical Models.* University of California, Berkeley. June 30, 2003.
  - Martin J. Wainwright and Michael I. Jordan. *Graphical Models, Exponential Families, and Variational Inference*, Foundations and Trends in Machine Learning, V1 (1-2), 2008.
  - Trevor Hastie, Robert Tibshirani, Jerome Friedman. *Elements of Statistical Learning*. Springer, 2003.
  - Yoshua Bengio. *Learning Deep Architectures for AI*. Foundations and Trends in Machine Learning, V2 (1), 2009.
  - David J.C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

Feedback

- Any suggestions?
- Any problems?
- Anything else you want to learn in class?

Say something in WeChat!



Thanks!

HP: http://keg.cs.tsinghua.edu.cn/jietang/

Anything you want to say…