上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

上海交通大学
约翰·霍普克罗夫特计算机科学中心
John Hopcroft Center for Computer Science

VE445: Introduction to Machine Learning

Shuai Li

John Hopcroft Center, Shanghai Jiao Tong University

https://shuaili8.github.io

JOINT INSTITUTE
交大密西根学院

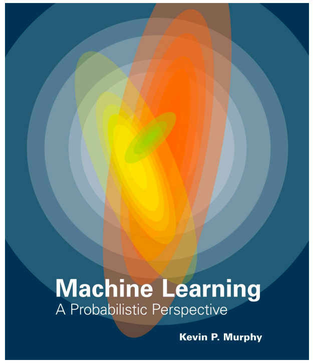
https://shuaili8.github.io/Teaching/VE445/index.html

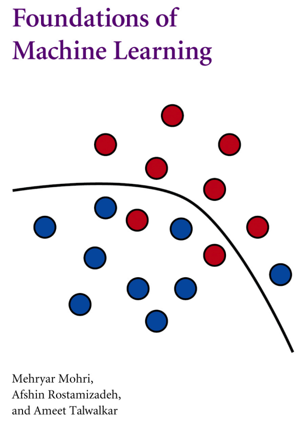
Teaching Assistant

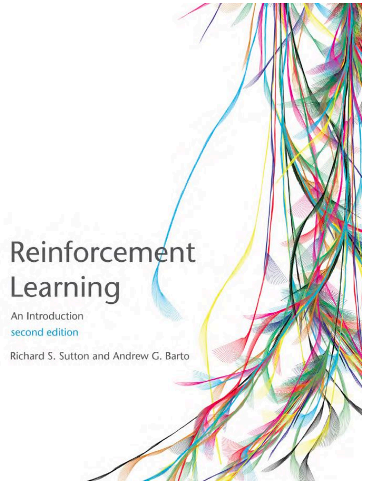
- Jingying Wang (王菁滢)
  - Email: wjymonica@sjtu.edu.cn
  - Senior student major in ECE
  - Research on Crowd Counting, Hand Pose Estimation, and Action Quality Assessment
  - RC time & location: TBD
  - OH time & location: TBD

References (will add more during course)

- Machine Learning : A Probabilistic Perspective
  by Kevin P. Murphy
- Foundations of Machine Learning
  by Mehryar Mohri, Afshin Rostamizadeh and Ameet Talwalkar
- Reinforcement Learning: An Introduction
  by Richard S. Sutton and Andrew G. Barto
- 周志华《机器学习》清华大学出版社， 2016.







Goal

- Know what is machine learning and what it usually covers
- Familiar and understand popular machine learning algorithms
- Be able to build machine learning models in applications
  - Know which algorithms to adopt and when to adopt
- Could start machine learning research
  - Read machine learning papers

Prerequisites

- Basic computer science principles
  - Big-O notation
  - Comfortably write non-trivial code in Python/numpy
- Probability
  - Random Variables
  - Expectations
  - Distributions
- Linear Algebra & Multivariate/Matrix Calculus
  - Gradients and Hessians
  - Eigenvalue/vector

Slide credit: Anand Avati

Grading

- Labs/HomeWorks: 20%

- Midterm: 25%

- Project: 20%

- Final exam: 35%

Honor code

- Discussions are encouraged
- Independently write-up homework and code
- Same reports and homework will be reported

Course Outline

- Supervised learning
  - Linear/Logistic regression
  - SVM and Kernel methods
  - Generative models
  - Tree ensembles
- Deep learning
  - Neural Networks
  - Backpropagation
  - Convolutional Neural Network
  - Recurrent Neural Network
- Unsupervised learning
  - K-means, PCA, EM, GMM
- Reinforcement learning
  - Multi-armed bandits
  - MDP
  - Bellman equations
  - Q-learning
- Learning theory
  - PAC, VC-dimension, bias-variance decomposition

Introduction

What is Artificial Intelligence

- Describe machines (or computers) that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem solving”.

  --Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited.

- Intelligence is the computational part of the ability to achieve goals in the world. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.

  --By Prof. John McCarthy of Stanford University

  -- http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html

What is Machine Learning

- Term “Machine Learning” coined by Arthur Samuel in 1959.
  - Samuel Checkers-playing Program
- Common definition (by Tom Mitchell):
  - ***Machine Learning is the study of computer algorithms that improve automatically through experience***
- Subfield of Artificial Intelligence (AI)
  - The hottest subfield - reinvigorated interest in AI due to deep learning!

Slide credit: Anand Avati

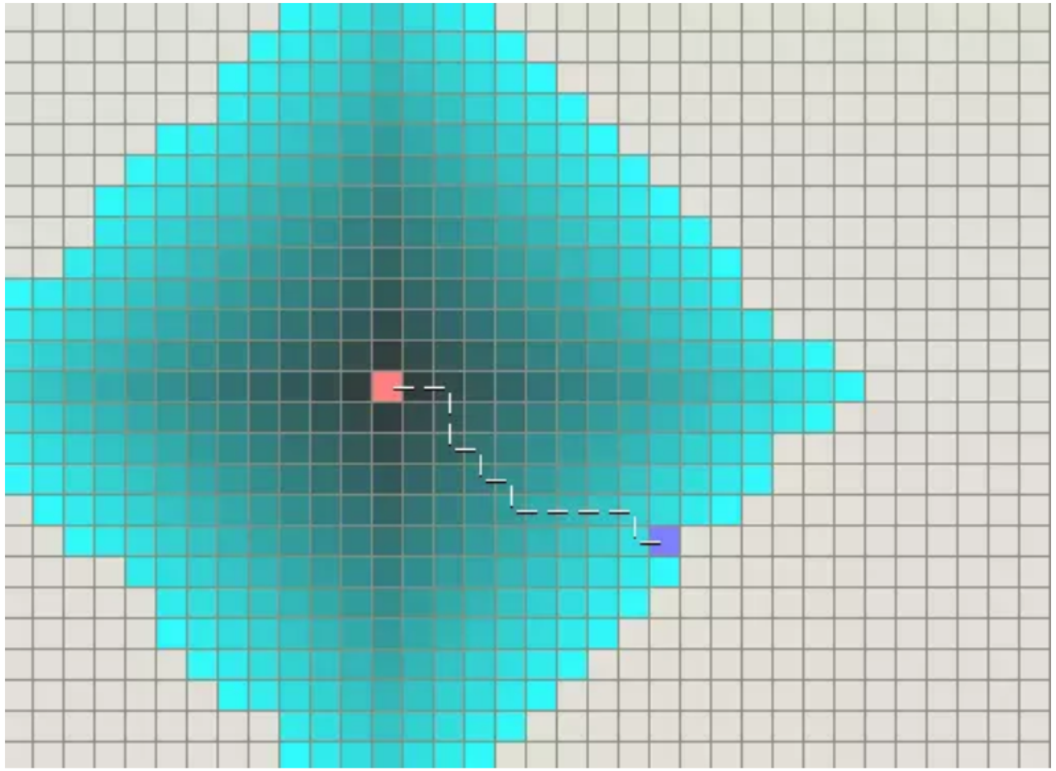
An example of AI but is not machine learning

- A* search algorithm
  - Objective: Find the shortest path between two nodes of a weighted graph

- Compare with Breadth First Searching and Greedy Searching

Breadth First Searching

- Pink: start point, Purple: end point;
- Blue: visited points, the darker the earlier



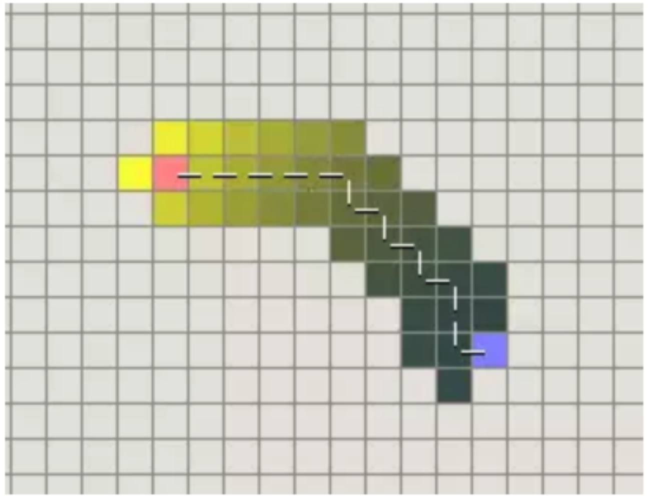
Each time it visits, or expand the point with least $g(n)$ value

- $g(n)$ is the distance from start point to point n.

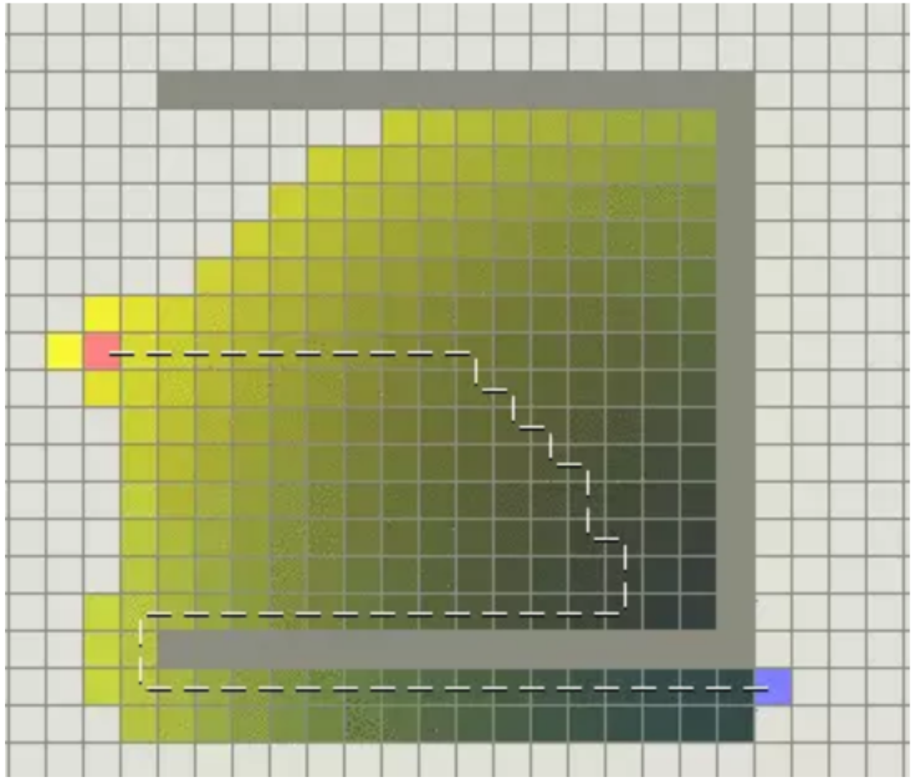
Short comings: computing burden is too high, it visited too many points before getting the end point.

Greedy Searching

- Each time it visit or expand the point with least $h(n)$ value
  - $h(n)$ is the distance from point n to end point. It works fine when there is no obstacles.

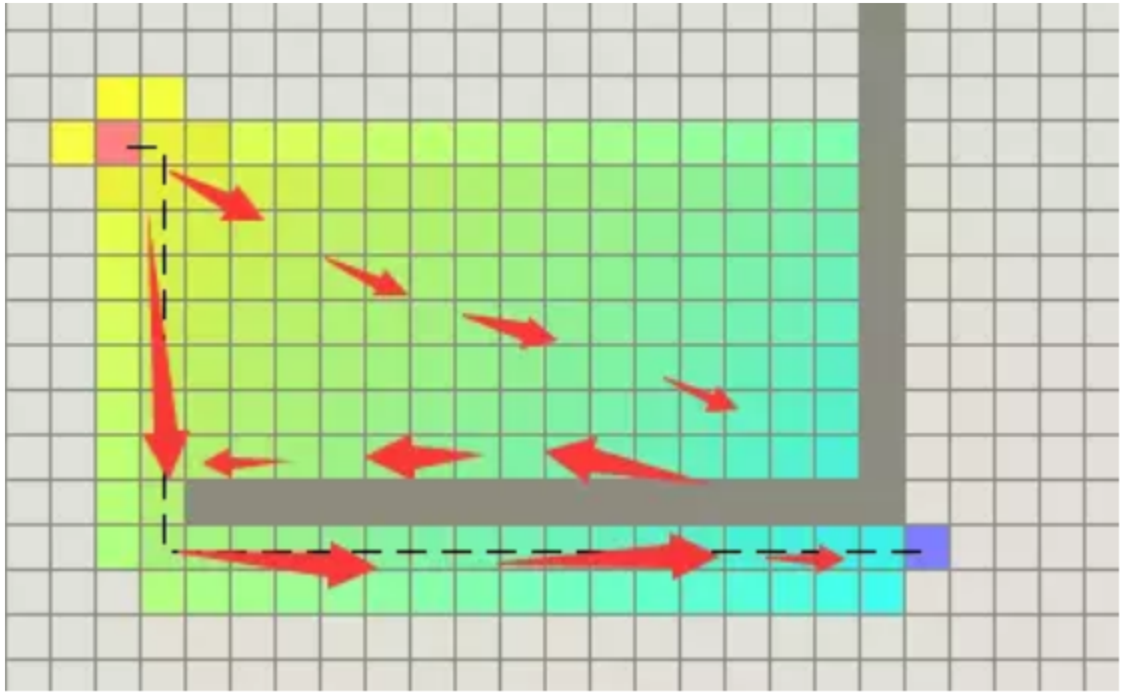


- The cost doubles when there is obstacles

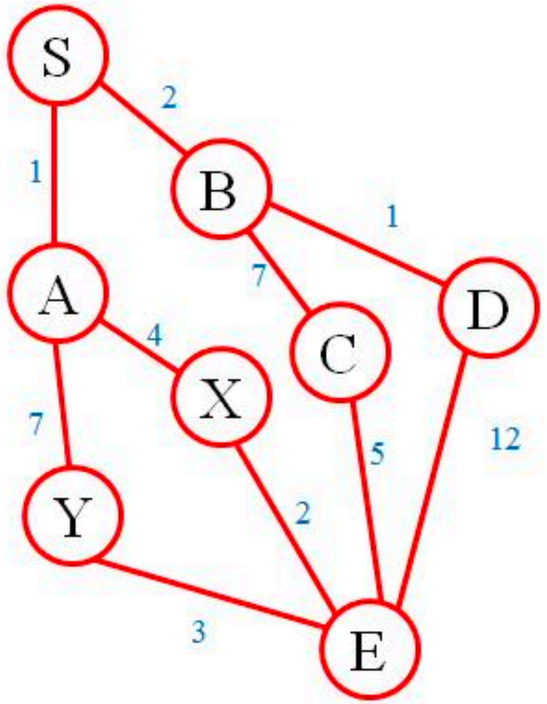


A* algorithm

- It combines the stability of BFS and the heuristics in greedy searching.
- Each time it visits point with the least $f(n) = g(n) + h(n)$ value.



Example of A* algorithm



- Values for h:

  A:5, B:6, C:4, D:15, X:5, Y:8

Expand S

{S,A} *f*=1+5=6

{S,B} *f*=2+6=8

Expand A

{S,B} *f*=2+6=8

{S,A,X} *f*=(1+4)+5=10

{S,A,Y} *f*=(1+7)+8=16

Expand B

{S,A,X} *f*=(1+4)+5=10

{S,B,C} *f*= (2+7)+4=13

{S,A,Y} *f*=(1+7)+8=16

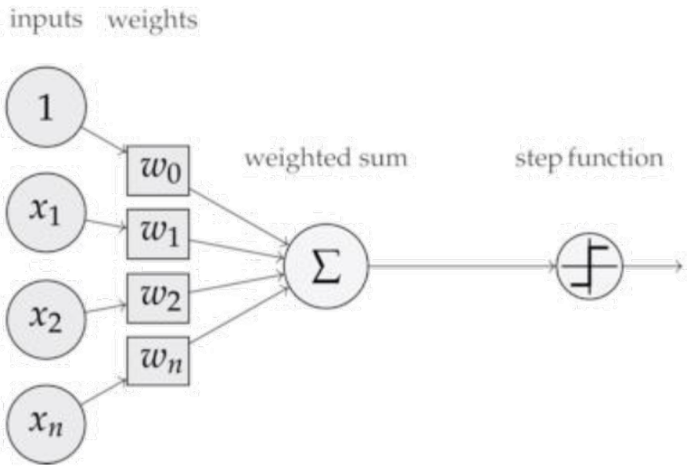
{S,B,D} *f*= (2+1)+15=18

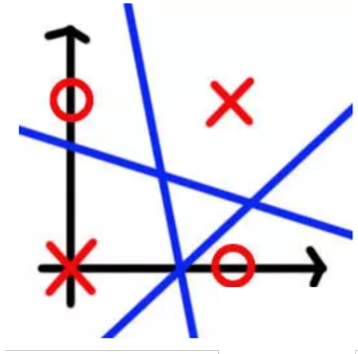
Expand X

{S,A,X,E} is the best path... (costing 7)

History of Machine Learning

- 1950s
  - Samuel's checker player
  - Machine learning term created
- 1960s
  - Neural networks: Perceptron
  - Pattern recognition
  - Minsky and Papert prove limitations of Perceptron





$$f(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{w} \cdot \mathbf{x} + b > 0, \\ 0 & \text{otherwise} \end{cases}$$

History of Machine Learning (cont.)

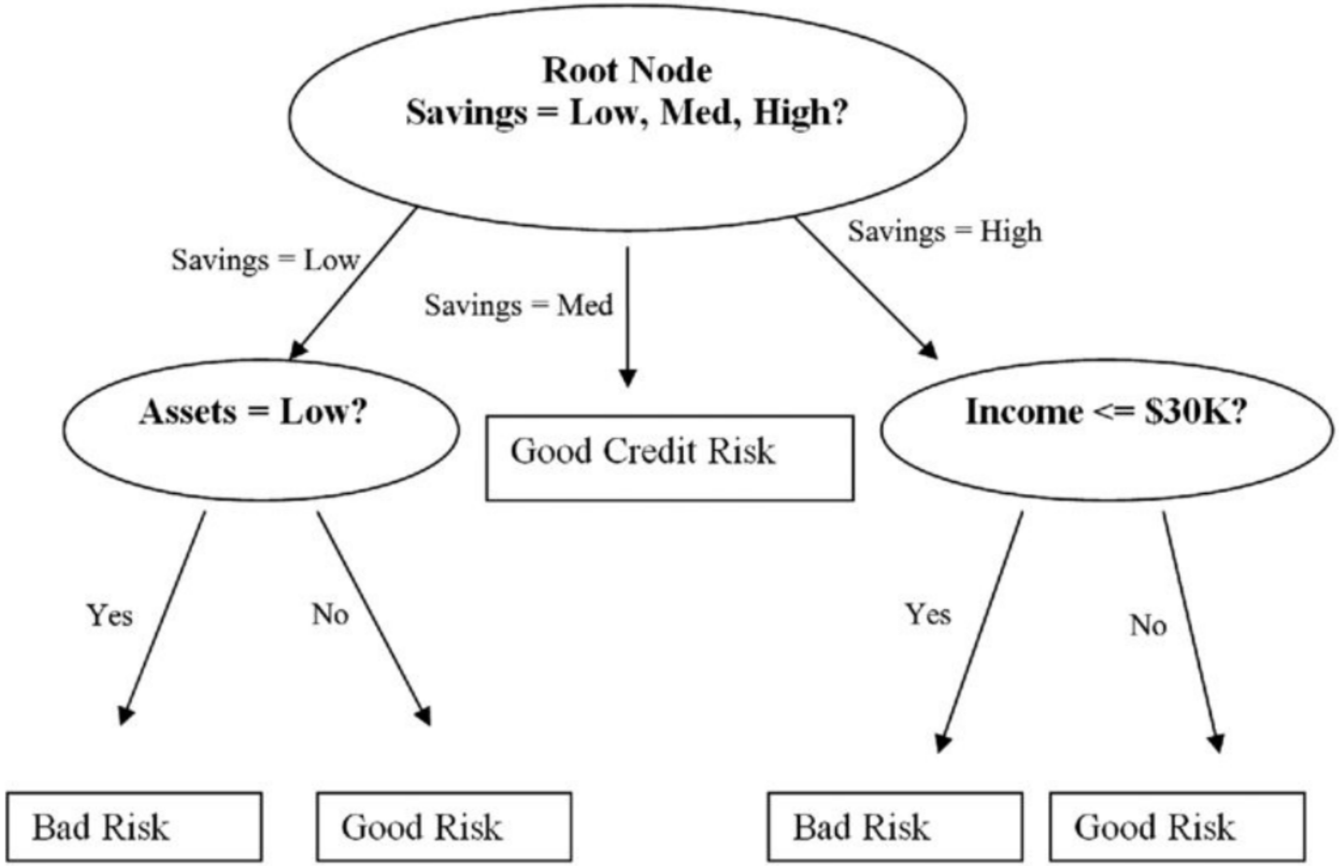
- 1970s
  - Symbolic concept induction
    - “Logic theorist”: We can give machine intelligence if we give them logic.
  - Winston’s arch learner
  - Expert systems and the knowledge acquisition bottleneck
    - Only with logic is far from intelligence
    - Machines need knowledge
    - Then find it is hard to teach knowledge summarized by humans to machines
    - It would be better if machines can learn knowledge by themselves!
  - Quinlan’s ID3
  - Mathematical discovery with AM

History of Machine Learning (cont.)

- 1980s
  - Advanced decision tree and rule learning
    - Learning from samples
    - Simple and efficient, still popular
    - Good ability to represent knowledge
    - Easy to demonstrate complicated data structure
    - But is hard to learn for large dataset
  - Explanation-based Learning (EBL)
  - Learning and planning and problem solving
  - Utility problem
  - Analogy
  - Cognitive architectures
  - Resurgence of neural networks (connectionism, backpropagation)
    - Learning from samples
    - Limits: rely heavily on parameters. Results could vary a lot even if parameters change a little
  - Valiant's PAC Learning Theory
  - Focus on experimental methodology

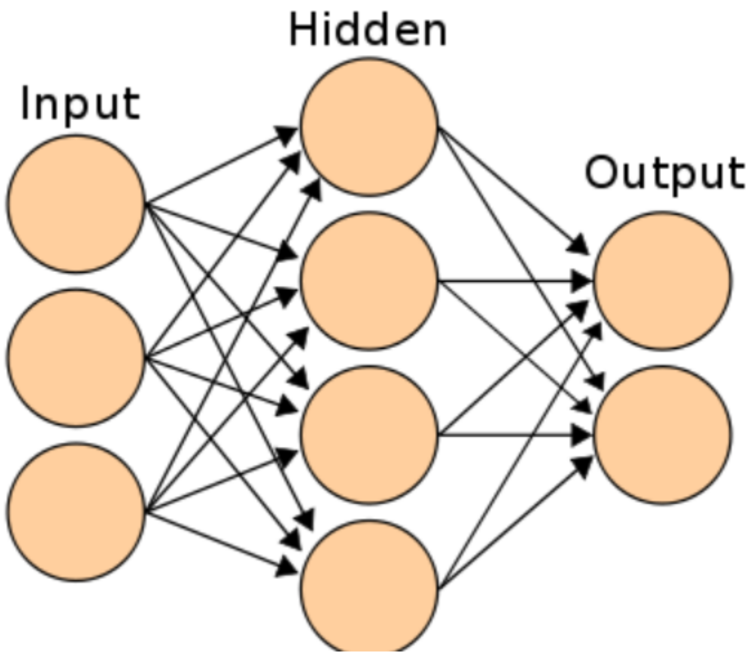
Example of Decision Tree

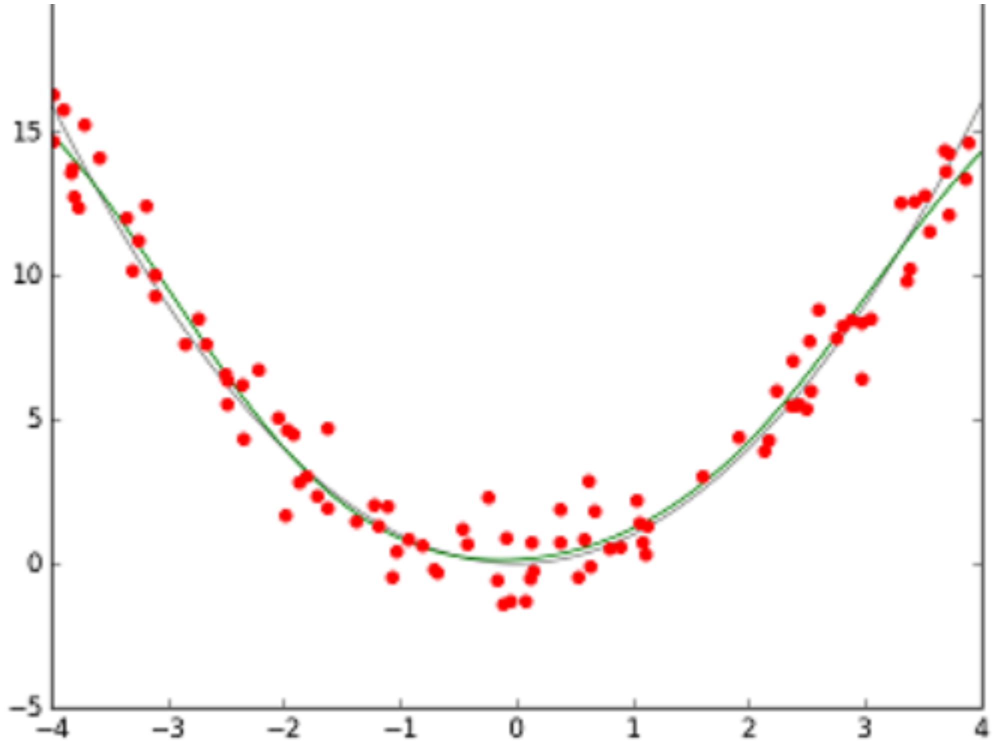
- Split the data by informative questions.



Example of Neural Network

- It approximates the function in high dimensional space.



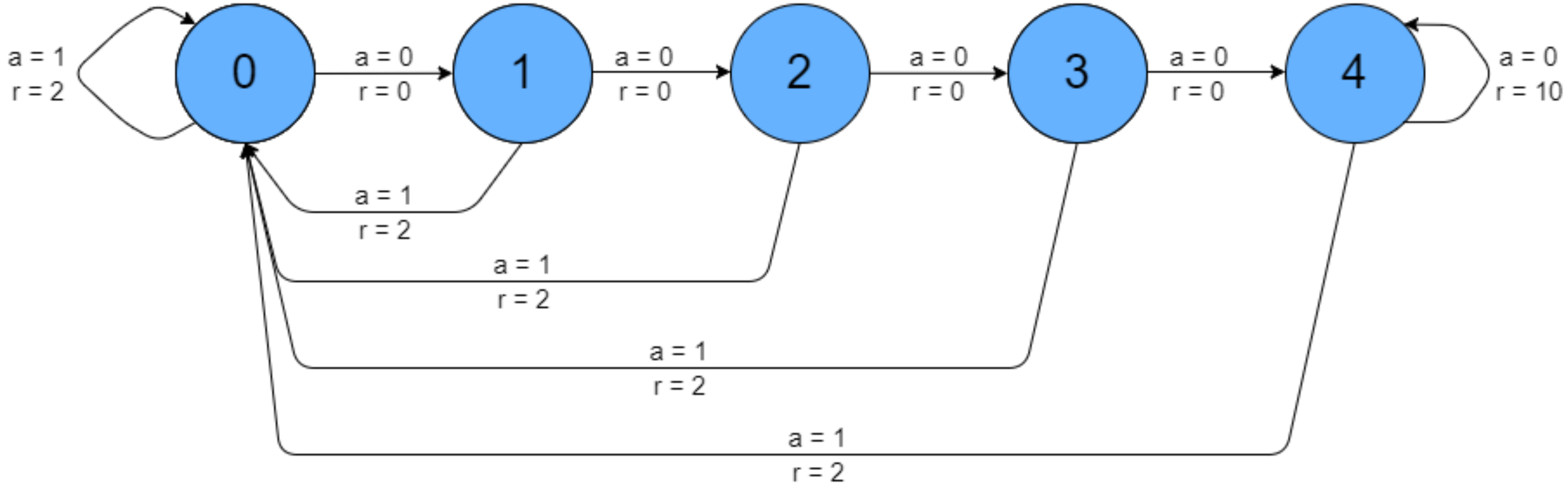


History of Machine Learning (cont.)

- 1990s
  - Data mining
  - Adaptive software agents and web applications
  - Text learning
  - Reinforcement learning (RL)
  - Inductive Logic Programming (ILP)
  - Ensembles: Bagging, Boosting, and Stacking
  - Bayes Net learning
  - Support vector machines
    - Statistical learning
  - Kernel methods

Examples of Reinforcement Learning

- Learn which action brings the highest reward at each state, based on your experience.



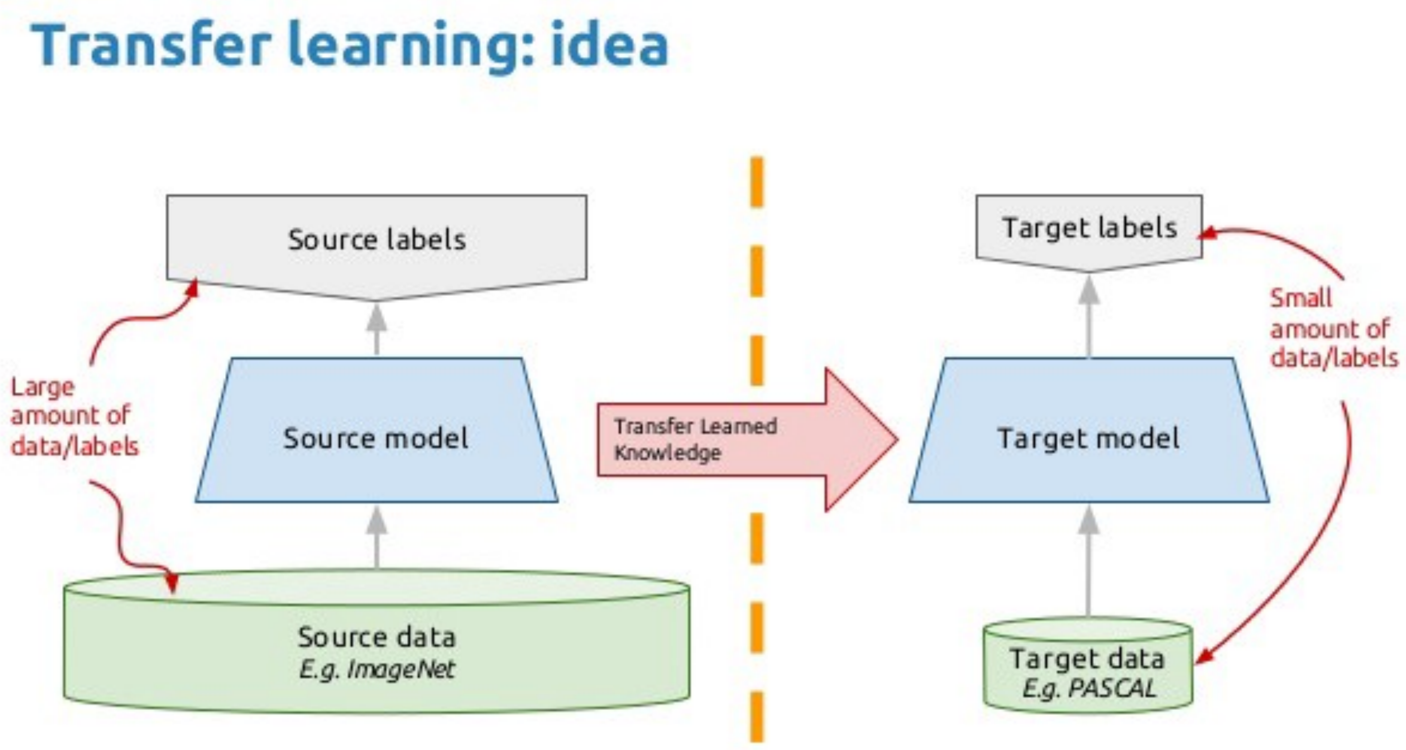
History of Machine Learning (cont.)

- 2000s
  - Graphical models
  - Variational inference
  - Statistical relational learning
  - Transfer learning
  - Sequence labeling
  - Collective classification and structured outputs
  - Computer systems applications
    - Compilers • Debugging • Graphics • Security (intrusion, virus, and worm detection)
  - Email management
  - Personalized assistants that learn
  - Learning in robotics and vision

Slide credit: Raymond J. Mooney

Transfer Learning

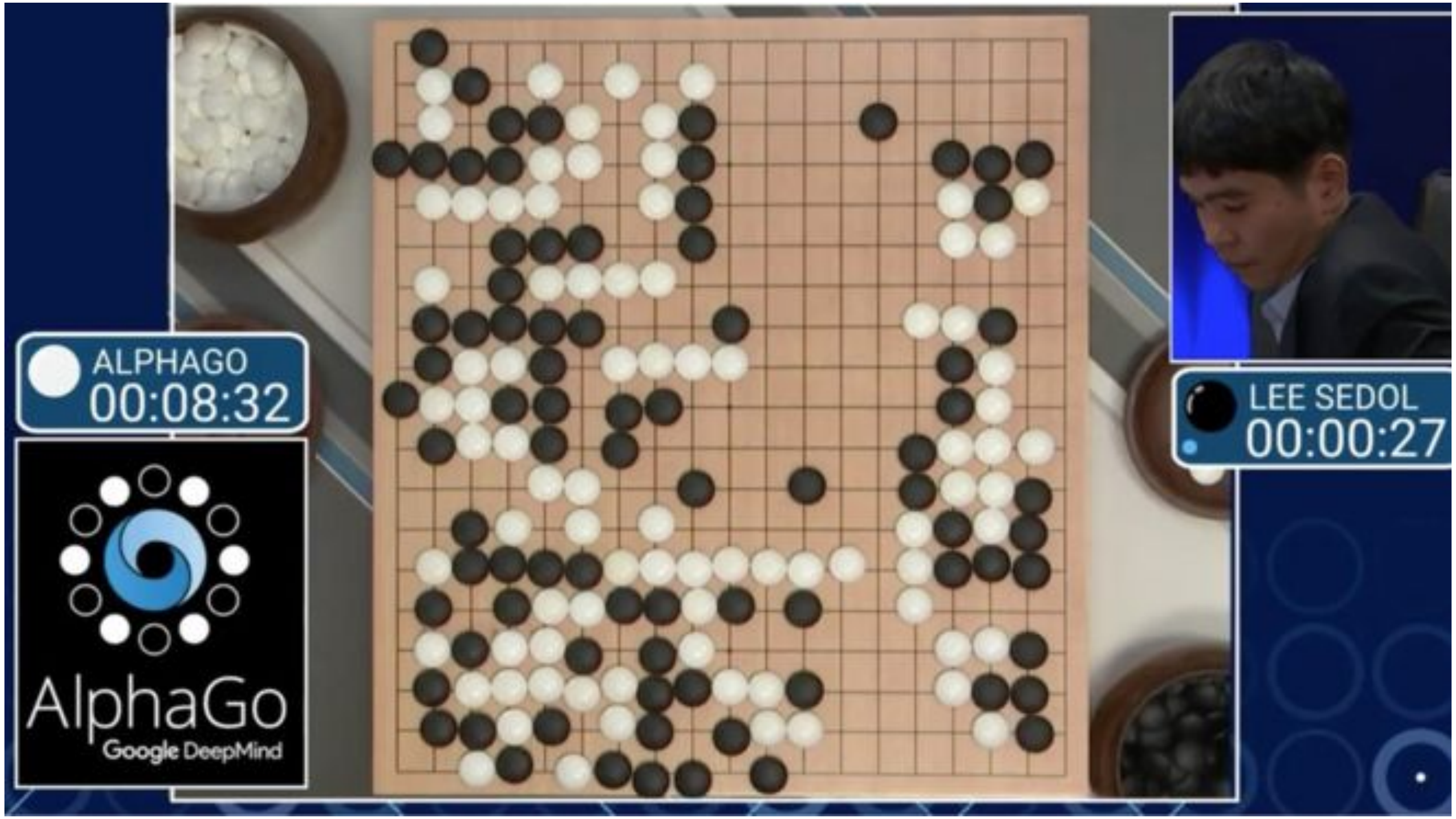
- Learn from source data and apply the knowledge on target data



History of Machine Learning (cont.)

- 2010s
  - Deep learning
    - Good performances in images/speeches
    - Rely on good parameters (compared to good user previously)
    - Lack of theoretical guarantees but lower threshold to users
  - Learning from big data
  - Learning with GPUs or HPC
  - Multi-task & lifelong learning
  - Deep reinforcement learning
  - Massive applications to vision, speech, text, networks, behavior etc.
  - Meta-learning and AutoML
  - …

Breaking through by DRL

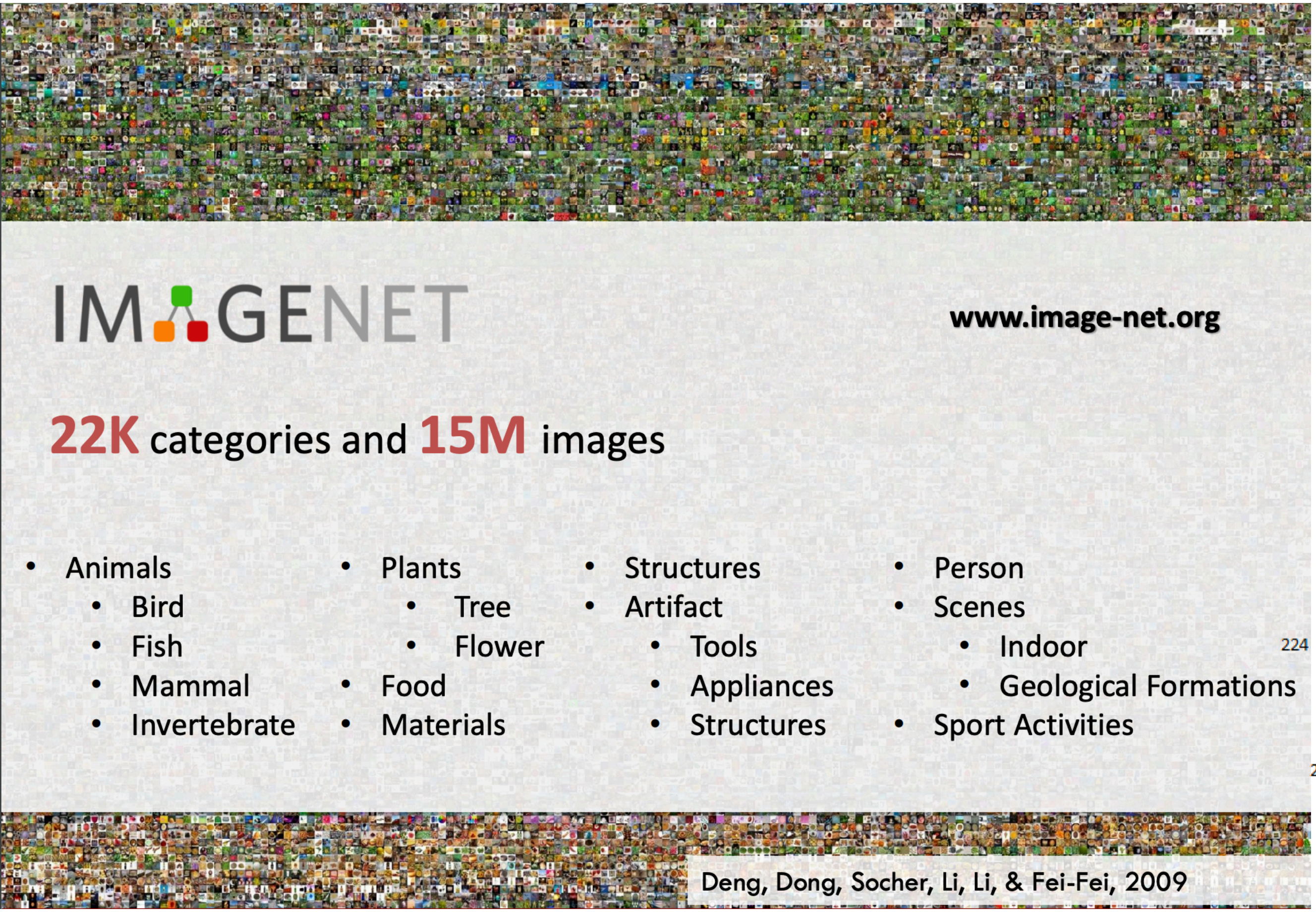


Machine Learning Trend

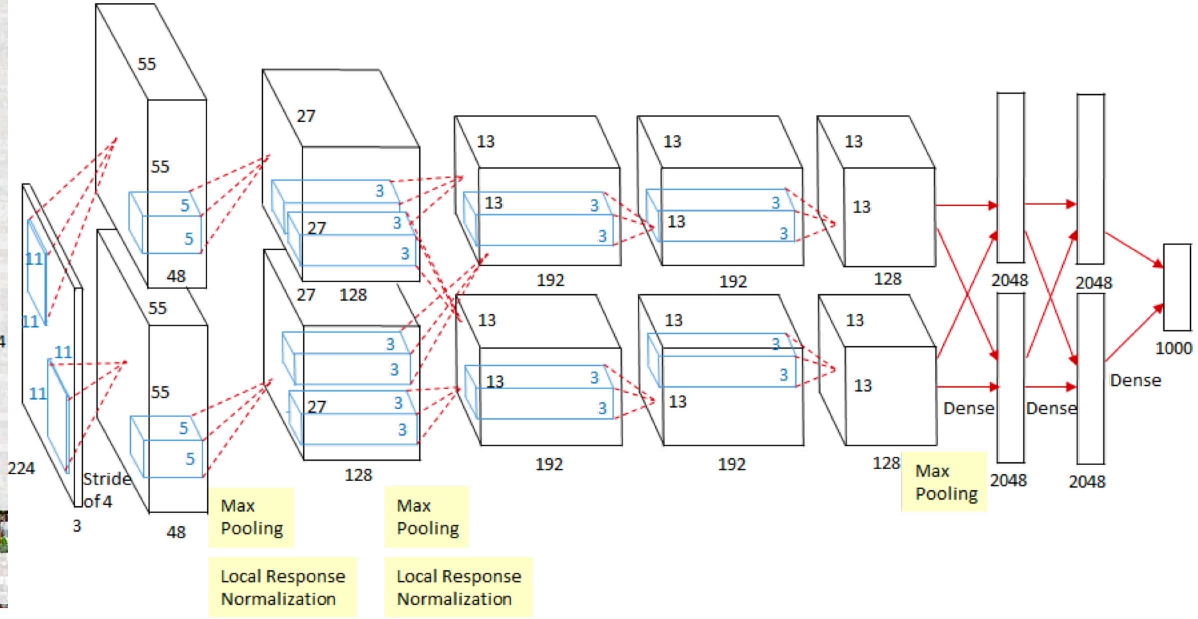
https://www.google.com/trends

Recent Progress

Computer Vision (CV) -- ImageNet, AlexNet



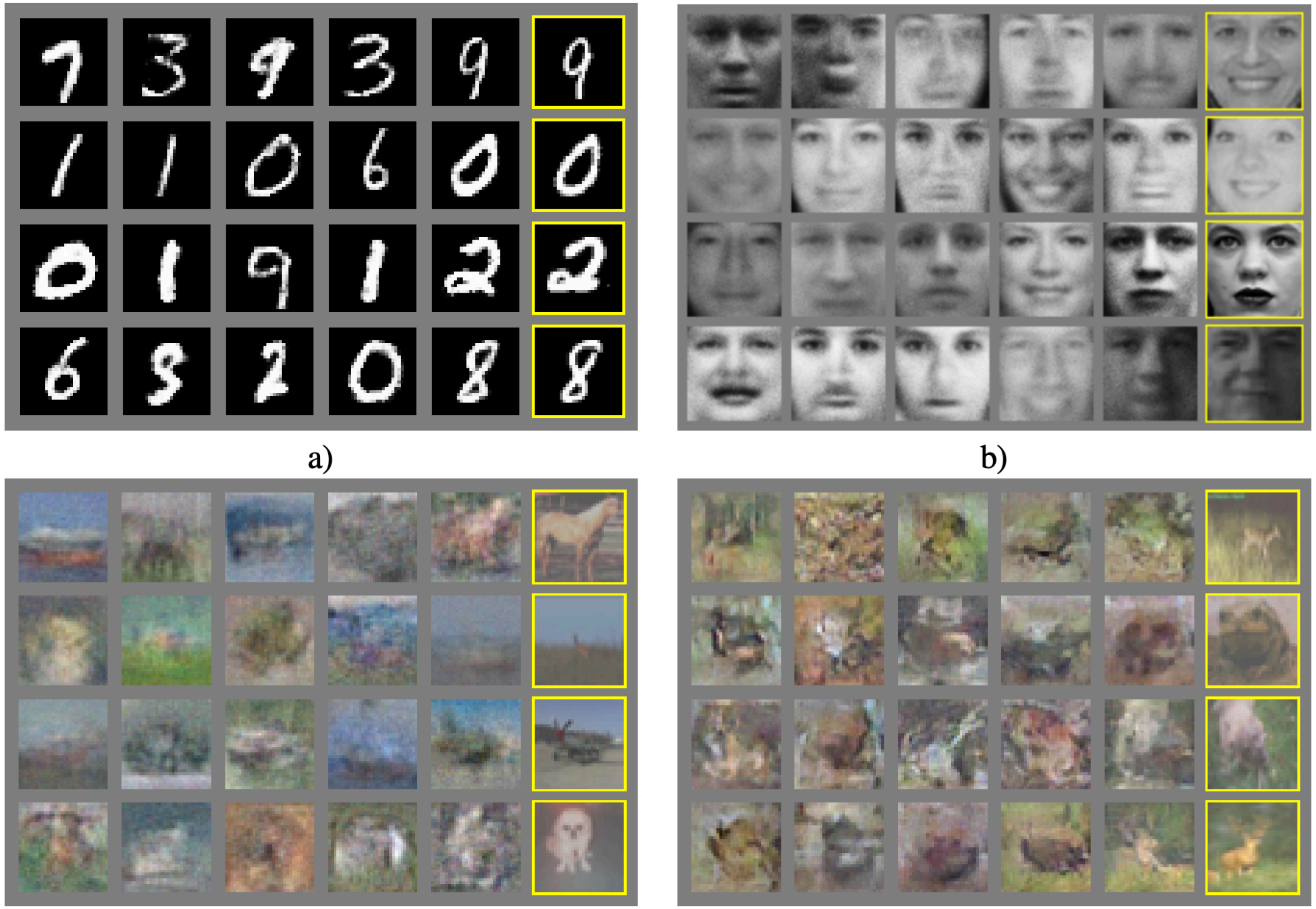
AlexNet, CNN



Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). IEEE.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

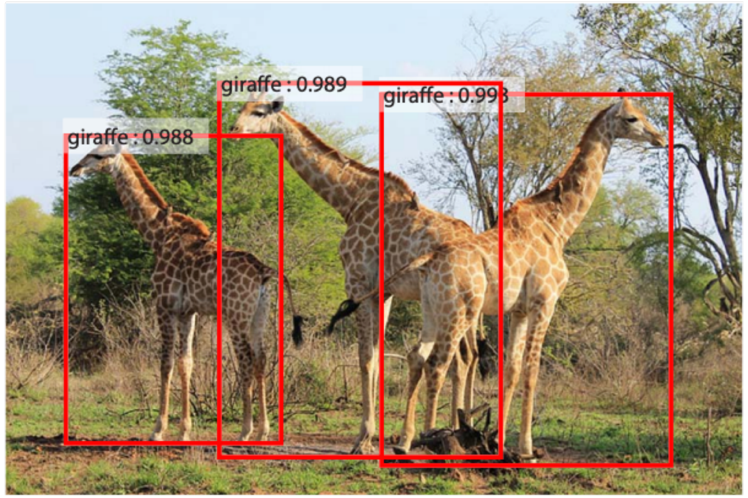
CV -- GAN

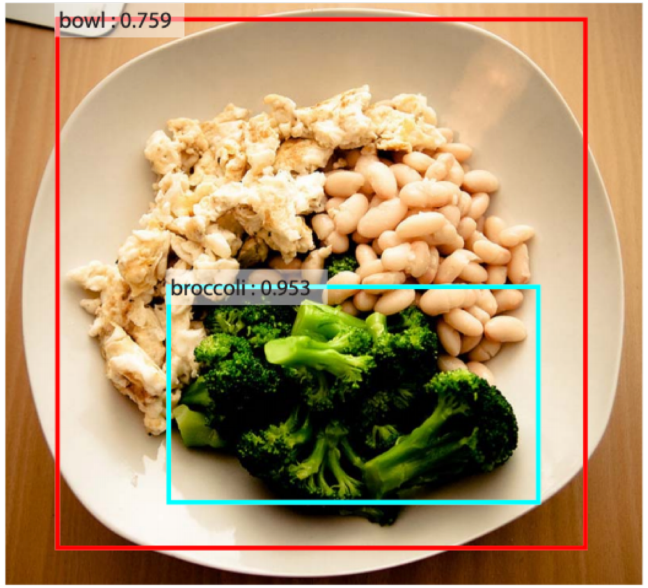


a) b) c) d)

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672-2680).

CV (Detection) -- R-CNN, Fast R-CNN, Faster R-CNN





1.Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).
2. Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448).
3. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91-99).

Speech recognition (Unsupervised, ICA)

Mixed

Separated

Speech recognition (Unsupervised, ICA, cont.)

Mixed

Separated

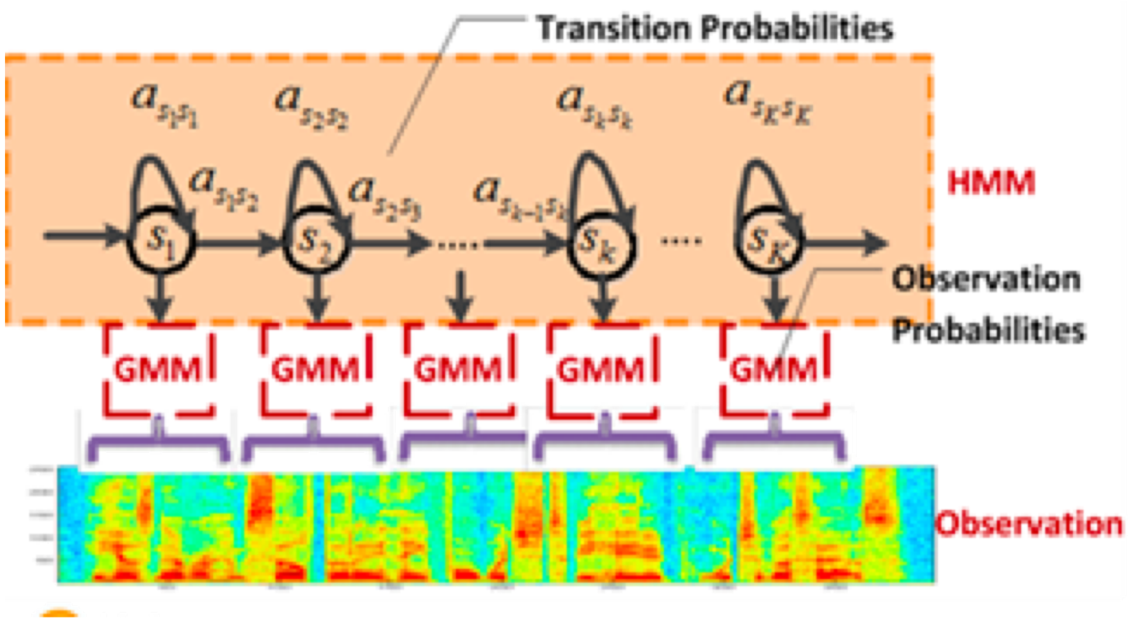
Speech recognition

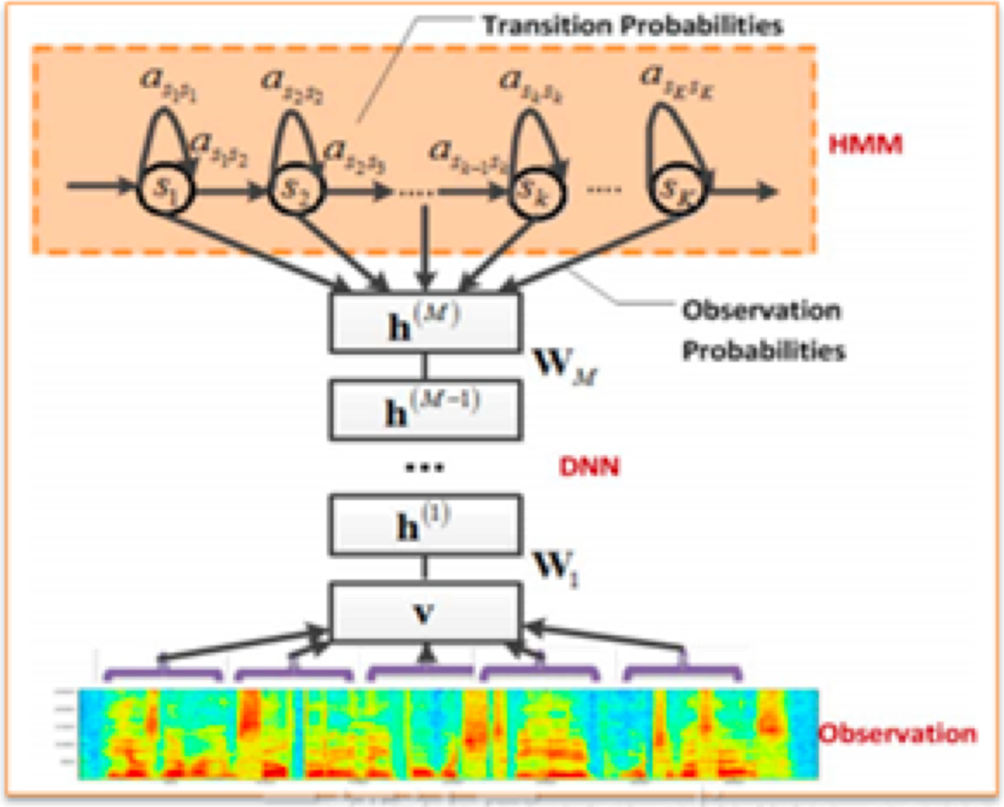
- Previous works use
  - Hidden Markov models (HMMs)
    - Deal with the temporal variability of speech
  - Gaussian mixture models (GMMs)
    - Determine how well each state of each HMM fits a frame or a short window of frames of coefficients that represents the acoustic input
- New
  - Feed-forward neural network
    - Takes several frames of coefficients as input and produces posterior probabilities over HMM states as output

Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A. R., Jaitly, N., ... & Sainath, T. (2012). Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal processing magazine*, *29*.

Speech recognition

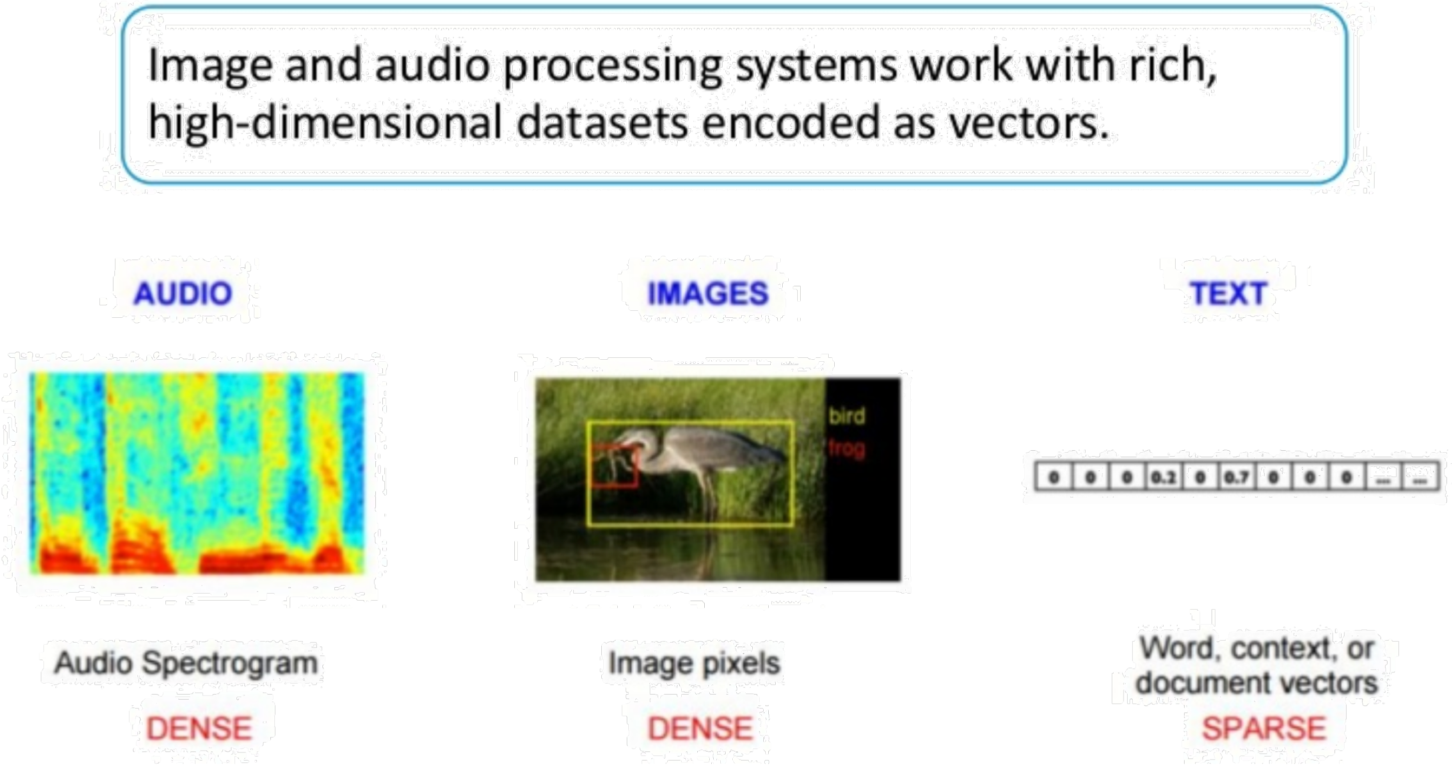
■ **Deep Learning: From GMM-HMM to DNN-HMM**





Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A. R., Jaitly, N., ... & Sainath, T. (2012). Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal processing magazine*, *29*.

Natural Language Processing (NLP) -- Word2Vec



Pennington, J., Socher, R., & Manning, C. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).

Natural Language Processing (NLP) -- Word2Vec (cont.)

Word Analogies

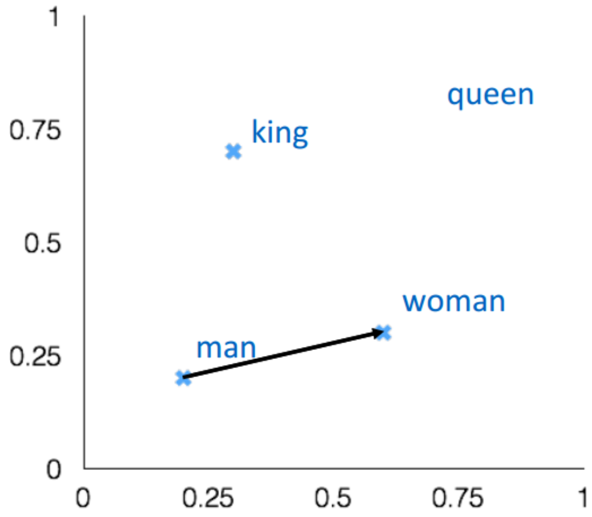
Test for linear relationships, examined by Mikolov et al. (2014)

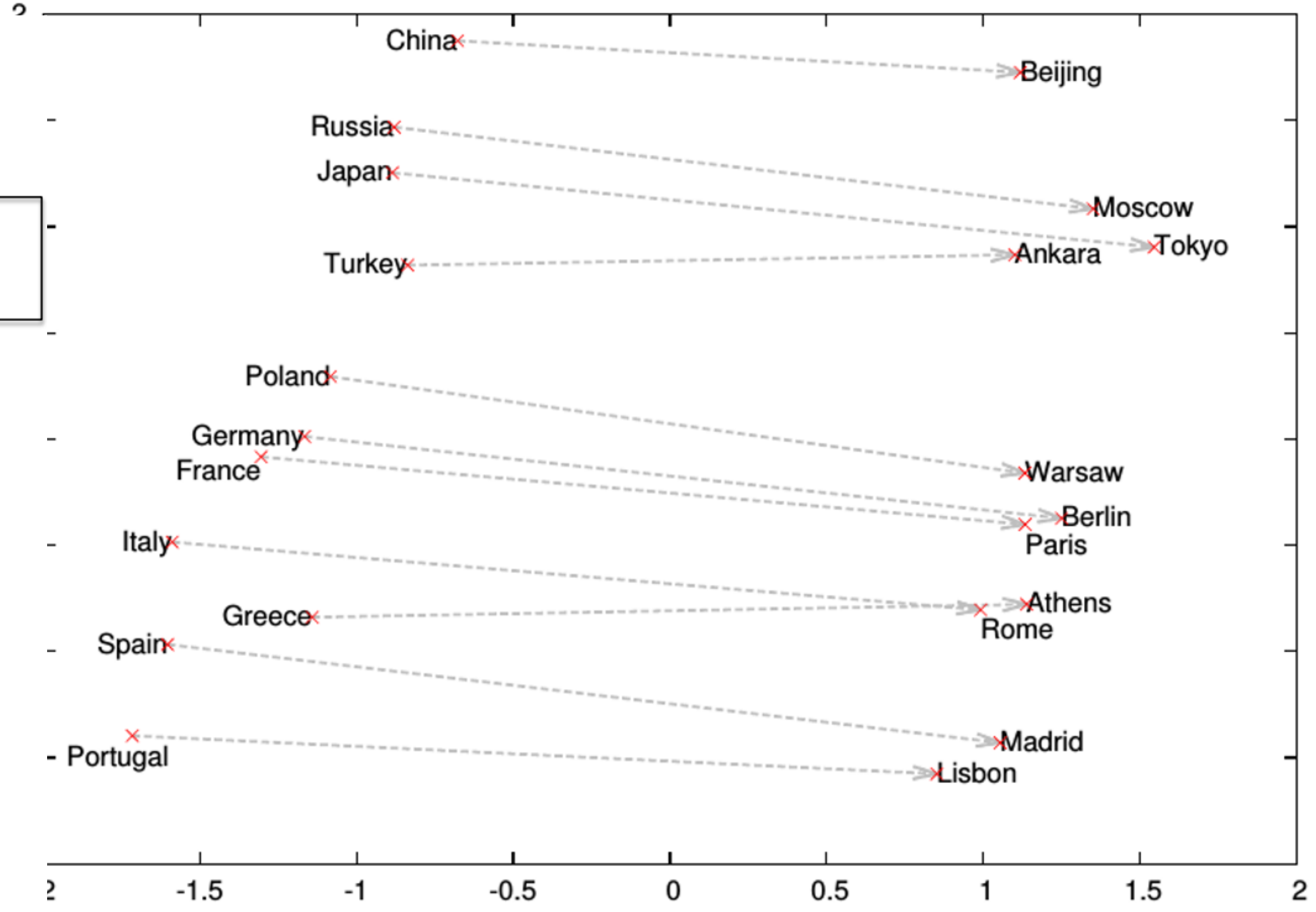
a:b :: c:?

$$d = \arg\max_x \frac{(w_b - w_a + w_c)^T w_x}{||w_b - w_a + w_c||}$$

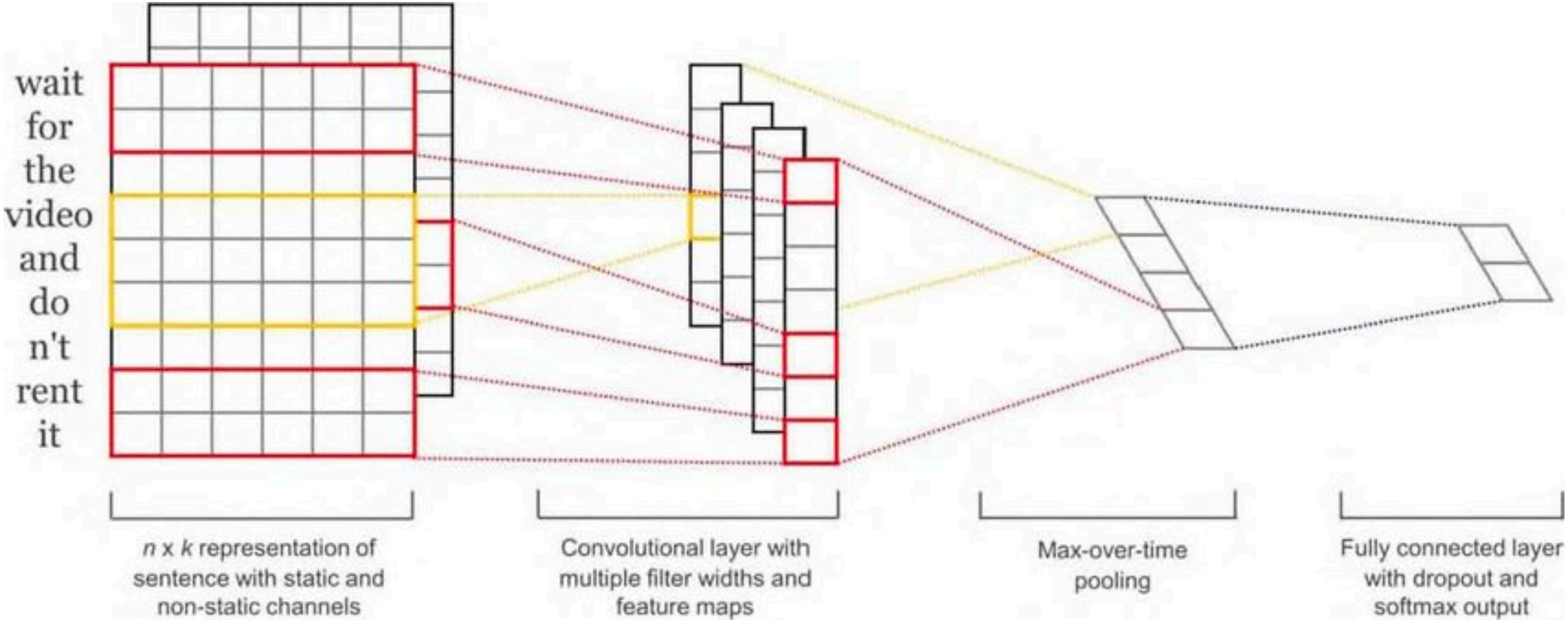
man:woman :: king:?

+	king	[0.30 0.70]
-	man	[0.20 0.20]
+	woman	[0.60 0.30]
	queen	[0.70 0.80]





NLP -- CNN



Kim, Y. (2014, October). Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1746-1751).

NLP -- BERT

- BERT
  - Bidirectional Encoder Representations from Transformers
  - The pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers
  - The pre-trained BERT model can be finetuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial taskspecific architecture modifications
  - It obtains new state-of-the-art results on eleven natural language processing tasks

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

NLP -- BERT

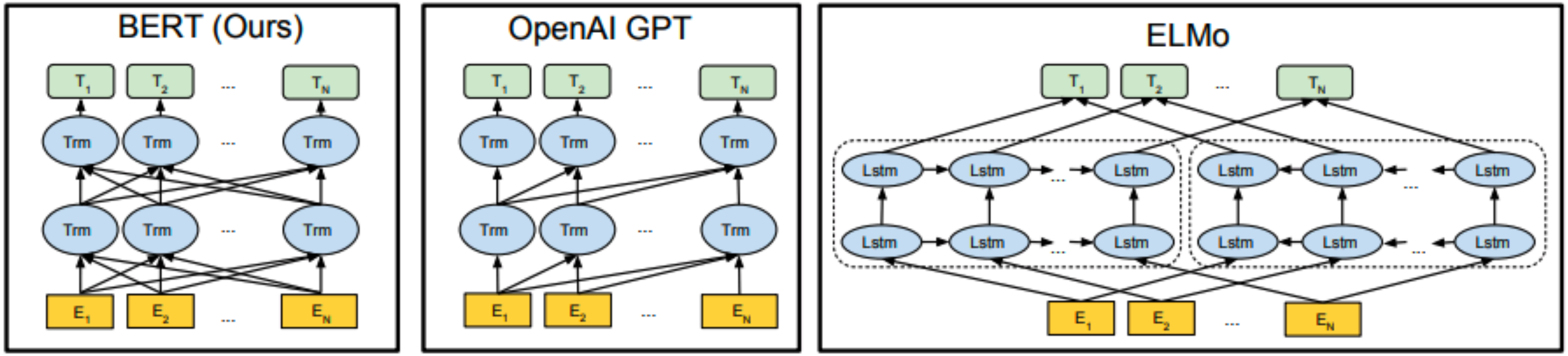


Figure 1: Differences in pre-training model architectures. BERT uses a bidirectional Transformer. OpenAI GPT uses a left-to-right Transformer. ELMo uses the concatenation of independently trained left-to-right and right-to-left LSTM to generate features for downstream tasks. Among three, only BERT representations are jointly conditioned on both left and right context in all layers.

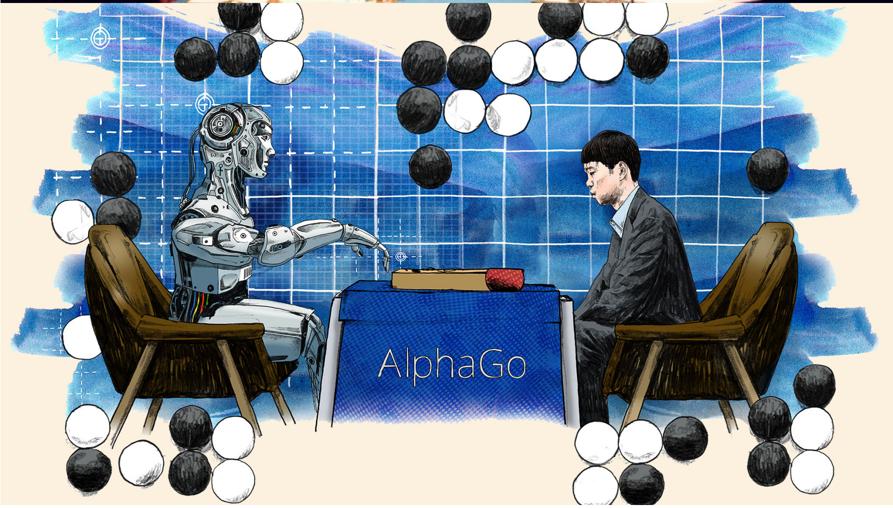
Game Playing -- Atari

Deep Reinforcement Learning

Trained separate DQN agents for 50 different Atari games, without any prior knowledge of the game rules

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Petersen, S. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529.

Game Playing

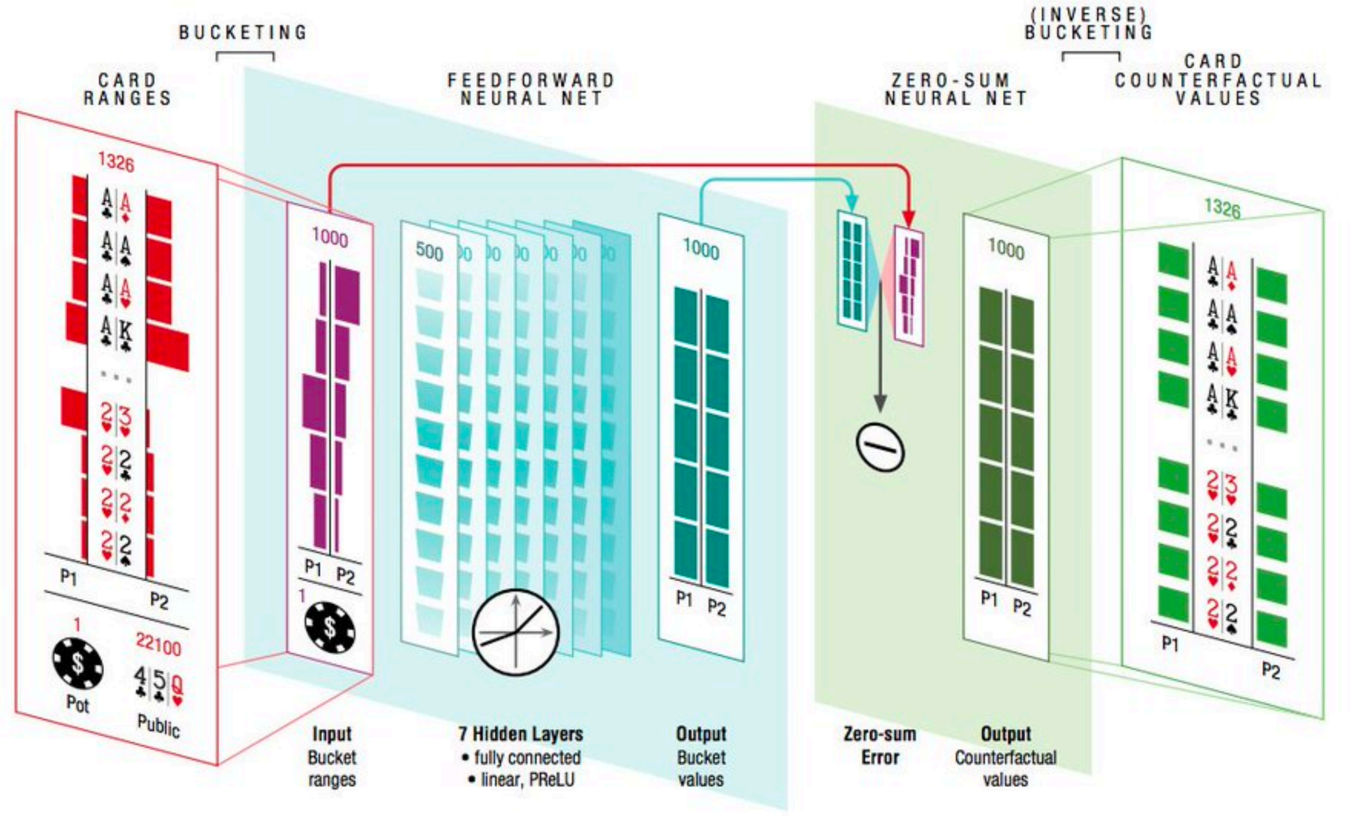


- IBM Deep Blue (1996)
  - Win Garry Kasparov by 3.5:2.5 on Chess
  - Search over 12 following steps

- AlphaGo (2016)
  - Win Lee Sedol by 4:1 on Go
  - Efficient search on large solution space

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Dieleman, S. (2016). Mastering the game of Go with deep neural networks and tree search. *nature*, *529*(7587), 484.

Game Playing -- Texas hold’em



DeepStack

- In a study involving 44,000 hands of poker, DeepStack defeated with statistical significance professional poker players in heads-up no-limit Texas hold’em
- Imperfect information setting

Moravčík, M., Schmid, M., Burch, N., Lisý, V., Morrill, D., Bard, N., ... & Bowling, M. (2017). Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, *356*(6337), 508-513.

History of Game AI

1956 checkers

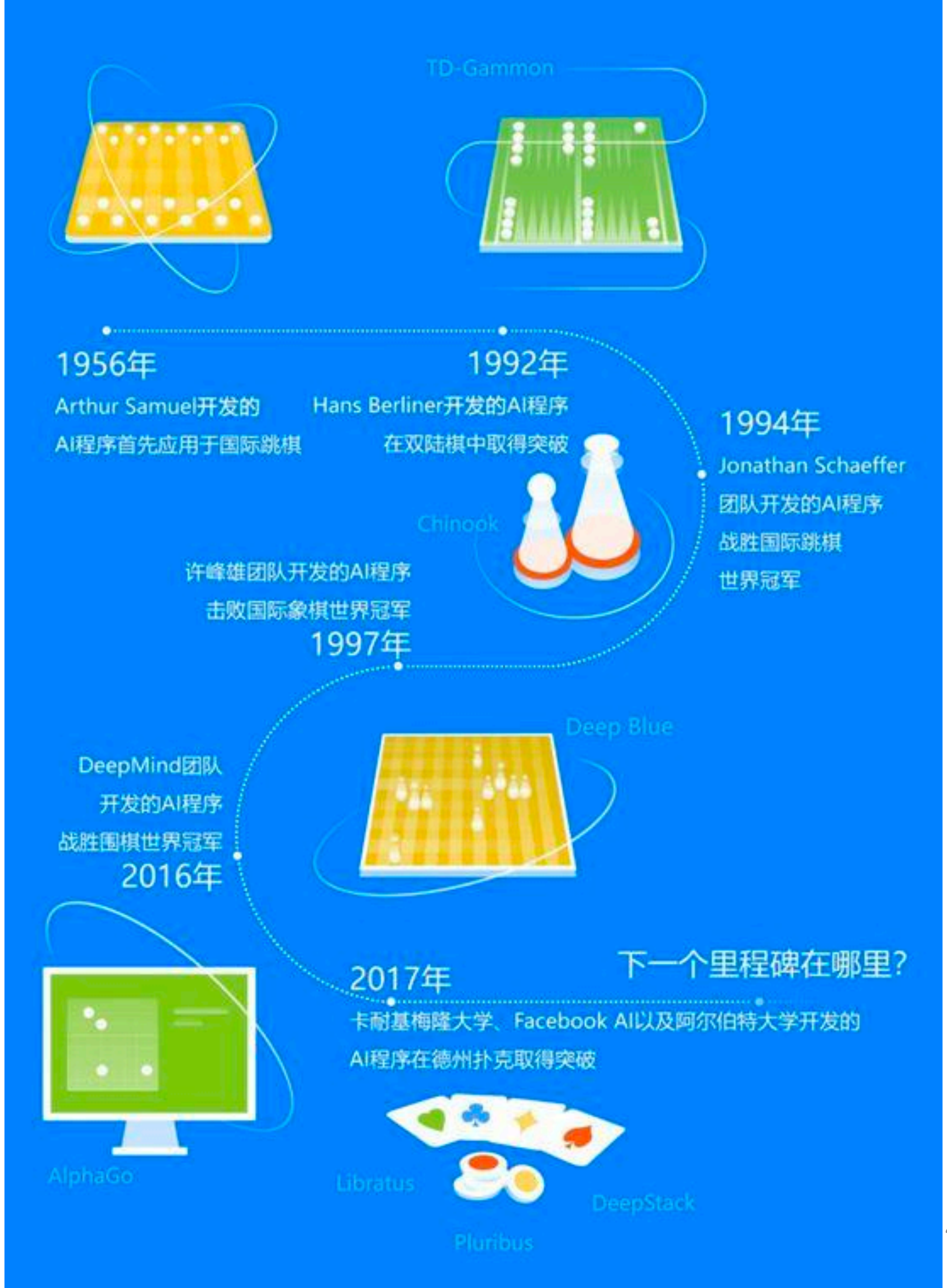
1992 backgammon

1994 checkers

1997 chess

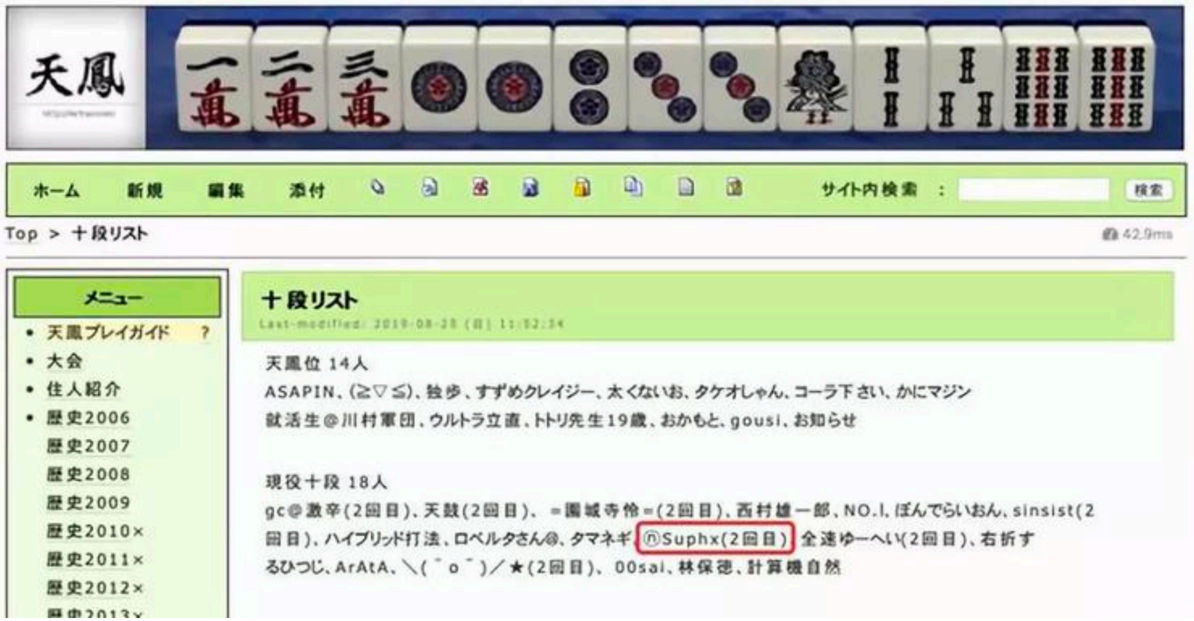
2016 Go

2017 Texas hold’em



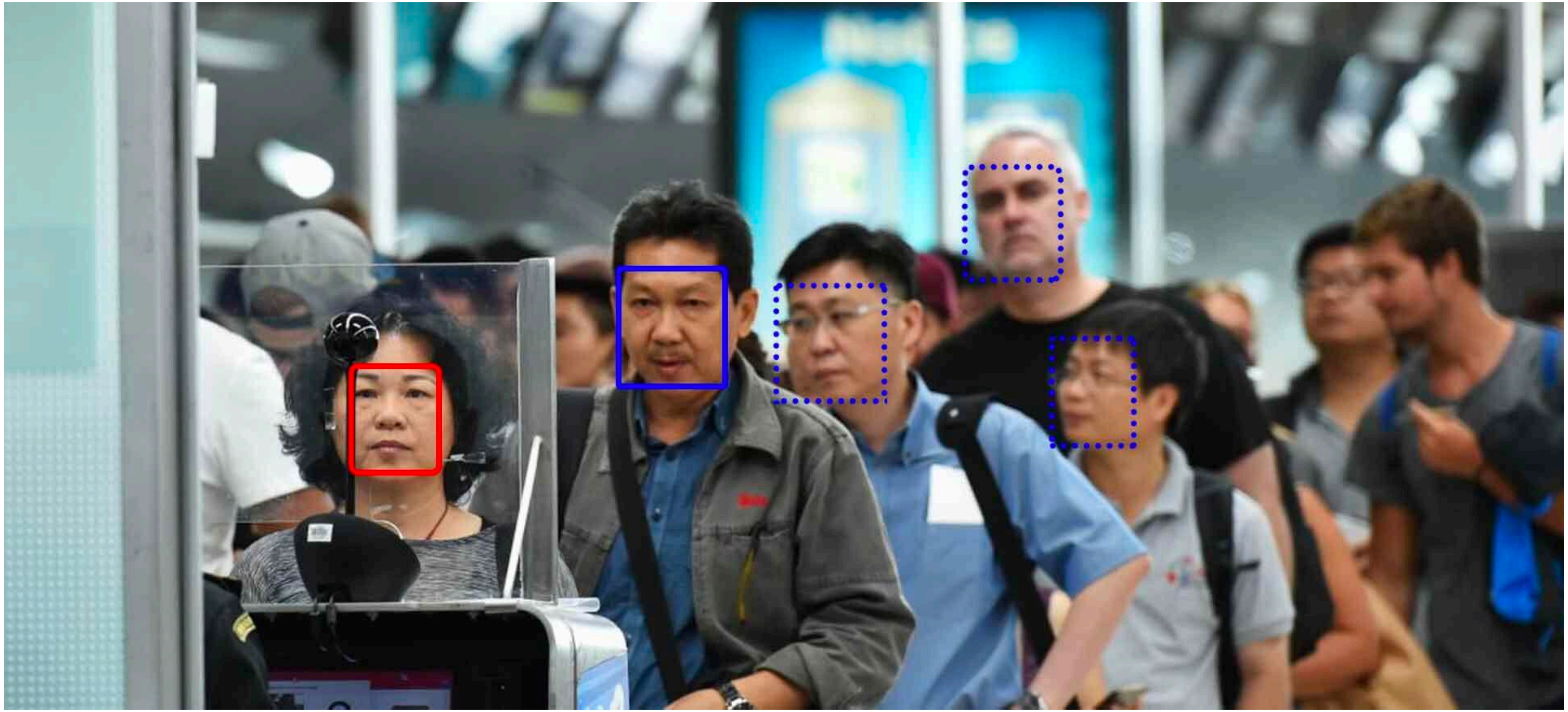
Game Playing -- Majiang (Aug 29, 2019)

- Microsoft Suphx
  - Professional level



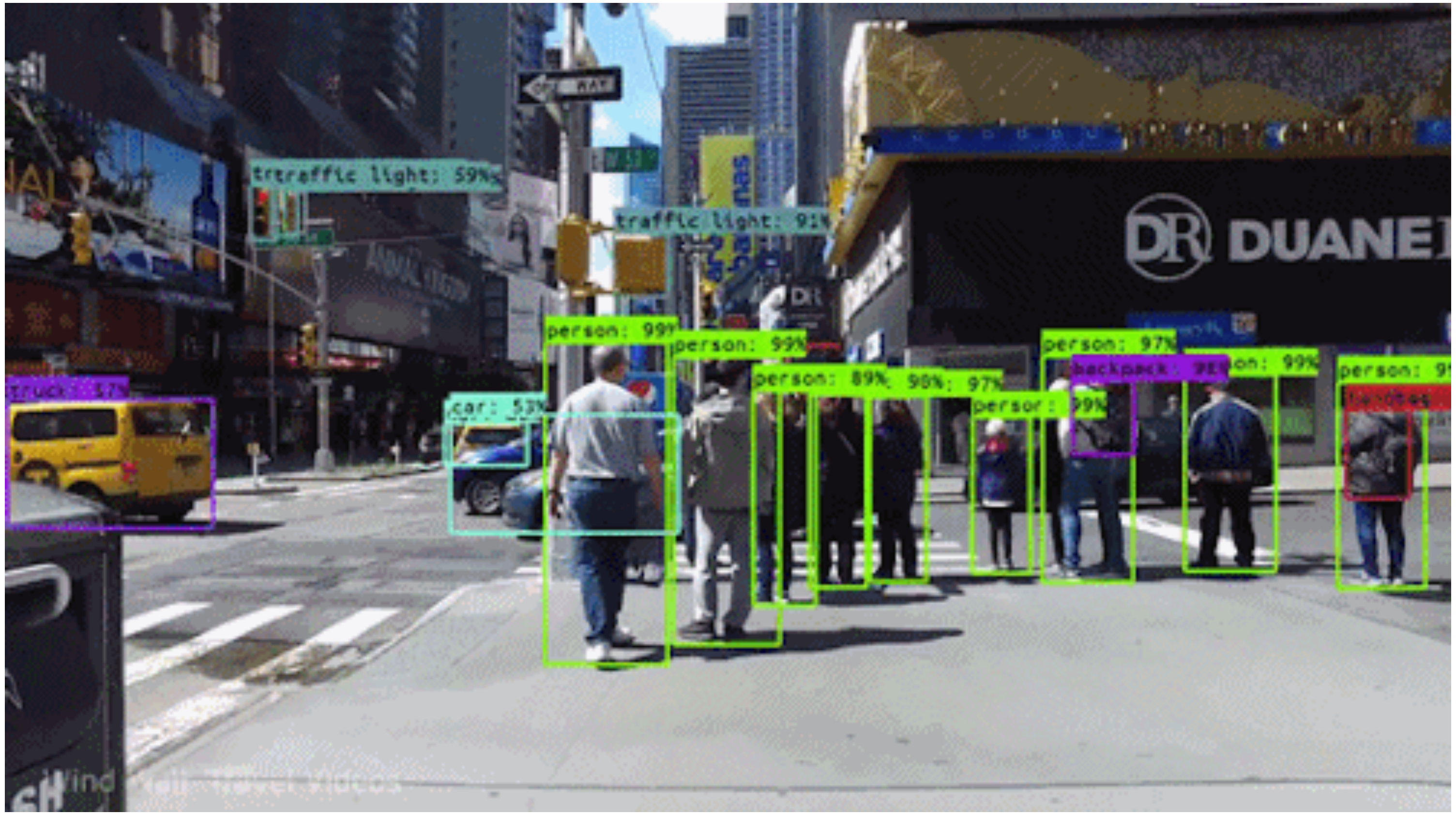
Applications

Face Recognition in Customs



https://bitrefine.group/home/transportation/face-recognition-support-system

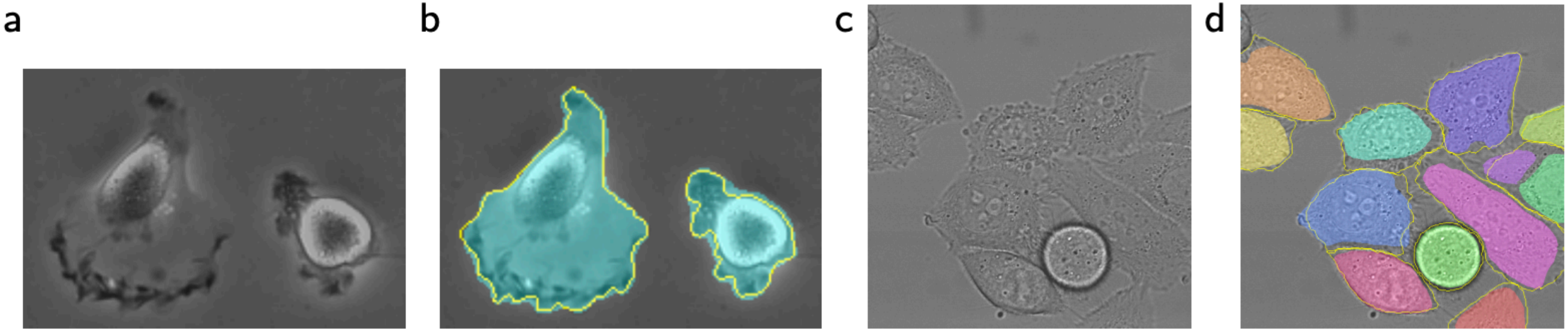
Autonomous Self-driving Cars



https://cdn-images-1.medium.com/max/1600/1*q1uVc-MU-tC-WwFp2yXJow.gif

Medical image analysis

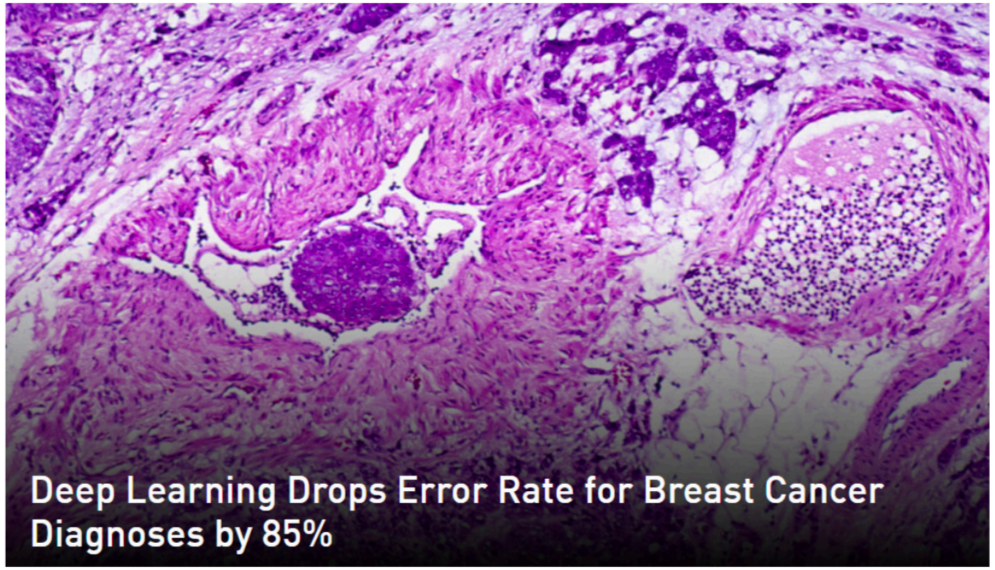
- Segmentation results on ISBI cells and DIC-HeLa cells



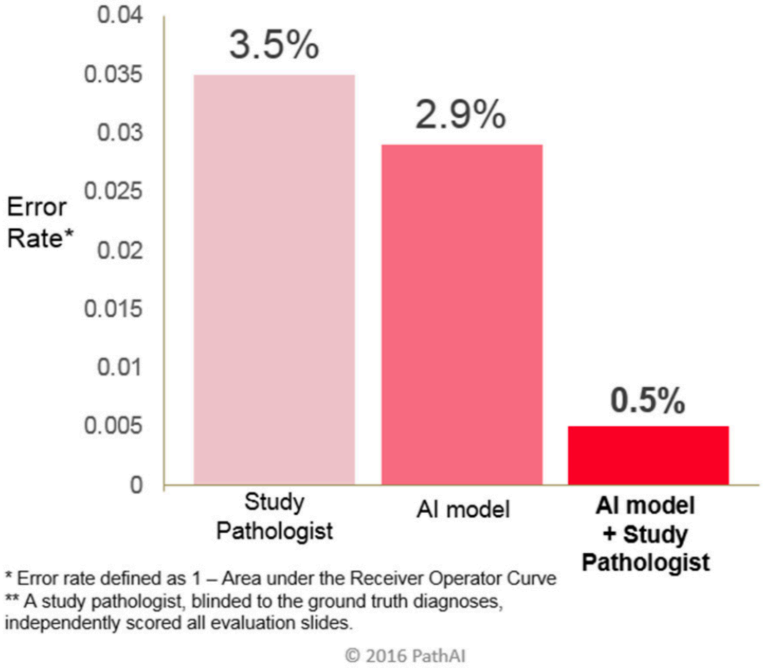
Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

Medical image analysis

- Breast Cancer Diagnoses



(AI + Pathologist) > Pathologist



Wang, Dayong, et al. "Deep learning for identifying metastatic breast cancer." arXiv preprint arXiv:1606.05718 (2016).
https://blogs.nvidia.com/blog/2016/09/19/deep-learning-breast-cancer-diagnosis/

Voice assistants

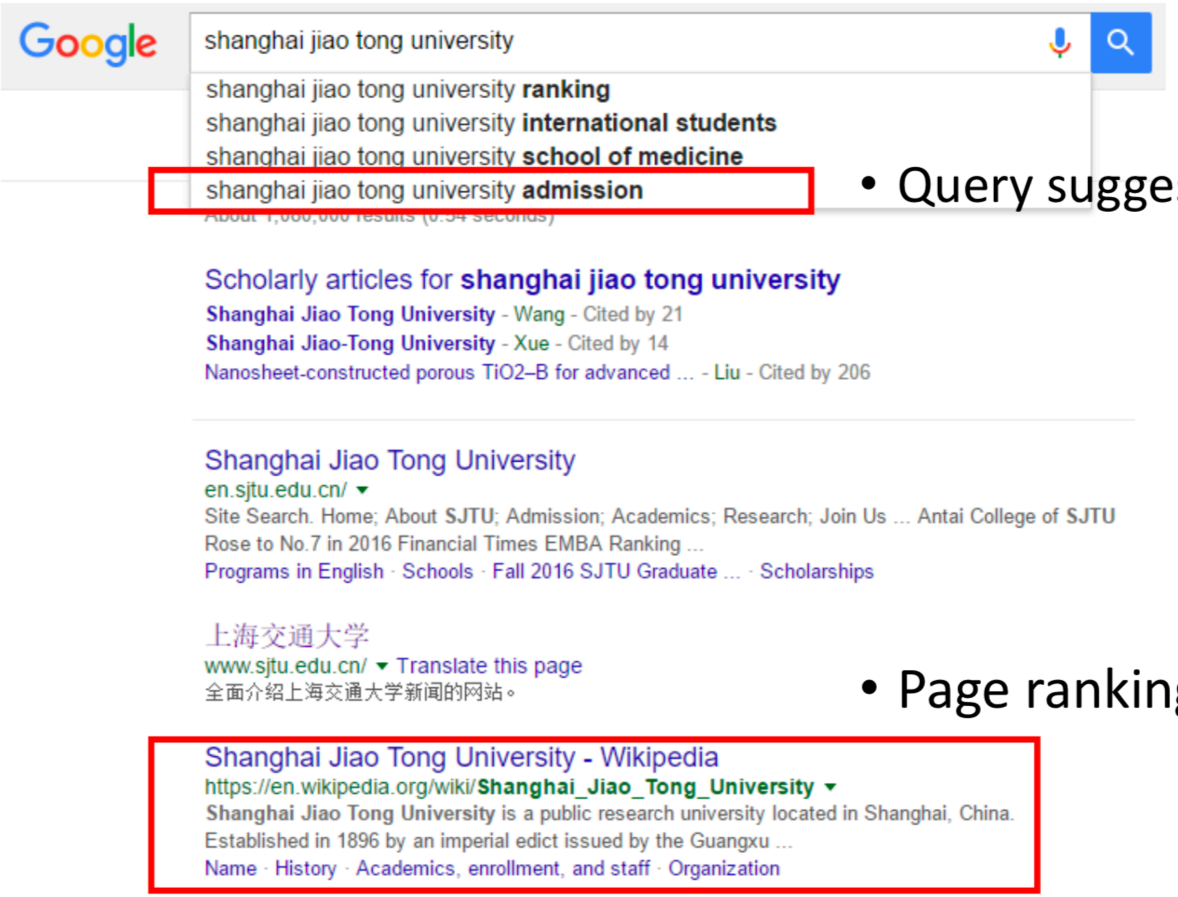
- Google AI can make complete phone calls (2018)

Voice assistants

Alexa Siri Google Now Cortana

- Alexa/Siri/Google/Cortana
- XiaoAI (Xiaomi)/
  HiAssistant (EMUI)/Siri/
  Bixby (Samsung)/
  Jovi (vivo)

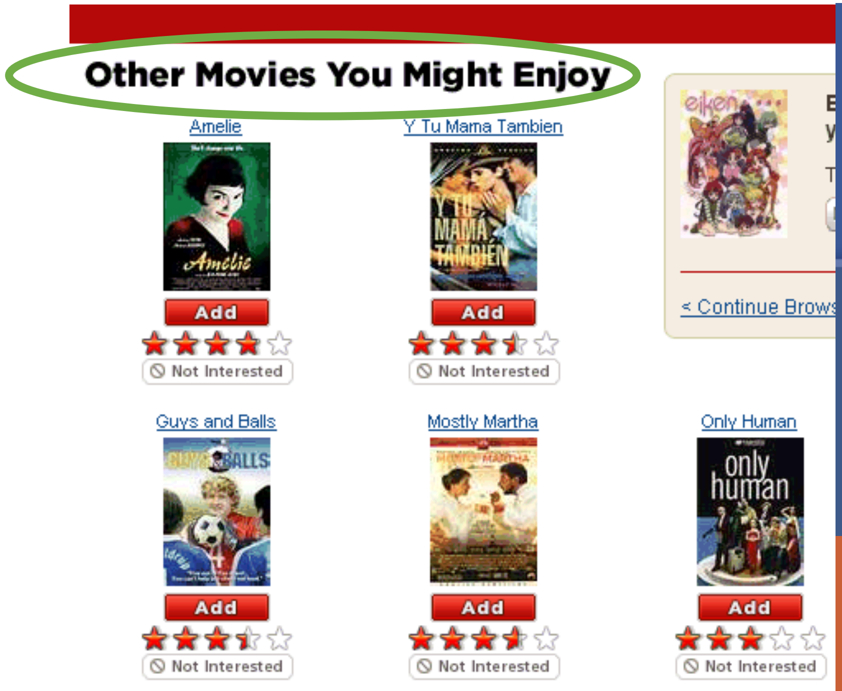
Web search



- Query suggestion
- Page ranking

Slide credit: Weinan Zhang

Web recommendation



Recommend movies/events/products based on history records

Netflex/Facebook/Amazon

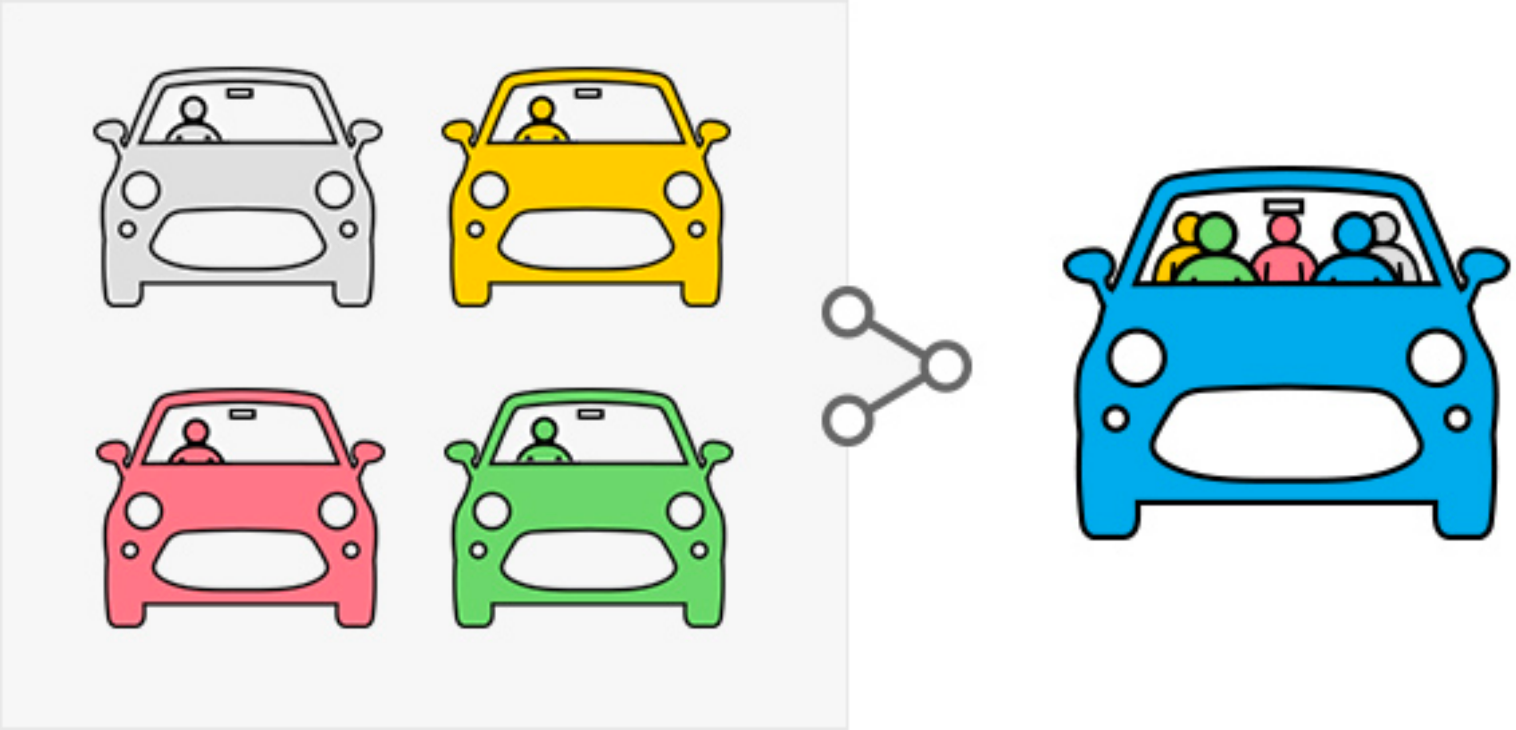
Online advertising

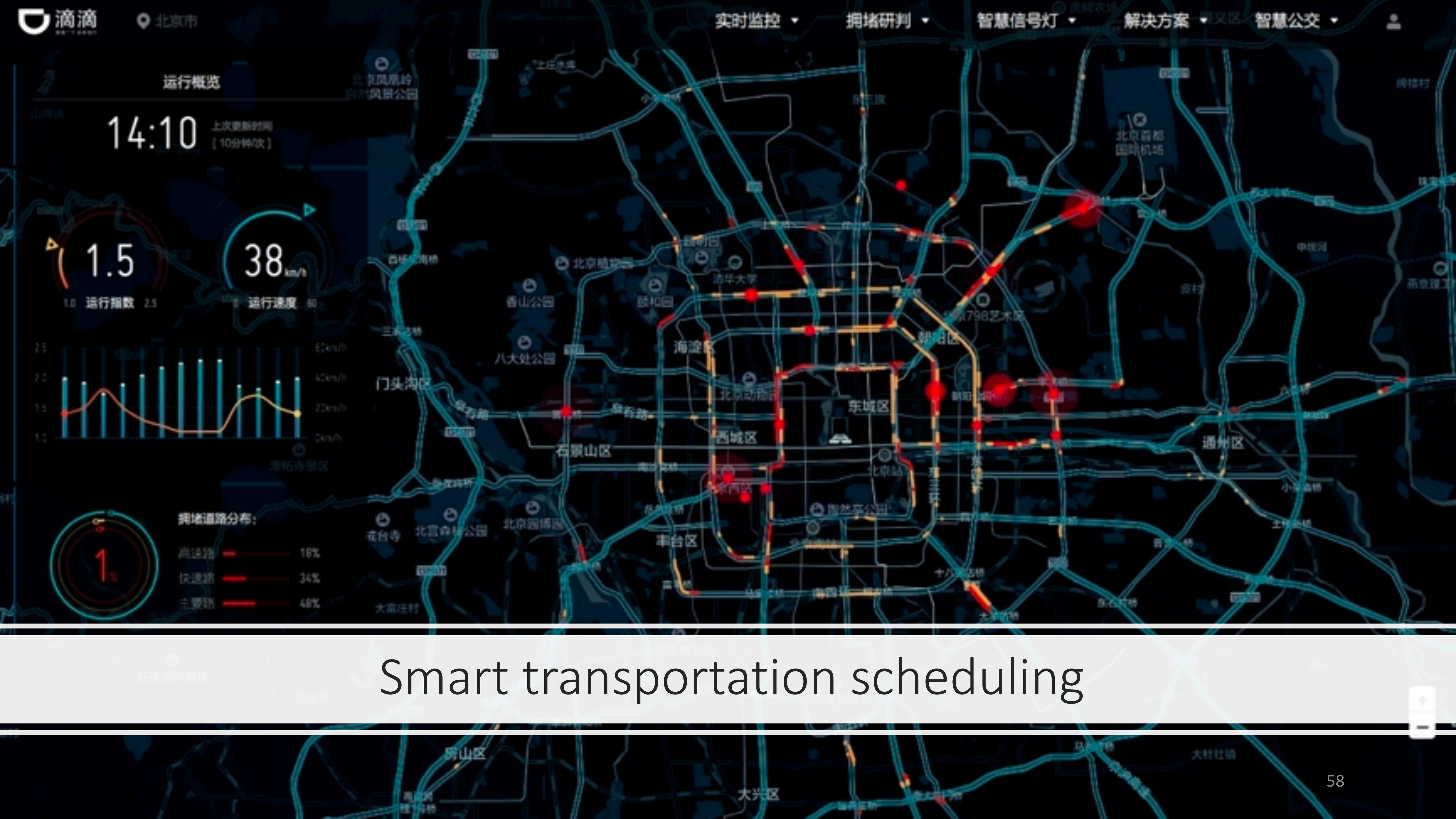
Taobao

- Which ad to show
  - Could attract users
- How to set up the bid price
  - for both the platform and the advertisers

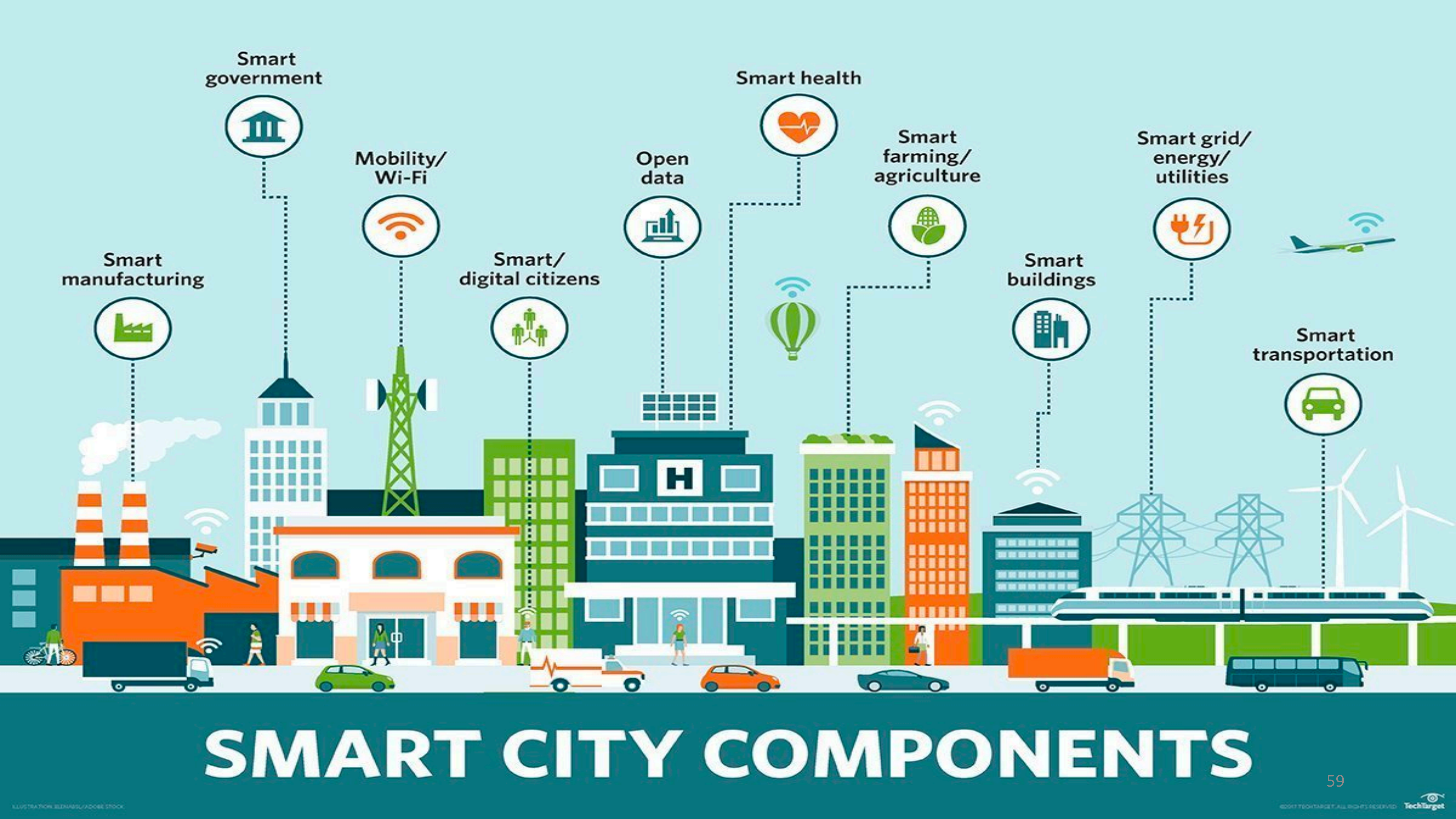
Ride sharing

- Improve traffic





Smart transportation scheduling


Smart government
Mobility/ Wi-Fi
Open data
Smart health
Smart farming/ agriculture
Smart grid/ energy/ utilities
Smart manufacturing
Smart/ digital citizens
Smart buildings
Smart transportation
SMART CITY COMPONENTS

WAIC 2019 Shanghai (Aug 29-31)

WAIC
世界人工智能大会
WORLD ARTIFICIAL INTELLIGENCE CONFERENCE

Strategic partners

Alibaba Group 阿里巴巴集团
IBM
国家电网 STATE GRID
科大讯飞 IFLYTEK
滴滴
国盛集团 Guosheng Group
上汽集团 SAIC MOTOR
商汤 sensetime
Tencent 腾讯
vanke
Microsoft
aws
ICBC 中国工商银行
中国建设银行 China Construction Bank
中国平安 PING AN 金融·科技
中国移动 China Mobile 5G

Summary

- What is Machine Learning and what is Artificial Intelligence
- An example of AI but not ML
  - A* algorithm
- History of ML
  - Deduction
  - Learning from samples (deep learning)
- Recent progress
  - Computer vision/speech recognition/natural language processing/game AI
- Many applications
  - Many industries/many aspects of life

Shuai Li

https://shuaili8.github.io

Questions?

https://shuaili8.github.io/Teaching/VE445/index.html