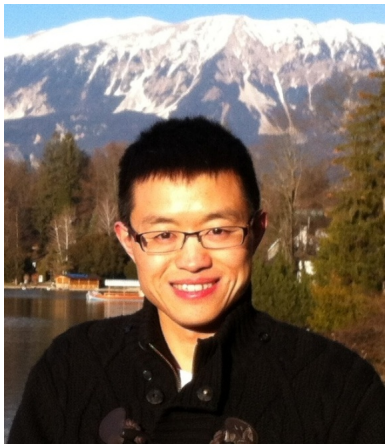
Advanced Machine Learning

Jie Tang
Department of Computer Science & Technology
Tsinghua University

A bit about Jie…



- Jie Tang, Associate Professor, Department of Computer Science of Tsinghua University. My research interests include social network, data mining, and machine learning.
- 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 CCF Young Scientist Award, NSFC Excellent Young Scholar, Newton Advanced Fellowships Award, IBM Innovation Faculty Award, and New Star of Beijing S&T.
- Have published more than 100 paper on major international conf/journals, including KDD (11), IJCAI/AAAI (10), IEEE TKDE (5), ICML, Machine Learning
- #Citation: 4,651 and H-index: 33
- Have a notable system, AMiner.org for academic researcher network analysis. The system has attracted 5.43 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

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 3:00pm-6:00pm

Teaching Assistant

- Head TA: Jing Zhang
  - FIT 1-308
  - E-mail: zhangjing0544@gmail.com
  - Phone: 62788788-20
  - Open hours: 15:00pm-18:00pm Tuesday
  - TAs:
    - Yang Yang, stack@live.cn
    - Zhilin Yang, kimiyoung@yeah.net
    - Cong Ma, 976983150@163.com

Resources

- Course homepage
  - http://arnetminer.org/THU_AML/
- 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

- Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2007.
- Yoshua Bengio, Ian J. Goodfellow, and Aaron Courville. Deep Learning. 2015.
- Daphne Koller and Nir Friedman. Probabilistic Graphical Models. MIT Press, 2009

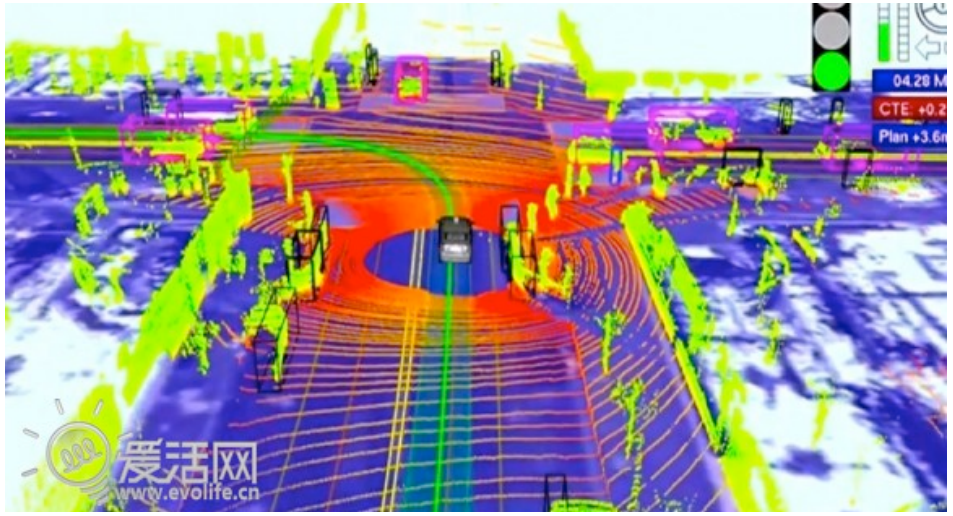
Machine Learning Trends

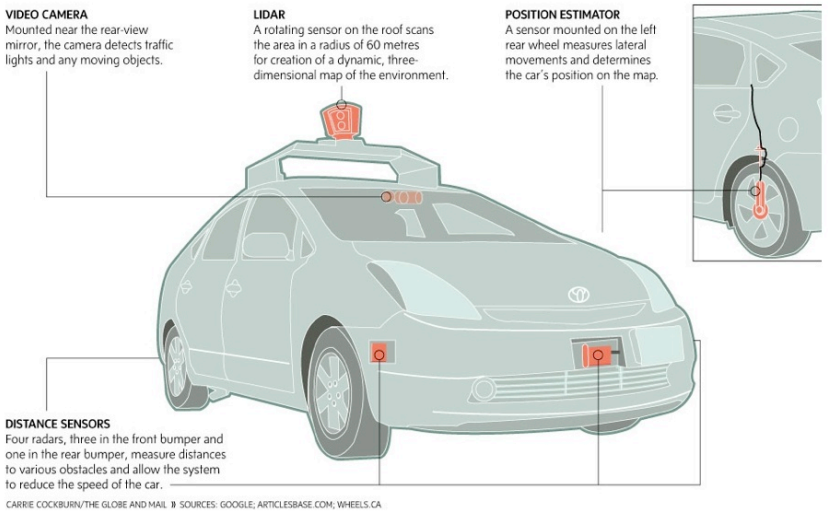
Big Data

Deep Learning

Collective Learning

BIG DATA—Google driverless car





Google's self-driving cars last month hit a 480,000-miles-driven milestone.

--2012/09

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—IBM Watson

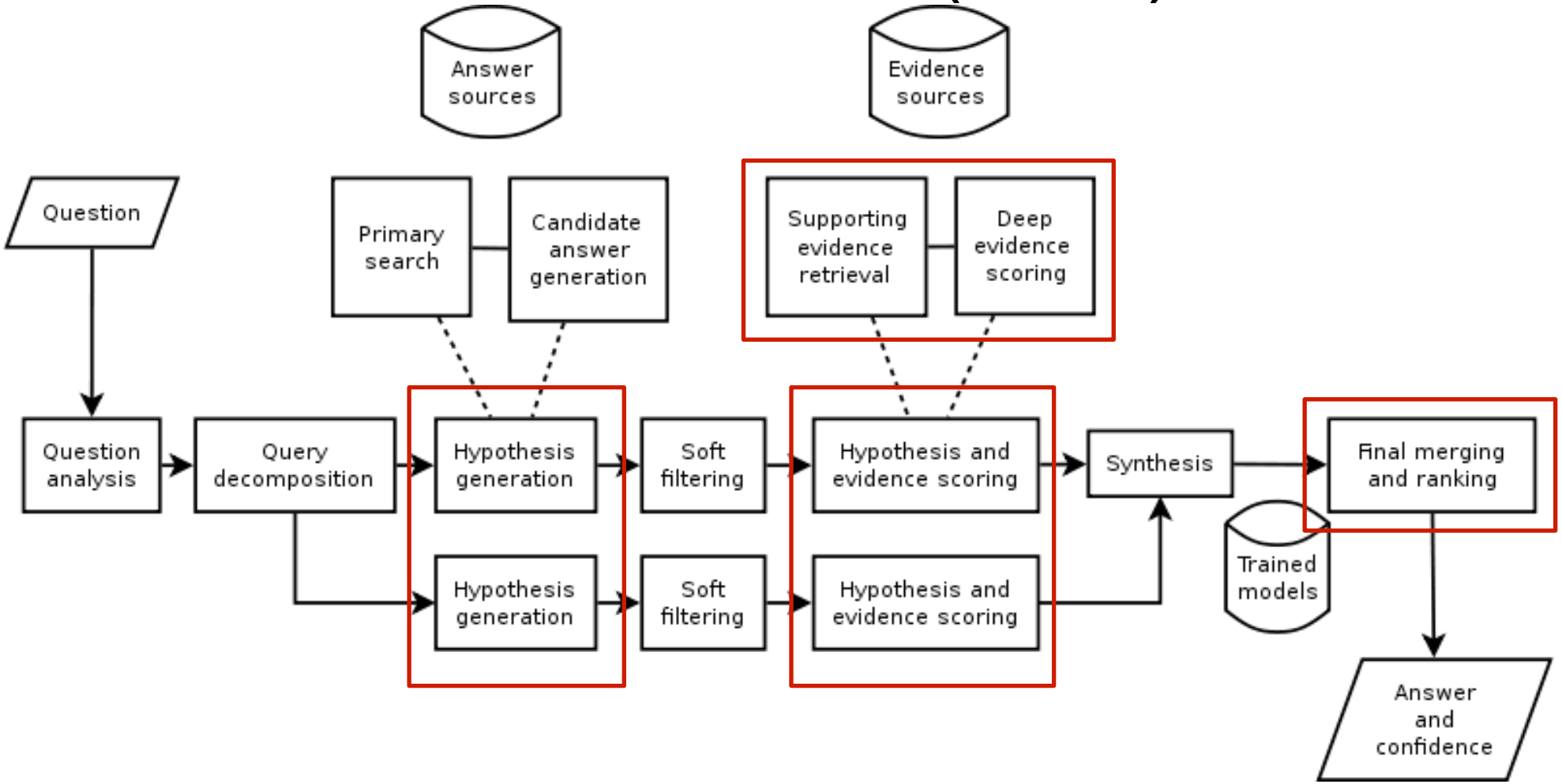


In 2011, as a test of its abilities, Watson competed on the quiz show Jeopardy!, in the show's only human-versus-machine match-up to date

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

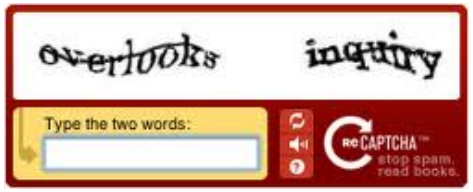
Watson had access to 200 million pages of structured and unstructured content consuming four terabytes of disk storage including the full text of Wikipedia, but was not connected to the Internet during the game. For each clue, Watson's three most probable responses were displayed on the television screen.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

IBM Wastson (cont.)



MECHANISM—Collective Learning

- Learning in Social Networks
  - CAPTCHA + OCR



Lius von Ahn
--AP@CMU

  - DUOLINGO



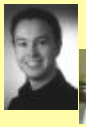
The State of Machine Learning

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 

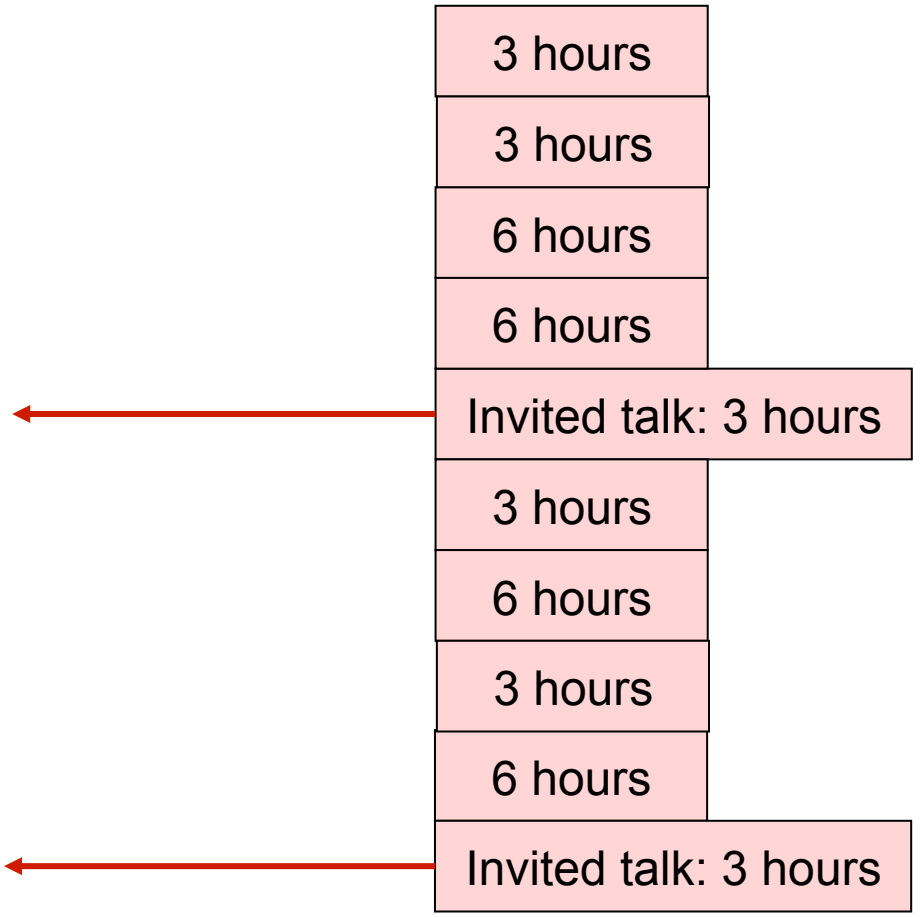
   

For more: http://arnetminer.org/expertisesearch.do?keyword=Machine%20Learning

Overview of Class

Overview of Class

- Introduction
- Basic algorithm review
- Support vector machines
- Probabilistic topic model
- Conditional random fields
- Bayesian nonparametrics
- Sparse learning
- Deep learning



Agenda (3 hour)

- Overview
  - Course overview
  - ML Introduction

Agenda (3 hours)

- Basic algorithms review
  - Data representation
  - Supervised learning
    - Evaluation
    - Perceptron, Bayesian classification, Decision tree, etc.
  - Unsupervised learning
    - Evaluation
    - K-means
  - Reinforcement learning
    - Adaboost
  - Semi-supervised learning
    - Learning with labeled and unlabeled data
    - Active learning, Transfer learning

Agenda (6 hours)

- Support Vector Machine
  - What is a Support Vector Machine?
  - Kernel Trick
  - Extensions of the Binary Classifier
  - The Property of Piecewise-Linear
  - Parallelization & Fitting SVM for large-scale data
  - What are Support Vector Machines Used for?

Agenda (6 hours)

- Probabilistic Topic model
  - Probabilistic Latent Semantic Indexing
  - Latent Dirichlet Allocation
  - Author Topic model
  - Derivation & Implementation

Agenda (6 hours)

- Conditional random fields
  - Hidden Markov model
  - Maximum entropy Markov models
  - Conditional random fields
  - Loopy belief propagation
  - Max-margin Markov network

Agenda (6 hours)

- Non-parametric learning
  - Mont-Carlo methods
  - Gaussian processes
  - Dirichlet process mixture models
  - Indian Buffet processes

Agenda (3 hours)

- Sparse learning
  - Sparse least squares
  - Sparse coding
  - Low-rank matrix factorization

Agenda (6 hours)

- Deep learning
  - RBMs and Deep RBMs
  - Convolutional Networks
  - Recurrent Nets
  - Deep Auto-Encoders
  - Large scale deep learning

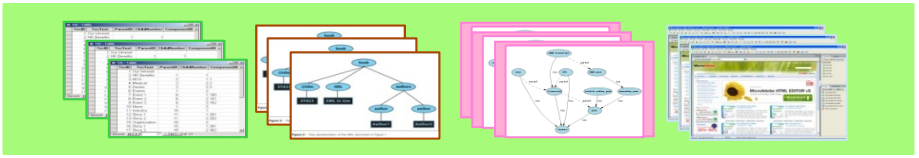
Prerequisites

- 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)

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

Overview of Class (cont.)



Supervised learning

Unsupervised learning

Machine Learning

Lectures on ML

Generative learning

Topic model
Non-parametric

Discriminative learning

SVM
CRF
Max-margin
Sparse coding
etc.

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
  - Hw1: Logistic regression implementation (10)
    - Bonus: Neural networks
  - Hw2: PLSI implementation (10)
    - Bonus: LDA by Gibbs sampling
  - Hw3: Monte Carlo inference for DP mixtures (10)
  - Hw4: Deep RBMs(10)
  - HW5: Link prediction/recommendation (contest) (10)
    - Bonus: Top 10
- Projects
  - Prj1: To design new ML algs. to real problems
  - Prj2: To solve novel problems using ML algs.

4+1

References

- Handbooks:
  - Christopher M. Bishop. *Pattern Recognition and Machine Learning*, Springer, 2007.
  - Yoshua Bengio, Ian J. Goodfellow, and Aaron Courville. Deep Learning. 2015.
- 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.

Thanks!

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