

Introduction to MSBD 5001

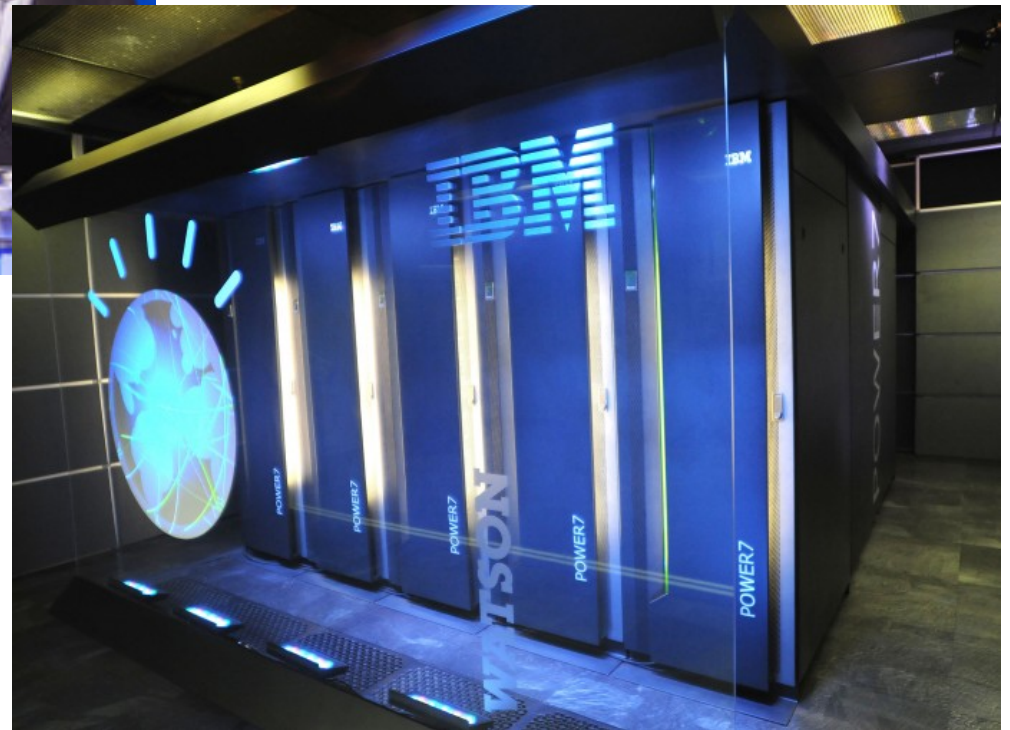
Qiang Yang, CSE, HKUST

Cecia Chan, CSE, HKUST

TA: Bo Liu, CSE HKUST



Big Data in the News



Crowd Intelligence

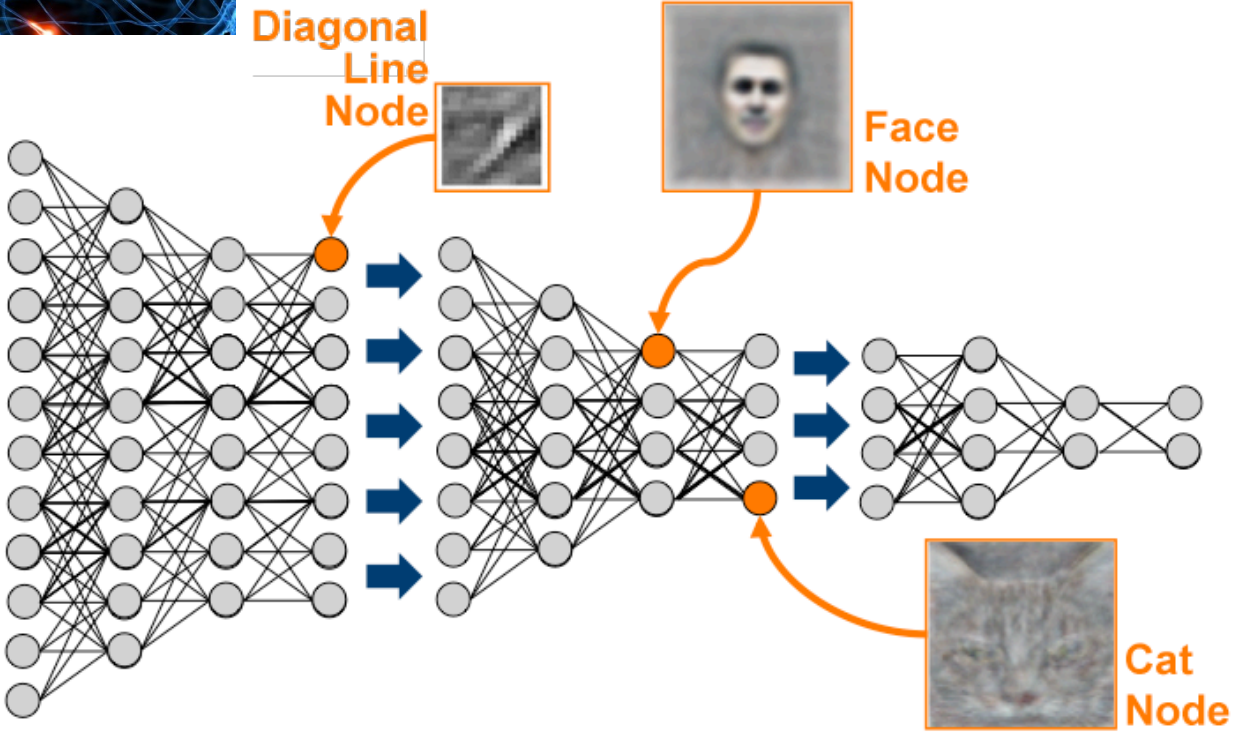
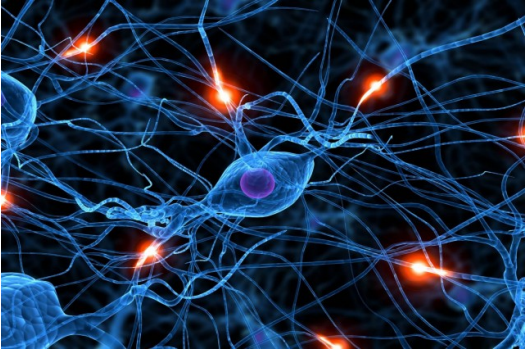


- Natural language
- Conversational
- Contextual
- Personal
- Works with built-in apps
- Dictation anywhere
- 3G + Wi-Fi

engadget



Machine Learning



ABILITY

ARTIFICIAL INTELLIGENCE

Prediction
预测未来的能力



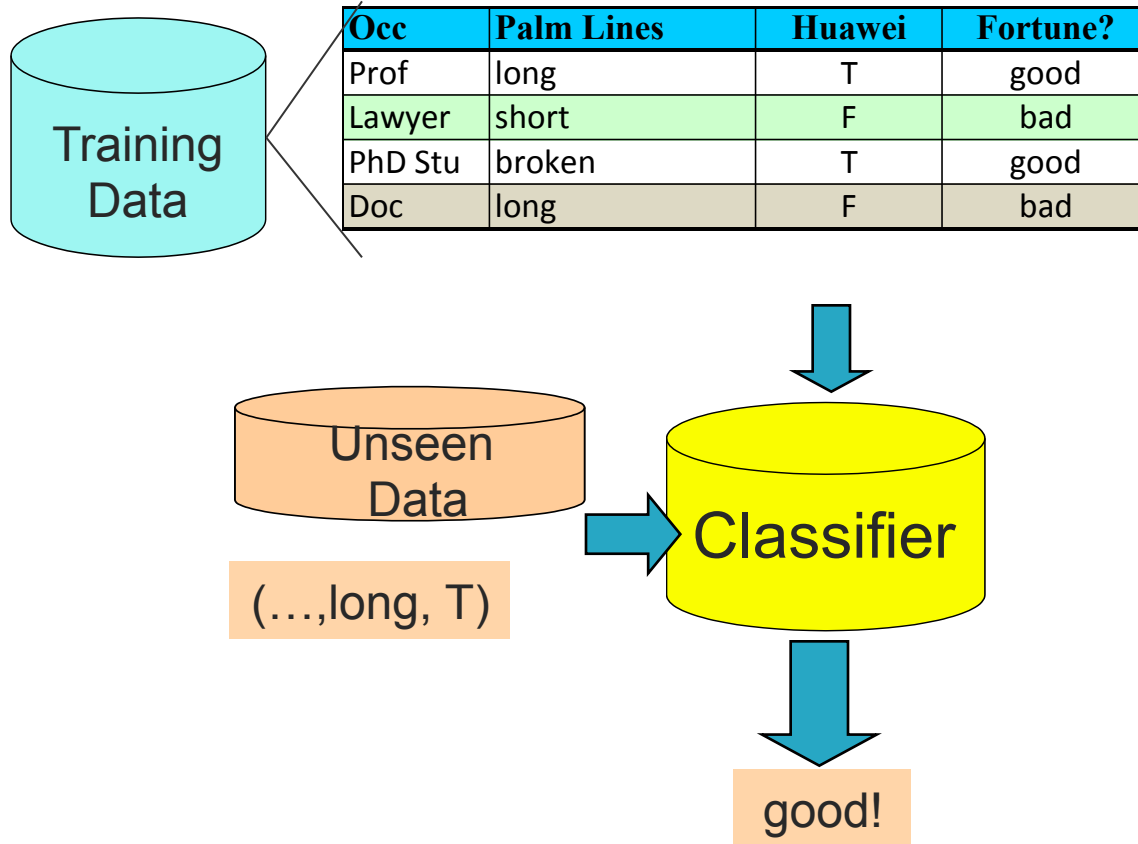
Data Integration
数据、知识的整合的集成与融合



Lifelong Learning
终身学习

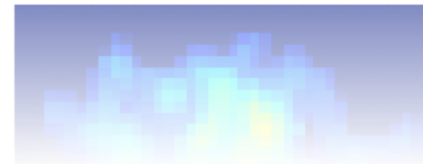
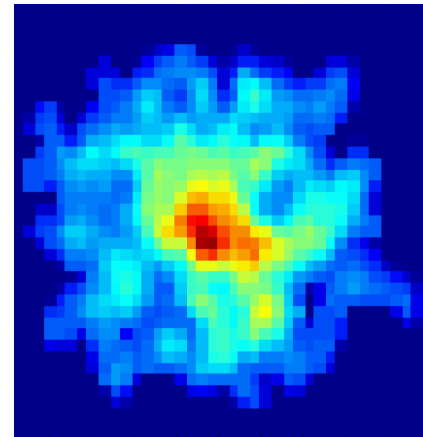
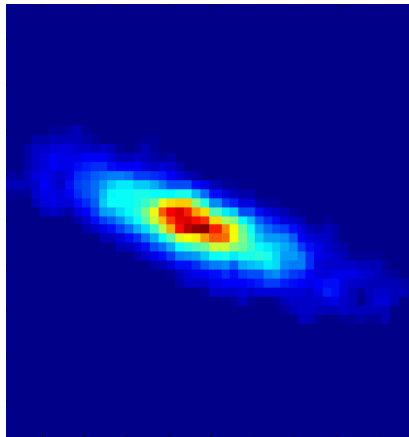
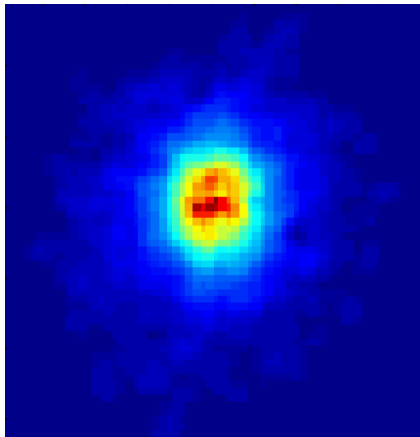


Question: What is common between Machine Learning and Fortune Tellers?

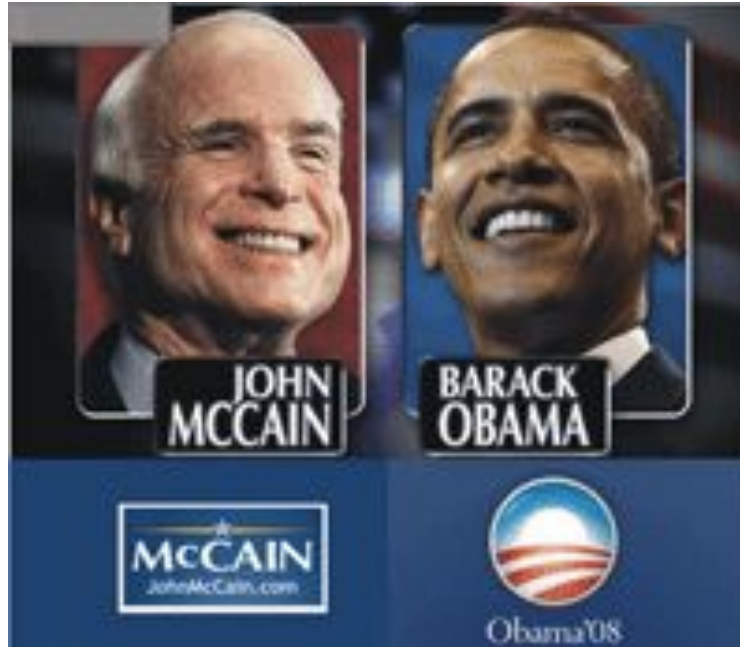


Birth of KDD

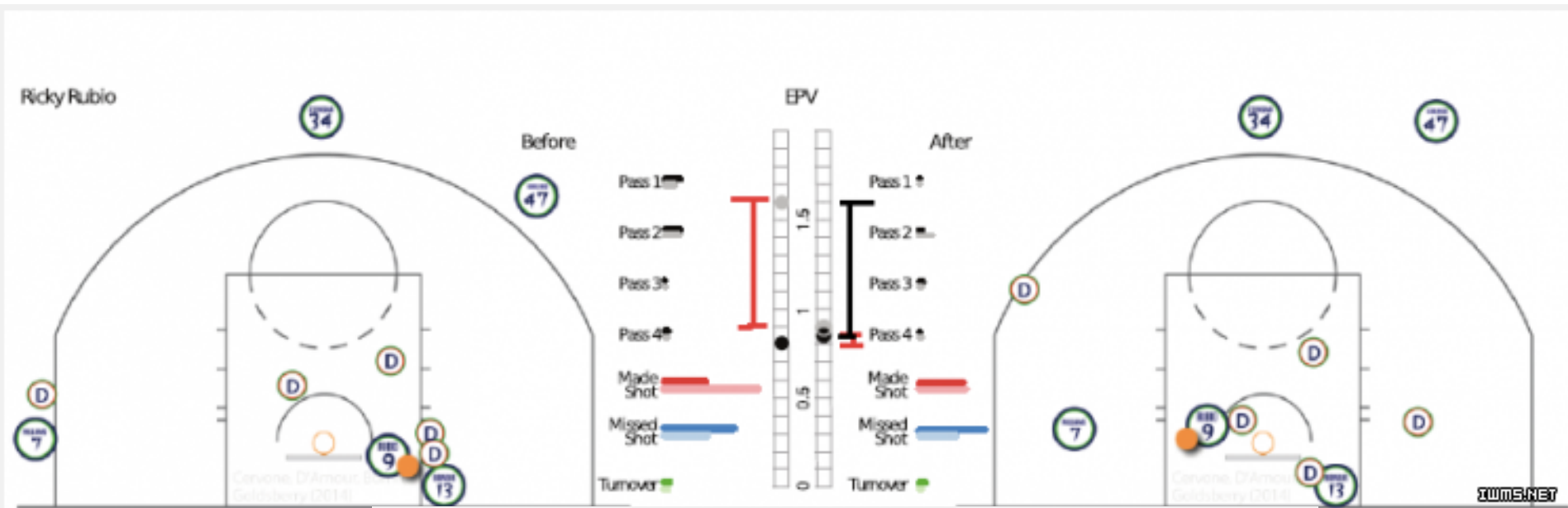
- Usama Fayyad
@NASA: New
Galaxies
- Data
 - 70 Million Stars, 9 GB
 - Photos: 150 GB



Case: US Election



Case: NBA



Player	EPVA	Player	EPVA
Chris Paul	3.48	Ricky Rubio	-3.33
Dirk Nowitzki	2.60	Kevin Love	-2.38
Deron Williams	2.52	Russell Westbrook	-2.07
Stephen Curry	2.50	Evan Turner	-1.90
Jamal Crawford	2.50	Austin Rivers	-1.84
Greivis Vasquez	2.46	Rudy Gay	-1.75
LaMarcus Aldridge	2.40	Jrue Holiday	-1.51
Steve Nash	2.09	Paul George	-1.49
Wesley Matthews	2.06	Chris Singleton	-1.48
Damian Lillard	1.95	Roy Hibbert	-1.44

案例（医学） 基因测序与中医



Case: Law



Case: Education



How to discover cheaters in teachers?

Case: Transfer Learning

A Picture is Worth a Thousand Words

苹果

The apple is the pomaceous fruit of the apple tree, species *Malus domestica* in the rose family *Rosaceae* ...



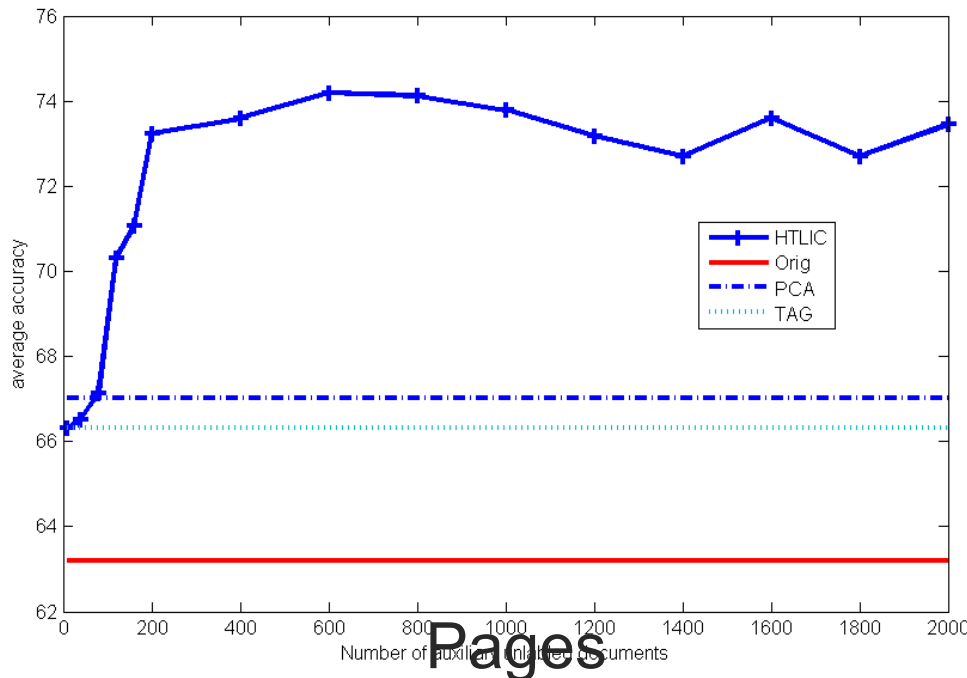
香蕉

Banana is the common name for a type of fruit and also the herbaceous plants of the genus *Musa* which produce this commonly eaten fruit ...



A Picture is Worth a Thousand Words

Acc



➤ How many books is good enough?

➤ 200 Pages

➤ = 100 Photos

➤ 1Photo = 2 Pages

1Photo=1000字！

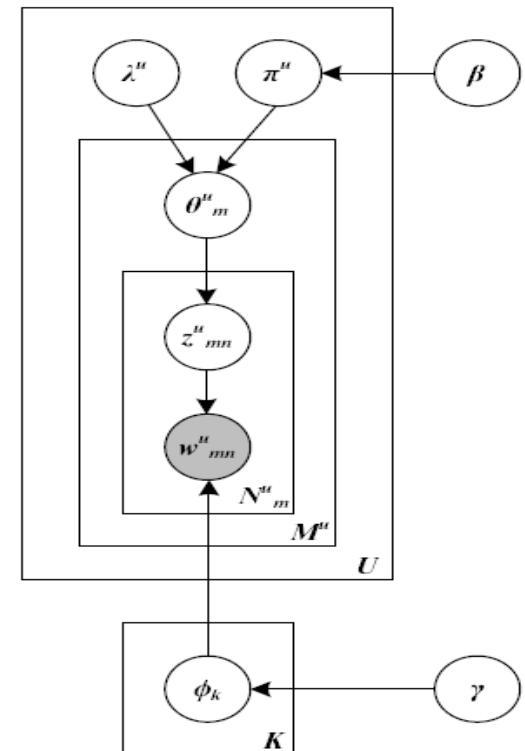
Topic Models and Social Media

User Models

用户模型1 用户模型2

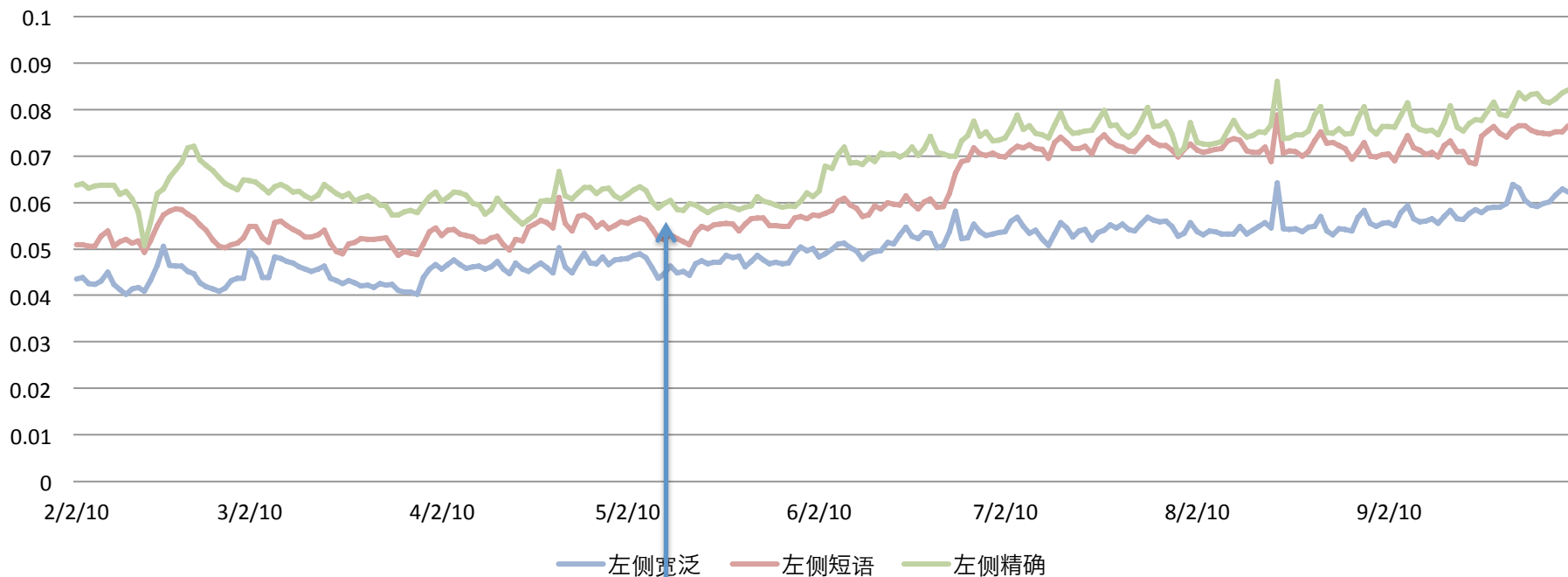


.....



终身学习案例：百度搜索广告

左侧分匹配模式CTR2



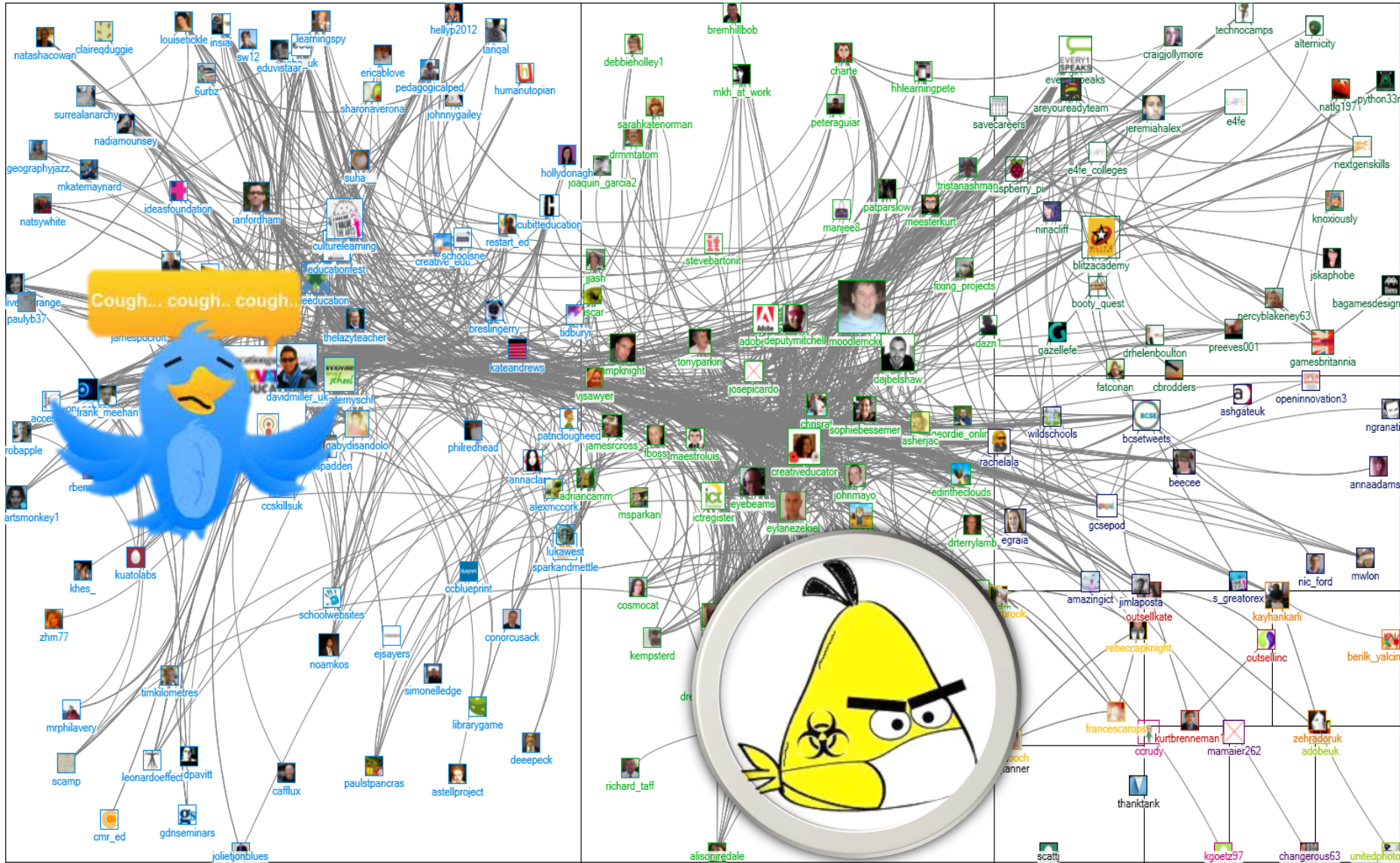
Lifelong Machine Learning算法上线



草堂山水下

渔艇鸟花边







CSE Department Research Areas

- **AI: Artificial Intelligence**
 - Machine learning, Natural Language Understanding
- **Databases:**
 - Database systems, Search Engines, Knowledge
- **NE: Network and Computer Systems**
 - Wireless networks, data centers, cloud computing
- **ST: Software Technologies**
 - Software engineering, multimedia and Internet
- **TH: Theoretical Computer Science**
 - Computational geometry, algorithms
- **VG: Vision and Graphics**
 - Computer vision, graphics, Data Visualization
- **HCI: Human Computer Interaction**
 - Ease of Use, Emotion, Attention

Data Science

Security and Privacy

Big Data at HKUST: Industry Supported Centers

- Smart Cities
 - Thales (ITF)
 - Digital China (ITF)
- Taobao
 - Cybersecurity, Privacy
- HK Jockey Club:
 - Strong Supporter for Big Data Program in Education and Research
- BDBI Center:
 - Big Data for Bio Intelligence (Raymond Chu, \$1M USD/5 years)
- WHAT Lab
 - AI Lab (Wechat, \$10M / 5 years)
- BGI Shenzhen:
 - A team under our supervision



Online Shopping



Smart Cities



Health



Finance & Risk Management



Genetic Farming

Data Mining & Visualization

Big Data Computing & Data Management

Data Integration

Data Extraction



Web Data



City Data



Financial Data



Logistics Data



Agricultural Data



Sensor Data

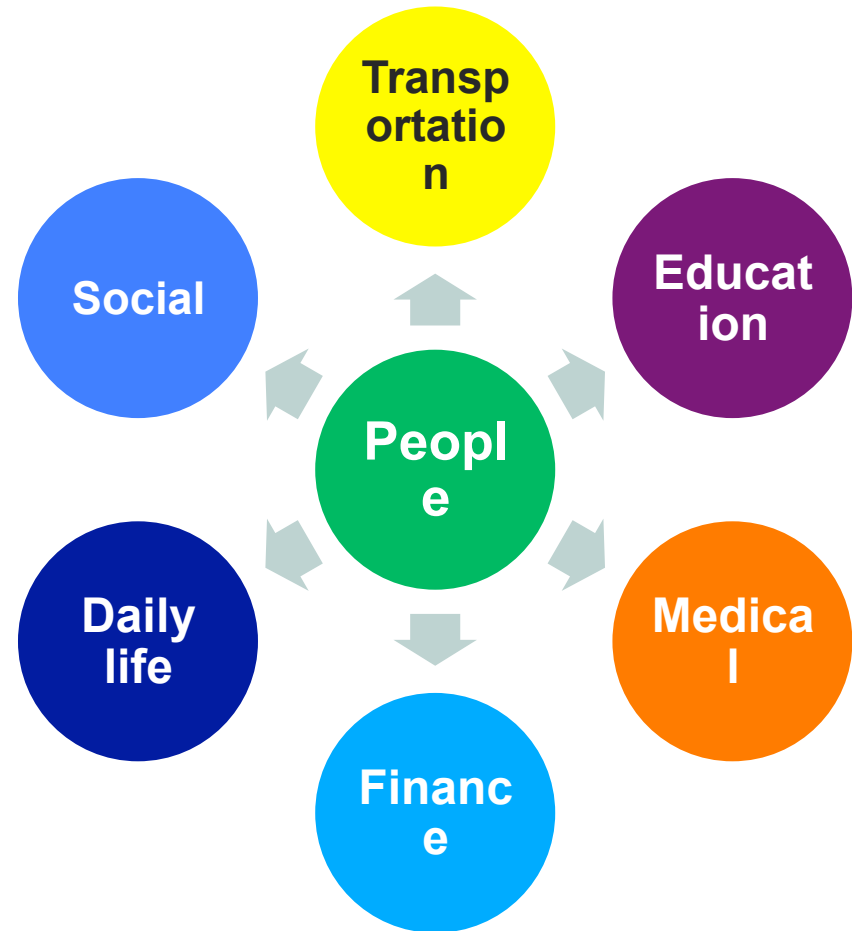
Privacy & Security

Big Data Theory

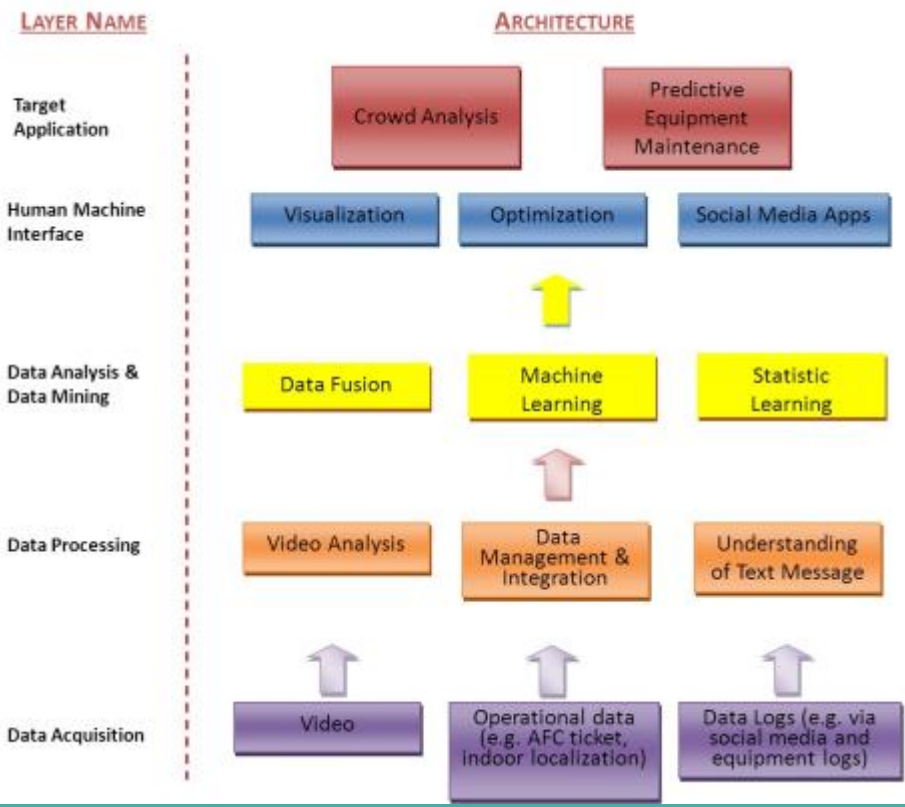
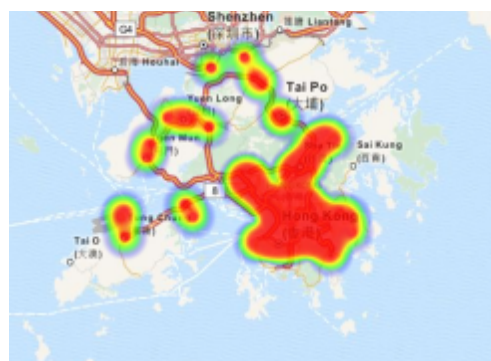
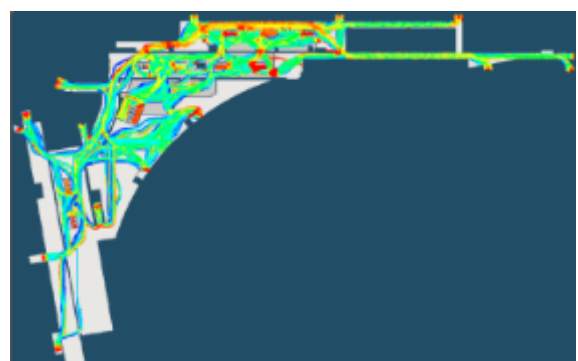


Big Data and Smart City @ CSE

- MOU with Digital China
- Kowloon East Gov
- 20 M HKD Project



Joint Project with Thales and Hong Kong ITF: Big Data Platform for Smart Transportation Applications with Heterogeneous Data Sources



HKSUT + Thales Expertise

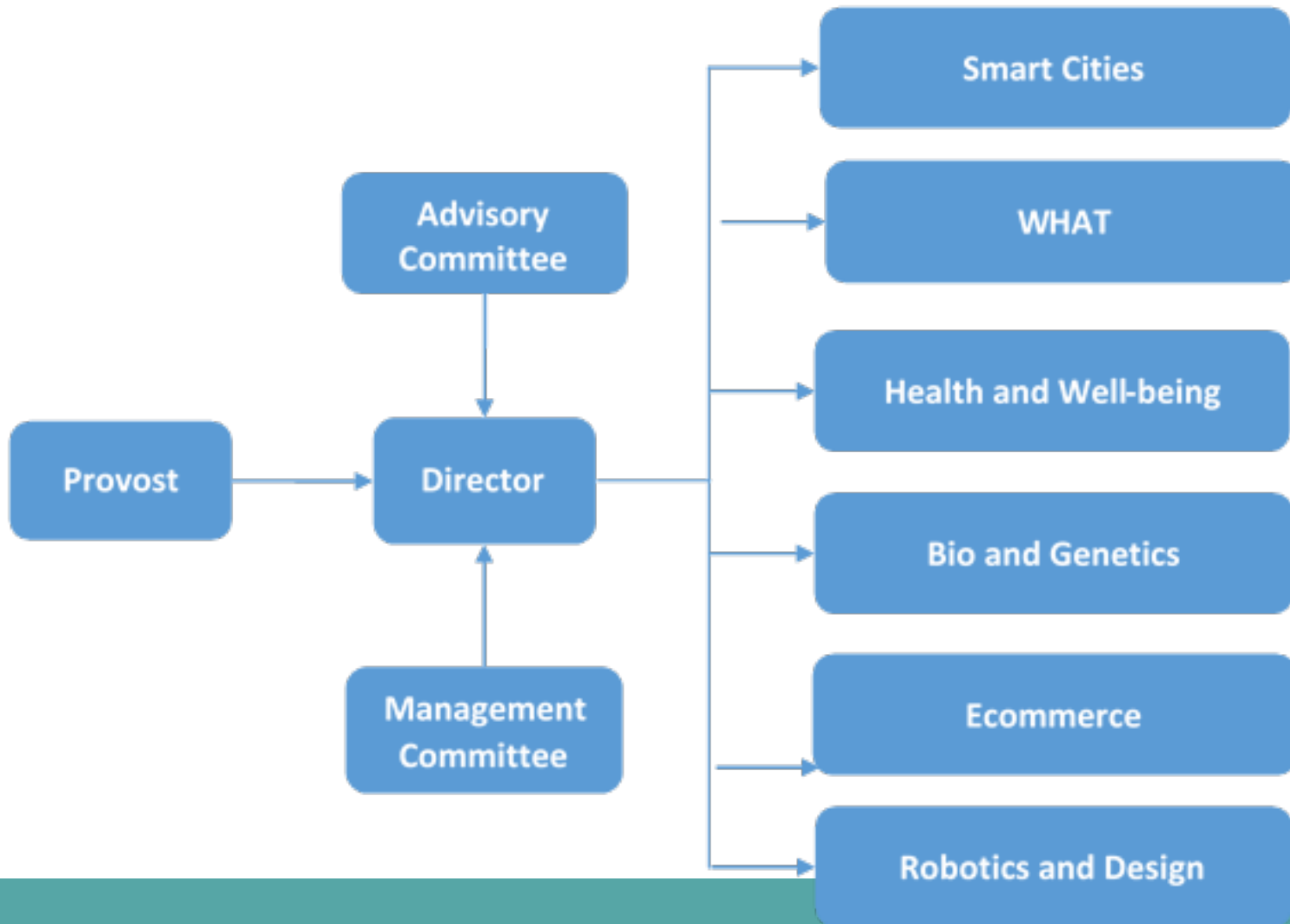
HKUST Expertise

HKUST Expertise

Thales Expertise



Big Data Institute @ HKUST



Course Work

- Lectures and Demos
- Midterm (20%)
- Final (40%)
- Group Projects (40%)
 - Project term paper
 - Presentations and Demos