



Connecting with the Customer

Speaker:

Professor Lionel M. Ni

Chair Professor, Hong Kong University of Science and Technology



Making Big Profits from Big Data

Opportunities and Challenges in Business

Lionel M. Ni
Hong Kong University of Science and Technology



Google Knows You Better Than You Do

What Google Knows

Google compiles enough data to build comprehensive portfolios of most users—who they are, where they go and what they do—and the information is all available at google.com/dashboard. Here are just a few things WSJ reporter Tom Gara found out about himself.

GMAIL
134,966

All of Tom's emails since he first got a Gmail account in 2004. Google also stores his 6,147 chats.

CONTACTS
2,702

Google knows the people that Tom emails the most. At the top is a friend in Egypt.

YOUTUBE
9,220

Videos Tom has watched, listed in chronological order, including a series viewed in June about canoes.

GOOGLE SEARCH
64,019

Google thinks Tom performs most of his searches around 8 a.m. ET, but this is probably skewed by years spent outside the U.S.

ANDROID DEVICES
3

Google knows all of Tom's synched Android phones, including the old Nexus S phone that he gave to his mom.

WALLET
3

Credit cards (two expired) saved in Google Wallet, plus two shipping addresses and 13 itemized purchases since June 2009.

DOCS
855

Documents Tom has created, plus the 115 he has opened that belong to other people.

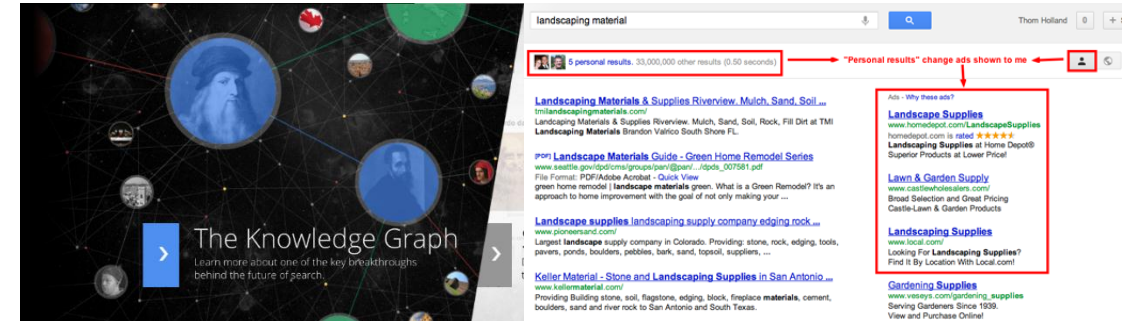
Graphic by
Alberto Cervantes/
The Wall Street Journal

LOCATION

**Willunga,
South Australia**

Due to an unknown glitch, Google bases Tom's location from one of his old Android phones, which he gave to his mother in Australia.

Google
is watching



Personalized ads based on **YOUR** data



**Enhanced click-through rate =
MORE profits!**



Brand Recommendation in Tmall.com (Alibaba)

天猫推荐总体情况

推荐产品 超过**40+** 推荐产品
每天服务 **10M+** 用户, 双11当天服务**36M+** 用户

推荐算法 User2Items Item2Items Personalized Ranking Others

推荐实体



品牌推荐



商品推荐



促销活动

等等



- Serving **10 million+** users every day, 36 million+ users at the busiest day
- **40+ recommended commodities for each user**
- Predicting user preferences from behaviors data

Due to better models, the revenue of Tmall.com increased by 20%+



Alibaba Big Data Competition

User behavior data from Tmall.com (Alibaba)

User	Brand	Date	Behavior
Alice	Lenovo	2014-04-18	Click
Bob	Sony	2014-04-20	Click
Bob	Sony	2014-04-20	Buy
...

- Alibaba organizes a competition to look for **even better prediction models**
- Given the log of user behaviors (including **CLICK**, **BUY**, **BOOKMARK**, and **ADD-TO-CART**) of a certain period
- Predict which users **will buy** which brands at a later time

天池 大数据科研平台						
首页 比赛介绍 赛题和数据 排行榜 社区 FAQ 登录						
排行榜						
每天凌晨更新（第一赛季周排行榜只展示TOP500的成绩，更多成绩请至个人中心查看）						
队名	所在组织	F1评分	准确率	召回率	最优成绩提交日	排名
HKUST	香港科技大学	8.02%	8.13%	7.91%	2014年04月20日	1
StayAwake	中国科学技术大学	7.86%	7.83%	7.88%	2014年04月18日	2
Goahead	中国科学院大学	7.82%	7.65%	8%	2014年04月18日	3
wuyang	香港科技大学	7.79%	7.62%	7.97%	2014年04月18日	4
只有这点程度吗	复旦大学	7.77%	8.31%	7.29%	2014年04月20日	5

We are the No.1 out of 7000+ teams in season 1.

Smart Tour in Canton Tower



A mobile phone app with **augmented reality**



Smart Shopping Mall

Your smartphone will accompany you go shopping

- Indoor navigation
- Personalized and location-based recommendation
- Finding your cars in big parking lots



Better shopping experience with very low cost!



Fraud Detection for Automobile Insurance



- It is estimated that there are **more than 2500** false insurance claims each year in UK.
- **Each** false claim could cause a loss of **up to \$300,000 HKD**.
- In the market, there are **old and expensive** anti-fraud solutions by FICO, SPSS, etc.
- Our solution uses **new machine learning technologies**.

Our solution detects more frauds while cuts the cost by more than a half.



Data-as-a-Service: Revitalizing Payment Transaction Log

FB: \$16B

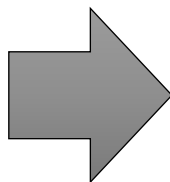
>\$10B



New IT Business



Data Center Infrastructure



Smart Shopping Mall



Microfinance Companies



Data-based Precise Marketing



Credit Scoring Service

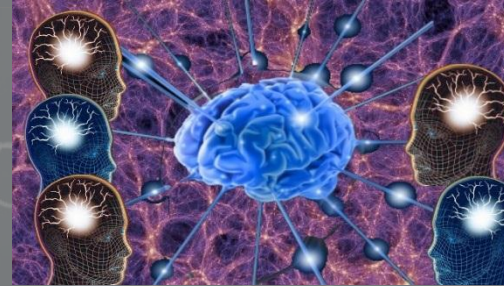
It requires new information technologies



Big Data Collection



Large-scale Data Mining



Advanced Machine Learning



Data



Knowledge



Insight

The point of Big Data is to make sense of it





Challenges of Harnessing Big Data for Business Value

Data Collection

- Extracting siloed data
- **Understanding metadata**
- Mixed structured and unstructured data
- Data cleaning
- **Data calibration**
- Data integration
- ...

Data Processing

- Data loading
- Building large indexes
- **Parallel algorithms design**
- **Fault-tolerance**
- Adapt to new hardware
- Data compression
- **Real-time response**
- ...

Data Mining

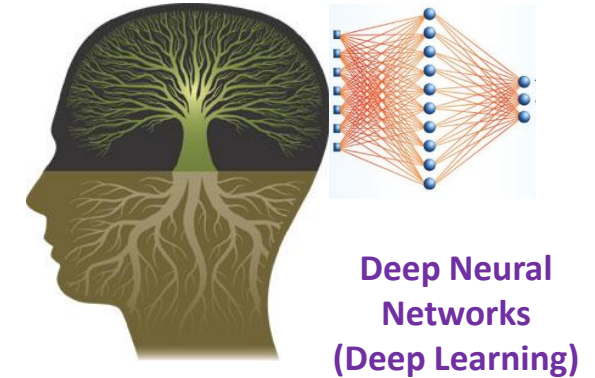
- Statistical analysis
- Data clustering
- **Predictive modeling**
- Ensemble of models
- Abnormal detection
- **Unsupervised learning**
- Data visualization
- ...



“Big Data is a Revolution that Will Transform How We Live, Work, and Think”

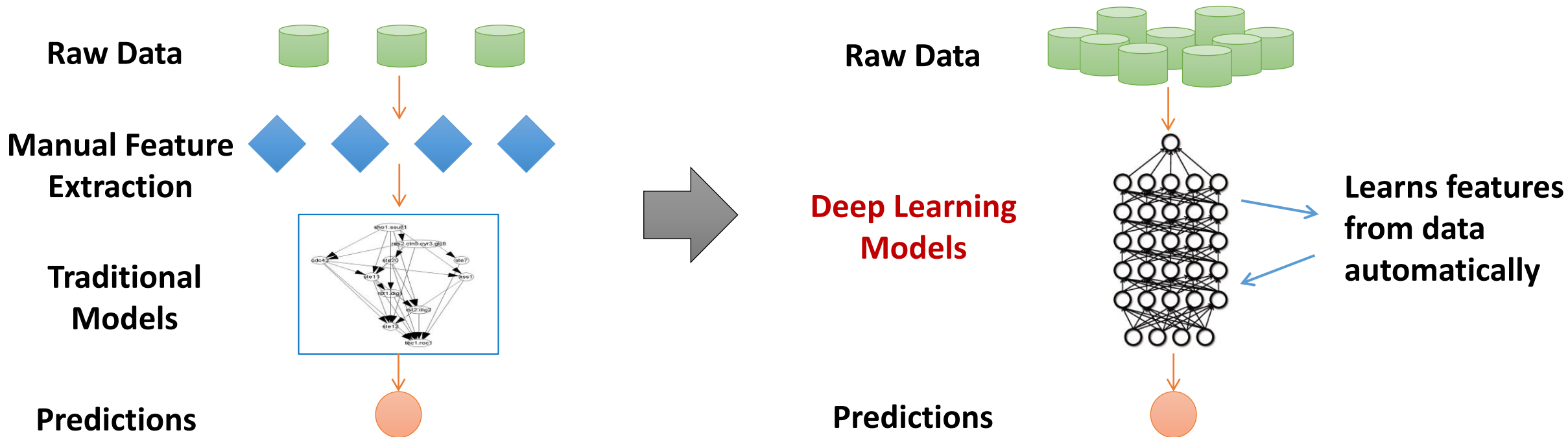
Winners in the Big Data era would be those who:

- Have **as much as possible data**
- Have a creative mind of what the data can do (**the value of the data**)
- Know how to **extract knowledge and gain insight from data**
- Have the computation resources and know how to **process huge amounts of data**





Deep Learning: A Frontier of Today's Artificial Intelligence



Amazing accuracy (> 85%) for image recognition; yet many works remain to be done to apply it for business data.



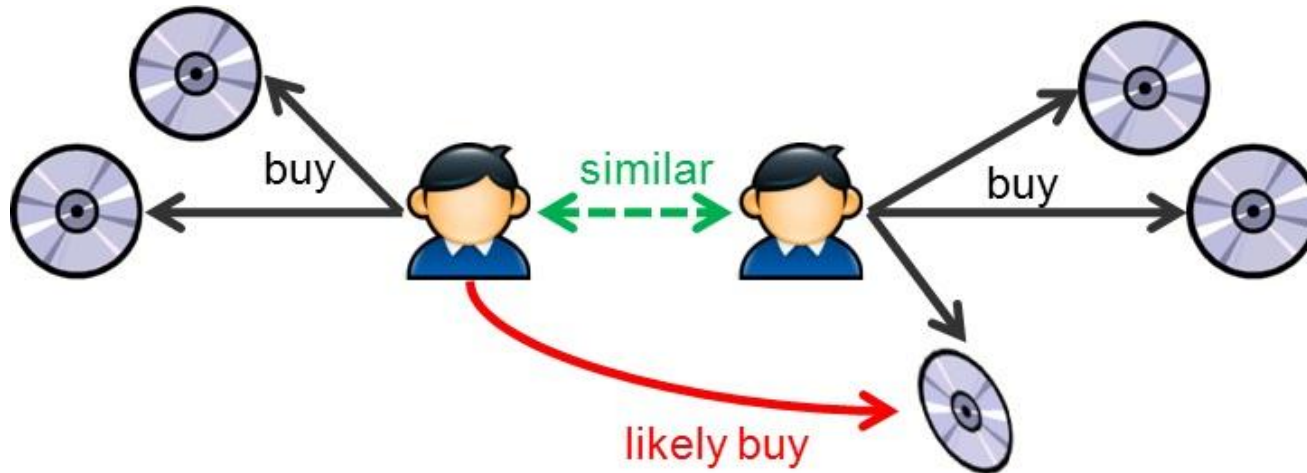
Alibaba Big Data Competition Revisited

Models are no secret

- They are all textbook knowledge.
- Alibaba even let the players know what models they use.

The point is how to well use them

- How similar is similar?
- How many similar users/brands to use?
- How to tune the parameters?
- ...



**Rocket science?
No! It's all art!**



What Companies Should Do

To collect more data

- “Data will **become more valuable than you thought** when you collect it.”
— Dr. Jian Wang, CTO of Alibaba Corp.

To profit from data

- Treating big data as **profitable assets**
- Figuring out **how data will help the business**

To hire experienced data scientists

- Data science is more of **an art** than a science.

The screenshot shows the Harvard Business Review website. The header includes the logo and a search bar. A navigation bar lists categories: THE MAGAZINE, BLOGS, VIDEO, BOOKS, CASES, WEBINARS, and COL. The main article title, "Data Scientist: The Sexiest Job of the 21st Century" by Thomas H. Davenport and D.J. Patil, is highlighted with a red box. Below the title are social media sharing icons and a comment count. A breadcrumb trail shows the path: Newsroom \ Announcements \ Gartner Says Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally... The article is categorized as a "Press Release" and has social media engagement counts (Like: 38, Tweet: 36, Share: 36, +4). The date "ORLANDO, Fla., October 22, 2012" is shown. The article title is repeated in green text and highlighted with a red box. The summary text below reads: "Analysts Discuss Key Issues Facing the IT Industry During Gartner Symposium/ITxpo 2012, October 21-25, in Orlando".

Harvard Business Review

THE MAGAZINE BLOGS VIDEO BOOKS CASES WEBINARS COL

Data Scientist: The Sexiest Job of the 21st Century
by Thomas H. Davenport and D.J. Patil

Comments (87)

Newsroom \ Announcements \ Gartner Says Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally...

Press Release Share: Like 38 Tweet 36 Share 36 +4

ORLANDO, Fla., October 22, 2012 View All Press Releases

Gartner Says Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally to Support Big Data By 2015

Analysts Discuss Key Issues Facing the IT Industry During Gartner Symposium/ITxpo 2012, October 21-25, in Orlando

- New technologies do bring other concerns, such as privacy. **We should face them and solve them, other than escape from them.**
- Technologies will move the world forward, and **there is no way back.**
- The history has told us that those companies **who overlook the technology advances will be out of business, faster than you thought.**



**Do You Still
Remember Kodak?**

Big Data Will Change Business

In the near future:

**Big Data Should Be and Could Be
A Core Competency of Your Enterprise**

Thank You !