## BIG DATA:
## TECHNOLOGIES AND APPLICATIONS

5 Data-Driven Paradigm and Big Data Use Cases

Il-Yeol Song, Ph.D.
College of Computing & Informatics
Drexel University
Philadelphia, PA 19104

DREXEL UNIVERSITY
College of
Computing & Informatics

---

## Data-Driven Paradigm

- How to utilize those raw data
  - *to learn new insights*,
  - *predict trends and changes*,
  - *introduce innovation and market leads*, and
  - *create new opportunities*?

Il-Yeol Song, Ph.D.    7/8/2016    | 2

---

## Data-Driven Paradigm (DDP)

- Big data will change the way we live, work and think (Mayer-Schonberger & Cukier, 2013):



Il-Yeol Song, Ph.D.    7/8/2016    | 3

---

## Data-Driven Paradigm (DDP)

- Changes introduced by DDP (Mayer-Schonberger & Cukier, 2013):
  - (1) Use the whole data, rather than sampling (eg, Walmart)
  - (2) Accept correlation; identification of causality may not be possible (eg. Amazon sales recommendation by people vs products)
  - (3) Accept the probability of prediction (eg, Moneyball)
  - (4) Big data outperforms experts (eg, Amazon reviewers, Wisdom of crowd)
    - Emerging Data-driven marketers
  - *(5) Datafication of things* (eg, car seat data)
  - (6) Leverage data as it is captured, process in real-time
  - (7) Movement of value from physical items to brands, ideas, and intellectual rights

Il-Yeol Song, Ph.D.    7/8/2016    | 4

## Data-Driven Paradigm (DDP)

(5) *Datafication of things*: digital footprints

- Quantify/measure as many *granular data* as possible
- Transform a phenomenon into *quantified data* for tabulation and analysis
  - Data consumable by algorithms
  - Text searchable
    - Enable c*ulturomics* that study human behavior and cultural trends
- Digitization is not the same as datafication
  - Google books: only for reading
- Spice up your proprietary data with third-party data
- Let the data speak for itself

## Data-Driven Paradigm (DDP)

(5) *Datafication of things*: Examples

- Google's page rank algorithm
- Facebook's "likes"
- Two hedge funds started analyzing tweets as signals for investment in the stock market
- Asthma hospitalization rate analysis: EHR data + Twitter data + Weather data
- GE's Leap aircraft engine design research:
  - Trajectory mechanical data + weather data + sensor data
    - Add sensors to engine: generates 1TB data per day
  - An aircraft engine becomes smaller, cleaner and more efficient
  - Saving $3M per year, with less noise and pollution

## Slide 1

**Data-Driven Paradigm**
**(7) Leverage data as it is captured, real-time, no latency**

**TRADITIONAL APPROACH**

Data → Repository → Analysis → Insight

**BIG DATA APPROACH**

Data
Analysis
Insight

Analyze data *after* it's been processed
and landed in a warehouse or mart

Analyze data *in motion* as it's generated,
in real-time

@graemeknows

IBM

## Slide 2

### Strategies of Increasing Value of the Data in DDP

1. Reuse
   - Let others see your data
   - Ex: *Mobile phone companies sell their geo-location data to another company*

2. Merging/Combining different data sets
   - Merge data from multiple sources
   - Ex: *IBM merged data from Honda, Pacific Gas & Electric company – even the weather to predict the best places to install car charging points*

3. Twofers with multiple application options
   - Find a secondary use of data
   - Ex: Google street view: map service, driverless cars, real-estate service

DREXEL UNIVERSITY College of Computing & Informatics · Il-Yeol Song, Ph.D. · 7/8/2016 · 10

## Slide 3

### Data Market

1. Those that own the data
   - Linkedin, Facebook, Amazon, Banks
   - They will control the market

2. Those that have the capacity to analyze the data
   - Decide.com (Purchased by eBay), Technical consulting companies

3. Those with big data mentality
   - Apply unique ideas on how to take advantage of the data and increase value
   - Business consulting companies

DREXEL UNIVERSITY College of Computing & Informatics · Il-Yeol Song, Ph.D. · 7/8/2016 · 11

## Slide 4

### The 4th Paradigm

- Data-Intensive scientific discovery is called the 4th paradigm of science:
  - Discovery is guided by data rather than by a model

    "*The world of science has changed ... data-intensive science [is] so different that it is worth distinguishing [it] ... as a new, fourth paradigm for scientific exploration.*" - **Jim Gray**

  - Examples:
    - Recommender systems in Netflix and Amazon: Pure data (user ratings of movies or products) allows an empirical prediction of what users like.
  - The End of Traditional Science
    http://www.wired.com/wired/issue/16-07

DREXEL UNIVERSITY College of Computing & Informatics

## Dark Side of Data-Driven Paradigm (DDP)

(1) Privacy exposed during data collection
(2) "Probability and Punishment"
(3) Data Dictatorship as an absolute truth

## Dark Side of Data-Driven Paradigm (DDP)

(1) Privacy exposed during data collection

- The classic notice '*I have read and accept the conditions*' is no longer valid.
  – The user only gives permission for the first use made of the information, but not the next, or any unknown use in the future.

## Dark Side of Data-Driven Paradigm (DDP)

(2) "Probability and Punishment"

- Preventive policies already use Big Data to establish which individuals, neighbourhoods, or cities should be subject to increased surveillance to prevent crime

- Big Data's predication ability may be mis-used
  – Ex: Criminal prediction

- **There must be monitoring performed by people to ensure that individuals are judged on their deeds, and not by what an algorithm predicts they might do**.

## Dark Side of Data-Driven Paradigm (DDP)

(3) Data Dictatorship as an absolute truth

- Big Data analysis results are not the absolute truth
- Monitoring and transparency are emerging solutions
  – A new types of professional called "the Algorithmist" is needed to perform the monitoring and transparency
  – They should **guarantee confidentiality and impartiality with a Hippocratic oath similar to doctors and lawyers**.
  – Their work should focus on monitoring current Big Data practices

## Values in Big Data Projects

- **Automate decision-making**
- **Generate deeper business insights**
  - Who are my most valuable customers?
  - Who are my most influential customers?
  - What are most important products?
  - What are my most successful campaigns?
- **Optimize**
  - Reduce costs
  - Mitigate risks across operational and financial aspects
- **Personalize**
- **Design new processes**
  - Uncover new revenue opportunities
  - Garner new products, customers, markets

Il-Yeol Song, Ph.D.  
7/8/2016  
17

## Creating Big Data Projects

1. **Add more detailed transaction data**
2. **Add unstructured data**
3. **Add low-latency, real-time, data**
4. **Integrate predictive analytics**

Il-Yeol Song, Ph.D.  
7/8/2016  
18

## Getting *Values* in Big Data Projects

1. Data-driven culture
   - Look at the data dispassionately, without replying on intuition
   - Emergence of data-driven marketers and data-driven economy
   - Ex: Apple, Google, Wal-mart, with their supply-chain management
2. Analytical talent around data science
   - Innovations can come from doing diagnostics, predictive, and prescriptive analytics
3. Understanding on big data solutions
   - Hadoop ecosystems, NoSQL
5. Leadership
   - Data analytic life cycle: from business strategy, questions, data, solution, evaluation, monitoring): CDO disciplines

Il-Yeol Song, Ph.D.  
7/8/2016  
19

## Chief Data Officer (CDO)

- A CDO provides **vision** and **strategy** for all data management initiatives:
  - Is a champion for **global data policies, standards, governance, quality**, data source management, education, and vendor relationships across the enterprise
  - Identifies **business questions** and **metrics** in business context
  - **Oversees big data project roadmap and workflows** from conceptual analysis to deployment

Il-Yeol Song, Ph.D.  
7/8/2016  
20

## Chief Data Officer (CDO)

- A CDO provides **vision** and **strategy** for all data management initiatives:
  - Is a champion for **global data policies, standards, governance, quality**, data source management, education, and vendor relationships across the enterprise
  - Identifies **business questions** and **metrics** in business context
  - **Oversees big data project roadmap and workflows** from conceptual analysis to deployment

Il-Yeol Song, Ph.D.    21

## Who is a Chief Data Officer (CDO)?



Il-Yeol Song, Ph.D.    22
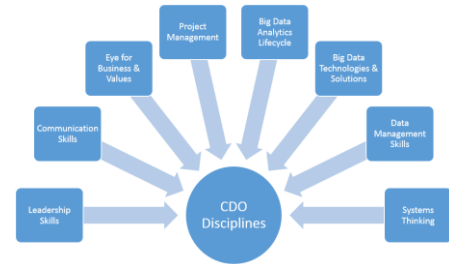
## Use Cases in Big Data Projects

- **Marketing and sales growth**
- **Operational and financial performance improvement**
- **Risk and compliance management**
- **New product and service innovation**
- **Direct/Indirect data monetization**
- **-- (Gartner, 2015)**

Il-Yeol Song, Ph.D.    23

## Big Data Use Cases

- Healthcare
  - Early detection of a disease (e.g., Alzheimer)
  - Customized drugs based on patient's history
  - Early detection of epidemics with crowdsourcing
  - Smart health projects with care networks for older people
  - Integrating genomic analysis with healthcare
  - Disease prevention, flu forecast and prevention
  - Detecting abnormal situations in ICU
  - IBM Watson (Seton Health Care Family use Watson to learn 2M patient data annually)

Il-Yeol Song, Ph.D.    7/8/2016   | 24

6

## Big Data Use Cases

- **Customer analysis**
  - Personalized coupon
  - Fraud detection for IRS, social security claims
  - Churn analysis
  - Better user profiling, more targeted marketing
  - More customized (optimized) pricing and automated bidding on a number of exchanges
  - Geo-marketing via cell phone (restaurants, retail)
  - Expand your existing customer analytics with social media data
    - They influence each other

## Big Data Use Cases

- **Web**
  - Better taxonomies to classify text (news, user reports, published articles, blog posts, yellow pages etc.)
  - Detection of duplicate and fake accounts on social networks

- **Education and Academic**
  - Customized, on-demand, online education with automated grading
  - Better detection of fake reviews for systems based on collaborative filtering (in short, superior collaborative filtering technology)

- **Crime prevention**
  - Criminal protection by predicting likely locations of criminal activities

## Big Data Use Cases

- **Disaster/risk prevention/detection**
  - Fire prevention based on geodata, household data, lifestyle data
  - Preparation against Tsunami, taipoon
  - Detection of earthquakes, solar flares - including intensity forecasting
- **Public service improvement**
  - Optimizing electricity trading and delivery
  - Smart utility meters with sensors
  - Drafting new laws from complaints and social phenomenon
  - Personalized labor support system (Germany, saving 10B euro saving)
  - Compliance detection using its event management solution (HP)
- **Terror prevention**
- **Defense application**

## Big Data Use Cases

- **Technical applications**
  - Image classification (Facebook, Flickkr)
  - Voice recognition and NL processing (Siri, Watson)
  - Body movement recognition (Robot)