Fall 11-19-2014

Should We Love or Hate Big Data? The Good, the Bad, and the Ugly

Dennis Crossen M.Sc., MBA
La Salle University, crossend@lasalle.edu

Karti Puranam PhD
La Salle University, puranam@lasalle.edu

Madjid Tavana PhD
La Salle University, tavana@lasalle.edu

Follow this and additional works at: http://digitalcommons.lasalle.edu/explorercafe

Part of the Applied Statistics Commons, Instrumentation Commons, Technology and Innovation Commons, and the Vital and Health Statistics Commons

Recommended Citation
Crossen, Dennis M.Sc., MBA; Puranam, Karti PhD; and Tavana, Madjid PhD, "Should We Love or Hate Big Data? The Good, the Bad, and the Ugly" (2014). Explorer Café 28.
http://digitalcommons.lasalle.edu/explorercafe/28

This Article is brought to you for free and open access by the Explorer Connection at La Salle University Digital Commons. It has been accepted for inclusion in Explorer Café by an authorized administrator of La Salle University Digital Commons. For more information, please contact careyc@lasalle.edu.
“Should we love or hate big data?”

Professors Tavana, Crossen, Puranam
Business Systems & Analytics Department
What is Big Data?
What is Big Data?

- **It is the accumulation and analysis of information.**
- **Lots of information. Oceans of information!!**
  - Every time someone clicks on something at Amazon, it’s recorded and another drop of is added to the ocean.
  - Every time a scanner beeps at the supermarket checkout.
  - Every time a home electricity meter reports a reading.
  - Every time a parcel passes a FedEx checkpoint.
  - Every time a customs officer checks a passport.
  - Every time someone posts to Facebook
  - Every time someone does a Google search

— *the ocean swells*

---Dan Gardner, “An Ocean of Data”
Big Data

- In the 15 minutes it takes to give this presentation 20 petabytes of data are generated.

- 20 Petabytes
  - =1 trillion GB
  - It is three times the amount of information currently in the Library of Congress.
  - About one-half of all written works from the beginning of recorded history in all languages.

- The world’s total data is doubling every two years.
THE AVERAGE PERSON TODAY PROCESSES MORE DATA IN A SINGLE DAY THAN A PERSON IN THE 1500’S DID IN AN ENTIRE LIFETIME.
Defining Dimensions of Big Data (The 3 Vs):

**Volume**
- Data, data, everywhere!
  - Transaction-based data, social media data, machine data...

**Velocity**
- Data is streaming in at unprecedented speed - near-real time
  - Smartphones, RFID tags, sensors, smart metering...

**Variety**
- Data today comes in all types of formats
  - Numerical data, unstructured text documents, email, video, audio...

The 4th dimension -> Veracity.
Worlds are a-changin’

- We are shifting from a world in which we “know,” because we sampled a little and extrapolated a lot, into a world in which we know
- where all data is collected, analyzed, and stored.
- We are all becoming citizens of the new realm of truly Big Data.
- The optimal decision maker will combine the insights from data analysis and his/her gut-feeling.
Big Data in Politics
President Obama’s re-election campaign

- His 2008 campaign used a lot of data but they had too many disparate databases. (get-out-to-vote and fundraising)
- His 2012 campaign combined everything into one and were able to draw correlations and conclusions like never before.
  - E.g. On the west coast females aged 40-49 would attend a fundraiser dinner with George Clooney with a much higher probability than others
  - Who should an individual receive emails from (Obama, Biden or Michelle?)
President Obama’s re-election campaign

- **Predicting turnout by districts or even neighbourhoods**
  - E.g. They collected data from 29,000 Ohio citizens allowing for deep analysis of who was going to vote and where volunteers would be needed for get-out-the-vote.

- **Predicting responses to campaign events**
  - It was here they learned Obama’s lacking performance in the first presidential debate hadn’t been as detrimental as pundits claimed.

- **Targeted ad-buying**
  - Instead of relying solely on ad-consultants they used data to figure out where to buy ads.
  - Unconventional ad slots e.g. “Sons of Anarchy”, “The Walking Dead” and “Don’t Trust the B—- in Apt. 23” instead of the evening news.
Predicting Election Results

- **Nate Silver** predicts election results accurately
  - 49 out of 50 states correctly in the 2008 election
  - 2010 midterm election
  - 50 out of 50 states correctly in 2012.

- The Princeton election consortium also predicted these successes with high accuracy.

- The idea is to predict an unobservable variable (the intended voting behaviour in each state) which was used to predict the actual vote which is observable.

- The other aspect was deciding which states were safe for which candidate and which states are toss-ups.
Big Data in Healthcare
Big Data in Healthcare

- Epidemiological studies like never before.
  - E.g. flu patterns, TB mutations.
- Innovations in treatment
  - What treatments are most effective for particular conditions,
  - Identify patterns related to drug side effects or hospital readmissions
  - Help patients and reduce costs using EHRs.
- Standardization of medical prices
  - The ACA requires more transparency in prices paid to healthcare providers
- Nutrition
  - The idea of food deserts and identifying where they exist is a byproduct of Big data.
  - Targeted food drives
Big Data
... Across the Universe ...
Good Data

- Arecibo, Puerto Rico
  - 1974 to closest star cluster (at that time)
  - Communicate with ET
  - 25,000 Light-Years
  - (146,965,638,531,210 miles)
  - Transmitted (23x73 pixels)
  - FM (2380 MHz) / 1679 binary digits
  - Duration = ..... <Guess>
- Cray Supercomputer
  - 2.3 petaflops
  - 1 petaflop = quadrillion floating point operations per second
Good

- **Kwajalein**
  - Deep-Space Tracking
  - Satellite Communications
  - Missile Launching
  - Weather Relays

- **Space Fence**
  - Track debris
  - Detect “Space junk” approx. 1200 miles
  - Detects “baseball size” material
**Good**

- **NASA**
  - Now (in sec): 1.73 Gb
  - Deep-Space: Mb/s
  - Earth orbiters: Gb/s
  - RF (slow) ... Lasers (faster)
  - 24 TBs/day (approx. 2.4 Library of Congress/day)
  - (1 TB = 1,000,000,000,000 bytes)
  - Climate Change Data: 350 PBs by 2030
    (5 PB = all USPS letters in 1-yr)
• ESA Rosetta
  ○ 10 year journey
  ○ 6.4 Billion KM
  ○ 40,000 mph
  ○ Ground communications approx. 50 min

THE 4BILLION-MILE SPACE RACE

1. Mar 2004: Launched aboard Ariane 5 from French Guiana
2. Mar 2007: Accelerated using Mars’s gravity to pick up speed
3. Aug 6, 2014: Rosetta catches up with Comet 67P after 4 billion mile journey
4. TODAY: Philae landing craft will be sent to surface of comet, 315 million miles from earth

ROSETTA FACTFILE
Size: 10ft x 6ft x 6ft
Wingspan: 105ft
Weight: 3.3 tons
Speed: 23.4 miles/sec (84,000 mph)
Payload: Philae lander and a nickel disk engraved with first three chapters of Genesis, microscopically engraved in 1,000 languages

PHILAE LANDER FACTFILE
8:35am GMT: The lander will be dropped from a height of 13.6 miles. It is expected to arrive at the comet at 3:30pm. It will be anchored by harpoons driven into surface
Size: 3ft x 3ft x 3ft
Weight: 220lb (but low gravity means it is equivalent to 1/30th ounce on the comet)
Power: Battery and solar panels
Payload: 20 instruments including cameras and a drill to take samples of rock

COMET FACTFILE
Name: 67P/Churyumov–Gerasimenko
Discovered by: Klim Ivanovych Churyumov in 1969
Length: 3.1 miles
Width: 1.6 miles
Surface temp: -70°C (South Pole average in winter -60°C)
Orbit of Sun: Every 6.4 years

© MailOnline
• **Easter Island**
  - 1996: Crash of Russian orbiter
  - Magnetic Observatory
  - Infra-sounds (spikes @ 1 hz)
  - Space-weather analytics
  - Emergency Landing site for Space Shuttle
Good

- Easter Island
  - Meteorological station (not really)
Bad

- Malaysian Airline
  - Transponder turned off ???
  - Satellite handshakes ???

What happened to MH370?

1. Last signal received by air traffic control.
2. Last military radar contact. Communication system disabled. Classic Aero system still sends hourly 'pings' to satellite.
3. The pings to the satellite continued for at least five hours. 'Doppler effect' (changes in the signals) allowed south location corridor to be mapped.
4. Data suggests MH370 was travelling south at a steady cruising altitude above 30,000 feet before crashing into the ocean.

Search areas

Southern corridor (last signal transmitted to satellite)
Bad

- Governments Over-reaching
  - Edward Snowden
  - Wiki-Leaks
  - UK “Big-Brother” alert
• **“Spygate”**
  - Using data to illegally obfuscate (aka: “steal”) proprietary information from competitors

• **Remedy**
  - Usually jail-time
  - ... unless ... you are ...
  - Lord Sidious
Questions