(Big?) Data Driven Research

Vinayak Deshpande
Kenan-Flagler Business School
University of North Carolina
Three Research Projects

• Linking Airline flight delays to flight schedules
  - Joint work with Mazhar Arikan

• Supply Chain Management at US Coast Guard
  - Joint work with Ananth Iyer

• Enhancing customer experience at retail stores: Using video feed data
  - Joint work with Saravanan Kesavan
Airline Flight Delays: Data from BTS

- **Airline On-time Performance Data**: Published by the Bureau of Transportation Statistics (BTS).
- **Aircraft Registry Database**: Published by the Federal Aviation Administration (FAA).
- **Airlines Fares Consumer Report**: Published by the Department of Transportation (DOT).
- **T-100 Domestic Market Data**: Published by the BTS.
- **ASPM (Aviation System Performance Metrics) Database**: Published by the FAA.
- **National Climatic Data Center Database**: Published by the National Oceanic and Atmospheric Administration (NOAA).
Is this Big Data?

• Data set consists of approximately 100 million records (one for each flight flown in the US for the last 15 years) with about 10,000 columns. Data set is close to one Terabyte in size.

• Size and type of data poses some unique challenges
Big Data is Dirty!

• Valdosta, GA (VLD) – Atlanta, ATL. July 15 2005, Atlantic Southeast Airlines Flight Number =413, Air-time= -1405.

• Baltimore, MD (BWI) – Austin, TX (AUS). July 29 2005, Southwest Airlines Flight Number =111, Age of the aircraft = 68, Aircraft Type = Balloon.

• Boston, MA (BOS) – Norfolk, VA(ORF). July 19 2005, Comair Airlines Flight Number =495, Distance/Airtime= 468 miles/minute (36 times faster than speed of sound!)
Big Data brings computational challenges

- A simple linear regression model takes more than 10 days on a desktop PC
- Can be speeded up by using a parallel computing cluster
- Data may make exact analysis infeasible
  - Log-Laplace vs Exponential distributions to model block time
- Need to develop approximations that can be efficient on large data sets
Big data can be used for predictive as well as prescriptive research

• Predictive Models - Forecasting on-time arrival probability for flights

• Prescriptive Models – Building robust flight schedules that minimizes flight delays
US Coast Guard Supply Chain

- 26 Airstations operating 4 aircraft types
- One stop shop for all aviation logistics support
  - Depot level maintenance
  - Supply engineering
  - Spare parts inventory management
  - Component repair
  - Information services
- 60,000 individual parts,
  - Inventory value over $937 million
Un-Scheduled Failures

Air Station Repair

ACMS Data

AMMIS Data

Scheduled Maintenance

Maintenance Shop

Warehouse
- Air Station
- ARSC

Item Managers
Procurement Specialists

Component Repair (Internal)

Vendors (OGA & Commercial)

failed parts

good parts
Features of Available Data

- Each part is tracked through a serial number, making each copy of a part uniquely identifiable and traceable.
- Part age (in flight hours and number of takeoffs and landings) available for every part, significant predictor of part failures.
- New sensor technologies could make this and more information available on a real-time basis, making dynamic real-time decision making possible.
Research Questions?

• What is the value of merging information/data from different sources/organizational boundaries?

• Can this data be used for making the supply chain efficient?
  - Using part age information can significantly improve part availability and simultaneously reduce inventory

• Can this data be used for dynamic decision making?
  - Real-time information can be used to plan inventory stocking at warehouses, repair planning at repair shops, and sourcing parts from vendors
Enhancing customer experience at retail stores using video feed data
Challenges posed by unstructured data?

• How can unstructured data such as video feeds be utilized for enhancing customer experience?

• How can information on customer pathways be utilized?

• Can this data be used for dynamic decision making (e.g., customized pricing?)
Big Data can be dangerous!

http://www.youtube.com/watch?v=HNaCf_Fnp6E