Kalkidan Ashenafi

Use Cases and explanation of the Big Data Characteristics involved

1. Category: Social Media

An example Use Case for Social Media is *Truthy Twitter Data Analysis* by Filippo Menczer, Alessandro Flammini and Emilio Ferrara from Indiana University. The goal of this Use Case is “Understanding how communication spreads on socio‐technical networks. Detecting potentially harmful information spread at the early stage (e.g., deceiving messages, orchestrated campaigns, untrustworthy information, etc.”[1]

A templated compilation of Use Cases from NIST Big Data Interoperability Framework: Volume 3, Big Data characteristics of this Use Case are the following:

*Volume*: “approximately 30TB/Year compressed data.” [1]

*Velocity*: “Near real time data storage, querying and analysis.” [1]

*Variety*: “Data schema provided by social Media data source. Currently using Twitter only with a plan to expand incorporating Google+, Facebook.” [1]

*Variability*: “Continuous real-time data stream incoming from each source.” [1]

1. Category: Medicine and Healthcare

An example use case for Medicine and Health care is *Electronic Medical Record (EMR)* Data by Shaun Grannis from Indiana University. The goals of this Use Case is to “use advanced methods for normalizing patient, provider, facility and clinical concept identification within and among separate health care organizations to enhance models for defining and extracting clinical phenotypes from non‐standard discrete and free‐text clinical data using feature selection, information retrieval and machine learning decision‐models. Leverage clinical phenotype data to support cohort selection, clinical outcomes research, and clinical decision support” [2]

A templated compilation of Use Cases from NIST Big Data Interoperability Framework: Volume 3, Big Data characteristics of this Use Case are the following:

*Volume*: “More than 12 million patients, more than 4 billion discrete clinical observations. Over 20 TB raw data.” [2]

*Velocity*: “Between 500,000 and 1.5 million new real‐time clinical transactions added per day.” [2]

*Variety*: “The integration of a broad variety of clinical datasets from multiple sources.” [2]

*Variability*: “Data from clinical systems evolve over time because the clinical and biological concept space is constantly evolving: new scientific discoveries lead to new disease entities, new diagnostic modalities, and new disease management approaches.” [2]

1. Category Scientific Research

An example of a Use Case for Scientific Research is *Radar Data Analysis for The Center for Remote Sensing of Ice Sheets* by Geoffrey Fox from Indiana University. The goal of this Use Case is to “determine the depths of glaciers and snow layers to be fed into higher level scientific analyses.” [3]

A templated compilation of Use Cases from NIST Big Data Interoperability Framework: Volume 3, Big Data characteristics of this Use Case are the following:

*Volume*: “approximately 0.5 Petabytes per year of raw data.” [3]

*Velocity*: “All data gathered in real time but analyzed incrementally and stored with a GIS interface.” [3]

*Variety*: “Lots of different datasets – each needing custom signal processing but all similar in structure. This data needs to be used with wide variety of other polar data.” [3]

*Variability*: “Data accumulated in approximately 100 TB chunks for each expedition.” [3]

1. Category Sales and Marketing

An example of a Use Case for Sales and Marketing is *Netflix Movie Service by Geoffrey* Fox from Indiana University. The goal of this Use case is to “allow streaming of user selected movies to satisfy multiple objectives (for different stakeholders) ‐‐ especially retaining subscribers. Find best possible ordering of a set of videos for a user (household) within a given context in real time; maximize movie consumption.” [4]

A templated compilation of Use Cases from NIST Big Data Interoperability Framework: Volume 3, Big Data characteristics of this Use Case are the following:

Volume: “Summer 2012. 25 million subscribers; 4 million ratings per day; 3 million searches per day; 1 billion hours streamed in June 2012. Cloud storage 2 petabytes (June 2013)” [4]

Velocity: “Media (video and properties) and Rankings continually updated” [4]

Variety: “Data varies from digital media to user rankings, user profiles and media properties for content‐based recommendations.” [4]

Variability: “Very competitive business. Need to aware of other companies and trends in both content (which Movies are hot) and technology. Need to investigate new business initiatives such as Netflix sponsored content.” [4]

1. Category Politics

A good example of use case in this category is Statistical *Survey Response Improvement by Cavan Capps* from U.S. Census Bureau. The goal of this Use Case is to “use advanced methods, that are open and scientifically objective, the statistical agencies endeavor to improve the quality, the specificity and the timeliness of statistics provided while reducing operational costs and maintaining the confidentiality of those measured.” [5]

A templated compilation of Use Cases from NIST Big Data Interoperability Framework: Volume 3, Big Data characteristics of this Use Case are the following:

*Volume*: “For this particular class of operational problem approximately one petabyte.” [5]

*Velocity*: “Varies, paradata from field data streamed continuously, during the decennial census approximately 150 million records transmitted.” [5]

*Variety*: “Data is typically defined strings and numerical fields. Data can be from multiple datasets mashed together for analytical use.” [5]

*Variability*: “Varies depending on surveys in the field at a given time. High rate of velocity during decennial census.” [5]

Citations:

1. *U.S. Department of Commerce. National Institute of Standards and Technology. “*NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements*.” NIST Special Publication 1500-3 (September 2015); Available from* [*http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-3.pdf*](http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-3.pdf) *pg.A-73*
2. *U.S. Department of Commerce. National Institute of Standards and Technology. “*NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements*.” NIST Special Publication 1500-3 (September 2015); Available from* [*http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-3.pdf*](http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-3.pdf) *pg.A-46*
3. *U.S. Department of Commerce. National Institute of Standards and Technology. “*NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements*.” NIST Special Publication 1500-3 (September 2015); Available from* [*http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-3.pdf*](http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-3.pdf) *pg.A-115*
4. *U.S. Department of Commerce. National Institute of Standards and Technology. “*NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements*.” NIST Special Publication 1500-3 (September 2015); Available from* [*http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-3.pdf*](http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-3.pdf) *pg.A-24*
5. *U.S. Department of Commerce. National Institute of Standards and Technology. “*NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements*.” NIST Special Publication 1500-3 (September 2015); Available from* [*http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-3.pdf*](http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-3.pdf) *pg.A-9*