Introduction to Big Data: Scalable Representation and Analytics for Data Science

* Steve Kaisler  <Skaisler1@comcast.net>
Frank Armour <farmour@american.edu>
Albert Espinosa alberto@american.edu

Big data is an emerging phenomenon characterized by the three Vs: volume, velocity, and variety. The volume of data has increased from terabytes to petabytes and is encroaching on exabytes. Some pundits are suggesting that zettabytes ($10^{21}$) are reachable within the next several years. Velocity is concerned with not only how fast we accumulate data, but also how fast some of the data that we already have is changing. Some systems accumulate data at the rate of multiple petabytes per year; some systems have stored data that changes at the rate of terabytes per year. Changing data usually lags accumulating data by several orders of magnitude. Data accumulating at a multiple petabyte rate requires terabits to petabits of transport capacity. Finally, the variety and modality of data is continually evolving; it may be both structured and unstructured.

This minitrack addresses several issues associated with the representation and use of Big Data for deriving new information.

An Agent Model for Incremental Rough Set-based Rule Induction: A Big Data Analysis in Sales Promotion
Yu-Neng Fan and Ching-Chin Chern

Fan and Chern propose a new mechanism for inducing rules that uses five different agents to develop the rules, manage the rules set, and evaluate rules within the rule set. New rules are extracted by modeling the reducts when the database is updated. A case study based on a Home Shopping company is used to address the database issues and demonstrate efficiency through incremental re-computation of the dataset when the database is updated.

Kaisler, Armour, Espinosa, and Money introduce a set of issues that will frame the analysis of big data. Their goal is to motivate the discussion of the representation, analysis, navigation and visualization of Big Data and to address scaling performance of algorithms supporting these operations. The authors are beginning a collaborative project to examine advanced analytics in business environments beyond the traditional business intelligence analytics.

Enterprise BigGraph
Aisha Naseer and Loredana Laera

Naseer and Laera describe BigGraph – a method for organizing and representing information of many types using Linked Data principles. Using a graph-based approach to manage big data, a variety of navigation, discovery, and visualization techniques become immediately available for use. The authors discuss how BigGraph supports dynamic updating of the information repository and address its use in processing the growing volumes of enterprise “social” data.

Big Data: Issues and Challenges Moving Forward
Stephen Kaisler, Frank Armour, Alberto Espinosa, and William Money