Research Roadmap: Big Data in Healthcare

Joseph M. Woodside[1]
[1]Department of Decision and Information Sciences, Stetson University, DeLand, Florida, US

Abstract - *Healthcare generates 30 percent of the world's data with a value of $300 billion in the next decade. Despite this, many healthcare providers have not developed a strategy for handling this data, and realizing the full set of opportunities to reduce costs and improve quality of care delivery. This Research Roadmap is set in major sections of structured and unstructured data and incorporates analytical decision making to improve the healthcare industry through Big Data.*

Keywords: Big Data, Healthcare, Structured Data, Unstructured Data

1 Big Data in Healthcare

Healthcare generates 30 percent of the world's data according to estimates, with 65 percent of survey respondents indicating their data storage will grow at a rate between 25-50 percent per year, driven by imaging files, electronic health records, personal health records, and scanned documents. Big Data in healthcare reflects the volume, growth, and types of data, as well as the tremendous opportunities available to unlock the potential value. In one report Big Data in healthcare was expected to be valued at $300 billion in the next decade, with annual growth between 1.2 – 2.4 Exabytes per year. Despite this growth, full integration and coordination has yet to occur and few healthcare providers have developed a formal strategy for handling the increasing amounts of data [1,2,3].

Healthcare organizations are under both industry and government pressure to reduce costs and improve quality of care delivery. With ever increasing amounts of data healthcare organizations are identifying the importance of business intelligence and analytics for decision making. Gartner recognized this as one of the fastest growing areas despite minimal economic growth, as organizations seek to compete and differentiate themselves through data based decisions [4,5]. The Big Data in Healthcare knowledge and research roadmap represents an opportunity to take advantage of the growing healthcare field and big data sources. The research stream seeks to develop innovate research, theory-building, publications, and grant funding.

This research output can be utilized to provide the management methods to improve the healthcare industry through Big Data decision making aimed at reducing costs and improving quality through use of information systems and technology. The research stream incorporates business intelligence, data analysis, and data mining to develop fact based decision making, through analysis and major research sections of structured and unstructured data.

2 Structured Data

Structured healthcare data is typically available in real-time, and includes established systems such as medical claims, pharmacy, lab, and financial data. Data is stored in data warehouses, and typically range from terabytes to petabytes in size. Decision making may take the form of reporting, alerts, performance management, and financials [6,7]. Future research streams include:

- Hospital Readmission – identify the factors leading to hospital readmission, predictive modeling of readmission, and successful methods to reduce readmission rates.
- Machine Learning Outcomes – development of machine learning algorithms to learn and predict outcomes automatically, significantly reducing costs and increasing quality.
- Geographic Information Systems – identify the spatial and temporal trends through cluster analysis to predict healthcare utilization/costs. This research is able to be applied to new to population management and database applications.
- Pay for Performance – formulation of incentive programs that can be utilized for providers and healthcare entities as a cloud service.
- Fraud Detection – neural network and pattern detection of patients, providers, and healthcare entities for fraud as well as training or errors improvement.

3 Unstructured Data

Unstructured healthcare data is typically available in non real-time, and includes data sources such as electronic medical records, radiology images, and clinical notes. Data is stored outside of the data warehouse in file systems, and commonly larger in size than structured systems ranging from terabytes to petabytes in size. Decision making may take the

form of data mining, text mining, and other batch analysis methods [6,7]. Future research streams include:

- Risk Management – through classification validate existing and identify potential future care management needs for patient conditions based on imaging, notes, and historical medical records.
- Wellness – application architecture for a patient centric framework, to improve virtual health management and monitoring.
- Social Media – development of social media tools and evaluation of current tools for improving member's general health and condition specific areas.
- Workflow – monitoring of clinical and non-clinical activities, for improving process intelligence through best practices, and efficiency of care.
- Communities of Care – analysis of web logs and outcomes information for measuring self-management of conditions by the patient and community of patients.

4 References

[1] Healthcare IT News survey on healthcare data growth: Miliard, M. "'Deluge of data' has hospitals swimming upstream" H.I. News, 2012.

[2] Healthcare Informatics survey results on data and migration: DeGaspari, J. "Healthcare Data Experiencing Explosive Growth" Healthcare-Informatics, 2010.

[3] Description of healthcare Big Data estimates: Hughes, G. "How big is 'Big Data ' in healthcare," SAS, 2011.

[4] Description of Big Data: "Big Data in Healthcare - where does one start?" Schick, C. IBM, 2012.

[5] Description of Growth: "Gartner Forecasts Global Business Intelligence Market to Grow 9.7 Percent in 2011" Gartner, 2011.

[6] Description of Data: "Healthcare Data's Perfect Storm." Burns, W. Hitachi, 2013.

[7] Description of Data: "Big Data in Healthcare - where does one start?" Schick, C. IBM, 2012.