Can We Apply Principles From Social Networking To Healthcare Informatics For Intelligent Data Analytics?

Dr Shoumen Datta, Research Scientist, Engineering Systems Division, Department of Civil and Environmental Engineering; Co-Founder and Research Director, Forum for Supply Chain Innovation, School of Engineering, Massachusetts Institute of Technology, Room 1-179, 77 Massachusetts Avenue, Cambridge, MA 02139, USA (shoumen@mit.edu) and Co-founder, Centre for Innovation in Distributed Systems at IT Tralee (www.cids.ie)

and

Mr Andrew Shields, Pre-doctoral Research Associate, Centre for Innovation in Distributed Systems at IT Tralee

ABSTRACT

Extracting the principles associated with complexity theory and swarm intelligence has offered practical solutions for routing and scheduling. Reality mining and its link with social networking relationships may yield pragmatic ideas applicable to many fields including business services and healthcare analytics. In healthcare, the focus is on the patient and physiological systems generate patient data. Since human physiology is highly integrated and always strives to maintain biologically relevant equilibrium, it follows, naturally, that physiological data and variables are likely to be co-integrated. Because physiological systems always strive to maintain homeostasis, it follows, that the focus of physiology is to attain equilibrium where various interacting components possess information about the functional status of other components. The physiological state, may, therefore, be viewed as a healthcare service system which is amenable to data analytics using the principles of Nash Equilibrium. In sharp contrast, the healthcare industry, like most businesses, suffers from chronic information asymmetry of data and information about its supply chain network. Information asymmetry in the complex and inter-related healthcare network may offer grounds to explore potential benefits to the industry if information asymmetry could be reduced through appropriate acquisition of data including real-time data. Availability of high volume data may improve forecasting in healthcare sectors related both to service and the business. But, while the innovation for service may draw inspiration from information symmetry, the techniques to reduce volatility in the healthcare business must address information asymmetry. In both cases, potential use of advanced econometric tools (generalized autoregressive conditional heteroskedascity or GARCH) may be applicable. The network in healthcare relates both to physiological circuitry and the business relationships. Both networks may be evaluated in the context of the structures of social networking to elucidate if such principles may add value to healthcare analytics.

KEYWORDS

Intelligent Decision Systems, Systems Engineering, Remote Monitoring, Glucose Sensors, ICT, Semantics, Informatics, Analytics, Electronic Medical Records, EMRS, Interoperability, Mash-Up, SOA, Agents, AI, Biomarkers, Cardiovascular Diseases, Early Detection, Preventative Medicine, Social Networking, Reality Mining, Data Mining, Linguistics, RFID, Supply Chain Management, WSN, Nash Equilibrium, Information Asymmetry, Generalized Autoregressive Conditional Heteroskedasticity, VAR, ARMA, Nano-sensors

INTRODUCTION

This work-in-progress is based on and is an extension of a previous work (Datta 2008). One issue of interest is the acquisition of medical data to improve healthcare. Sensor technology may be one solution to capture this data without significant intrusion. Given the potentially vast amount of data stream from sensors, an increasingly important part of wireless sensor networks (WSN) will include applications and algorithms to fuse, interpret, augment and present information. Data mining may be one tool of choice.

Data mining as applied to "business intelligence" applications may play a role but may be inadequate to address the service part of healthcare because "service" relates to an individual or patient-centric data. On the other hand, the healthcare industry may have distinctive needs pertaining to business and operational efficiencies. Data, information and knowledge hold the potential to improve both healthcare service and the healthcare industry, but the tools and applications are expected to differ, in their pursuit of these varied goals and functions. This is where the prevalent view of data mining may diverge.

EVOLUTION OF DATA MINING

The tools of data mining were enriched when principles associated with complexity theory and swarm intelligence (Bonabeau, Dorigo and Theraulaz 1999) emerged to offer business solutions (Icosystem 2008) for a wide variety of routing and scheduling needs. A similar wave (Finin *et al* 2008) is imminent under the generic banner of data mining tools that may stem from reality mining (Eagle and Pentland 2005) and its link with social networking relationships.

From one perspective, reality mining is viewed as data mining of sensor streams. The data provides an insightful infrastructure between detection and action allowing organizations to use data to add value (Boone 2004). For example, attaching sensors to patients recovering from acute myocardial infarction may allow medical practitioners to monitor heart rate and rhythm. Combining these sensor readings with models of heart rhythm, projections of rehabilitation may be generated. It may allow the medical practitioner to prescribe appropriate medical treatments and enable hospital managers or social services to make resource allocations for optimal patient care with maximal cost efficiencies.

At the point of contact where the sensors are deployed, data steams may be subjected to in-network processing or gross level filtering may be necessary to manage the data load transmitted to the data mining processor. Filtering, averaging, removing outliers and other techniques eliminates some data but could create a composite view of the physical phenomena by blending the readings from sensor clusters. Transmission of data may incur a transaction cost for using wireless networks (GSM/GPRS/4G). Edge functions with embedded filtering logic may process the raw sensor data before transmission to analytic engines, if transaction cost is a function of data load. The criticism of this reductionist approach is the perceived inadequacy of the filtering logic at the edge. Can we rely on intelligent learning systems to understand and distinguish which outlier data may be an aberration or is the first clue to dysfunction?

Decision makers are less interested in sensor data and more interested in trying to answer questions either in the context of healthcare service or industry. In this respect, data mining seeks to create useful information from structured data while reality mining seeks to create usable insight from data. The healthcare industry is prone to information asymmetry (Akerlof 1970) and may benefit if information asymmetry could be reduced through appropriate acquisition of data and making sense of the data.

The strength of this emerging vision of healthcare may be determined by exploring :

1. Quality of care improvements
2. Impact on human resources in terms of time savings for medical professional
3. Reduction in cost and potential for savings
4. Length of time required for return on investment
5. Profitability of businesses (SME) and growth of high potential start-ups
6. Economic benefits for the nation’s healthcare system
7. Reproducibility, portability and sustainability of the services model as a global template
8. Business opportunities to implement similar services in other communities or nations

TECHNOLOGY IMPACT ON HEALTHCARE SPENDING

As a population we are healthier than ever but more worried about our health. Longevity in the West continues to rise. A typical British woman can now expect to become an octogenarian, which is double the life expectancy when Queen Victoria ascended the throne in 1837. In 1950, UK reported 26,000 infant deaths but within half a century that number decreased by 80%. Between 1971 and 1991 stroke related deaths dropped by 40% and coronary heart disease fatalities were down 19% (Porter 1997). However, there is a pervasive sense that our well-being is imperiled by 'threats' from the air we breathe to the food in our stores. The global and local public chooses to be more troubled about current air pollution than during the urban smog of the 1950's in the UK, when tens of thousands died of bronchitis.

Improvements in healthcare and life expectancy are also due to technology and at least half of the cost is a result of increased use of technology (Cutler and McClellan 2001, Fuchs and Garber 2003). Medical technology advances applied in one year can lead to higher or lower spending in succeeding years. Costs may increase if people live longer to receive more care or may be reduced if people live healthier lives and as a result require less care due to investment in remote monitoring or preventative medicine. These long term impacts of new technology need to be evaluated in addition to the initial costs. For example, as a consequence of a myocardial infarction, many people will receive bypass surgery or angioplasty (collectively termed revascularization) to restore blood flow to the heart. Each is expensive and the value is debated (Skinner, Staiger and Fisher 2007). The central empirical issue in evaluating the impact of any technology is the determination of who receives the care. Weaker patients or those with poor health may not be strong enough to withstand revascularization. The relatively healthy patients may not need it. Thus, the set of people receiving revascularization is not a random set of people with a myocardial infarction. Using knowledge generated from sensor data using data mining and reality mining algorithms may provide useful metrics in determining patient selection for treatments and if used in conjunction with advanced econometric techniques to improve the accuracy of analytics.

In the early 1980's hospital admission for *Diabetes mellitus* declined due to the development of effective methods to monitor blood sugar at home (Sennett 1990). The paradigm shift in moving health care outside of the hospital may be further accelerated if, for example, pathology information systems can provide key databases for health services and new informatics-based approaches can educate the public what to monitor using over-the-counter tools, such as glucose and cholesterol measuring kits.

ACQUISITION AND KNOWLEDGE DISCOVERY OF MEDICAL DATA TO IMPROVE HEALTHCARE

Currently, it is possible to determine, among others, heart rate and blood glucose, with small, non-invasive wireless sensors. Devices such as the Ring Sensor can measure oxygen saturation and transmit the real-time data through a wireless up-link (Sokwoo *et al* 1998). Companies such as Nonin (NONIN 2008) and Numed (Numed 208) have developed wireless vital sign sensors based on bluetooth while Radianse (Radianse 2008) has developed RF-based location-tracking. EU's MobiHealth is developing a 3G-enabled Body-Area Network (Fraunhofer-Gesellschaft 2008).

Innovation in data collection calls for innovation in analytics (Kamel, Hetherington and Wheeler 2007; Kamel et al 2007; Konstantas 2007) in addition to exploring non-traditional tools. For example, clues to diagnosing depression, Depressed people may speak slowly, a change that speech analysis software on a phone might analyse (Greene 2008).

Mobile phones may be useful in gathering health related information, as well as providing information focused on the individual mobile phone user or large groups and their interactions. For example, researches can use data on a sample population over a given period and then running simulations to determine how a disease may spread or build models to predict epidemics or pandemics, for example, SARS and H5N1 (Eagle and Pentland 2004).

TECHNOLOGY IN PREVENTATIVE MEDICINE

Systems physiology of the 21st century is set to become highly quantitative and one of the most computer-intensive disciplines (Noble 2002). Since human physiology is *highly* integrated, it may follow, naturally, that the physiological variables (for example, blood pressure, heart rate, pulse rate) are likely to be co-integrated (Granger and Swanson 1996). In other words, because physiological systems strive to maintain homeostasis, it follows, that the goal of physiology is to attain equilibrium. When one variable is affected, for example, pulse rate, its effect is "integrated" or reflected or related to another linked variable, for example, blood pressure. Physiological mechanisms *in vivo* will attempt to rectify this situation and may try to restore the blood pressure of the individual to 120/80 mm Hg (normal reading).

Due to the innate physiological drive to restore equilibrium, data analysis in healthcare service may benefit from a potential exploration and application of the principles of Nash Equilibrium (Nash 1950) to predict from a set of patient data what other parameters (co-integrated) are likely to change or may be influenced by the change documented (data at hand). It may provide clues to improve diagnosis. Many other tools exist for processing data, such as artificial neural networks (ANN) and mining. Convergence of diverse multi-disciplinary research in analytics may be of immense benefit for medical informatics.

HEALTHCARE VALUE NETWORK

Healthcare value chain consist of three major players: producers (product manufacturers), purchasers (group purchasing organizations and wholesalers/distributors) and healthcare provider (hospital systems and integrated delivery networks or IDN).

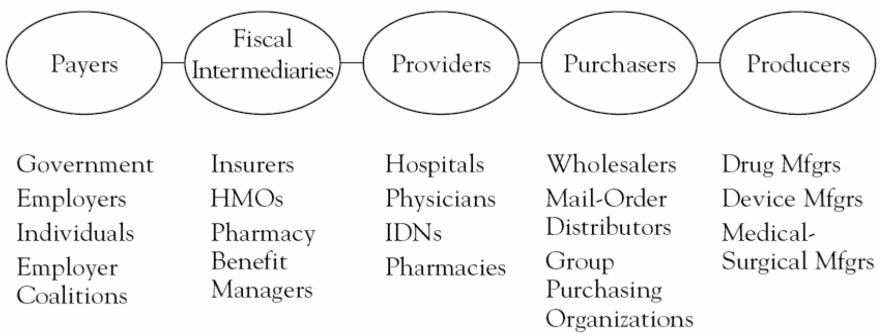

Figure 1: Healthcare Value Chain [1]

Healthcare supply chains need to balance the push-pull model and offer greater visibility to channel participants to successfully lower costs and increase service levels. Process discrepancies between organizations increases risk when ICT systems lack in interoperability. The use and analysis of object data in a model that captures the end-to-end business network (as well as links to other factors that may impact the function of a specific node) may help to reduce risk. It is in this context that a combinatorial use of MGARCH and VAR techniques may offer value hitherto unimaginable.

[1] http://media.wiley.com/product_data/excerpt/17/07879602/0787960217.pdf

CONCLUSION

Ubiquitous sensor networks have the potential to generate vast streams of data. Making sense of the data represents a considerable challenge. In one approach, some may react only to exceptions. A low-level data mining layer may abstract data into "cubes" of information suitable for reality mining agents to analyse and catalyse decision support. The future of an integrated reality-online analytical framework may concomitantly support remote experts, real-time teams and insight-based or intelligence-trained feedback for point-of-contact (POC) decision makers to improve patient-specific services.

The ill-formulated amorphous challenge presented in this working paper relies on (and expects to trigger further development by) individuals who have a breadth of interest and are gifted with the ability to construct fruitful analogies between fields while extracting pragmatic applications to balance cost versus ethics while serving the vastly divergent domains spanned by healthcare: healthcare industry economics conscious of cost and healthcare as a patient-centric service defined by ethical globalization.

REFERENCES

Akerlof, G. (1970) The Market For Lemons: Quality Uncertainty And The Market Mechanism. *Quarterly Journal of Economics* **84** 488-500

Bonabeau, E., Dorigo, M. and Theraulaz, G. (1999) *Swarm Intelligence: From Natural to Artificial Systems.* Oxford University Press, NY.

Boone, G. (2004) *Reality Mining: Browsing Reality with Sensor Networks.* www.sensorsmag.com/articles/0904/14/main.shtml

Byrnes, J. (2004) *Fixing the Healthcare Supply Chain.* http://hbswk.hbs.edu/archive/4036.html

Cutler, D.M. and McClellan, M. (2001) Is Technological Change In Medicine Worth It? *Health Affairs*

Datta, S. (2008) Auto ID Paradigm Shifts from Internet of Things to Unique Identification of Individual Decisions in System of Systems. *MIT Engineering Systems Division Working Paper* http://esd.mit.edu/WPS/2008/esd-wp-2008-09.pdf

Datta, S., Granger, C. W. J., Barari, M. and Gibbs, T. (2007) Management of Supply Chain: An Alternative Modeling Technique for Forecasting. *Journal of the Operational Research Society* **58** 1459-1469

Datta, S. (2008) Will Nano-Butlers Work for Micro-Payments? Innovation in Business Services Model may Reduce Cost of Delivering Global Healthcare Services. http://hdl.handle.net/1721.1/41901

Eagle, N. and Pentland, A. (2005) Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing.* http://reality.media.mit.edu/pdfs/realitymining.pdf

Eagle, N. and Pentland, A. (2004) *Social Serendipity: Proximity Sensing and Cueing.* Technical Note 580. MIT Media Lab

Emery, F E. (1976) *Systems Thinking.* Penguin, UK.

Finin, T., Joshi, A., Kolari, P., Java, A., Kale, A. And Karandikar, A. (2008) The Information Ecology of Social media and Online Communities. *Artificial Intelligence Magazine* **28** 1-12 http://ebiquity.umbc.edu/_file_directory_/papers/376.pdf

Fraunhofer-Gesellschaft. (2008) *Body Area Network.* www.ban.fraunhofer.de/index_e.html

Fuchs, V. and Garber, A.M. (2003) Medical Innovation: Promises & Pitfalls. *Brooking Review*

Fuchs, V. and Garber, A.M. (2003). *Medical Innovation: Promises & Pitfalls.* www.brookings.edu/articles/2003/winter_technology_fuchs.aspx

Granger, C. W. J., and Swanson, N.R. (1996) Further developments in the study of cointegrated variables. *Oxford B Economics and Statistics* **58** 374-386

Greene, K. (2008) Reality Mining. *Technology Review*

Icosystem www.icosystem.com/about_management.htm

Joshi, A., Finin, T., Java, A., Kale, A. and Kolari, P. (2008) Web (2.0) Mining: Analyzing Social Media. http://ebiquity.umbc.edu/_file_directory_/papers/379.pdf

Kamel Boulos, M.N. *et al.* (2007) CAALYX: a new generation of location-based services in healthcare *International Journal of Health Geographics* 6-9

Kamel, Boulos M.N., Hetherington, L. and Wheeler, S. (2007) Second Life: an overview of the potential of 3-D virtual worlds in medical and health education. *Health Information and Libraries Journal* 233-245.

Konstantas, D. (2007) An Overview of Wearable and Implantable Medical Sensors. *Methods of Information in Medicine* 66-69

Nash, J. F. (1950) Equilibrium Points in N-Person Games. *Proceedings of the Nationall Academy of Sciences (USA)* http://www.pnas.org/misc/Nash.pdf

Noble, D. (2002) Modeling the Heart-- from Genes to Cells to the Whole Organ. *AAAS* www.sciencemag.org/cgi/content/abstract/295/5560/1678

NONIN www.nonin.com

Numed www.numed.co.uk

Porter, R. (1997) *The Greatest Benefit to Mankind: A Medical History of Humanity.* www.nytimes.com/books/first/p/porter-benefit.html

Radianse www.radianse.com

Schletzbaum, A. (2007) King County Healthcare System Emergency Response Planning http://nvac.pnl.gov/meeting_spring07/presentations/Schletzbaum_Healthcare_Coalition.ppt

Sennett, C. (1990) *Why hospital admissions have declined : technologic change and changing patterns of care for cataract and diabetes mellitus.* PhD Thesis, MIT

Skinner, J., Staiger, D. and Fisher, E. (2007) Is Technological Change in Medicine Always Worth It? The Case of Acute Myocardial Infarction." *Health Affairs* February 2007.

Sokwoo, R., Boo-Ho, Y., Kuowei, C. and Asada, H.H. (1998) *The Ring Sensor: a New Ambulatory Wearable Sensor for Twenty-Four Hour Patient Monitoring.* d'Arbeloff Laboratory and Department of Mechanical Engineering. Ph.D. Thesis. Massachusetts Institute of Technology Cambridge, Massachusetts.