Some Major Biomedical Informatics Challenges

Kenneth Baclawski, Mieczylaw Kokar and Christopher Matheus

1. Medical Records.

Significant efforts are new being undertaken to transform the US health care delivery system. "The Institute of Medicine has highlighted how between 44,000 and 98,000 Americans die each year from medical errors. This heightened the issue of patient safety and quality for the public and decision-makers. Early in 2004, President Bush called for widespread adoption of interoperable electronic health records within the next 10 years, believing that electronic health records will reduce medical errors, cut healthcare costs through increased efficiency, and ultimately result in improved patient care... Over 80% of health care providers in the US in 2005 did not have electronic health record systems. Of the systems that do exist few are interoperable."[1]

Even if one presumes that one can solve the massive problem of health care providers developing electronic health record systems, little will be accomplished if the resulting systems are not interoperable. Indeed, for all its disadvantages, a paper-based medical record system allows for interoperation, which cannot be said for electronic record systems in general. Furthermore, experience has shown that simply automating paper-based processes has relatively little impact on productivity. To achieve significant gains in efficiency and improved patient care, the overall process of medical care delivery must be based on web-centric information technologies.

Unfortunately, the Health IT problem is currently defined as the problem of providing electronic health *records*. This effectively mandates a closed-world, relational solution to the problem of health care information, which is closely tied with existing workflows and processes. Simply converting such records to electronic form is likely to have the unintended consequences of imposing inappropriate workflows and making it even more difficult to improve processes.

The research issue is to develop standard medical event ontologies that support interoperability, are independent of workflows and processes, and are compatible with existing processes. This last point is important as there must be a path whereby the existing infrastructure can evolve to the new standard. While this challenge is difficult, it is certainly possible if the existing standards bodies are willing to accept the need for this approach.

2. Uncertainty.

The World Wide Web has enabled an unprecedented level of interoperability and connectedness of information and systems. The Semantic Web is an extension of the World Wide Web in which information is given a well-defined meaning, so that computers and people may more easily work in cooperation. This is done by introducing a formal logical layer to the web in which one can perform rigorous logical inference. However, the Semantic Web does not include a mechanism for dealing with the uncertainties that are common in biology and medicine. As a result, conclusions of scientific research papers are informally stated using natural language, making interoperation difficult. As a result, meta-analysis techniques that combine results from multiple sources require a great deal of effort and skill, yet are still error-prone and can omit relevant information.

Unfortunately, while there has been some research on the topic of representing uncertainty using the semantic web,[2] it has not yet progressed to the point where one can routinely represent the stochastic models required for scientific research results. For example, the only standard for Bayesian networks is a defacto standard developed by Microsoft,[3] which can only represent finite discrete random variables. An extension to the Semantic Web, called the Bayesian Web,[4] would integrate statistical inference with logical inference, but it is currently only a proposal.

The research and development challenge is to develop a full-featured stochastic reasoning infrastructure, comparable to the logical reasoning infrastructure of the semantic web. This infrastructure should support features such as the following:

- A common interchange format for stochastic models and statistical test results
- Allow specification of the context of a model or result
- Open hierarchy of probability distribution types
- Component based construction of stochastic models
- Stochastic inference engines

3. Data Fusion and Situation Awareness.

Meta-analysis is the integration of data from disparate sources. Such sources are generally subject to uncertainty due to the lack of full knowledge about the phenomena being observed as well as the limitations of the measuring devices being used. One can reduce these uncertainties by making a series of independent observations. Meta-analysis is the process of combining the evidence afforded by the observations.

Meta-analysis is only one level of the process in which observations are ultimately used to make scientific and clinical decisions. The entire process is known as data fusion. The levels of data fusion have been formalized in the JDL model,[5] which is primarily used in military and emergency response contexts. While meta-analysis (i.e., level 1 data fusion) is well-developed in biomedicine, the higher levels of data fusion are not integrated in biomedical contexts. Level 2 data fusion is called situation awareness and roughly corresponds to medical diagnosis.

The research challenge is to develop data fusion tools for biomedicine that support researchers and clinicians in the task of situation awareness as well as higher levels of data fusion. Some examples of applications of such a tool include:

- Tracking epidemics
- Monitoring the patient during surgery
- Meta-analysis services for researchers

- Assessing the health of populations by region or recognized group

[1] Health IT in Government: Transforming health care and empowering citizens. Health IT in Government
[2] Workshop on Uncertainty Reasoning for the Semantic Web, Fourth International Semantic Web Conference, Galway, Ireland, 2005.
[3] XML Belief Network File Format, Microsoft Decision Theory and Adaptive Systems Group, April, 1999.
[4] K. Baclawski and T. Niu, Ontologies for Bioinformatics, MIT Press, October, 2005.
[5] A. Steinberg and C. Bowman, Revision to the JDL data fusion model, Handbook of Multisensor Data Fusion, page 2-1, CRC Press, 2001.