Modeling Medical Imaging and Molecular Biology Correlates from Literature

William Hsu, Ph.D., James Sayre, Ph.D., Alex A.T. Bui, Ph.D., and Ricky K. Taira, Ph.D.
Medical Imaging Informatics Group, Department of Radiological Sciences, University of California, Los Angeles, CA

Abstract

Advances in high-throughput sequencing and computational biology have provided a clearer understanding of linkages between genotypes and clinical phenotypes. Results are typically disseminated through publications, but it is often difficult for practitioners to use this information towards personalized medicine. We describe a framework for modeling findings using a dependency graph to represent variables and interactions abstracted from papers. A model-driven visualization assists users with querying the contents of multiple papers and identifying causal associations.

Introduction. The large number and diversity of topics in published literature makes it difficult for any one person to stay up-to-date on all relevant findings. For example, a search of "brain cancer" on ClinicalTrials.gov returns some 1,731 studies. The traditional way of presenting knowledge in published literature has many limitations: 1) information is described in natural language and is fraught with imprecision and ambiguity; 2) interpreting conclusions often requires understanding mathematical formalisms in which the reader may not be well-versed; and 3) it is not always apparent how to interpret results of a study in the context of a complex disease. We present work on a computational infrastructure that integrates results of randomized controlled trials (RCT) and observational studies published in literature. In particular, we are developing a dependency model that captures variables, interactions, and contextual information (e.g., hypothesis, statistical method, variables, interpretation) from papers and a visualization that facilitates tasks such as: 1) searching/filtering studies based on patient/treatment criteria; 2) comparing and contrasting findings across multiple studies; and 3) interpreting explanations generated by the model. Efforts such as NeuroScholar [1], Information Hyperlinked over Proteins (IHOP) [2], and Ontology for Clinical Research (OCR) [3] have resulted in methods for extracting, representing, and visualizing information in published literature. Our work complements these efforts by developing approaches to represent detailed information abstracted from multiple papers for a given disease. We focus on several key questions such as: how do we convey what is the major new contribution of a study; how do we facilitate comparison of similar RCT studies (i.e., identify contributions which support existing causal associations or new studies that refute prior explanatory hypotheses); and how to link fragmented evidence presented within each paper into a unified representation.

Methods and discussion. Towards these aims, we have developed a graph-based dependency model that integrates evidence from multiple papers. Each fragment consists of variables derived from clinical observables, imaging features, physiological processes (e.g., angiogenesis, cellular mitosis), and gene expression (e.g., EGFR overexpression) related by a specific type of interaction (e.g., causes, inhibits, decreases). We have generated a model from ten papers describing linkages among imaging findings (e.g., contrast enhancement, edema), physiological processes, genes, and interventions (e.g., rapamycin) in the domain of neuro-oncology. We incorporate knowledge of known molecular pathways from sources such as [4] to bridge gaps in understanding between fragments of evidence. Based on this model, we have created an interactive visualization that allows users to navigate and query the graph. Selecting an edge between two nodes brings up supporting data describing how the relationship is measured (e.g., statistical test, study population, raw data). Controls to filter the display by variable type, physiological process, relation to drug, or other attributes are provided. We are exploring how the model can also be used to generate explanations based on the encoded knowledge; a Bayesian belief network derived from the dependency graph and parameterized using results from meta-analysis of the papers is being explored for this purpose.

References

[1] Burns G, Feng D, Hovy E. Intelligent approaches to mining the primary research literature: techniques, systems, and examples. Computational Intelligence in Medical Informatics. 2008:17-50.

[2] Hoffmann R, Valencia A. Implementing the iHOP concept for navigation of biomedical literature. Bioinformatics. 2005;21(suppl 2):ii252-8.

[3] Sim I, Carini S, Tu S, Wynden R, Pollock BH, Mollah SA, et al. The Human Studies Database project: federating human studies design data using the Ontology of Clinical Research. AMIA Summits on Translational Science Proceedings. 2010:51-5.

[4] Hanahan D. The hallmarks of cancer. Cell. 2000;100(1):57-70.