To appear in AI Magazine, 2013.

The AAAI Fall Symposium on "Discovery Informatics: The Role of AI Research in Innovating Scientific Processes"

Yolanda Gil
Information Sciences Institute
University of Southern California
gil@isi.edu

Haym Hirsh
Department of Computer Science
Rutgers University
hirsh@cs.rutgers.edu

The AAAI symposium on Discovery Informatics was held in Arlington, VA on November 2-4, 2012. The goal of this symposium was to discuss the role of AI techniques in improving or innovating scientific discovery processes. This paper summarizes the discussions at the meeting, an overview is also available at the web site [DIS 2012]. Figure 1 shows an illustration of the most popular words in the symposium notes, highlighting data as well as the importance of people, models, and systems and a number of AI research areas of interest.

The increasing complexity of scientific processes offers new opportunities for intelligent systems, information sciences, and human-computer interaction. The symposium was kicked off with an overview of Discovery Informatics by the organizers, focusing on improvements and innovations across the spectrum of scientific activities that could have significant impact on the rate of scientific discoveries. The symposium included several invited presentations, two panels, and reviewed papers. Participants came from academia, industry, and several government agencies.

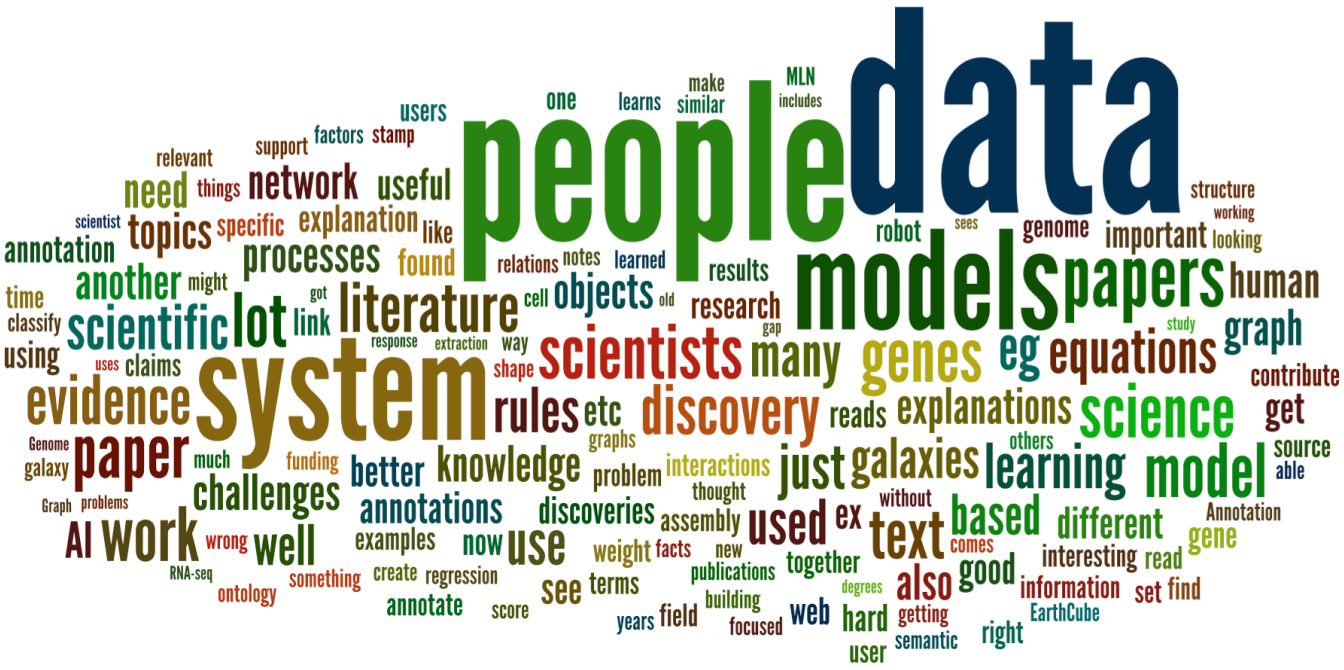

Figure 1. Most popular words in the Discovery Informatics symposium notes.

Science is a great challenge domain for AI. Scientists are overwhelmed by the growing availability of data, analytic methods, and publications. Many scientific questions require reaching outside of a given discipline, but it is often hard to identify relevant expertise. Intelligent systems that can assist or automate any aspects of scientific processes are of great value to scientists. Pat Langley of Carnegie Mellon University gave a historical perspective of AI research in computational discovery, and presented his most recent work on explanation and causal modeling. Larry Hunter of the University of Colorado proposed biomedicine as a grand challenge for AI, as it is an AI-complete problem with some simplifying features and great potential for profound impact in science and society. Vasant Honavar of Iowa State University argued that there are hard AI problems in all aspects of scientific discovery processes. Barbara Ransom of the National Science Foundation discussed opportunities within EarthCube, a new NSF initiative to transform scientific practice across all geosciences through new capabilities that vastly improve productivity and support knowledge management for cross-disciplinary earth, atmospheric, and ocean sciences.

Coping with the growing availability of data is an active area of research. Automated techniques are a way to approach this challenge. Hod Lipson of Cornell University described feature-free learning and symbolic regression techniques that enable the discovery of natural laws and invariants in a variety of large scientific datasets. A current challenge for these techniques is that when scientists are presented with interesting findings they want to see an explanation for the findings, which is hard to generate. Susan Epstein of City University of New York presented a novel clustering algorithm and its application to protein discovery. Taha Bahadori of the University of Southern California presented a new approach to extracting causal models from temporal data. Arman Masoumi from Ryerson University presented the use of automated planning to generate novel chemical compounds. An alternative to automated techniques is to recruit volunteers to process data. Chris Lintott of Oxford University talked about the expanding Zooniverse citizen science framework and the lessons learned about people's motivation and interests as an unprecedented amount of volunteers classified millions of galaxy images. Citizen scientists want to investigate things they notice in the data and make discoveries for themselves, so a great challenge is to make science data and analytic software accessible to them. A panel discussed additional challenges in "big data," including relational data, genomic data, and spatio-temporal data.

The growing size of the scientific literature is another major challenge. Jude Shavlik of the University of Wisconsin talked about machine reading, where humans provide initial rules that are refined through examples extracted from text. Evaluations in web documents and geoscience articles show that human rules increase performance, but the challenge is to allow non-AI experts to specify them. William Cohen of Carnegie Mellon University described several uses of text mining in science: recommending articles to scientists, extracting and integrating facts, and integrating extracted facts with hand-coded biomedical databases. He pointed out that specialized techniques are needed to handle biomedical text. Tim Clark of Harvard University described an ontology for annotating scientific papers using Semantic Web infrastructure, and the challenges in creating knowledge bases that allow scientists to explore related evidence from different publications. Gully Burns of the University of Southern California described an ontology-based approach for meta-analysis of the literature applied to HIV vaccines.

Intelligent support for data analytics is also an active area of research. Leonardo Salayandia and Nicholas del Rio of the University of Texas El Paso presented workflow management and provenance support tools, and discussed the challenges of assisting scientists to create visualizations. A panel discussed the opportunities in supporting individual scientists to manage metadata, publish data, and benefit from existing but difficult to use science infrastructure.

All the reviewed papers and abstracts from invited talks are available as a AAAI Technical report.

The symposium was held as part of the Fall Symposium Series of the Association for the Advancement of Artificial Intelligence (AAAI). The symposium was organized by Will Bridewell (Stanford University), Yolanda Gil (University of Southern California), Haym Hirsh (Rutgers University), Kerstin Kleese van Dam (Pacific Northwest National Laboratory), and Karsten Steinhaeuser (University of Minnesota).

References

[DIS 2012] Web site of the 2012 AAAI Fall Symposium on Discovery Informatics: http://www.discoveryinformaticsinitiative.org/dis2012

[Bridewell et al 2012] Will Bridewell, Yolanda Gil, Haym Hirsh, Kerstin Kleese van Dam, and Karsten Steinhaeuser (Eds). "Discovery Informatics: The Role of AI Research in Scientific Discovery." AAAI Fall Symposium Series Technical Report, 2012.