

Title: Clinical Natural Language Processing with Machine Learning

Summer 2019, DMIS Invited Lecture

Speaker: Timothy Miller (Harvard Medical School)

Date: 2019.08.29 (Thursday) 16:00 ~

Location: 우정정보관 B101호



Abstract: Unstructured text in electronic health records holds a wealth of information that can potentially improve clinical research and patient care. Reliably extracting this information into a usable form is the goal of clinical natural language processing (NLP). In this talk, Dr. Miller will give an overview of his research that is primarily focused on two difficult problems in clinical NLP: Learning representations of language that can apply to multiple tasks, and adapting machine learning models to work on data from multiple institutions and medical domains.

Speaker Bio: I am part of the Computational Health Informatics Program (CHIP) located in Boston Children's Hospital. I work in natural language processing (NLP) of clinical text, extracting information from medical records to facilitate clinical research and make the healthcare system more efficient.

I am trained as a computer scientist, with a BS from Marquette University and MS and PhD from the University of Minnesota, Twin Cities. I did my thesis on the linear time recognition of the syntax of disfluencies in spontaneous speech. In the general domain, natural language processing (NLP) is usually applied to standard corpora including financial newswire text and a few other canonical sources, to the extent that the whole field is probably overtrained on these few data sources. I am interested in doing clinical NLP research to broaden the usage and development of NLP models to new domains, and the domain of clinical research is especially exciting because of the direct impact it can have on people's lives.

I am interested in applying statistical models of human language to data in the electronic health record. Specifically, I am currently interested in tree kernels for support vector machines, constituency parsing and features derived therefrom, clinical domain adaptation, generative (Bayesian) models of text generation, and coreference resolution.