NLP for the Institute

Developing and Deploying an NLP Capability to Accelerate Cancer Research

Aaron Cohen

Oregon Health & Science University
Portland, Oregon, United States
cohenaa@ohsu.edu

Abstract— It has been well documented that a great deal of data useful for medical research is present in clinical narrative text.

There is perhaps less discussion about how often what was structured data at its origin has become inaccessible except in free text form. This problem is further compounded in tertiary care institutions, like the OHSU Knight Cancer Institute, where the entire history of a referred patient's condition may only be present in the electronic health record (EHR) as free text.

At the same time, future medical advances, such as in cancer research, will require much more complete patient data than has been previously available. Such advances include the discovery of new cures, expanding early detection, and realizing the promise of precision medicine. Phenotype description and outcome characterization are two areas in particular where text sources could greatly supplement our current data.

The OHSU Knight Cancer Institute has begun a program to create a natural language processing (NLP) capability to extract, store, and link data from free text sources at the patient level, and make this data available to researchers in a continuous, reusable, efficient and timely manner through services delivery from the Translational Research Hub (TRH). This talk will present the challenges, progress, and future goals of our program to build NLP capabilities that can help us use free text from the EHR to first support the transformation of cancer research with the hopes of positively impacting clinical care in the future.

Keywords— Text mining; Cancer research; Translational medicine