Inferring Methodological Meta-knowledge from Large Biomedical Corpora

Goran Nenadic
University of Manchester
gnenadic@manchester.ac.uk

Abstract

Large amounts of biomedical corpora have emerged from different sources, including scientific literature, lab notes, patents and electronic health records. Most of the efforts in biomedical text mining have focused on the extraction and linkage of specific facts, such as molecular interactions, links between genes and diseases, or patients' symptoms. Such facts are rarely contextualised using the associated scientific or professional methodology (e.g. what methods were used to detect particular interaction, or to diagnose a particular disease). However, methods are the vital, but often neglected, under-pinning of science and practice. Given enough data, the ability to extract methodological knowledge would allow us to "infer" common (and possibly best) practice for a given task, and thus indeed learn from vast amount of text. This is obviously a complex task that involves identification, representation and linking of steps in associated methods, requiring a series of NLP methods such as temporal information extraction and discourse analysis. In this talk we will explore finding out what methods are being used to do what experiment from the literature, or to infer what clinical pathways patients have followed, based on the notes in their electronic health records. We will illustrate some of the work in the context of bioinformatics (e.g. recovering a general view of the methods described in the literature) and clinical practice (e.g. reconstruction of patient journeys). We will also discuss how feasible this task is given the known issues with the lack of reported details needed for understanding and reproducibility of associated methods (i.e. how much of a method is indeed present in the literate or clinical records).

30th Pacific Asia Conference on Language, Information and Computation (PACLIC 30)
Seoul, Republic of Korea, October 28-30, 2016

5