

A Method for Clinical Concept Extraction from Medical Text

Authors : Moshe Wasserblat, Jonathan Mamou, Oren Pereg

Abstract : Natural Language Processing (NLP) has made a major leap in the last few years, in practical integration into medical solutions; for example, extracting clinical concepts from medical texts such as medical condition, medication, treatment, and symptoms. However, training and deploying those models in real environments still demands a large amount of annotated data and NLP/Machine Learning (ML) expertise, which makes this process costly and time-consuming. We present a practical and efficient method for clinical concept extraction that does not require costly labeled data nor ML expertise. The method includes three steps: Step 1- the user injects a large in-domain text corpus (e.g., PubMed). Then, the system builds a contextual model containing vector representations of concepts in the corpus, in an unsupervised manner (e.g., Phrase2Vec). Step 2- the user provides a seed set of terms representing a specific medical concept (e.g., for the concept of the symptoms, the user may provide: 'dry mouth,' 'itchy skin,' and 'blurred vision'). Then, the system matches the seed set against the contextual model and extracts the most semantically similar terms (e.g., additional symptoms). The result is a complete set of terms related to the medical concept. Step 3 -in production, there is a need to extract medical concepts from the unseen medical text. The system extracts key-phrases from the new text, then matches them against the complete set of terms from step 2, and the most semantically similar will be annotated with the same medical concept category. As an example, the seed symptom concepts would result in the following annotation: "The patient complaints on fatigue [symptom], dry skin [symptom], and Weight loss [symptom], which can be an early sign for Diabetes." Our evaluations show promising results for extracting concepts from medical corpora. The method allows medical analysts to easily and efficiently build taxonomies (in step 2) representing their domain-specific concepts, and automatically annotate a large number of texts (in step 3) for classification/summarization of medical reports.

Keywords : clinical concepts, concept expansion, medical records annotation, medical records summarization

Conference Title : ICAIH 2019 : International Conference on Artificial Intelligence for Healthcare

Conference Location : Paris, France

Conference Dates : December 30-31, 2019