NLP in Electronic Health Records: A Decade Perspective

TREO Talk Paper

James Boit
Dakota State University
James.boit@trojans.dsu.edu

Omar El Gayar
Dakota State University
Omar.El-Gayar@dsu.edu

Hassan Alyami
Dakota State University
hmalyami@pluto.dsu.edu

Abstract

With the increased adoption of electronic health record (EHR) systems, the exponential growth of healthcare big data in both structured and unstructured forms have created a repository of untapped insights. The EHR data such as clinical, radiological, patient and medical narratives are heterogeneous, complex, unstructured and sometimes poorly annotated thus presenting challenges for evidence-based analysis. Today’s information explosion and infoxication give rise to emerging technologies such as natural language processing (NLP) which help in automating the process of information extraction. NLP’s use in healthcare seeks to improve the interaction between clinicians and patients by enhancing the process of clinical information extraction and documentation.

According to MarketsandMarkets (2016), the global market for NLP in healthcare is projected to increase from 7.63 billion US dollars in 2016 to 16.07 billion dollars by 2021. The application of NLP technologies has had a wide-reaching implication in clinical research and practice. The potential uses of NLP in healthcare to improve the quality of patient care and outcomes are gaining traction. This research paper therefore, aims to present evidence on the current state of the art studies investigating the application of NLP in EHR in the last decade with a focus on technical aspects, the performance of algorithms, and application domain. Also, it offers to prescribe suggestions for practitioners and researchers.

References