Text and Ontology Driven Clinical Decision Support System

Deepal Dhariwal[1], Anupam Joshi, PhD[1], Michael A. Grasso, MD, PhD[2]
[1]University of Maryland Baltimore County, Baltimore, MD; [2]University of Maryland School of Medicine, Baltimore, MD

Abstract

In this paper, we discuss our ongoing research in the domain of text and ontology driven clinical decision support system. The proposed framework uses text analytics to extract clinical entities from electronic health records and semantic web analytics to generate a domain specific knowledge base (KB) of patients' clinical facts. Clinical Rules expressed in the Semantic Web Language OWL are used to reason over the KB to infer additional facts about the patient. The KB is then queried to provide clinically relevant information to the physicians.

Introduction

Vast amounts of information are present in unstructured format in physicians' notes. Text processing techniques and medical thesauri such as Unified Medical Language System (UMLS)[1] can be used to extract clinically relevant entities from such data. The extracted entities can then be mapped to concepts from medical ontologies to generate a structured KB of patient facts. Clinical Rules written over this KB could then be used to develop systems that can help with a variety of clinical tasks such as detecting adverse events , decision support alerts in diagnostic process, summarizing patient's medical history, identifying patients with high risk for certain events.

We propose a generic text and ontology driven information extraction framework which will be useful in clinical decision support systems. In the first phase, standard preprocessing techniques such as section tagging, dependency parsing, gazetteer lists are used filter clinical terms from the raw data. In the second phase, a domain specific medical ontology is used to establish relation between the extracted clinical terms. The output of this phase is a Resource Description Framework (RDF http://www.w3.org/RDF/) KB that stores all possible medical facts about the patient. In the final phase, an OWL reasoner and clinical rules are used to infer additional facts about patient and generate a richer KB. This KB can then be queried for a variety of clinical tasks. For instance, each physician or group could establish their own queries to encode the common reasoning used by them help in the differential diagnosis process. In effect, we are creating an automated "Dr. House" that can read information about the patient from prior physicians' notes and test results and make suggestions about possible diagnoses to the physician.

To demonstrate a proof of concept of this framework, we have used discharge summaries from the cardio- vascular domain. In the text processing phase, the clinical records are parsed using the Clinical Text Analysis and Knowledge Extraction System (http://incubator.apache.org/ctakes/) annotators, to extract demographic information, prior medical history, medications, observations, tests taken, laboratory results, surgeries and coronary risk factors. For every concept we consider its polarity, section in which the concept occurs, the associated numerical value, its UMLS synonyms, and the lookup window for the concept. In the knowledge generation phase, we use heart failure ontology[2] to assert cardiovascular related facts about patients. The KB has all the information present in the discharge summaries expressed as RDF triples which correlate the patient, the medical term and a property from the ontology. The KB is then queried to determine the TIMI Risk Score and San Francisco Syncope score for a patient. These are evidence-based guidelines used to categorize the risk of death and ischemic events in patients experiencing unstable angina and to evaluate the risk of adverse outcomes after a syncopal event, respectively.

The goal of this research is to combine factual knowledge about patients, procedural knowledge (clinical rules), and structured knowledge (medical ontologies) to develop a clinical decision support system

Acknowledgement
Deidentified clinical records used in this research were provided by the i2b2 National Center for Biomedical Computing funded by U54LM008748 and were originally prepared for the Shared Tasks for Challenges in NLP for Clinical Data organized by Dr. Ozlem Uzuner, i2b2 and SUNY. The first two authors were supported by funds from the Oros Professorship endowment.

References

1. US Department of Health and Human Services, National Institutes of Health, National Library of Medicine. Unified Medical Language System (UMLS). Available from: www.nlm.nih.gov/research/umls, 2012AB.
2. Jovic A, Gamberger D, Krstacic G. Heart Failure Ontology. Bio-Algorithms & Med-Systems. 2011;7(2):101-110