201905 Augmenting NLP Results by Leveraging SNOMED CT Relationships for Identification of Implantable Cardiac Devices from Patient Notes

Sina Madani, Vanderbilt University Medical Center (US)

Co-authors

1. Dario Giuse
2. Asli Weitkamp
3. Michael McLemore

Summary

Immediate identification of infection in patients with cardiac devices is important in reducing patient's morbidity & mortality. We hypothesized by leveraging SNOMED CT relationships we can amplify the results of an NLP pipeline to identify patients with infection who have cardiac devices.

Audience

Clinical, Research/academic, Technical

Learning Objectives

1. Understand the current issues in ontology-based NLP approaches
2. Formulate an approach to expand clinical concepts by using SNOMED CT attribute and is-a relationships
3. Learn possible solutions in creating and maintaining domain ontologies as sharable knowledge objects

Abstract

Introduction:

Cardiac Implantable Electronic Devices (CIEDs) are commonly used in patients with cardiovascular diseases to increase life expectancy and preventing sudden death. However, implementation of such devices, both during and after the operation, could pose significant complications such as pneumothorax, haematoma, haemothorax, and infection. While the infection rate after the operation is relatively low, considerable morbidity and mortality have been reported with patients who have positive blood culture and identified as high risk of death. Therefore, immediate identification and treatment of infection in patients with CIED is the most critical factor in reducing the morbidity and mortality as well as the cost associated with explantation of the hardware.

At Vanderbilt University Medical Centre we have implemented a Natural Language Processing (NLP) pipeline that parses patient notes real-time once clinical narratives enter the Electronic Medical Record (EMR) system and creates a WordCloud view of patient data. The Word Cloud project began in May 2018 as an effort to create a new visual presentation of the "patient story". As part of the project, every clinical document (discrete data, structured data, reports, and free-text notes) is analyzed to extract clinical concepts mentioned anywhere in the document; the concepts are mapped to the Unified Medical Language System (UMLS) coded terminology. The Word Cloud database lists every occurrence of every concept with the relative timestamp in over 1 billion documents, covering a population of 3.4 million patients. We hypothesized by translating UMLS codes to SNOMED CT concepts and leveraging their relationships we can amplify the results of an NLP system to identify patients with cardiac device as soon as a positive blood culture is transmitted through the messaging platform and send an alert to the attending physicians.

Approach:

We retrospectively extracted all concepts from every clinical document from 1,459 patients whose blood cultures were reported positive between August 2018 and March 2019. Two primary CIED related SNOMED CT concepts were vetted by subject matter experts (SMEs); Permanent Pacemaker and Implantable Cardioverter Defibrillator. Subsequently, we expanded the initial list to 37 concepts based on SNOMED CT is-a (n=29) and attribute (n=23) relationships through iterative reconciliation with SMEs. The SNOMED CT relationships that were used for extending the original two concepts and the corresponding number of associated concepts were: sub class of (29), direct device of (10), has direct device (10), uses device (1), has procedure device (1), and procedure device of (1). In total, 38,779 cardiac device related concepts were extracted by NLP engine for all patient, from which 33.5% of the extracted data (13,306) were attributable to the concepts included in the extension category (n=37).

We are currently evaluating the initial data from this study to implement a real-time alert system for patients with CIEDs, whose infection is reported with positive blood culture from laboratory system. This study will validate the feasibility of using SNOMED CT relationships for augmenting NLP results by expanding domain concepts for an NLP pipeline. We are also considering using the same framework as a decision support tool for clinical abstractors to verify quality metrics extracted from clinical narratives.