Ontology-Based Information Extraction of Quality Metrics from Clinical Narratives

Sina Madani, MD,MS[1,2], Dean F. Sittig, PhD[1], Hua Xu, PhD[1], Parsa Mirhaji, MD, PhD[1], Victoria Jordan[2], PhD, Kim Dunn, MD, PhD[1]

1- The School of Biomedical Informatics |The University of Texas Health Science Center at Houston
2- MD Anderson Cancer Center |The University of Texas

Introduction

The Institute of Medicine reports a growing demand in recent years for quality improvement.

Numerous organizations are involved in the development and reporting of quality measurement metrics.

Disparate data models shift the burden of accurate and reliable metrics extraction and reporting to healthcare providers.

Subjective & Manual abstraction of quality metrics deepens the complexity of consistent, valid, explicit, and comparable quality measurement reporting.

We propose an ontology-based concept extraction framework to utilize unstructured clinical text for defining and reporting quality of care metrics that are interpretable and comparable across boundaries of different institutions.

Methods

- Clinical notes from 2000 patients who had undergone surgery in 2011 at MD Anderson Cancer Center were processed by MetaMap v2012.
- MetaMap XML outputs were converted into Resource Description Framework (RDF) format.
- Three ontologies were developed:
  - o Section header ontology
  - o Concept ontology
  - o Clinical note ontology

Methods (continued)

- Concept ontology contained entities representing five quality metrics from SNOMED terminology (Diabetes, Hypertension, Cardiac Surgery, Transient Ischemic Attack, CNS tumor).
- All ontologies and patient notes (RDFs) were imported into AllegroGraph® triple store as classes and instances respectively (Figure 1).
- SPARQL information retrieval protocol was used for reporting extracted concepts under four settings:
  - Baseline NLP output
  - Inclusion of a concept ontology
  - Exclusion of negated concepts
  - Inclusion of a section header ontology
- Existing manual abstraction data from surgical clinical reviewers, on the same set of patients and documents, was considered as the gold standard.
- Statistical agreement tests (precision, recall, F Measure) were applied for comparing the results of manual abstraction and the proposed ontology-based approach.

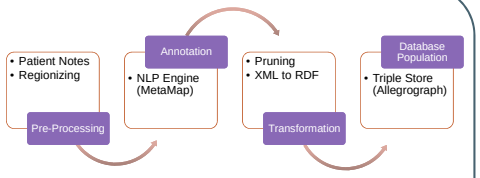

Figure 1 – Process pipeline of patient notes and complementary ontology components

Results

- The overall (micro-average) F-Measure increased from 68% to 82% (Chart 1)

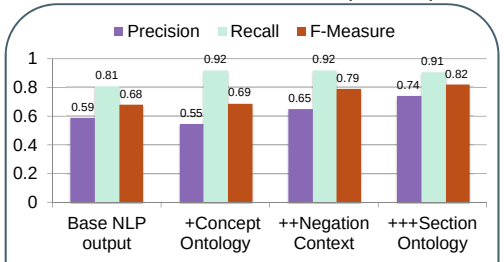

Chart 1- Cumulative ontology effect on base NLP output

- F-Measure increase was highest for CNS tumors (63%), Cardiac Surgery (57%), and TIA (32%)
- F-Measure increase was lowest for Hypertension (9%) and Diabetes (1%)
- We also compared the isolated effect of each ontology layer to the base NLP output (Chart 2)

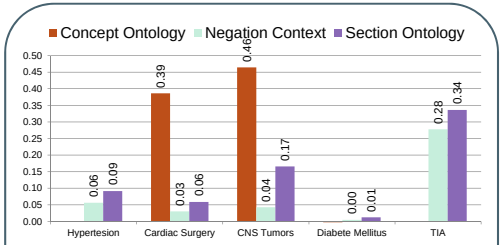

Chart 2- Isolated ontology effect on base NLP output

- The overall (micro-average) effect of section ontology was 11% and 5%
- (F-Measure) higher compared to concept ontology and negation context respectively

Summaries of Conclusions

- The application of ontology-based approach for natural language processing in our study has provided complementary mechanisms for increasing the performance of NLP systems.
- The pivot point for extracting more meaningful quality metrics from clinical narratives is the abstraction of contextual semantics hidden in the notes.
- We have defined some of these semantics and quantified them in multiple complementary layers
- Rigorous evaluations are still necessary to
  - o Create evaluation guidelines for assessment of performance of ontology-based information extraction systems
  - o Provide a consistent baseline for the purpose of comparing alternative approaches

Reference

Committee on Redesigning Health Insurance Performance Measures Payment and Performance Improvement Programs. Institute of Medicine. Rewarding provider performance. incentives in Medicare. Washington, DC: National Academies Press; 2007.

Acknowledgement

This project was supported by Grant No. 10510592 for Patient-Centered Cognitive Support under the Strategic Health IT Advanced Research Projects (SHARP) from the Office of the National Coordinator for Health Information Technology.

Please contact the first author via email: sina.madani@uth.tmc.edu