Improving Automated Clinical Follow-up through Optimization of Natural Language Processing (NLP) Methods

Mark D. Kovacs, MD, Medical University of South Carolina (Presenter); Joseph Mesterhazy; David E. Avrin, MD, FSIIIM; Thomas Urbania, MD; John Mongan, MD

Background

A major challenge for radiologists is obtaining relevant clinical follow-up for the cases they read. We developed a PACS-accessible, Electronic Medical Record (EMR)-integrated system, Correlate, which automatically finds answers to follow-up questions entered by radiologists by using NLP to prospectively search the EMR until an answer is identified. Correlate builds on prior work in EMR data searching by its novel use of NLP. The web-based application is launched from the PACS and a query is entered along with a set of criteria that help define relevant results. The system then searches the EMR each day, interfacing using MIRTH connect (a cross-platform HL7 interface engine). The question and criteria, as well as the retrieved unstructured medical record results, are mapped to Unified Medical Language System (UMLS) concepts using the MetaMap program. The Correlate algorithm then ranks retrieved results in order of likely relevance. If a retrieved patient record meets criteria for relevance, Correlate sends an email alerting the user with a link to the result.

Case Presentation

We will briefly describe the Correlate system architecture and workflow and provide statistics regarding the performance of the system in returning clinically relevant results from EMR searches. We will describe several methods used to optimize the relevance of returned results and the efficiency of the search.

Outcome

Since the beginning of May 2015, a total of 89 queries were submitted through Correlate, of which 58 returned results from the EMR. The total number of searches yielding relevant results (defined as relevance score above 60%) was 47. By this estimate, 81% of queries returning results were clinically relevant.

Regarding MIRTH connect statistics, searches were made on a total of 81 patients, and 1616 individual EMR records were parsed and analyzed through MetaMap and the Correlate scoring algorithm.
Correlate Client and Server Architecture

**PACS Diagnostic Workstation**
- PACS Client
- Web Integration

**PACS Server**
- Exams
- SOAP API

**MIRTH Interface**
- Patients with Questions
- Patient Results
- Email Notification

**Correlate Web Server**
- Question
  - Question + Keywords
  - Exam Metadata + Report
  - Mapped UMLS Concepts
- JavaScript Scoring Algorithms

**Answer**
- Result
- Mapped UMLS Concepts

**Correlate application architecture:**
- Grails 2.5 + jQuery UI, Apache Tomcat 8
- PostgreSQL 9.5, MIRTH Connect 3.2
- MetaMap v.2014 + Node.js

**PACS integration:**
- Server SOAP API
- Diagnostic Workstation Client API

**MetaMap Server**
- Node.js REST Interface

**MetaMap NLP**
- UMLS Metathesaurus Concepts
- Worse Sense Disambiguation / Part of Speech Tagger
- Medical Datasets (SNOMED CT)

---

**Figure 2**

CT scan image with annotations.
Figure 3

CT

Impression:
1. Moderate right hydronephrosis and hydronephrosis with transition point in the mid ureter likely secondary to mass effect from adjacent small retroperitoneal lymph nodes and soft tissue stranding.
2. Occlusion and thrombosis of inferior and superior branches of the left inferior mesenteric and superior mesenteric of the left hepatic vein branch to the medial segment of the left hepatic lobe.
3. Internal development of small volume ascites with hyper-enhancing fluid collection, may be secondary to portal vein versus umbilical vein.

END OF IMPRESSION:

Figure 4

CT

Impression:
1. Moderate right hydronephrosis and hydronephrosis with transition point in the mid ureter likely secondary to mass effect from adjacent small retroperitoneal lymph nodes and soft tissue stranding.
2. Occlusion and thrombosis of inferior and superior branches of the left inferior mesenteric and superior mesenteric of the left hepatic vein branch to the medial segment of the left hepatic lobe.
3. Internal development of small volume ascites with hyper-enhancing fluid collection, may be secondary to portal vein versus umbilical vein.

END OF IMPRESSION:
Discussion

Correlate builds on previous work in EMR data searching / automated clinical follow-up in its novel use of NLP. We improved the ease of use of our web-based search query page, including adding a decision tree whereby one can search for a specific type of result (e.g. follow-up imaging study or a pathology result) versus broadly search the EMR for relevant results. We added functionality to restrict results by modality. We improved utilization of semantic types to further streamline results. We identified certain types of results (e.g. Nursing notes) to exclude as generally not being relevant.

We improved the EMR search by updating Metamap 2013 to Metamap 2014, resulting in a significant increase in speed and accuracy. For example, a surgical report that previously took 1 minute 20 seconds to parse now takes 20 seconds. We also replaced the MetaMap Java API that processes reports with a custom node.js process that parses results in parallel, resulting in near-real-time processing.
Conclusion

Following early success using Correlate at our institution, further improvements to the search algorithm resulted in improved relevancy of returned results. Updating the program to use MetaMap 2014, as well as changing the way EMR reports are processed, also resulted in significant improvements in program efficiency.

References


Keywords

Correlate; Natural Language Processing (NLP); Electronic Medical Record (EMR); PACS; Automated Clinical Follow-Up