Building a Corpus for Sentiment Analysis of Surgical Operation Notes

Jake Albert,[1] Dragomir Radev PhD,[1] and Wade Schulz MD PhD [2,3]

[1]Department of Computer Science, Yale University; [2]Department of Laboratory Medicine, Yale School of Medicine;
[3]Yale-New Haven Hospital Center for Outcomes Research and Evaluation

LILY Lab

Introduction

There are many uses for sentiment analysis in the medical domain. One such task is assessing patients' risk from their medical history. For example, the Society of Thoracic Surgeons (STS) developed a model that, given information about a patient's reaction to previous cardiothoracic surgeries, assesses risk of complications arising should the patient undergo another operation. A human judge reads an unstructured surgical operation note in order to extract this information, but sentiment analysis of the same text can contribute to this model.

This project sought to analyze the sentiment of surgical operation notes based on factors that are more relevant to assessing future patient risk than the "good to bad"` polarity used in domains such as reviews. Two factors were selected:

- How complicated a surgery was
- How concerned the surgical team is assumed to be about the patient after completing the surgery

Problem Description

The domain of surgical operation notes presents two main challenges for sentiment analysis:

- Surgical operation notes are written objectively to detail how the surgery progressed and thus are sparse in sentiment terms such as emotion words.
- Many words in the medical domain connote a different sentiment to what they connote in everyday domains. Thus, large sentiment lexicons such as SentiWordNet and the Subjectivity Lexicon can be used with only limited success for medical texts.

To address these difficulties, surgical notes were annotated for the above sentiment categories at the word, sentence, and note level to create a domain-specific corpus for sentiment analysis.

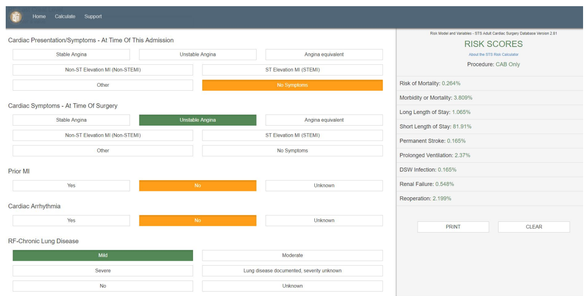

Figure 1: STS Risk Calculator. As an example, one fields specifies for "Cardiac Symptoms - At Time of Surgery." Here, unstable angina during the surgery indicates increased risk in future surgeries.

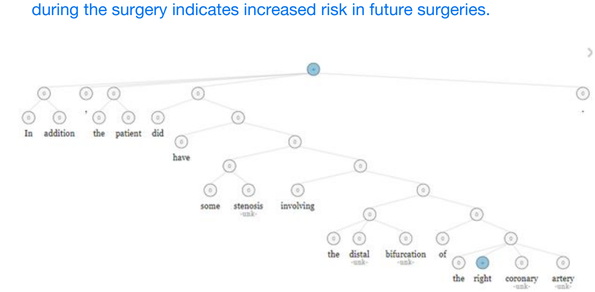

Figure 2. Stanford Sentiment Analysis Tool's incorrect labeling of the word "right," and consequently the sentence as a whole, as positive. The word "right" is used more often in its sense as a direction than its sense as "correct" in surgical operation notes.

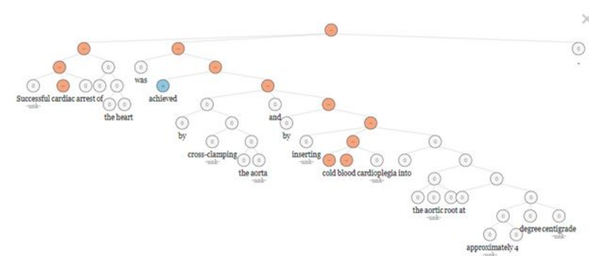

Figure 3. Stanford Sentiment Analysis Tool's incorrect labeling of the words "cold," "blood," and "cardiac," and the sentence as a whole, as negative. The sentence actually describes a successful step in the surgery.

	COMPLICATED: NO	COMPLICATED: YES
CONCERNING: NO	"All hemodynamic monitors, including intraoperative TEE, baseline EKG, and pulmonary monitoring were found to be normal."	"First, the PDA was exposed, and the right coronary artery was found to have moderate calcification throughout its course to the terminal bifurcation."
CONCERNING: YES	"Dressings were applied, and the patient was transferred to the CTICU on multiple drips of epinephrine, neosynephrine, and vasopressin for hypotension."	"The aortic cross-clamp was removed, however the heart was not noted to resume spontaneous coordinated contractions."

Figure 4. Examples of sentences that show a surgery is a) neither complicated nor concerning, b) complicated but not concerning, c) not complicated but concerning, and d) both complicated and concerning.

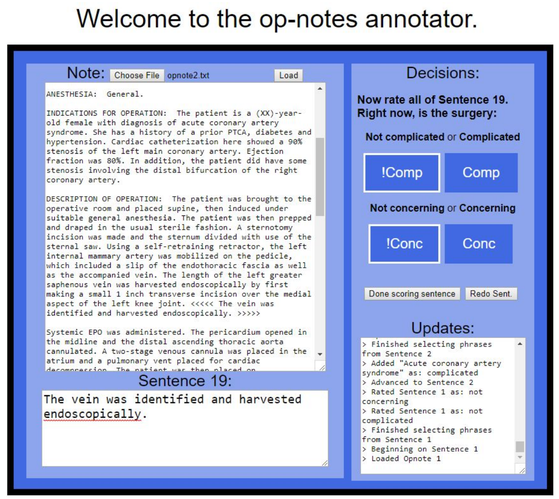

Figure 5. The Op-Notes Annotator Tool. Annotators label 1) words and sequences of words, 2) full sentences, and 3) the full note with a score indicating whether the surgery was complicated or not, and whether it was concerning or not.

Methods and Data

Surgical operation notes were drawn from Yale New Haven Hospital (YNHH) at the Center for Outcomes Research and Evaluation (CORE) and stored as JSON files on secured servers. They were annotated by doctors associated with CORE using an annotator tool written in JavaScript. Word, sentence, and document labels were stored as JSON.

Labels of sentences from the annotated notes were used for supervised learning of a binary classifier by means of an RNN implemented with LSTM units in TensorFlow. The annotated sentences were divided 90% into a training set 10% into a testing set. The RNN was trained on both word embeddings generated from the training set, and on word embeddings imported from BioNLP, a collection of word vectors trained on PubMed and PMC biomedical papers.

Testing accuracy for both RNN training schemes was compared against a baseline result of a Naive Bayes Classifier written in Python using the Natural Language Toolkit.

Conclusion

This framework offers promise if trained on sets of several thousand surgical operation notes. There are three areas of suggested investigation:

- Developing a sentiment lexicon from the annotated notes for complicated/not complicated and concerning/not concerning
- Continuing to improve the sentence classifier
- Determining note-level sentiment as a function of sentence sentiment (mean, mode, or a more complex but learnable function)

Acknowledgement

Thank you to Professor Radev and to Dr. Schulz for their guidance throughout this project. I have learned a lot through this work.