

A New Corpus for Clinical Events with Change of State

Meliha Yetisgen-Yildiz, PhD^{1,2}, Prescott Klassen, MS², Lucy Vanderwende, PhD^{3,1},
Fei Xia, PhD^{2,1}

¹Biomedical and Health Informatics, ²Department of Linguistics, University of Washington, Seattle, WA; ³Microsoft Research, Redmond, WA

Abstract

Understanding the event structure of sentences and the whole documents is an important step in being able to extract meaningful information from text. Our task is the identification of critical illness phenotypes, specifically pneumonia, from clinical narratives. To capture those phenotypes, it is important to identify the change of state for events, in particular events that measure and compare multiple states across time. In this abstract, we describe a corpus annotated for events with change of state information. Our corpus is comprised of chest x-ray reports, where we find many descriptions of change of state comparing the volume and density of the lungs and surrounding areas.

Introduction

The narrative accompanying chest X-rays contains a wealth of information that is used to assess the health of a patient. X-rays are obviously a single snapshot in time, but the report narrative often makes either explicit or, more often, implicit reference to a previous X-ray. In this way, the sequence of X-ray reports is used not only to assess a patient's health at a moment in time but also to monitor change. Critical illness phenotypes such as pneumonia are consensus-defined diseases, which means that the diagnosis is typically established by human inspection of the data rather than by means of a test. We are in the process of developing a phenotype detection system for pneumonia. In order to train and evaluate the system, we asked medical experts to annotate the X-ray report with phenotype labels and to highlight the text snippets in the report that supported the phenotype labeling. Analysis of the text snippets revealed that most of these snippets mention a change of state or lack of a change of state (i.e., persistent state). We created a corpus from the 1008 highlighted text snippets and annotated them for events with change of state.

Annotation

In our annotation schema¹, an event in our corpus is represented as a (loc, attr, val, cos, ref) tuple, where *loc* is the anatomical location (e.g., "lung"), *attr* is an attribute of the location that the event is about (e.g., "density"), *val* is a possible value for the attribute (e.g., "clear"), *cos* indicates the change of state for the attribute value compared to some previous report (e.g., "unchanged"), and *ref* is a link to the report(s) that change of state is compared to (e.g., "prior examination"). Not all the fields in the tuple are required to be present in an event. When a field is absent, either it can be inferred from the context or it is unspecified. Figure 1 includes example chest x-ray report sentences with event annotations. More detailed information can be found in our annotation guideline downloadable from UW-BioNLP website (<http://depts.washington.edu/bionlp/index.html>).

Three annotators who are graduate students annotated events in the 1008 snippets. 100 of the snippets were annotated by all three annotators. The inter-rater agreement at the tuple level was 0.85 macro f-score and 0.89 micro f-score.

Future Work

Our ultimate goal is to train a statistical event detection approach and use it in phenotype detection and other NLP systems to monitor patients' medical conditions over time and prompt physicians with early warning, expecting that this will improve patient health care quality while reducing the overall cost of health care.

Acknowledgements

Microsoft Research Connections, University of Washington Research Royalty Fund, UL1TR000423

References

1. Vanderwende L, Xia F, Yetisgen-Yildiz M. Annotating Change of State for Clinical Events. Proceedings of the 1st Workshop on EVENTS: Definition, Detection, Coreference, and Representation Workshop of NAACL'2013, 2013.

- (1) The *lungs* are clear.
(lungs, <density>, clear, -, -)
- (2) *Lungs*: No focal opacities.
(lung ... focal, opacities, no, -, -)
- (3) The *chest* is **otherwise unchanged**.
(chest, -, -, otherwise unchanged, -)
- (4) *Left base* opacity has **increased** and *right based* opacity **persists** which could represent atelectasis, aspiration, or pneumonia.
(left base, opacity, -, increased, -)
(right base, opacity, -, persists, -)
- (5) **Since the prior examination** *lung volumes* had **diminished**.
(lung, volumes, -, diminished, prior examination)

Figure 1. Example snippets with event annotations.