Annotating Change of State for Clinical Events

Lucy Vanderwende
One Microsoft Way
Redmond, WA 98052
lucyv@microsoft.com

Fei Xia
University of Washington
Seattle, WA 98195
fxia@uw.edu

Meliha Yetisgen-Yildiz
University of Washington
Seattle, WA 98195
melihay@uw.edu

Abstract

Understanding the event structure of sentences and whole documents is an important step in being able to extract meaningful information from the text. Our task is the identification of phenotypes, specifically, pneumonia, from clinical narratives. In this paper, we consider the importance of identifying the change of state for events, in particular, events that measure and compare multiple states across time. Change of state is important to the clinical diagnosis of pneumonia; in the example "there are bibasilar opacities that are unchanged", the presence of *bibasilar opacities* alone may suggest pneumonia, but not when they are unchanged, which suggests the need to modify 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. We propose an annotation schema to capture this information as a tuple of <location, attribute, value, change-of-state, time-reference>.

1 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 X-ray 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. 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. Our recent efforts have focused on building a phenotype detection system. 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 X-ray report that supported their phenotype labeling.

Analysis of the text snippets that support the labeling of pneumonia and the Clinical Pulmonary Infection Score (CPIS) reveal that most of these snippets mention a change of state or the lack of a change of state (i.e. persistent state). This is understandable given our task, which is to monitor patients for ventilator associated pneumonia (VAP), which can develop over time as a patient is kept on a ventilator for medical reasons.

Change of state (COS) is most often understood as an aspectual difference that is reflected in verb morphology (Comrie, 1976), where a state is described as initiating, continuing or terminating (see also Quirk et al., 1973, Section 3.36). In our corpus, however, COS is often reflected not in verbs, but more frequently in nouns. A careful analysis of our data indicates that the states expressed as nouns don't have the traditional aspects but rather exhibit COS more closely associated with comparatives, as they are susceptible to subjective and to objective measurement (Quirk et al., 1973, Section 5.38). These events compare two states across time or comparing one state against an accepted norm. Monitoring the state of the patient, and therefore comparing current state with previous states, is of paramount importance in the clinical scenario. We therefore propose in this paper to expand the annotation of COS to include the comparison of states over time.

2 The Task

Early detection and treatment of ventilator associated pneumonia (VAP) is important as it is the most common healthcare-associated infection in critically ill patients. Even short-term delays in appropriate antibiotic therapy for patients with VAP are associated with higher mortality rates, longer-term mechanical ventilation, and excessive hospital costs. Interpretation of meaningful information from the electronica medical records at the bedside is complicated by high data volume, lack of integrated data displays and text-based clinical reports that can only be reviewed by manual search. This cumbersome data management strategy obscures the subtle signs of early infection.

Our research goal is to build NLP systems that identify patients who are developing critical illnesses in a manner timely enough for early treatment. As a first step, we have built a system that determines whether a patient has pneumonia based on the patient's chest X-ray reports; see Figure 1 for an example.

01 CHEST, PORTABLE 1 VIEW
02 INDICATION:
03 Shortness of breath
04 COMPARISON: July 16 10 recent prior
05 FINDINGS:
06 Left central line, tip at mid-SVC.
07 Cardiac and mediastinal contours as before
08 No pneumothorax.
09 Lungs: Interval increase in right lung base
10 pulmonary opacity with air bronchograms,
11 increasing pneumonitis / atelectasis.

Figure 1. Sample chest X-ray report

2.1 Annotation

To train and evaluate the system, we created a corpus of 1344 chest X-ray reports from our institution (Xia and Yetisgen-Yildiz, 2012). Two annotators, one a general surgeon and the other a data analyst in a surgery department, read each report and determined whether the patient has pneumonia (PNA) and also what the clinical pulmonary infection score (CPIS) is for the patient. The CPIS is used to assist in the clinical diagnosis of VAP by predicting which patients will benefit from obtaining pulmonary cultures, an invasive procedure otherwise avoided. There are three possible labels for PNA: (2a) no suspicion (negative class), (2b) suspicion of PNA, and (2c) probable PNA (positive class). Likewise, there are three labels for CPIS: (1a) no infiltrate, report can include mention of edema or pleural effusion, (1b) diffuse infiltrate or atelectasis (i.e. reduced lung volume), and (1c) localized infiltrate, where one opacity is specifically highlighted and either PNA or infection is also mentioned.

In addition to the labels, we also asked the annotators to highlight the text snippet they used to assign the CPIS and PNA categories to reports (see (Yu et al., 2011) for similar approach to capturing expert knowledge). Thus, the snippets represent the support found for the CPIS and PNA label determination. The snippet found in lines 9-11, in figure 1, for example, was support for both the CPIS (1c) and the PNA label (2c).

2.2 Preliminary Results

We used this corpus to train two SVM classifiers, one for CPIS and the other for PNA, and evaluated them using 5-fold cross validation (for details, see Tepper et al., 2013). The micro F1-score of the CPIS classifier was 85.8% with unigram features and 85.2% with unigram+bigram features. The micro F1-score of the PNA classifier was 78.5% with unigrams and 78.0% with unigram+bigrams.

We analyzed errors made by the CPIS and PNA classifiers and observed that many of them were due to lack of in-depth semantic analysis of text. Consider the snippet "*The previously noted right upper lobe opacity consistent with right upper lobe collapse has resolved*", which is labeled in the gold standard 1A (no infiltrate). The system mislabeled it 1C, (localized infiltrate), because the snippet supports 1C entirely up until the crucial words "has resolved". This error analysis motivated the clinical event annotation task described in this paper.

3 Change of State for Clinical Events

In our data, clinically relevant events are often expressed as nouns. A text that mentions "a clear lung", for instance, implicitly describes the event of checking the lung density for that patient and finding it to be clear[1]. The TimeML annotation guidelines (Saurí et al., 2012) specify that states are to be annotated when they "identifiably change over the course of a document being marked up". In our scenario, where the document is the collection of the patient's medical notes during hospital stay, a noun phrase such as "lung capacity" is then a state that can certainly change over the course of the document.

Our corpus contains radiology reports and highlighted snippets of text where annotators found support for their finding. It is noteworthy that these snippets frequently describe observations of change, either in lung volume or in density. In fact, these changes of state (henceforth COS) appear more often in these snippets than non-snippets. Taking a random sample of 100 snippets, we found that 83/100 included some signal for COS, while a random sample of 100 non-snippet sentences included only 61/100 mentions of COS.

Let us consider some examples of snippets in which the clinical events, in italics, are referred to using nouns, a shorthand for examination / measurement of the noun in question. We have marked the signal words expressing a comparison across time in bold.

1. The *lungs* are clear.
2. *Lungs*: No focal opacities.
3. The *chest* is **otherwise unchanged**.
4. *Left base* opacity has **increased** and *right base* opacity **persists** which could represent atelectasis, aspiration, or pneumonia.

Snippets 1 and 2 describe states in the current X-ray report and do not express a COS. A close look at 3 and 4, however, reveals language that indicates that the experts are comparing the state in the current X-ray with at least one other X-ray for that patient and in doing so, are describing a COS. Consider the phrases "otherwise unchanged" in snippet 3, and "increased" and "persists" in snippet 4. Such words signal that the radiologist is examining more than one report at a time and making comparisons across these X-rays, without explicit reference to the other X-rays. There are other examples which exhibit explicit reference, for example, snippets 5 and 6, where the signal words and the explicit reference are in boldface, and the clinical events in italics:

5. *Bilateral lower lobe* opacities are **similar** to those seen on **DATE**
6. **Since the prior examination** *lung volumes* have **diminished**

Previous COS analyses (e.g., (Sun et al., 2013; Saurí, 2005)) have largely been limited to an analysis where events are expressed as verbs, and so is usually restricted to aspectual distinctions such as start, stop, and continue. In our data, however, many of the events are expressed as nouns and so we propose to extend the COS analysis to include measurements comparing two or more successive states and so will include concepts such as more, less, and equal[2].

4 Annotating change of state

While previous event annotation (Uzuner et al., 2010; Uzuner et al., 2011; Albright et al., 2013) marks multiple types of events, temporal expressions, and event relations, our annotation focuses on tracking changes in a patient's medical conditions. 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 the change of state is compared to (e.g., "prior examination"). Not all the fields in the tuple will be present in an event. When a field is absent, either it can be inferred from the context or it is unspecified.

[1] The guidelines for the 2012 i2b2 temporal relation challenge define events as "clinically relevant events and situations, symptoms, tests, procedures, …" (Sun et al., 2013)

[2] In English, the morphology provides evidence, though rarely, that the comparative is a property of the change of state of an adjective. Consider the verb "redden", a derived form of the adjective "red", which means "to become more red", combining the inchoative and comparative (Chris Brockett, pc.)

The annotations for Snippets 1-6 are as follows: a dash indicates that the field is unspecified, and <…> indicates the field is unspecified but can be inferred from the location and the attribute value. For instance, the attribute value *clear* when referring to the location *lungs* implies that the attribute being discussed is the *density* of the lung.

Ex1: (lungs, <density>, clear, -, -)
Ex2: (lungs, <density>, no focal opacities, -, -)
Ex3: (chest, -, -, unchanged, -)
Ex4: (left base, <density>, opacity, increased, -), and (right base, <density>, opacity, persists, -)
Ex5: (Bilateral lower lobe, <density>, opacities, similar, DATE)
Ex6: (lung, volumes, -, diminished, prior examination)

A few points are worth noting. First, the mapping from the syntactic structure to fields in event tuples is many-to-many. For example, a noun phrase consisting of an adjective and noun may correspond to one or more fields in an event tuple. For instance, in the NP *left base opacity* in example 4, *left base* is *loc*, and *opacity* is *val*. In example 6, the NP *lung volumes* will be annotated with *lung* as loc and *volumes* as *attr*, but no *val*. Similarly, an adjective can be part of a *loc* (e.g., *bilateral* in example 5), a *val* (e.g., *clear* in example 1), or a *cos* (e.g., *unchanged* in example 3). Finally, the cos field may also be filled by a verb (e.g., *increase* and *persists*, in example 4). Making such distinctions will not be easy, especially for annotators with no medical training.

Second, events often have other attributes such as polarity (positive or negative) and modality (e.g., factual, conditional, possible). Most events in X-ray reports are positive and factual. We will add those attributes to our representations if needed.

5 Summary

Annotating events in a general domain without targeting a particular application can be challenging because it is often not clear what should be marked as an event. Our annotation focuses on the marking of COS in medical reports because COS is an important indicator of the patient's medical condition. We propose to extend COS analysis to include comparison of state over time.

We are currently annotating a corpus of X-ray reports with the COS events. Once the corpus is complete, we will use it to train a system to detect such events automatically. The events identified by the event detector will then be used as features for phenotype detection. We expect that the COS features will improve phenotype detection accuracy, in the same way that using features that encode negation and assertion types improves classification results as demonstrated by Bejan et al. (2012).

Our ultimate goal is to use event detection, phenotype detection, and other NLP systems to monitor patients' medical conditions over time and prompt physicians with early warning, and thus improve patient healthcare quality while reducing the overall cost of healthcare.

Acknowledgments

We wish to thank the anonymous reviewers for their comments and also our colleagues Heather Evans at UW Medicine, and Michael Tepper, Cosmin Bejan and Prescott Klassen at the University of Washington. This work is funded in part by Microsoft Research Connections and University of Washington Research Royalty Fund.

References

Daniel Albright, Arrick Lanfranchi, Anwen Fredriksen, William F. Styler IV, Colin Warner, Jena D. Hwang, Jinho D. Choi, Dmitry Dligach, Rodney D. Nielsen, James Martin, Wayne Ward, Martha Palmer, and Guergana K. Savova. 2013. *Towards comprehensive syntactic and semantic annotations of the clinical narrative*. Journal of American Medical Informatics Association (JAMIA). [Epub ahead of print].

Cosmin A. Bejan, Lucy Vanderwende, Fei Xia, and Meliha Yetisgen-Yildiz. 2013. *Assertion modeling and its role in clinical phenotype identification*. Journal of Biomedical Informatics, 46(1):68-74.

Bernard Comrie. 1976. *Aspect*. Cambridge Textbooks in Linguistics.

Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik, 1973. *A Grammar of Contemporary English*. . Longman Group Ltd, London

Roser Saurí, Jessica Littman, Bob Knippen, Robert Gaizauskas, Andrea Setzer, and James Pustejovsky. 2005. TimeML Annotation Guidelines Version 1.2.1. Manuscript, Available at http://www.timeml.org/site/publications/specs.html

Weiyi Sun, Anna Rumshisky, Ozlem Uzuner. 2013. Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. In Journal of the American Medical

Informatics Association (JAMIA). Published Online First: 5 April 2013 10.1136/amiajnl-2013-001628.

Michael Tepper, Heather. Evans, Fei Xia, and Meliha Yetisgen-Yildiz. 2013. *Modeling Annotator Rationales with Application to Pneumonia Classification*. In Proceedings of Expanding the Boundaries of Health Informatics Using AI Workshop in conjunction with AAAI'2013.

Özlem Uzuner, Imre Solti, Fei Xia, and Eithon Cadag. 2010. *Community annotation experiment for ground truth generation for the i2b2 medication challenge*. Journal of American Medical Informatics Association (JAMIA), 17(5):519-23.

Özlem Uzuner, Brent R. South, Shuying Shen, and Scott L. DuVall. 2011. *2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text*. Journal of American Medical Informatics Association (JAMIA), 18(5):552-556.

Fei Xia and Meliha Yetisgen-Yildiz. 2012. *Clinical corpus annotation: challenges and strategies*. In Proceedings of the Third Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM'2012) in conjunction with the International Conference on Language Resources and Evaluation (LREC), Istanbul, Turkey.

Shipeng Yu, Faisal Farooq, Balaji Krishnapuram, and Bharat Rao. 2011. *Leveraging Rich Annotations to Improve Learning of Medical Concepts from Clinical Free Text*. In Proceedings of the ICML workshop on Learning from Unstructured Clinical Text. Bellevue, WA.