Temporal Relation Extraction from Medical Discharge Summaries

Emily Silgard, Melissa Tharp, Rutu Mulkar-Mehta
esilgard@yahoo.com, mtharp@ucsd.edu, me@rutumulkar.com
University of California, San Diego, CA, USA

***Abstract*—In this paper we discover temporal relations in patient discharge summaries, when the relevant medical events and temporal expressions were provided in the training data.**

Keywords: Temporal Relation, Biomedical Informatics

I. Introduction

In this paper we describe our work on extracting temporal relations from patient discharge summaries. We simplify the procedure for extracting temporal relations, by using pre-specified temporal events (such as medications, diagnosis and symptoms) and temporal markers (such as admission date and discharge date) for discovering the link between events and timestamps using statistical classifiers and rule-based systems.

II. Related Work

There have been numerous initiatives for extracting temporal relations from discourse, such as the works by Mani et al. (2006), Chambers, Wang, Jurafsky (2007), Zhou et al. (2005) to name a few. However, none of the above approaches can be applied to discovering temporal links in medical discharge summaries, which remains a challenging problem.

III. Approach

The focus of our work was on discovering temporal relations between clinically relevant events within the scope of a single sentence. The data for analysis (patient discharge summaries) was 190 articles selected from the MIMIC-II corpus, pre-annotated with events, and temporal anchors. The task was to identify temporal relations, or "TLINKs" that represent a relative ordering, when relating either two events, or an event and a temporal anchor. For instance, in the sentence, "She was ***transferred*** to the ***CSRU***", ***transferred*** and ***CSRU*** are the two events related by a 'before' relation, i.e. the ***transfer*** happens before going to ***CSRU***.

We use machine learning systems to classify relationships between the events and temporal expressions within single sentences. The feature set used for the events include polarity (positive or negative), modality (factual or proposed), POS tags, other lexical items that appear in the sentence, and whether they fall before after or between the events under consideration. The feature vectors were classified using a Naïve Bayes Model (with delta=.1 smoothing), where the probability of a potential temporal relation belonging to a class (before, overlap, or none) is equal to the product of the probability of each feature in that link given the class, times the probability of the class itself.

A parallel system deals with words in different sentences throughout the discharge summary that could potentially refer to the same admission event. We detect these events with a rule-based system, looking for partial string matches, where there was no evidence of a second iteration of the same event (such as "prior", "another") and where the event is not referring to a specific hospital or department.

Admission date and discharge date are the only temporal anchors annotated in the text. The majority of events listed in the section - patient's medical history - happen before admission, and similarly the majority of events discussed in the section - hospital course - happen before discharge. We use a simple rule based approach to capture these "before" relations. Exceptions occur when the event in question is "admission" or "discharge" itself, in which case the temporal relation would be "overlap". Other exceptions include the presence of the trigger words, such as "follow-up", which mostly likely signals a future event.

IV. Results

Each of the three parts of our system attempted to classify a subset of the potential temporal relations in the document. As a result, recall of the individual components judged against all of the gold standard links is not a very informative evaluation measure. We report precision of the components as follows: 64% for admission co-reference detection, 90% for anchoring events to the admission or discharge date, and 60% for discovering temporal links within a sentence using machine learning.

References

[1] Chambers, N.; Wang, S.; and Jurafsky, D. 2007."Classifying temporal relations between events". In the proceedings of The Association for Computer Linguistics.

[2] Mani, I.; Verhagen, M.;Wellner, B.; Lee, C. M.; and Pustejovsky, J. 2006. "Machine learning of temporal relations", Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meetingof the ACL 753–760.

[3] Zhou, L.; C. Friedman, C; Parsons, S.; Hripcsak,G.; 2005, "System Architecture for Temporal Information Extractino, Representation and Reasoning in Clinical Narrative Reports" AMIA Annu Symp Proc. V.2005; 2005. PMC1560