Temporal Classification of Medical Events

Preethi Raghavan[*]**, Eric Fosler-Lussier**[*]**, and Albert M. Lai**[†]
[*]Department of Computer Science and Engineering
[†]Department of Biomedical Informatics
The Ohio State University, Columbus, Ohio, USA
{raghavap, fosler}@cse.ohio-state.edu, albert.lai@osumc.edu

Abstract

We investigate the task of assigning medical events in clinical narratives to discrete time-bins. The time-bins are defined to capture when a medical event occurs relative to the hospital admission date in each clinical narrative. We model the problem as a sequence tagging task using Conditional Random Fields. We extract a combination of lexical, section-based and temporal features from medical events in each clinical narrative. The sequence tagging system outperforms a system that does not utilize any sequence information modeled using a Maximum Entropy classifier. We present results with both hand-tagged as well as automatically extracted features. We observe over 8% improvement in overall tagging accuracy with the inclusion of sequence information.

1 Introduction

There has been a lot of interest in building timelines of medical events from unstructured patient narratives (Jung et al., 2011; Zhou and Hripcsak, 2007). Creating a timeline from longitudinal clinical text requires learning temporal relations such as *before, simultaneous, includes, overlaps, begins, ends* and their inverses between medical events found within and across patient narratives (Allen, 1981). However, learning temporal relations for fine-grained temporal ordering of medical events in clinical text is challenging: the temporal cues typically found in clinical text may not always be sufficient for this task.

An important characteristic of a clinical narrative is that the medical events in the same narrative are more or less semantically related by narrative discourse structure. However, medical events in the narrative are not ordered chronologically. Thus, the clinical narrative structure is not always temporally coherent.

Moreover, extracting precise temporal features for highly accurate temporal ordering of medical events is difficult as the temporal relationship between medical events is varied and complicated. Zhou and Hripcsak (2007) identify six major categories of temporal expressions from a corpus of discharge summaries: "date and time," "relative date and time," "duration," "event-dependent temporal expression," "fuzzy time," and "recurring times." Their study of temporal expressions in clinical text indicates that relative time (e.g., ever since the *episode* 2 days ago) may be more prevalent than absolute time (e.g., 06/03/2007). Further, temporal expressions may be fuzzy where "history of *cocaine use*" may imply that *cocaine use* started 2 years ago or 10 years ago.

In this paper, we address a relatively simpler task of assigning medical events to coarsely defined time-bins. The time-bins, *way before admission, before admission, on admission, after admission, after discharge*, are defined based on the relative temporal distance of the medical event from the admission date, which is the only explicit date almost always found in each clinical narrative. We extract features based on narrative structure as well as temporal expressions to label a sequence of medical events from each clinical narrative with a highly probable

HISTORY PHYSICAL   DATE: 06/03/2007
NAME: Smith Jack   MR#: XXX-XX-XXXX
ATTENDING PHYSICIAN: Bill Payne MD   DOB: 02/28/1960
CHIEF COMPLAINT
Chest pain and arm infection.
HISTORY OF PRESENT ILLNESS   ①   ②   ③
Patient is a 48-year-old male with history of cocaine use hypertension who presents with chest pain ④
⑤ which started 2 days ago . He did not having chest pain yesterday but ever since the episode 2 days ago
⑥ he has felt a little weaker. He did have chest pain today and this is what prompted him to come to the
ER. He also notices that he has had some infections under his arms. He states that he had to have an ⑧
⑦ abscess I and D 3 or 4 months ago under his arm and 2 to 3 weeks ago he noticed some more spots and
these spots have now grown and now are under both arms. Currently he is chest pain free. His blood
pressure upon presentation was 189/106.   ⑨
REVIEW OF SYSTEMS
On exam initial blood pressure was 189/106 current blood pressure 148/83 with heart rate of 74
respirations 16. Heart regular rhythm. No murmurs. Arms: He does have tender areas right greater
than left under the arm. Difficult to tell if there is any erythema but obvious cellulitis sludge abscess
under the right arm which is tender.
ASSESSMENT/PLAN
1. Chest pain history of cocaine with T-wave inversions in the inferior leads. Currently he is chest pain
    free. We will check a 2-D echocardiogram. Consult Cardiology for a stress test.
2. Axillary abscesses. Consult Surgery for I and D. We will place on IV vancomycin pain control.
3. Cocaine abuse. Encouraged to quit.

Figure 1: Excerpt from a de-identified clinical narrative (cn1) written for a patient in 2007. Medical events are underlined. Enumerated events (in circles) are used as an example later in Table 1.

HISTORY PHYSICAL   DATE: 06/17/2007
NAME: Black Jack   MR#: XXX-XX-XXXX
ATTENDING PHYSICIAN: Jack Payne MD   DOB: 02/28/1960
He is a 48-year-old African American gentleman with a history of cocaine use and hypertension. He
has hidradenitis of both axilla resected. The patient is MRSA positive on IV antibiotics at the present
time. The patient's physical condition is excellent but he had MRSA in the axilla for hidradenitis that
was devastating. The wounds now are very large but he is wound vac and being changed to alginate.
Both axilla show major wounds of 20-25 cm in diameter and 4-5 cm deep in overall size and he has
excoriations on his chest from the tape. The plan is to change him from vac to alginate and see him
in a week.

Figure 2: Excerpt from another de-identified clinical narrative (cn2) for the same patient written in later in 2007. Medical events are underlined.

sequence of time-bins using Conditional Random Fields (CRFs). The learned time-bins can be used as an informative temporal feature for tasks such as fine-grained temporal ordering of medical events and medical event coreference resolution.

2 Motivation

Clinical narratives are medical reports that contain unstructured text documenting the medical history of the patient. Medical events are temporally-related concepts in clinical narratives that describe medical conditions affecting the patient's health, or tests and procedures performed on a patient. Sample excerpts from two different clinical notes (cn1 and cn2) of the same patient, generated over time, are shown in Figures 1 and 2. We can see from the examples that narrative structure moves back and forth in time and is not temporally coherent. We use cn1 and cn2 as running examples throughout the paper.

The medical events assigned to time-bins in each clinical narrative allow us to derive a coarse temporal order between medical events within and across the longitudinal medical history of the patient. Since we learn time-bins centered around admission in each narrative and we also know the admission date and perhaps the discharge dates in cn1 and cn2, we can derive a coarse partial order across the medi-

cal events in cn1 and cn2. This is shown in Figure 3. Even if the discharge dates are not known, we still know that the admission date (A1) of cn1 is 6/03/2007 and A2 of cn2 is 06/17/2007. Thus, $A2 > A1$, and all the time-bins in cn2 that are on or after admission would have happened after A2. The partially ordered time-bins can now be used for tasks such as medical concept coreference resolution.

In cross narrative coreference resolution tasks, we can prune the space of candidate pairs of medical events by ruling out portions of clinical narratives that will not have any coreferring medical events. For example, in the timeline shown in Figure 3, the medical events in time-bins *admission*, *after admission* and *discharge* of cn2 will not corefer with any medical event in cn1. Further, when mentions of the same medical events occur in different time-bins, it could mean that they are the same instance of the medical event and they corefer. For instance, *cocaine abuse* and *cocaine use*. Similarly, *MRSA positive* is assigned to time-bin *on admission* whereas *MRSA* is assigned to *before admission* and both mentions of *MRSA* corefer.

3 Related Work

The Timebank (Pustejovsky et al., 2003) corpus of annotated newswire text is widely used for temporal relation learning. The TempEval challenges have often focused on extracting different types of temporal relations from Timebank (Verhagen et al., 2009). In Timebank, events are typically verbs that denote change in state. Since the notion of an event in Timebank is different from medical events in clinical text, it is not possible to directly train models on Timebank and apply them to clinical text. The THYME work (Savova et al., 2009) extends TimeML to the
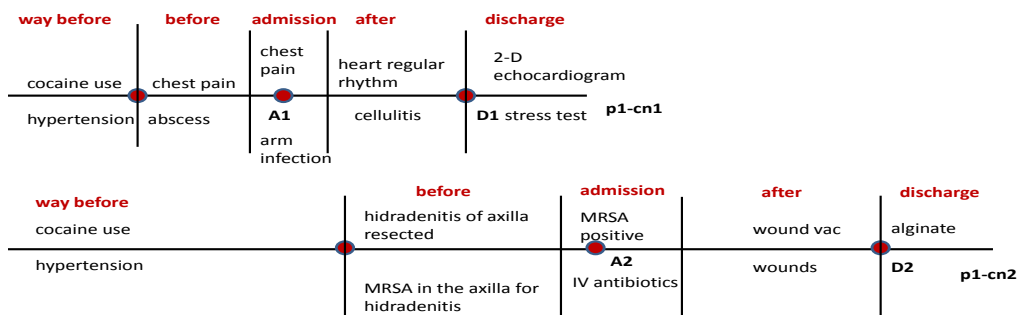
Figure 3: Medical events in clinical narratives cn1 and cn2 for patient p1 assigned to time-bins. A1 is the admission date in cn1 and D1 is the discharge date. Similarly A2 is the admission date in cn2 and D2 is the discharge date. Thus, we have, $A1 < D1, D1 < A2, A2 < D2$

medical domain to create layered annotation to be used for event linking. Boland et al. (2012) identify the temporal knowledge representation requirements of clinical eligibility criteria and develop a frame-based representation designed to support semantic annotation for temporal expressions in eligibility criteria. However, the nature of data found in eligibility criteria is different from clinical narratives.

Previous attempts at learning temporal relations between medical events in clinical text include Jung et al. (2011) and Zhou et al. (2006). Gaizauskas et al. (2006) learn the temporal relations *before, after, is_included* between events from a corpus of clinical text much like the event-event relation TLINK learning in Timebank (Pustejovsky et al., 2003). However, the corpora used in these studies are not freely available. A comprehensive survey of temporal reasoning in medical data is provided by Zhou and Hripcsak (2007).

The task addressed in this paper is at a higher level than the temporal relation learning or temporal ordering task. Without getting into fine-grained temporal ordering, we define coarse time-bins and classify medical events into one of the time-bins.

We work with a similar motivation of being able to answer clinical trial eligibility criteria with temporal constraints. However, while they model the temporal information in eligibility criteria, we process the temporal information and medical events in the clinical narrative to assign events to time-bins. The learned time-bins are a step towards fine-grained temporal ordering of medical events in clinical text. More importantly, we also demonstrate

how automatic feature extraction for this task gives us promising results, though not as good as using hand-tagged features.

4 Problem Description

A patient could have multiple clinical narratives, generated over a period of time, representing the patient's longitudinal medical history. Returning to the examples in Figures 1 and 2, in this section we describe how such clinical narratives are translated into a temporal-bin assignment problem.

4.1 Medical event representation

Medical events in clinical narratives often have a time duration with a corresponding start and stop time, for example, *history of hypertension* (Zhou et al., 2006). In this example, *hypertension* started at some point before admission and is present to the current date. Time duration based representation is essential to learning the exact fine-grained temporal order of medical events within and across clinical narratives. In order to keep the task of classifying medical events into coarse time-bins relatively simple and easy to learn, we use a time-point notation for representing medical events. Each mention of a medical event is assigned to a time-bin without taking into consideration whether it denotes the beginning or end of that event. We also do not differentiate between coreferences of the same medical event. Thus, if *chest pain* is mentioned in the past medical history and the same *chest pain* continues to persist in the after admission time-bin, the two different mentions of chest pain get anchored to dif-

ferent time-bins. Similarly, *cocaine use* started in the history of the patient and *cocaine abuse* still persists. We assign the two different mentions of this medical event into different time-bins.

4.2 Time-bins

As mentioned earlier, we learn to classify medical events into one of the following time-bins: *way before admission, before admission, on admission, after admission, after discharge*. The intuition behind each time-bin label is as follows. The time-bin *way before admission* is intended to capture all medical events that happened in the past medical history of the patient but are not mentioned as being directly related to the present illness. *Before admission* captures events that occurred before admission and are related to the present illness. *On admission* captures medical events that occur on the day of admission. *After admission* captures medical events that occur between admission and discharge (during the hospital stay or clinic visit). Finally, medical events that are supposed to occur in the future after the patient is discharged belong to the class *after discharge*.

Further, the time duration of each time-bin varies based on the patient. For instance, the hospital stay of a patient could be 4 days or 1 month or a year. This makes it very difficult to define exact time-bins based on the intuitions described above. In order to make the problem more precise and consistent across different patients, we restrict *way before admission* to events that happened more than a year ago and *before admission* to events that occurred in the same year before admission. If it is unclear as to when in the past the medical event occurred, we assume it happened *way before admission*.

5 Learning time-bin assignments

We model the problem of classifying medical events to time-bins as a sequence tagging task using CRFs (Lafferty et al., 2001). CRFs are a joint model of label sequence conditioned on the observation.

For the task proposed in this paper, an observation sequence is composed of medical events in the order in which they appear in a clinical narrative, and the state sequence is the corresponding label sequence of time-bins. Each label in the label sequence could be any one of the time-bins *way before admission (wa), before admission (ba), on admission (a), after*

admission (aa), after discharge (ad). Thus, given a sequence of medical events in narrative order we learn a corresponding label sequence of time-bins {wb, b, a, aa, ad}.

The probability of time-bin (label) sequence y, given a medical event (input) sequence x, is given by,

$$P(Y|X) = \exp \sum_i (S(x, y, i) + T(x, y, i)) \quad (1)$$

where i is the medical event index and S and T are the state and transition features respectively. State features S consider the label of a single medical event and are defined as,

$$S(x, y, i) = \sum_j \lambda_j s_j(y, x, i) \quad (2)$$

Transition features consider the mutual dependence of labels y_{i-1} and y_i (dependence between the time-bins of the current and previous medical event in the sequence) and are given by,

$$T(x, y, i) = \sum_k \mu_k t_k(y_{i-1}, y_i, x, i) \quad (3)$$

where s_j and t_k are the state and transition feature functions. Above, s_j is a state feature function, and λ_j is its associated weight and t_j is a transition function, and μ_j is its associated weight. In contrast to the state function, the transition function takes as input the current label as well as the previous label, in addition to the data. The mutual dependence between the time-bins of the current and previous medical events is observed frequently in sections of the text describing the history of the patient. Around 40% of the medical events in gold standard corpus demonstrate such dependencies.

The Maximum Entropy (MaxEnt) model (Berger et al., 1996) estimates the probability of a time-bin given the observed medical event. In this case, we are interested in finding the time-bin with the maximum estimated probability.

6 Feature Space

We extract features from medical event sequences found in each clinical narrative. The extracted feature-set captures narrative structure in terms of the narrative type, sections, section transitions, and

position in document. The medical event and the context in which it is mentioned is captured with the help of lexical features. The temporal features resolve temporal references and associate medical events with temporal expressions wherever possible.

6.1 Section-based features

Determining the document-level structure of a clinical narrative is useful in mapping medical events to time-bins. This can be achieved by identifying different sections in different types of clinical narratives and relating them to different time-bins. The section in which the medical event is mentioned tells us something about when it occurred. Li et al. (2010) train a hidden Markov model (HMM) to map a sequence of sections to 15 possible known section types in free-text narratives with high accuracy.

Commonly found sections in discharge summaries and history and physical reports include: "past medical history," "history of present illness," "findings on admission," "physical examination," "review of systems," "impression," and "assessment/plan." On the other hand, radiology notes tend to have sections describing "indication," "'comparison," "findings" and "impression". Similarly, pathology notes may have sections including "clinical history," "specimen received," "laboratory data" and "interpretation." While some sections talk about patient history, some other sections describe the patient's condition after admission, or plans after discharge. However, some clinical notes like cn2 in Figure 2 may not have any section information.

The combined feature representing the type of clinical narrative along with the section can be informative. Section transitions may also indicate a temporal pattern for medical events mentioned across those sections. For instance, "past medical history" (*way before admission*), followed by "history of present illness" (*way before admission*), followed by "findings on admission" (on admission), followed by "physical examination" (after admission), followed by "assessment/plan" (*discharge*). Medical events in different types of sections may also exhibit different temporal patterns. A "history of present illness" section may start with diseases and diagnoses 30 years ago and then proceed to talk about them in the context of a medical condition that happened few years ago and finally describe the patient's condition on admission.

In addition to the section information, we also use other features extracted from the clinical narrative structure such as the position of the medical concept in the section and in the narrative.

6.2 Lexical features

Bigrams are pairs of words that occur in close proximity to each other, and in a particular order. The bigrams preceding the medical event in the narrative can be useful in determining when it occurred. For instance, "**history of** *cocaine use and hypertension*," "**presents with** *chest pain*," "**have** *chest pain*," "**since the** *episode*," etc. If the preceding bigram contains a verb, we also extract the tense of the verb as a feature. However, tense is not always helpful in learning the time of occurrence of a medical event. Consider the following line from cn2 in Figure 2, "He has *hidradenitis of both axilla resected*." Though "has" is in present tense, the medical event has actually occurred in the history and is only being observed and noted now. Additionally, we also explicitly include the preceding bigrams and the tense of verb for the previous and next medical event as a feature for the current medical event.

Every medical event that occurs above a certain frequency threshold in all the clinical narratives of a particular patient is also represented as a binary feature. More frequent medical events tend to occur in the history of the patient, for example, *cocaine use*. We use a threshold of 3 in our experiments. The medical event frequency in also calculated in combination with other features such as the type of clinical narrative and section type.

6.3 Dictionary features

The UMLS[1] includes a large Metathesaurus of concepts and terms from many biomedical vocabularies and a lexicon that contains syntactic, morphological, and orthographic information for biomedical and common words in the English language. We map each medical event to the closest concept in the UMLS Metathesaurus and extract its semantic category. The semantic categories in UMLS include Finding, Disease or Syndrome, Therapeutic or Preventative procedure, Congenital abnormality,

---

[1]https://uts.nlm.nih.gov/home.html

and Pathologic Function. The intuition behind this is that medical events associated with certain semantic categories may be more likely to occur within certain time-bins. For instance, a medical event classified as "Congenital abnormality" may be more likely to occur *way before admission*.

6.4 Temporal features

Temporal features are derived from any explicit dates that fall in the same sentence as the medical concept. The gold-standard corpus contains annotations for temporal anchors for events. Although there are no explicit dates in cn1 and cn2, there may be narratives where there are mentions of dates such as *fever* on June 7th, 2007. In some cases, there may also be indirect references to dates, which tell us when the medical event occurred. The reference date with respect to which the indirect temporal reference is made depends on the type of note. In case of history and physical notes, the reference date is usually the admission date. For instance, *chest pain which started 2 days ago*, this would mean *chest pain* which started 2 days before admission. Since the admission date is 06/03/2007, *chest pain* would have started on 06/01/2007. Similarly, 3 to 4 months ago resolves to February 2007 or March 2007 and 2 to 3 weeks ago resolves to first or second week of May 2007. Whenever, the exact date is fuzzy, we assume the date that is farthest from the reference date as accurate. So in case of these examples, February 2007 and first week of May 2007 are assumed to be correct. We also calculate the difference between admission date and these dates associated with medical events. Another fuzzy temporal expression is "history of," where history could mean any time frame before admission. We assume that any medical event mentioned along with "history of" has occurred *way before admission*.

Other implicit temporal expressions can be found in phrases such as *upon presentation yesterday, today, at the present time,* and *now*. *Upon presentation, at the present time, today*, and *now* resolve to the admission date 06/03/2007 and *yesterday* resolves to the day before admission 06/02/2007. There are some other implicit temporal expressions expressed relative to medical events, for example, *ever since the episode 2 days ago he has felt a little weaker.* Here, *episode* refers to *chest pain* and since

chest pain happened 2 days ago, ever since then up to the present time would resolve to the time period between 06/01/2007 and 06/03/2007. This time period is associated with *weaker*.

7 Corpus

We use annotators that are students or recently graduated students from diverse clinical backgrounds with varying levels of clinical experience to annotate a corpus of clinical narratives from the medical center. The corpus consists of narratives specifically from MRSA cases and consists of admission notes, radiology and pathology reports, history and physical reports and discharge summaries. The features marked by the annotators include medical events; corresponding time-bin; corresponding UMLS concept identifier; the UMLS semantic category; temporal expressions; the link between temporal expressions and medical events, if any; and the section under which the medical event is mentioned, if any. The annotators marked 1854 medical events across 5 patients and 51 clinical narratives. The annotation agreement across our team of annotators is high; all annotators agreed on 89.5% of the events and our overall inter-annotator Cohen's kappa statistic (Conger, 1980) for medical events was 0.865.

While we found the inter-annotator agreement for medical event UMLS concept identifiers to be lower than for medical events and temporal expressions, agreement was still very high. We discovered that in many cases there was either a discrepancy in the granularity to which the medical events were coded or whether or not clinical judgment was used in selecting the concept identifier. For example, all of our annotators marked "B-Cell CLL" as an event. Three of them coded this term as "C0023434: Chronic Lymphocytic Leukemia." Two others coded this event as "C0475774: B-cell chronic lymphocytic leukemia variant." While both could be considered correct annotations for "B-Cell CLL," C0475774 is the more specific term. In another example, all of the annotators marked the phrase "white blood cell count of 10,000." For this situation, one of them selected "C0750426: white blood cell count increased," while another selected "C0023508: White Blood Cell count procedure." In contrast, the other three selected different concept

identifiers, applying clinical judgment to the medical events. One other annotator selected "C0860797: differential white blood cell count normal."

We use this gold-standard corpus for our experiments. We conduct two sets of experiments with the clinical narratives in this corpus: 1) Medical event, Time-bin experiments using hand-tagged features from the corpus and 2) Medical event, Time-bin experiments using automatically extracted features from the corpus.

8 Experiments

We first conducted experiments using the hand-tagged features in our corpus. Based on these features, we generated the section-based, lexical, dictionary and temporal features described in the previous sections. We used 10-fold cross validation in all our experiments. We use the Mallet[2] implementation of CRFs and MaxEnt. CRFs are trained by Limited-Memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) for our experiments. The per-class accuracy values of both sequence tagging using CRFs and using a MaxEnt model are indicated in Table 3.

When modeled as a multi-class classification task using MaxEnt, we get an average precision of 81.2% and average recall of 71.4% whereas using CRFs we obtain an average precision of 89.4% and average recall of 79.2%. In order to determine the utility of temporal features, we do a feature ablation study with the temporal features removed. In this case the average precision of the CRF is 79.5% and average recall is 67.2%. Similarly, when we remove the section-based features, the average precision of the CRF is 82.7% and average recall is 72.3%. The section-based features seems to impact the precision of the *on admission* and *after admission* time-bins the most.

We compare our approach for classifying medical events to time-bins with the following baseline model. We assign medical events to time-bins based on the type of narrative, any explicit dates and section in which they occur. Each section is associated with a pre-defined time-bin. In the case of the sections in cn1, any medical event under "history of present illness" is *before admission*, "review of sys-

Medical Event	Baseline	MaxEnt	CRF	Gold
①cocaine use	ba	wa	wa	wa
②hypertension	ba	wa	wa	wa
③chest pain	ba	ba	ba	ba
④episode	ba	ba	ba	ba
⑤chest pain	ba	ba	a	a
⑥infections	ba	wa	ba	ba
⑦abscess	ba	ba	ba	ba
⑧spots	ba	ba	ba	ba
⑨chest pain free	ba	wa	a	a

Table 1: Time-bin predictions by the section baseline method, MaxEnt model and CRF for a subset of medical events marked in cn1 in Figure 1.

Class(time-bin)	Section baseline	
	P	R
way before admission (wa)	56.3	61.4
before admission (ba)	60.2	57.5
on admission (a)	63.8	59.1
after admission (aa)	57.5	68.2
after discharge (ad)	52.3	55.1

Table 2: Per-class precision (P) and recall (R) for medical events, time-bins using hand-tagged extracted features.

tems" is *after admission* and "assessment/plan" is *discharge*. If the narrative has a "past medical history" or a similar section, the events mentioned under it would be assigned to *way before admission*. Partial results of (medical event, time-bin) assignment in cn2 as per this baseline can be seen in Table 1. However, this baseline does not work for clinical narratives like cn2 that do not have any section information. This model gives us an average precision of 58.02% and average recall of 60.26% across the 5 time-bins. Per-class predictions for the baseline are shown in Table 2.

The most common false positives for the *before admission* class are medical events belonging to *on admission*. This may be due to lack of temporal features to indicate that the event happened on the same day as admission. Frequently, medical events that belong to the *aa*, *ba* and *wa* time-bin get classified as *after discharge*. One of the reasons for this could be misleading section information in case of historical medical events mentioned in the assessment/plan section.

Next, we conduct experiments using automatically extracted features. This is done as follows. The medical events are extracted using MetaMap, which recognizes medical concepts and codes them using

---

[2]http://mallet.cs.umass.edu/

Class(time-bin)	MaxEnt		CRF	
	P	R	P	R
way before admission (wa)	72.4	63.5	79.8	66.7
before admission (ba)	83.4	80.8	92.0	92.4
on admission (a)	76.6	72.1	87.5	75.2
after admission (aa)	88.6	82.1	93.6	99.1
after discharge (ad)	85.2	58.7	94.3	62.5

Table 3: Per-class precision (P) and recall (R) for medical events, time-bins using hand-tagged extracted features.

UMLS (Aronson, 2001). Based on this UMLS code, we can extract the semantic category associated with the code. Compared to the 1854 medical events marked by the annotators, MetaMap identifies 1257 medical events, which are a subset of the 1854. The UMLS coding by the annotators is more contextually relevant and precise. We use a rule-based algorithm to identify and extract document structure based features such as sections from clinical narratives. The rules are formulated based on commonly occurring sections in our corpus. We extract lines that are all upper-case, and longer than a word and use their stemmed representation to sort them by frequency of occurrence in the corpus. While parsing the text in each clinical narrative, on encountering a line that matches a section title from the frequent list, all subsequent lines are associated with that title until a new section title is encountered. In case of the lexical features, we extract bigrams and calculate the tense of the verb preceding the medical event using the Stanford NLP software.[3] The temporal features are extracted with the help of TimeText developed by Zhou and Hripcsak (2007) that automatically annotates temporal expressions in clinical text. However, it is not able to capture many of the implicit temporal references. Following this, a temporal expression is linked to a medical event if it occurs in the same sentence as the medical event.

The average precision and recall of the MaxEnt model using automatically extracted features is 74.3% and 66.5% respectively. Sequence tagging using CRFs gives us an average precision and recall of 79.6% and 69.7% respectively. Although the results are not as good as using hand-tagged features, they are certainly promising. One reason for the loss in accuracy could be because the automatically calculated temporal features are not as precise as the

---
[3] http://nlp.stanford.edu/software/

Gold-standard Features		
	P	R
ME	81.2	71.4
CRF	89.4	79.2
CRF(no temp. feats)	79.5	67.2
CRF(no section feats)	82.7	72.3
Automatic Features		
	P	R
ME	74.3	66.5
CRF	79.6	69.7
Baseline (P;R)	58.02	60.26

Table 4: Overall Result Summary: Average precision (P) and recall (R) with manually annotated gold-standard features, automatically extracted features and the baseline.

hand-tagged ones. These results are summarized in Table 4.

9 Conclusion

We investigate the task of classifying medical events in clinical narratives to coarse time-bins. We describe document structure based, lexical and temporal features in clinical text and explain how these feature are useful in time-binning medical events. The extracted feature-set when used in a sequence tagging framework with CRFs gives us high accuracy when compared with a section-based baseline or a MaxEnt model. The learned time-bins can be used as an informative feature for tasks such as fine-grained ordering of medical events and medical event coreference resolution. We also experiment with hand-tagged vs. automatically extracted features for this task and observe that while automatically extracted features show promising results, they are not as good as using hand-tagged features for this task.

Acknowledgments

References

James F. Allen. 1981. An interval-based representation of temporal knowledge. In *IJCAI*, pages 221–226.

Alan R. Aronson. 2001. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc of AMIA Symposium*, pages 17–21.

Adam Berger, Stephen Della Pietra, and Vincent Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22:39–71.

Mary Regina Boland, Samson W. Tu, Simona Carini, Ida Sim, and Chunhua Weng. 2012. EliXR: An Approach to Eligibility Criteria Extraction and Representation. *Proc of AMIA Clinical Research Informatics Summit.*

Anthony J. Conger. 1980. Integration and generalization of kappas for multiple raters. In *Psychological Bulletin Vol 88(2)*, pages 322–328.

Rob Gaizauskas, Henk Harkema, Mark Hepple, and Andrea Setzer. 2006. Task-oriented extraction of temporal information: The case of clinical narratives. In *Proceedings of the Thirteenth International Symposium on Temporal Representation and Reasoning*, TIME '06, pages 188–195.

Hyuckchul Jung, James Allen, Nate Blaylock, Will de Beaumont, Lucian Galescu, and Mary Swift. 2011. Building timelines from narrative clinical records: initial results based-on deep natural language understanding. In *Proceedings of BioNLP 2011 Workshop*, BioNLP '11, pages 146–154.

John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, pages 282–289.

Ying Li, Sharon Lipsky Gorman, and Noemie Elhadad. 2010. Section classification in clinical notes using supervised hidden markov model. In *IHI*, pages 744–750.

James Pustejovsky, José M. Castaño, Robert Ingria, Roser Sauri, Robert J. Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir R. Radev. 2003. TimeML: Robust specification of event and temporal expressions in text. In *New Directions in Question Answering '03*, pages 28–34.

Guergana Savova, Steven Bethard, Will Styler, James Martin, Martha Palmer, James Masanz, and Wayne Ward. 2009. Towards temporal relation discovery from the clinical narrative. *Proc of AMIA Symposium*, pages 568–572.

Marc Verhagen, Robert J. Gaizauskas, Frank Schilder, Mark Hepple, Jessica Moszkowicz, and James Pustejovsky. 2009. The TempEval challenge: identifying temporal relations in text. *Language Resources and Evaluation*, 43(2):161–179.

Li Zhou and George Hripcsak. 2007. Temporal reasoning with medical data - a review with emphasis on medical natural language processing. *Journal of Biomedical Informatics*, pages 183–202.

Li Zhou, Genevieve B. Melton, Simon Parsons, and George Hripcsak. 2006. A temporal constraint structure for extracting temporal information from clinical narrative. *Journal of Biomedical Informatics*, pages 424–439.