TN-TIES: A System for Extracting Temporal Information from Emergency Department Triage Notes

Ann K. Irvine, Stephanie W. Haas, Ph.D., Tessa Sullivan
University of North Carolina, Chapel Hill, NC

The triage note field of the Emergency Department (ED) patient record describes the reason for the patient's visit, including specific symptoms and incidents. Here we present the Triage Note Temporal Information Extraction System (TN-TIES), which systematically processes triage note text and outputs a human and machine readable interpretation of the timing of the events leading up to the ED visit. TN-TIES consists of chunking, classification, and interpretation processing stages. The results at each stage are promising. This system is a first step towards a complete interpretation and timeline presentation of all events that occurred before a patient's visit to the ED, which could help clinicians, public health officials, and others understand and visualize the data.

I. Introduction

Identification and extraction of event descriptions is an ongoing field of investigation in natural language processing and text mining[1-4], especially for text drawn from patient records[5-9]. Identifying temporal relationships among important medical events could lead to better representation of individual patients' experiences, as well as patterns of events across patients who share some common characteristics.

In this paper, we report on a system that we have built to process temporal information found in the triage note (TN), a single field in the Emergency Department (ED) patient record. The TN is recorded at the beginning of the patient's visit to the ED, describing the reason the patient came to the ED. Figure 1 shows an example record, containing the visit timestamp, the Chief Complaint (CC) field, and the TN. The CC is a brief field, often just a noun phrase with 1 or 2 concepts. The TN provides additional detail such as the history of the present illness, symptom duration, or how an injury occurred. As in the example, the TN often includes one or more temporal references, often stated in relation to the time the note is recorded.

The long-term goal of our research is to automatically identify all events of interest to clinicians and other stakeholders in the CC and TN and place them in their correct relative positions on a timeline. Analysis of pre-triage timelines that share some characteristic, such as final diagnosis or symptom set, could lead to the identification of

```
2006-11-22 00:54:00          COUGH

Pt Has had cough and cold for last 1 1/2
weeks. Seen here last night for same. Pt
coughing and vomiting, can't sleep. Pt
has clear bil BS. No fever
```

Figure 1: Record showing timestamp, CC and TN.

symptom/event patterns, thus a better understanding of patients' experiences leading up to their decision to come to the ED. This could aid clinicians in patient treatment and also help a variety of public health functions, including biosurveillance. Furthermore, the timelines may also be useful in educating ED clinicians.

II. Background

<u>Description of dataset</u> With permission from the North Carolina Division of Public Health, we obtained 598 patient records of visits to North Carolina EDs from the North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC-DETECT: http://www.ncdetect.org/). At the time of data collection, NC DETECT gathered records from 94 hospitals in North Carolina (about 85% of the hospitals in the state) on a daily basis. The de-identified records in the corpus contain the timestamp of the patient visit, the CC, and the TN. Records were randomly chosen from 2 days of records collected in 2006 from hospitals that include the TN in their data feed. We manually annotated all events and temporal information in the records according to the annotation scheme that we developed (see below) to create a gold standard. TN text, like other clinical note genres, differs from "standard English" in its structure and vocabulary. Triage nurses rarely record complete sentences and use a variety of shorthand notations and medical terms[10]. TNs often include direct quotations from patients in addition to nurses' reports. The notes do, however, share many common phrases and, to some degree, punctuation.

<u>Annotation Scheme</u> Our system for identifying and classifying TN temporal expressions (TEs) is based upon the work that Zhou and others have done on hospital discharge summaries[6-8]. The temporal classes and their frequencies in the 598 TN sample are shown in Table 1, along with an example of each. Many of the 895 TEs in the sample display

Class	Example TE	Number of Tags	Percent of Tags
Relative Date and Time (RDT)	involved in MVC 1 week ago	362	35%
Duration (DR)	fever of 103 for 3 days	207	20%
Key Event (KE)	Pain level now: 8/10	202	19%
Date and Time (DT)	Discharged from hospital 11-17 for same	151	15%
Fuzzy Time (FT)	history of constipation	77	7%
Other Event (OE)	took two Vicodin after accident	28	3%
Recurring Time (RT)	dizzy one day every month	14	1%
Total TE Tags		1041	

Table 1: Temporal classes, example TEs, and frequencies among the 598 TN sample

characteristics of multiple classes; these are classified into more than one group. We used the Zhou work as a model both because of the similarity in their medical domains and the relatively straightforward types of temporal relationships we have encountered in TNs. We manually annotated the TEs in the 598-sample and performed several iterations of modifications to the gold standard classes and annotation rules. In the final round of annotation, two coders found a total of 1041 TE tags. Both coders tagged 945 (90.8%) of the final set of tags, and the coders agreed upon 100% of those TE tags that they both annotated. All of the tag discrepancies were for TEs with multiple tags. The discrepancies were a result of simple human errors in applying the categorization rules rather than disputed classifications, and the coders resolved them in consultation.

Although our system's output could be extended to be compliant with the TimeML standards[11], at present such a representation is much more complex than needed for our goals. The coding scheme that we have developed allows us to create interpretation rules for each type of TE that specify how to place the associated events on a timeline. Multiple tagging has aided us in creating and implementing an architecture of interpretation that contains a series of simple steps. Our system's representations of points in time and intervals over time are based upon Allen's explanations of temporal logic[12-13].

III. Methods

In order to identify and interpret temporal expressions that appear in the text of triage notes, we have developed the Triage Note Temporal Information Extraction System (TN-TIES). As shown in Figure 2, TN-TIES includes three main processing stages: a chunker, a classifier, and an interpreter. TN-TIES currently implements classifiers for the five largest TE classes and an interpreter for the largest – Relative Date and Time. Because triage note text is very different from general English text, the system is heavily tailored to the domain and the specifics of the training corpus. In developing TN-TIES, we used

80% of our 598 TN sample as training data (712 TEs) and 20% as testing data (183 TEs). The 598 records, taken over a two day period, included more than four times as many records from the first day as from the second. The training and testing sets roughly correspond to the first and second days, respectively.

Chunker TN-TIES performs a partial parse of the text of a TN in order to identify coherent phrases. Examples of individual phrases, or chunks, include single noun phrases, a noun phrase and a prepositional phrase, and a single verb phrase. However, our gold standard TE annotations are semantically, rather than syntactically, determined. The rule-based chunker uses only keywords (generally common function words) and punctuation to extract chunks. For example, the comma in the following TN text triggers the chunker to split it into two chunks: *headache since Tuesday, vomited twice*. Adding part-of-speech tags did not improve chunking performance. In accordance with the gold standard annotations, TN-TIES outputs individual TE chunks that include both an event and an explicit time reference. Non-TE chunks often consist of a single event or verb phrase. The six substantial (having multiple content words) chunks are underlined in the following sample TN. Three of the chunks contain temporal information:

pt has been having pain in her chest tonight. Mom stated her fever started tonight. Pts grandmother died yesterday and mom states pt has been crying a lot

Classifier After chunking the text of a triage note, TN-TIES employs a series of binary classifiers that predict whether or not each chunk belongs in each of the temporal classes. The classifiers use a combination of lexical features and regular expressions. About 50 keyword features, not including morphological variants, were identified through manual and automatic means. Regular expressions identify structures such as explicit date and time references. We used the TN training set and the Oracle Data Mining Software
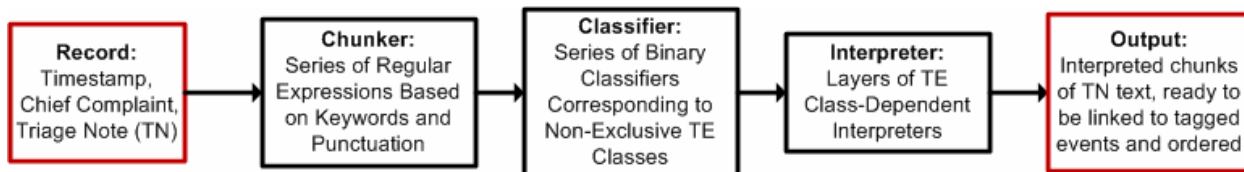
Record: Timestamp, Chief Complaint, Triage Note (TN) → **Chunker:** Series of Regular Expressions Based on Keywords and Punctuation → **Classifier:** Series of Binary Classifiers Corresponding to Non-Exclusive TE Classes → **Interpreter:** Layers of TE Class-Dependent Interpreters → **Output:** Interpreted chunks of TN text, ready to be linked to tagged events and ordered

Figure 2: TN-TIES architecture

(http://www.oracle.com/technology/products/bi/odm/odminer.html) to train Decision Tree and Naïve Bayes classifiers for each of the five major temporal classes.

Interpreter The TN-TIES interpreter consists of a series of interpretation steps. Because many chunks (16% of the training data) are classified into multiple categories, the interpreter addresses each component in a specified order. It identifies the "temporal zone" and then narrows the estimated relevant point or time interval. In example (a), the *yesterday* temporal zone is interpreted first, and then *3pm*. Some classes, especially the Duration class, demand a layer of qualitative interpretation that is a different task from identifying a single point in time. Two examples of our systematic, layered approach to interpretation follow. All interpretation actions are triggered by keywords or key phrases.

(a) *seizure 3pm yesterday*
(1) chunk classified both as a RDT and as a Date and Time (DT) TE
(2) *yesterday*, as an RDT trigger word, is interpreted relative to the ED visit timestamp and has an interval of roughly 24 hours
(3) *3pm*, as a DT trigger word, is interpreted within the bounds of the zone found in (2) – the 3pm within the 'yesterday' zone
(b) *migraine since Friday*
(1) chunk classified as both a DT and a Duration TE
(2) *Friday*, as a DT trigger word, is interpreted as the most recent Friday relative to the timestamp
(3) *since* triggers a certain duration quality, indicating that the migraine began on the Friday identified and has continued to the time indicated by the timestamp

Relative Date and Time temporal expressions are, in nearly all cases, interpreted before the classes with which they co-occur. Thus, we built the RDT interpreter first. The interpreter takes advantage of key trigger words and phrases, and temporal zones are estimated based on the visit timestamp. The

interpretation rules are derived from empirical observations of usage variation of key RDT concepts such as "yesterday" and "this morning,"[14] as well as our own intuition.

IV. Results

Chunker Table 2 presents the results of the TN-TIES chunker with regard to the TEs in the testing data. Overall, 91% of the testing set TEs were chunked perfectly. Several TEs (4.5% of gold standard chunks) were split into two pieces and several chunks included two TEs (4.5% of gold standard chunks).

	Number (chunks)	Percent (chunks)
Good chunks	162	91%
Time reference and event split in two chunks	8	4.5%
2 temporal chunks included in a single output chunk	8	4.5%
Total chunks in testing set	178	

Table 2: Chunking accuracy on testing dataset

Classifier Figures 3 and 4 show the performance of the decision tree (DT) and Naïve Bayes (NB) classifiers, respectively, for each of the five major temporal classes (listed in Table 1). In general, the DT classifiers outperformed the NB classifiers. The RDT DT classifier achieved 94% precision and 86% recall on the RDT TEs, the most frequent type of TE in the dataset. The classifiers achieve high levels of precision for the Duration, Key Event, and Date and Time classes, but recall levels are notably lower. The low recall levels are likely indicative of the limitations of the feature set. We expect that expanding the feature set to include additional words and abbreviations would increase recall.

Interpreter Interpretation is based upon the TN timestamp. In the example below, the interpreter reports that the seizure occurred sometime between 5:00 a.m. on 11/21/06 and 5:00 a.m. on 11/22/06. The RDT interpreter also includes rules for phrases that indicate concepts such as "last night", "this morning", and "today." Several trigger words and phrases may refer to a single concept. For example, the trigger phrases "this morning" and "this am" refer

to a single concept. The RDT interpreter's accuracy is dependent upon the accuracy of the rules that we constructed manually. Although our interpretation is based on manual analysis of the training set, defining temporal boundaries (e.g. between one day and the next, or between afternoon and evening) is necessarily somewhat arbitrary. Our forthcoming consensus study will inform these definitions, and any adjustments would require minimal changes to the system.

2006-11-22 09:25:00   seizure

Pts mom reports pt had seizure yesterday. Treatment PTA: none.

TN-TIES output:
  seizure yesterday:
    between 2006-11-21 05:00 and 2006-11-22 05:00

Figure 5: Example record and interpreted output of RDT TE

V. Discussion and Future Work
Overall, TN-TIES performs quite well when compared to the gold standard manual annotations. The TN Chunker performs well in processing entire triage notes. The Classifier is also quite promising, particularly for the classes most represented in the training sample and the class on which we have focused much of our feature engineering efforts. Finally, our implementation of the RDT Interpreter provides readable output estimating the time zone during which events occurred according to the TN text.

An important contribution of our work is developing a system that successfully extracts and interprets TEs from note-style text, rather than standard English. Another contribution is the persistent system architecture that we developed and implemented for an end-to-end extraction and interpretation system targeted at the specific and important domain of TNs. We have created a robust system that, as we have shown for the RDT class, should be able to identify and interpret all important TEs in any TN.

In the future, we plan to extend TN-TIES to accurately identify and interpret all TE classes. Doing so will require two efforts: (1) improving the classifier's recall performance for the remaining classes and (2) examining the data corpus and creating rules for interpreting the remaining classes. Improving classification recall performance for the other classes will require extending the feature set. We performed several iterations of adding keywords and patterns to the feature set and have observed incremental increases in performance. So far we have executed those iterations with the intention of improving the RDT class recall. We expect to see similar improvements in recall for the other classes as we continue to manually and automatically extend the classification feature set. Similarly, we have focused our work on writing interpretation rules for the TEs in the RDT class. We also expect to be able to produce appropriate, readable interpretations of the other classes.

We plan a study with domain experts (beyond those on our research team) to refine interpretation rules, e.g., defining the boundaries of temporal zones such as *yesterday*.

Our project's long term goals include creating a visual display of events leading up to a patient's visit to the ED on a timeline. These goals will require us to recognize and tag events. This work will draw on our experience in classifying TEs in preparation for interpretation.

References

1. Ahn, D. The stages of event extraction. In: Boguraev B, Numoz R, Pustejovsky J, editors. Proceedings of the Association for Computational Linguistics Workshop on Annotating and Reasoning about Time and Events; 2006 July 23; Sydney, Australia. Stroudsburg PA: ACL 2006. 1-8.
2. Han B, Gates D, Levin L. Understanding temporal expressions in emails. In Moore RC, editor. Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL; 2006, June 4-9; New York, NY. Stroudsburg PA: ACL 2006. p. 136-45.
3. Allen RB. A focus-context browser for multiple timelines. In JCDL '05, Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries, June 7-11, 2005, Denver, CO. ACM, p. 260-261.
4. Adlassnig KP, Combi C, Das AK, Keravnou ET, Pozzi G. Temporal representation and reasoning in medicine: research directions and challenges. Artif Intell Med 2006 Oct; 38(2):101-13.
5. Bramsen P, Deshpande P, Lee Y, Barzily R. Finding temporal order in discharge summaries. In AMIA Annu Symp Proc 2006; 81-5.
6. Zhou L, Friedman C, Parsons S, Hripcsak G. System architecture for temporal information extraction, representation and reasoning in clinical narrative reports. In AMIA Annu Symp Proc 2005; 869-73.
7. Zhou L, Melton G, Parsons S, Hripcsak G. A temporal constraint structure for extracting temporal

information from clinical narrative. J. Biomed Inform 2006 Aug; 39(4): 424-439.

8. Hripcsak G, Zhou L, Parsons S, et al. Modeling electronic discharge summaries as a simple temporal constraint satisfaction problem. J Am Med Inform Assoc 2005 Jan-Feb; 12(1): 55-63.

9. Chapman WW, Chu D, Dowling JN. ConText: An algorithm for identifying contextual features from clinical text. In BioNLP: Biological, translational, and clinical language processing. 2007 June; Prague, Czech Republic: ACL 2007. 81-88.

10. Travers DT, Haas SW. Using nurses' natural language entries to build a concept-oriented terminology for patients' chief complaints in the emergency department. J. Biomed Inform 2003 Aug-

Oct; 36(4-5): 260-270.

11. Hobbs JR, Pustejovsky J. Annotating and reasoning about time and events. In Proceedings of AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning; 2003 March; Stanford, CA.

12. Allen J, Hayes P. Moments and points in an interval-based temporal logic. Comput. Intell. 1989 Nov; 5(4): 225-238.

13. Allen J. Towards a General Theory of Action and Time. Artif Intell 1984; 23(2): 123-154.

14. Sullivan T, Irvine A, Haas SW. It's all relative: usage of relative temporal expressions in triage notes. In Proceedings of ASIST, 2008 Oct.; Columbus, OH: (to appear)
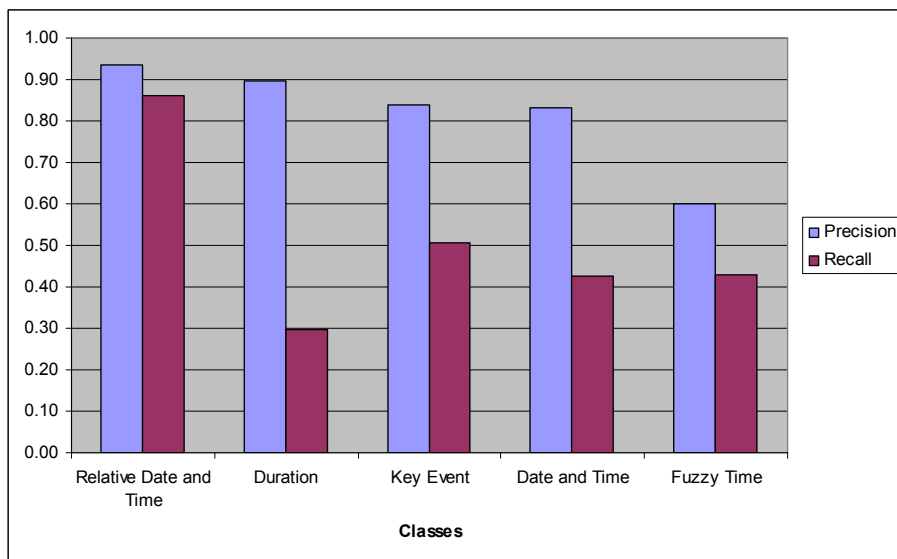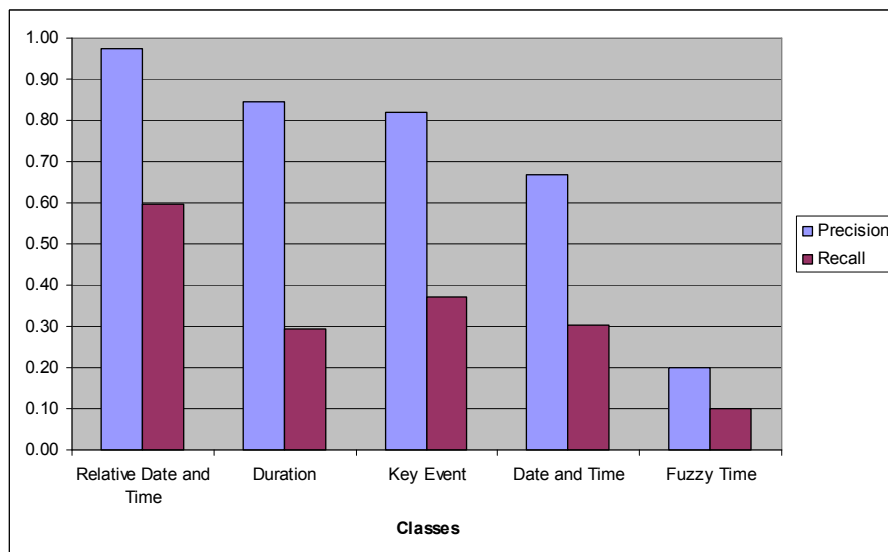
Figure 3: Decision Tree Classifier Results

Figure 4: Naïve Bayes Classifier Results