Published in final edited form as:

J Biomed Inform. 2013 December ; 46(0): S1–S4. doi:10.1016/j.jbi.2013.11.005.

Chronology of your health events: Approaches to extracting temporal relations from medical narratives

Özlem Uzuner, Amber Stubbs, and Weiyi Sun

State University of New York at Albany, Albany, New York USA

The narrative portions of medical records complement structured electronic health records and provide valuable, longitudinal health information that can guide research and clinical care. Even in a single record there is significant longitudinal medical information that outlines the sequence of clinically significant events and the medical history for a patient. Implied in the temporal sequence of events are causality and correlations, which can inform future treatments not only of that specific patient but also of patients in similar conditions.

Over the past decades, Natural Language Processing (NLP) technologies have come a long way towards interpreting the contents of narrative medical records and translating them into a structured format that lends itself to automated decision support. Systems that can extract key clinical concepts [1,2], their negation and uncertainty [3,4], and their relationships with each other [5,6] have collectively improved access to some of the most important pieces of information buried in these narratives. The increase in the availability of annotated gold-standard corpora for such systems encouraged and supported the resulting improvement [7-11]. A remaining challenge, the focus of this supplement and the topic of the 2012 i2b2 Shared-Task and Workshop on Challenges in Natural Language Processing for Clinical Data, is temporal relations, i.e., determination of the time sequence of clinically significant events presented in medical records [12-15]. We refer to this shared task as the 2012 i2b2 Challenge.

In order to foster collaboration and research regarding temporal relations in medical records, i2b2 annotated and distributed 310 gold-standard records from Partners Healthcare and the Beth Israel Deaconess Medical Center [16]. TimeML [17] and an intermediate version of the THYME¹ guidelines provided the basis for the annotations that featured:

1. **EVENTs**, which indicate clinically-relevant events such as surgeries, symptoms, and treatments. EVENTs have the following attributes: *type* (clinical concepts, clinical departments, evidentials, occurrence), polarity (positive or negative), and modality (if an event actually happened, might happen etc.).
2. **TIMEX3s**, which indicate temporal expressions such as times, dates, durations, and frequencies (e.g., “Last Monday”, “10/5/2002”, admission and discharge dates, etc.). TIMEX3 attributes: type, value (containing a normalized temporal expression), and modifier.

ouzuner@albany.edu

¹https://clear.colorado.edu/TemporalWiki/index.php/Main_Page

3. **TLINKs**, which identify temporal relations in TIMEX3/EVENT, EVENT/EVENT, and TIMEX3/TIMEX3 pairs. The TLINK's type attributes in the distributed corpus were BEFORE, AFTER, and OVERLAP.

A sample i2b2 temporal annotation (modified for print) from a report dated August 1990 is shown below:

7/13: Developed chest pain

<TIMEX3 text="7/13" type="DATE" value="1990-07-13" modifier="n/a">

<EVENT text="chest pain" type="problem" mod="ACTUAL" polarity="POS">

<TLINK from="chest pain" to "7/13" type="DURING">

The development of the gold standard was manual, involved double annotation with adjudication, and took 8 annotators 568 hours. Inter-annotator agreement (IAA) for this process prior to adjudication was 0.87 average precision and recall on EVENTS, 0.89 on TIMEX3s, 0.86 for TLINK extent matches, and 0.73 accuracy for TLINK type matches. These agreement numbers are on par with the proven TimeBank [18] temporal annotation corpus in the news domain.

The resulting 2012 i2b2 challenge corpus included 310 annotated records of 178 thousand tokens and 55 thousand TLINKs before temporal closure (355 TLINKs after temporal closure) between 31 thousand EVENTS and TIMEX3s. In order to evaluate the various approaches to extraction of temporal relations and the determination of the state of the art, i2b2 released 190 of these records to the community for system development. The resulting systems were evaluated on the remaining 120 records. The systems were evaluated in three tracks:

Track 1 - EVENT and TIMEX3 recognition

Participants, given un-annotated free text narrative medical records, developed automatic methods for the extraction of EVENTS, TIMEX3s, and their attributes.

Track 2 - TLINK creation

Given medical records with the gold standard TIMEX3s and EVENTS, the participants built systems that determined their time ordering.

Track 3 - End-to-end track

Given un-annotated medical records, participants implemented solutions for both Track 1 and Track 2.

Although the participants used a variety of techniques to address the 2012 i2b2 challenge, two major trends emerged. Most systems:

- Utilized a hybrid of machine learning (ml)-based and rule-based approaches

- Leveraged knowledge sources such as the Unified Medical Language System (UMLS) [19] and the Systematized Nomenclature of Medicine--Clinical Terms (SNOMED CT) [20].

The 2012 i2b2 Challenge was the first shared task to address temporal relations in clinical records. Outside of the clinical domain, TempEval challenges [21-23] have focused on newswire texts. The most recent TempEval challenge systems were either ml-based or rule-based; relatively few hybrid systems were built. Additionally, use of world knowledge was much less common in the TempEval systems, though participants often used other linguistic features such as parts of speech. Clinical narratives are often not as well-formed as newswire texts; they do not always adhere to formal grammar or standard narrative structures [14,15], making them difficult to interpret even for humans. Such characteristics of clinical text require automatic language processing tools to be trained and tuned to that text [13,24,25]. The difficulty in interpreting clinical narratives is evident in temporal relations.

In a new methodological review article in this issue, Sun et al [26] discuss and demonstrate with examples that 1) temporal annotation tends to be time-consuming: TIMEX3 normalization and TLINK annotation involve non-trivial logical inferences from human annotators to specify otherwise implicit or vague temporal relations in the text; and 2) there is no one correct way to annotate temporal relations in a record: the same temporal relation can be represented by very different TLINK assignments, which complicate the adjudication process. In order to create accurate annotations in a reasonable time, i2b2 simplified the annotation guidelines, adopted a flexible and user-friendly annotation tool (MAI/MAE [27]), and streamlined the annotation procedure. Only adjudicated annotations were included in the 2012 i2b2 Challenge corpus.

Track 1 - EVENT and TIMEX3 recognition

The complexity of the temporal relation extraction is reflected in the systems developed for the 2012 i2b2 Challenge. As examples, Jindal and Roth [28] and Lin et al. [29] approached EVENT and TIMEX3 recognition in two distinct ways.

Jindal and Roth designed a system that handles TIMEX3s and EVENTS independently of each other [28]. For EVENTS, they first identified EVENT candidates by filtering the output of a shallow parser, then they used a series of external resources, including MetaMap [30], the Medical Subject Headings (MeSH) [31] and SNOMED CT ontologies, and Negex output [3] as features in a series of Support Vector Machine (SVM) classifiers [32] in order to determine the type, polarity, and modality of each EVENT. Finally, the authors used a set of hand-coded rules for sentence-level inference over EVENT types. These rules reflected the observation that attributes belonging to EVENTS “which appear close to one another are sometimes closely related” and would therefore likely be of the same type; an optimization system (“Integer Quadratic Program”) determined how the rules were applied to the data.

For TIMEX3 extraction, Jindal and Roth leveraged HeidelTime [33], the best-performing temporal expression tagger from TempEval-2. The authors modified some of the output for HeidelTime, and added a series of rules to expand the system’s recognition of medical

TIMEX3s (e.g., “POD#n”, “HD”). In addition, they hand-coded rules for recognizing Admission and Discharge dates.

In contrast to Jindal and Roth’s approach of handling EVENT and TIMEX3 extraction separately, in MedTime, Lin et al. [29] used a combination of ml- and rule-based functions in a single pipeline that identified both. However, this is not to suggest that the pipeline was simple: MedTime included rules for pre-processing and cleaning data, output from the Stanford CoreNLP system [34], Wikipedia medical abbreviations and MetaMap for feature generation, HeidelTime for TIMEX3 identification, and NegEx for polarity marking. In addition, Conditional Random Fields (CRFs) [35] determined TIMEX3s and EVENTS, a rule-based system normalized TIMEX3s, and an SVM classified EVENT modality. MedTime may seem excessively complex, but it justified itself by tying for 4th place in both EVENT and TIMEX3 recognition.

Track 2 - TLINK creation

The 2012 i2b2 Challenge corpus presents two broad categories of TLINKs: 1) links that connect EVENTS and TIMEX3s to their Section Times (i.e., Admission or Discharge dates) (hereafter referred to as “SecTime TLINKs”), and 2) links that connect EVENT/EVENT, EVENT/TIMEX3, and TIMEX3/TIMEX3 pairs within and between sentences (“Sentence TLINKS”). The rich set of subtasks of the TLINK creation track allowed the 2012 i2b2 challenge participants to conceptualize the task in different ways, to tackle the various subtasks in different ways and in different orders, resulting in very different approaches that utilized existing NLP tools as much as possible. All the systems described in this issue used combinations of rules and machine learning for TLINK creation, and all but one discussed the use of additional NLP systems such as cTAKES [37] (Savova et al., 2010), Stanford CoreNLP, and discourse parsers [37].

D’Souza and Ng [38] created a hybrid system that took different approaches for each of the TLINKs types. For the SecTime TLINKs, the authors viewed the problem as a sequence labeling task and trained CRF++² as a sequence learning system. For the Sentence TLINKs, they identified four different categories of links and trained a specialized classifier for each: intra-sentence EVENT/EVENT, intra-sentence EVENT/TIMEX3, inter-sentence EVENT/EVENT (adjacent sentences), and inter-sentence co-referent EVENTS. Each specialized classifier trained on an extensive feature set that included lexical, grammatical, dependency, and semantic information as well as entity features (such as EVENT and TLINK attributes). The authors also used a set of 665 hand-coded rules to augment the performance of the classifiers, placing 5th in the Track 2 evaluation with an F-measure of 0.61.

In contrast to the four TLINK categories created by D’Souza and Ng, Nikfarjam et al. [39] divided up the TLINKs into three: 1) between-sentence TLINKs, 2) within-sentence TLINKs, and 3) SecTime/EVENT TLINKs. For between-sentence TLINKs, the authors developed a small set of highly accurate heuristic rules. For within-sentence TLINKs, they turned individual sentences into temporal graphs using rules and trained two SVMs to check

²<http://crfpp.sourceforge.net>.

the resulting graph-based TLINKs: one SVM for EVENT/EVENT links and one for TIMEX3/EVENT links. Both of these SVMs used EVENT and TIMEX3 attributes, lexical information, and dependency-based information as features. For SecTime/EVENT TLINKs, the authors trained a separate SVM, overall placing 4th (F-measure = 0.63) and achieving the highest precision out of all Track 2 participants.

Cheng et al. [40] divided the TLINKs in two: 1) between-sentence TLINKs, including SecTime/EVENT links, and 2) within-sentence TLINKs. A set of rules targeting co-referential EVENTS and SecTimes created the between-sentence TLINKs. The Maximum Entropy classifier in MALLET [35] used a wide-ranging set of features based on the output of cTAKES to generate type attributes for within-sentence TLINKs. The features included information such as the distance between two entities, tokens occurring between the entities, information about the entities heads' and the paths between them, part of speech, tense of relevant verbs, and semantic type.

In their TEMPTing (TEMPoral relaTion extractING) system, Chang et al. [41] applied a rule-based algorithm and an ml-based algorithm in parallel, then merged the results from the two. The rule-based algorithm categorized TLINKs into three types, and applied three sets of corresponding rules: one for intra-sentence TLINKs, one for inter-sentence TLINKs, and one for SecTime TLINKs. In contrast, the ml-based algorithm used a binary classifier to remove candidate TLINK pairs that were not likely to be connected, then used a multi-class classifier to determine the type of the TLINK that connected the remaining pairs. In the end, the output from the rule-based and the ml-based algorithms were integrated using a rule set that, broadly, favored the pairs from the rule-based system but the type attributes from the ml-based system. TEMPTing ranked 3rd in Track 2, with an F-measure of 0.68.

Summary

Participants in the 2012 i2b2 Shared-Task and Workshop on Challenges in Natural Language Processing for Clinical Data created a variety of systems for processing temporal relations in clinical records. The different ways of conceptualizing the shared-task Tracks reflects the complexity of temporal analysis of narratives even for humans, and the use of hybrid systems, world knowledge, and other sources of linguistic information reflect the difficulty of formulating temporal analysis for automated methods. Despite their promising results, and significant advancement on the state of the art in temporal relations in medical records, the 2012 i2b2 challenge systems only scratched the surface in this task. Open questions remain about the applicability of the developed systems for real life practical questions, such as the determination of the progression of diseases in patients, for example, heart disease in diabetic populations. Nevertheless, the 2012 i2b2 Challenge corpus of temporal annotations remains a valuable asset to the medical NLP community, and we hope will serve as the basis for further innovation in temporal relations, resulting in systems that can be applied to real life clinical problems.

Works cited

- [1]. de Bruijn B, Cherry C, Kiritchenko S, Martin J, Zhu X. Machine-learned solutions for three stages of clinical information extraction: the state of the art at i2b2 2010. *J Amer Med Inform Assoc.* 2011; 18:557–562. [PubMed: 21565856]
- [2]. Jiang M, Chen Y, Liu M, Rosenbloom ST, Mani S, Denny JC, Xu H. A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries. *J Amer Med Inform Assoc.* 2011; 18:601–606. [PubMed: 21508414]
- [3]. Chapman W, Bridewell W, Hanbury P, Cooper G, Buchanan B. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform.* 2001; 34(5):301–10. [PubMed: 12123149]
- [4]. Chapman WW, Chu D, Dowling JN. ConText: An algorithm for identifying contextual features from clinical text. *BioNLP 2007: Biological, Translational, and Clinical Language Processing.* 2007:81–88.
- [5]. Xu Y, Liu J, Wu J, Wang Y, Tu Z, Sun J-T, Tsujii J, Chang E. A classification approach to co-reference in discharge summaries: 2011 i2b2 Challenge. *J Am Med Inform Assoc.* 2012; 19:897–905. [PubMed: 22505762]
- [6]. Chowdhury MFM, Zweigenbaum P. A controlled greedy supervised approach for co-reference resolution on clinical text. *J Biomed Inform.* 2013; 46:506–515. [PubMed: 23562650]
- [7]. Uzuner O, Goldstein I, Luo Y, Kohane I. Identifying patient smoking status from medical discharge records. *J Amer Med Inform Assoc.* 2007; 15:14–24. [PubMed: 17947624]
- [8]. Uzuner O, Solti I, Cadag E. Extracting medication information from clinical text. *J Amer Med Inform Assoc.* 2010; 17:514–8. [PubMed: 20819854]
- [9]. Uzuner O, Solti I, Xia F, Cadag E. Community annotation experiment for ground truth generation for the i2b2 medication challenge. *J Amer Med Inform Assoc.* 2010; 17:519–23. [PubMed: 20819855]
- [10]. Uzuner Ö, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *J Amer Med Inform Assoc.* 2011; 18:552–556. [PubMed: 21685143]
- [11]. Uzuner O, Bodnari A, Shen S, Forbush T, Pestian J, South BR. Evaluating the state of the art in co-reference resolution for electronic medical records. *J Amer Med Inform Assoc.* 2012; 19(5): 786–91. [PubMed: 22366294]
- [12]. Hripcsak G, Zhou L, Parsons S, Das AK, Johnson SB. Modeling electronic discharge summaries as a simple temporal constraint satisfaction problem. *J Amer Med Inform Assoc.* 2005; 12(1):55–63. [PubMed: 15492038]
- [13]. Zhou L, Melton GB, Parsons S, Hripcsak G. A temporal constraint structure for extracting temporal information from clinical narrative. *J Biomed Inform.* 2006; 39(4):424–39. [PubMed: 16169282]
- [14]. Irvine, AK.; Haas, SW.; Sullivan, T. TN-TIES: A system for extracting temporal information from emergency department triage notes; *Proceedings of the AMIA Annual Symposium*; 2008; p. 328-32.
- [15]. Li, M.; Patrick, J. Extracting Temporal Information from Electronic Patient Records; *Proceedings of the AMIA Annual Symposium*; 2012; p. 542-551.
- [16]. Sun W, Rumshisky A, Uzuner O. Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. *J Amer Med Inform Assoc.* 2013; 20:806–813. [PubMed: 23564629]
- [17]. Sauri, R.; Littman, J.; Knippen, B., et al. TimeML annotation guidelines, V.1.2.1. 2005. Available at: <http://www.timeml.org/timeMLdocs/AnnGuide14.pdf>
- [18]. Pustejovsky J, Hanks P, Sauri R, See A, Gaizauskas R, Setzer A, Radev D, Sundheim B, Day D, Ferro L, Lazo M. The TIMEBANK Corpus. *Proceedings of Corpus Linguistics.* 2003:647–656. 2003.
- [19]. Bodenreider O. The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Research.* 2004; 32:D267–D270. [PubMed: 14681409]
- [20]. Cornet R, de Keizer N. Forty years of SNOMED: A literature review. *BMC Medical Informatics and Decision Making.* 2008; 8(Suppl 1):S2. [PubMed: 19007439]

- [21]. Verhagen, M.; Gaizauskas, R.; Schilder, F.; Hepple, M.; Katz, G.; Pustejovsky, J. SemEval-2007 Task 15: TempEval Temporal Relation Identification; Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007); 2007; p. 75-80. Association for Computational Linguistics
- [22]. Verhagen, M.; Sauri, R.; Caselli, T.; Pustejovsky, J. SemEval-2010 Task 13: TempEval-2; Proceedings of the 5th International Workshop on Semantic Evaluation; 2010; p. 57-62. Association for Computational Linguistics
- [23]. UzZaman, N.; Llorens, H.; Derczynski, L.; Allen, J.; Verhagen, M.; Pustejovsky, J. SemEval-2013 Task 1: TempEval-3: Evaluating Time Expressions, Events, and Temporal Relations. Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013); 2013; p. 1-9. Association for Computational Linguistics
- [24]. Savova, G.; Bethard, S.; Styler, W.; Martin, J.; Palmer, M.; Masanz, J.; Ward, W. Towards temporal relation discovery from the clinical narrative; Proceedings of the AMIA Annual Symposium; 2009; p. 568-72.
- [25]. Ong, FR. Master's thesis. Vanderbilt University; 2009. The Tarsqi Toolkits Recognition of Temporal Expressions within Medical Documents.
- [26]. Sun W, Rumshisky A, Uzuner O. Annotating Temporal Information in Clinical Narratives. J Biomed Inform. this issue.
- [27]. Stubbs, A. MAE and MAI: Lightweight Annotation and Adjudication Tools; Proceedings of the Linguistic Annotation Workshop V; Portland, Oregon. July 23-24, 2011; Association of Computational Linguistics
- [28]. Jindal P, Roth D. Extraction of Events and Temporal Expressions from Clinical Narratives. J Biomed Inform. this issue.
- [29]. Lin Y-K, Chen H, Brown RA. MedTime: A Temporal Information Extraction System for Clinical Narratives. J Biomed Inform. this issue.
- [30]. Aronson AR, Lang F-M. An overview of MetaMap: historical perspective and recent advances. J Amer Med Inform Assoc. 2010; 17(3):229–236. [PubMed: 20442139]
- [31]. Nelson, Stuart J.; Zeng, Kelly; Kilbourne, John; Powell, Tammy; Moore, Robin. Normalized Names for Clinical Drugs—RxNorm at 6 years. J Amer Med Inform Assoc. 2011; 18(4):441–8. [PubMed: 21515544]
- [32]. Chang CC, Lin CJ. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology. 2011; 2:1–27.
- [33]. Strötgen, J.; Gertz, M. Language Resources and Evaluation. Vol. 47. Springer; 2013. Multilingual and Cross-domain Temporal Tagging; p. 269-298.
- [34]. Klein, D.; Manning, CD. Accurate unlexicalized parsing; Proceedings of the 41st annual Meeting of the Association for Computational Linguistics; 2003; p. 423-430.
- [35]. McCallum, A. Mallet: a machine learning for language toolkit. 2002. <http://mallet.cs.umass.edu>
- [36]. Savova G, Masanz J, Ogren P, Zheng J, Sohn S, Kipper-Schuler K, Chute CG. Mayo Clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, component evaluation and applications. J Amer Med Inform Assoc. 2010; 17:507–13. [PubMed: 20819853]
- [37]. Lin Z, Ng HT, Kan M-Y. A PDTB-styled end-to-end discourse parser. Natural Language Engineering. 2012:1–34. FirstView.
- [38]. D'Souza J, Ng V. Classifying Temporal Relations in Clinical Data: A Hybrid, Knowledge-Rich Approach. J Biomed Inform. this issue.
- [39]. Nikfarjam A, Emadzadeh E, Gonzalez G. Towards Generating Patient's Timeline: Extracting Temporal Relationships from Clinical Notes. J Biomed Inform. this issue.
- [40]. Cheng Y, Anick P, Hong P, Xue N. Temporal Relation Discovery between Events and Temporal Expressions Identified in Clinical Narrative. J Biomed Inform. this issue.
- [41]. Chang Y-C, Dai H-J, Wu JC-Y, Chen J-M, Tsai RT-H, Hsu W-L. TEMPTing System: A Hybrid Method of Rule and Machine Learning for Temporal Relation Extraction in Patient Discharge Summaries. J Biomed Inform. this issue.