23rd International Conference of the European Federation for Medical Informatics
User Centred Networked Health Care - A. Moen et al. (Eds.)
MIE 2011 / CD / Posters

Empirical Approach to Event Sequencing in Automatic Analysis of Patient Records

Dimitar TCHARAKTCHIEV [a], Svetla BOYTCHEVA [b, c], Galia ANGELOVA [b]
[a] *University Specialized Hospital for Active Treatment of Endocrinology (USHATE), Medical University, Sofia, Bulgaria, dimitardt@gmail.com*
[b] *Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria*
[c] *University of Library Studies and IT, Sofia, Bulgaria*

Abstract. Patient Records (PRs) convey important patient-related findings. The automatic analysis of the PR free text descriptions is a challenging research task. This poster presents an empirical strategy to structure temporal information in hospital PRs in Bulgarian language. Temporal markers are defined and the case history is structured accordingly. Current evaluation results are presented as well.

Keywords. Natural Language Processing, Patient Records, Extraction of Temporal Information

1. Background and Objectives

Time is a fundamental entity in clinical narratives but automatic analysis of temporal information is a relatively recent activity in biomedical text processing [1]. There are rare attempts to build formal models of time, which process clinical narratives; this prevents reasoning about temporal relations between events discussed in the text. Recent results include *(i)* a suggestion for an annotation scheme of temporal relations in clinical narratives [2]; *(ii)* an algorithm which infers status of a condition with regard to four properties from simple lexical clues occurring in the context of the condition. The properties are *negated*, *hypothetical*, *historical*, or experienced by someone *other* than the patient. The accuracy is more than 97% for *negated* condition and more than 73% for *historical* condition [3]; *(iii)* evaluation of TimeText, a temporal reasoning system, which generates temporal relations about endpoints (start or finish) of pairs of medical events. The system captured 79% of 307 temporal relations determined to be clinically important by the raters and answered correctly to 84% of the temporal questions [4].

Our approach for recognition of temporal markers exploits the structure of the hospital PRs in Bulgaria. In addition we rely on the fact that the University Hospital USHATE treats the most complex cases from allover the country and all PRs (including discharge letters) contain a human-written summary of the case history in the PR section 'Anamnesis'. The PRs contain a relatively compact, sketchy abstracts where the major illness phases are discussed together with treatment and medication changes. These fragments form our training corpus for extracting temporal markers and developing strategies of how to order the respective events.

2. Methods and Results

In medicine, an episode comprises all activities that are performed between diagnosis of a disease and its cure; the patient-related documentation reflects this default fragmentation of healthcare tasks [5]. Studying various approaches to determine and annotate the granularity of PR temporal intervals, when important clinical events occur, we consider *episodes* as *sets of events defined via the explicit temporal markers uttered by the physicians* who examine and treat patients. We believe that human experts declare explicitly the most important temporal markers to adequately communicate the case history to another medical doctor. Therefore, we consider the markers as primary signals for diseases progression phases. Our model is framed using three tags suggested in [6]: *(i) reference point*; *(ii) direction*; and *(iii) temporal expression* plus additional tags needed for our project: *(iv) diagnoses, complains or symptoms* (i.e. what happens, occurs or is found during the episode); *(v) episode end point*; *(vi) drugs/treatment applied during the episode*; and *(vii) medication effect*. There could be several diagnoses or symptoms enumerated in one episode as well as more than one drug correspondingly prescribed to the patient.

We have developed a prototype for automatic recognition of temporal markers and sequencing episodes by calculation of actual dates or periods for relative temporal clauses, e.g. ′3 years ago′. The prototype integrates components for high-accuracy extraction of drugs and diagnoses [7]. We evaluated manually current temporal information processing. 7149 temporal markers were identified in 1374 anonymised discharge letters of diabetic patients. The percentage recall in recognition of temporal markers in about 57% and precision is 84%. Mistakes are most often due to *abbreviations*, sophisticated *prepositional phrases* for marking start, duration, cycle or interval, *ambiguity*, *references* to previously introduced moments of time or to multiple moments in one token etc. The temporal markers are identified by an empirically-elaborated context-free grammar, which is under incremental development.

Acknowledgements. The on-going work presented in this poster is supported by grant DO 02-292 "Effective search of conceptual information with applications in medical informatics", funded by the Bulgarian National Science Fund in 2009-2012.

References

[1] Zhou L. and G. Hripcsak. Temporal reasoning with medical data - a review with emphasis on medical natural language processing. *Journal of Biomedical Informatics* 2007, 40(2), pp. 183-202.

[2] Savova, G., S. Bethard, W. Styler, J. Martin, M. Palmer, J. Masanz, and W. Ward. Towards Temporal Relation Discovery from the Clinical Narrative. In *Proc. AMIA Annual Symposium 2009*, pp. 568–572.

[3] Harkema, H., J. Dowling, T. Thornblade, and W. Chapman. Context: An Algorithm for Determining Negation, Experiencer, and Temporal Status from Clinical Reports. J Biomed Inf. 2009, 42(5): 839–851.

[4] Zhou L., S. Parsons, and G. Hripcsak. The evaluation of a temporal reasoning system in processing clinical discharge summaries. *Journal of AMIA*, 2008, 15(1), pp. 99-106.

[5] Tcharaktchiev, D. Hospital Information Systems. *Sofia, Kama*, 2003 (in Bulgarian).

[6] Hyun S., S. Bakken and S.B. Johnson. Markup of temporal information in electronic health records. In *Stud. Health Technologies and Informatics* Vol. 122, 2006, pp. 907-908.

[7] Tcharaktchiev, D., G. Angelova, S. Boytcheva, Z. Angelov and S. Zacharieva. Completion of Structured Patient Descriptions by Semantic Mining. To appear in the *Proc. 2nd PSIP Int. Workshop 'Patient Safety through Intelligent Procedures in Medication',* Paris, May 16-17 2011, to be published by IOS Press.