Relationship identification model: how do medical concepts relate in patient records?

Rezarta Islamaj Doğan, PhD, Aurélie Névéol, PhD, Zhiyong Lu, PhD
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD

Abstract

Automatic recognition of clinically important concepts and their inherent relationships are preliminary steps for many applications in medical informatics. Our approach combines a statistical model, natural language processing features and conditional random fields for concept recognition, as well as a context-blocks scheme and support vector machines for relationship identification. In addition to serving as a benchmark for comparison, this system may also serve as a preliminary step for other discovery tasks in medical informatics.

The extraction of medical concepts and their inherent relationships are primary tasks for any processing system of patient records. In this work, we present an end-to-end system that can be used to extract concepts such as medical problems, treatments and tests, and furthermore identify different relationships between them [1]. We represent a relationship with a scheme of five distinct context-blocks determined by the position of concepts in the text, as illustrated in Figure 1. We learned various features for this model and built support vector machine classifiers for each relationship. For experiments, our system was trained and evaluated on the 4th i2b2 challenge set of 826 patient records. Our results include F-measures of 0.870 and 0.774 for exact concept detection and relationship extraction respectively. Compared with the best concept detection results obtained during the challenge, our method achieved a significant improvement in performance. Most importantly, we report robust results of our relationship extraction system when using automatically extracted concepts. This system may also serve as a preliminary step for other discovery tasks in medical informatics.

Figure 1. Relationship representation between two concepts as five context-blocks. The phrases: a partial biliary obstruction and magnetic resonance cholangiopancreatography correspond to concept categories “medical problem” and “medical test” respectively. The represented relationship is “medical test is requested to investigate medical problem”.

Acknowledgements

Funding: This research was supported by the Intramural Research Program of the NIH, National Library of Medicine.

References