MEDINFO 2004
M. Fieschi et al. (Eds)
Amsterdam: IOS Press

Automatic Summarization of Patient Discharge Summaries to Create Problem Lists using Medical Language Processing

Hui Cao, MD, MS, Michael F. Chiang, MD, James J. Cimino, MD, Carol Friedman, Ph.D., George Hripcsak, MD, MS

Department of Biomedical Informatics; Columbia University, New York, NY, USA

Abstract: The recognition of an "information overload" problem has stimulated research involving automated text summarization. We describe a system combining traditional statistical methods with natural language technology to automatically generate a patient problem list based on clinical discharge summaries dictated by physicians. Our evaluation shows that this system has moderate recall (69%) and precision (61%). It captured over 95% of the diagnoses, and over 90% of the symptoms and findings associated with the diagnoses for each patient.

Introduction: When a patient is admitted, physicians always have to spend a significant amount of time reading through details of the medical record that may or may not be relevant to this admission. Physicians often do not have enough time to read all these details, but must make critical decisions in a timely manner and would benefit from accurate summaries of patient medical records. To date, this problem has received relatively little attention from researchers. In this study, we describe a novel method for generating patient problem lists from text-based discharge summaries. Our system uses robust natural language processing technology to facilitate a classical statistical approach in weighting findings that comprise the problem list. The input to our system is a set of discharge summaries and the task for the system is to summarize the medical problems of the patient. The output is a list of medical problems that the patient has.

Methodology: The Medical Language Extraction and Encoding System (MedLEE) is a medical language processing system in real use at Columbia Presbyterian Medical Center. For each patient, all the discharge summaries are obtained, parsed by MedLEE and then transformed to text knowledge representation structures in XML format. The system takes XML as input, extracts all the findings that belong to pre-selected semantic types, and weights the findings based on frequency and semantic type. For refining the result, we used a pragmatic filter in extraction stage and a controlled vocabulary built using UMLS and manual review in post-extraction processing stage.

Evaluation: We evaluated our currently finished work to examine performance of the system and to identify existing problems. A set of cases (n=9) was randomly selected from among all patients admitted to New York Presbyterian Hospital in 2000. We used a reference standard generated by experts through Delphi method as a "gold standard". The lists generated by our system were then compared with the reference standard. Each discrete finding was classified either as: (1) "True positive", (2) "False positive". Two metrics, recall and precision, were used to assess system performance for every case.

Results: Recalls and precisions of the problem lists generated by the system for nine cases ranged from 55% to 80%. The average recall and precision were 61.9% and 60.9%, respectively. Our system captured over 95% of the diagnoses, and over 90% of the symptoms and findings associated with the diagnoses for each patient.

Discussion: There are mainly two reasons that the recall was moderate. The first reason is that MedLEE does not define a number of medical phrases. Consequently, our system failed to catch them. The second reason is that the semantic classes were pre-selected, which limits the information to be gathered from the analyzed text. The precision of 61% does not mean that the system produced wrong findings 39% of the time. Most of findings that were in the system response but not in the reference were true findings. They were not included in the reference generated by experts due to the following reasons: (1) the finding was very common, hence clinical uninformative or (2) the finding could be easily explained by the patient's underlying diseases and didn't have to appear in the problem list.

Three interesting features of the experts' responses were noted. The first is that diseases, symptoms, signs and medical procedures are well modeled in the experts' responses. Experts intend to list the major underlying problems first and then organize the other problems as the subset of underlying problem. The second is that experts distinguish between transient problems and ongoing problems. The third is that experts order the findings according to the severity and the importance to patient care. To generate a clinically useful problem list, further work will be required.

Conclusion: We have introduced a system for generating problem lists based on discharge summaries, which has potential to automatically generate summaries of medical records.

Acknowledgement: This study was funded by National Library of Medicine grant R01 LM06910 "Discovering and applying knowledge in clinical databases."