Towards Chronological Summary of Medical Records

Hee Kyong Park¹, Jinwook Choi, MD, PhD¹
¹Department of Biomedical Engineering, Seoul National University, Seoul, South Korea

Abstract

We developed a semantic tool (MedinfoParser) that provides a chronological summary from medical records. It consists of a Korean parser, an English parser, a numeric detector, and a noun phrase detector. It provides Part-Of-Speech (POS) tags, and extracts medical concepts and numeric concepts. In order to identify medical terms we developed a module for connecting UMLS knowledge server. A rule parser and rule patterns were also designed for the numeric term detection.

Introduction

In the era of the electronic medical record (EMR), a huge amount of electronic medical documents are produced everyday. Most clinical notes are still written in free text. In order to catch the patient’s clinical status in a short period of time, we designed MedinfoParser, a system for making a summary of medical records in a chronological order.

Methods

MedinfoParser parses medical documents by annotating POS tags, understanding semantics of medical terms and dates relevant to each event. It consists of a Korean parser, an English parser, a numeric detector, and a noun phrase detector (Figure 1).

![Figure 1. A class diagram of MedinfoParser](image)

The Korean parser tokenizes Korean medical documents and puts POS tags after the each parsed word. We have modified and used KLT parser¹ for the tokenizing and parsing process.

In general, Korean medical records are written in Korean and English mixed. However, most medical expressions are written in English. To parse those key medical expressions, for every alphabetic word, English parser identifies medical concepts by sending queries to UMLS knowledge source server. If it gets UMLS CUIs, then it puts a ‘MED’ tag after each medical term. As medical documents contain a lot of abbreviations and locally defined terms which are not searched through UMLS, we made the English parser use user dictionaries.

As numeric expressions contain very important meaning in medical records, we have developed a numeric detector that extracts numeric terms and finds clinical events which are relevant to the terms. We designed numeric concept detection rules based on data frame². Each data frame contains functions, other concept classes, relations, expressions to extract single numeric concept. We defined seventeen numeric concepts that are critical in medical field, and other concept classes needed for the numeric concept extraction. And we have been developing the rule patterns.

The noun phrase detector detects boundaries of the medical noun phrases which are consisted of more than two words and tags as ‘NP’. We made principles to clearly identify boundaries of ambiguous cases. About one hundred noun phrases were extracted from the discharge summaries of Seoul National University Hospital.

Conclusion

We have developed a semantic tool that extracts key medical concepts from Korean medical documents. The MedinfoParser generates lexical and semantic information needed for extracting key medical information. The information produced by the parser will be useful in other semantic applications in the medical field. For the future work, we plan to develop a visualization module for showing the chronological summary.

References

1. NLP laboratory of Kookmin University. http://nlp.kookmin.ac.kr/