Title	
Author(s)	Moharasan, Gandhimathi
Citation	
Issue Date	2019-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/15784
Rights	
Description	Supervisor: Dam Hieu Chi, ,

EXTRACTION OF TEMPORAL INFORMATION FROM CLINICAL NARRATIVES

Abstract

Keywords: Temporal information extraction, electronic medical records, clinical narratives, conditional Random fields, semi-supervised learning, Naive Bayes classifier.

The Electronic Medical Records (EMRs), is a digital version of patient medical history written and stored by medical professionals at the hospital. It has many components such clinical narratives, radiology reports, laboratory results and etc. Among these components, text data describing medical observation of patients by doctor's and nurse's known as clinical narratives and it covers most significant information about patient health status. EMR clinical narratives consists with large unstructured data become promising resources to support advancement and enhancement of clinical studies as it obtained by doctor's and nurse's with their medical expertise and observing on each individual patient for one medical practice during treatment activities. One perspective of the exploiting the potential clinical narratives in EMRs is clinical decision support systems, can vary in the areas include prognosis, disease monitoring, adverse drug effects and drug development, etc.

In spite of having many advantages, EMR clinical narratives remain with many challenges for exploitation. One major problem is structured representation of longitudinal clinical narrative of patients for further utilization in medical care. As all the significant medical information of patient's have noted with the notion of time, temporal reasoning plays a key role in structured representation. Temporal reasoning is a fundamental, yet vital skill that requires understanding the natural language text. Therefore temporal reasoning is key task for temporal information extraction. Temporal information in clinical text perform the crucial role in interpretation of the patient clinical information such as progress of disease, frequencies of medication information, to detect treatment pattern and adverse drug events.

The dissertation studies for basic steps on reconstructing clinical narratives to structured representation. In other words, the thesis studies to propose methods that most appropriately extract temporal information from clinical narratives by utilizing annotated and unannotated data. To this end, the thesis systematically approaches the three fundamental problems: extraction of implicit and explicit temporal expressions, extraction of temporal events by using annotated data along with the support of unannotated data and detection of temporal relation and classification. Our main targets are to seek provably stable and reliable models that can effectively extracts information from clinical narratives.

The first contribution comes from temporal expression extraction. A novel feature set has been proposed to address the problem of temporal expressions extraction. Our proposed framework has following key theoretical properties: (1) new proposed feature set is obtained from raw clinical text and (2) adopted HeidelTime features that are appropriate

for temporal expressions extraction from clinical narratives. Existing methods are either having the advantage of HeidelTime or developing rules/ machine learning models, but not the integrated components of both. Hence the integrated properties help to extract temporal expressions effectively.

The second contribution is stemming from temporal event extraction in clinical narratives. The introduction of a novel semi-supervised framework to exploit abundant unannotated data for extracting temporal events from clinical narratives.

To best of our knowledge and survey from literature, this work is the first to propose semi-supervised method for extracting temporal event from clinical text. This approach innovated the novel idea of gradually extending the training corpus by adding annotated data obtaining from unannotated clinical narratives. When working with very high dimensional medical data, our proposed method effectively extracts temporal events. The main result of this study is a novel semi-supervised method that can reach state-of-art performance with stable improvement than existing methods.

The third contribution is from temporal relation identification and classification. We formulated a new assumption on generating and identifying the potential candidate pairs from list of temporal events or expressions that can appropriately relate events/expressions in clinical narratives based on their attributes. Moreover, to address the problem of temporal relation detection, we exploited Naive Bayesian Classifier to detect the temporal relationship among the identified pair's. The effective candidate pair's generation helps to improve the relation classification performance.

In conclusion, our proposed methods with novel feature sets can effective extract temporal expression in clinical narratives. A proposed novel semi-supervised framework for temporal events extraction successfully utilized unannotated clinical narratives along with annotated data and enhanced event extraction performance. In case of temporal relation detection, novel assumption for candidate generation pairs along with adopted dependency parsing approach can improve the quality of candidate pairs and consequently temporal relation classification with Naive Bayesian classification. Finally, this study accomplishes our objectives prosperously as stated.