Temporal taggings of Entities in Text

Haozhou Pang
Department of Computing Science
University of Alberta
haozhou@ualberta.ca

Yourui Guo
Department of Computing Science
University of Alberta
yourui@ualberta.ca

Abstract

TODO

1 Introduction

Entities in text can be related to some time points or intervals mentioned in an article, and the demand of knowing temporal expressions that are significant to entities in articles is increasing over the years. For example, utilizing the temporal expressions can help us to look for the named entity it truly refers to. When we read a newspaper about Bush, we can interpret the named entity as George W. Bush knowing his birthdate being July 6, 1946 mentioned in the article. Also, extracting temporal expressions related to entities helps us to construct knowledge bases. As we know that most of knowledge bases contain a small quantity of temporal taggings for named entities. Thus, knowledge bases can be expanded by adding more related temporal taggings detected in the plain text.

Temporal tagging detection is the process of looking for expressions in text that refers to time points or time intervals. Few works have studied on detecting temporal taggings [1, 2]. SUTime [1] is a library for recognizing and normalizing time expressions. It transforms time expressions like October 1963 to the normalized value of 1963-10 and type of DATE. Normalizing temporal expressions reduces the chance of mis-recognizing the same date value to different ones, and the result of detecting the pairs of entity and temporal tagging can be more accurate as well.

The concept of knowledge base is addressed to solve this problem where individual entities and relationships among them are stored in a large repository. The problem then arises: can we find out a method to detect entities and temporal taggings, along with observe their relations in plain text?

Our task could be formulated in such way: given a set of articles with entities and meaningful temporal expressions, we first pre-process the text and extract the candidate entities and temporal expressions, then we calculate the importance score by using a series of features for each entity - temporal expression pair so that the output will be a list of ranked entity - temporal expression pairs. In this way, we can extract the most important temporal information along with their corresponding entities from a article (plain text) without assistance of any external resources.

This task is important as it can be applied to many other fields in Information Retrieval, for example, our model can be used to enrich the content of a existing knowledge base by providing temporal information that was taken from any source of text.

2 Related Work

TODO

3 Method

The section presents information about how we approached the current task with more detailed description about the data set.

3.1 Data

TODO

3.2 Lexical Feature

TODO

3.3 Syntactic Feature

TODO

3.4 Evaluation

TODO

4 Result

The section present the results of current task followed by detailed error analysis. TODO

5 Conclusion

TODO

Acknowledgments

First and far most, we would like to thank prof. Davood Rafiei for the advice and help on this project, also special thank to Michael Su for helpful discussions.