

Using Language Models for Improving Sentence Retrieval in TAC KBP

Ilyas CELIK

Free University of Bozen-Bolzano
University of Saarland

Abstract

Sentence retrieval is one of the most important components of slot filling task which has crucial effects on overall results of TAC KBP. Defining a well-constructed slot filling query, brevity of the sentences and use of co-references are some of major challenges one has to cope with in sentence retrieval in order to retrieve well-ranked relevant sentences in response to the queries.

In this study, we took two datasets that are built from TAC KBP 2011 data collection. Initially, we analyzed the performance of the document retrieval for both datasets. After we verified the performance of the document retrieval for each of the datasets we focused on the sentence retrieval to improve the sentence rankings and the number of the relevant sentences retrieved per queries. We have used triggering language modeling as our baseline and we applied language modeling with four different smoothing techniques (Absolute discounting, Bayesian smoothing with Dirichlet priors, Linear interpolation and Log Linear Interpolation) to get a better ranked sentences. Using the dataset-2, which gave the best results in document retrieval, we observed that using Dirichlet language modeling outperforms on our baseline model in sentence retrieval task. We obtained a 10.9% improvement in mean average precision (MAP) value.

To boost the performance of the best performing language model, we proposed a new approach that requires defining relation based patterns for each slot types. By using different number of patterns for seven different relations we observe about 3% increment in MAP results. In total, we improved the MAP results by 13.76% using Dirichlet language modeling along with relation based patterns.

Key Words:

Information Retrieval, Information Extraction, Language Modeling, Question Answering, Slot Filling, Pattern Matching