Simple and Efficient Dual-FOFE-net Neural Models for Entity Linking with PageRank

Feng Wei, Uyen Trang Nguyen, Hui Jiang
Dept. of EECS, York University, Canada
{fwei, utn, hj}@cse.yorku.ca

This paper presents a simple and computationally efficient approach for entity linking, which achieves state-of-the-art performance on both TAC2016 and TAC2017 task datasets.

PageRank based Distillation

PageRank based distillation not only enhances the candidate coverage, but also speeds up the whole model.

Limitations of conventional candidate generation module:
• Too many candidates
• Too much noise

We rank the candidates of each mention based on their numbers of outbound links to all the recognized mentions in the same document, and keep the top $\tau$ candidates for each mention.

Simple and Efficient Dual-FOFE-net

• Fixed-size ordinarily-forgetting encoding (FOFE) codes, as an alternative sequence embedding representations, can almost uniquely encode any variable-length sequence of words into a fixed-size representation without losing any information.
• A smaller forgetting factor is helpful for representing the positional information of all words in the sequence; a larger forgetting factor is useful for maintaining the long-term dependency of past context.

Results on TAC-KBP Task Datasets

On the TAC2016 dataset, our model outperforms the best system by 1.2% (NERL C F1) and 1.1% (CEAFm C F1), respectively, in terms of the overall trilingual entity linking performance.

On the TAC2017 dataset, our model achieves higher $F_1$ scores, by 0.2% (NERL C F1) and 0.4% (CEAFm C F1), respectively, compared to the best system in terms of the overall trilingual entity linking performance.