Name Entity Linking on Historical Newspaper Data

University of Zurich[UZH]

Chao Feng, Feifei Xia, Petra Abramovic

Team red

Introduction

Entity linking is an important part of text mining application.EL models are expected to accurately link the ambiguous mentions to entities in the knowledge base.

In this project, we adapted a neural EL model for the material of historical collections. This dataset introduces new challenges with historical and noisy inputs. Our results showed that this historical noise has a great influence on the model performance.

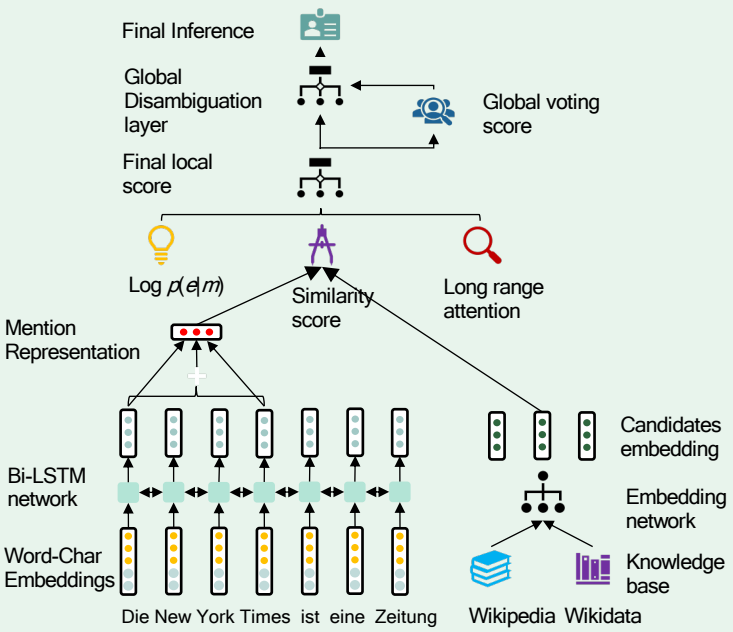
Task

HIPE is a shared task which aims in identifying historical entities and linking them with Wikidata knowledge base. This shared task corpora are composed of newspaper articles from several Swiss, Luxembourg and American historical newspapers.

Method

In this project, we adapted a neural entity linking method [Kolitsas:2018] to link the marked mentions with their ground truth entities.

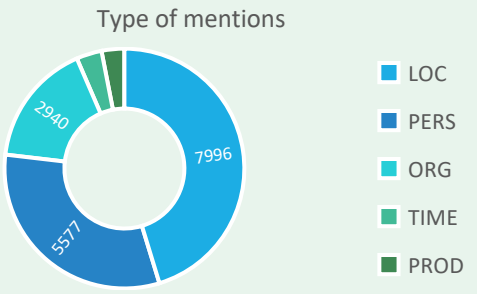
Model structure



Experiment and Resulte

We used the HIPE dataset as the corpus of our experiment. There are 162 documents in training set, 14 documents in dev set and 14 document in test set.

Number of mentions by type



Results of the final model

Data set	Micro/macro	Precision	Recall	F1
HIPE test	Micro	83.5	80.3	81.9
	Macro	84.0	82.3	83.1

Discussion

Examples of prediction

Mention	Prediction
Rath der 500	Rat der F¨unfhundert
Rarhder Alten	NIL
Tuillerien	Palaisdes Tuileries
Platz des Carrousel	Platz in Frankre-ich
Hause Coigny	Der Glanz desHauses Amberson

Green means prediction is right, red means prediction is wrong, and NIL means the model cannot find any candidate.

- About 15% of the mentions cannot find any candidates in our model.
- This model relies much on embedding similarity, if the similarity between the mention and the candidate is high, the model has a great chance to find the right entity.
- The historical noise has a huge influence on the results. Most of the LOC entities are predicted right, but some PERS entities are failed. Maybe the name of historical location did not change a lot, but the name (e.g. title) of historical people changed more frequently.

Acknowledgements:
This poster is the outcome of advanced techniques text mining course project.
We thank Dr. Jan Deriu and Nico Colic for their insightful comments.

References:
Kolitsas .et al. 2018. "*End-to-End Neural Entity Linking.*"Proceedings of the 22nd Conference on Computational Natural Language Learning.