

Biomedical Entity Linking Based on Less Labeled Data

Yu HU, Derong SHEN, Tiezheng NIE, Yue KOU, Ge YU

Frontiers of Computer Science, DOI: [10.1007/s11704-022-1192-8](https://doi.org/10.1007/s11704-022-1192-8)

Problems & Ideas

- Problems of weakly labeled biomedical entity linking:
 - The ambiguity of natural language descriptions, including polysemy and abbreviation, is an urgent problem to be solved.
 - Biomedical texts lack detailed annotations.
- Ideas: The algorithm, Relation-aware Biomedical Entity Linking(BioEL), is proposed to address these problems.
 - A multi-layer probabilistic network is proposed for entity linking, which generates the corresponding candidate concepts with less labeled entity annotation.
 - Jointly capture the entity and relation information for entity linking is important for this problem.
 - To reduce the ambiguity, all possible candidate concepts to each entity is computed, then rank the candidate concepts according to evaluation scores.

Main Contributions

- Contributions:
 - Entity linking and interaction relationship classification is jointly learned to improve the quality of entity linking based on weakly labeled data.
 - The results show that this work achieves a good experimental scores while the provision of the reliable labeled anchors.

<i>Method</i>	Acc	Acc-10	Rec
<i>Least Squares</i>	0.32	0.75	0.62
<i>CCA</i>	0.60	0.81	0.60
<i>Standard LSTM</i>	0.68	0.86	0.59
<i>MS-LSTM</i>	0.71	0.91	0.65
<i>KERN</i>	0.73	0.86	0.70
<i>BioEL</i>	0.86	0.90	0.79

We evaluate our approach on two datasets: the GENIA corpus as the training data and a set of PUBMED ab stracts(PubAbs) as the testing data. The BioER achieves 86.1% and 90.21% in accuracy and accuracy-10, while LSTM and MS-LSTM methods achieve 68.91% and 71.75% in strict accuracy and 86.15% and 87.53% in accuracy-20, respectively. Several other sets of experimental data show similar results.