

An integrated pipeline model for biomedical entity alignment

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

Frontiers of Computer Science, DOI: [10.1007/s11704-020-8426-4](https://doi.org/10.1007/s11704-020-8426-4)

Problems & Ideas

- **Problems of biomedical entity alignment**
 - The errors generated in the entity and relation extraction step may propagate to hidden information mining step.
 - Biomedical entity has distinctive features compared with common entity
- **Ideas: Dividing the entity alignment into two sub-tasks, Entity Identification and Entity-Concept Mapping**
 - we follow the recent works on deep neural networks and introduce the third-party knowledge base, then propose a pipeline model to extract biomedical entities.
 - we push forward the establishment of entity description standardization and give entity relation a structural expression via third-party knowledge base.
 - Each identified entity is regarded as a mapped concept in the concept graph.

Main Contributions

- The proposed biomedical entity alignment model performs better than state-of-the-art algorithms

| | | EXTRACTION | | | CLASSIFICATION | | |
|-----|-------|------------|------|------|----------------|------|------|
| | | P | R | F | P | R | F |
| STL | A-B-R | 75.5 | 79.1 | 77.3 | 62.7 | 70.9 | 66.5 |
| | A-B-L | 73.2 | 80.6 | 76.7 | 61.8 | 69.3 | 65.3 |
| | B-R | 71.5 | 75.5 | 73.5 | 59.1 | 67.5 | 63.0 |
| | B-L | 75.8 | 79.8 | 77.7 | 60.3 | 68.7 | 64.2 |
| MTL | A-B-R | 79.6 | 83.6 | 81.5 | 67.3 | 75.7 | 71.2 |
| | A-B-L | 80.2 | 82.7 | 81.4 | 65.9 | 72.5 | 69.0 |
| | B-R | 79.1 | 79.5 | 79.2 | 58.7 | 70.3 | 63.9 |
| | B-L | 79.9 | 81.3 | 80.5 | 60.4 | 69.1 | 64.4 |

| | | EXTRACTION | | | CLASSIFICATION | | |
|-----|-------|------------|------|------|----------------|------|------|
| | | P | R | F | P | R | F |
| STL | A-B-R | 68.5 | 75.2 | 71.6 | 60.5 | 69.9 | 64.8 |
| | A-B-L | 71.3 | 74.5 | 72.8 | 61.3 | 70.3 | 65.5 |
| | B-R | 63.4 | 70.7 | 66.8 | 59.5 | 67.1 | 63.0 |
| | B-L | 65.7 | 72.8 | 69.0 | 59.3 | 68.5 | 63.5 |
| MTL | A-B-R | 75.6 | 79.4 | 77.4 | 63.1 | 70.5 | 66.5 |
| | A-B-L | 73.5 | 80.3 | 76.7 | 63.7 | 71.6 | 67.4 |
| | B-R | 73.2 | 76.5 | 74.8 | 60.7 | 68.3 | 64.2 |
| | B-L | 74.9 | 76.3 | 75.6 | 62.5 | 70.6 | 66.3 |

| DATASET | METHOD | ENTITY EXTRACTION | | |
|---------|------------|-------------------|-------------|-------------|
| | | P | R | F |
| DATA1 | our method | 80.2 | 82.7 | 81.4 |
| | Li's | 78.6 | 79.6 | 79.1 |
| | Huang's | 78.5 | 80.3 | 79.3 |
| DATA2 | our method | 73.4 | 80.2 | 76.6 |
| | Li's | 73.2 | 77.9 | 75.4 |
| | Huang's | 72.7 | 79.5 | 75.9 |

| DATASET | METHOD | ENTITY CLASSIFICATION | | |
|---------|------------|-----------------------|-------------|-------------|
| | | P | R | F |
| DATA1 | our method | 65.9 | 74.3 | 69.8 |
| | Li's | 63.2 | 74.6 | 68.4 |
| | Huang's | 65.5 | 73.5 | 69.2 |
| DATA2 | our method | 63.7 | 68.6 | 66.0 |
| | Li's | 63.1 | 65.7 | 64.3 |
| | Huang's | 62.4 | 67.5 | 64.8 |