Analysis of NIL Results in an Entity Linking System by Mai Linh Pham

Anton Serjogin
Centre for Information and Speech Processing, LMU
anton.serjogin@gmail.com

12.06.2017

Entity linking is a process of linking mentions from text to a corresponding entity in knowledge base, entities that are not linked to the knowledge base are called NIL results. Fine-grained entity annotation is an expanded set of tags in comparison to the standard named entity recognitions tags that feature location, organisation, person and miscellaneous. Entity linking systems can not link all entities and in order to improve them, NIL results should be analyzed, which is why the aim of this work is to determine whether fine-grained types are useful for clustering and analyzing NIL mentions.

Task description: firstly, combine the outputs of an entity annotation tool and entity linking system, secondly, extract NIL output, thirdly, cluster the NIL output and finally, use fine-grained types for clustering task.

FIGER is a fine-grained entity recognition system which features more than a hundred labers arranged in a hierarchical structure, it allows overlapping types and is better in uncommon entity recognition.

WAT is a entity linking system, which consists of: spotter (scans the input text for mentions and retrieves a list of candidate entities), disambiguator (ranks candidate entities with different disambiguation algorithms), pruner (removes useless annotations, aims at increasing precision). As mentioned before, the outputs of these both tools should be combined and in order to combine them, these outputs should be in the same format.

The next step is to extract the NIL mentions. Some mentions that are annotated by FIGER are not linked by WAT, which is why a list with all entity names in the knowledge base should be created (all wikipedia titels + all wikipedia redirect links) and then these unlinked mentions, with a corresponding KB entry, should only be regarded.

- Coarse-grained type (features a large number of single-level tags that are mainly divided into several groups)
- Fine-grained type (division into multi-level and single-level types, semantical clustering, mapping of single types to multi-level types, multi-level type clustering)
- Top-level type (a large number of first-level tags in multi-level tags are identical, grouping results in cluster reduction, too specific types are clusterd as ”undefined”, new domains are created to reduce these undefined types)

In conclusion, fine-grained entity types:

- Can be used for NIL mention semantic clustering
- Take into account both lexical and contextual entity properties
- Information anchord in tags can be used for analysis
- More informative than coarse-grained types