

Related Entity and Role Identification

Problem Statement:

In a given event, there could be many entities involved. But not all entities are related to an event. [Define what it means related]. If we want to summarize the event, we need to identify which entities are involved and their respective role.

Pre-processing:

Stop word removal, Lemmatization and POS Tagging and NER tagging for person; location and all.

Methods:

1. Sequence Tagging or Extended NER Model

Similar to Named Entity Recognition, we can model it as classification problem. Where we take a sequence of words and provide tag to them.

A) This could be learned on different granularity level.

- Sentence Level

- Paragraph Level

- Document Level

- Event Level

- We can combine these granularity to predict role.
- We can also look for multimodal information related to an event, somehow improve accuracy by learning from different doc of related event.
- We can also combine NER tagger identification of Person, Location and Date to give score.
- A sentence/sentence can be given weight-age if it has a sequence of entities or roles which are more important.
- Look for entity co-reference, relation extraction models and understand them.
- See if LDA could be used in this.
- An entity present in all the articles related to event is more important.

2. Lexicon and Syntax Based

1. Apply NER tagger and relation extraction. Have relations built between entities.

2. Build phrases using dependency parser and extract sentences using that phrases and find the related entities.

3. Somehow score the sentences with weights, which will

4. Apply LDA get the words for topic and look for proximity and get the entities.

3. Semantic Role Labelling

Accused will be actor and Victim will be participant.

4. Across Document

a) Take all the articles related to an event. The entity which is being affected in any way will be occurring in all the documents. And also will occur in the same document again and again. **Frequency** will be one factor to give importance. Second **Position** of entity matters. Mostly important entities will be in Topic and in the start.

Approach

Figure out some baseline methods for doing this. HMM CRF, MEMM, LSTM, BLSTM.

Either we can improve these methods or give comparisons.

Look for meta information, related event articles, multimodal data video image texts, social media tweets.

*** Other since our dataset is small. using other dataset to improve performance

Few Fundamental Concepts that could be used to solve the problem:

1. Named Entity Recognition

2. Semantic Role Labelling

3. Word Sense Disambiguation

4. Entity co-reference resolution

5. Entity Relation Extraction

6. Topic and word relation

7. Event Summarization