Related Entity and Role Identification

Alternate Title: [Context Based Event Centric Fine Grained Named Entity Recognition]

Problem Statement:

In a given event, there could be many entities could be mentioned in online media. But not all entities are related to an event. Related may be defined as event caused by that entity, entity affected in the event or associated to that event. We need to identify such entities are involved and their respective role.

Why our problem is complex

In NER methods, previous words has been taken and also the grammatical structure has been used. Here the entities are identified and classified with an assumption that entities follow the similar sentence structure. To learn this structure, we need huge training set and that we have.

In other entity classification methods like fine grained entity classification, people use wiki dataset and free base kind of data sets where the tags are already given to known entities.

Our problem is more complex because we are focusing on event specific roles to be assigned to entities. Apart from influenced by sentence structure role is influence by domain of event. Within domain also which type of event. In a terrorist attack domain, bomb blast, hijack, danga etc could be different type. Same entity could play different role depending upon how it is involved which makes the problem more complex. Also dataset is big problem for such problems.

How this problem will help in my Actual problem of event forecasting in Social Unrest?

In event forecasting in final event generation, I need to display date, location, person and organization at least. Role may also be shown. We need to show entities involved.

- All the entities having roles can be called related entities.

- In case we can't generate roles for Social unrest from bomb blast, we still can use entity ranking functions to get top rated entities class wise and show them.

- We can identify entities whose activity/presence needs to be tracked for forecasting. Like Jantar Mantar, Some nuisance persons and others.

Pre-processing:

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

Method:

Sequence Models - Sentence Level

- Have annotated articles so that it can be used as Supervised classification problem.

- Baseline methods used sequence models. Use NER methods. HMM, CRF, LSTM, BLSTM

- Also use Fine Grained Entity where number of classes are more.

- See why these sequence methods failed and identify such characteristics,

Document Level

- Identify document level attributes to increase the performance. Like same entity may have assigned different labels.

- May involve context of the article to have proper role. **Example -** Obama addressing a rally. In a rally he launched his book. With sentence level Obama is politician as well as author. But from whole document he is Politician in this context.

Event Level

- Have multiple documents of the same event and try to identify the role.

- Context should be factor in label assignment.

Sentence Level - Syntactic Structure

- Use syntax features to improve the NER method. Use Semantic Role Labelling

- Identify lexicons and structure to identify roles. Example: 'Wasim was injured in blast at Assam'. Due to injured , 'Wasim' is accused.

- Not only take previous words but also the next set of words as context.

Giving weight age to entities words than other words and also rank entities too.

- Giving weight age to entities will help identify entities.

- Also ranking entities will help assign correct labels at document level and event level in case of multi label.

- Ranking entities have other use cases also. Sometimes we want normal NER tagger with person, org and location tag. But we want to have only relevant person.

- Entity Ranking can be modelled as IR problem where entity is query and article is doc

- LDA ranking of entities could be used.

Generalize Model/Method to be used on other domain (Transfer Learning)

- Understand characteristic of dataset to identify which methods works better and why.

- Domain adaption. Model trained on one domain could be used on other domain. Model to label the role can be used on other domain or similar domain or dataset having similar characteristic.

- NER tagger + Scoring method should work on all the domains.

- Alternatively create tagged dataset manually and test method.