A) Analysis:

1. Failing to capture relation between entity roles and entity mention:

One of the major reason for failure of these methods is that, they fail to identify certain semantic relationship between entity-roles and entity-mentions. The way roles are defined and used while writing an article are not covered by these methods. Sometimes a particular role is defined in the relation with another role. e.g. *Yunus was arrested in the case of Mumbai blast. It is believed that he was a mastermind of Indian Mujahiddin who got training in Pakistan.* Here Yunus is identified as *PER_Accused* and this identification leads to the related entities *Indian Mujahiddin* as *ORG_Accused* and *Pakistan as LOC_Accused*. While writing an article different mentions of the same entity could be spread across document. One mention *m1* of an entity could be assigned as correct role using sentence level context and other by referring to this mention *m1*.

Capturing Relation Between Entity Mentions

i) Use 'true type' hypothesis which says that an entty could have only one type/role in an event. Define a preference order in the roles, and if entity is identified by that role, update the other mentions. e.g. if a mention is previously identified as xxx_Others and later the same entity is identified by xxx_ Accused, update the other mentions.

- Manually define the preference.
LOC_Event > LOC_Accused > LOC_Others (Obvious LOC_Event plays more importance than other LOC tags)
PER_Accused > PER_Victim > PER_Others (If the person is accused, he could be victim also cases like fiyadin attack or got injured in the blast triggered by him/herself)
ORG_Accused > ORG_Victim > ORG_Others

- Try out all the permutations and get the best one.

Identifying Multiple mentions of Same entity:

i) Use any method of corefeernce resolution OR
ii) Use following rules simply,
- Use partial names based resolution. e.g Combine 'Narendra Modi' and 'Modi'
- For organization use initial letters. 'United Liberation Front of Assam' and ULFA are same.

Identify sentences which are useful for Entity Typing

Divide the sentences into two groups:
- Which defines the role for entity mention

- Sentence having entity mentions which refer to mentions in other sentences.

Linear Regression or Simple Neural Net or RNN Based Score Calculation

Information flow from one mention to other back and forth. (RNN principle)
Final Score of a mention *Mi*, *Tj* = Sentence level score of Mi, Tj + Other mention score wrt all the types.

Learn all types together and seperate

Together because roles are dependent on each other.

2. Inherent issue with HMM, CRF and Skip-Gram(Need to check this)**:** These statistical natural language processing models use only local features and this makes them unable to fully account for the long distance structure that is prevalent in language use. Ref: Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling

3. Similar context of Person and Orgnanization for victim and accused role. They could be differentiated by top level types, so NER and Entity retrieval should be combined somehow.

4. Inspect the similar words of the Roles and Type vector similarities

Cosine Distances among Types

Role	Role	Cosine Distances
PER_Others	PER_Victim	0.45
PER_Others	PER_Accused	0.17
PER_Accused	PER_Victim	0.42
LOC_Event	LOC_Accused	0.21
LOC_Event	LOC_Victim	0.99
LOC_Event	LOC_Others	0.19
LOC_Accused	LOC_Others	0.09
LOC_Accused	LOC_Victim	0.93
LOC_Others	LOC_Victim	0.94
ORG_Others	ORG_Victim	0.28
ORG_Others	ORG_Accused	0.15
ORG_Accused	ORG_Victim	0.28

5. Analyze the final results.

ev_015_st_009.txt

LOC_Event:

Dimapur_railway_station[LOC_Event] Nagaland[LOC_Event]
Nagaland[LOC_Others] Dimapur[LOC_Event] Assam[LOC_Event]

LOC_Others

Nagaland[LOC_Others] Dimapur[LOC_Event]
Dimapur_railway_station[LOC_Event] Kohima[LOC_Others]
Nagaland[LOC_Event]

2011_6_29_st-235.txt

PER_Victim V.J._Chandiran[PER_Others] Jyothi[PER_Victim]
Chandiran[PER_Others]
PER_Others V.J._Chandiran[PER_Others] Chandiran[PER_Others]
Jyothi[PER_Victim]

Analysis: Jyothi above Chandiran in PER_Vcitim and Chandiran above Jyothi in PER_Others

In general current method and model tries to give correct results by giving actual labels higher score, but the boundary is still blur among. The reason could be:
Type representation are learned by replacing all the mentions. We should take those sentences to learn representation of Type which contain the few seed words. Means the type is having significant words in the representation but majorily dominated by common words.

Using CNN could adjust these weights of the words giving better results.

6. Dataset based observations:
i) LOC_Victim is very less and conflict with LOC_Event. Drop this role.
ii) ORG_Victim and LOC_Event conflicts. e.g Attack happened near BJP headquarters. Somewhere ‘BJP headqurters’ is tagged as LOC_Event and somewhere ‘BJP’ as orgnization victim.

iii) LOC vs ORG: Bomb blasts in Iraq. Then Iraq is a LOCATION. Iraq asked for help from Kuwait to deal with attackers. Iraq is not location but the Organization here.
iv) Loose Tag Definitions of LOC_Accused, PER_Victim, ORG_Victim and ORG_Accused causing ambiguity in tagging.

B) Further Experiments:

i) Ranking Mechanisms

New Ranking Mechanism

1. Use BM25 Ranking
2. Use CNN architecture from 'Learning to Rank Short Text Pairs with CNN'. Design this architecture.
3. Include 'Entity Salience' features as described in Entity Salience papers.

Changes in Existing Ranking Mechanism

i) Dont include the role xxx_Others
ii) Don't include duplicate results. Take the first one only.

ii) Entity Representation

- Learn entity in similar fashion of learning Type.
- Use bigram instead unigram in context of entity.
- Remove non-informative words from the context.

iii) Others

1. XXX_Others are giving very high accuracy. These entities are not relevant for us or for our task. Remove them from Ranking. Exclude XXX_Others from the answer
2. Get the statistics of the roles.
- Fequency of each role
- Number of role per document. Mean, Average
2. Break sentence on '.   O'

3. Stop words and punctuation.

C) Paper Writing

1. Our event is events of bomb blasts be in in present, past or future.
2. Common framework for NER and Entity Ranking. Use precision as a measure to get the results for NER based methods.

3. Get the results for only PERSON, LOCATION and ORGANIZATION in NER and compare the changes in the subtype inclusion.
4. Experiment with different parameters also.
5. Write Related works

Paper Reading to write the paper:

Information Retreival and Short Text Matching

1. Convolutional Neural Network Architectures for Matching Natural Language Sentences, NIPS 2014

2. Neural Netwroks for IR, WSDM'18

NER:

1. https://blog.paralleldots.com/data-science/named-entity-recognition-milestone-models-papers-and-technologies/
2. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling
3. NLP from Scratch

Entity Retrieval

1. Role based Named Entity Retrieval
2. Context-based Named Entity Retrieval

Question Answering

1. SLP chapter
2. Deep learning for Answer sentence selection
3. Convolutional Neural Networks vs. Convolution Kernels: Feature Engineering for Answer Sentence Reranking
4. SemEval-2017 task3: Community Question Answering

List QA and Set Completion

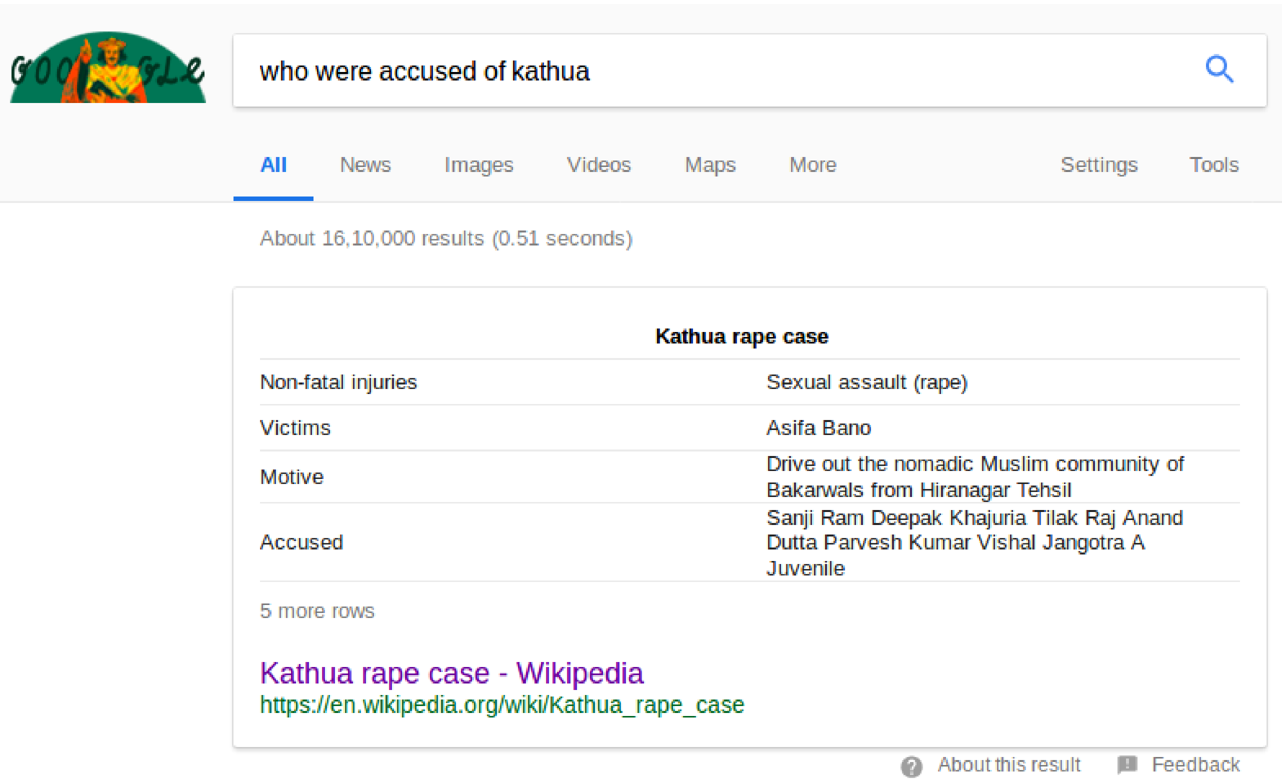
D) Dataset correction

Till ev_15, done. Do more.
Tag document with event

Misc

Current Q&A system relied on explit tagging of answers from Wikipedia or IMDB
 Examples:

i) Accused of Kathua rape case:



Kathua rape case

Location	Kathua, Jammu and Kashmir
Coordinates	32.385°N 75.517°E
Date	10 January 2018- 17 January 2018
Target	Asifa Bano
Attack type	Abduction, Rape, and Murder
Non-fatal injuries	Sexual assault (rape)
Victims	Asifa Bano
Motive	Drive out the nomadic Muslim community of Bakarwals from Hiranagar Tehsil[1]
Accused	Sanji Ram Deepak Khajuria Tilak Raj Anand Dutta Parvesh Kumar

Google

Where is the Louvre Museum located?

All Maps News Images Videos More Settings Tools

About 2,93,00,000 results (1.74 seconds)

Paris

The Louvre (US: /ˈluːv(rə)/), or the Louvre Museum (French: Musée du Louvre [myze dy luvʁ] (listen)), is the world's largest art museum and a historic monument in **Paris**, France. A central landmark of the city, it is located on the Right Bank of the Seine in the city's 1st arrondissement (district or ward).

Louvre - Wikipedia

https://en.wikipedia.org/wiki/Louvre

About this result Feedback