Cross-Supervised Joint-Event-Extraction with Heterogeneous Information Networks

Yue Wang[1,5], Zhuo Xu[1], Lu Bai[1], Yao Wan[2], Lixin Cui[1], Qian Zhao[5], Edwin R. Hancock[3], Philip S. Yu[4]

1. Central University of Finance and Economics, Beijing, China
2. Huazhong University of Science and Technology, Wuhan, China
3. Department of Computer Science, University of York, York, UK
4. University of Illinois at Chicago, Chicago, USA
5. State Key Laboratory of Cognitive Intelligence, iFLYTEK, Hefei, China

Outline

- Introduction
- Preliminaries
- Our Proposed Model
- Experiments and Analysis

Introduction

The New York Times Annotated Corpus contains over 1.8 million articles written and published by the New York Times between January 1, 1987 and June 19, 2007 with article metadata provided by the New York Times Newsroom, the New York Times Indexing Service and the online production staff at nytimes.com. The corpus includes:

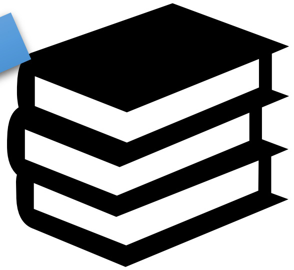
- Over 1.8 million articles (excluding wire services articles that appeared during the covered period).
- Over 650,000 article summaries written by library scientists.
- Over 1,500,000 articles manually tagged by library scientists with tags drawn from a normalized indexing vocabulary of people, organizations, locations and topic descriptors.
- Over 275,000 algorithmically-tagged articles that have been hand verified by the online production staff at nytimes.com.
- Java tools for parsing corpus documents from .xml into a memory resident object.

As part of the New York Times' indexing procedures, most articles are manually summarized and tagged by a staff of library scientists. This collection contains over 650,000 article-summary pairs which may prove to be useful in the development and evaluation of algorithms for automated document summarization. Also, over 1.5 million documents have at least one tag. Articles are tagged for persons, places, organizations, titles and topics using a controlled vocabulary that is applied consistently across articles. For instance if one article mentions "Bill Clinton" and another refers to "President William Jefferson Clinton", both articles will be tagged with "CLINTON, BILL".

The New York Times has established a community website for researchers working on the data set at http://groups.google.com/group/nytnlp and encourages feedback and discussion about the corpus.

Named entities:
e.g. person, company, organization, geographic name...

Event triggers:
e.g. bankrupt, conflict, marry, dispatch...



Relations:
e.g. /location/location/contains,
/people/person/place_of_birth,
/people/person/place_lived

Introduction

Named entities:
e.g. person, company, organization, geographic name...

Event triggers:
e.g. bankrupt, conflict, marry, dispatch...

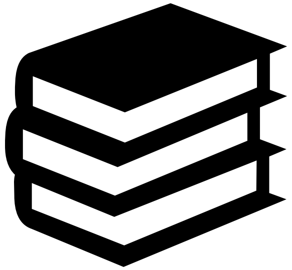
Relations:
e.g. /location/location/contains,
/people/person/place_of_birth,
/people/person/place_lived

The New York Times Annotated Corpus contains over 1.8 million articles written and published by the New York Times between January 1, 1987 and June 19, 2007 with article metadata provided by the New York Times Newsroom, the New York Times Indexing Service and the online production staff at nytimes.com. The corpus includes:

- Over 1.8 million articles (excluding wire services articles that appeared during the covered period).
- Over 650,000 article summaries written by library scientists.
- Over 1,500,000 articles manually tagged by library scientists with tags drawn from a normalized indexing vocabulary of people, organizations, locations and topic descriptors.
- Over 275,000 algorithmically-tagged articles that have been hand verified by the online production staff at nytimes.com.
- Java tools for parsing corpus documents from .xml into a memory resident object.

As part of the New York Times' indexing procedures, most articles are manually summarized and tagged by a staff of library scientists. This collection contains over 650,000 article-summary pairs which may prove to be useful in the development and evaluation of algorithms for automated document summarization. Also, over 1.5 million documents have at least one tag. Articles are tagged for persons, places, organizations, titles and topics using a controlled vocabulary that is applied consistently across articles. For instance if one article mentions "Bill Clinton" and another refers to "President William Jefferson Clinton", both articles will be tagged with "CLINTON, BILL".

The New York Times has established a community website for researchers working on the data set at http://groups.google.com/group/nytnlp and encourages feedback and discussion about the corpus.

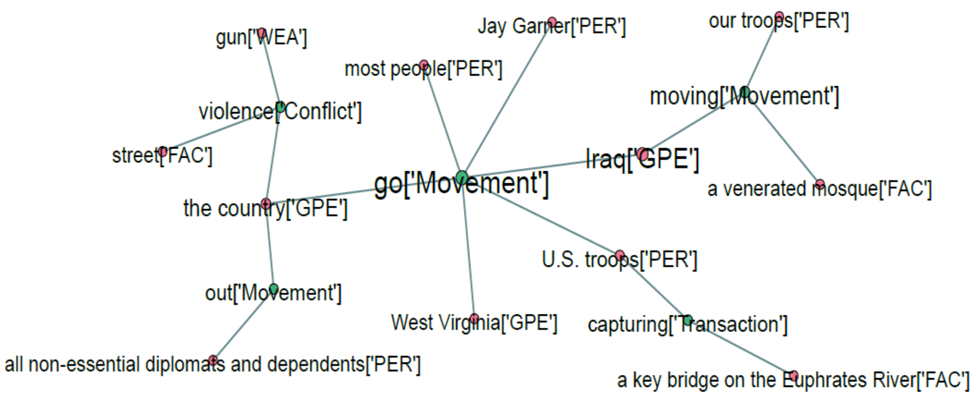


Introduction

- Problems of existing methods

The original positions for triggers (green) and entities (red).

Direct-adjacency-distribution for entities (Direct) v.s. meta-path-based distribution (Meta) based on a given trigger

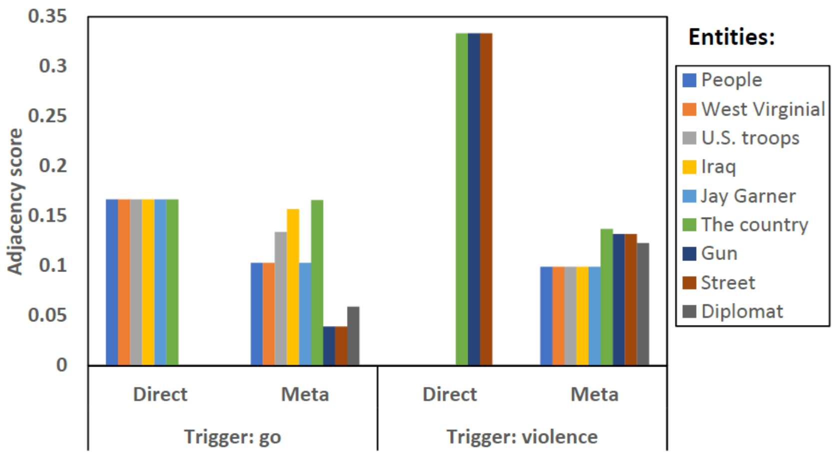


(a)

10 sentences from the ACE 2005 corpus:

1. At daybreak, U.S. troops pushed through the outskirts of Karbala on the road to Baghdad, capturing a key bridge on the Euphrates River.
2. And so I would like you to take a look at the CNN / " USA TODAY " / Gallup poll, taken last week, should U.S. troops to go to Iraq to remove Saddam Hussein from power.
3. Franks was in charge of the operation that was supposed to quickly go in, take over Iraq, and then start moving our troops out rapidly -- at this point I think we 're going to see our troops over there at least through the end of the decade.
4. Armed coalition soldiers moving toward a venerated mosque at the request of a Muslim cleric, but angry locals who didn't understand what they were trying to do.
5. Yeah, I did go through -- West Virginia, one time through -- from Pittsburgh.
6. Jay Garner the retired general will go into Iraq soon with his troops soon.
7. Harrods is Harrods though, and most people go there to be blown away by the prices.
8. But despite issuing a host of tough decrees, Bremer has failed to stem the rampant crime and street violence plaguing the country.
9. In a horribly deceitful manner, the Brady Campaign has released " report cards " for every state on their gun laws that supposedly shield children from gun violence.
10. the state department is ordering all non-essential diplomats and dependents out of the country.

(b)



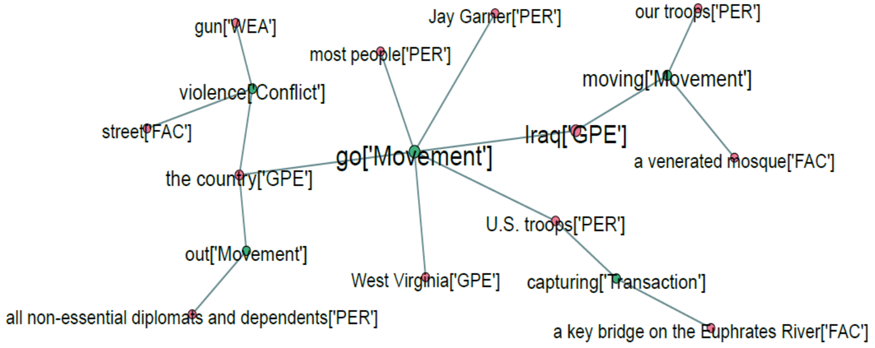
(c)

e.g. from "go" to "gun" along the meta-path "**Movement-GPE-Conflict-WEA**"

An example of co-occurred relationships between triggers and entities.

Introduction

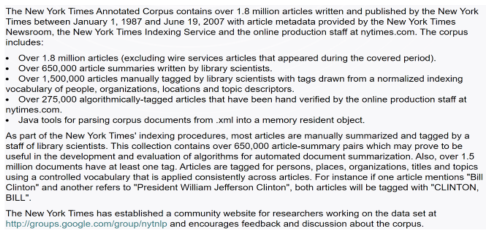
- Heterogeneous Information Network of entities and triggers



Helps

Named entities:
e.g. person, company, organization, geographic name...

Event triggers:
e.g. bankrupt, conflict, marry, dispatch...



The New York Times Annotated Corpus contains over 1.8 million articles written and published by the New York Times between January 1, 1987 and June 19, 2007 with article metadata provided by the New York Times Newsroom, the New York Times Indexing Service and the online production staff at nytimes.com. The corpus includes:

- Over 1.8 million articles (excluding wire services articles that appeared during the covered period).
- Over 650,000 article summaries written by library scientists.
- Over 1,500,000 articles manually tagged by library scientists with tags drawn from a normalized indexing vocabulary of people, organizations, locations and topic descriptors.
- Over 275,000 algorithmically-tagged articles that have been hand verified by the online production staff at nytimes.com.
- Java tools for parsing corpus documents from .xml into a memory resident object.

As part of the New York Times' indexing procedures, most articles are manually summarized and tagged by a staff of library scientists. This collection contains over 650,000 article-summary pairs which may prove to be useful in the development and evaluation of algorithms for automated document summarization. Also, over 1.5 million documents have at least one tag. Articles are tagged for persons, places, organizations, titles and topics using a controlled vocabulary that is applied consistently across articles. For instance if one article mentions "Bill Clinton" and another refers to "President William Jefferson Clinton", both articles will be tagged with "CLINTON, BILL".

The New York Times has established a community website for researchers working on the data set at http://groups.google.com/group/nytnlp and encourages feedback and discussion about the corpus.

Introduction

Comparison with different models without any pre-defined features

TABLE II: Comparison on real-world datasets

Model	ACE 2005			NYT			CoNLL			WebNLG		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Seq2Seq	0.442±0.025	0.493±0.0272	0.466±0.026	0.818±0.012	0.832±0.012	0.825±0.012	0.709±0.015	0.852±0.011	0.774±0.013	0.851±0.009	0.910±0.007	0.880±0.008
CRF	0.434±0.031	0.478±0.033	0.455±0.032	0.813±0.011	0.828±0.011	0.821±0.01	0.718±0.016	0.867±0.013	0.785±0.014	0.864±0.005	0.921±0.005	0.892±0.005
GCN	0.435±0.030	0.487±0.032	0.459±0.031	0.804±0.013	0.819±0.013	0.811±0.013	0.706±0.015	0.871±0.014	0.780±0.013	0.884±0.008	0.931±0.008	0.907±0.008
JEE	0.423±0.023	0.468±0.030	0.443±0.026	0.717±0.009	0.645±0.014	0.679±0.012	0.713±0.019	0.814±0.013	0.76±0.015	0.775±0.015	0.818±0.012	0.796±0.013
JT	0.469±0.003	0.426±0.005	0.447±0.004	0.725±0.012	0.691±0.006	0.708±0.009	0.738±0.025	0.837±0.006	0.784±0.021	0.818±0.011	0.829±0.007	0.823±0.008
CSM_{DA}	0.455±0.024	0.494±0.022	0.474±0.023	0.835±0.012	0.847±0.012	0.841±0.012	0.730±0.017	0.856±0.021	0.788±0.019	0.908±0.005	0.941±0.004	0.924±0.004
CSM_{HIN}	0.477±0.030	0.533±0.033	0.503±0.031	0.859±0.007	0.870±0.008	0.865±0.008	0.754±0.018	0.890±0.020	0.816±0.017	0.923±0.004	0.953±0.003	0.937±0.003

Model	Entity extraction			Trigger extraction		
	Precision	Recall	F1	Precision	Recall	F1
Seq2Seq	0.494	0.489	0.49	0.383	0.426	0.403
CRF	0.502	0.483	0.491	0.395	0.473	0.431
GCN	0.508	0.491	0.499	0.381	0.443	0.410
JEE	0.451	0.497	0.472	0.407	0.411	0.409
JT	0.492	0.458	0.474	0.447	0.414	0.432
CSM_{DA}	0.509	0.535	0.52	0.404	0.442	0.422
CSM_{HIN}	0.512	0.552	0.532	0.464	0.484	0.474

Our models

Introduction

Main contributions of this paper.

- First to use the indirect “entity-trigger” cooccurrence relationships (encoded in HIN) to improve the performance of the joint-event-extraction task.
- Proposed **Cross-Supervised-Mechanism.**
- Verify the indirect “entity-trigger” cooccurrence relationships is helpful.

Preliminaries

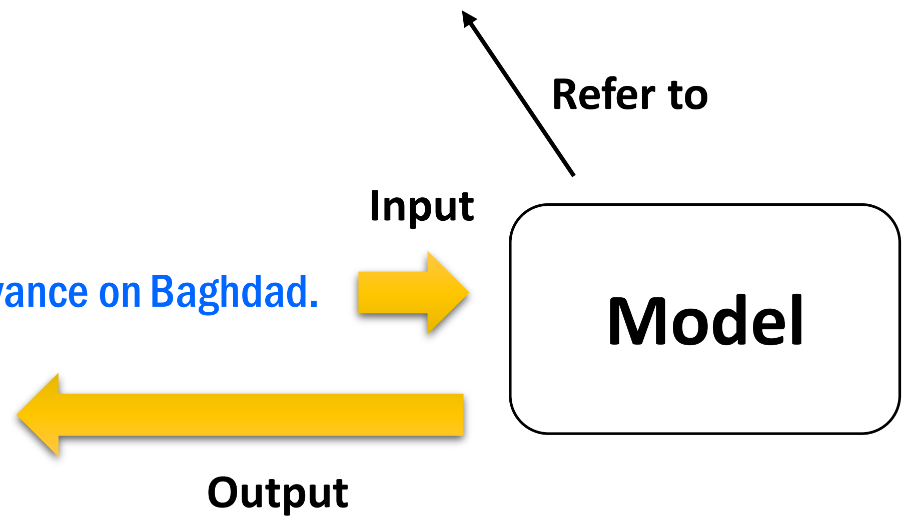
- The Joint-Event-Extraction Task

Sequence-to-Sequence Labeling. The goal of joint-event-extraction is to train a machine learning model under the supervision of a pre-annotated corpus. Minimizing the cross-entropy loss function [15] has always been introduced to achieve this goal. The cross-entropy loss function is defined as follows:

$$\mathcal{L} = \arg\min \sum_{\forall i \in [1,n]} \sum_{\forall y_i \in \mathcal{A}} -Pr(y_i|w_i) \log(\hat{Pr}(y_i|w_i)), \quad (1)$$

Combined tag-set (example)

B_{ORG}, I_{ORG}, B_{PER}, I_{PER}, B_{ATTACK}, O,…



Armored forces destroyed dozens of Iraqi tanks and personnel carriers in their advance on Baghdad.

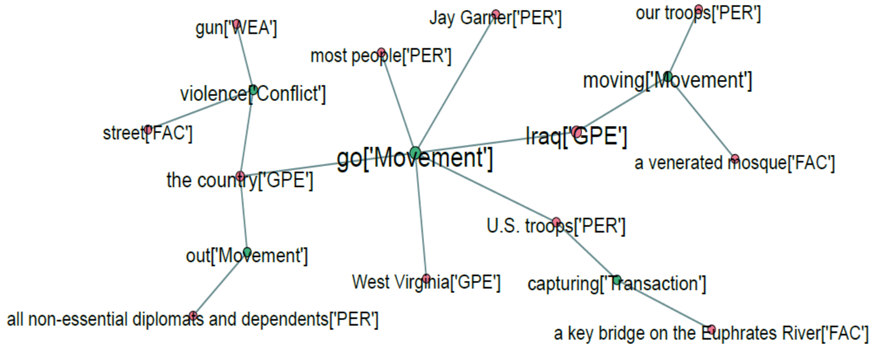
$B_{Conflict}$ B_{VEH} I_{VEH} I_{VEH} I_{VEH} B_{VEH} I_{VEH}

trigger

entities

Preliminaries

- “Entity-Trigger” Heterogeneous Information Network
  - It is defined as a weighted graph
  - **G = <V, E, W>**
  - **V** is a node set of entities and triggers
  - **E** is the co-occurred relationships between nodes



Two maps

$$\phi : V \to \mathcal{A}$$

$$\psi : E \to \mathcal{R}$$

Preliminaries

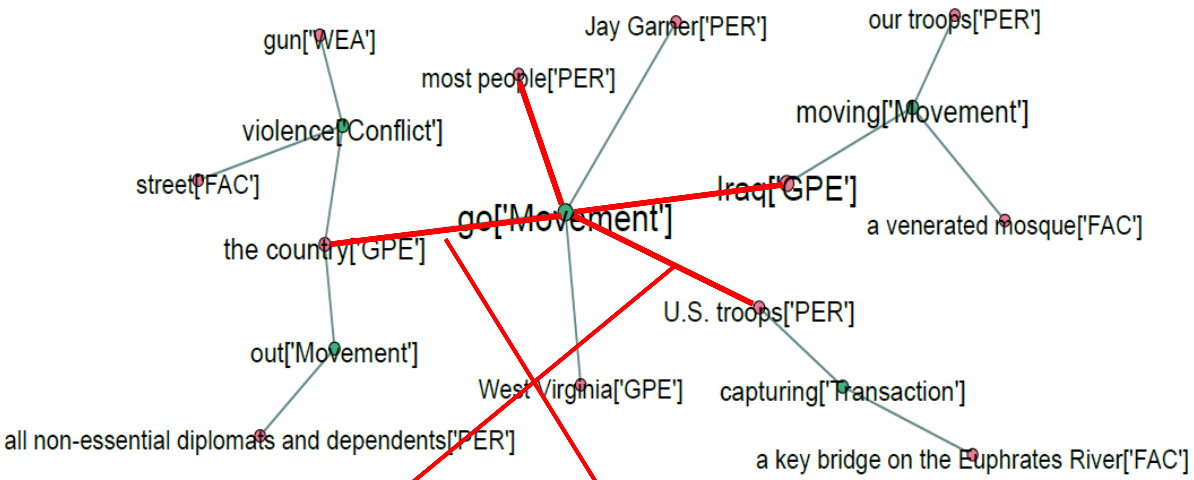
- “Entity-Trigger” Heterogeneous Information Network

  Meta-path between two nodes (entities or triggers)

$V_1 \; \mathcal{A}_1\mathcal{A}_2 \ldots \mathcal{A}_{l+1} \; V_2$

An example of meta-path

“PER-Movement-GPE”



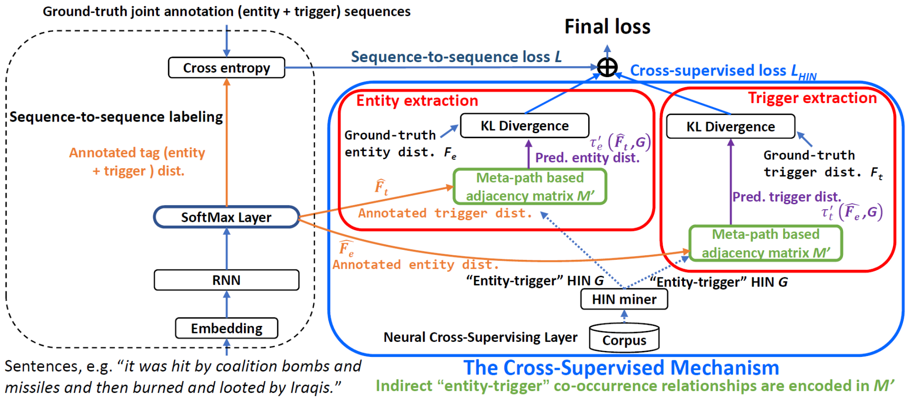
“U.S. troops-go-Iraq”, “most people-go-the country”

Our Proposed Model

Entity Distribution
Trigger Distribution

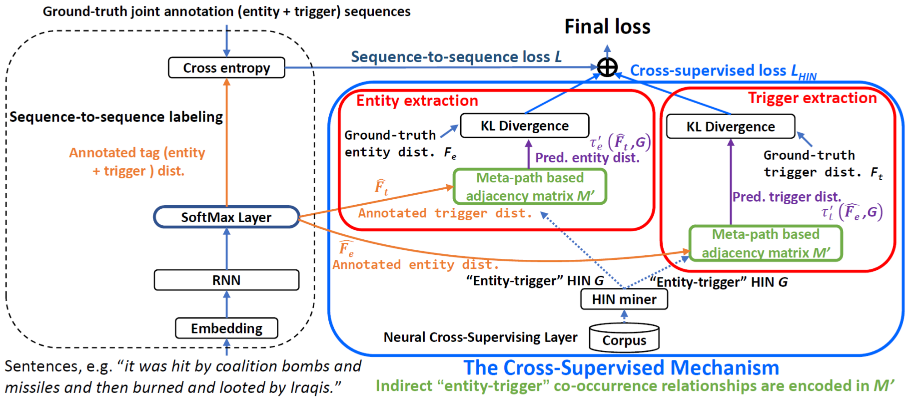
Frequencies for diff. named entities

Frequencies for diff. event triggers



Our Proposed Model

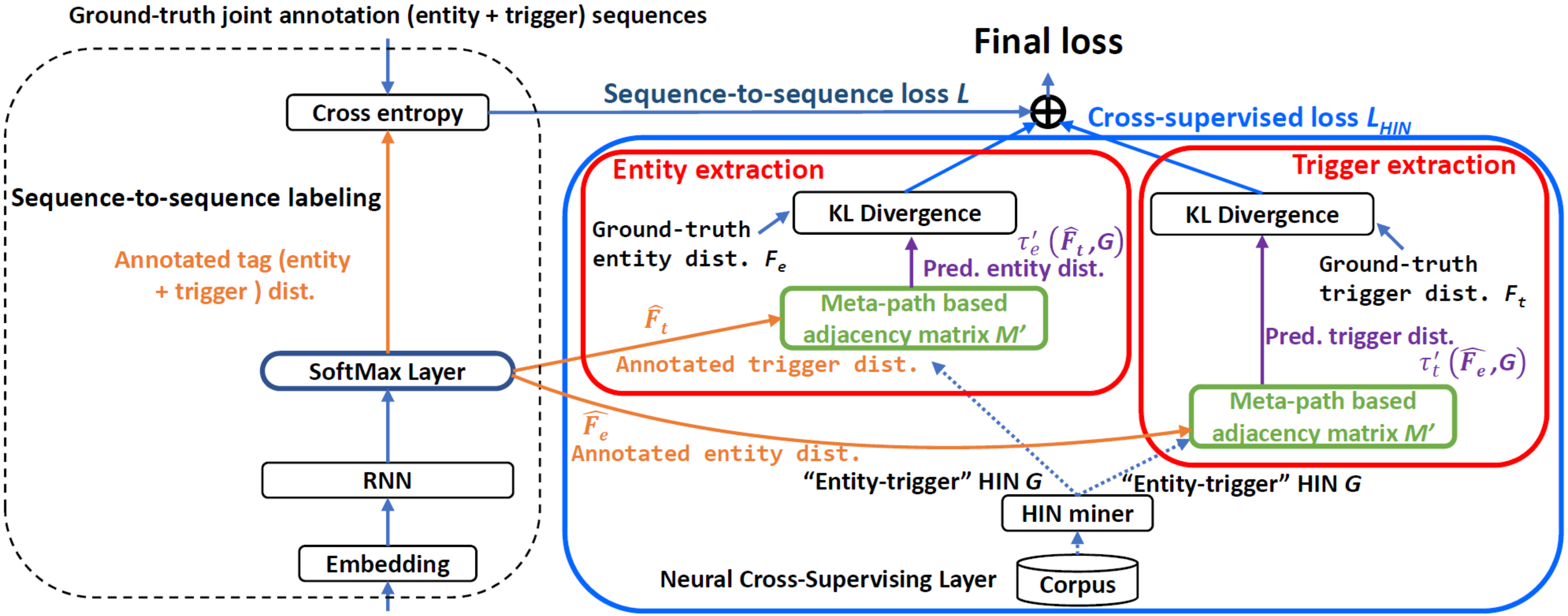
Entity Distribution
Trigger Distribution



Cross-Supervised Mechanism

$$\mathcal{L}_{HIN} = F_t \log \frac{F_t}{\tau'_t(\hat{F}_e, G)} + F_e \log \frac{F_e}{\tau'_e(\hat{F}_t, G)},$$

Our Proposed Model



Sentences, e.g. *"it was hit by coalition bombs and missiles and then burned and looted by Iraqis."*

The Cross-Supervised Mechanism

Indirect "entity-trigger" co-occurrence relationships are encoded in *M'*

Experiments and Analysis

- **Datasets**

From Linguistic Data Consortium (LDC)

New York Times Newsroom

A Spanish corpus made available by the Spanish EFE News Agency

A challenge of natural language generation

TA LE I: Da et statistics

	ACE2005	NYT	CoNLL	WebNLG
sentences	2,107	6,304	3,932	10,165
entities	4,590	12,643	13,511	2,217
triggers	1,921	6,355	3,903	1,309
entity types	11	17	4	9
event types	8	4	11	289
meta-paths (l=3)	4,459	18,035	22,399	12,675

Experiments and Analysis

A basic model in seq2seq framework with a combined tag-set

- Sequence-to-Sequence Joint Extraction (Seq2Seq)
- Conditional Random Field Joint Extraction (CRF)
- GCN
- Joint Event Extraction (JEE)
- Joint Transition (JT)
- CSM_{DA}
- CSM_{HIN}

An improved seq2seq method with CRF constraints

A state-of –the-art method, based on GCN

A classic joint method for extracting entities and triggers together.

The proposed model with ADJ. matrix

A latest joint method for extracting entities and triggers together.

The full proposed model with HIN

Experiments and Analysis

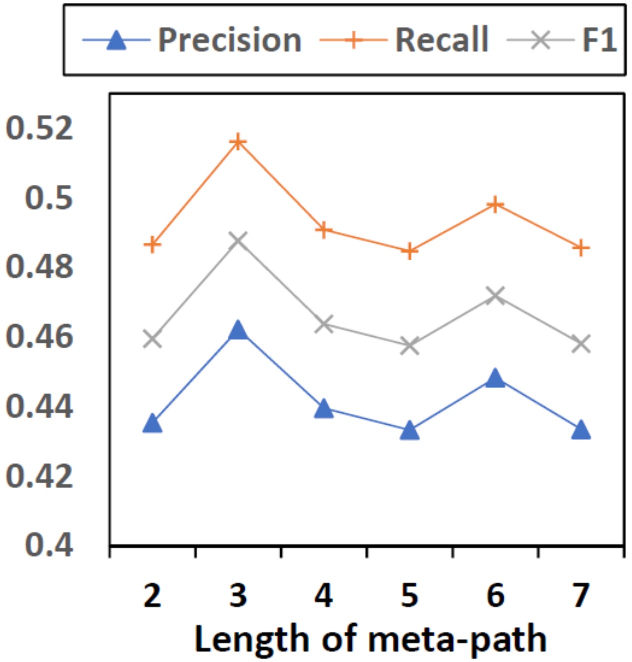
- Results.

Model	ACE 2005			NYT			CoNLL			WebNLG		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Seq2Seq	0.442±0.025	0.493±0.0272	0.466±0.026	0.818±0.012	0.832±0.012	0.825±0.012	0.709±0.015	0.852±0.011	0.774±0.013	0.851±0.009	0.910±0.007	0.880±0.008
CRF	0.434±0.031	0.478±0.033	0.455±0.032	0.813±0.011	0.828±0.011	0.821±0.01	0.718±0.016	0.867±0.013	0.785±0.014	0.864±0.005	0.921±0.005	0.892±0.005
GCN	0.435±0.030	0.487±0.032	0.459±0.031	0.804±0.013	0.819±0.013	0.811±0.013	0.706±0.015	0.871±0.014	0.780±0.013	0.884±0.008	0.931±0.008	0.907±0.008
JEE	0.423±0.023	0.468±0.030	0.443±0.026	0.717±0.009	0.645±0.014	0.679±0.012	0.713±0.019	0.814±0.013	0.76±0.015	0.775±0.015	0.818±0.012	0.796±0.013
JT	0.469±0.003	0.426±0.005	0.447±0.004	0.725±0.012	0.691±0.006	0.708±0.009	0.738±0.025	0.837±0.006	0.784±0.021	0.818±0.011	0.829±0.007	0.823±0.008
CSM_{DA}	0.455±0.024	0.494±0.022	0.474±0.023	0.835±0.012	0.847±0.012	0.841±0.012	0.730±0.017	0.856±0.021	0.788±0.019	0.908±0.005	0.941±0.004	0.924±0.004
CSM_{HIN}	0.477±0.030	0.533±0.033	0.503±0.031	0.859±0.007	0.870±0.008	0.865±0.008	0.754±0.018	0.890±0.020	0.816±0.017	0.923±0.004	0.953±0.003	0.937±0.003

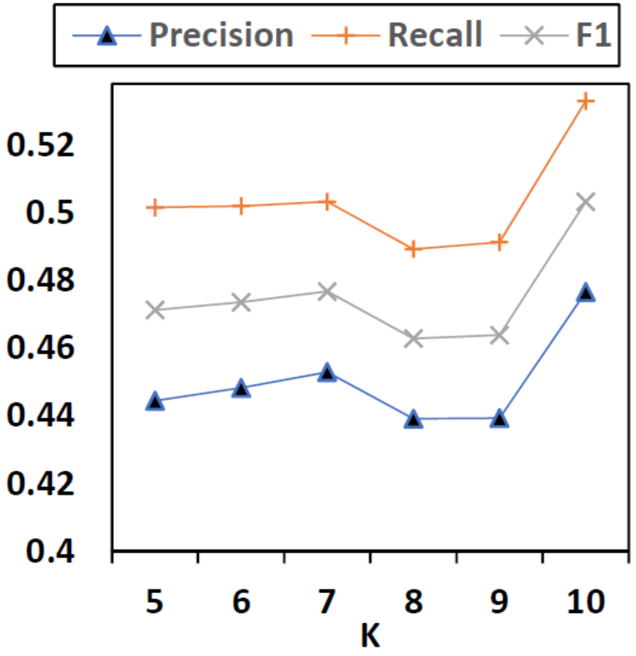
Model	Entity extraction			Trigger extraction		
	Precision	Recall	F1	Precision	Recall	F1
Seq2Seq	0.494	0.489	0.49	0.383	0.426	0.403
CRF	0.502	0.483	0.491	0.395	0.473	0.431
GCN	0.508	0.491	0.499	0.381	0.443	0.410
JEE	0.451	0.497	0.472	0.407	0.411	0.409
JT	0.492	0.458	0.474	0.447	0.414	0.432
CSM_{DA}	0.509	0.535	0.52	0.404	0.442	0.422
CSM_{HIN}	0.512	0.552	0.532	0.464	0.484	0.474

Our method excels other alternatives significantly

Experiments and Analysis



(a)



(b)

Fig. 3: Sensitivity in different parameters

Conclusion

- This is the first test to leverage this kind of information to the joint models.
- We verify the indirect entity-trigger co-occurred relationships are important to the joint-event-extraction task.
- Different from the distant supervision or other methods, this method tries to maximize the utilization of the currently available training data.

Thank You !

Any questions?

First author:
Yue Wang, wangyuecs@cufe.edu.cn
Department of Computer Science
Central University of Finance and Economics