Reasoning with Latent Structure Refinement for Document-level Relation Extraction

Guoshun Nan[1]*, Zhijiang Guo[1]*, Ivan Sekulić[2], Wei Lu[1]

[1]StatNLP, Singapore University of Technology and Design

[2]Università della Swizzera italiana

StatNLP Research
statnlp.org

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Università della Svizzera italiana

Sentence-level Relation Extraction

Input Sentence

Carey will succeed **Jack**, who held the position for 15 years and will take on a new role as **chairman**.

Given entity pair: <**Jack**,**chairman**>

Output

The relation: *per:title*

Document-level Relation Extraction

Input Document

Lutsenko is a former minister of **internal affairs**. **He** occupied this post in the cabinets of **Yulia Tymoshenko**. The ministry of **internal affairs** is the **Ukrainian** police authority.

Given entity pair: <**Yulia Tymoshenko**,**Ukrainian**>

Output

The relation: *country of citizenship*

StatNLP Research
statnlp.org

Document-level Relation Extraction

Challenge

Capturing the ***interactions*** of mentions of an entity ***across*** sentences, where an entity may involve ***multiple*** mentions.

Related Work

Sequence-based Models

Graph-based Models

Related Work

Sequence-based Models

Encoding the sentences of a document.

CNN

Zeng et al., 2014

RNN

Cai et al., 2016; Sorokin et al., 2017; Jia et al., 2019

Transformer

Verga et al., 2018

Related Work

Graph-based Models

Constructing document-level graphs for inference.

Syntax

Peng et al., 2017, Sahu et al., 2019

Attention

Veličković et al., 2019, Guo et al., 2019

Rules

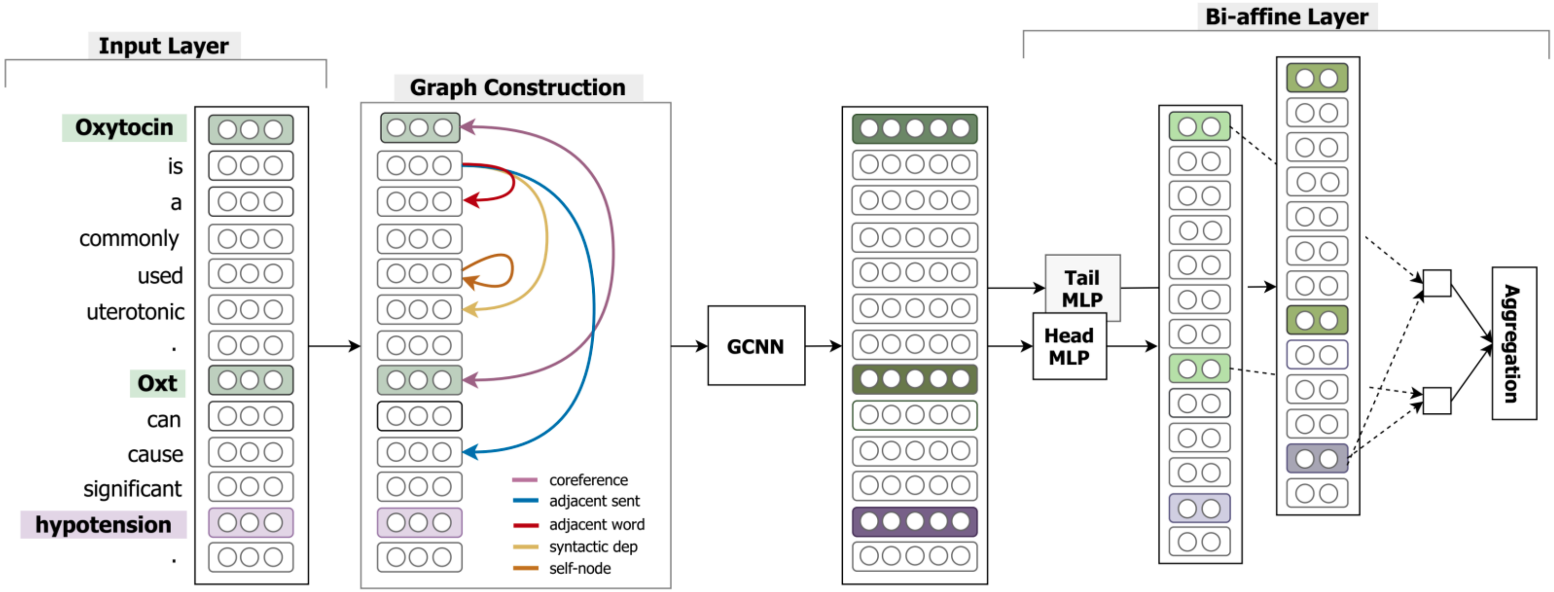
Christopoulou et al., 2019

Related Work

GCNN

Sahu et al., 2019

Constructing a graph based on **coreferences and syntax**.



Sahu et al., 2019. Inter-sentence relation extraction with document-level graph convolutional neural network. In *Proc. of ACL.*

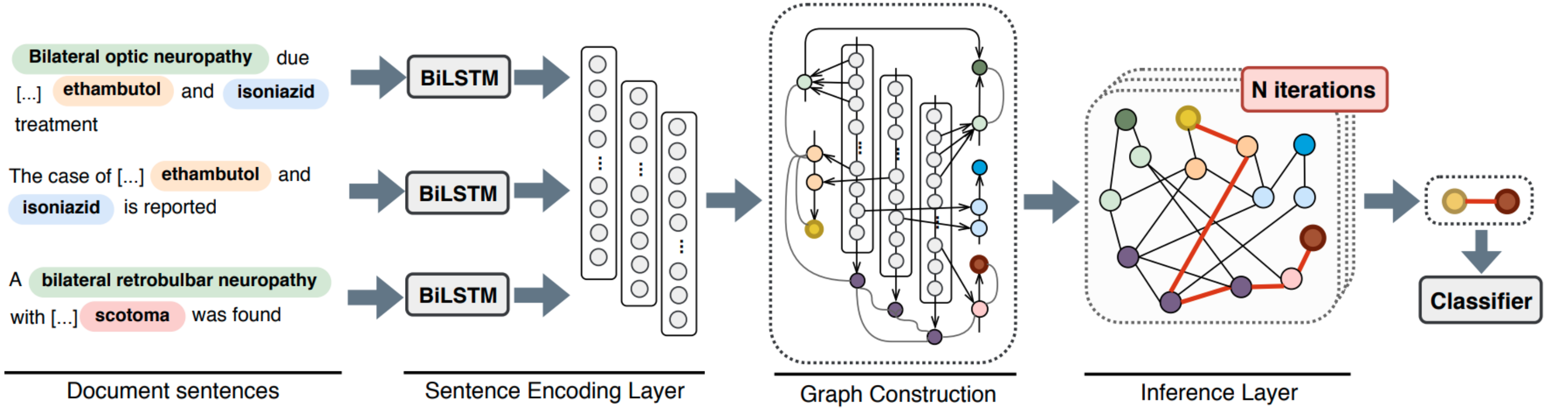
StatNLP Research
statnlp.org

Related Work

EoG

Christopoulou et al., 2019

Constructing a document graph based on a **predefined rule.**



Christopoulou et al., 2019. Connecting the dots: Document-level neural relation extraction with edge-oriented graphs. In *Proc. of EMNLP*.

Related Work

Summary

Graph-based models are able to capture more information that is obscure from plain text.

Structural information is effective for the relation extraction task.

Motivation

Previous works rely on dependency tree, coreference or rules to construct a static graph, and they may not capture rich interactions of entities and mentions in a document.

Can we **learn a task-specific structure** of a document from data to capture the interactions between entities ?

Motivation

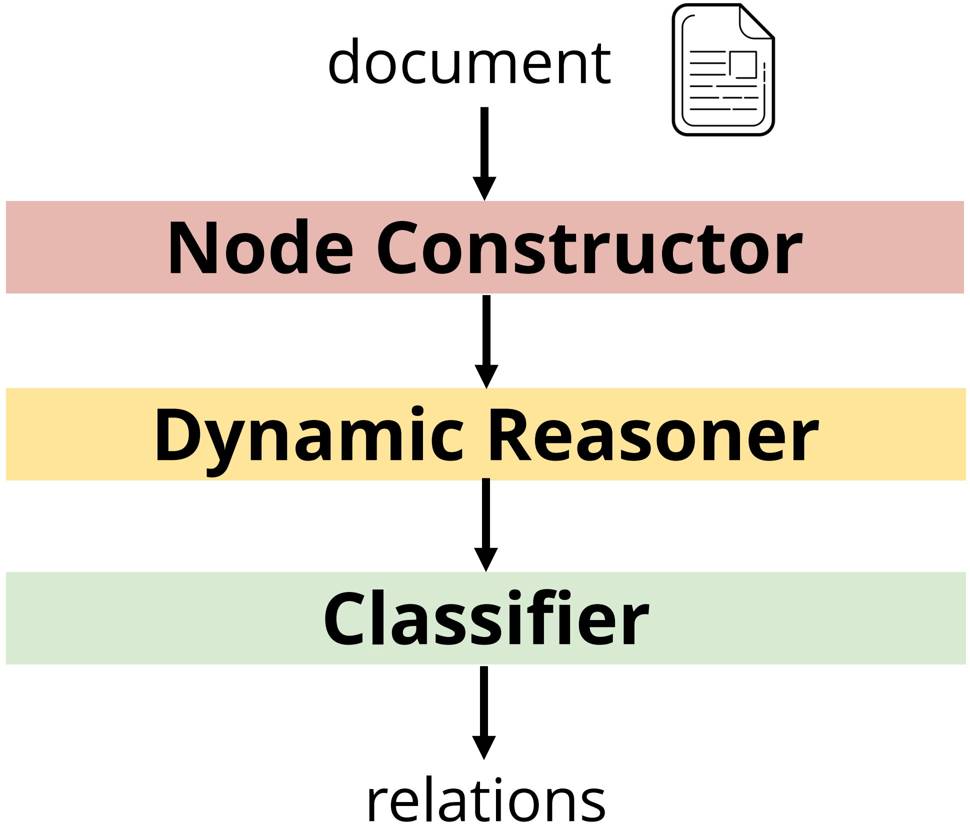
To induce a task-specific structure of a document:

How can we define the nodes of the graph?

How can we learn a task-specific structure?

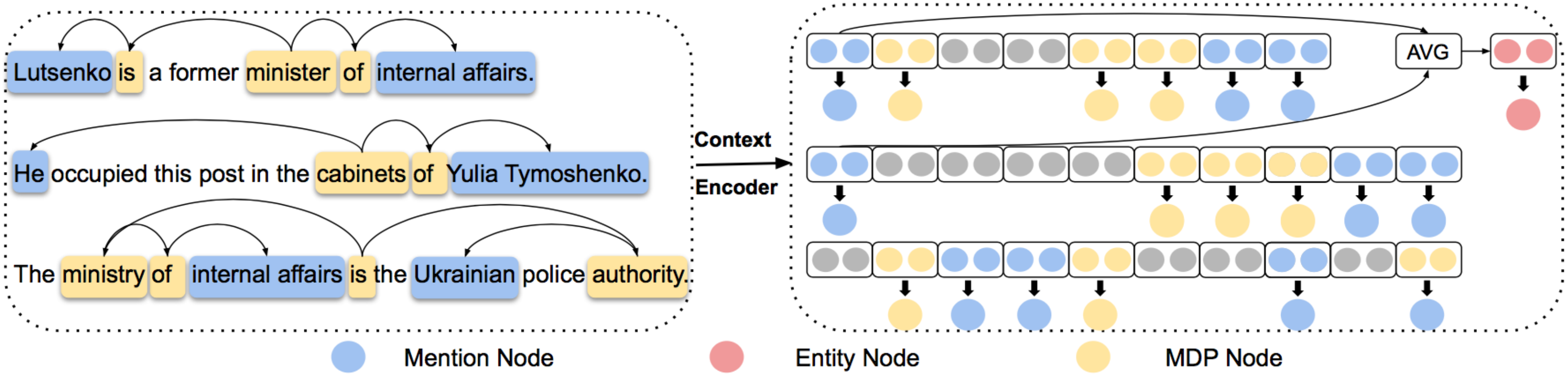
Model

We define three types of nodes for a document graph, and treat the graph structure as a latent variable and induce it in an end-to-end manner.



Node Constructor

Defining nodes for document-level graph



Mention Node: different mentions of entities in each sentence.

Entity Node: the average of its mention representations.

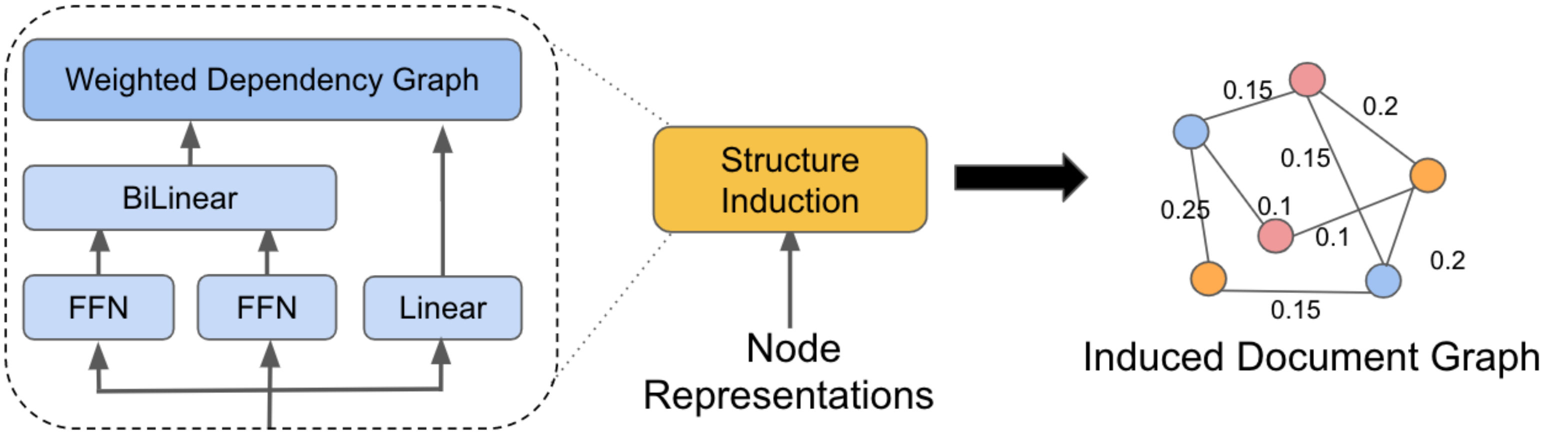
MDP(Meta Dependency Paths) node: a set of shortest dependency paths for all mentions in a sentence.

StatNLP Research
statnlp.org

Dynamic Reasoner

1. Structure Induction

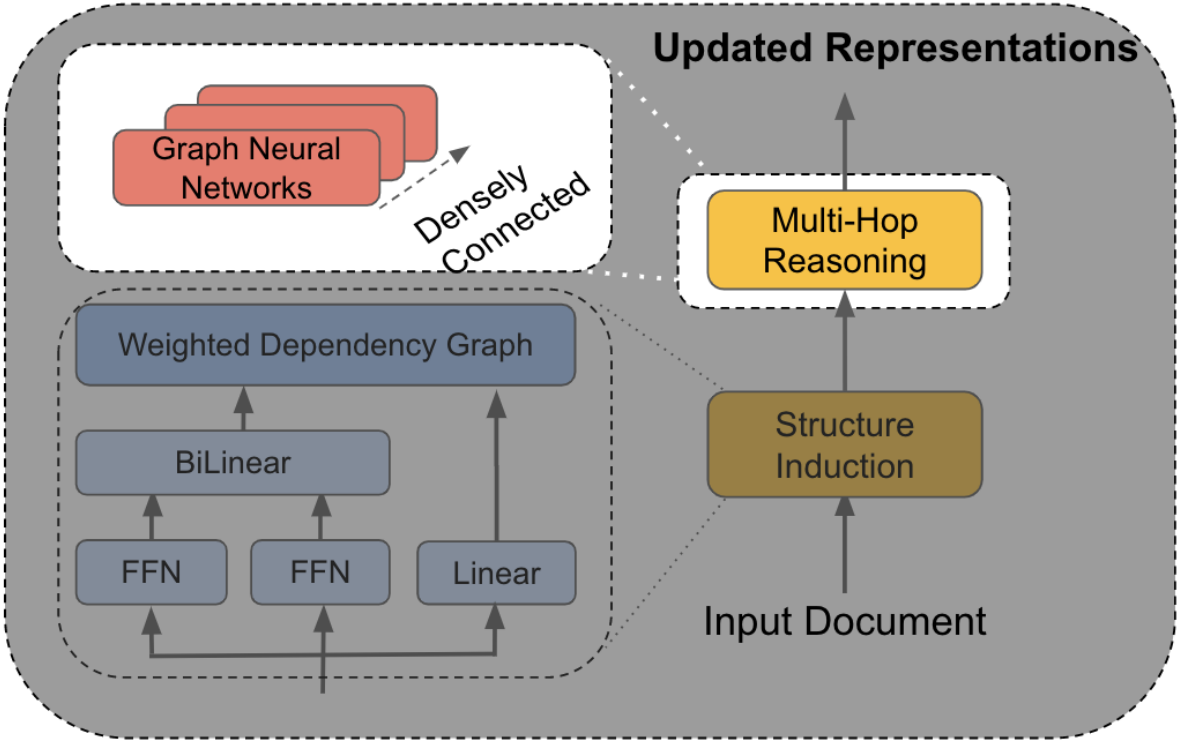
We opt for ***Matrix-Tree Theorem(MTT)*** to **induce a latent structure**, which is considered as **an adjacency matrix** of the document graph.



Dynamic Reasoner

2. Multi-hop Reasoning

We use densely connected graph convolution networks to **capture more structural information** on a large document graph.

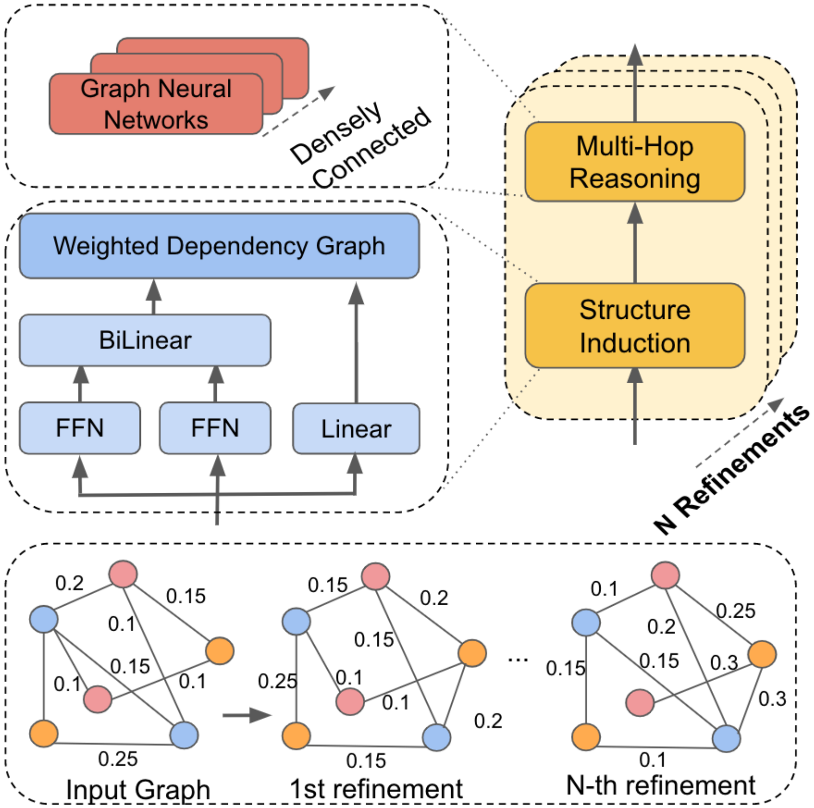


StatNLP Research
statnlp.org

Dynamic Reasoner

3. Iterative Refinement

We **stack *N*** blocks of the reasoner to **repeatedly refine** the document graph ***N*** times to generate a more informative structure.



StatNLP Research
statnlp.org

Experiments

Dataset

Dataset	Relation Types	Train	Dev	Test
DocRED	96	3,053	1,000	1,000
Chemical- Disease Reactions (CDR)	2	500	500	500
Gene-Disease Associations (GDA)	2	23,354	5,838	1,000

Experiments

Main Results on the DocRED Dataset

Model	**Type**	***Inter-F1***	***F1***
CNN (Yao et al., 2019)	Sequence-based	37.58	42.26
BiLSTM (Yao et al., 2019)	Sequence-based	43.49	51.06
ContextAware (Yao et al., 2019)	Sequence-based	42.46	50.70
GAT (Veličković et al., 2019)	Graph-based	43.94	49.51
GCNN (Sahu et al., 2019)	Graph-based	44.11	51.62
EoG (Christopoulou et al., 2019)	Graph-based	44.60	51.82
AGGCN (Guo et al,. 2019)	Graph-based	45.45	51.45

Graph-based models perform better than sequence-based models.

Experiments

Main Results on the DocRED Dataset

Model	**Type**	***Inter-F1***	***F1***
GAT (Veličković et al., 2019)	Graph-based	43.94	49.51
GCNN (Sahu et al., 2019)	Graph-based	44.11	51.62
EoG (Christopoulou et al., 2019)	Graph-based	44.60	51.82
AGGCN (Guo et al,. 2019)	Graph-based	45.45	51.45
Glove + LSR	Graph-based	**48.35**	**54.18**
BERT (Wang et al., 2019)	Sequence-based	47.15	53.20
Two-Phase BERT (Wang et al.,2019)	Sequence-based	47.28	53.92
BERT + LSR	Graph-based	**52.05**	**59.05**

LSR performs much better as it **dynamically refine a latent structure**, without relying on a static graph.

Experiments

Results on the biomedical dataset CDR

Model	**Type**	***F1***
Gu et al. (2017)	Sequence-based	61.3
Nguyen et al. (2018)	Sequence-based	62.3
Verga et al. (2018)	Sequence-based	62.1
Sahu et al. (2019)	Graph-based	58.6
Christopoulou et al. (2019)	Graph-based	63.6
LSR	Graph-based	61.2
LSR w/o MDP nodes	Graph-based	**64.8**

Experiments

Results on the biomedical dataset GDA

Model	**Type**	***F1***
NoInf (Christopoulou et al., 2019)	Graph-based	74.6
Full (Christopoulou et al., 2019)	Graph-based	80.8
EoG (Christopoulou et al., 2019)	Graph-based	81.5
LSR	Graph-based	79.6
LSR w/o MDP nodes	Graph-based	**82.2**

Conclusion

We treat the document-level graph **as a latent variable** and induce **the task-specific structure** in an end-to-end fashion.

We compare our model with the state-of-the-art models and perform deep analysis to show the **effectiveness of the latent structure**.

Thank You

Code Available:
https://github.com/nanguoshun/LSR