Graph Neural Networks in NLP

Yue Zhang

School of Engineering, Westlake University

西湖大學
WESTLAKE UNIVERSITY

Outline

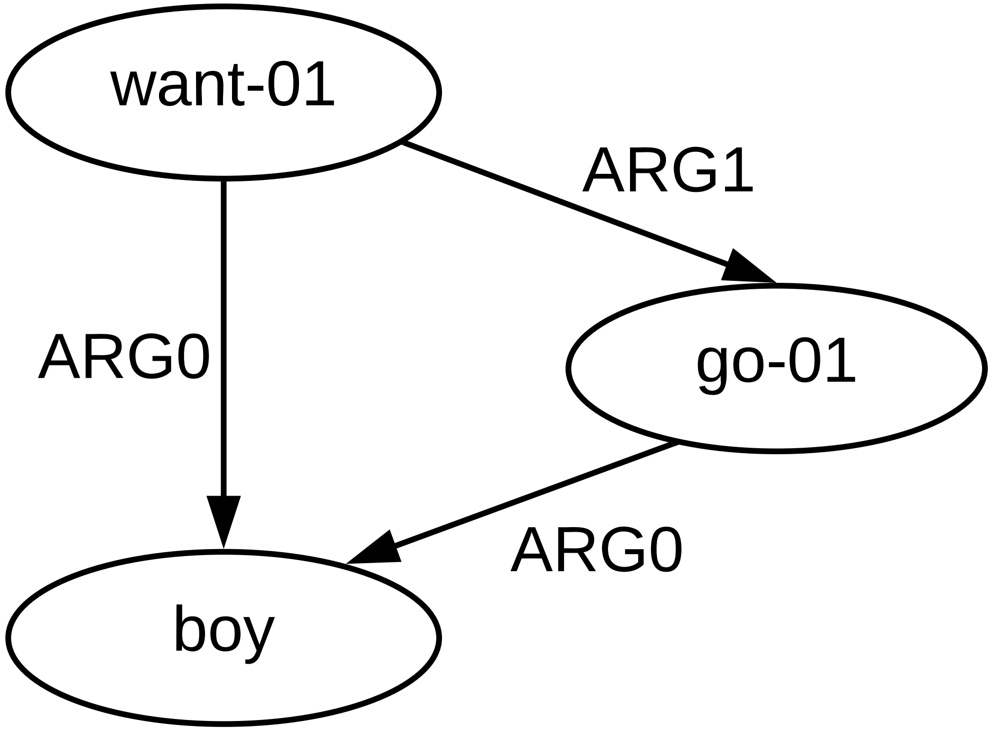
- Graph Structures in NLP
- Two major models for representing graphs
- Solving problems using graph encoding

Outline

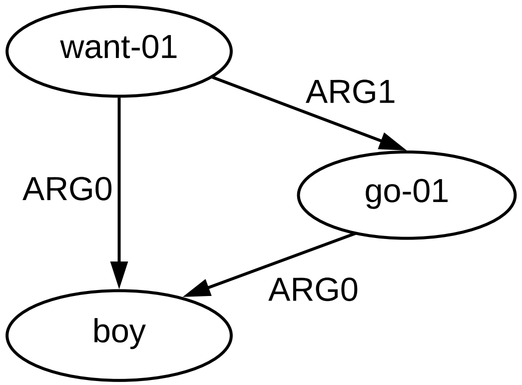
- Graph Structures in NLP
- Two major models for representing graphs
- Solving problems using graph encoding

Graphical Structures in NLP

Abstract Meaning Representation

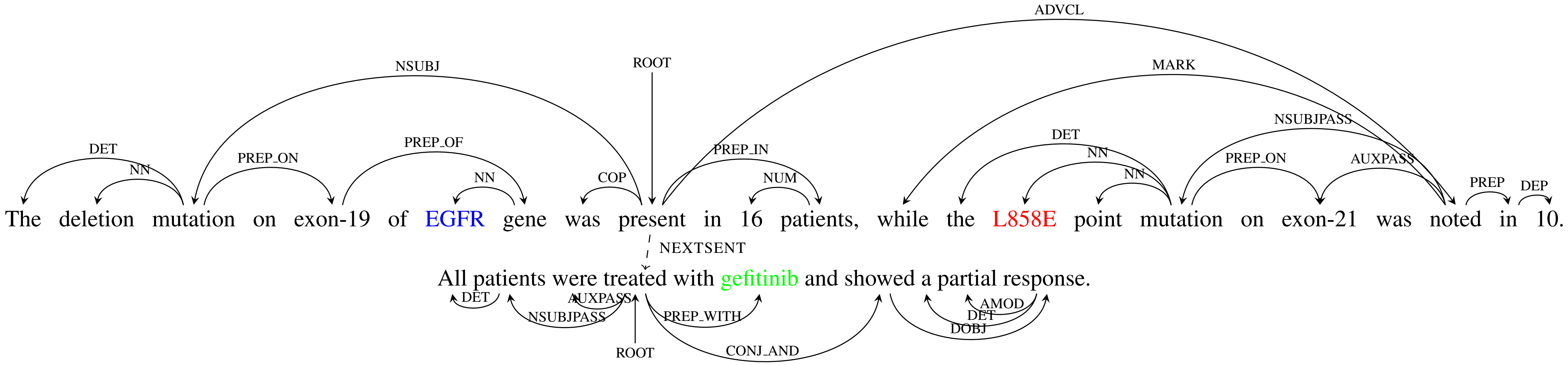


Graphical Structures in NLP

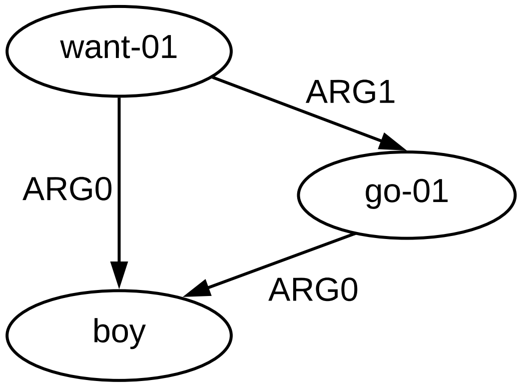


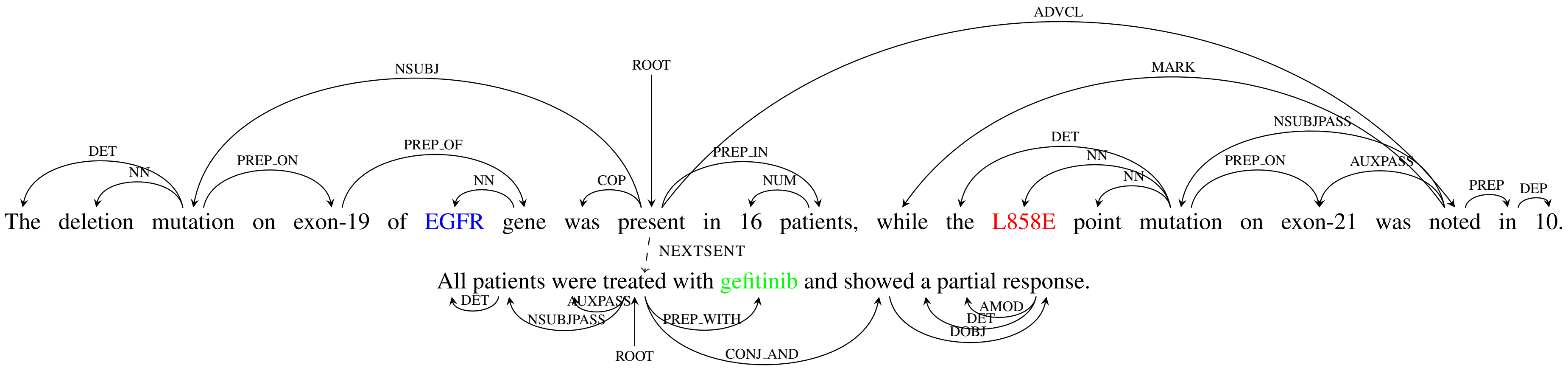
Graphical Structures in NLP

Dependency trees and discourse relations between sentences



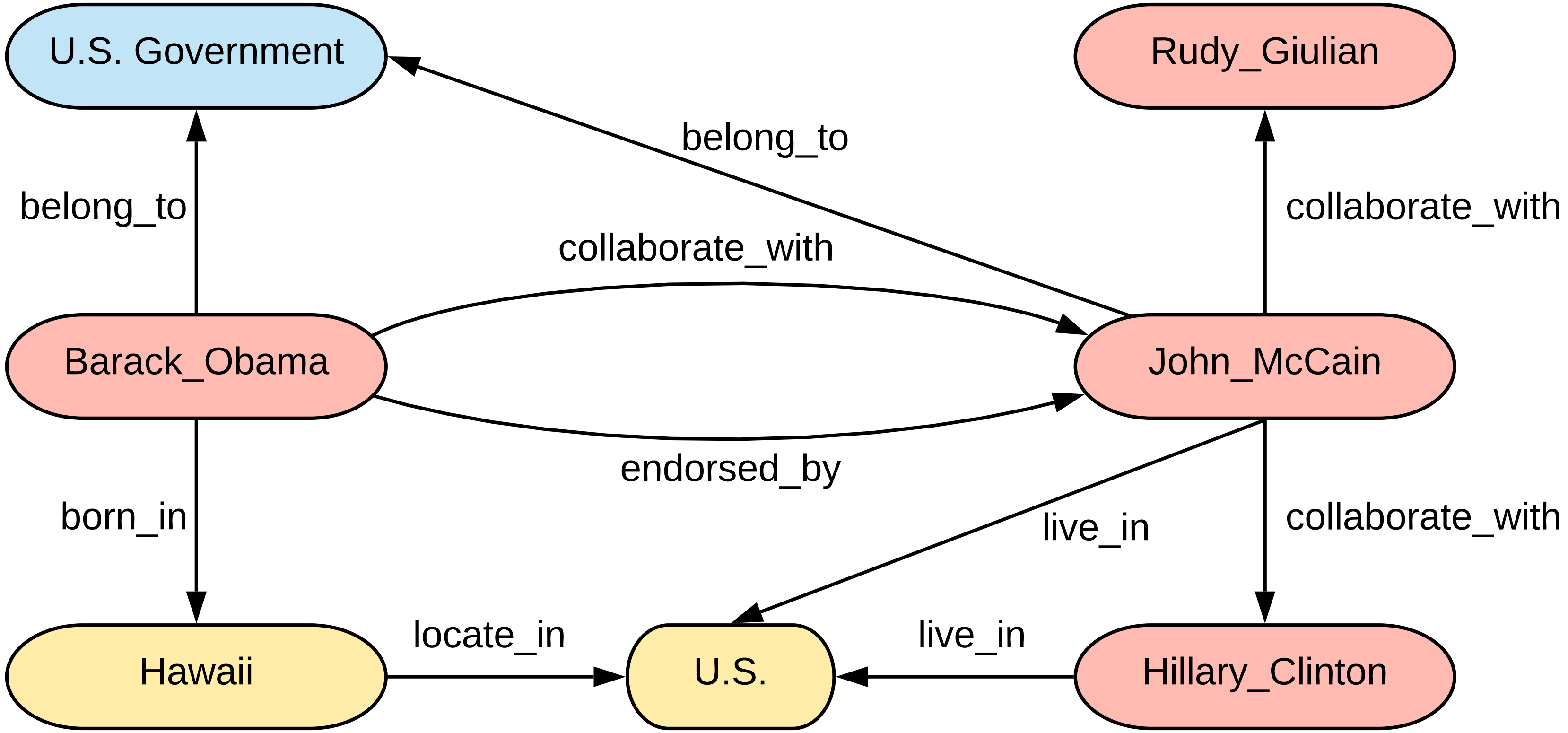
Graphical Structures in NLP



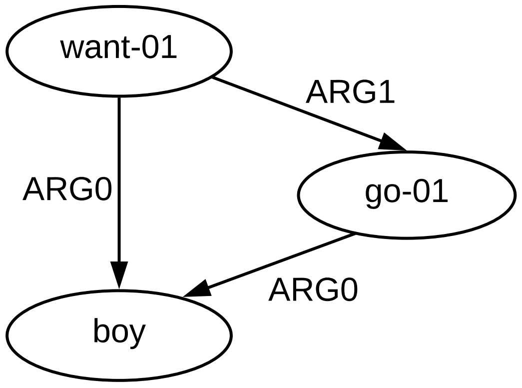


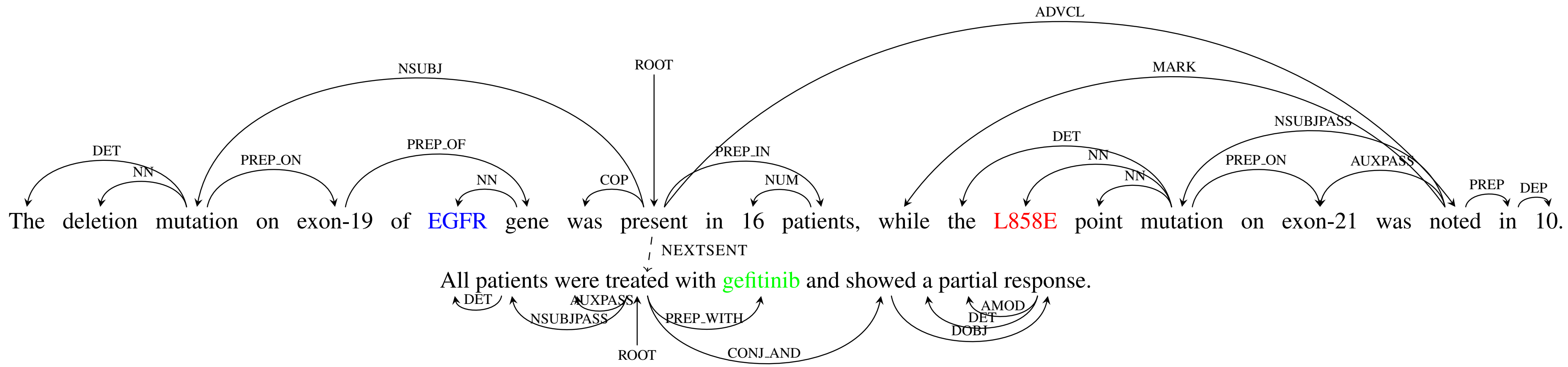
Graphical Structures in NLP

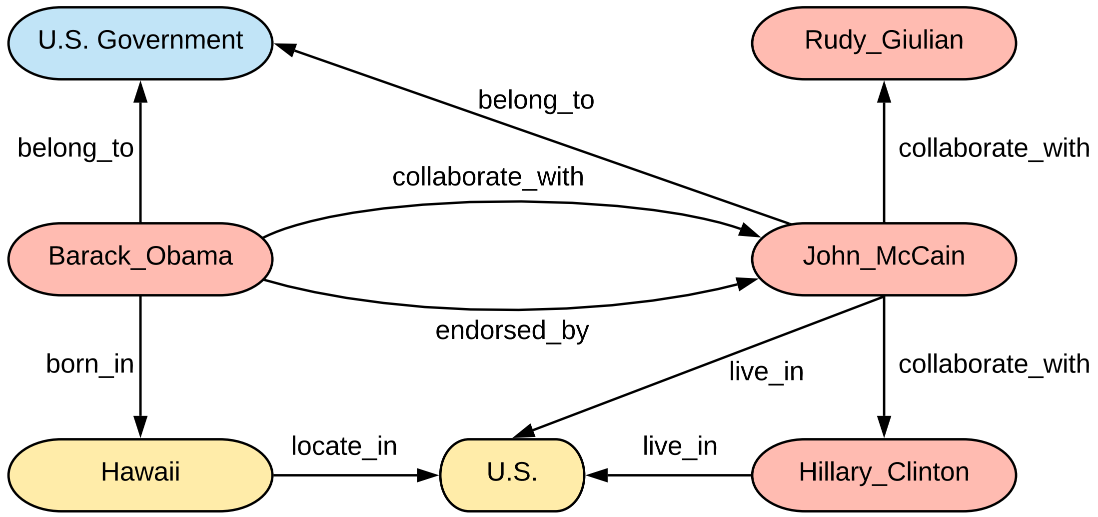
Knowledge



Graphical Structures in NLP

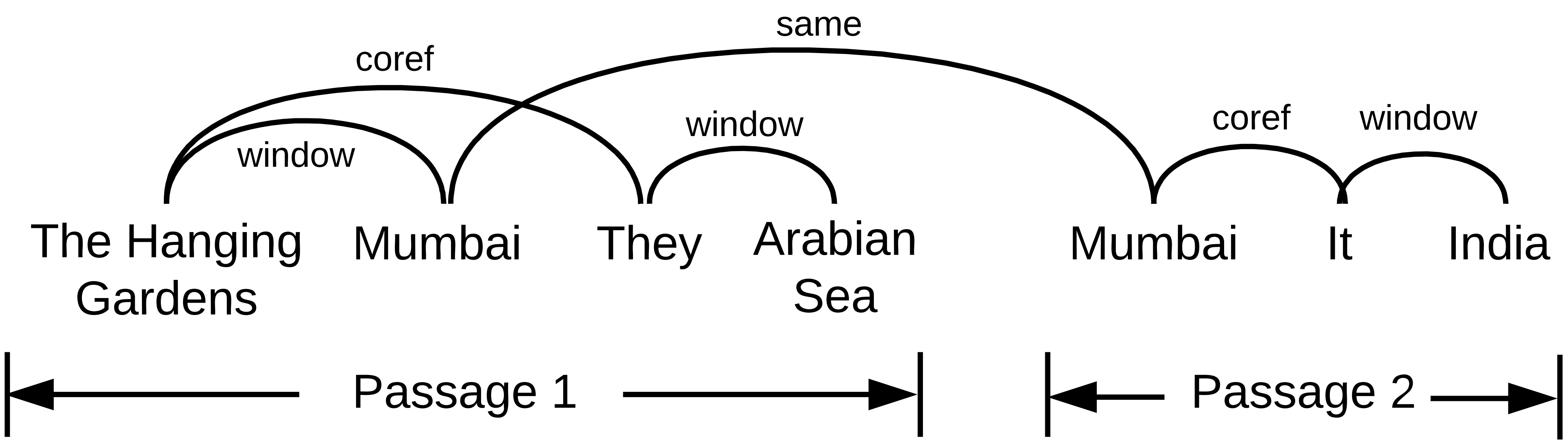




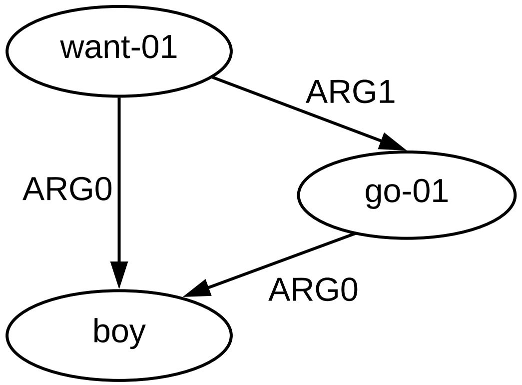


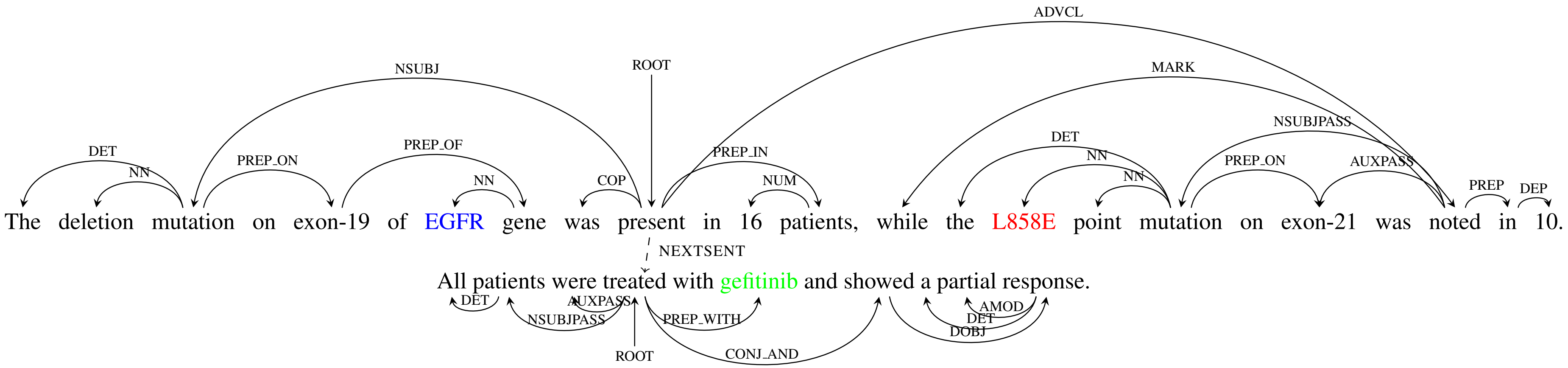
Graphical Structures in NLP

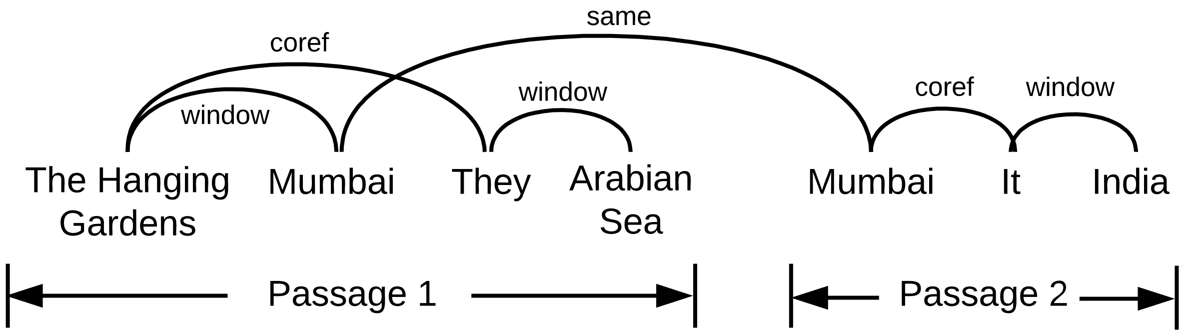
Entity and co-reference link

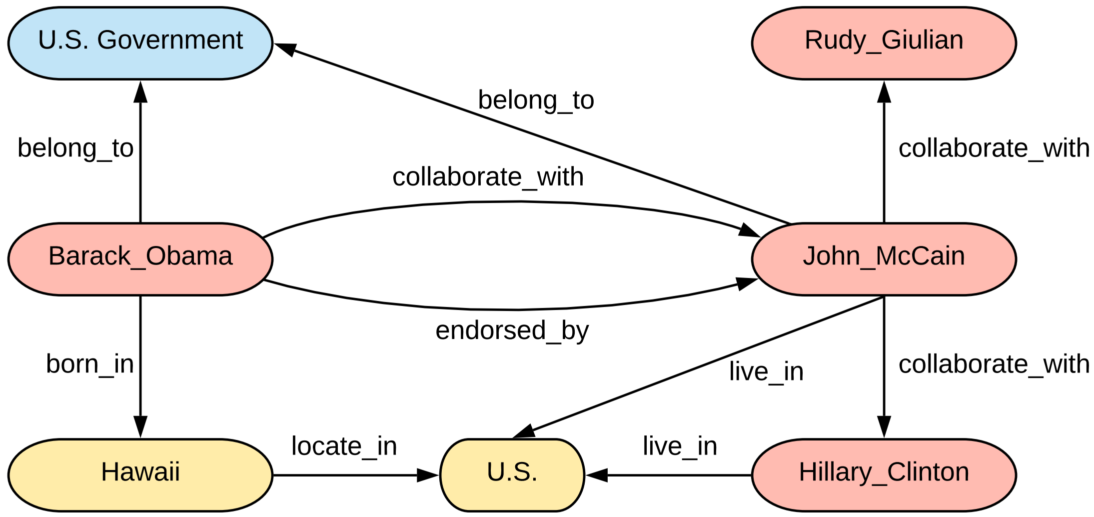


Graphical Structures in NLP



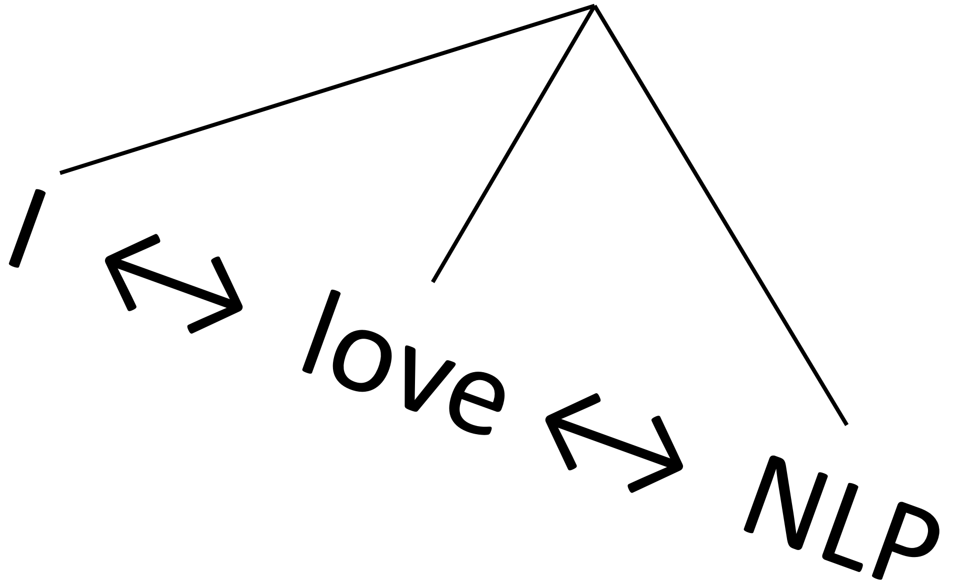




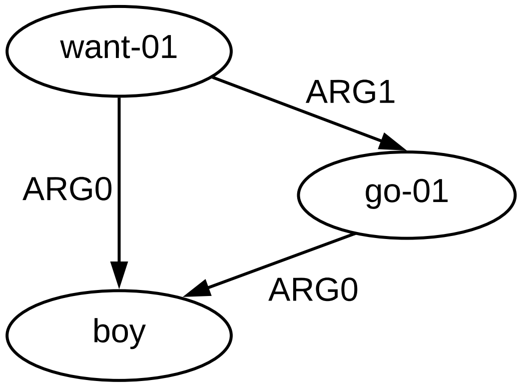


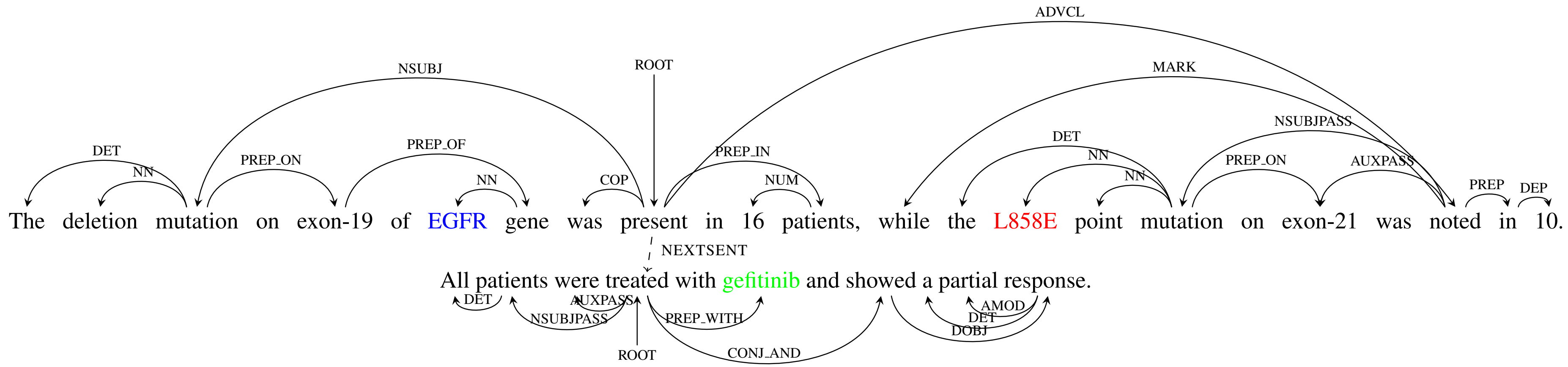
Graphical Structures in NLP

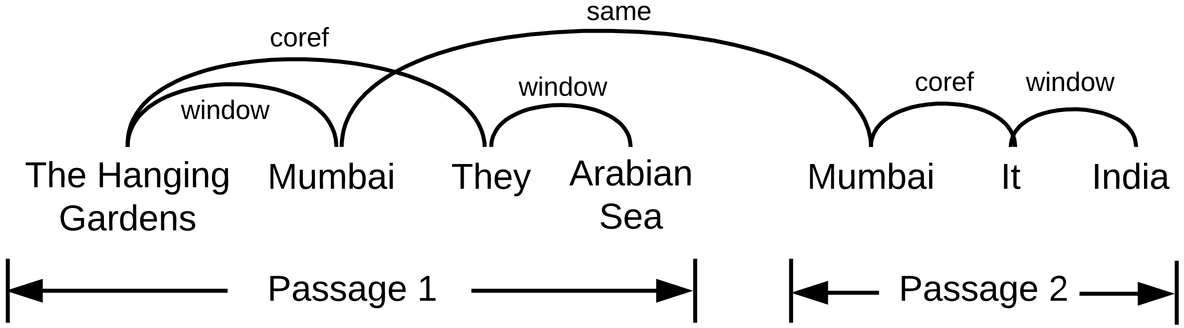
Sentence



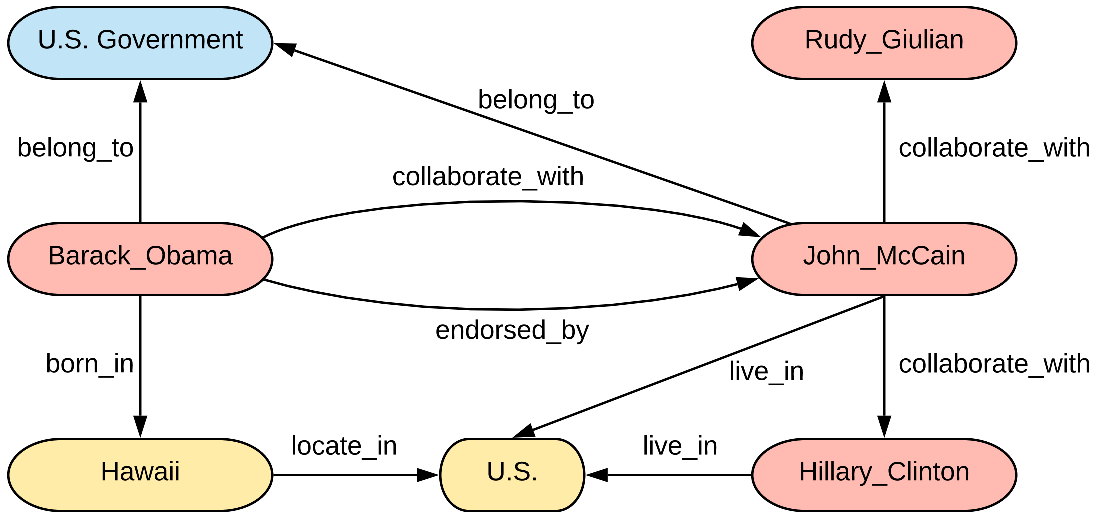
Graphical Structures in NLP



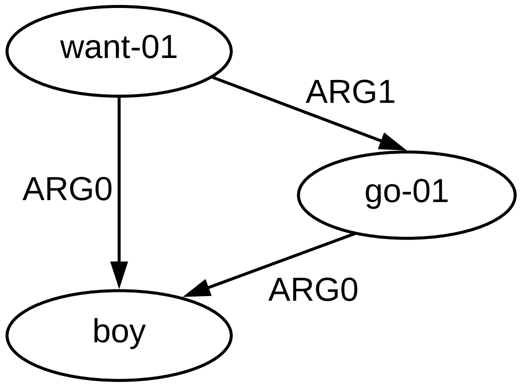


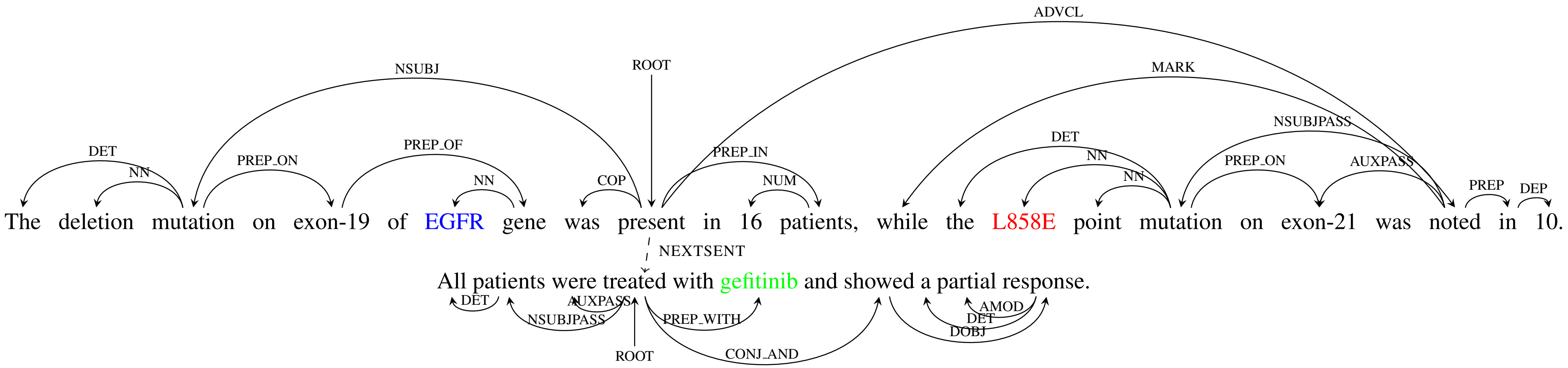


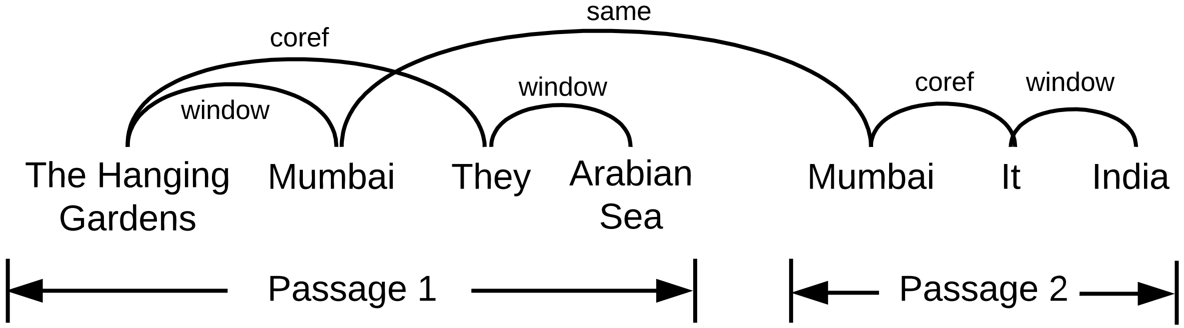
I ↔ love ↔ NLP

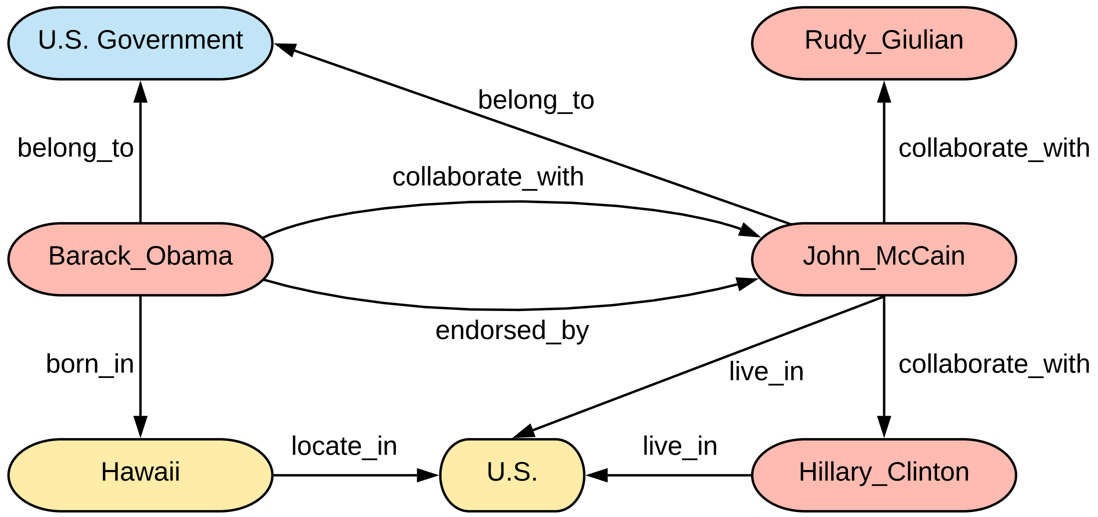


Graphical Structures in NLP





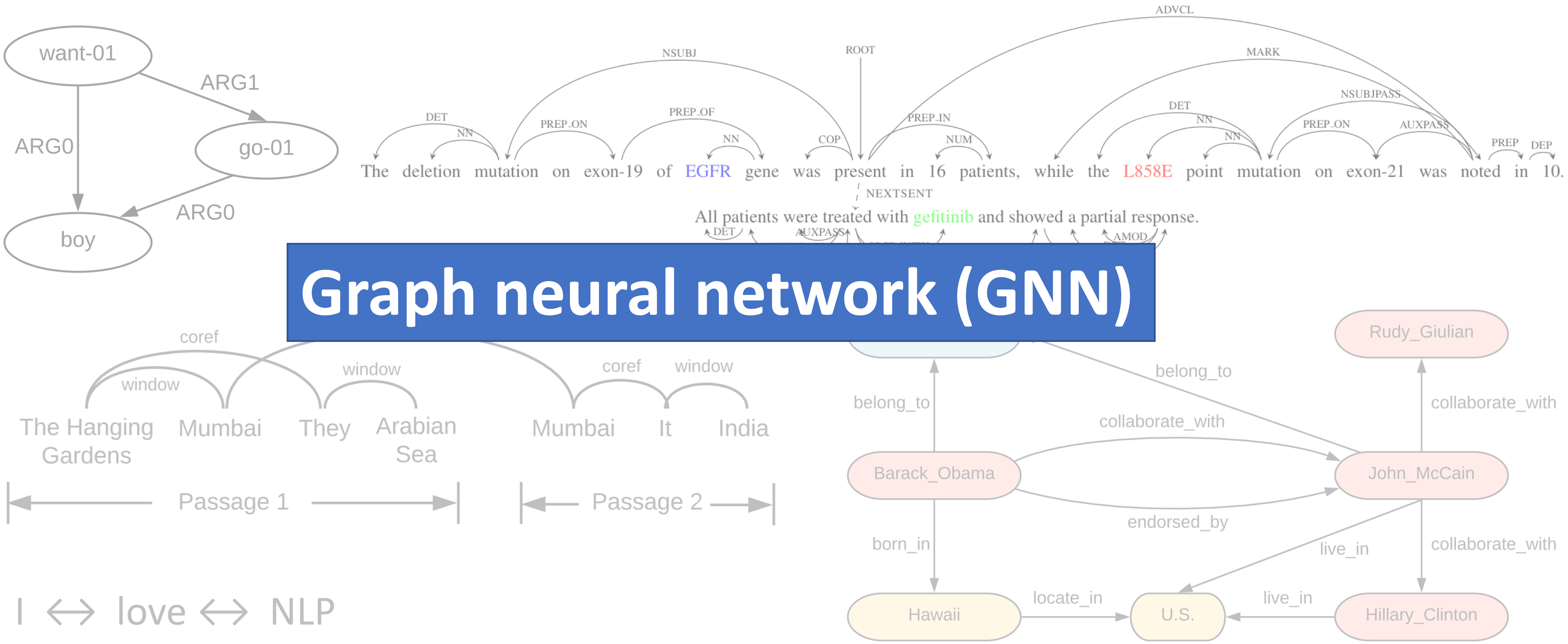




I ↔ love ↔ NLP

• • •

Graphical problems in NLP



Graphical problems in NLP

Song et al., ACL 2018
Song et al., TACL 2019
Ding et al., ACL 2019
Tu et al., ACL 2019

Song et al., EMNLP 2018
Zhu et al., ACL 2019
Guo et al., ACL 2019

Bastings et al., EMNLP 2017
Marcheggiani et al., NAACL 2018
Marcheggiani and Ivan Titov, EMNLP 2017

Graph neural network (GNN)

...

Song et al., Arxiv preprint 2018
Cao et al., NAACL 2019
Xiao et al., NAACL 2019

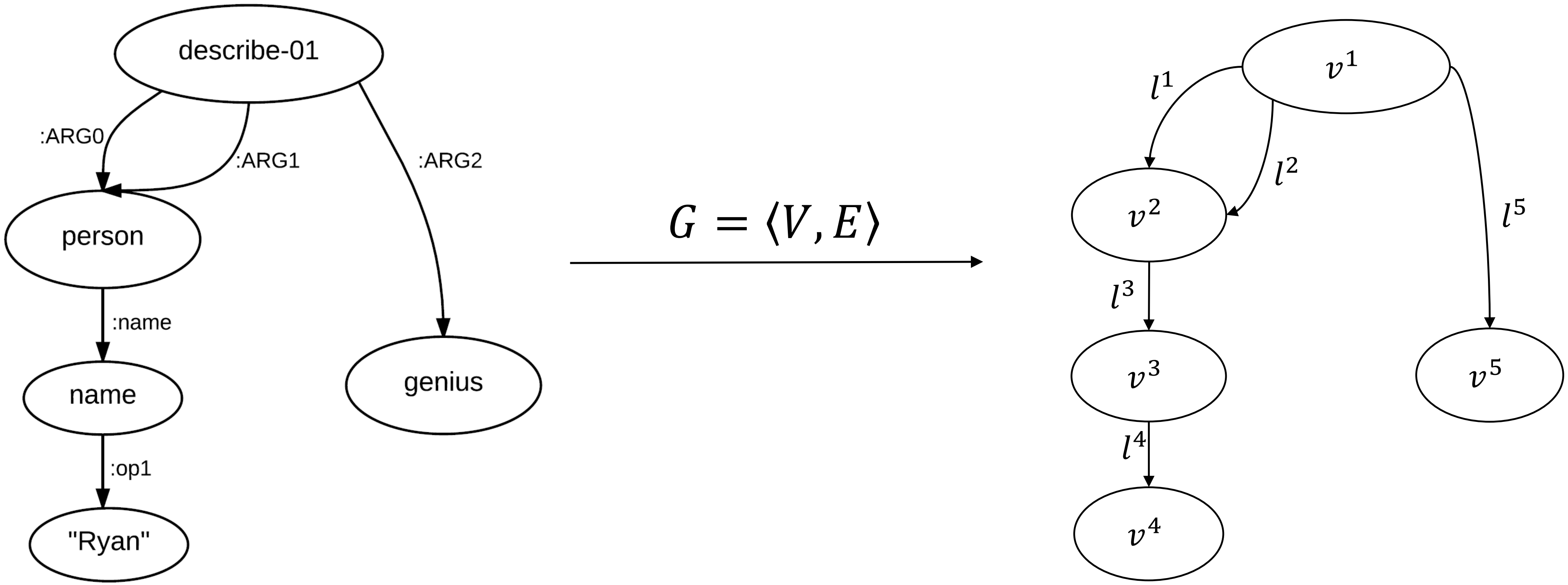
Damonte and Cohen, NAACL 2019
Qian et al., NAACL 2019
Koncel-Kedziorski et al., NAACL 2019

I ↔ love ↔ NLP Zhang et al., ACL 2018

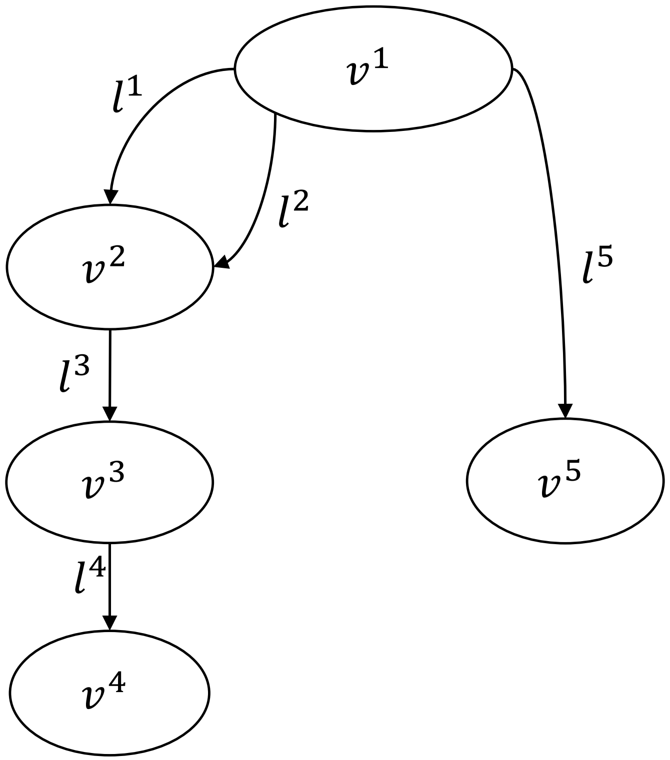
Outline

- Graph Structures in NLP
- Three major models for representing graphs
- Solving problems using graph encoding

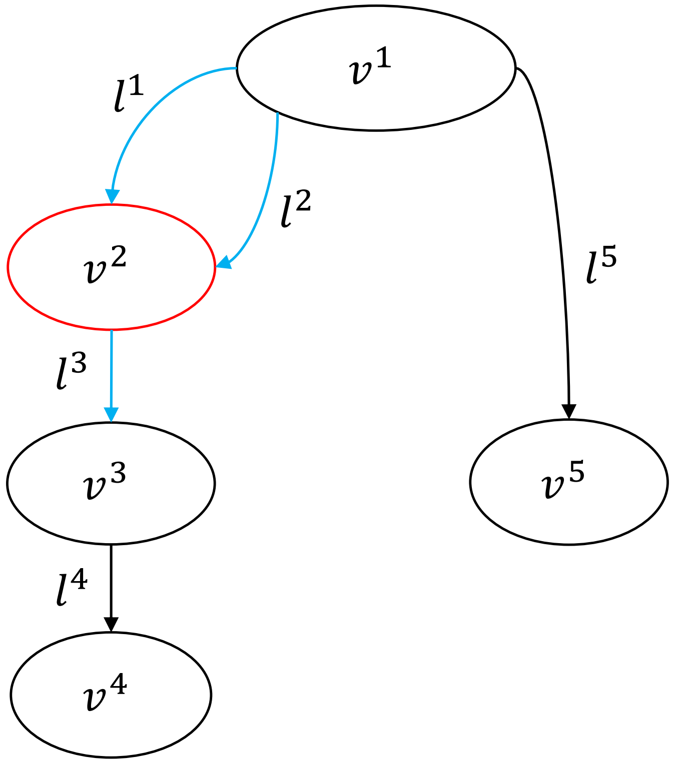
GNNs


describe-01
:ARG0
:ARG1
:ARG2
person
:name
name
genius
:op1
"Ryan"
$G = \langle V, E \rangle$
v^1
l^1
l^2
l^5
v^2
l^3
v^3
v^5
l^4
v^4

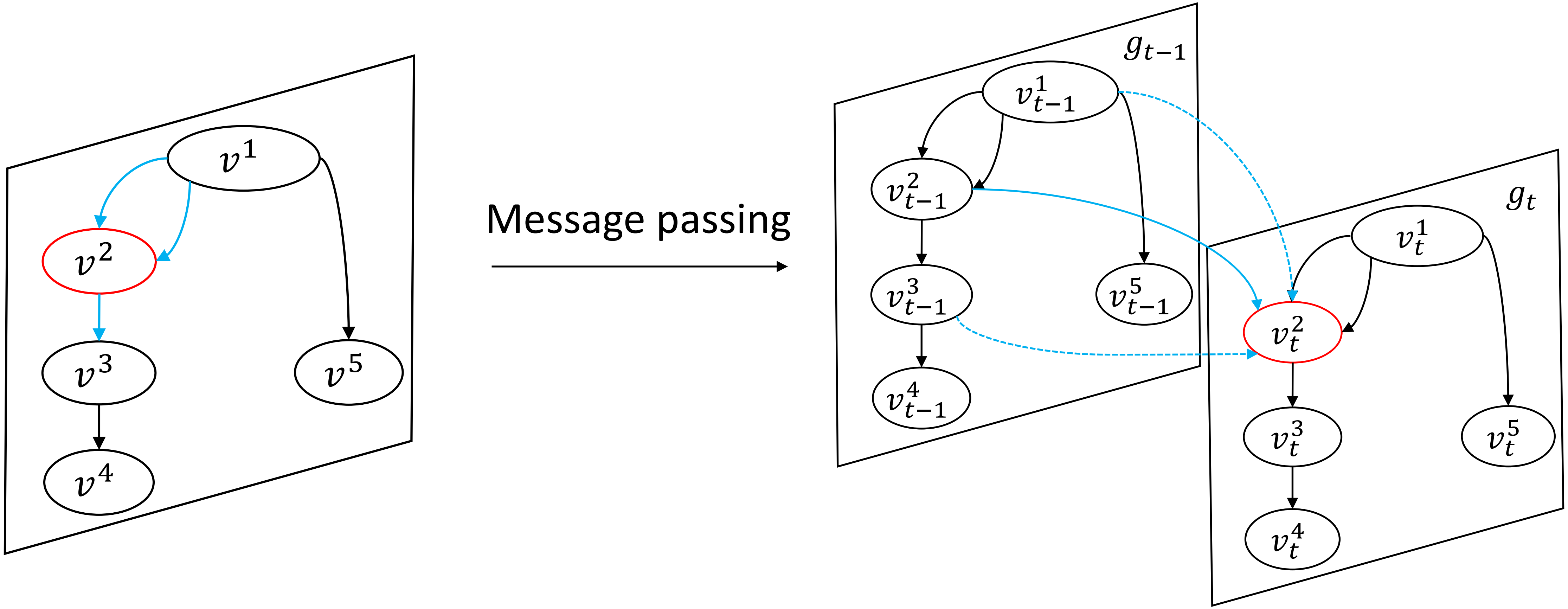
GNNs



For certain vertex

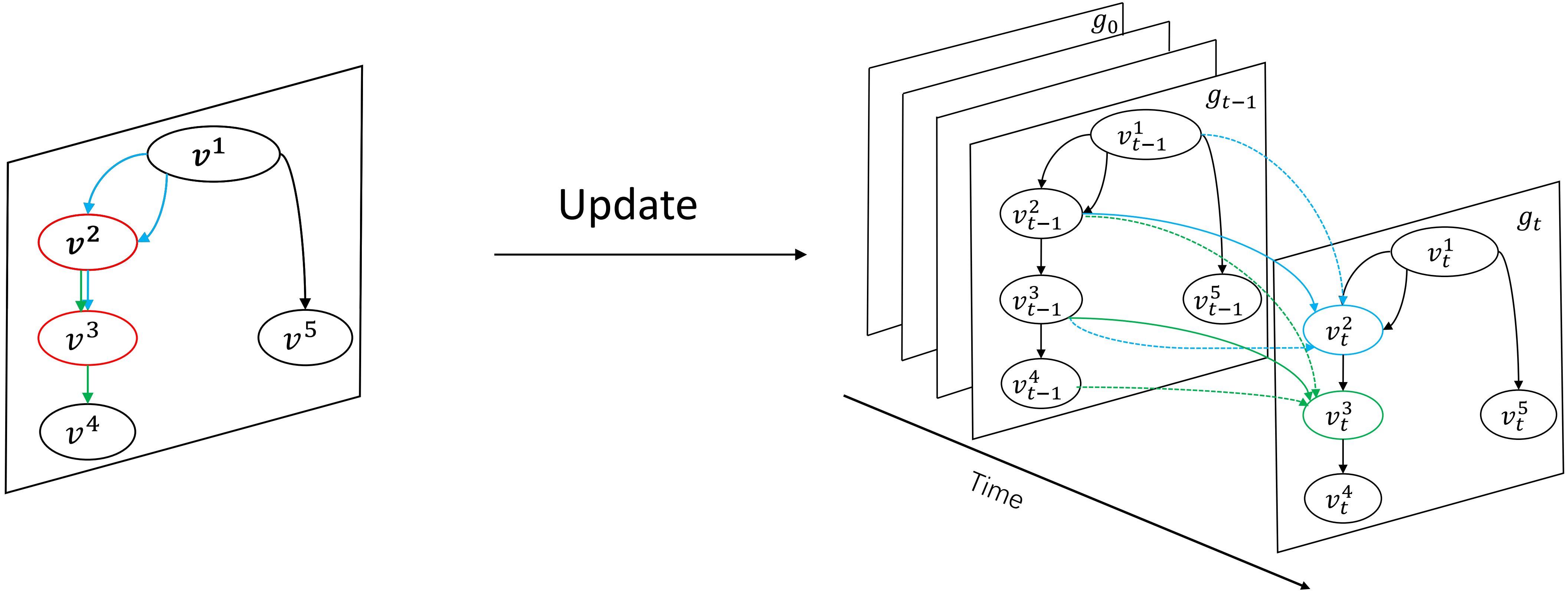


GNNs



v^1
v^2
v^3
v^4
v^5
Message passing
g_{t-1}
v_{t-1}^1
v_{t-1}^2
v_{t-1}^3
v_{t-1}^4
v_{t-1}^5
g_t
v_t^1
v_t^2
v_t^3
v_t^4
v_t^5

GNNs



v^1
v^2
v^3
v^4
v^5
Update
g_0
g_{t-1}
v_{t-1}^1
v_{t-1}^2
v_{t-1}^3
v_{t-1}^4
v_{t-1}^5
g_t
v_t^1
v_t^2
v_t^3
v_t^4
v_t^5
Time

Comparing GRN with other GNNs

	GRN (ACL 2018)	GCN (EMNLP 2017)	GAT (ICLR 2018)
Message calculation:	$m_j^t = \sum_{i \in N_j} h_i^{t-1}$		$\boldsymbol{a} = \alpha\left(h_j^{t-1}, h_i^{t-1}\right)$
State update:	$h_j^t, c_j^t = LSTM(m_j^t, [h_j^{t-1} c_j^{t-1}])$	$h_j^t = \sigma(W m_j^t + b)$	$h_j^t = \sigma(\sum_{i \in N_j} \boldsymbol{a} W h_i^{t-1})$
State memory:	both h and c	only h	only h

GRN: Linfeng song, Yue Zhang et al., A Graph-to-Sequence Model for AMR-to-Text Generation. ACL 2018.
GCN: Joost Basting, Ivan Titov et al., Graph Convolutional Encoders for Syntax-aware Neural Machine Translation, EMNLP 2017.
GAT: Peter Velickovic, Guillem Cucurull et al., Graph Attention Networks, ICLR 2018.

Comparing GRN with other GNNs

	GRN (ACL 2018)	GCN (EMNLP 2017)	GAT (ICLR 2018)
Message calculation:	$m_j^t = \sum_{i \in N_j} h_i^{t-1}$		$\boldsymbol{a} = \alpha\left(h_j^{t-1}, h_i^{t-1}\right)$
State update:	$h_j^t, c_j^t = LSTM(m_j^t, [h_j^{t-1} c_j^{t-1}])$	$h_j^t = \sigma(W m_j^t + b)$	$h_j^t = \sigma(\sum_{i \in N_j} \boldsymbol{a} W h_i^{t-1})$
State memory:	both h and c	only h	only h

Numerous variations in the representation of edges, vertexes, and graph information propagation

GRN: Linfeng song, Yue Zhang et al., A Graph-to-Sequence Model for AMR-to-Text Generation. ACL 2018.
GCN: Joost Basting, Ivan Titov et al., Graph Convolutional Encoders for Syntax-aware Neural Machine Translation, EMNLP 2017.
GAT: Peter Velickovic, Guillem Cucurull et al., Graph Attention Networks, ICLR 2018.

Outline

- Graph Structures in NLP
- Two major models for representing graphs
- Solving problems using graph encoding

Solving problems using graph encoding

- Neural Machine Translation
- Text Generation
- Question Answering
- Information Extraction
- Sentiment
- Social Classification
- Syntactic Parsing
- Semantic Role Labeling
- Word Embedding
- Sentence Representation

Solving problems using graph encoding

- Neural Machine Translation
- Text Generation
- Question Answering
- Information Extraction
- Sentiment
- Social Classification
- Syntactic Parsing
- Semantic Role Labeling
- Word Embedding
- Sentence Representation

Syntax-aware Neural Machine Translation

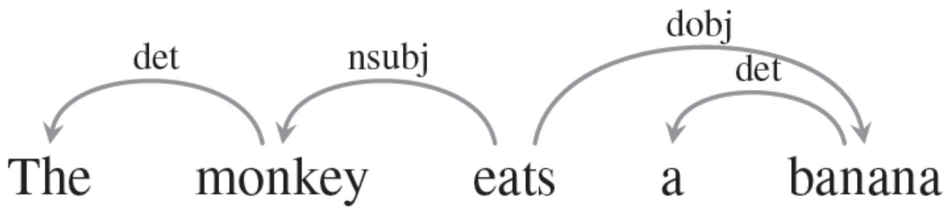
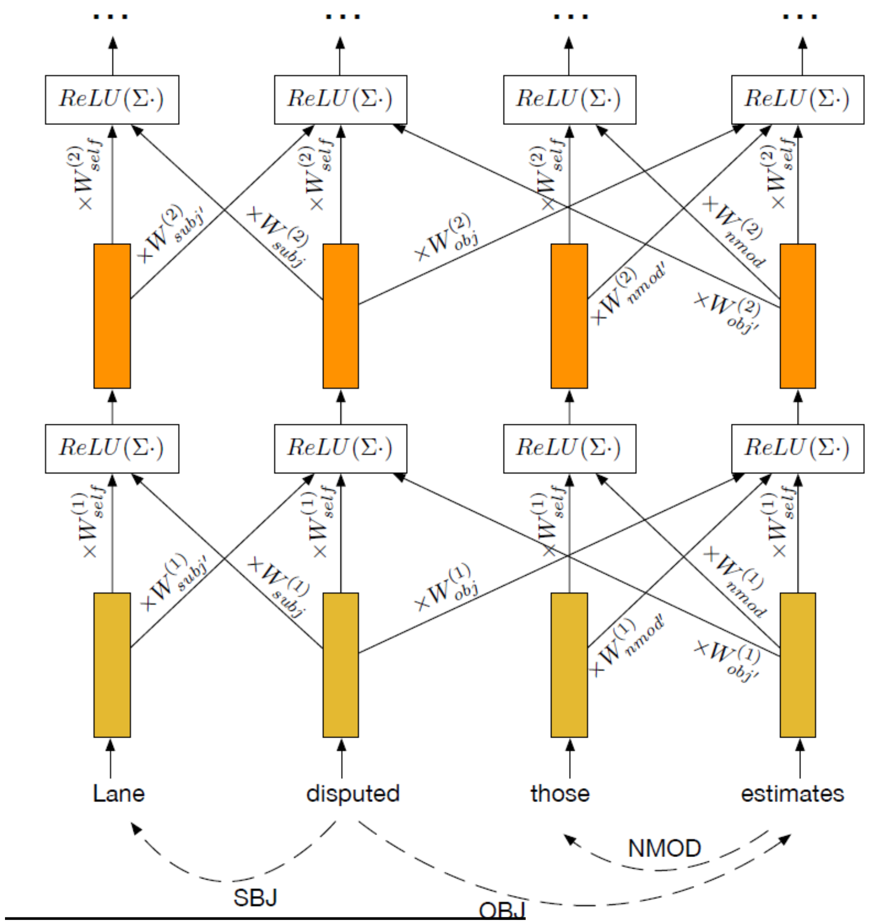


Figure 1: A dependency tree for the example sentence: "*The monkey eats a banana.*"

The goal is to provide the encoder with access to rich syntactic information but let it decide which aspects of syntax are beneficial for NMT, because rigid constraints on the syntactic typically hurt MT.

Joost Bastings, Ivan Titov et al., Graph Convolutional Encoders for Syntax-aware Neural Machine Translation, EMNLP 2017.

Syntactic GCNs



Standard GCN layer

$$h_v^{(k+1)} = ReLU\left(\sum_{u\in\mathcal{N}(v)} W^{(k)} h_u^{(k)} + b^{(k)}\right)$$

Syntactic GCN layer

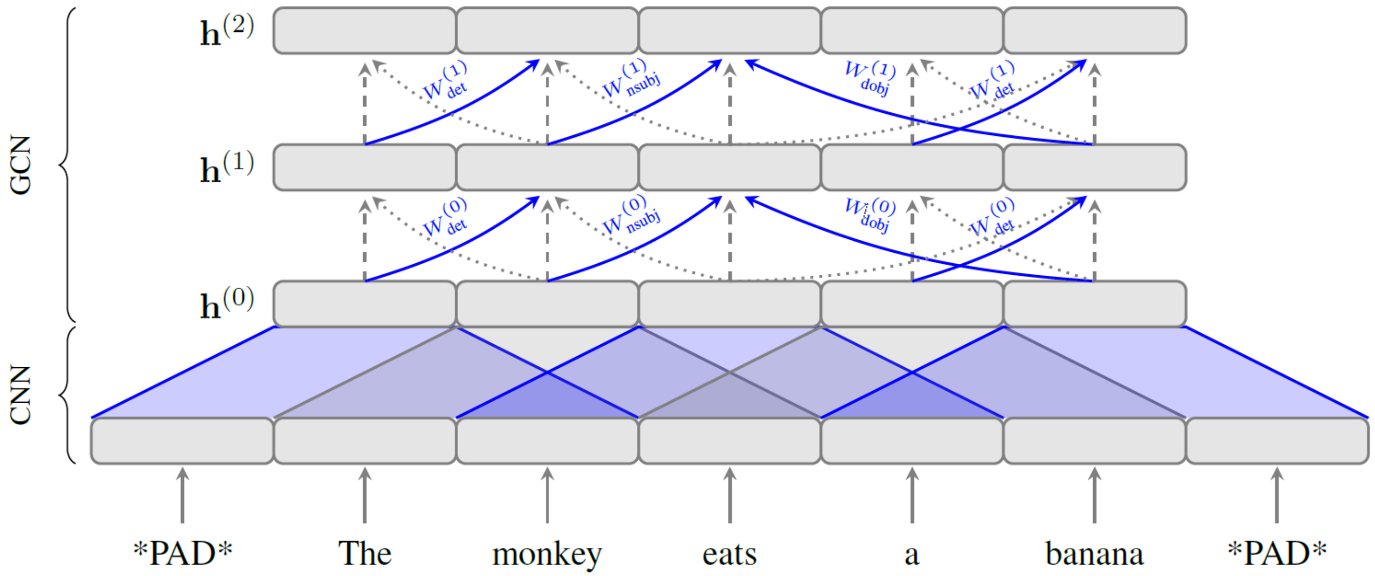
$$h_v^{(k+1)} = ReLU\left(\sum_{u\in\mathcal{N}(v)} W^{(k)}_{L(u,v)} h_u^{(k)} + b^{(k)}_{L(u,v)}\right)$$

Handle over-parameterized

$$W^{(k)}_{L(u,v)} = V^{(k)}_{dir(u,v)}$$

Joost Bastings, Ivan Titov et al., Graph Convolutional Encoders for Syntax-aware Neural Machine Translation, EMNLP 2017.

Syntax-aware Neural Machine Translation



Syntactic GCN layer

$$h_v^{(k+1)} = ReLU\left(\sum_{u \in \mathcal{N}(v)} W^{(k)}_{L(u,v)} h_u^{(k)} + b^{(k)}_{L(u,v)}\right)$$

A 2-layer syntactic GCN on top of a convolutional encoder.

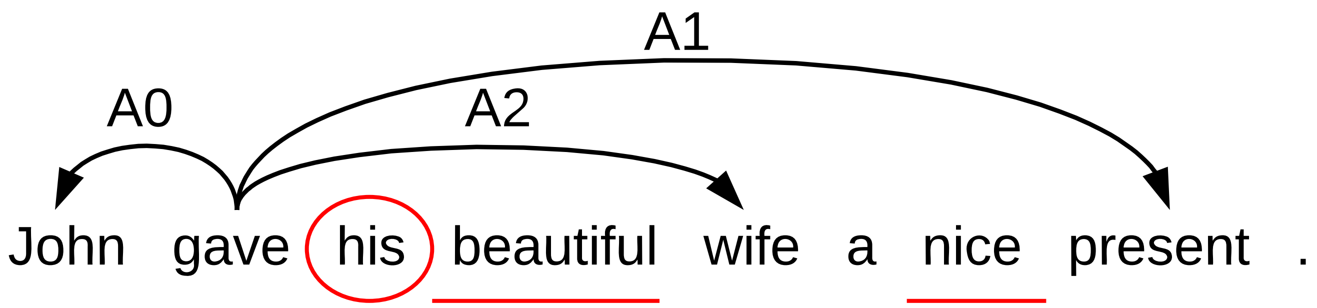
Joost Bastings, Ivan Titov et al., Graph Convolutional Encoders for Syntax-aware Neural Machine Translation, EMNLP 2017.

Experiments

	Kendall	$BLEU_1$	$BLEU_4$
BoW	0.3352	40.6	9.5
+ GCN	0.3520	44.9	12.2
CNN	0.3601	42.8	12.6
+ GCN	0.3777	44.7	13.7
BiRNN	0.3984	45.2	14.9
+ GCN	0.4089	47.5	16.1
BiRNN (full)	0.5440	53.0	23.3
+ GCN	0.5555	54.6	23.9

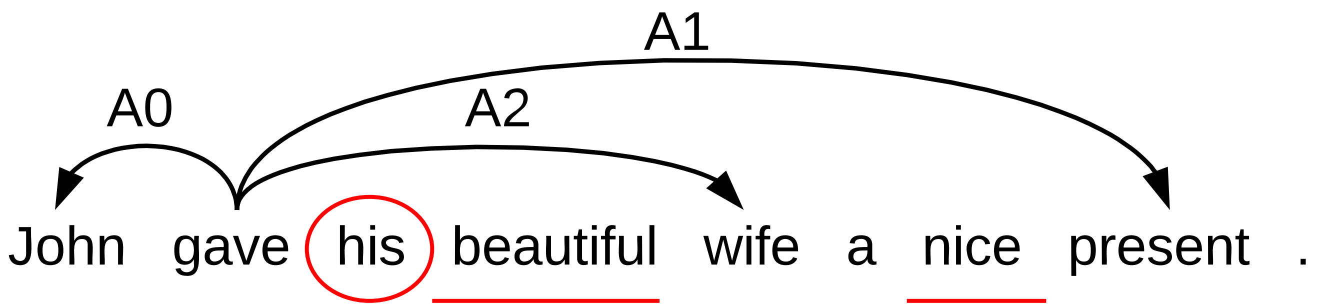
Joost Bastings, Ivan Titov et al., Graph Convolutional Encoders for Syntax-aware Neural Machine Translation, EMNLP 2017.

Exploiting Semantics in NMT with GCNs



Diego Marchenggiani, Joost Basting, Ivan Titov, Exploiting Semantics in Neural Machine Translation with Graph neural networks, NAACL 2018.

Exploiting Semantics in NMT with GCNs

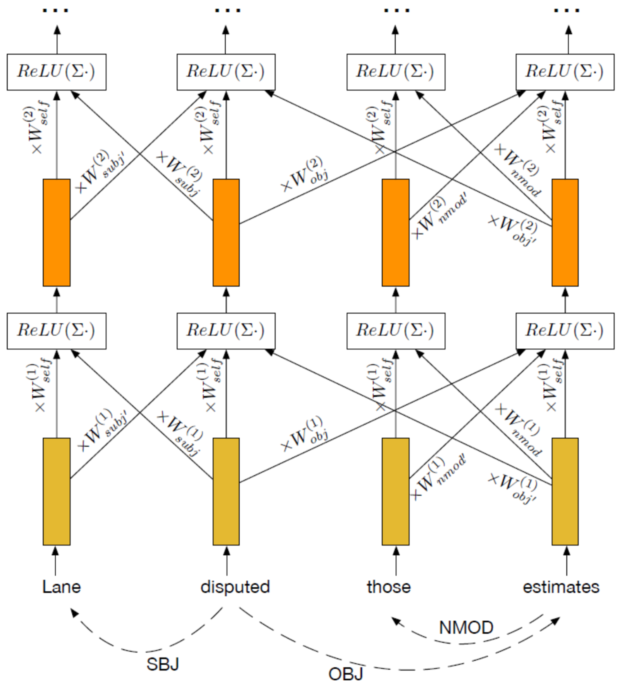


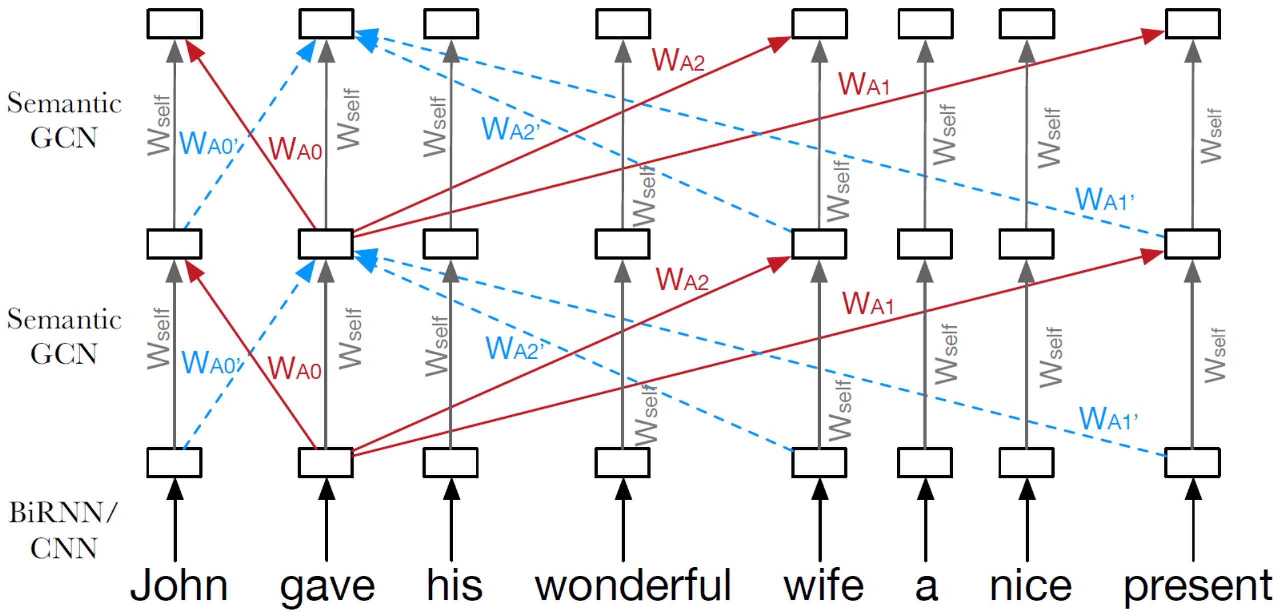
Fill the gap by considering semantic structures in NMT.

Sematic roles can be beneficial to “argument switching”.

Diego Marchenggiani, Joost Basting, Ivan Titov, Exploiting Semantics in Neural Machine Translation with Graph neural networks, NAACL 2018.

Exploiting Semantics in NMT with GCNs





Similar gating $h_v^{(k+1)} = ReLU\left(\sum_{u \in \mathcal{N}(v)} W_{L(u,v)}^{(k)} h_u^{(k)} + b_{L(u,v)}^{(k)}\right)$ $W_{L(u,v)}^{(k)} = V_{dir(u,v)}^{(k)}$

Diego Marchenggiani, Joost Basting, Ivan Titov, Exploiting Semantics in Neural Machine Translation with Graph neural networks, NAACL 2018.

Experiments

	BiRNN	CNN
Baseline (Bastings et al., 2017)	14.9	12.6
+Sem	15.6	13.4
+Syn (Bastings et al., 2017)	16.1	13.7
+Syn + Sem	15.8	14.3

Test BLEU, En-De, News commentary

Diego Marchenggiani, Joost Basting, Ivan Titov, Exploiting Semantics in Neural Machine Translation with Graph neural networks, NAACL 2018.

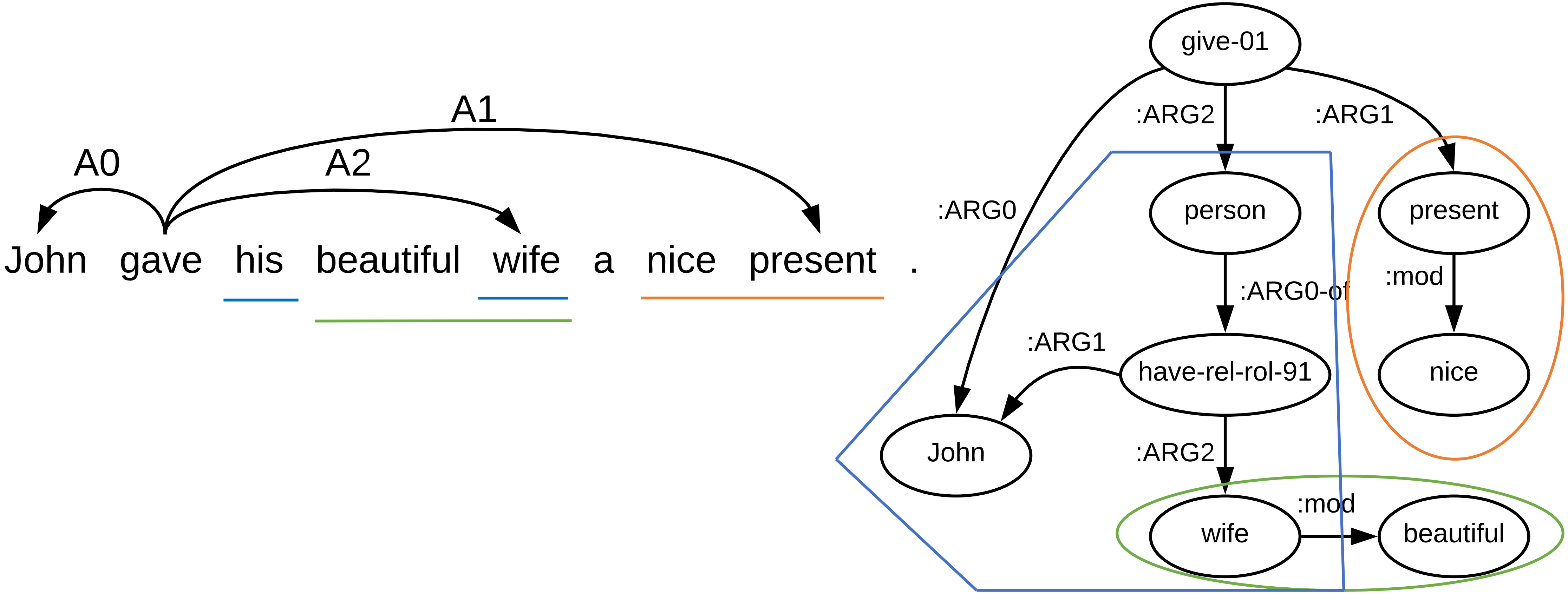
Experiments

	BiRNN
Baseline (Bastings et al., 2017)	23.3
+Sem	24.5
+Syn (Bastings et al., 2017)	23.9
+Syn + Sem	24.9

Test BLEU, En-De, full WMT16

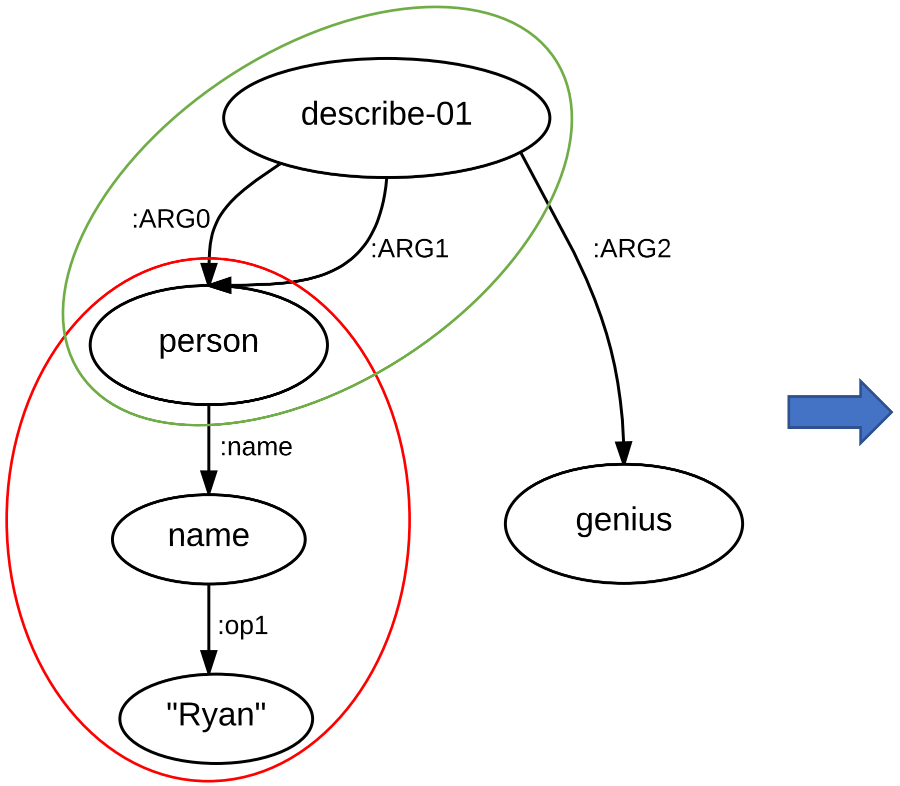
Diego Marchenggiani, Joost Basting, Ivan Titov, Exploiting Semantics in Neural Machine Translation with Graph neural networks, NAACL 2018.

Semantic NMT using AMR



Linfeng Song, Daniel Gildea, Yue Zhang, Semantic Neural Machine Translation using AMR, TACL 2019.

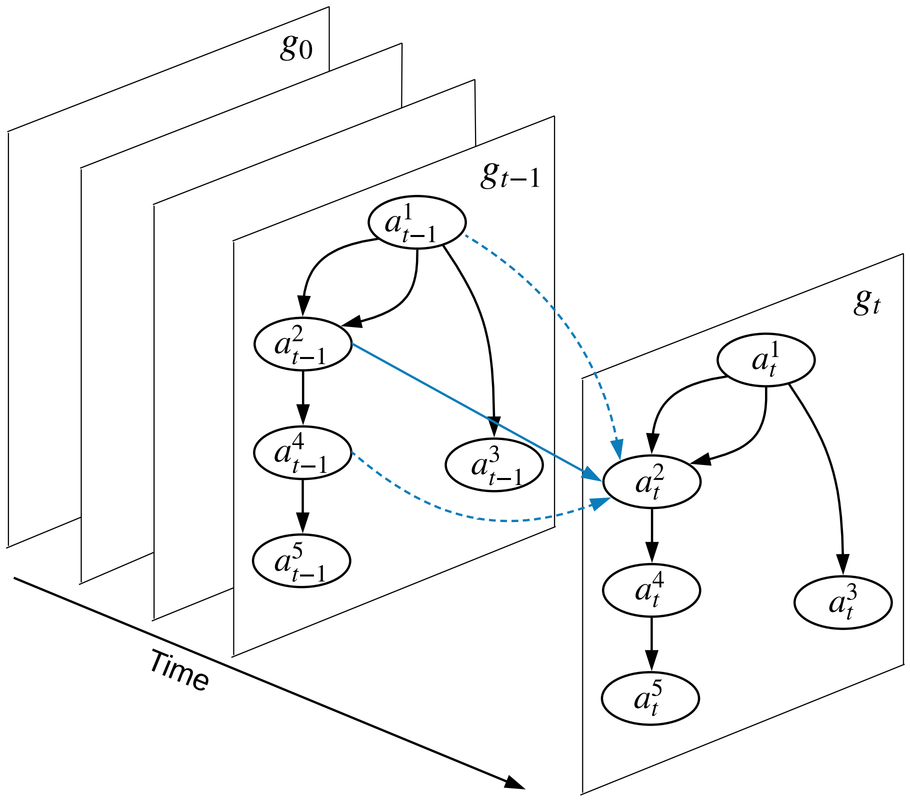
Abstract meaning representation (AMR)



Ryan's description of himself: a genius

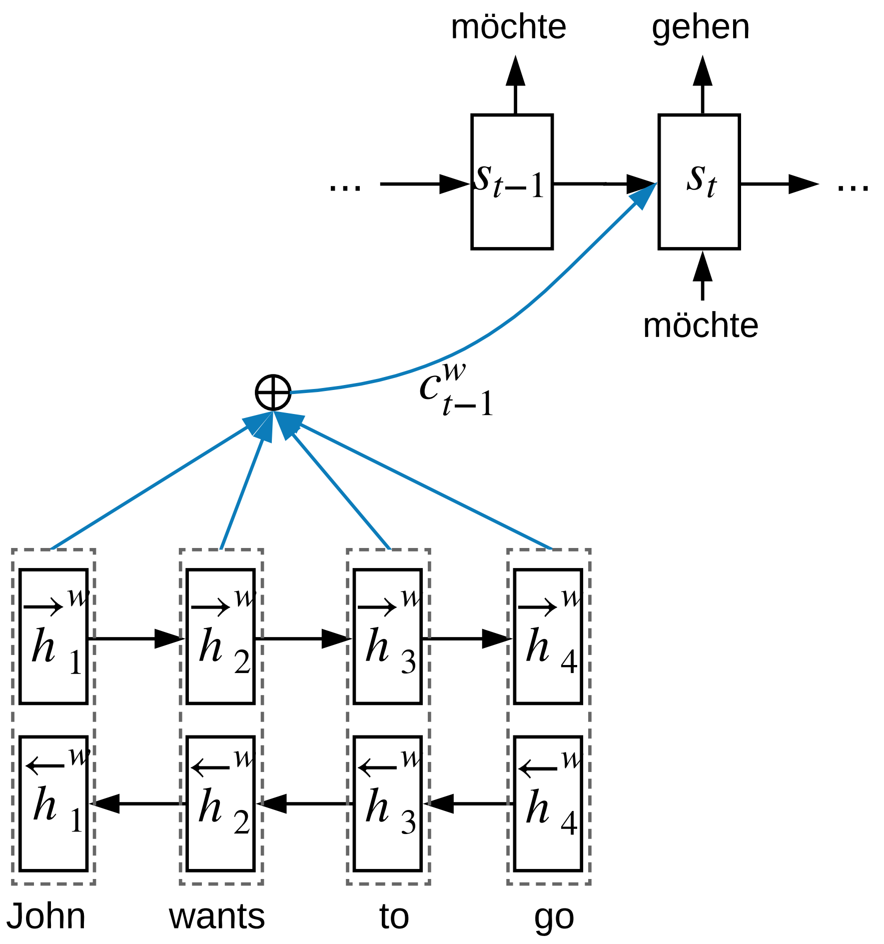
Linfeng Song, Daniel Gildea, Yue Zhang, Semantic Neural Machine Translation using AMR, TACL 2019.

Encoding AMRs with GRN



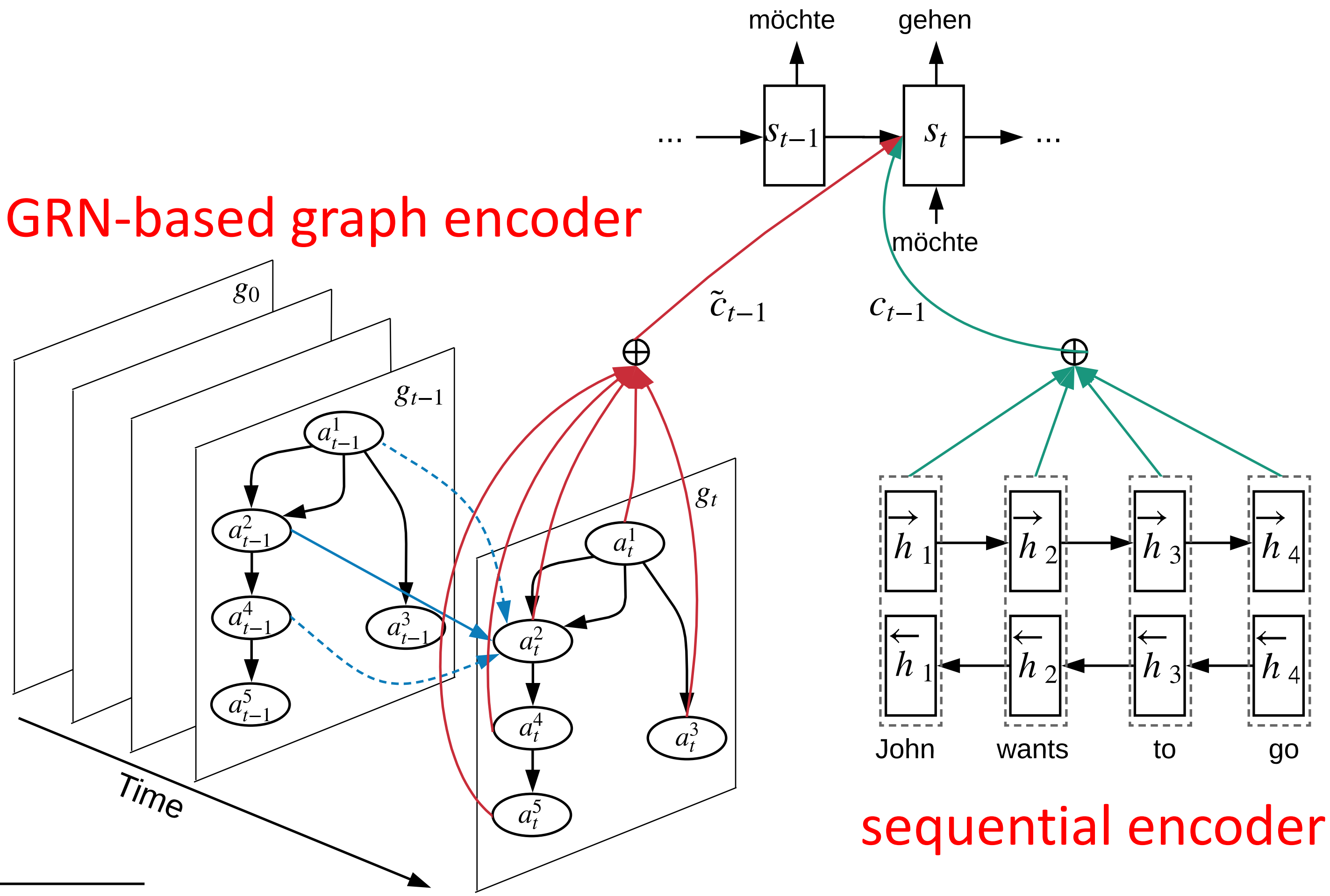
Linfeng Song, Daniel Gildea, Yue Zhang, Semantic Neural Machine Translation using AMR, TACL 2019.

Baseline: attention-based seq2seq



Linfeng Song, Daniel Gildea, Yue Zhang, Semantic Neural Machine Translation using AMR, TACL 2019.

Model: Dual2seq



Linfeng Song, Daniel Gildea, Yue Zhang, Semantic Neural Machine Translation using AMR, TACL 2019.

Experiments

- Benchmark (EN-DE):
  - Training: News commentary v11 (241K), full WMT 16 (4.5M)
  - Dev/Test: newstest2013/newstest2016
- Preprocessing:
  - Tokenization by Moses tokenizer
  - Training sentences with length ≥ 50 are filtered
  - AMRs (JAMR), dependency trees (CoreNLP), semantic roles (IBM SIRE)
- Report cased BLEU **(primary metric)**, Meteor and TER↓

Linfeng Song, Daniel Gildea, Yue Zhang, Semantic Neural Machine Translation using AMR, TACL 2019.

Main results

System	NC-v11			Full WMT 16		
	BLEU(%)	TER↓	Meteor(%)	BLEU(%)	TER↓	Meteor(%)
OpenNMT-tf	15.1	0.6902	30.4	24.3	0.5567	42.3
Seq2seq	16.0	0.6695	33.8	23.7	0.5590	42.6
Marcheggiani et al. (Dep)	16.1	--	--	23.9	--	--
Marcheggiani et al. (SRL)	15.6	--	--	24.5	--	--
Marcheggiani et al. (both)	15.8	--	--	24.9	--	--
Dual2seq-LinAMR	17.3	0.6530	36.1	24.0	0.5643	42.5
Duel2seq-SRL	17.2	0.6591	36.4	23.8	0.5626	42.2
Dual2seq-Dep	17.8	0.6516	36.7	25.0	0.5538	43.3
Dual2seq	**19.2** +3.2	**0.6305**	**38.4**	**25.5** +1.8	**0.5480**	**43.8**

Linfeng Song, Daniel Gildea, Yue Zhang, Semantic Neural Machine Translation using AMR, TACL 2019.

Main results

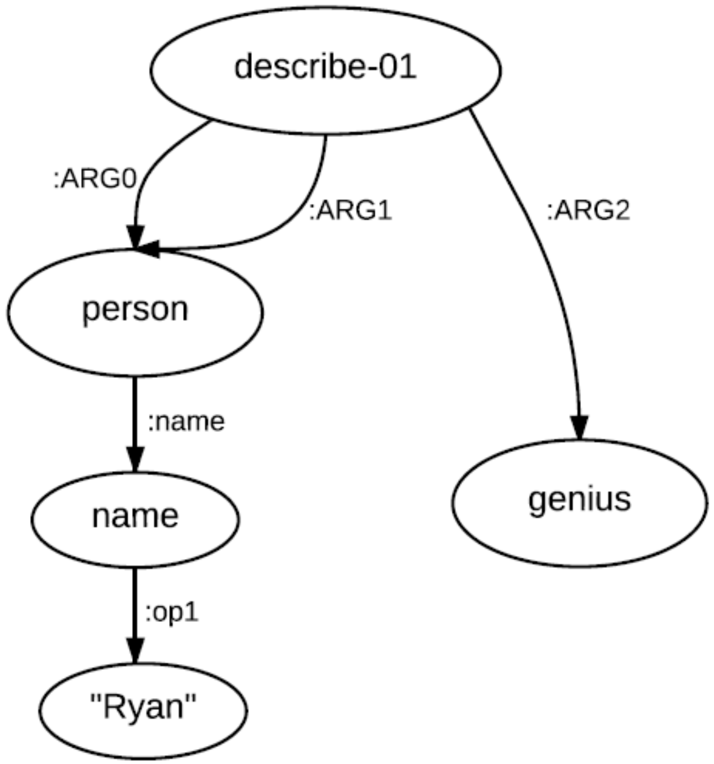
System	NC-v11			Full WMT 16		
	BLEU(%)	TER↓	Meteor(%)	BLEU(%)	TER↓	Meteor(%)
OpenNMT-tf	15.1	0.6902	30.4	24.3	0.5567	42.3
Seq2seq	16.0	0.6695	33.8	23.7	0.5590	42.6
Marcheggiani et al. (Dep)	16.1	--	--	23.9	--	--
Marcheggiani et al. (SRL)	15.6	--	--	24.5	--	--
Marcheggiani et al. (both)	15.8	--	--	24.9	--	--
Dual2seq-LinAMR	17.3	0.6530	36.1	24.0	0.5643	42.5
Duel2seq-SRL	17.2	0.6591	36.4	23.8	0.5626	42.2
Dual2seq-Dep	17.8	0.6516	36.7	25.0	0.5538	43.3
Dual2seq	**19.2**	**0.6305**	**38.4**	**25.5**	**0.5480**	**43.8**

Linfeng Song, Daniel Gildea, Yue Zhang, Semantic Neural Machine Translation using AMR, TACL 2019.

Solving problems using graph encoding

- Neural Machine Translation
- Text Generation
- Question Answering
- Information Extraction
- Sentiment
- Social Classification
- Syntactic Parsing
- Semantic Role Labeling
- Word Embedding
- Sentence Representation

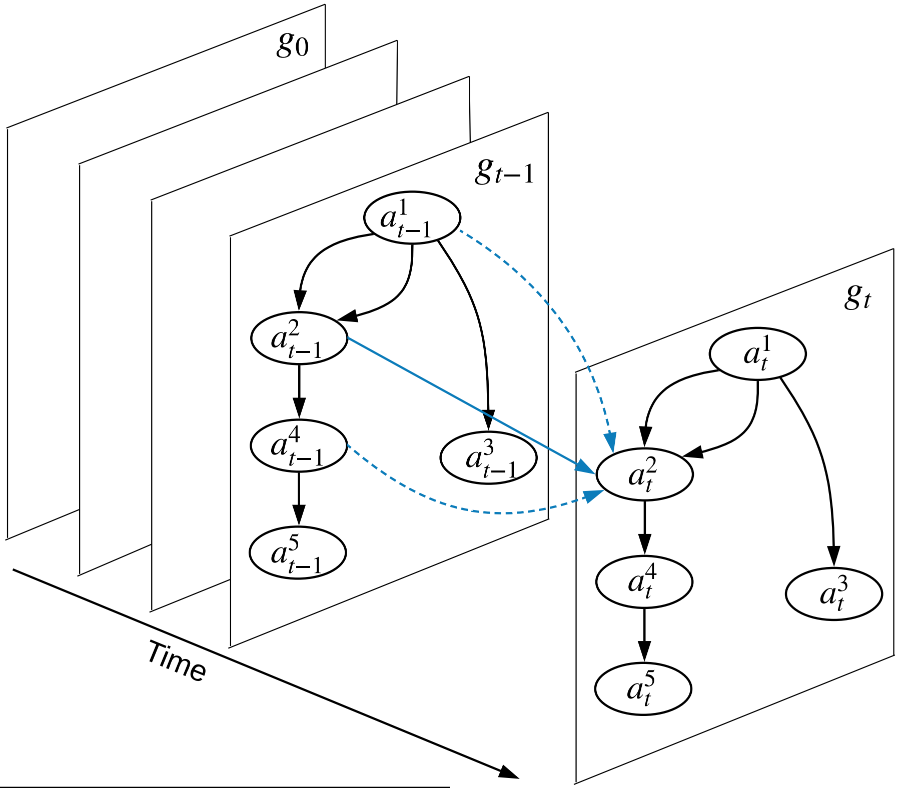
AMR-to Text Generation



- AMR: Abstraction meaning representation
- Differ from SRL, the task is challenging as word tenses and function words are abstracted away when constructing AMR graphs.

Linfeng Song, Yue Zhang et al., A Graph-to-Sequence Model for AMR-to-Text Generation, ACL 2018.
Daniel Beck, Gholamreza Haffari ey al., Graph-to-sequence learning using gated graph neural networks, ACL 2018.

Graph to sequence model



Same to above Graph state LSTM model

Encoding AMRs with GRN

Linfeng Song, Yue Zhang et al., A Graph-to-Sequence Model for AMR-to-Text Generation, ACL 2018.
Daniel Beck, Gholamreza Haffari ey al., Graph-to-sequence learning using gated graph neural networks, ACL 2018.

Experiments

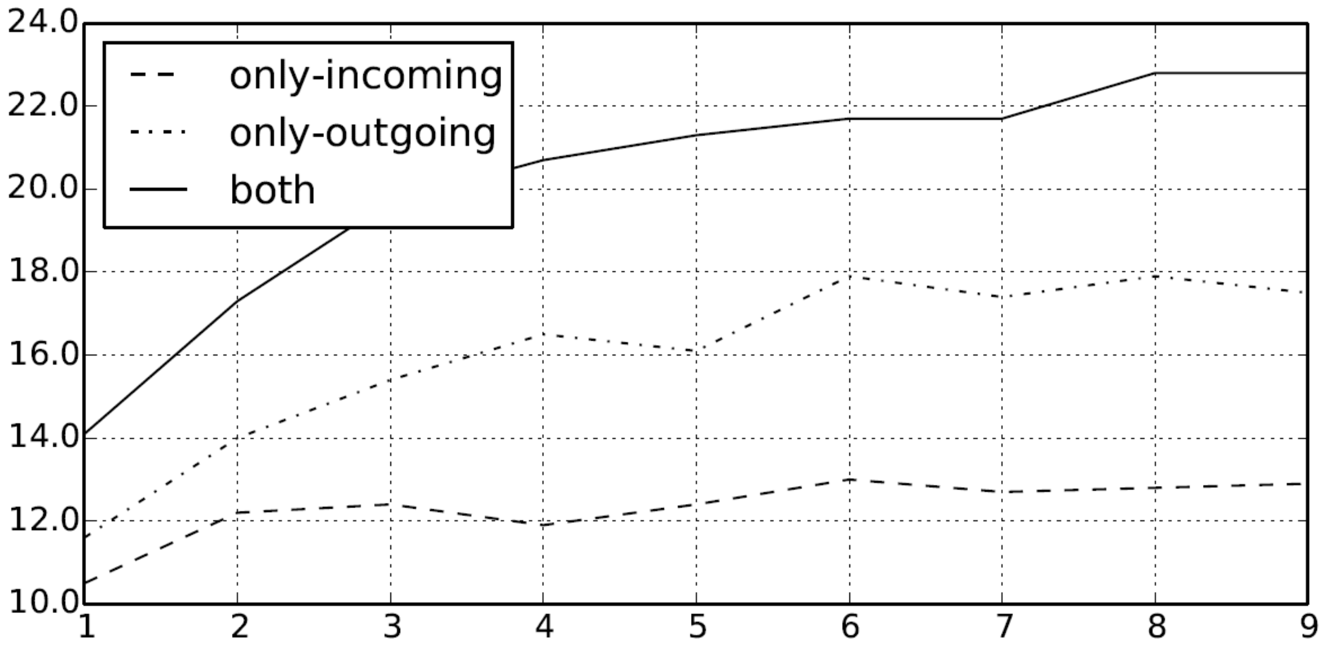
Model	BLEU	Time
Seq2seq	18.8	35.4s
Seq2seq+copy	19.9	37.4s
Seq2seq+charLSTM+copy	20.6	39.7s
Graph2seq	20.4	11.2s
Graph2seq+copy	22.2	11.1s
Graph2seq+Anon	22.1	9.2s
Graph2seq+charLSTM+copy	**22.8**	16.3s

3 times

AMR corpus(LDC2015E86)

Linfeng Song, Yue Zhang et al., A Graph-to-Sequence Model for AMR-to-Text Generation, ACL 2018.
Daniel Beck, Gholamreza Haffari ey al., Graph-to-sequence learning using gated graph neural networks, ACL 2018.

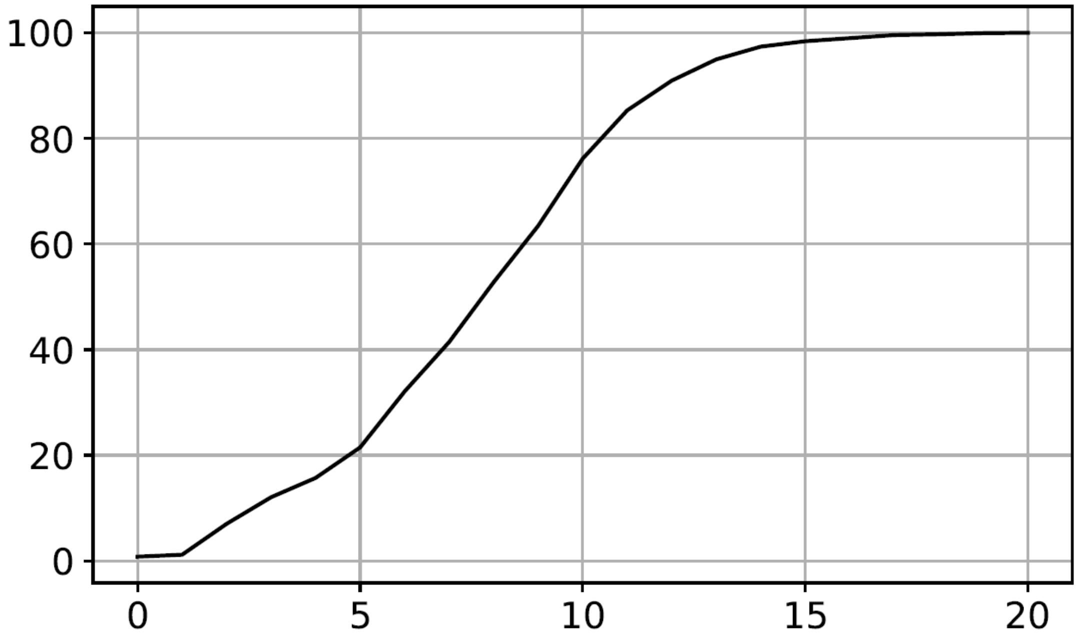
Experiments



Dev BLEU scores against transition steps for the graph encoder.

Linfeng Song, Yue Zhang et al., A Graph-to-Sequence Model for AMR-to-Text Generation, ACL 2018.
Daniel Beck, Gholamreza Haffari ey al., Graph-to-sequence learning using gated graph neural networks, ACL 2018.

Experiments



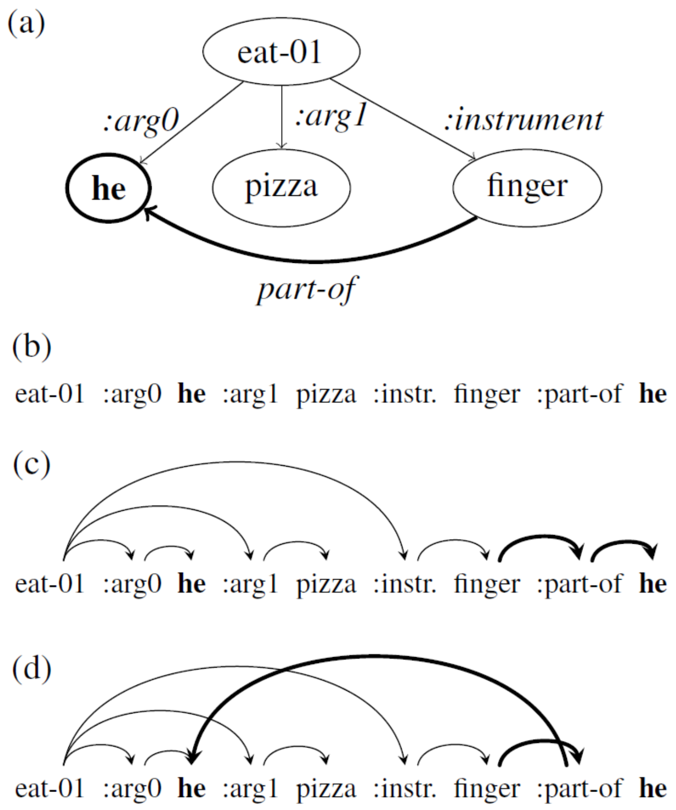
Percentage of Dev AMRs with different diameters.

Linfeng Song, Yue Zhang et al., A Graph-to-Sequence Model for AMR-to-Text Generation, ACL 2018.
Daniel Beck, Gholamreza Haffari ey al., Graph-to-sequence learning using gated graph neural networks, ACL 2018.

Structural Neural Encoders for AMR-to-text

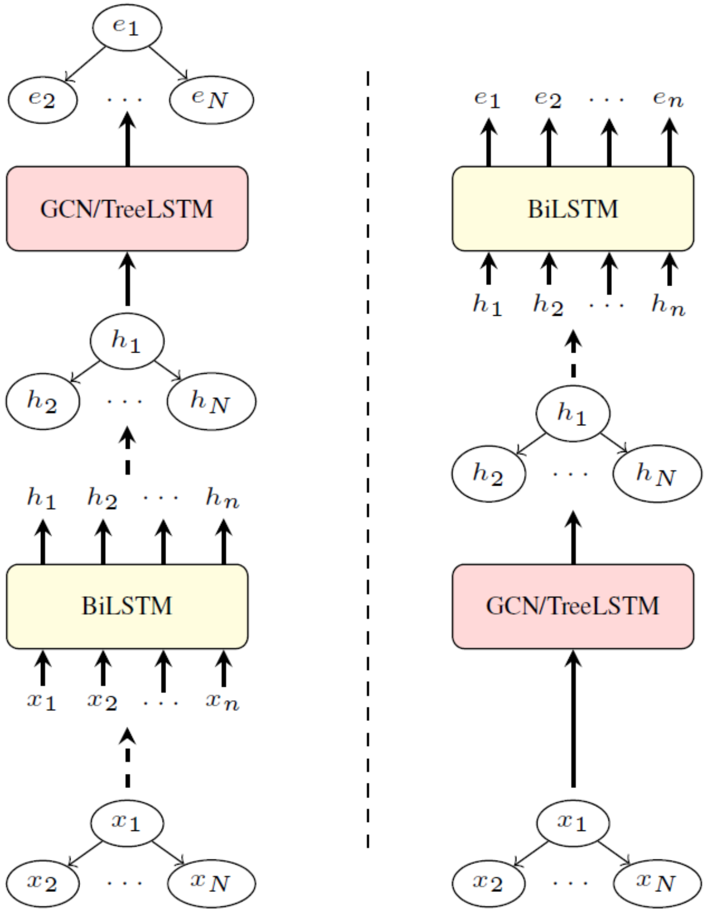
This paper investigated the different encoders between:

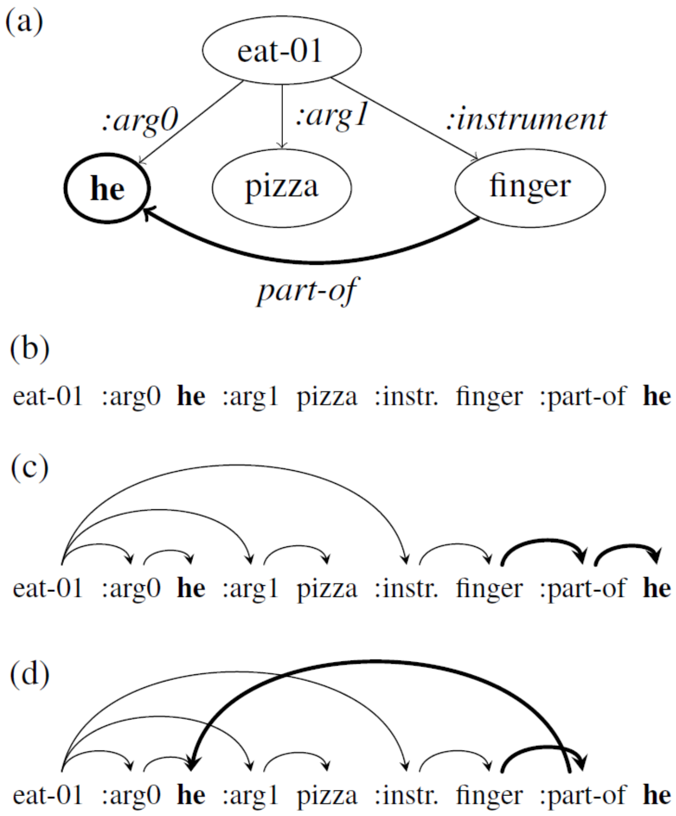
- Sequence
- Tree
- Graph



Damonte and Cohen, Structural Neural Encoders for AMR-to-text Generation, NAACL 2019.

Structural Neural Encoders for AMR-to-text





Damonte and Cohen, Structural Neural Encoders for AMR-to-text Generation, NAACL 2019.

Structural Neural Encoders for AMR-to-text

Input	Model	BLEU	Meteor
Seq	SEQ	21.40	22.00
Tree	SEQTREELSTM	21.84	22.34
	TREELSTMSEQ	22.26	22.87
	TREELSTM	22.07	22.57
	SEQGCN	21.84	22.21
	GCNSEQ	**23.62**	**23.77**
	GCN	15.83	17.76
Graph	SEQGCN	22.06	22.18
	GCNSEQ	**23.95**	**24.00**
	GCN	15.94	17.76

Damonte and Cohen, Structural Neural Encoders for AMR-to-text Generation, NAACL 2019.

Structural Neural Encoders for AMR-to-text

Input	Model	BLEU	Meteor
Seq	SEQ	21.40	22.00
Tree	SEQTREELSTM	21.84	22.34
	TREELSTMSEQ	22.26	22.87
	TREELSTM	22.07	22.57
	SEQGCN	21.84	22.21
	GCNSEQ	**23.62**	**23.77**
	GCN	15.83	17.76
Graph	SEQGCN	22.06	22.18
	GCNSEQ	**23.95**	**24.00**
	GCN	15.94	17.76

Evaluation script: https://github.com/sinantie/NeuralAmr

Damonte and Cohen, Structural Neural Encoders for AMR-to-text Generation, NAACL 2019.

Comment Generation

Several nontrivial challenges:

- The news articles can be very long while the title is can be too short to provide sufficient information.
- The title of the news sometimes uses hyperbolic expressions.
- Users focus on different aspects (topics) of the news.

Title

这部影片被称为“十年来最搞笑漫威电影”，你看了吗?

Have you seen the movie intitled as “the most hilarious Marvel movie”?

Content

点击“IPTV4K超高清”订阅，精彩内容等你共享《复仇者联盟3：无限战争》中的巅峰一役，将战火燃遍了整个宇宙...作为接档《复联3》的漫威电影，《蚁人2》的故事爆笑中带着温情，无疑成为了现阶段抚平漫威粉心中伤痛的一味良药...看过《复联3》的漫威粉们，心中都有同一个疑问：在几乎整个复仇者联盟都参与到无限战争的关键时刻，蚁人究竟去哪儿了？...

Click on the “IPTV4K ultra HD” to subscribe, fantastic contents are waiting for you to share. The battle in “Avengers: Infinity War” has spread the flames of war throughout the universe ... As the continuation Marvel movie to “Avengers 3”, the hilarious and warm “Ant-Man and the Wasp” is no doubt a good dose to heal the fans of Marvel at the time. ... Fans of the Marvel who have watched “Avengers 3” all have a doubt about where Ant-Man is when all other Avengers have been involved in the infinity war.

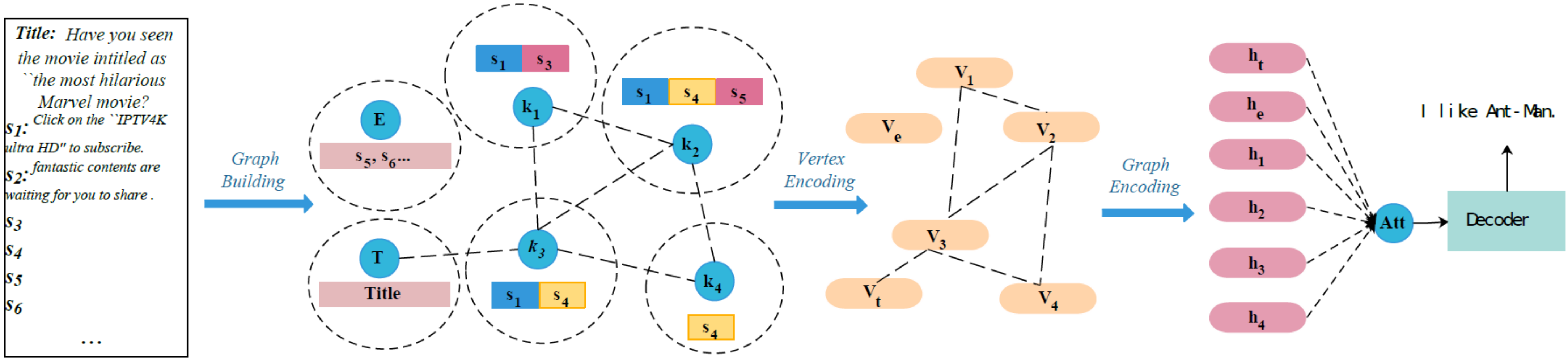
Comment

只有我觉得那个头盔像蚁人的头盔吗?

Am I the only one that thinks the helmet similar to the helmet of Ant-Man?

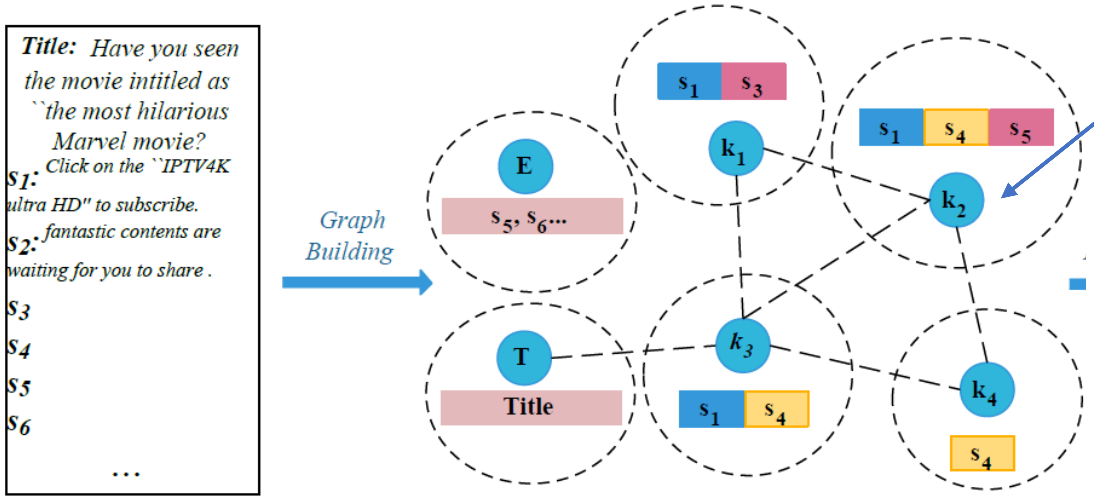
Coherent Comment Generation for Chinese Articles with a Graph-to-Sequence Model, Wei Li, et al., ACL2019

Coherent Comment Generation for Chinese Articles with a Graph-to-Sequence Model



Coherent Comment Generation for Chinese Articles with a Graph-to-Sequence Model, Wei Li, et al., ACL2019

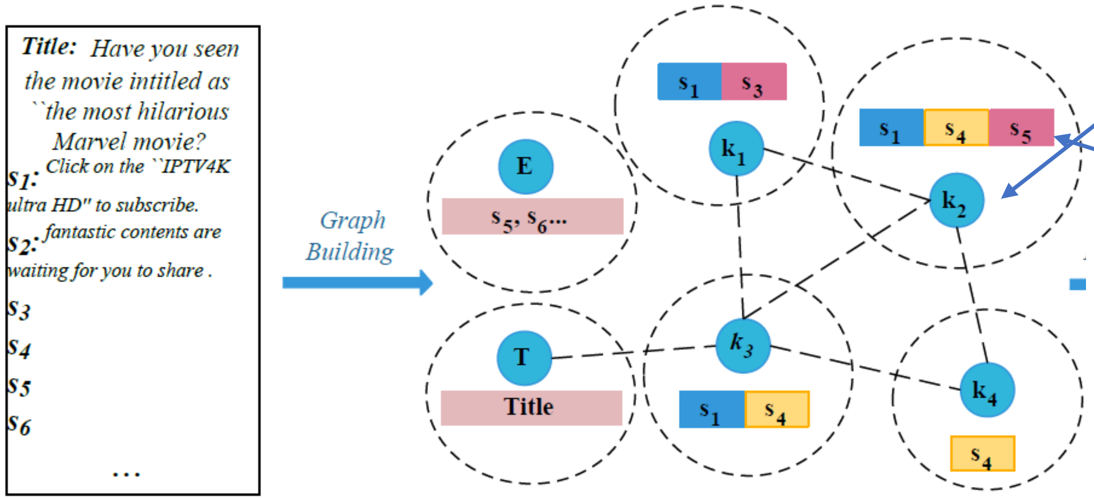
Graph Construction



- Do NER and Extract the keywords of the article as the topics, most of them are NERs.

Coherent Comment Generation for Chinese Articles with a Graph-to-Sequence Model, Wei Li, et al., ACL2019

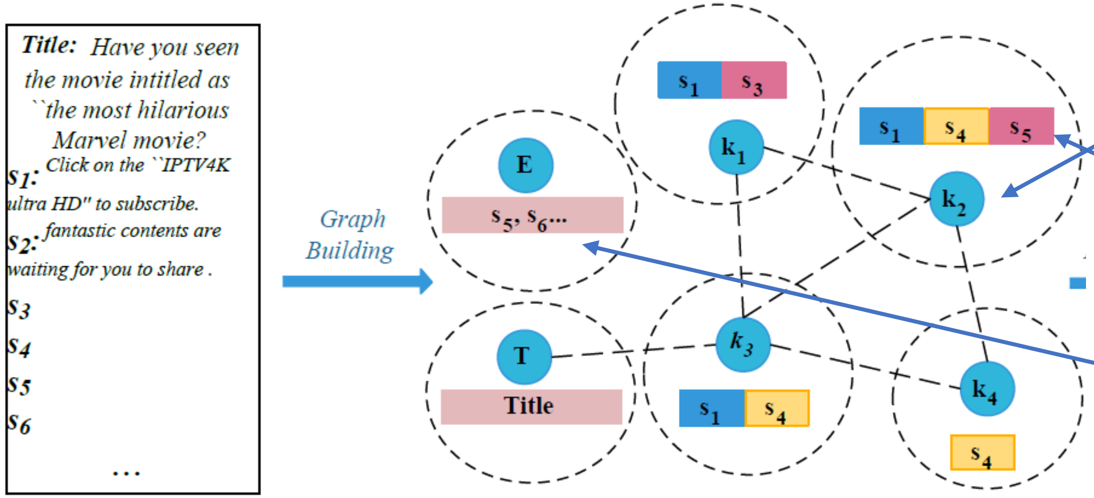
Graph Construction



- Do NER and Extract the keywords of the article as the topics, most of them are NERs.
- Associate each sentence to its corresponding keywords.

Coherent Comment Generation for Chinese Articles with a Graph-to-Sequence Model, Wei Li, et al., ACL2019

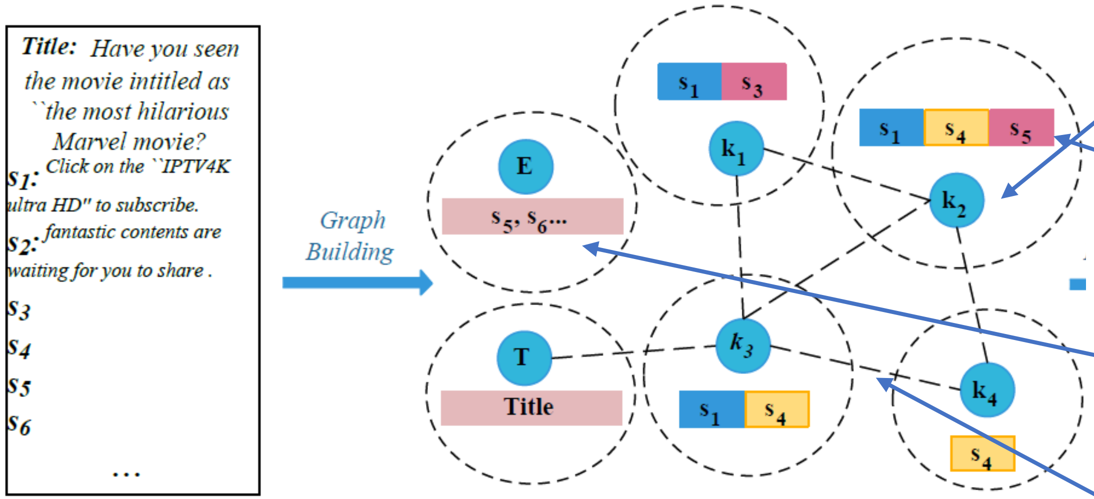
Graph Construction



- Do NER and Extract the keywords of the article as the topics, most of them are NERs.
- Associate each sentence to its corresponding keywords.
- Sentences that do mot contain any of the keywords.

Coherent Comment Generation for Chinese Articles with a Graph-to-Sequence Model, Wei Li, et al., ACL2019

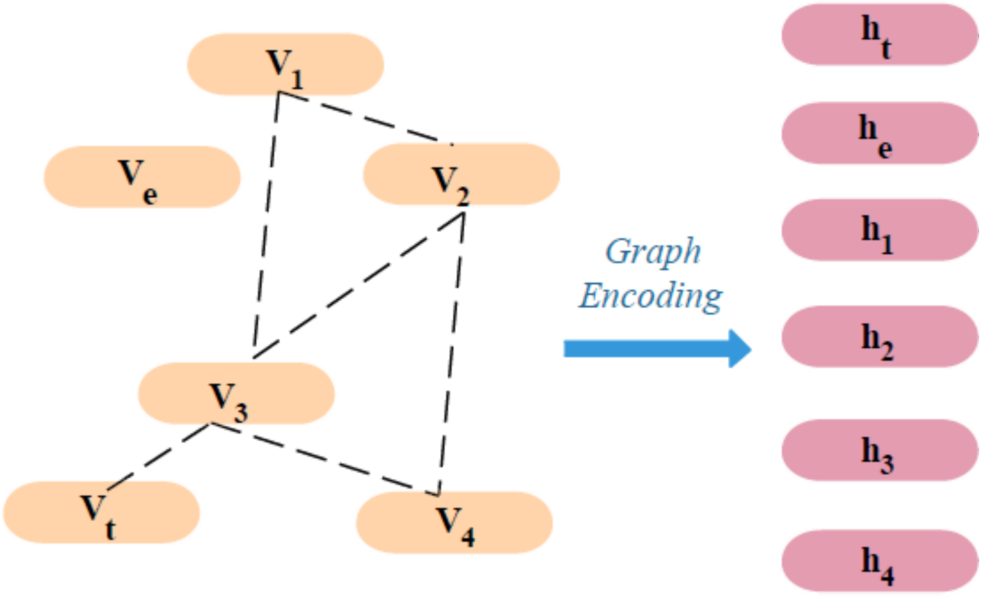
Graph Construction



- Do NER and Extract the keywords of the article as the topics, most of them are NERs.
- Associate each sentence to its corresponding keywords.
- Sentences that do mot contain any of the keywords.
- Use the number of shared sentences as edge weight.

Coherent Comment Generation for Chinese Articles with a Graph-to-Sequence Model, Wei Li, et al., ACL2019

Graph Encoder

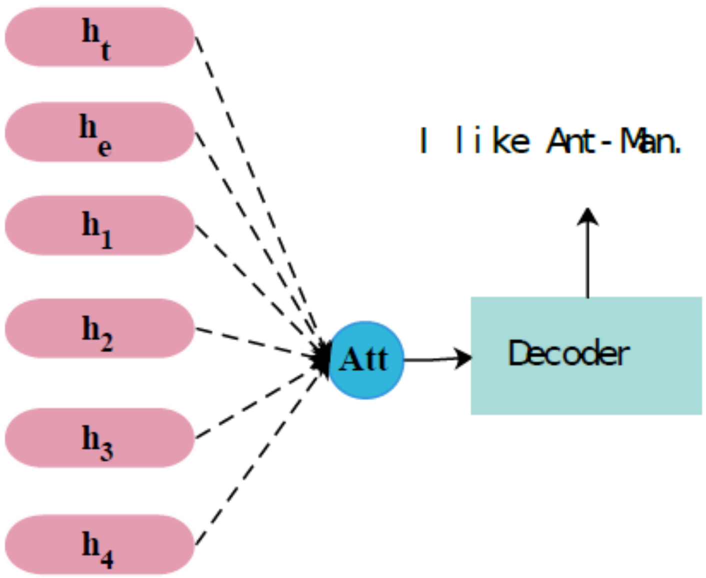


Standard GCN

$$h_v^{(k+1)} = ReLU\left(\sum_{u \in \mathcal{N}(v)} W^{(k)} h_u^{(k)} + b^{(k)}\right)$$

Coherent Comment Generation for Chinese Articles with a Graph-to-Sequence Model, Wei Li, et al., ACL2019

Decoder



Decoder: RNN with attention mechanism.

Coherent Comment Generation for Chinese Articles with a Graph-to-Sequence Model, Wei Li, et al., ACL2019

Experiments

Document and comment number of Entertainment and Sport.

Topic	document #	comment #
Entertainment	116,138	287,889
Sport	90,979	378,677

Length of content, title, comment and keyword.

	average word #		average character #	
	Ent	Sport	Ent	Sport
content	456.1	506.6	754.0	858.7
title	16.4	15.7	28.1	27.4
comment	16.3	19.4	26.2	31.2
keyword	8.4	9.0	-	-

Dataset : https://pan.baidu.com/s/1b5zAe7qqUBmuHz6nTU95UA5.

Coherent Comment Generation for Chinese Articles with a Graph-to-Sequence Model, Wei Li, et al., ACL2019

Experiments

Entertainment dataset

Models	Coherence	Informativeness	Fluency	Total
seq2seq-T (Qin et al., 2018)	5.38	3.70	8.22	5.77
seq2seq-C (Qin et al., 2018)	4.87	3.72	8.53	5.71
seq2seq-TC (Qin et al., 2018)	3.28	4.02	8.68	5.33
self-attention-B (Chen et al., 2018)	6.72	5.05	8.27	6.68
self-attention-K (Chen et al., 2018)	6.62	4.73	8.28	6.54
hierarchical-attention (Yang et al., 2016)	1.38	2.97	**8.65**	4.33
graph2seq (proposed)	**8.23**	**5.27**	8.08	**7.19**

Coherent Comment Generation for Chinese Articles with a Graph-to-Sequence Model, Wei Li, et al., ACL2019

Experiments

Sport dataset

Models	Coherence	Informativeness	Fluency	Total
seq2seq-T (Qin et al., 2018)	4.30	4.38	6.27	4.98
seq2seq-C (Qin et al., 2018)	3.88	3.85	6.02	4.58
seq2seq-TC (Qin et al., 2018)	4.70	5.08	6.37	5.38
self-attention-B (Chen et al., 2018)	5.15	5.62	6.28	5.68
self-attention-K (Chen et al., 2018)	6.68	5.83	**7.00**	6.50
hierarchical-attention (Yang et al., 2016)	4.43	5.05	6.02	5.17
graph2seq (proposed)	**7.97**	**6.18**	6.37	**6.84**

Coherent Comment Generation for Chinese Articles with a Graph-to-Sequence Model, Wei Li, et al., ACL2019

Solving problems using graph encoding

- Neural Machine Translation
- Text Generation
- Question Answering
- Information Extraction
- Sentiment
- Social Classification
- Syntactic Parsing
- Semantic Role Labeling
- Word Embedding
- Sentence Representation

Multi-hop reading comprehension

Q: (The Hanging Gardens, country, ?)
Candidates: {Iran, India, Pakistan, Somalia, ...}

The Hanging Gardens, in [**Mumbai**], also known as Pherozeshah Mehta Gardens, are terraced gardens ... [**They**] provide sunset views over [**the Arabian Sea**] ...

[**Mumbai**] (also known as Bombay, the official name until 1995) is the capital city of the Indian state of Maharashtra. [**It**] is the most populous city in [**India**] ...

[**The Arabian Sea**] is a region of the northern Indian Ocean bounded on the north by [**Pakistan**] and [**Iran**], on the west by northeastern [**Somalia**] and the Arabian Peninsula ...

Multi-hop reading comprehension

Q: (The Hanging Gardens, country, ?)
Candidates: {Iran, India, Pakistan, Somalia, ...}

The Hanging Gardens, in [**Mumbai**], also known as Pherozeshah Mehta Gardens, are terraced gardens ... [**They**] provide sunset views over [**the Arabian Sea**] ...

[**Mumbai**] (also known as Bombay, the official name until 1995) is the capital city of the Indian state of Maharashtra. [**It**] is the most populous city in [**India**] ...

[**The Arabian Sea**] is a region of the northern Indian Ocean bounded on the north by [**Pakistan**] and [**Iran**], on the west by northeastern [**Somalia**] and the Arabian Peninsula ...

Relevant evidence:

- The Hanging Gardens are in Mumbai.
- Mumbai is the most populous city in India.

Irrelevant evidence:

- The Hanging Gardens provide sunset views over the Arabian Sea.
- The Arabian Sea is bounded by Pakistan, Iran and Somalia.

Multi-hop reading comprehension

Q: (The Hanging Gardens, country, ?)
Candidates: {Iran, India, Pakistan, Somalia, ...}

The Hang[illegible] [Mumbai], [illegible] Pherozeshah Mehta Ga[illegible], are terraced gardens ... [They] provide sunset views over [**the Arabian Sea**] ...

(1) Structure creation

[**Mumbai**] (also known as Bombay, the official name until 1995) is the capita[illegible] [illegible] most populous city in [**India**] ...

(2) Evidence integration

[**The Arabian Sea**] is a region of the northern Indian Ocean bounded on the north by [**Pakistan**] and [**Iran**], on the west by northeastern [**Somalia**] and the Arabian Peninsula ...

Relevant evidence:

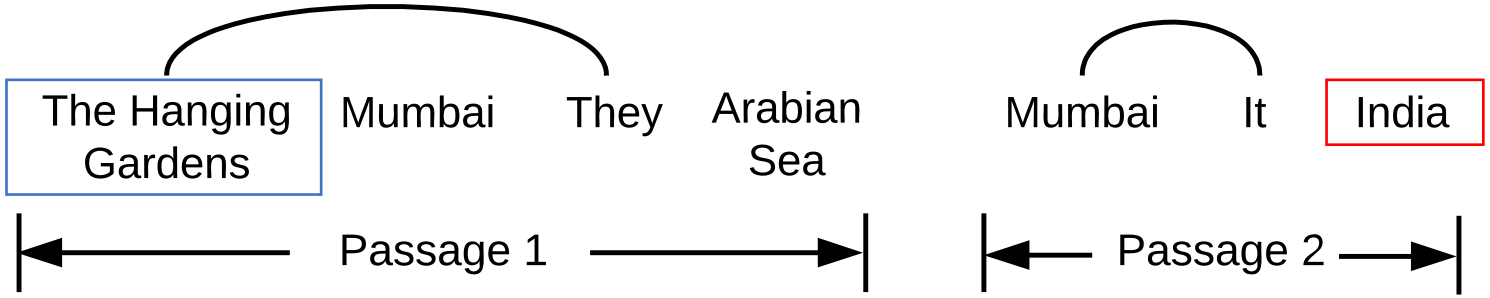
- The Hanging Gardens are in Mumbai.
- Mumbai is the most populous city in India.

Irrelevant evidence:

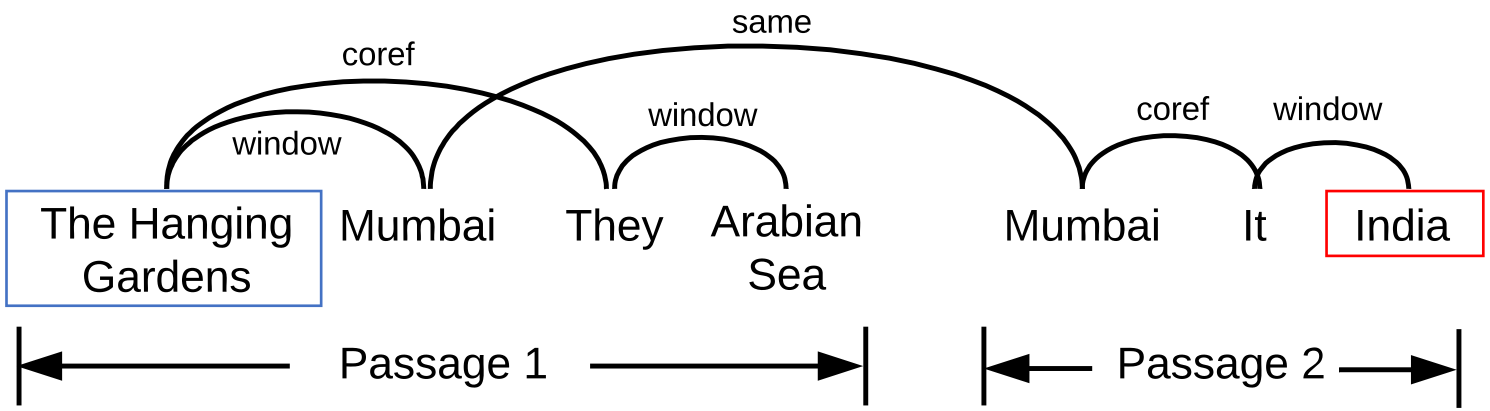
- The Hanging Gardens provide sunset views over the Arabian Sea.
- The Arabian Sea is bounded by Pakistan, Iran and Somalia.

Linfeng song, Zhiguo Wang et al., Exploring Graph-structured Passage Representation for Multi-hop Reading Comprehension with Graph Neural Networks, arXiv 2018.

Coref-DAG vs Evidence graph



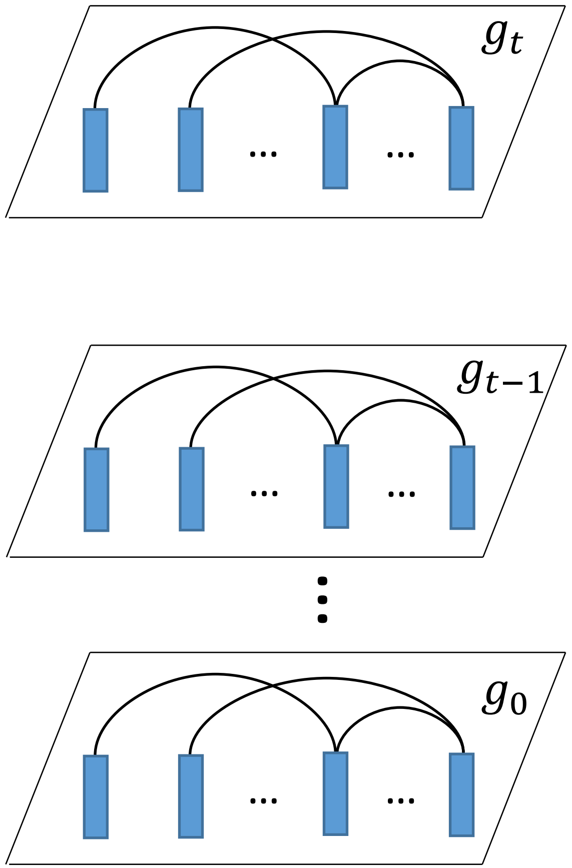
Coreference DAG



Evidence graph (Ours)

Linfeng song, Zhiguo Wang et al., Exploring Graph-structured Passage Representation for Multi-hop Reading Comprehension with Graph Neural Networks, arXiv 2018.
Bhuwan DhiNeural Models for Reasoning over Multiple Mentions using Coreference (Dhingra et al., NAACL 2018)

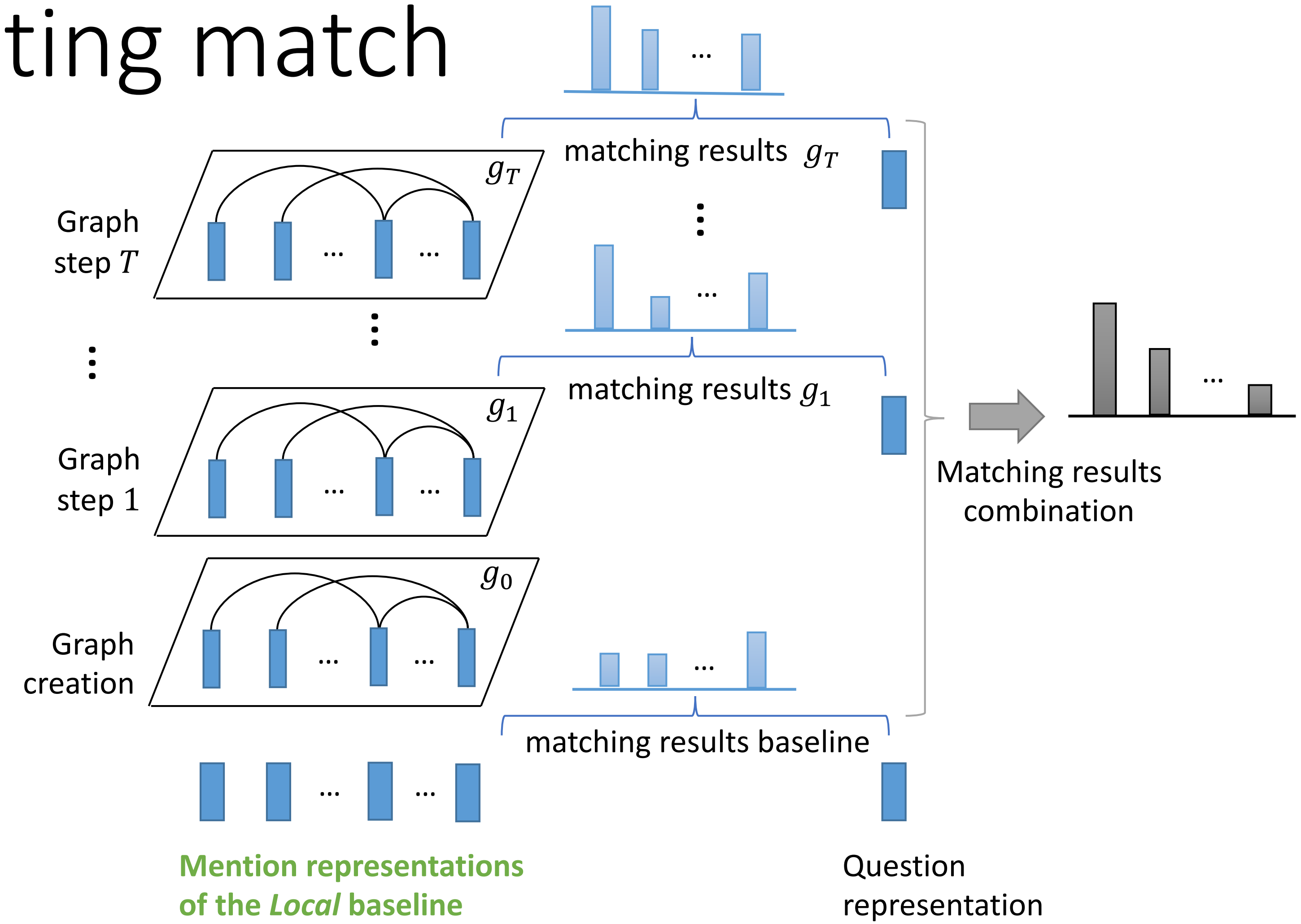
Graph recurrent network (GRN)



- GRN follows an iterative message passing process for updating each node state.
- Within each iteration, it takes two main steps:
  - Message calculation
  - Node state update

Linfeng song, Zhiguo Wang et al., Exploring Graph-structured Passage Representation for Multi-hop Reading Comprehension with Graph Neural Networks, arXiv 2018.

Calculating match



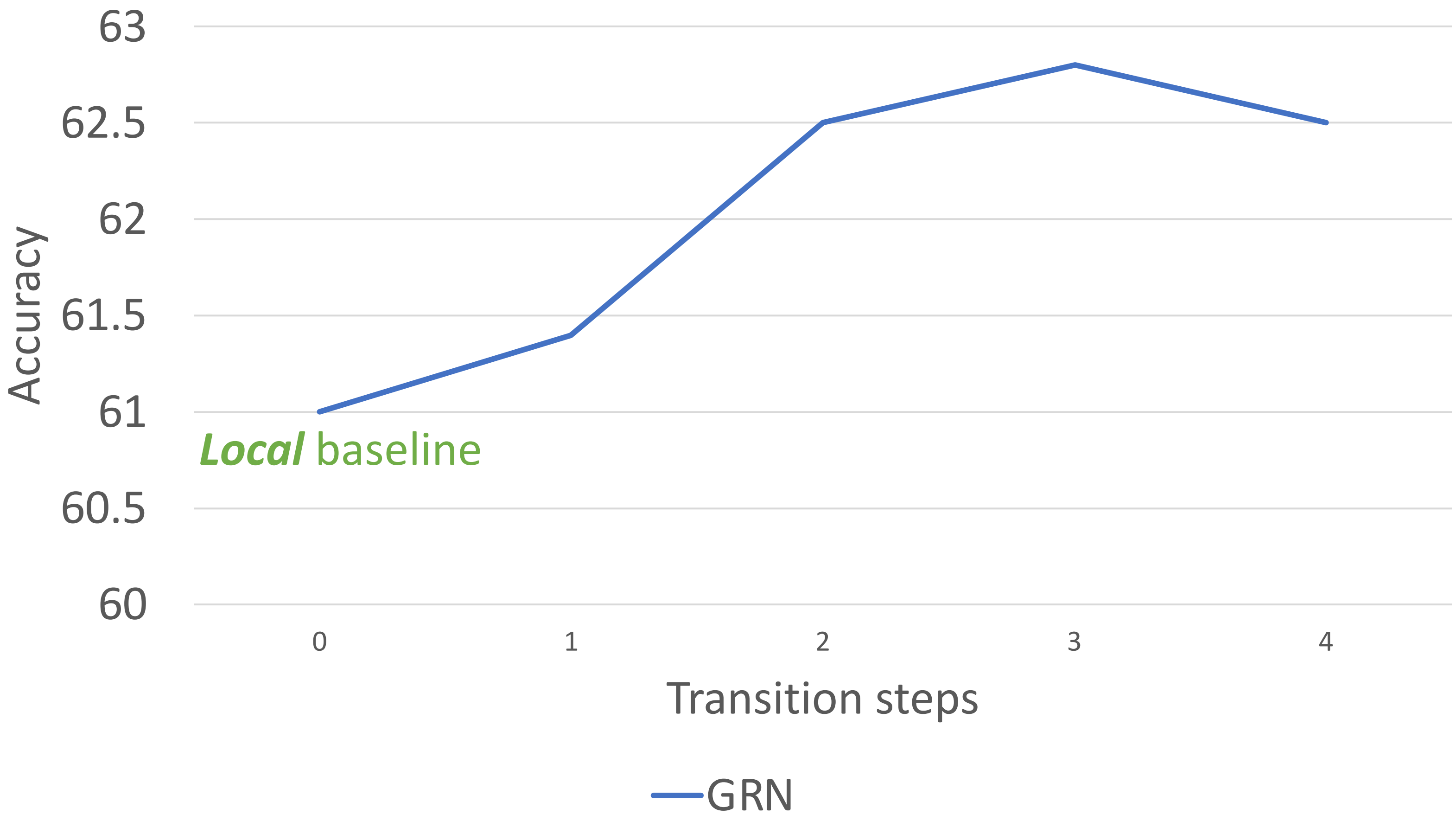
Linfeng song, Zhiguo Wang et al., Exploring Graph-structured Passage Representation for Multi-hop Reading Comprehension with Graph Neural Networks, arXiv 2018.

Experiments

- WikiHop (http://qangaroo.cs.ucl.ac.uk/)
  - 51K instances: 44K (training), 5K (dev), 2.5K (hold-out test)
  - Each instance is: $([p_1, p_2 \ldots p_L], q, C, a)$
  - Mentions are generated from automatic NER and coreference resolution, by Stanford CoreNLP

Linfeng song, Zhiguo Wang et al., Exploring Graph-structured Passage Representation for Multi-hop Reading Comprehension with Graph Neural Networks, arXiv 2018.

DEV experiment on message passing step (T)



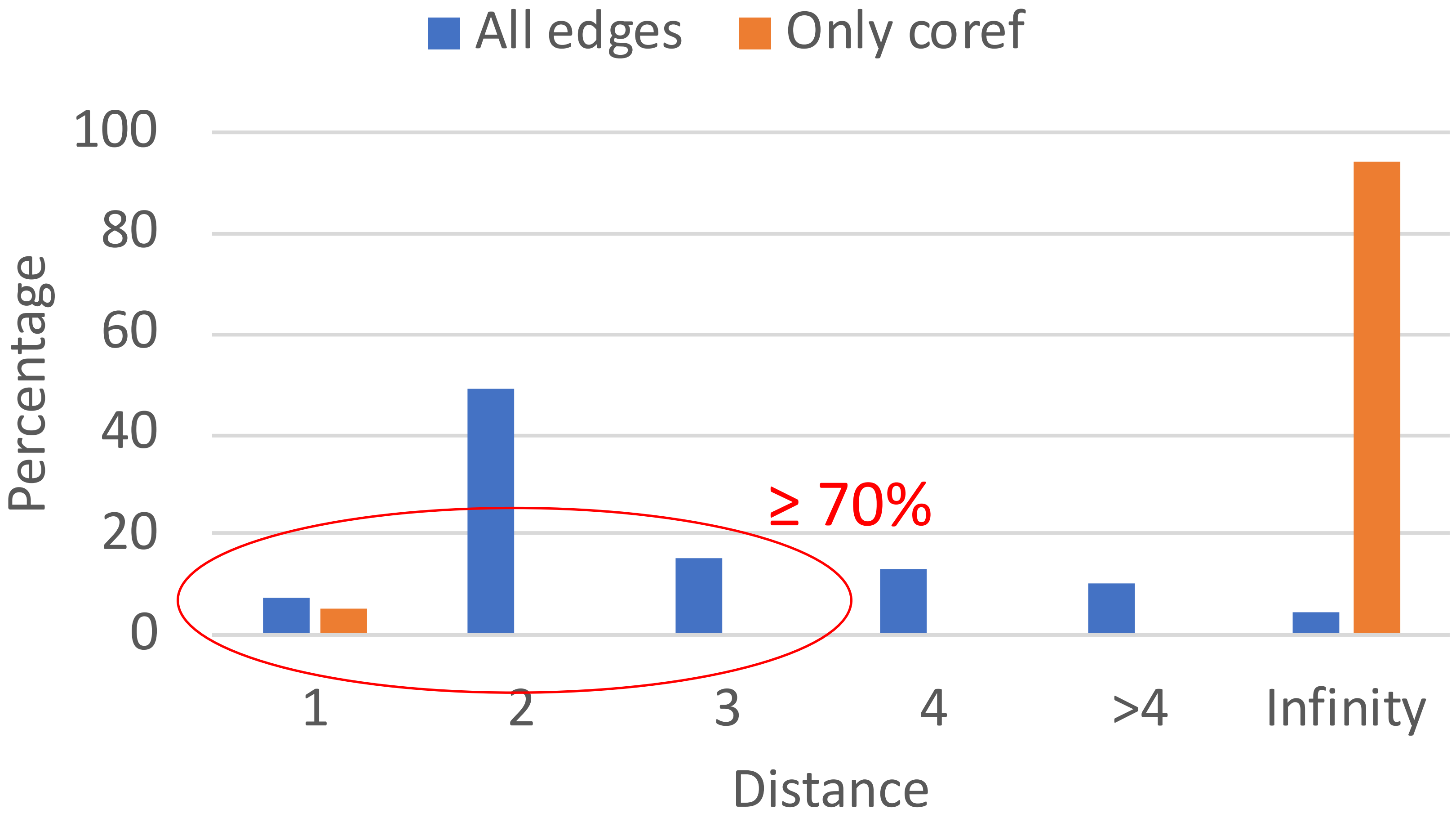
Linfeng song, Zhiguo Wang et al., Exploring Graph-structured Passage Representation for Multi-hop Reading Comprehension with Graph Neural Networks, arXiv 2018.

Main Comparison (accuracy)

Model	Dev	Test
GA w/ GRU (Dhingra et al., 2018)	54.9	--
GA w/ Coref-GRU (Dhingra et al., 2018)	56.0	59.3
Local	61.0	--
Local-2L	61.3	--
Coref-LSTM	61.4	--
Coref-GRN	61.4	--
Fully-Connect-GRN	61.3	--
MHQA-GRN	**62.8** +1.5	**65.4**

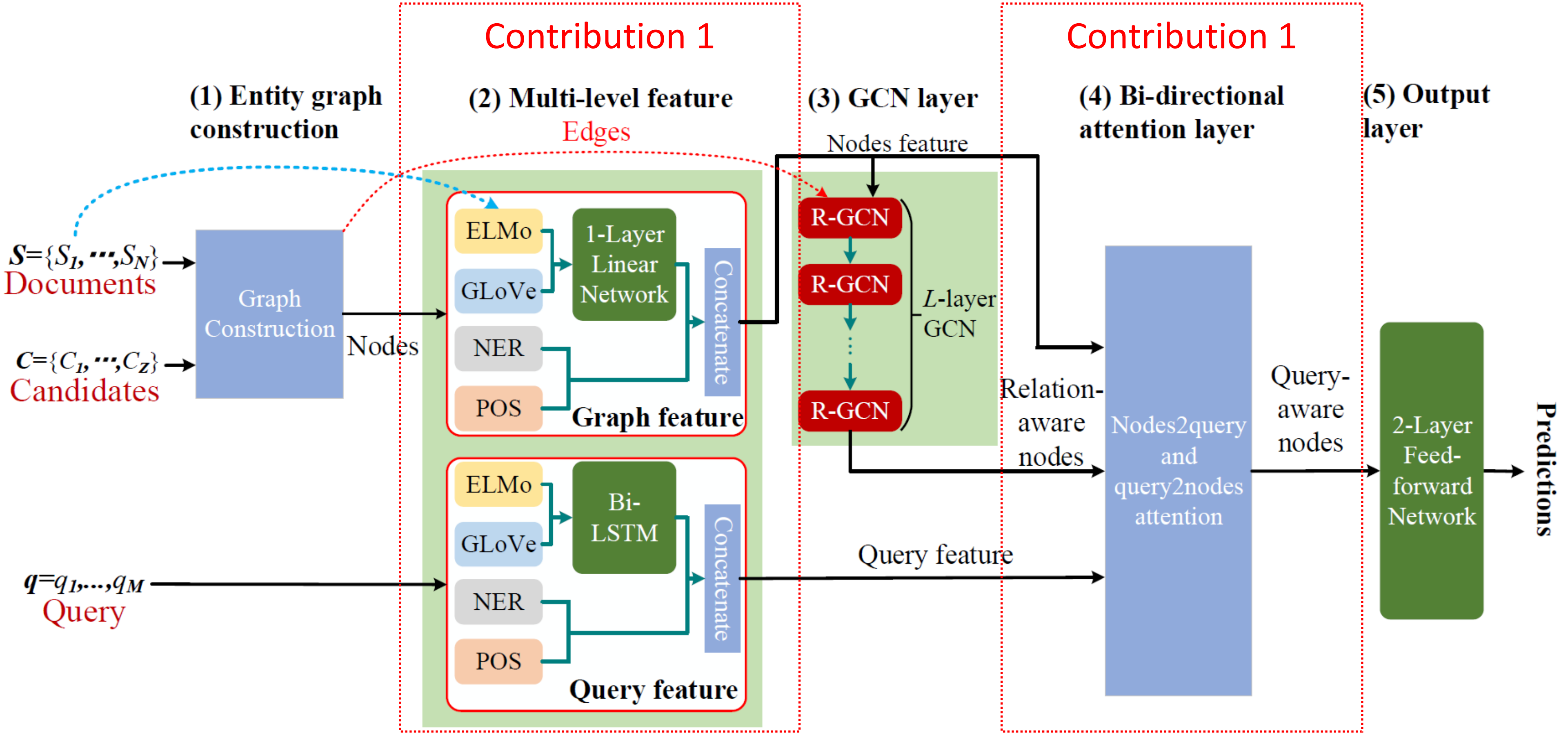
Linfeng song, Zhiguo Wang et al., Exploring Graph-structured Passage Representation for Multi-hop Reading Comprehension with Graph Neural Networks, arXiv 2018.

Distance between question and answer



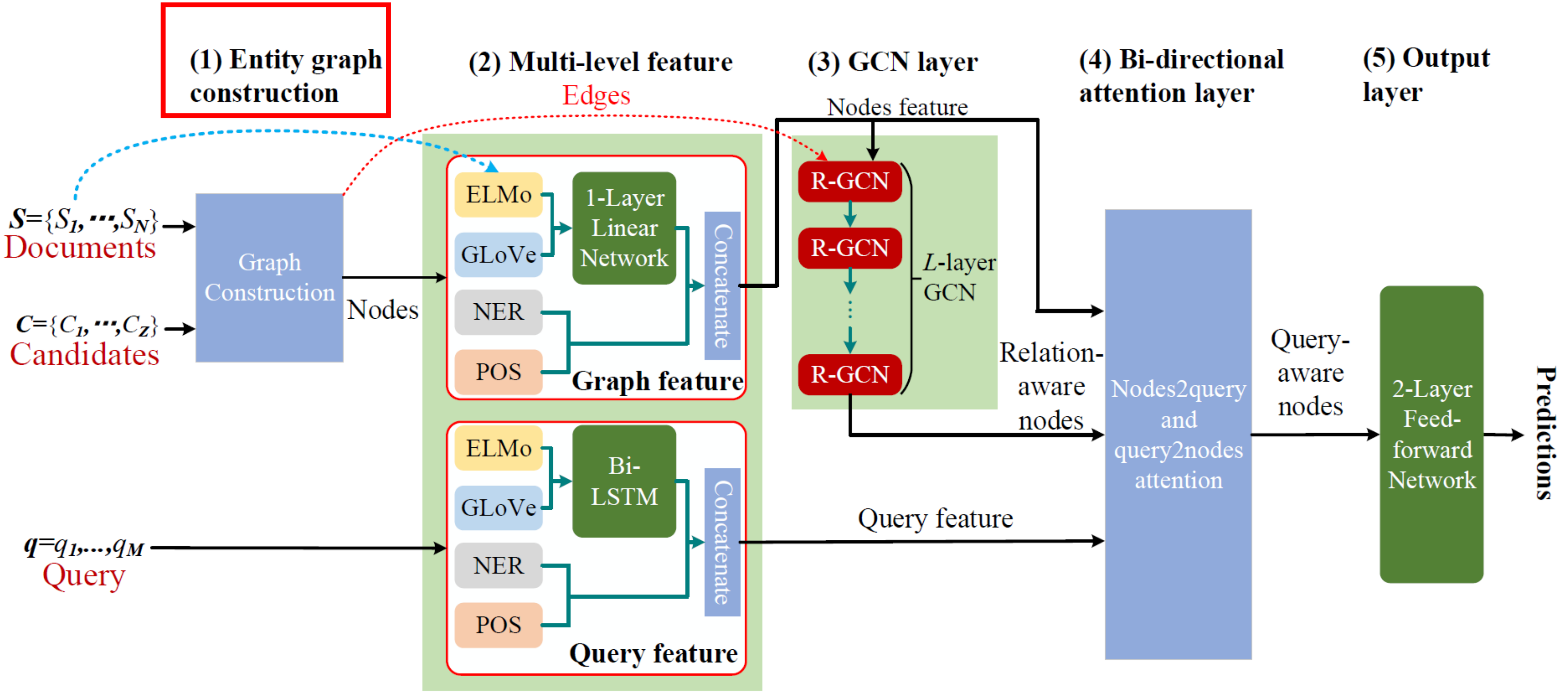
Linfeng song, Zhiguo Wang et al., Exploring Graph-structured Passage Representation for Multi-hop Reading Comprehension with Graph Neural Networks, arXiv 2018.

BAG: Bi-directional Attention Entity Graph



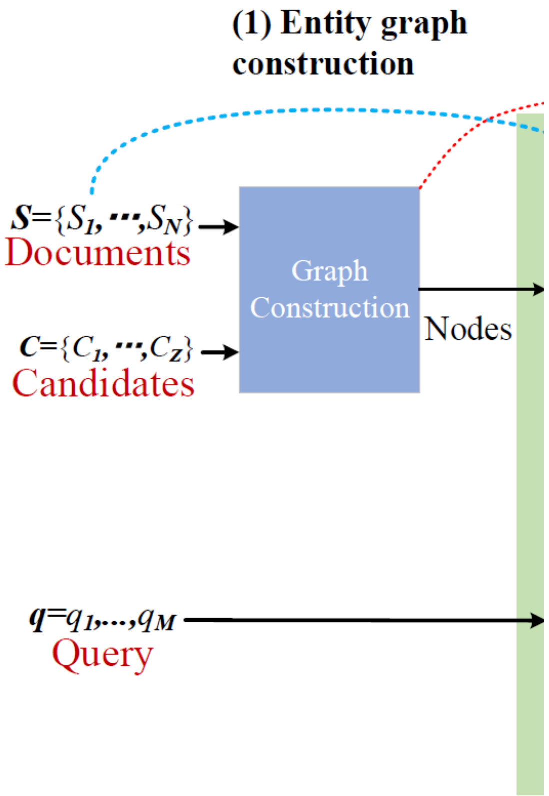
Yu Cao, Meng Fang et al., BAG: Bi-directional Attention Entity Graph Convolutional Network for Multi-hop Reasoning Question Answering, NAACL 2019.

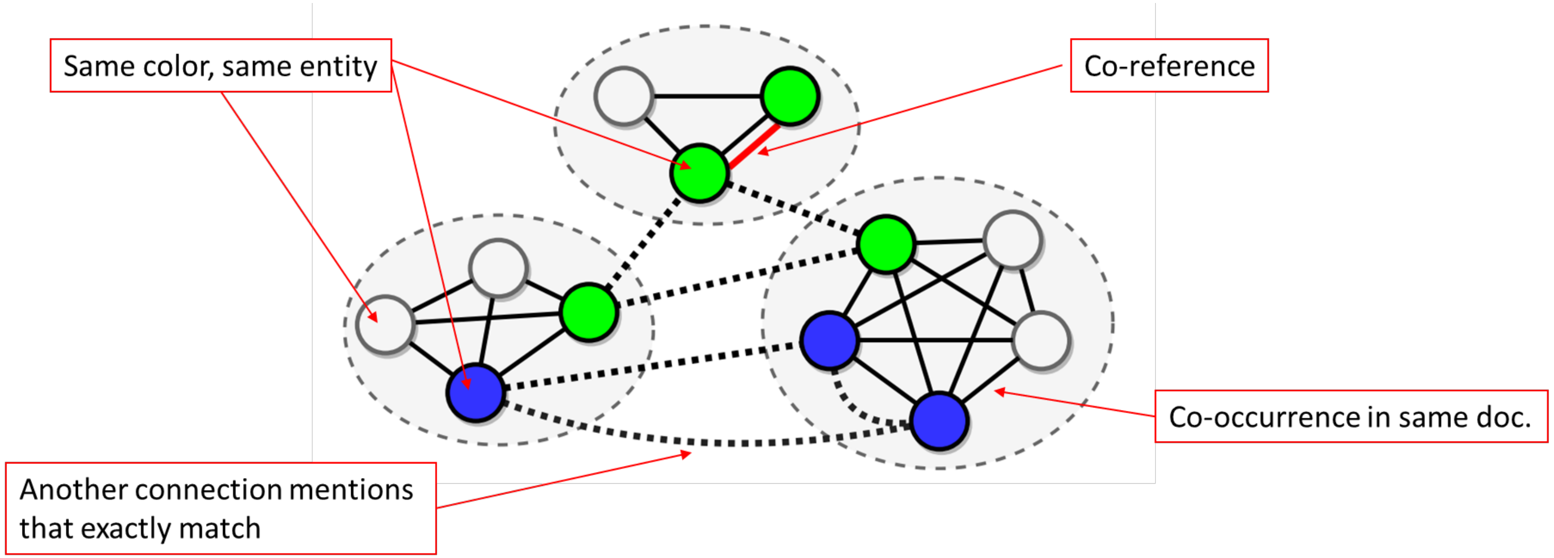
BAG: Bi-directional Attention Entity Graph



Yu Cao, Meng Fang et al., BAG: Bi-directional Attention Entity Graph Convolutional Network for Multi-hop Reasoning Question Answering, NAACL 2019.

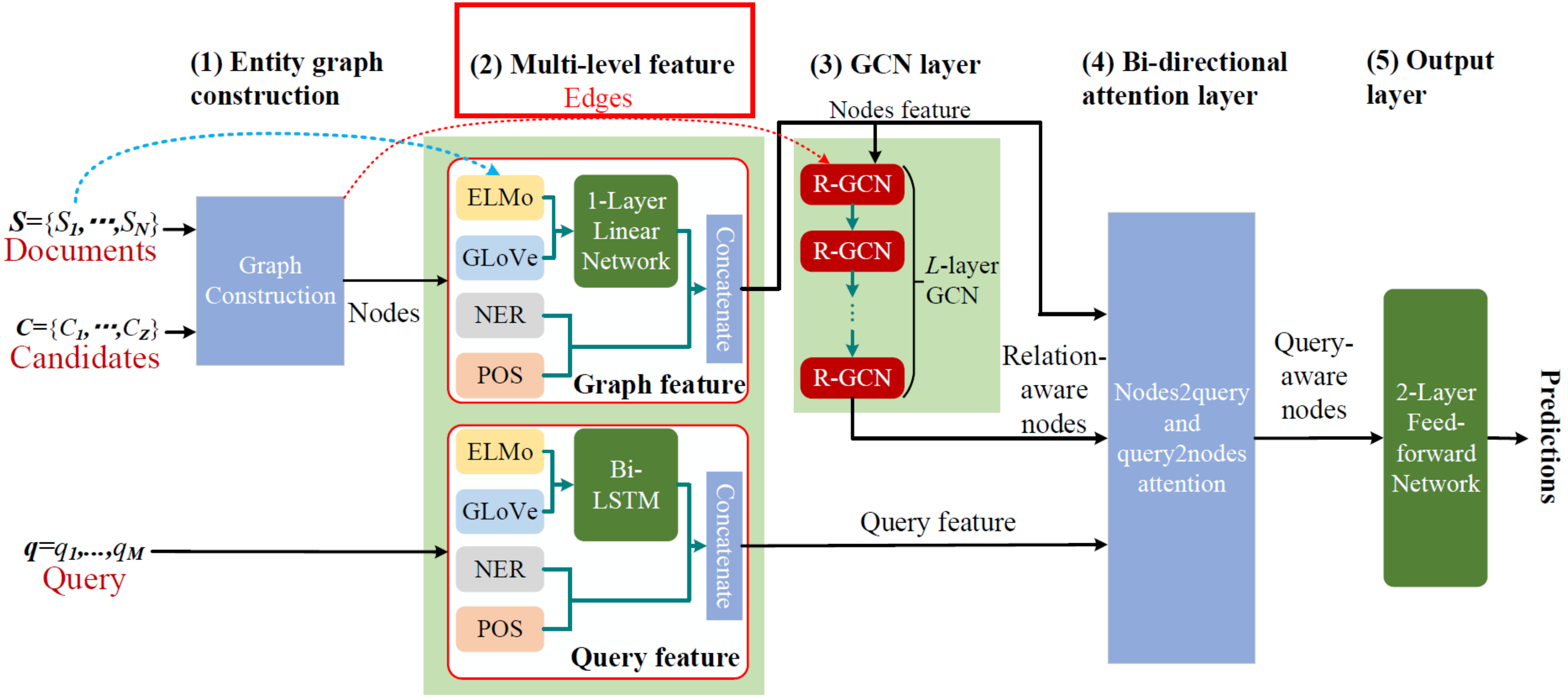
Entity Graph Construction using Entity-GCN





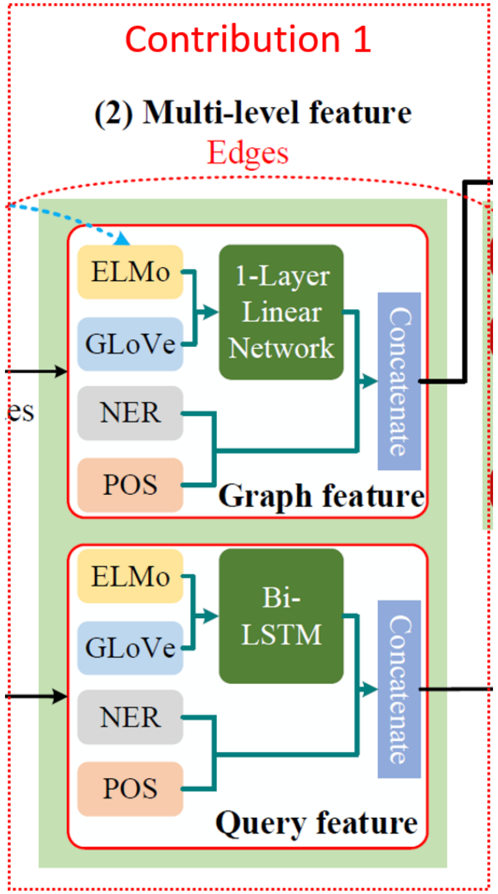
Yu Cao, Meng Fang et al., BAG: Bi-directional Attention Entity Graph Convolutional Network for Multi-hop Reasoning Question Answering, NAACL 2019.

BAG: Bi-directional Attention Entity Graph



Yu Cao, Meng Fang et al., BAG: Bi-directional Attention Entity Graph Convolutional Network for Multi-hop Reasoning Question Answering, NAACL 2019.

Multi-level features



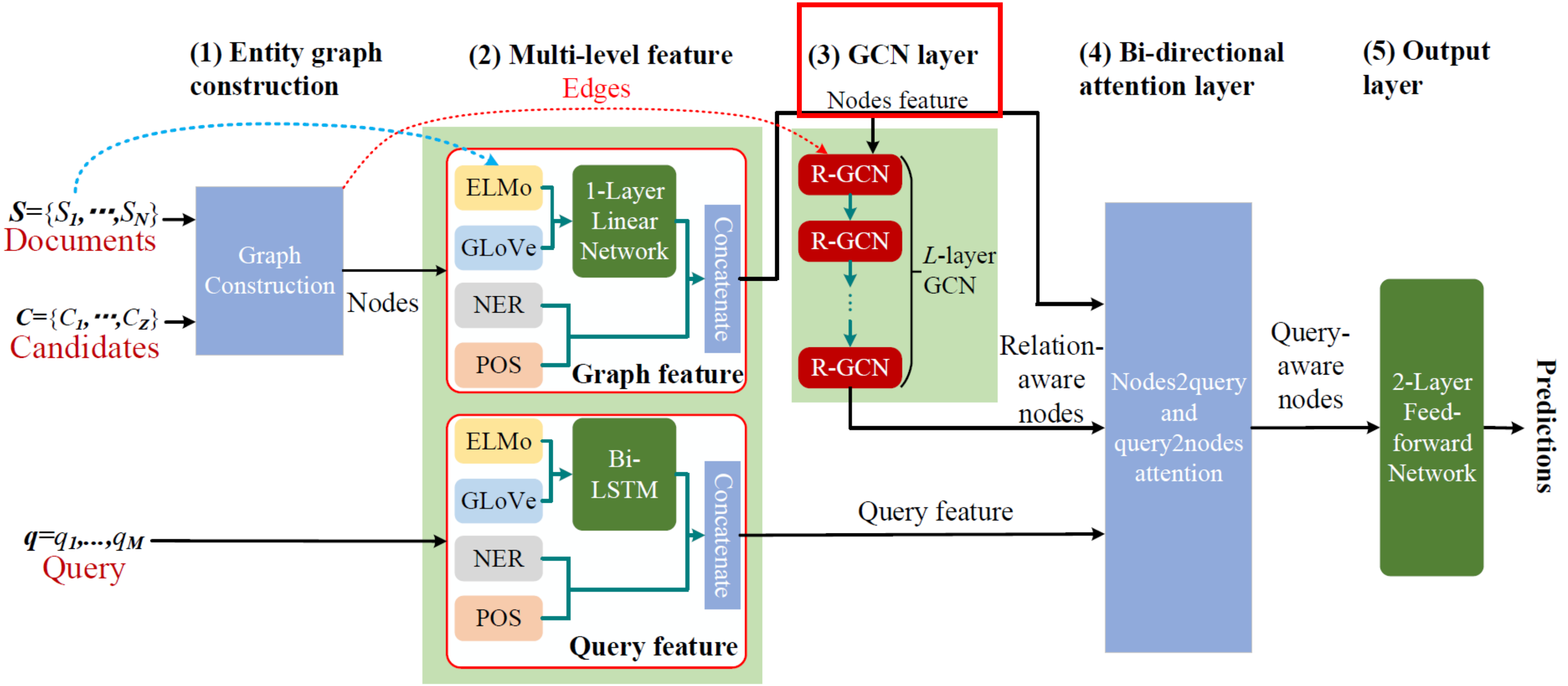
ELMo: represent tokens
GLoVe: represent contextual-level features
NER & POS ： represent semantic properties

For documents

For queries

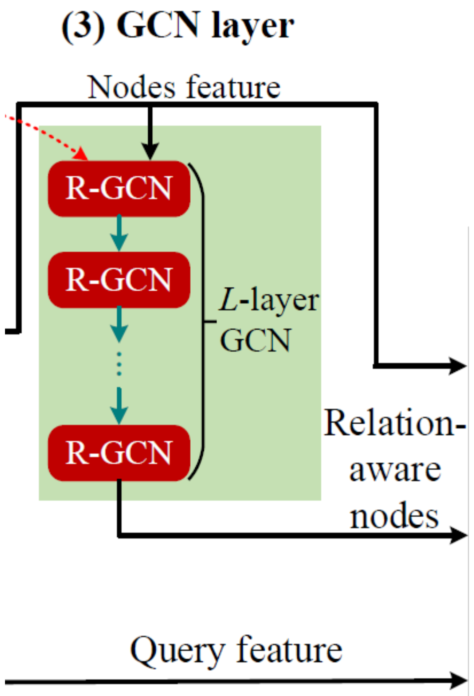
Yu Cao, Meng Fang et al., BAG: Bi-directional Attention Entity Graph Convolutional Network for Multi-hop Reasoning Question Answering, NAACL 2019.

BAG: Bi-directional Attention Entity Graph



Yu Cao, Meng Fang et al., BAG: Bi-directional Attention Entity Graph Convolutional Network for Multi-hop Reasoning Question Answering, NAACL 2019.

GCN layer using R-GCN



Standard GCN：

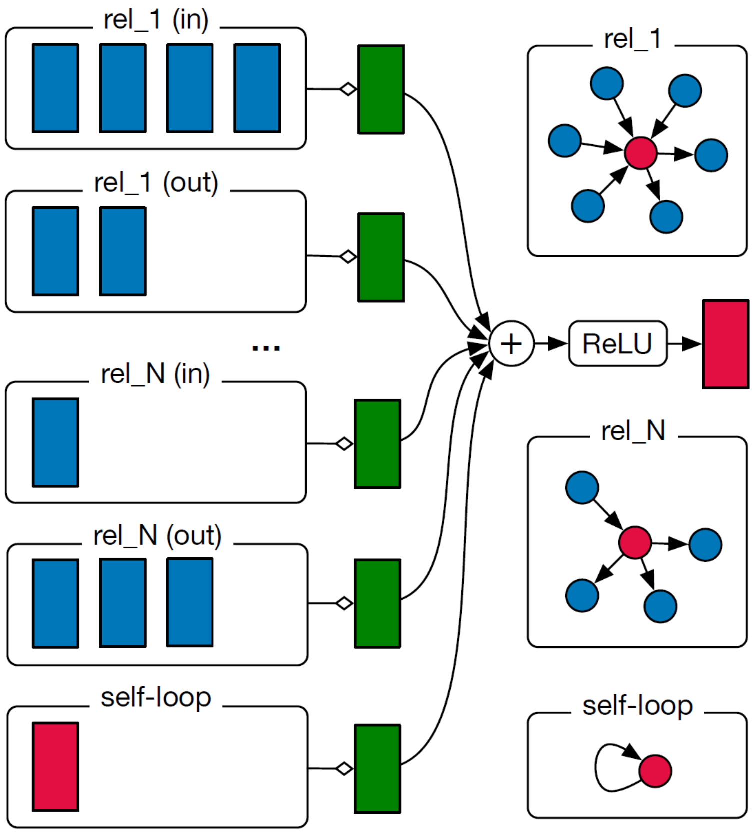
$$h_i^{(l+1)} = \sigma\left(\sum_{m \in \mathcal{M}_i} g_m(h_i^{(l)}, h_j^{(l)})\right)$$

$$g_m(h_i, h_j) = W h_j$$

R-GCN：

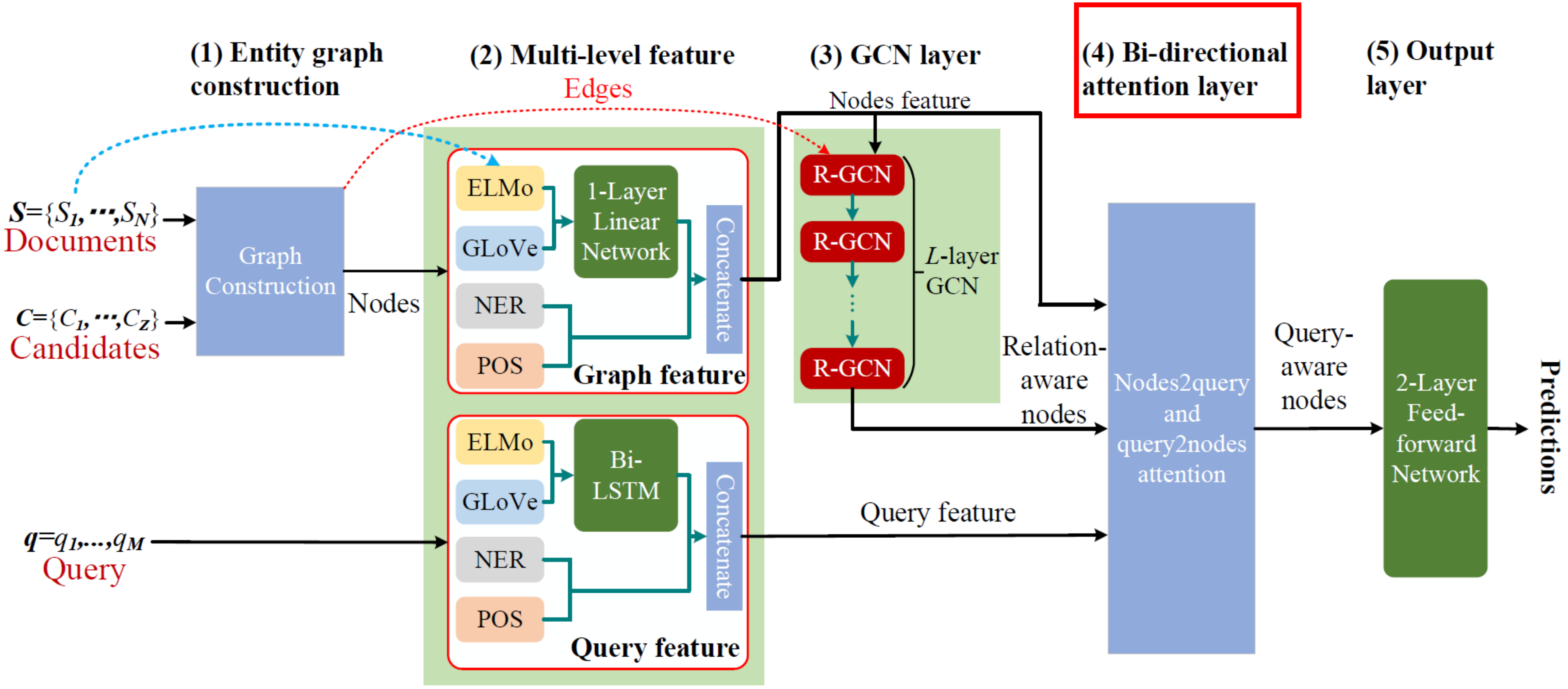
$$h_i^{(l+1)} = \sigma\left(\sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^r} \frac{1}{c_{i,r}} W_r^{(l)} h_j^{(l)} + W_0^{(l)} h_i^{(l)}\right)$$

Edges information



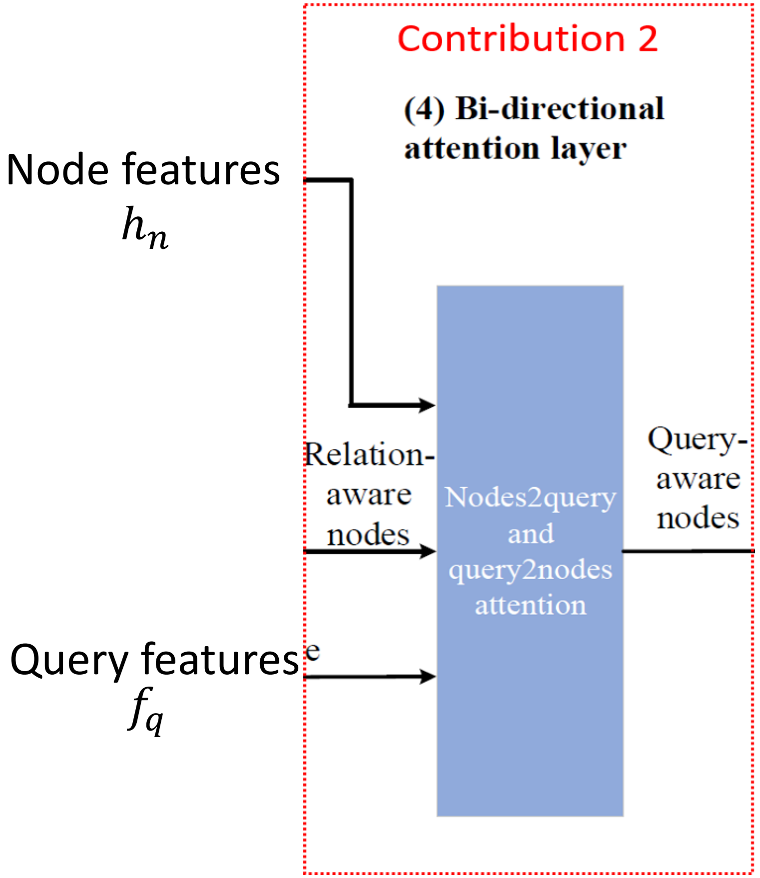
Yu Cao, Meng Fang et al., BAG: Bi-directional Attention Entity Graph Convolutional Network for Multi-hop Reasoning Question Answering, NAACL 2019.

BAG: Bi-directional Attention Entity Graph



Yu Cao, Meng Fang et al., BAG: Bi-directional Attention Entity Graph Convolutional Network for Multi-hop Reasoning Question Answering, NAACL 2019.

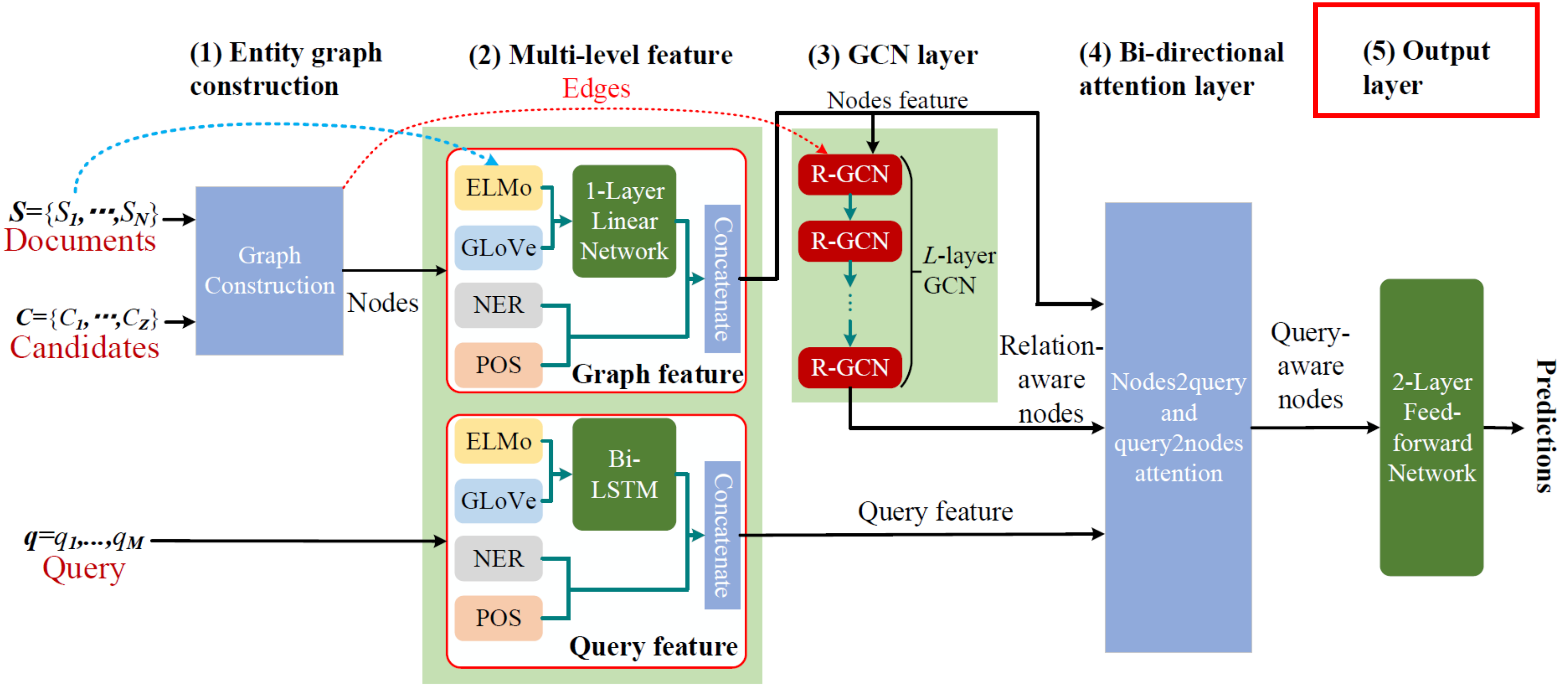
Bi-directional attention layer



Attention -- aggregation
Tradition attention: weighted sum of one variable
Bi-directional attention or co-attention:
- on two variables
- integrate both as key
- weighted sum of each variable

Yu Cao, Meng Fang et al., BAG: Bi-directional Attention Entity Graph Convolutional Network for Multi-hop Reasoning Question Answering, NAACL 2019.

BAG: Bi-directional Attention Entity Graph



Yu Cao, Meng Fang et al., BAG: Bi-directional Attention Entity Graph Convolutional Network for Multi-hop Reasoning Question Answering, NAACL 2019.

Experiments

Models	Unmasked			Masked	
	dev	test	test[1]	dev	test[1]
FastQA	27.2*	-	38.5	38.0*	48.3
BiDAF	49.7*	-	45.2	59.8*	57.5
Coref-GRU†	56.0*	59.3	57.2	-	-
MHQA-GRN‡	62.8*	65.4	-	-	-
Entity-GCN	64.8*	67.6	63.1	70.5*	68.1
BAG	**66.5**	**69.0**	**65.7**	**70.9**	**68.9**

WikiHop (http://qangaroo.cs.ucl.ac.uk/)
Yu Cao, Meng Fang et al., BAG: Bi-directional Attention Entity Graph Convolutional Network for Multi-hop Reasoning Question Answering, NAACL 2019.

Ablation Experiments of BAG model

Models	**Unmasked**
Without Attention	63.1
Using Single Attention	64.5
Without GCN	63.3
Without edge type	63.9
Without NER, POS	66.0
+Without ELMo	60.5
Full Model	**66.5**

Yu Cao, Meng Fang et al., BAG: Bi-directional Attention Entity Graph Convolutional Network for Multi-hop Reasoning Question Answering, NAACL 2019.

Multi-hop Reading Comprehension by Reasoning over Heterogeneous Graphs

Query: record_label get ready
Support doc 1: Mason Durell Betha (born August 27, 1977), better known by stage name Mase (formerly often stylized Ma$e or MA$E), is an American hip hop recording artist and minister. He is best known for being signed to Sean "Diddy" Combs's label Bad Boy Records. ...

Support doc 2: "Get Ready" was the only single released from Mase's second album, Double Up. It was released on May 25, 1999, produced by Sean "Puffy" Combs, Teddy Riley and Andreao "Fanatic" Heard and featured R&B group, Blackstreet, it contains a sample of "A Night to Remember", performed by Shalamar. ...

Support doc 3: Bad Boy Entertainment (also known as Bad Boy Records) is an American record label founded in 1993 by Sean Combs. ...

Candidates: bad boy records, record label, rock music, ...
Answer: bad boy records

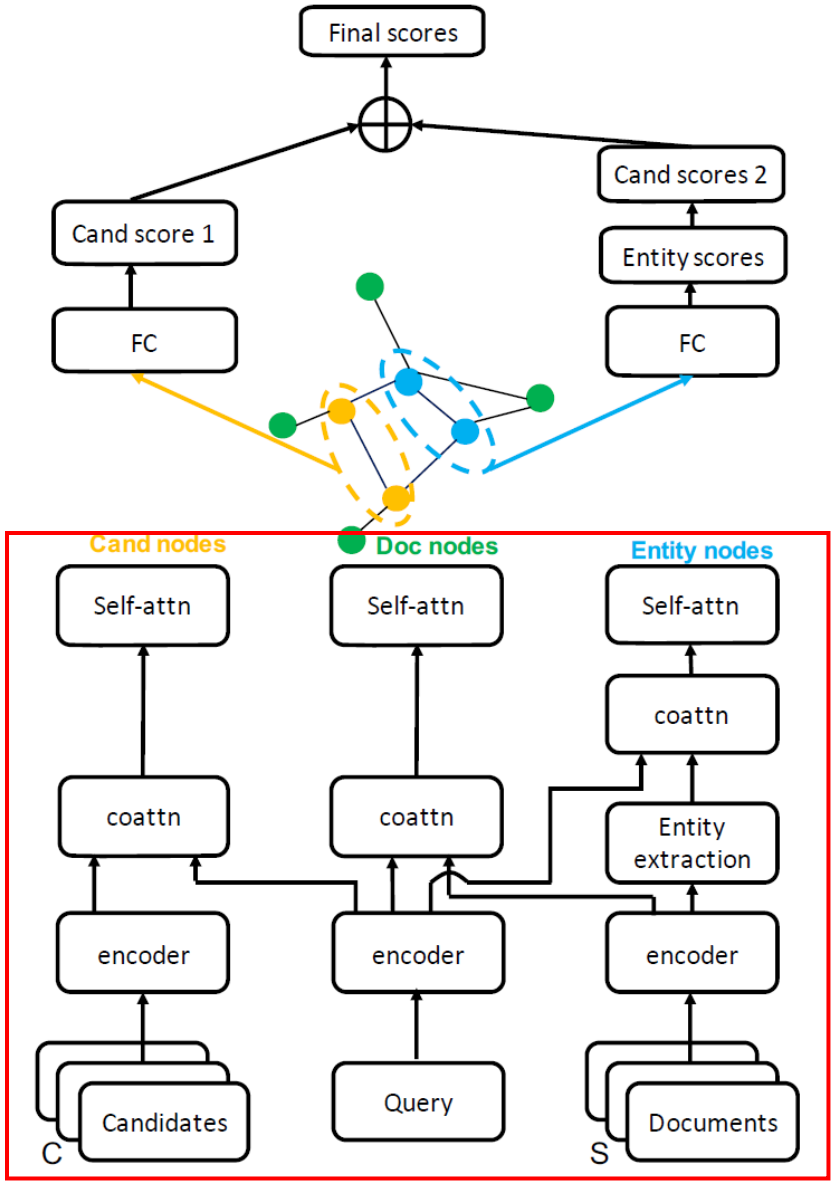
Heterogeneous Document-Entity (HDE) graph.

Feature integration over
- Candidates
- Entities
- Documents

Ming Tu et al., Multi-hop Reading Comprehension across Multiple Documents by Reasoning over Heterogeneous Graphs, ACL 2019.

HDE graph framework

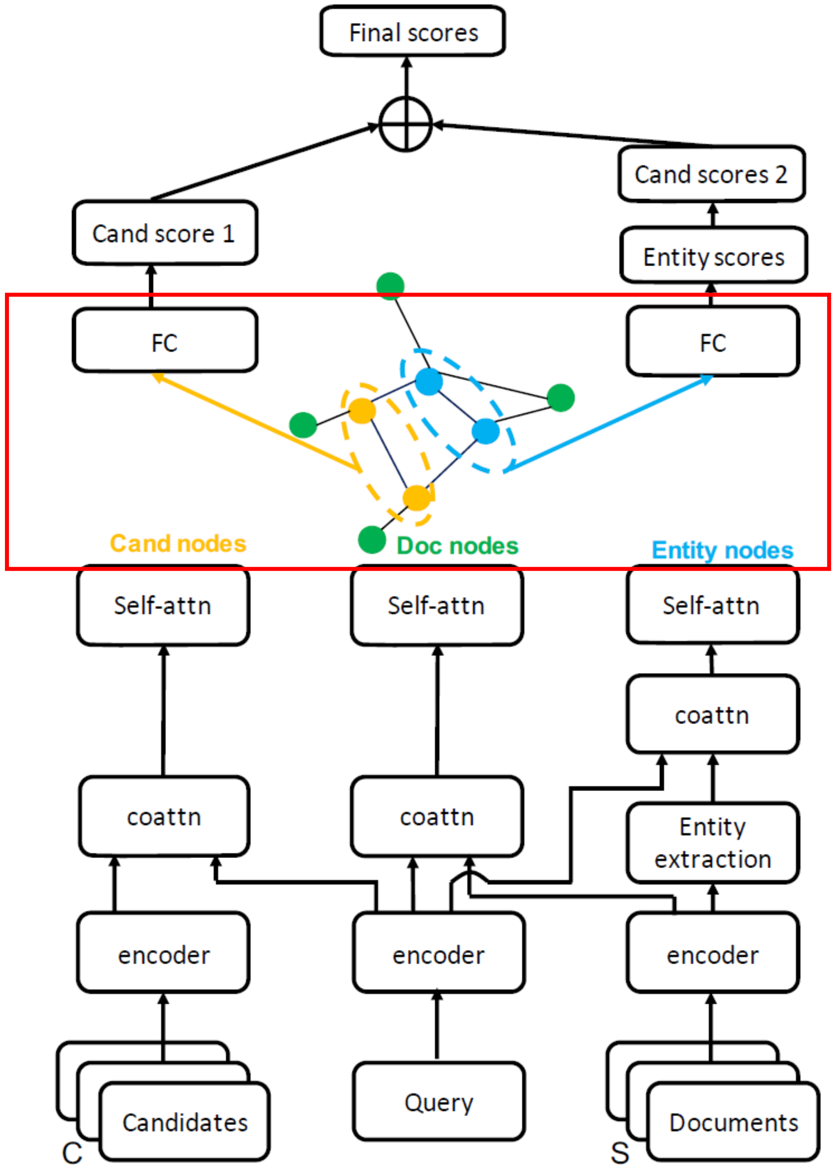
- Score accumulation.
- Reasoning over HDE graph with GNN.
- Initialize HDE graph nodes with co-attention and self-attention based context coding.



Ming Tu et al., Multi-hop Reading Comprehension across Multiple Documents by Reasoning over Heterogeneous Graphs, ACL 2019.

HDE graph framework

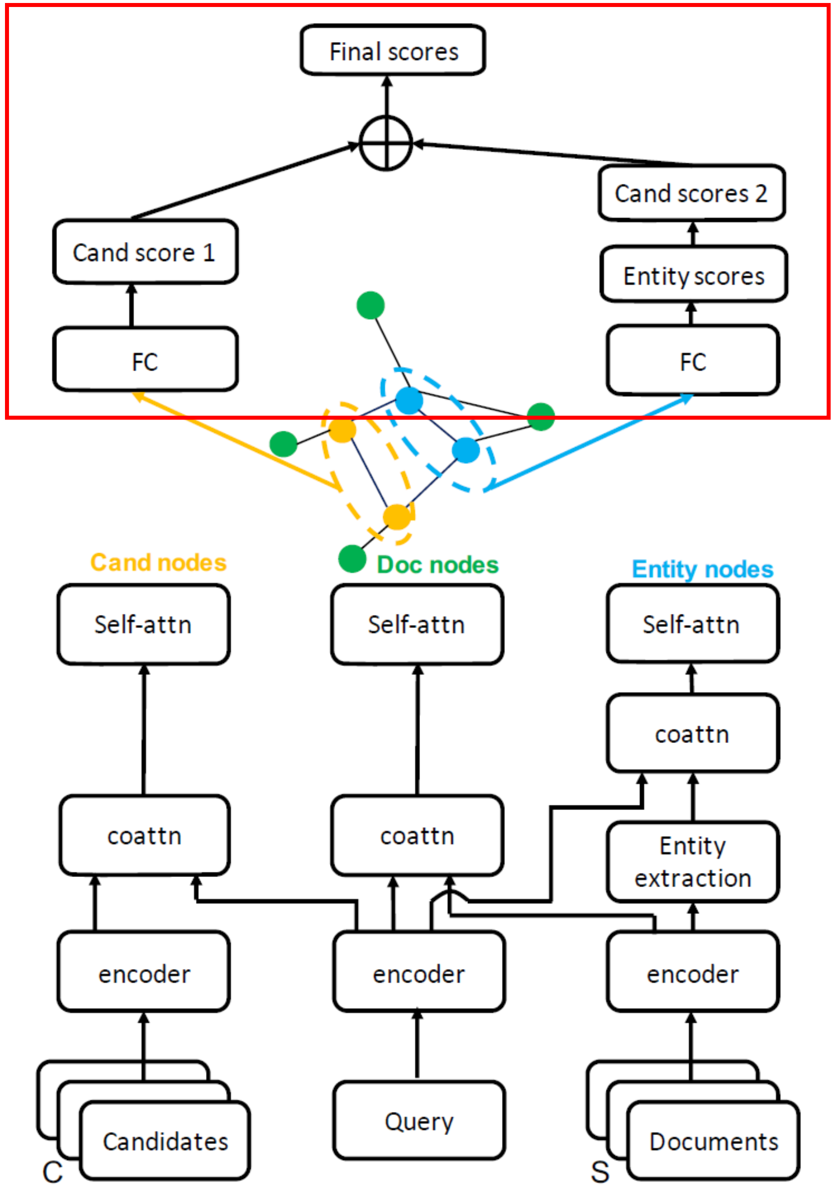
- Score accumulation.
- Reasoning over HDE graph with GNN.
- Initialize HDE graph nodes with co-attention and self-attention based context coding.



Ming Tu et al., Multi-hop Reading Comprehension across Multiple Documents by Reasoning over Heterogeneous Graphs, ACL 2019.

HDE graph framework

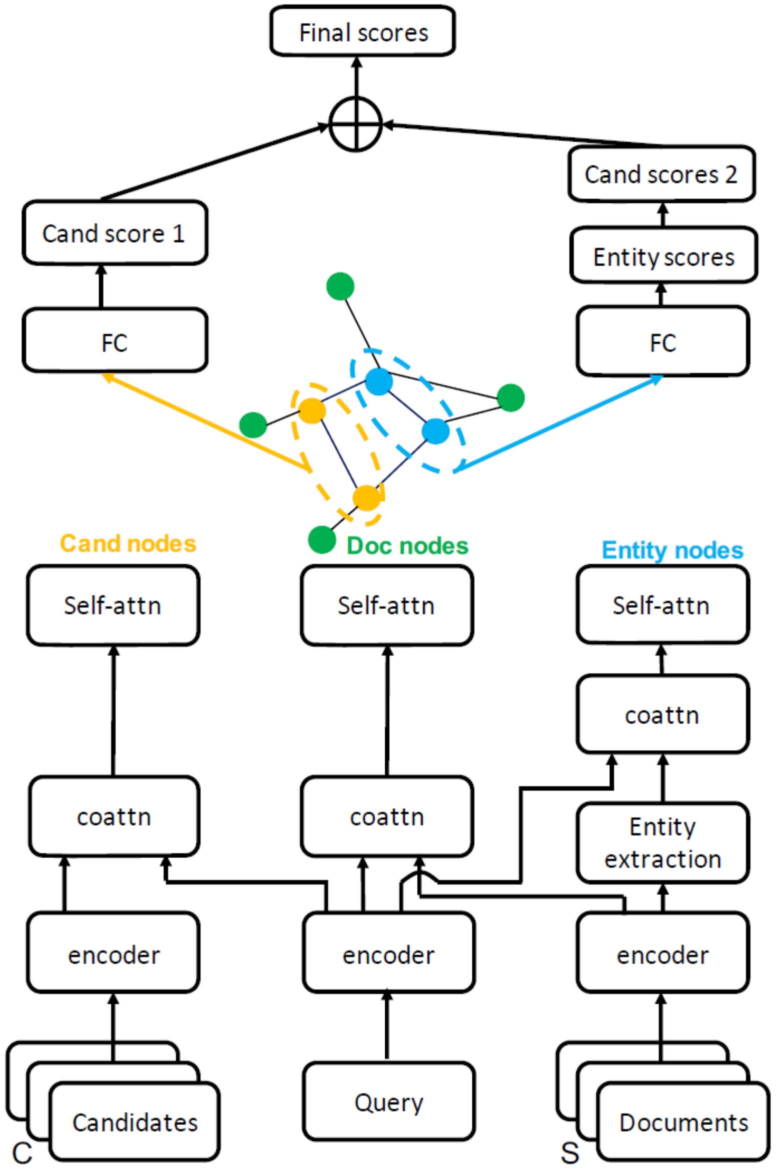
- Score accumulation.
- Reasoning over HDE graph with GNN.
- Initialize HDE graph nodes with co-attention and self-attention based context coding.



Ming Tu et al., Multi-hop Reading Comprehension across Multiple Documents by Reasoning over Heterogeneous Graphs, ACL 2019.

HDE graph framework

- Score accumulation.
- Reasoning over HDE graph with GNN.
- Initialize HDE graph nodes with co-attention and self-attention based context coding.

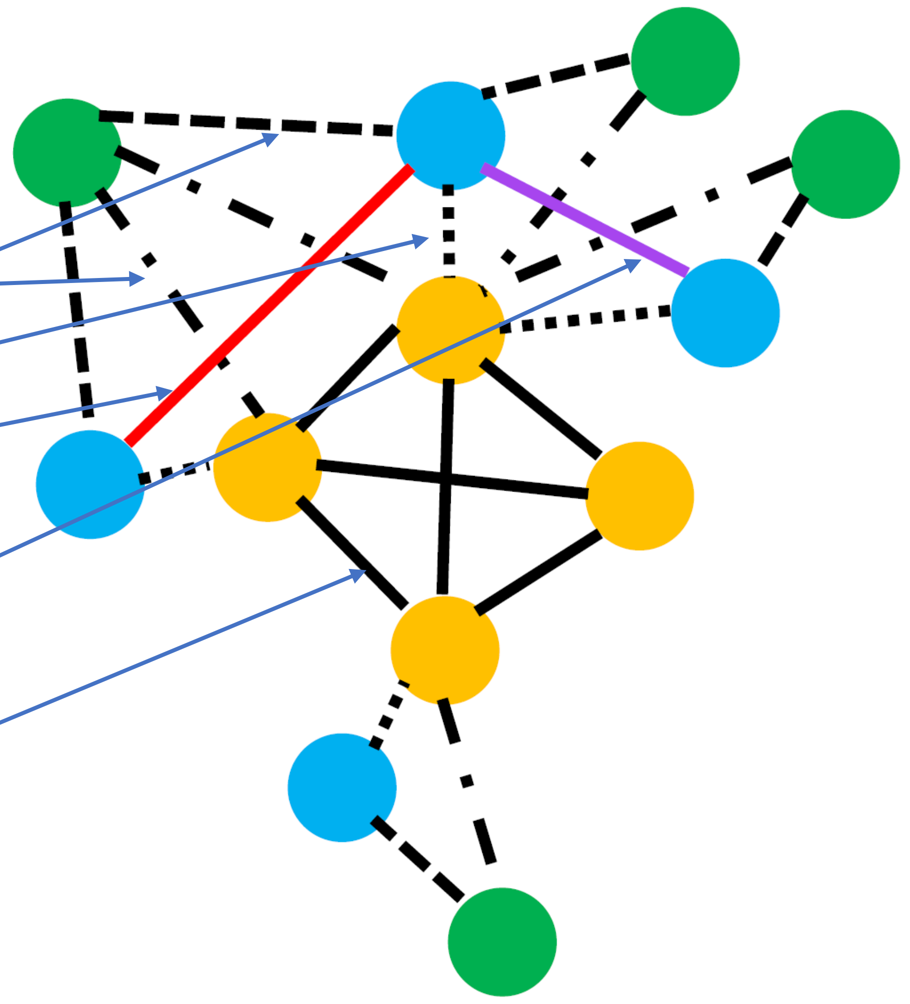


Ming Tu et al., Multi-hop Reading Comprehension across Multiple Documents by Reasoning over Heterogeneous Graphs, ACL 2019.

Reasoning with HDE graph

Documents Candidates Entities

1. candidate appear in the document.
2. the entity is extracted from the document.
3. the entity is a mention of the candidate.
4. Entities are extracted from the same document.
5. Entities are mentions of the same candidate or query subject and they are extracted from different documents.
6. all candidate nodes connect with each other.
7. entity nodes that do not meet previous conditions are connected. (ignored for good visualization)



Ming Tu et al., Multi-hop Reading Comprehension across Multiple Documents by Reasoning over Heterogeneous Graphs, ACL 2019.

Experiments

Single models	Accuracy (%)	
	Dev	Test
BiDAF	-	42.9
Coref-GRU(Dhingra et al., 2018)	56.0	59.3
MHQA-GRN(Song et al., 2018)	62.8	65.4
Entity-GCN(De Cao et al., 2018)	64.8	67.6
CFC(Zhong et al., 2019)	66.4	70.6
Kundu et al. (2018)	67.1	-
DynSAN*	-	**71.4**
Proposed	**68.1**	70.9
Ensemble models		
Entity-GCN(De Cao et al., 2018)	68.5	71.2
DynSAN*	-	73.8
Proposed	**70.9**	**74.3**

WikiHop

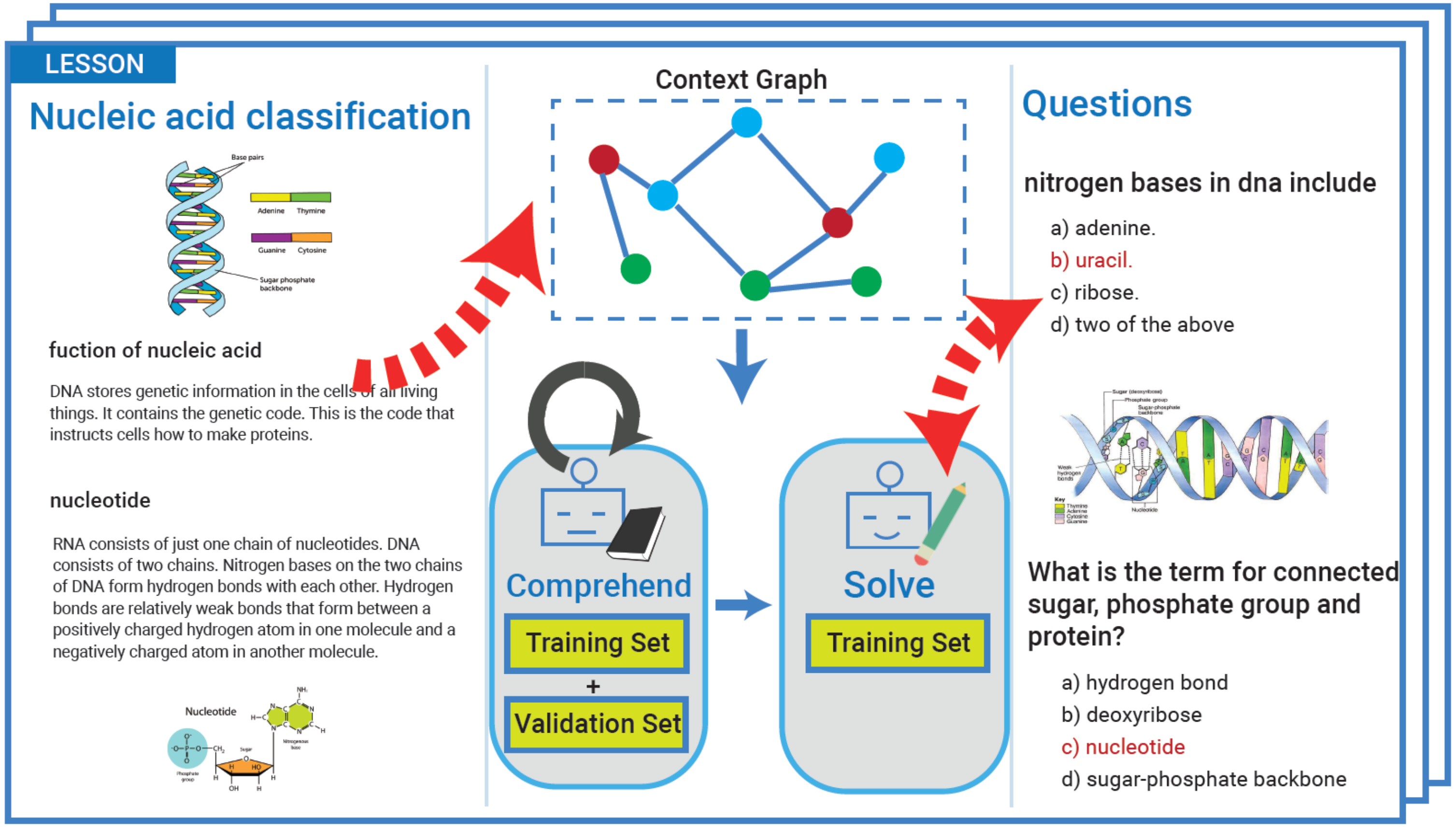
Ming Tu et al., Multi-hop Reading Comprehension across Multiple Documents by Reasoning over Heterogeneous Graphs, ACL 2019.

Experiments

Model	Accuracy (%)	
	Dev	Δ
Full model	**68.1**	-
- HDE graph	65.5	2.6
- different edge types	66.7	1.4
- candidate nodes scores	67.1	1.0
- entity nodes scores	66.6	1.5
- candidate nodes	66.2	1.9
- document nodes	67.6	0.5
- entity nodes	63.6	4.5

Ming Tu et al., Multi-hop Reading Comprehension across Multiple Documents by Reasoning over Heterogeneous Graphs, ACL 2019.

Textbook Question Answering



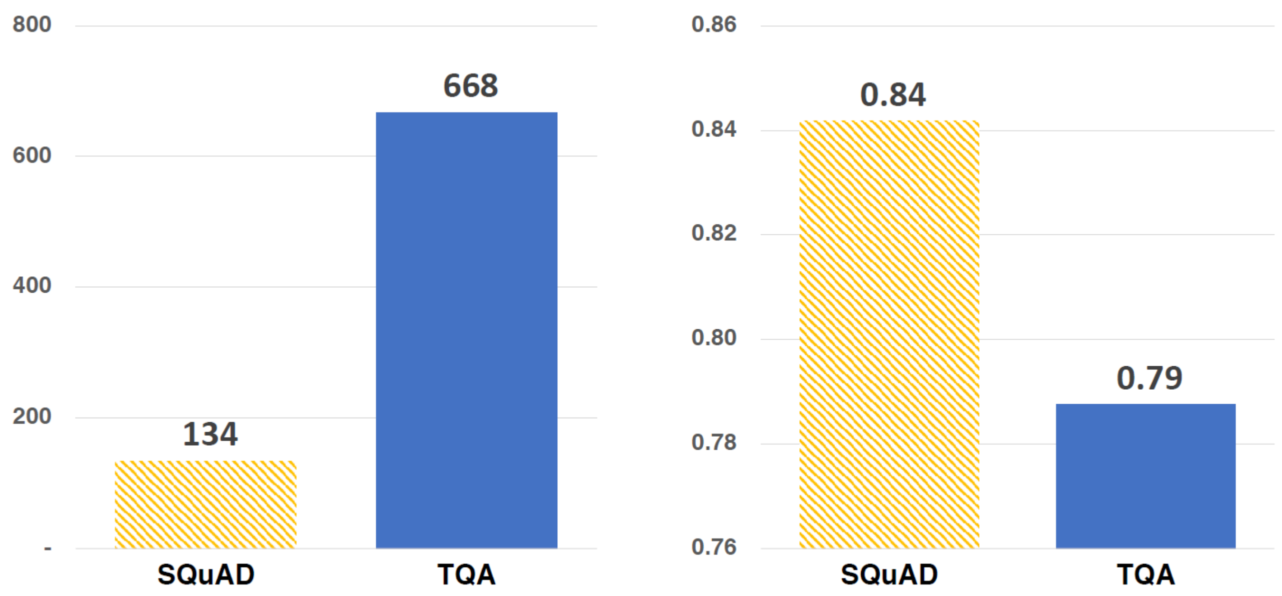
Daesik Kim, Seonhoon Kim and Nojun Kwak, Textbook Question Answering with Multi-modal Context Graph Understanding and Self-supervised Open-set Comprehension, ACL 2019.

Textbook Question Answering

Input Type		Context QA	Visual QA	Textbook QA
Context Part	Text	○	-	○
	Image	-	○	○
Question Part	Text	○	○	○
	Image	-	-	○

Daesik Kim, Seonhoon Kim and Nojun Kwak, Textbook Question Answering with Multi-modal Context Graph Understanding and Self-supervised Open-set Comprehension, ACL 2019.

Textbook Question Answering



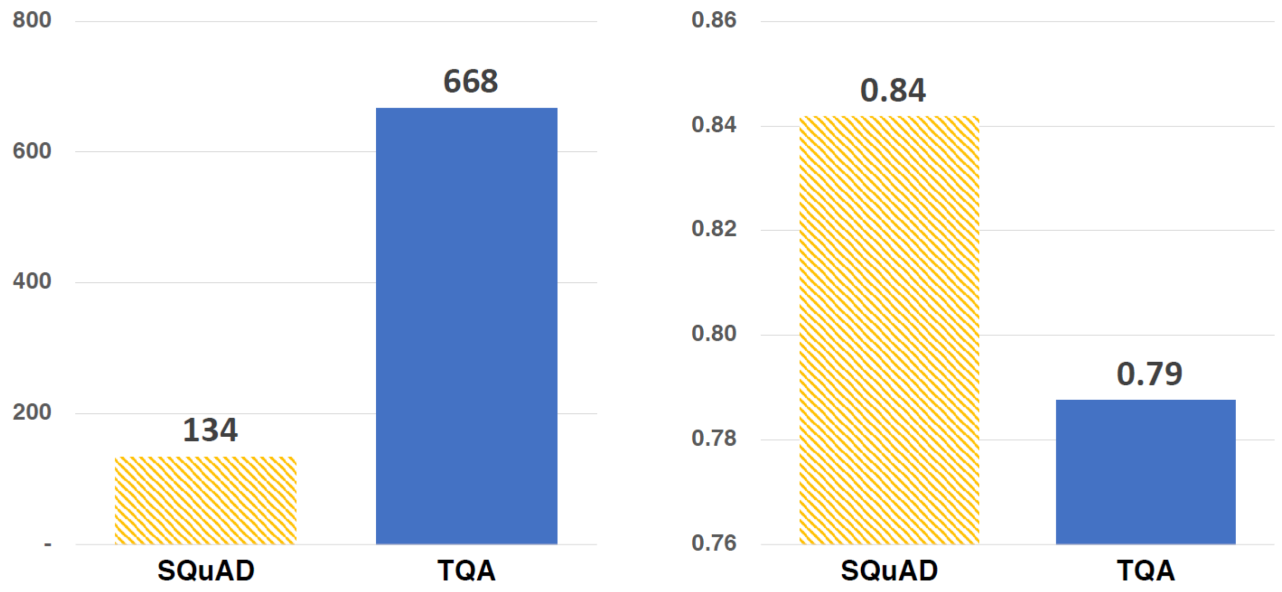
a) Average length of contexts

b) Ratio of words in valset that appear in trainset

- More complexity in data format and length.
- It is very difficult to solve problems that have not been studied before

Daesik Kim, Seonhoon Kim and Nojun Kwak, Textbook Question Answering with Multi-modal Context Graph Understanding and Self-supervised Open-set Comprehension, ACL 2019.

Textbook Question Answering



a) Average length of contexts

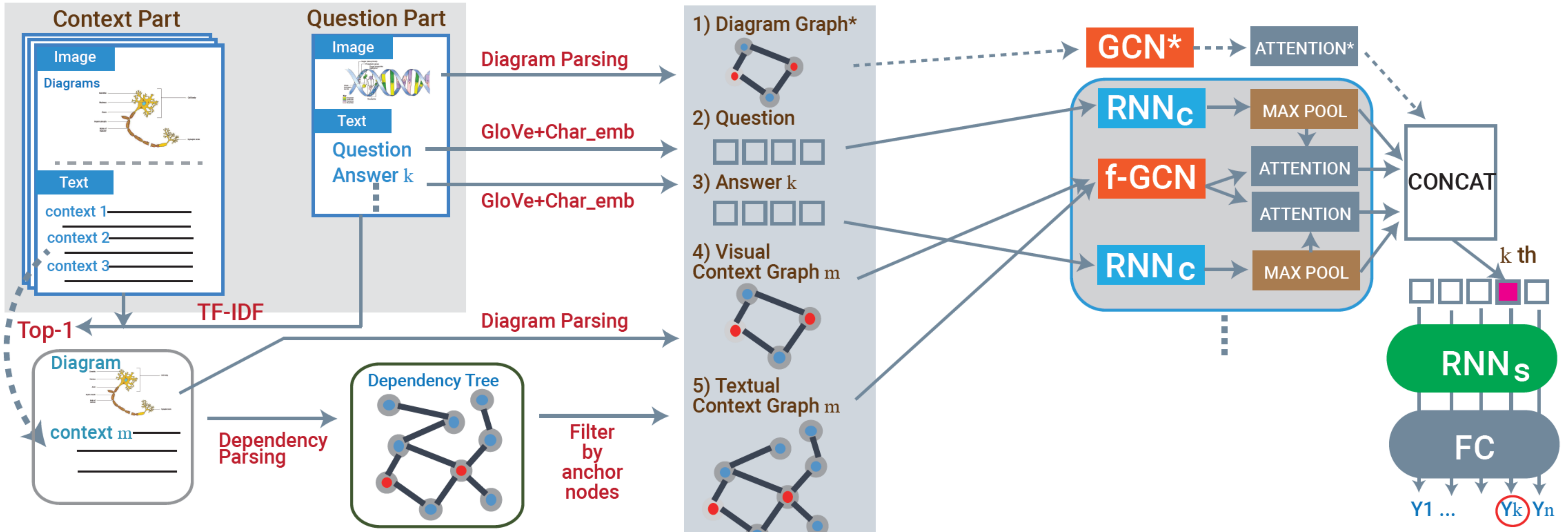
b) Ratio of words in valset that appear in trainset

- A fusion GCN (f-GCN)
- A novel self-supervised open-set comprehension (SSOC)

Daesik Kim, Seonhoon Kim and Nojun Kwak, Textbook Question Answering with Multi-modal Context Graph Understanding and Self-supervised Open-set Comprehension, ACL 2019.

Overall framework

$$\hat{a} = \underset{a \in \Omega_a}{\operatorname{argmax}}\, p(a|C, q; \theta)$$



a) Preparation step for k-th answer among n candidate

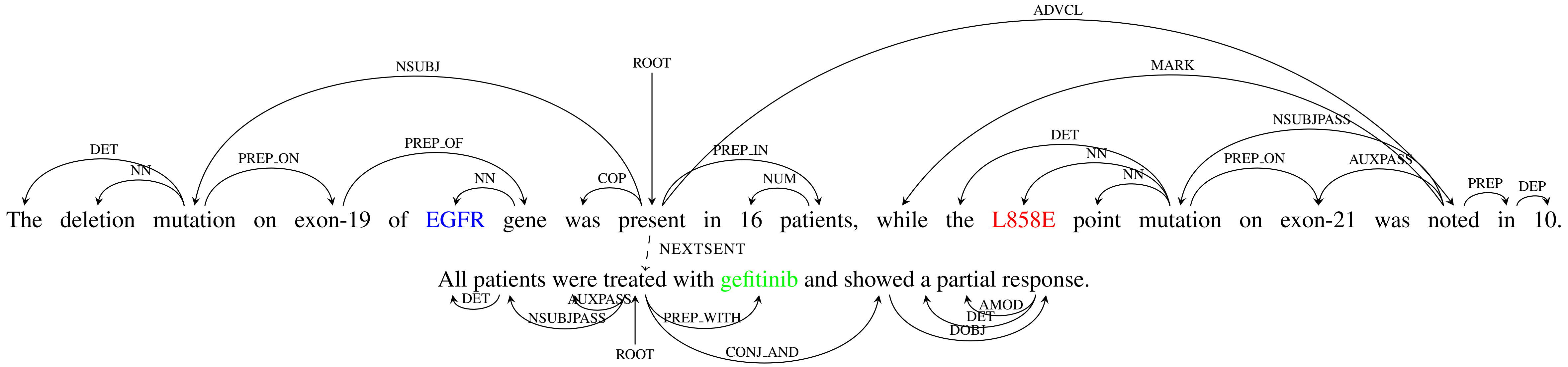
b) Embedding step and Solving step

Daesik Kim, Seonhoon Kim and Nojun Kwak, Textbook Question Answering with Multi-modal Context Graph Understanding and Self-supervised Open-set Comprehension, ACL 2019.

Solving problems using graph encoding

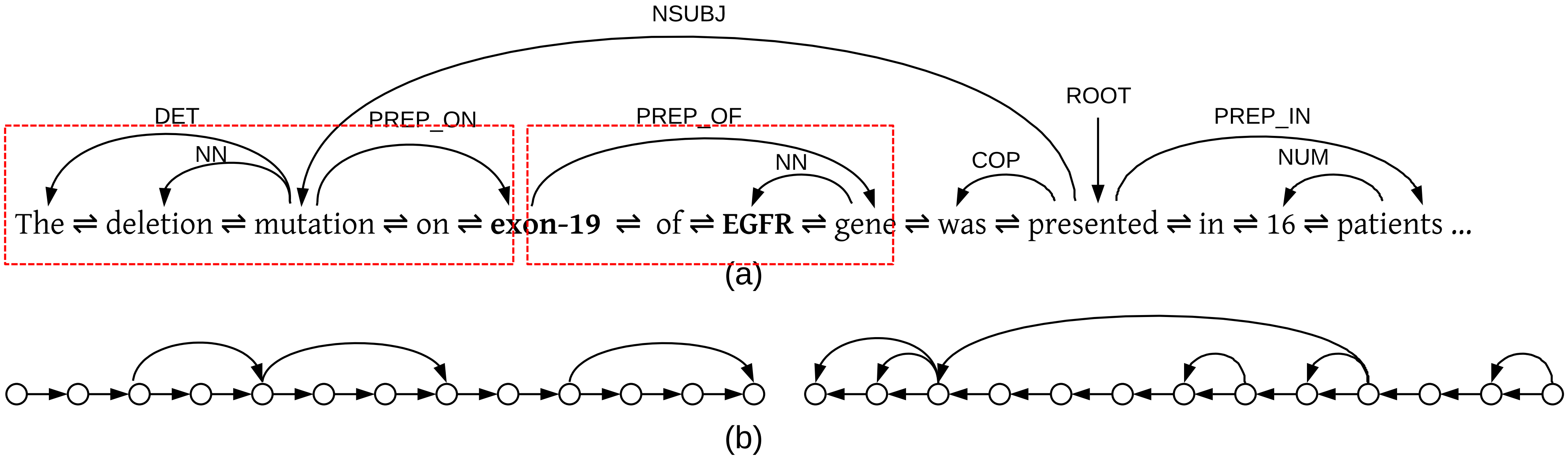
- Neural Machine Translation
- Text Generation
- Question Answering
- Information Extraction
- Sentiment
- Social Classification
- Syntactic Parsing
- Semantic Role Labeling
- Word Embedding
- Sentence Representation

Cross-sentence *N*-ary Relation Extraction



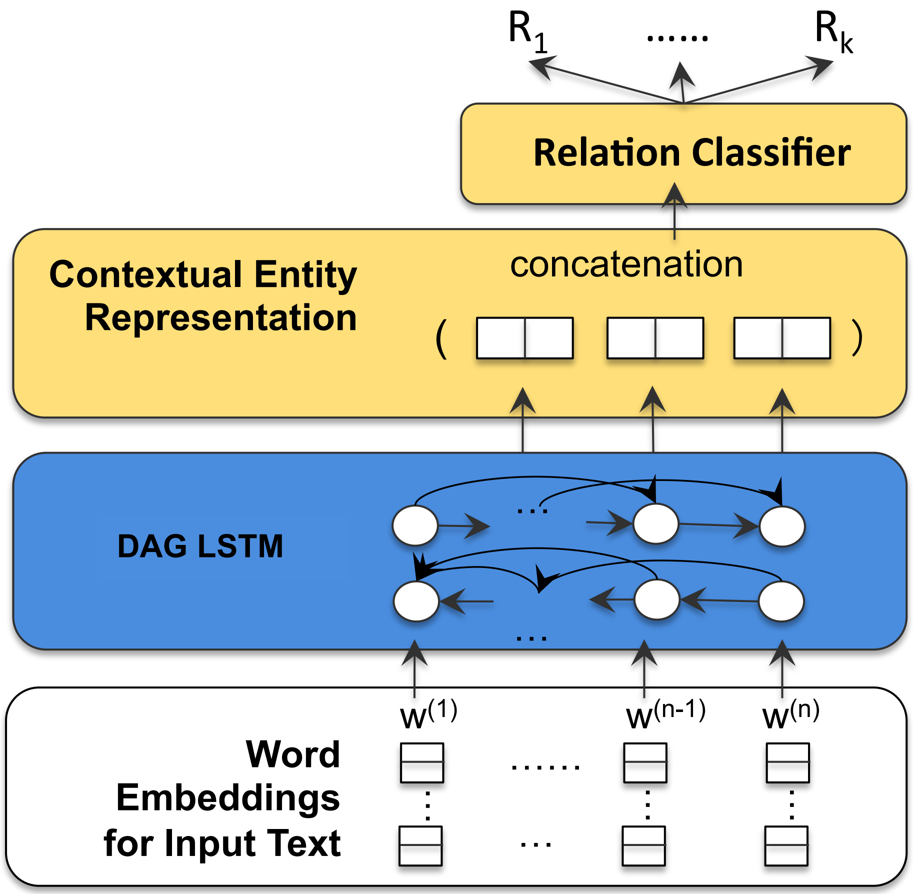
Linfeng Song, Yue Zhang et al., N-ary Relation Extraction using Graph State LSTM, EMNLP 2018.

Previous SOTA: DAG LSTM



Nanyun Peng, Hoifung Poon et al., Cross-sentence n-ary relation extraction with graph LSTMs, TACL 2017.
Linfeng Song, Yue Zhang et al., N-ary Relation Extraction using Graph State LSTM, EMNLP 2018.

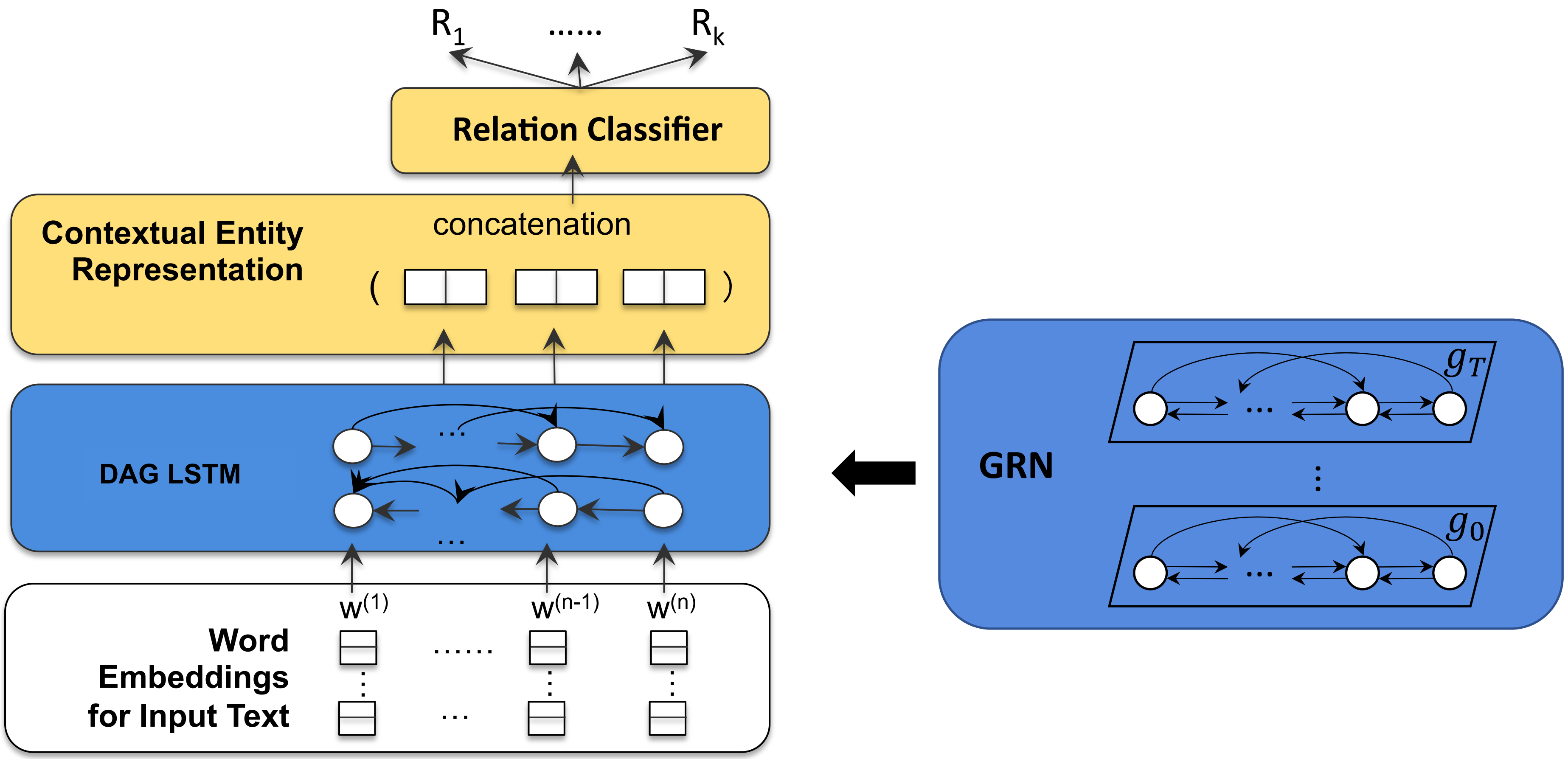
Overall framework



Linfeng Song, Yue Zhang et al., N-ary Relation Extraction using Graph State LSTM, EMNLP 2018.

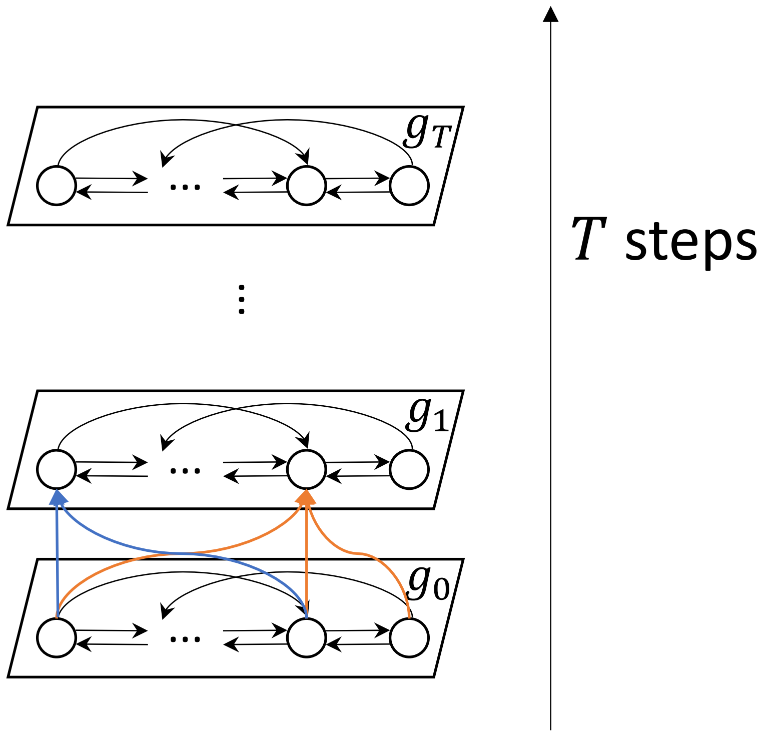
Overall framework

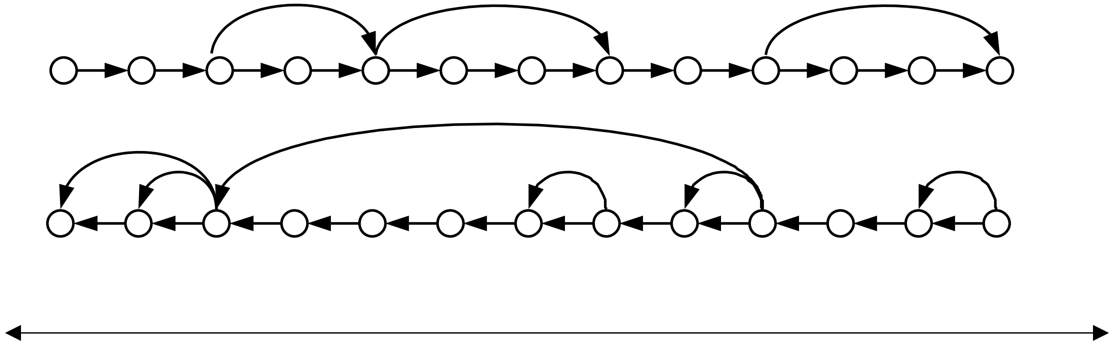
Code available at:
https://github.com/freesunshine0316/nary-grn



Linfeng Song, Yue Zhang et al., N-ary Relation Extraction using Graph State LSTM, EMNLP 2018.

Efficiency of GRN versus DAG networks





$T << N$

Linfeng Song, Yue Zhang et al., N-ary Relation Extraction using Graph State LSTM, EMNLP 2018.

Experiments

- Evaluate on the corpus by Peng et al., (2017), with annotations of dependency, discourse and entity boundaries.
  - Ternary (drug, gene, mutation): 6987 instances (Avg. length: 73.9)
  - Binary (drug, mutation): 6087 instances (Avg. length: 61.0)

- Message passing step T=5, as determined by a DEV experiment

- Evaluation (Peng et al., 2017):
  - 5-fold validation
  - Classification accuracy

Linfeng Song, Yue Zhang et al., N-ary Relation Extraction using Graph State LSTM, EMNLP 2018.

Main results

Model	Precision (%)
Peng et al. (2017)	80.7
Peng et al. (2017) + Multi-task	82.0
Bidir DAG LSTM	77.3
GRN	**83.2***

Ternary

Model	Precision (%)
Peng et al. (2017)	76.7
Peng et al. (2017) + Multi-task	78.5
Bidir DAG LSTM	76.4
GRN	**83.6***

Binary

Linfeng Song, Yue Zhang et al., N-ary Relation Extraction using Graph State LSTM, EMNLP 2018.

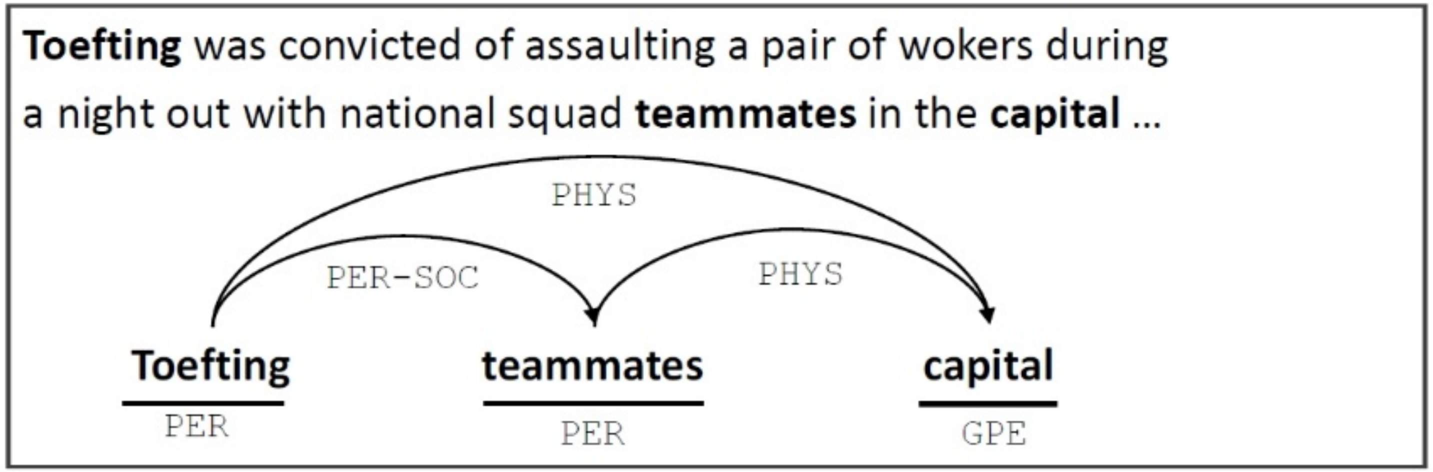
Efficiency (Ternary)

Model	Train	Decode
Bidir DAG LSTM	281s	27.3s
GRN	36.7s (7.7 times faster)	2.7s (10 times faster)

Average sentence length: 75
Message passing step: 5

Linfeng Song, Yue Zhang et al., N-ary Relation Extraction using Graph State LSTM, EMNLP 2018.

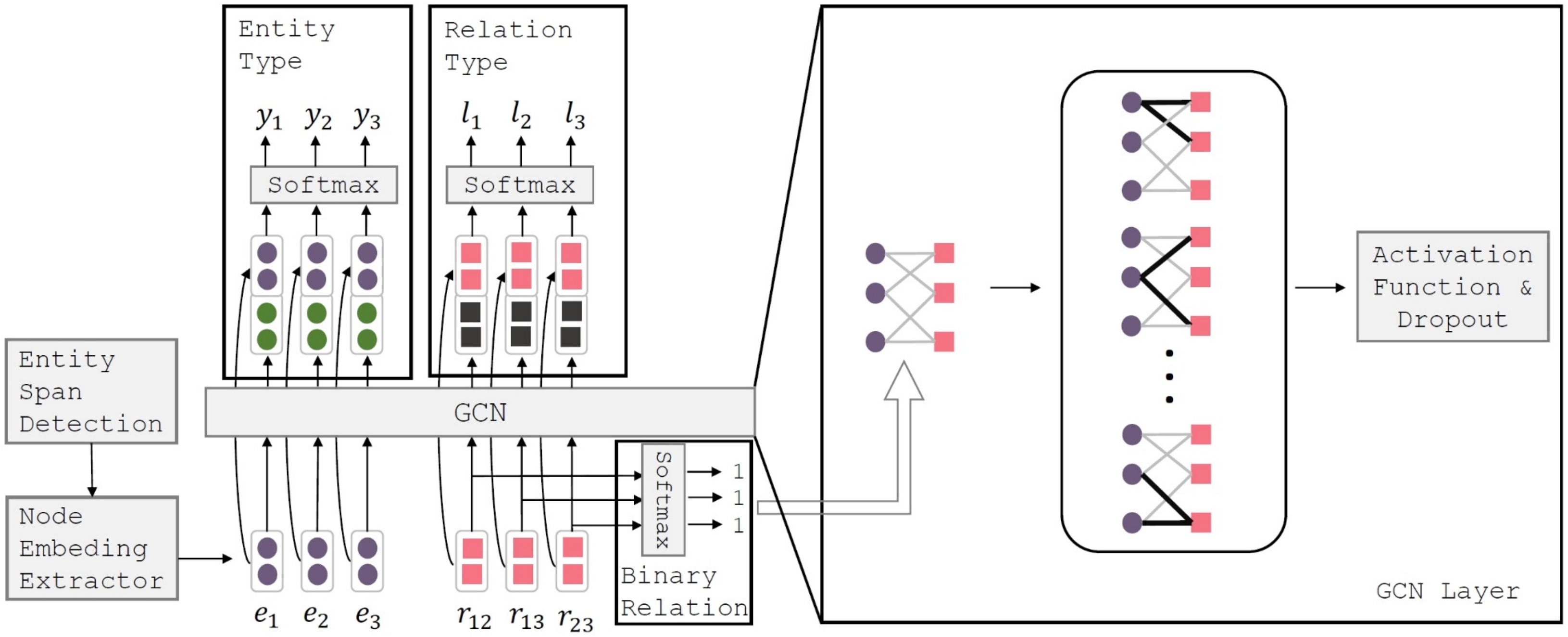
Joint Type Inference on Entities and Relation



Changzhi Sun, Yeyun Gong et al., Joint Type Inference on Entities and Relation via Graph Convolutional Networks.

Joint Type Inference on Entities and Relation

Four tasks: entity span → relation link → entity type → relation type



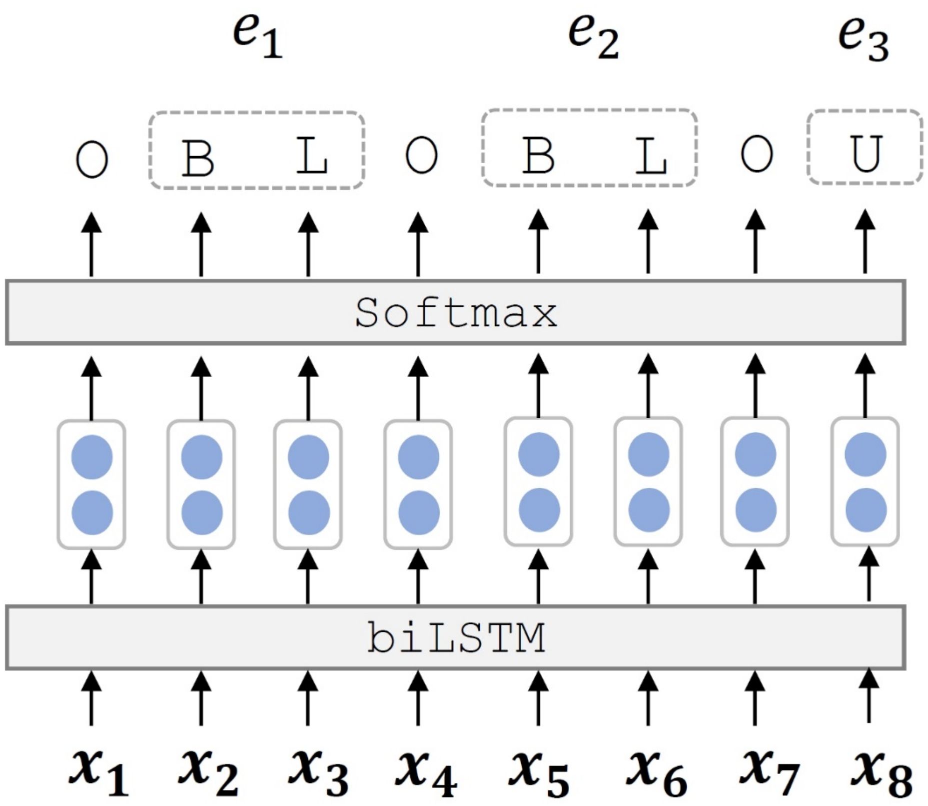
Changzhi Sun, Yeyun Gong et al., Joint Type Inference on Entities and Relation via Graph Convolutional Networks.

Entity span extraction

$$\mathbf{h}_i = \text{biLSTM}(\mathbf{x}_i; \theta_{\text{seq}})$$

$$P(\hat{t}_i | s) = \text{Softmax}(\mathbf{W}_{\text{span}} \mathbf{h}_i)$$

$$\mathcal{L}_{\text{span}} = -\frac{1}{|s|} \sum_{i=1}^{|s|} \log P(\hat{t}_i = t_i | s)$$

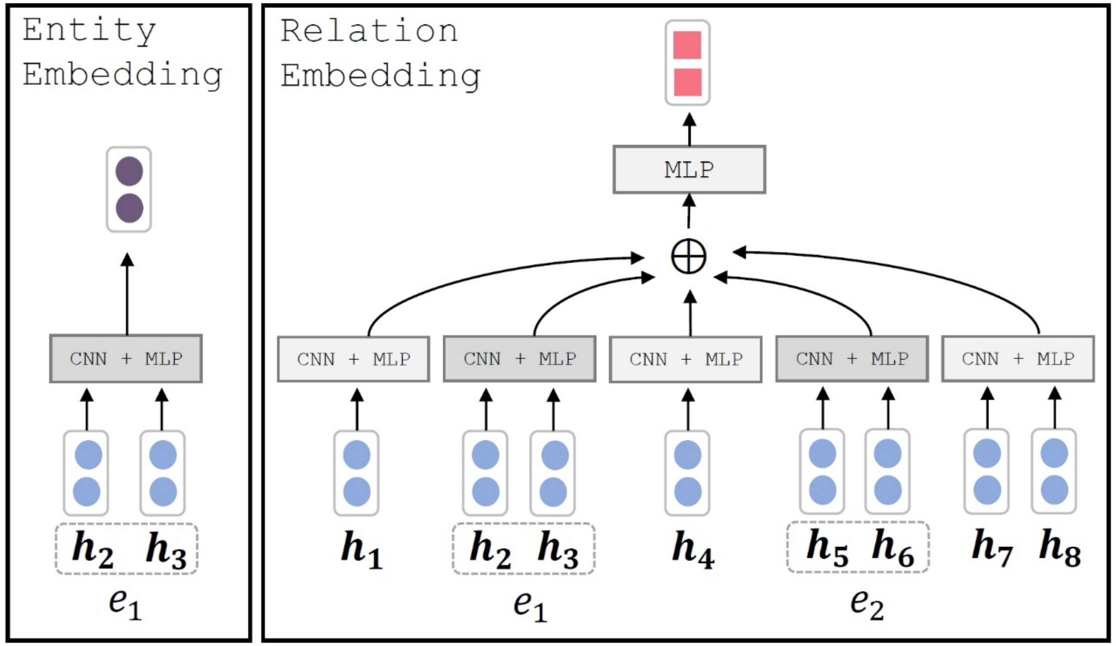


Changzhi Sun, Yeyun Gong et al., Joint Type Inference on Entities and Relation via Graph Convolutional Networks.

Node embedding

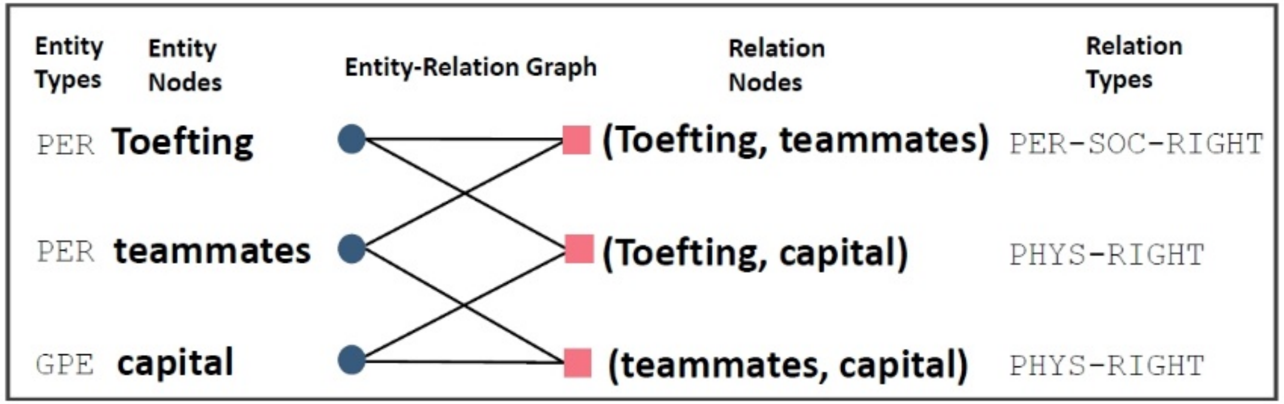
$$P(\hat{b}|r_{ij}, s) = \mathrm{Softmax}(\mathbf{W}_{\mathrm{bin}}\mathbf{H}_{r_{ij}})$$

$$\mathcal{L}_{\mathrm{bin}} = -\sum_{r_{ij}} \frac{\log P(\hat{b} = b|r_{ij}, s)}{\#\ \text{candidate relations}\ r_{ij}}$$



Changzhi Sun, Yeyun Gong et al., Joint Type Inference on Entities and Relation via Graph Convolutional Networks.

Entity-Relation Bipartite Graph



Standard GCN

Changzhi Sun, Yeyun Gong et al., Joint Type Inference on Entities and Relation via Graph Convolutional Networks.

Experiments

Model	**Entity**			**Relation**		
	P	R	F	P	R	F
L&J (2014)	85.2	76.9	80.8	65.4	39.8	49.5
Zhang (2017)	-	-	83.5	-	-	57.5
Sun (2018)	83.9	83.2	83.6	64.9	**55.1**	**59.6**
M&B (2016)	82.9	**83.9**	83.4	57.2	54.0	55.6
K&C (2017)	84.0	81.3	82.6	55.5	51.8	53.6
NN	85.7	82.1	83.9	65.6	50.7	57.2
GCN	**86.1**	82.4	**84.2**	**68.1**	52.3	59.1

Changzhi Sun, Yeyun Gong et al., Joint Type Inference on Entities and Relation via Graph Convolutional Networks.

Experiments

	1-layer	2-layer	3-layer
F1 of Entity Span	90.4	90.5	**90.7**
F1 of Binary Relation	61.5	**62.9**	62.8
F1 of Entity	81.6	82.1	**82.2**
F1 of Relation	**53.8**	53.5	53.6

Changzhi Sun, Yeyun Gong et al., Joint Type Inference on Entities and Relation via Graph Convolutional Networks.

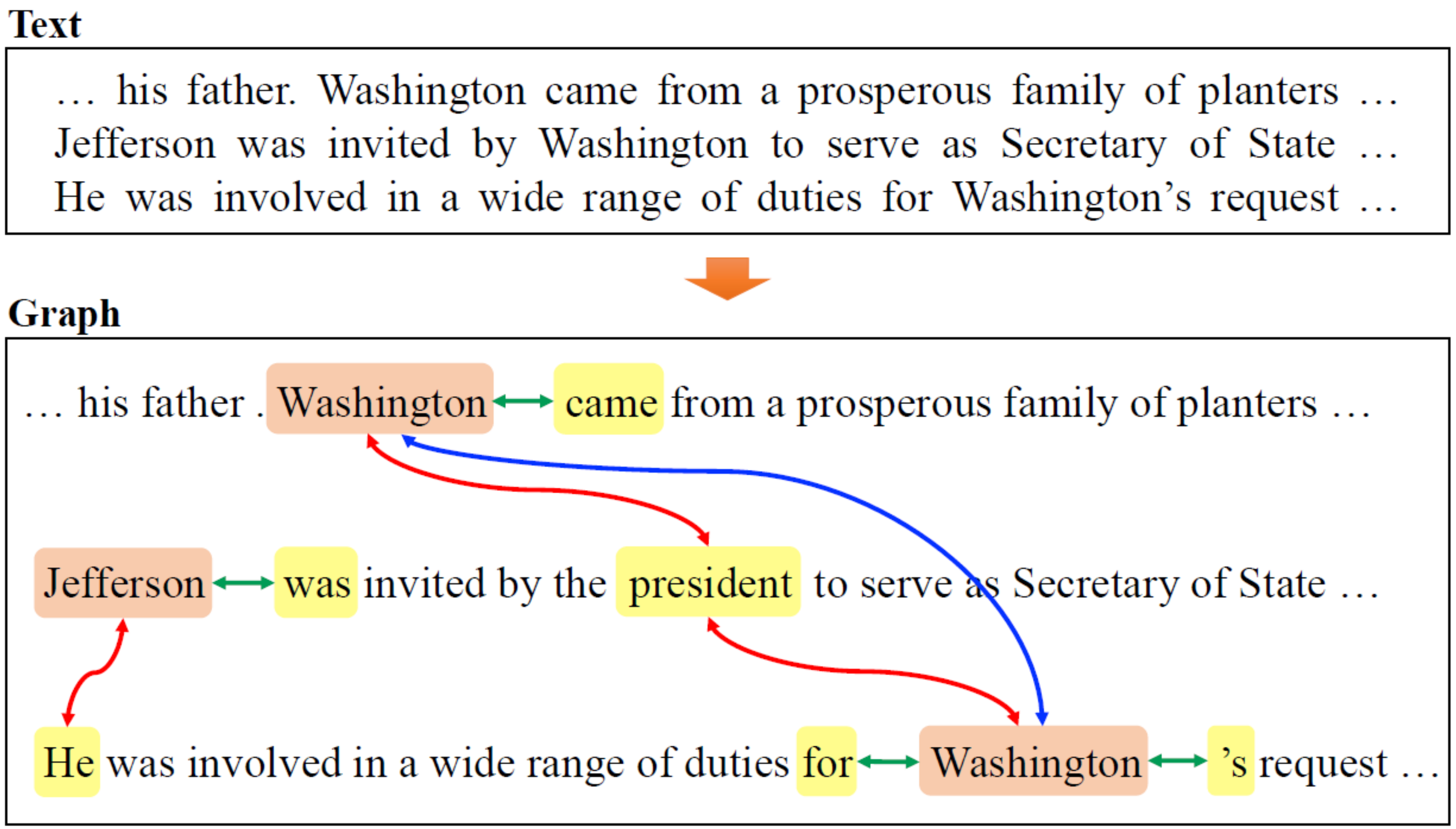
Graph Information Extraction

Shortage of Tree LSTM, Graph LSTM

- They can only model word level graph.
- They all use dependency trees.

Yujie Qian, Enrico Santus et al., GraphIE: A Graph-Based Framework for Information Extraction, NAACL 2019.

Graph Information Extraction



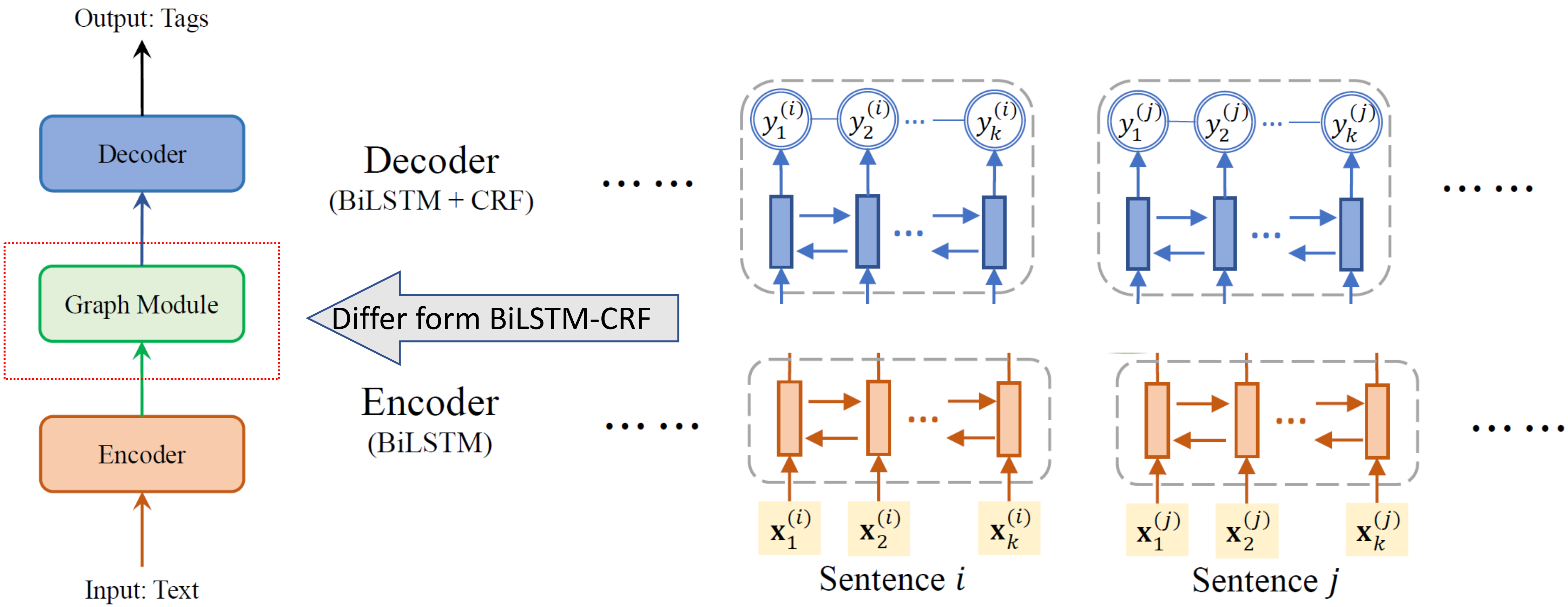
Green: local contextual info.

Red: co-referent edges

Blue: identical-mention edges

Yujie Qian, Enrico Santus et al., GraphIE: A Graph-Based Framework for Information Extraction, NAACL 2019.

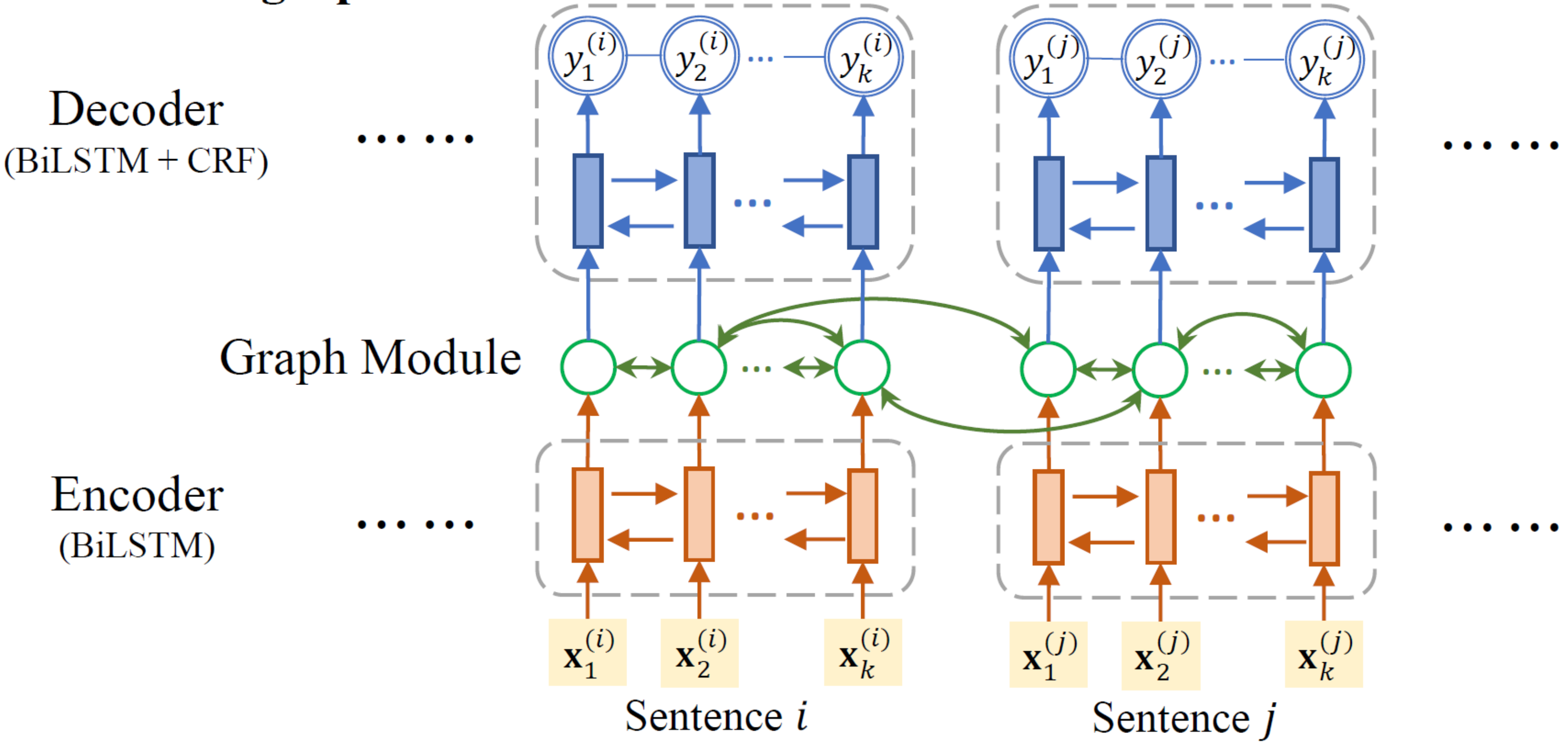
Graph Information Extraction



Yujie Qian, Enrico Santus et al., GraphIE: A Graph-Based Framework for Information Extraction, NAACL 2019.

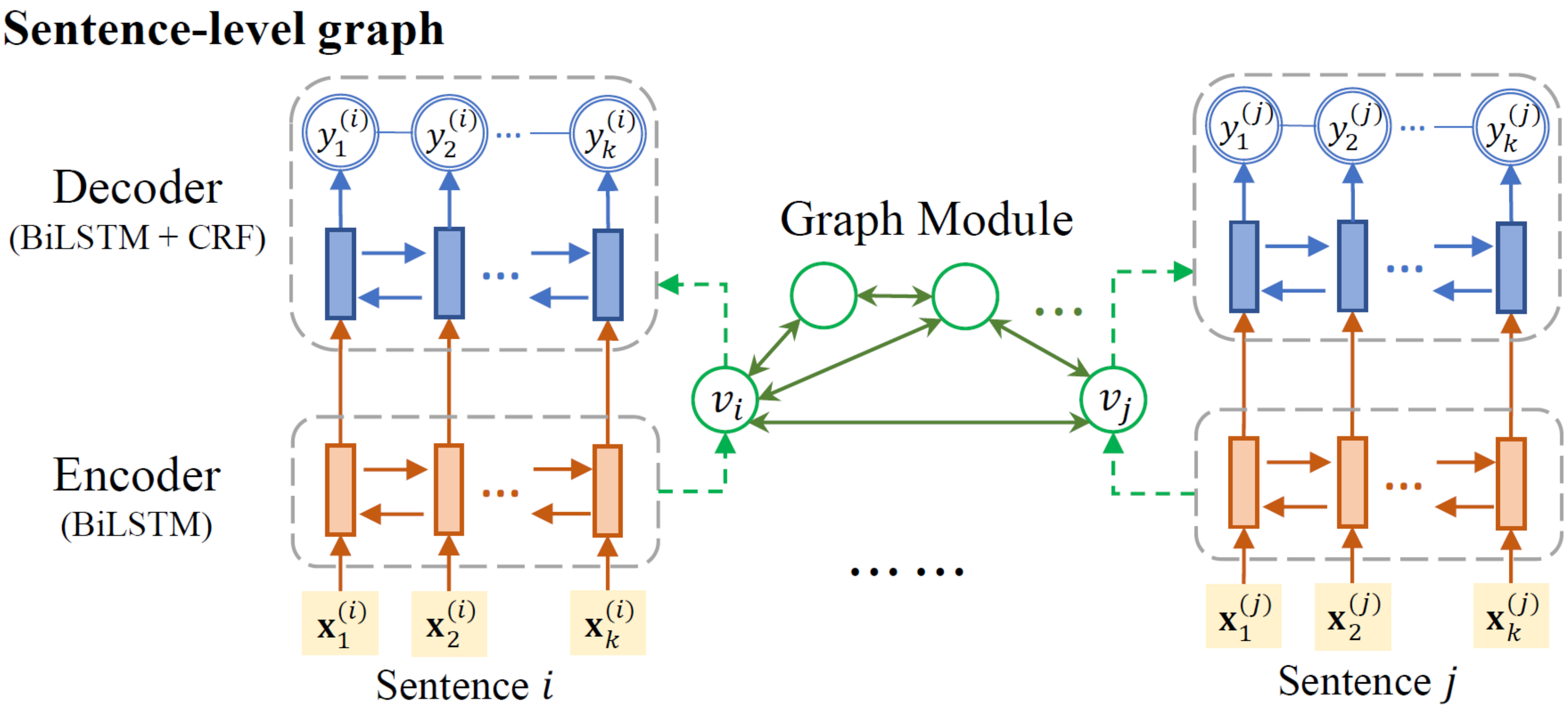
Graph Information Extraction

Word-level graph



Yujie Qian, Enrico Santus et al., GraphIE: A Graph-Based Framework for Information Extraction, NAACL 2019.

Graph Information Extraction



Yujie Qian, Enrico Santus et al., GraphIE: A Graph-Based Framework for Information Extraction, NAACL 2019.

Experiments

Evaluation Task	Graph Type	Node	Edge
Textual IE	word-level	word	1. non-local consistency (identical mentions) 2. local sentential forward and backward
Social Media IE	sentence-level	user's tweets	*followed-by*

Yujie Qian, Enrico Santus et al., GraphIE: A Graph-Based Framework for Information Extraction, NAACL 2019.

Experiments 1

DATASET		Train	Dev	Test
CoNLL03	#doc	946	216	231
	#sent	14,987	3,466	3,684
CHEMDNER	#doc	3,500	3,500	3,000
	#sent	30,739	30,796	26,399

Yujie Qian, Enrico Santus et al., GraphIE: A Graph-Based Framework for Information Extraction, NAACL 2019.

Experiments 1

DATASET	Model	F1
CONLL03	Lample et al. (2016)	90.94
	Ma and Hovy (2016)	91.21
	Ye and Ling (2018)	91.38
	SeqIE	91.16
	GraphIE	**91.74***
CHEMDNER	Krallinger et al. (2015)	87.39
	SeqIE	88.28
	GraphIE	**89.71***

Yujie Qian, Enrico Santus et al., GraphIE: A Graph-Based Framework for Information Extraction, NAACL 2019.

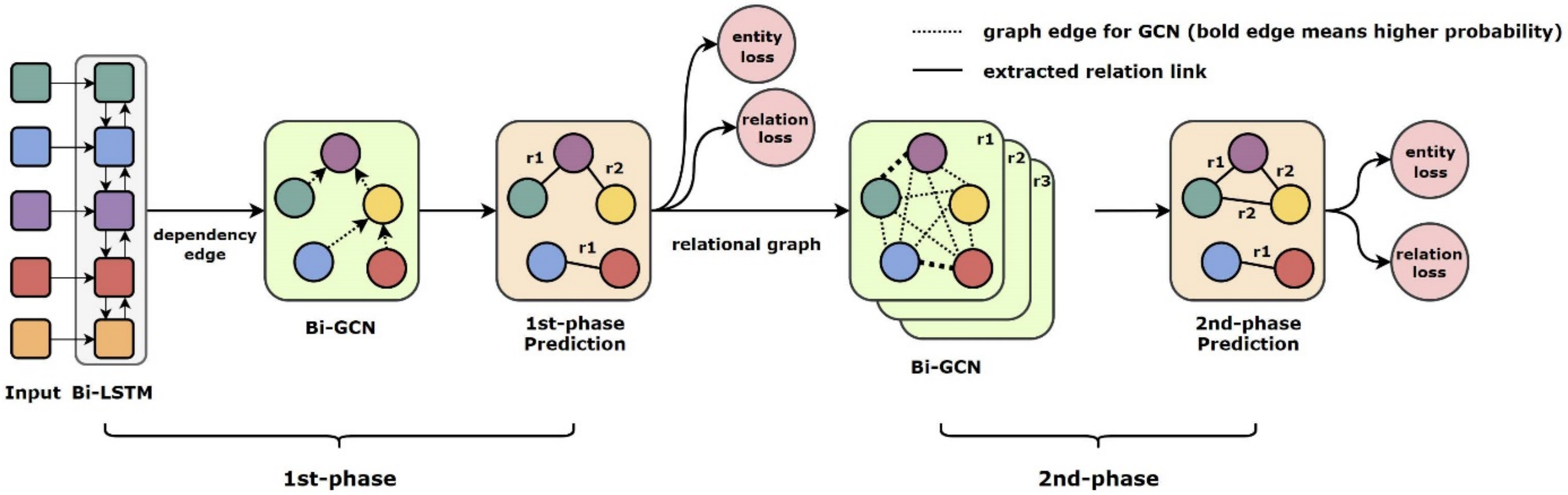
Experiments 2

Dataset	Dictionary			SeqIE			GraphIE		
	P	R	F1	P	R	F1	P	R	F1
Education	78.7	93.5	85.4	85.2	93.6	89.2	92.9	92.8	**92.9***
Job	55.7	70.2	62.1	66.2	66.7	66.2	67.1	66.1	**66.5**

3.7 up

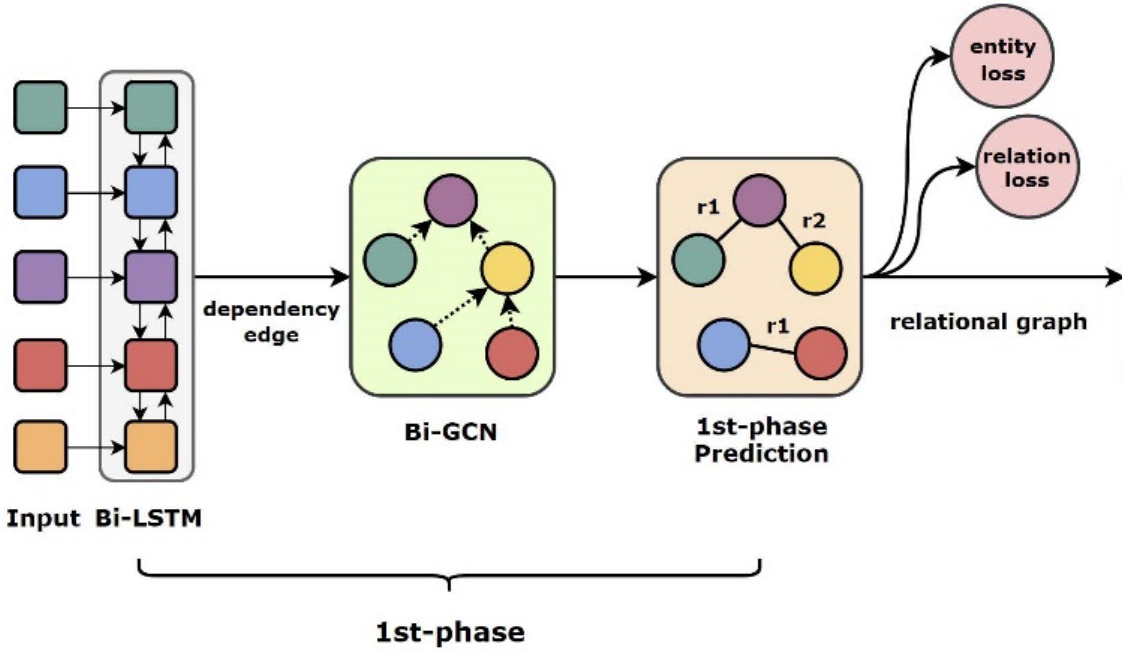
Yujie Qian, Enrico Santus et al., GraphIE: A Graph-Based Framework for Information Extraction, NAACL 2019.

GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction



Tsu-Jui Fu et al., GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction, ACL2019.

GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction



$$h_u^0 = Word(u) \oplus POS(u)$$

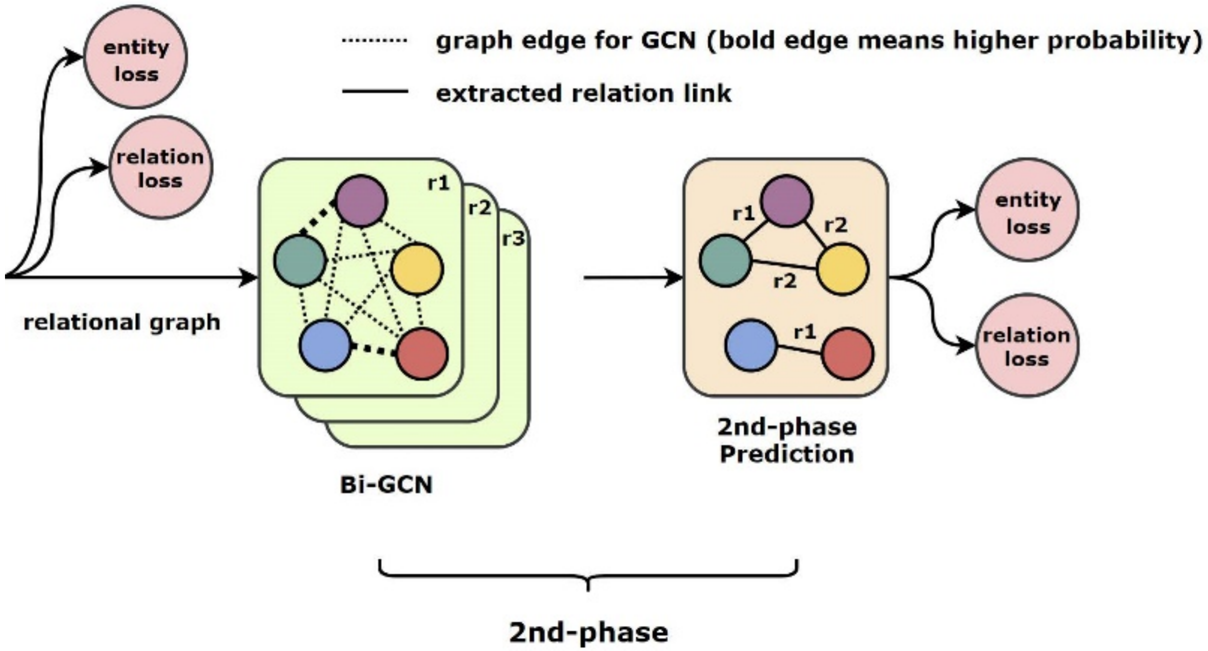
$$\overrightarrow{h}_u^{l+1} = ReLU\left(\sum_{v \in \overrightarrow{N}(u)} \left(\overrightarrow{W}^l h_v^l + \overrightarrow{b}^l\right)\right)$$

$$\overleftarrow{h}_u^{l+1} = ReLU\left(\sum_{v \in \overleftarrow{N}(u)} \left(\overleftarrow{W}^l h_v^l + \overleftarrow{b}^l\right)\right)$$

$$h_u^{l+1} = \overrightarrow{h}_u^{l+1} \oplus \overleftarrow{h}_u^{l+1},$$

Tsu-Jui Fu et al., GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction, ACL2019.

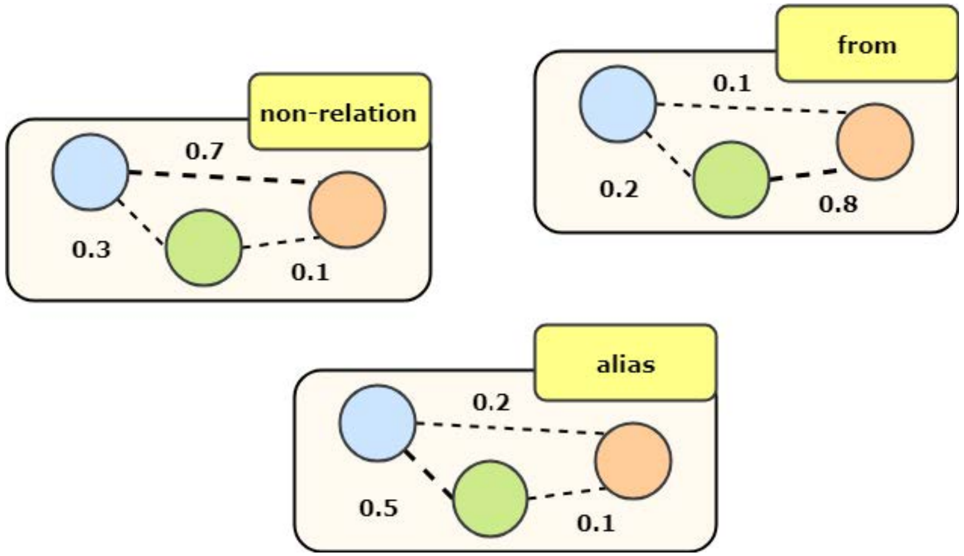
GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction



$$h_u^{l+1} = ReLU\left(\sum_{v \in V} \sum_{r \in R} P_r\left(u, v\right) \times \left(W_r^l h_v^l + b_r^l\right)\right)$$

Tsu-Jui Fu et al., GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction, ACL2019.

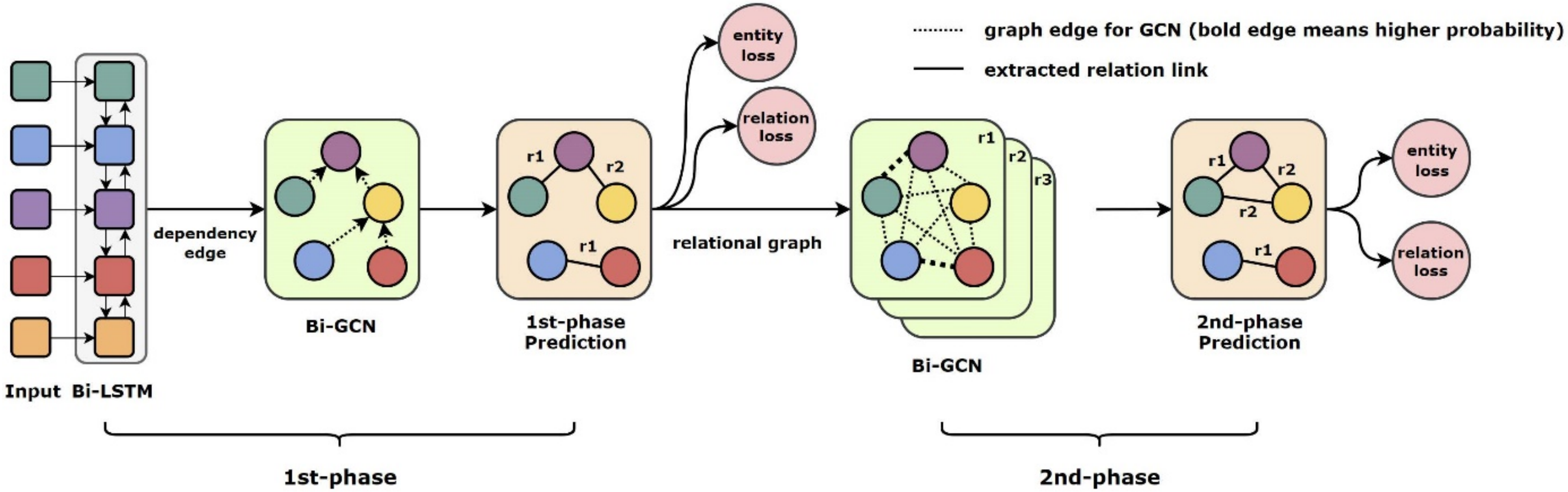
GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction



$$h_u^{l+1} = ReLU\left(\sum_{v\in V}\sum_{r\in R} P_r(u,v) \times \left(W_r^l h_v^l + b_r^l\right)\right)$$

Tsu-Jui Fu et al., GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction, ACL2019.

GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction



$$loss_{all} = (eloss_{1p} + rloss_{1p}) + \alpha\,(eloss_{2p} + rloss_{2p})$$

Tsu-Jui Fu et al., GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction, ACL2019.

Experiments

Method	NYT			WebNLG		
	Precision	Recall	F1	Precision	Recall	F1
NovelTagging	62.4%	31.7%	42.0%	**52.5%**	19.3%	28.3%
OneDecoder	59.4%	53.1%	56.0%	32.2%	28.9%	30.5%
MultiDecoder	61.0%	56.6%	58.7%	37.7%	36.4%	37.1%
GraphRel_{1p}	62.9%	57.3%	60.0%	42.3%	39.2%	40.7%
GraphRel_{2p}	**63.9%**	**60.0%**	**61.9%**	44.7%	**41.1%**	**42.9%**

Tsu-Jui Fu et al., GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction, ACL2019.

Experiments

Phase	#GCN layer	NYT	WebNLG
1st-phase	2	60.0%	**40.7%**
	3	60.0%	40.5%
2nd-phase†	1	**61.9%**	**42.9%**
	2	61.6%	42.4%
3rd-phase‡	1	61.8%	42.7%

Tsu-Jui Fu et al., GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction, ACL2019.

GCN for Multimodal Information Extraction

Value added tax invoice - extract buyer seller date amount.

Visually rich documents (VRDs)

Xiaojing Liu, Feiyu Gao et al., Graph Convolution for Multimodal Information Extraction from Visually Rich Documents, NAACL 2019.

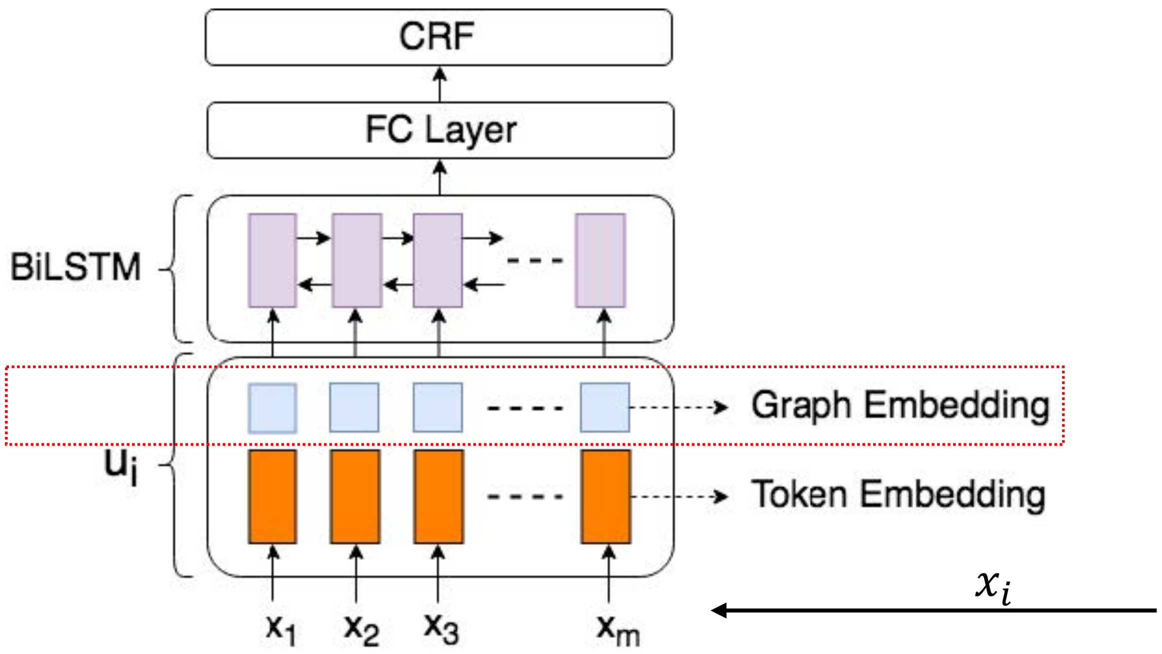
GCN for Multimodal Information Extraction

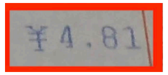
Feature from the image.
Text segments(4 coordinates), colors, fonts…

Xiaojing Liu, Feiyu Gao et al., Graph Convolution for Multimodal Information Extraction from Visually Rich Documents, NAACL 2019.

Baseline

BiLSTM-CRF baseline



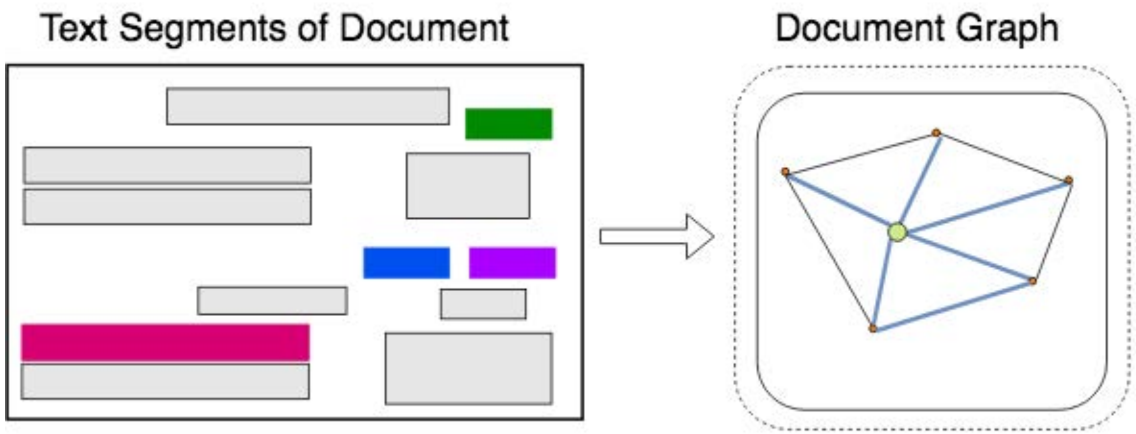


All above spectrum based approaches such as GCN do not work well on dynamic graph structures.

Differ from baseline.

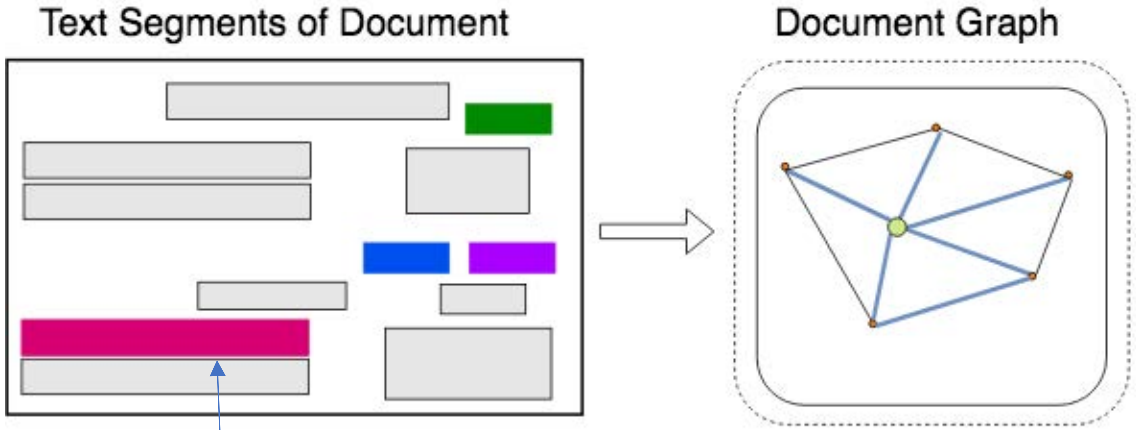
Xiaojing Liu, Feiyu Gao et al., Graph Convolution for Multimodal Information Extraction from Visually Rich Documents, NAACL 2019.

Model



Xiaojing Liu, Feiyu Gao et al., Graph Convolution for Multimodal Information Extraction from Visually Rich Documents, NAACL 2019.

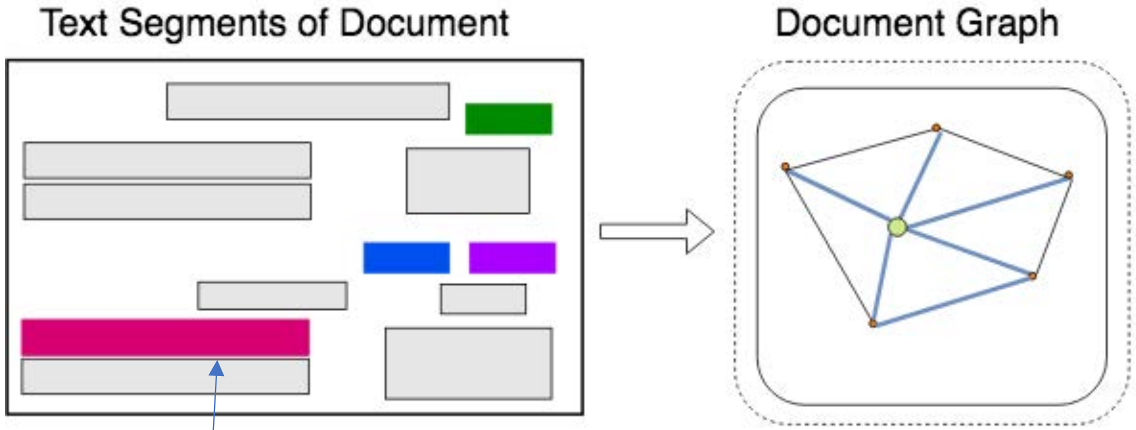
Model



Generated by OCR

Xiaojing Liu, Feiyu Gao et al., Graph Convolution for Multimodal Information Extraction from Visually Rich Documents, NAACL 2019.

Model



Generated by OCR

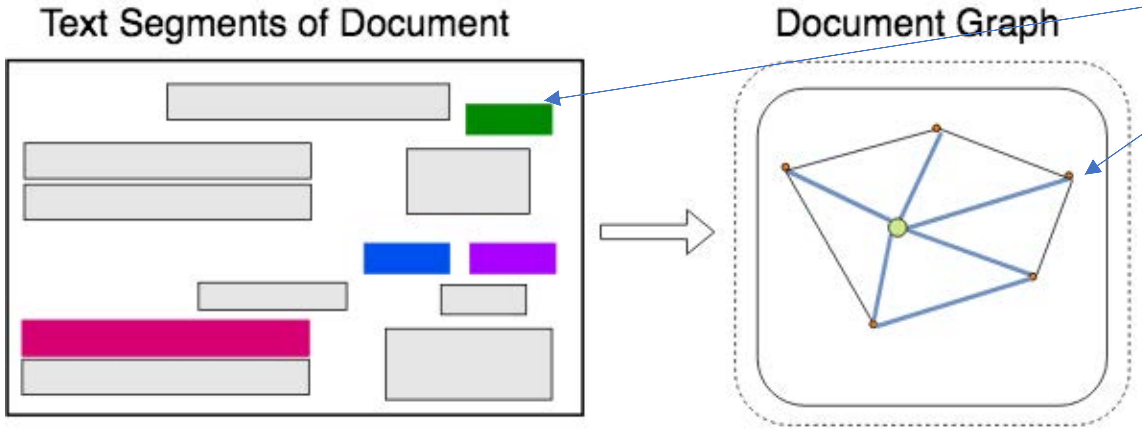
A graph model combine visual and textual context

Visual context: layout and relative positions

Textual context: the aggregate of text info.

Xiaojing Liu, Feiyu Gao et al., Graph Convolution for Multimodal Information Extraction from Visually Rich Documents, NAACL 2019.

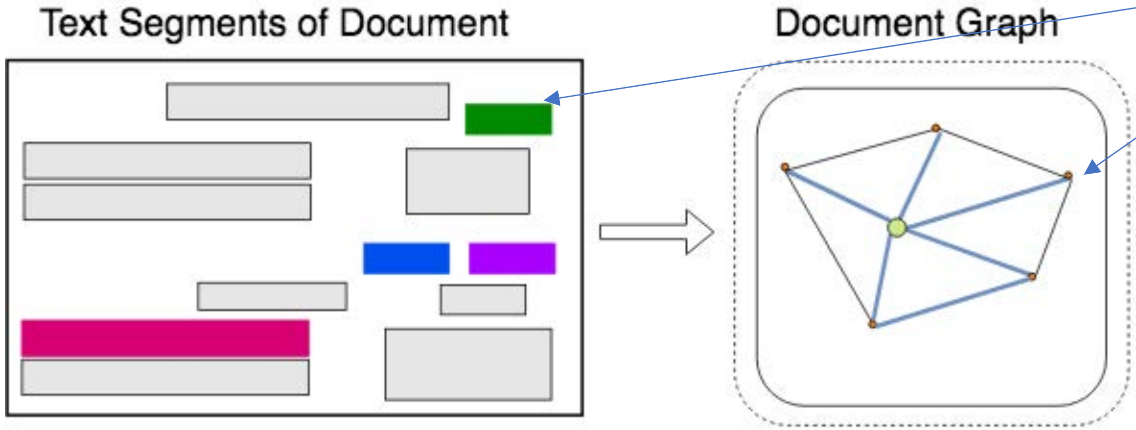
Model



For each node, calculate embedding using BiLSTM

Xiaojing Liu, Feiyu Gao et al., Graph Convolution for Multimodal Information Extraction from Visually Rich Documents, NAACL 2019.

Model



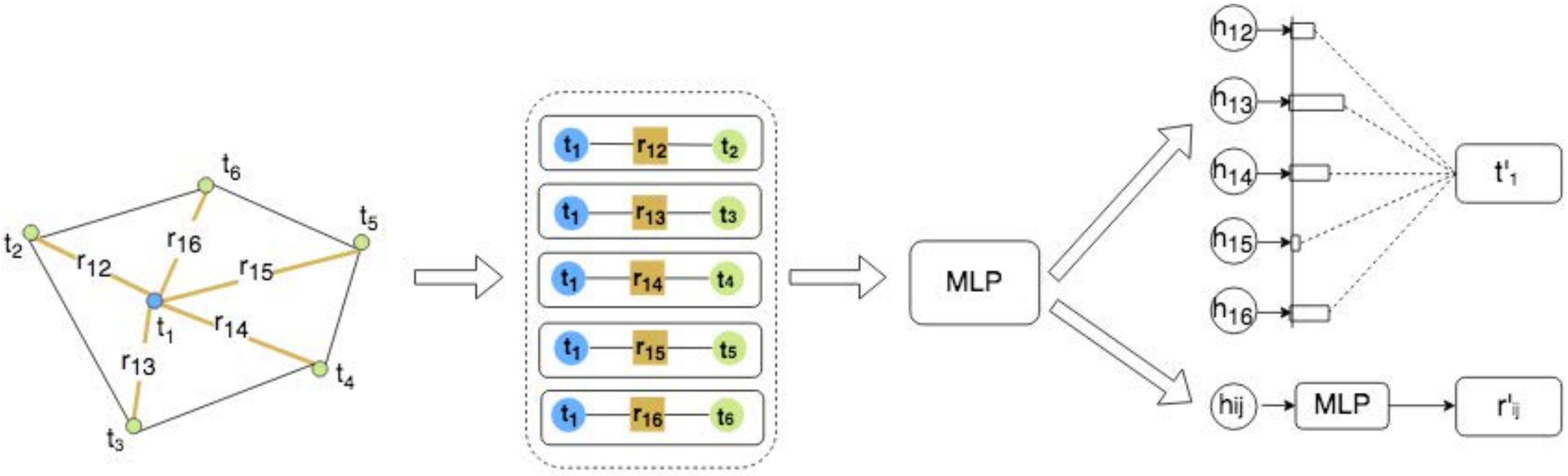
For each node, calculate embedding using BiLSTM

Between two nodes, edge embedding as follow:

$$\mathbf{r}_{ij} = [x_{ij}, y_{ij}, \frac{w_i}{h_i}, \frac{h_j}{h_i}, \frac{w_j}{h_i}]$$

Xiaojing Liu, Feiyu Gao et al., Graph Convolution for Multimodal Information Extraction from Visually Rich Documents, NAACL 2019.

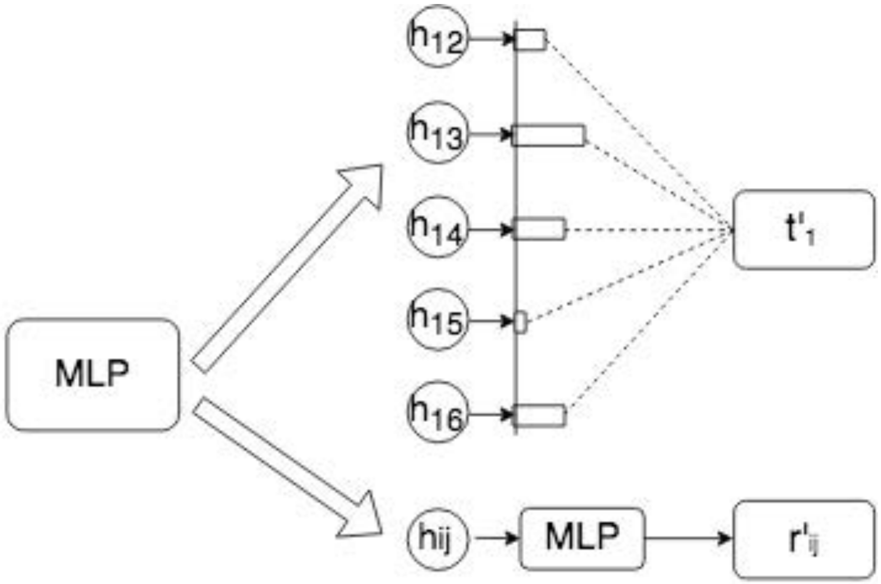
Model



$$\mathbf{h}_{ij} = g(\mathbf{t}_i, \mathbf{r}_{ij}, \mathbf{t}_j) = \mathrm{MLP}([\mathbf{t}_i \| \mathbf{r}_{ij} \| \mathbf{t}_j])$$

Xiaojing Liu, Feiyu Gao et al., Graph Convolution for Multimodal Information Extraction from Visually Rich Documents, NAACL 2019.

Using GAT as Graph Network module



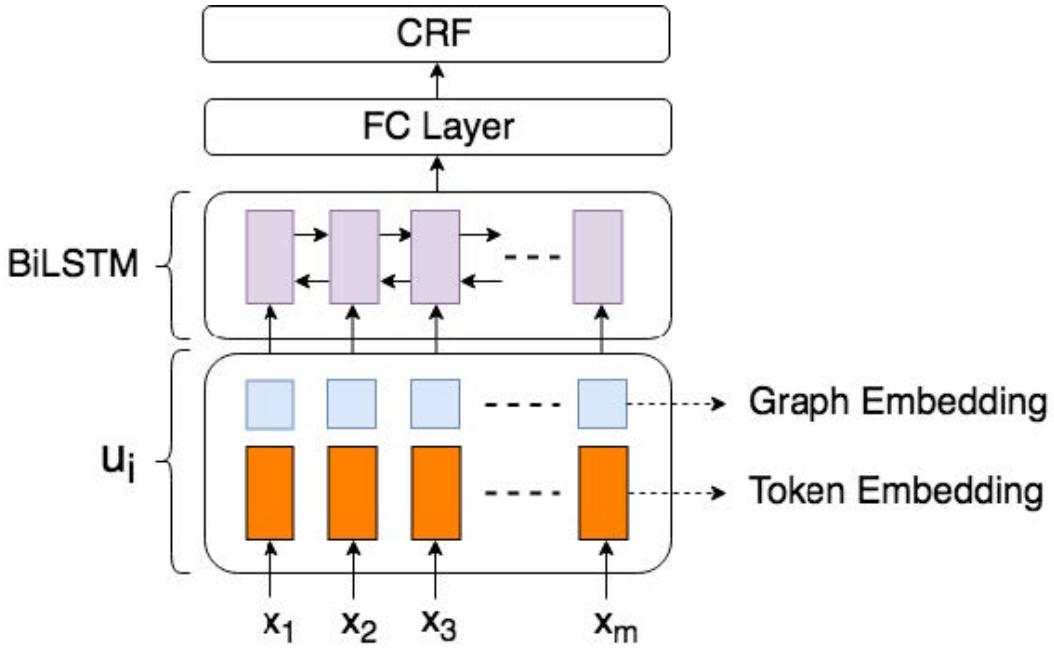
$$\mathbf{t}'_i = \sigma\Big(\sum_{j\in\{1,\cdots,n\}} \alpha_{ij}\mathbf{h}_{ij}\Big)$$

$$\alpha_{ij} = \frac{\exp(\mathrm{LeakyRelu}(\mathbf{w}_a^T\mathbf{h}_{ij}))}{\sum_{j\in\{1,\cdots,n\}}\exp(\mathrm{LeakyRelu}(\mathbf{w}_a^T\mathbf{h}_{ij}))}$$

$$\mathbf{r}'_{ij} = \mathrm{MLP}(\mathbf{h}_{ij})$$

Xiaojing Liu, Feiyu Gao et al., Graph Convolution for Multimodal Information Extraction from Visually Rich Documents, NAACL 2019.

Model



Final represntation of each segment is the concatenation of its embedding and graph information.

$$\mathbf{u}_i = e(x_i) \| \mathbf{t}'_i$$

Xiaojing Liu, Feiyu Gao et al., Graph Convolution for Multimodal Information Extraction from Visually Rich Documents, NAACL 2019.

Experiments

Model	VATI	IPR
Baseline I	0.745	0.747
Baseline II	0.854	0.820
BiLSTM-CRF + GCN	0.873	0.836

Xiaojing Liu, Feiyu Gao et al., Graph Convolution for Multimodal Information Extraction from Visually Rich Documents, NAACL 2019.

Result on different entities

Entities	Baseline I	Baseline II	Our model
Invoice #	0.952	0.961	0.975
Date	0.962	0.963	0.963
Price	0.527	0.910	0.943
Tax	0.584	0.902	0.924
Buyer	0.402	0.797	0.833
Seller	0.681	0.731	0.782

Xiaojing Liu, Feiyu Gao et al., Graph Convolution for Multimodal Information Extraction from Visually Rich Documents, NAACL 2019.

Experiments on GCN layers

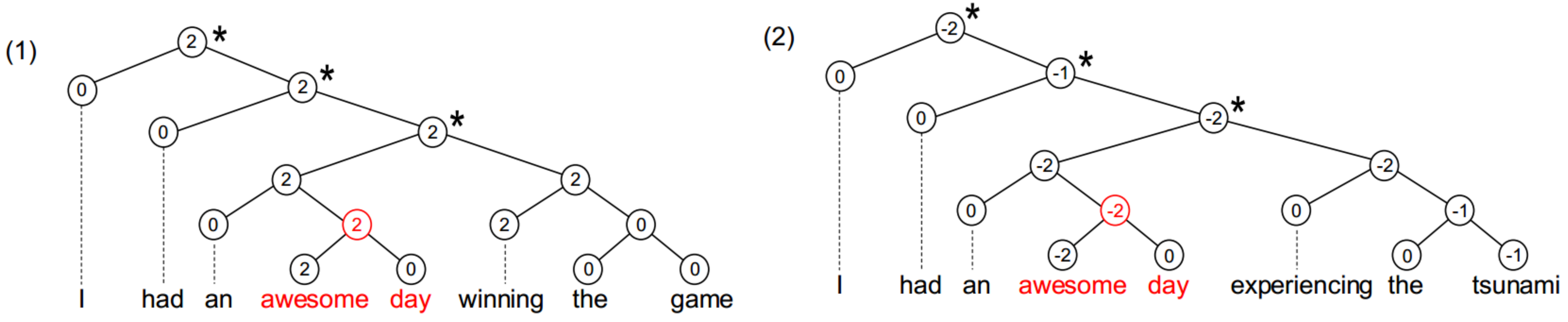
Entities	1 layer	2 layer	3 layer
Invoice #	0.959	0.975	0.964
Date	0.960	0.963	0.960
Price	0.931	0.943	0.931
Tax	0.915	0.924	0.917
Buyer	0.829	0.833	0.827
Seller	0.772	0.782	0.775

Xiaojing Liu, Feiyu Gao et al., Graph Convolution for Multimodal Information Extraction from Visually Rich Documents, NAACL 2019.

Solving problems using graph encoding

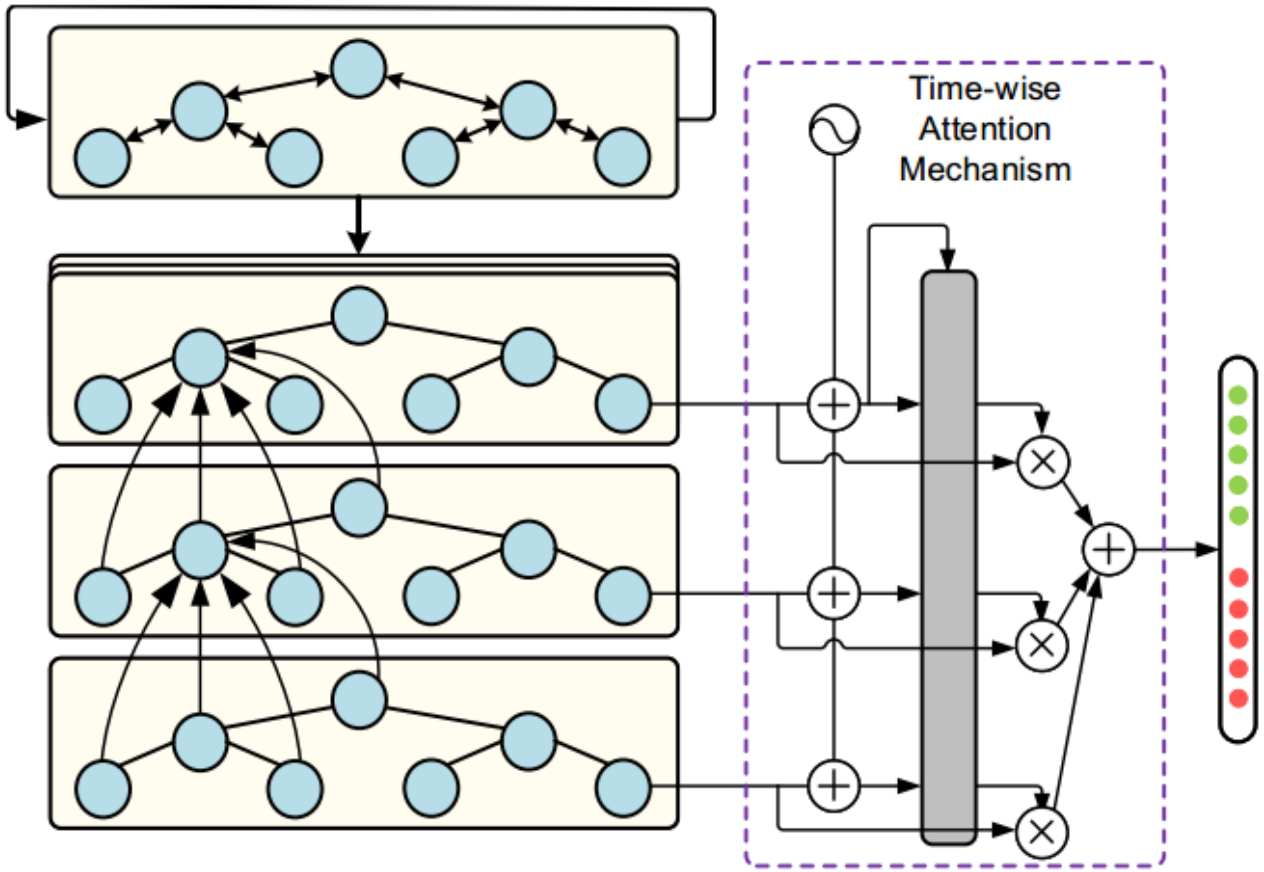
- Neural Machine Translation
- Text Generation
- Question Answering
- Information Extraction
- Sentiment
- Social Classification
- Syntactic Parsing
- Semantic Role Labeling
- Word Embedding
- Sentence Representation

Tree-based Sentiment Analysis



Yuan Zhang, Yue Zhang, Tree Communication Models for Sentiment Analysis, ACL 2019.

Communication over constituent tree with GRN



Yuan Zhang, Yue Zhang, Tree Communication Models for Sentiment Analysis, ACL 2019.

Experiments

Dataset: SST-5 & SST-2

Corpus	SST-5	SST-2
Classes	5	2
Sentences	11,855	9,613
Phrases	442,629	137,988
Tokens	227,242	185,621

Yuan Zhang, Yue Zhang, Tree Communication Models for Sentiment Analysis, ACL 2019.

Experiments

Main baselines:

- Stannard tree-LSTM (Zhu et al., 2015)
- Bidirectional tree-LSTM(Teng and Zhang, 2017)

Metrics:

- Root level accuracy
- Phrase level accuracy

Yuan Zhang, Yue Zhang, Tree Communication Models for Sentiment Analysis, ACL 2019.

DEV experiment

Block	Model	SST-5	SST-2
A	3 Step RTCM	83.1	92.3
	6 Step RTCM	83.2	92.7
	9 Step RTCM	83.4	92.9
	18 Step RTCM	83.2	92.8
B	Tree-LSTM	82.9	92.4
	CTCM	83.3	92.8
	RTCM	83.4	92.9
	RTCM+attention	83.5	93.3

Phrase level performances on the dev set.

R – GRN
C – GCN

Yuan Zhang, Yue Zhang, Tree Communication Models for Sentiment Analysis, ACL 2019.

Main results

Model	SST-5		SST-2	
	R	P	R	P
RNTN (S13)	45.7	80.7	85.4	87.6
BiLSTM (L15)	49.8	83.3	86.7	-
ConTree (LZ15)	49.9	-	88.0	-
ConTree (Z15)	50.1	-	-	-
ConTree (L15)	50.4	83.4	86.7	-
ConTree (T15)	51.0	-	88.0	-
Disan (S18)	51.7	-	-	-
RL LD/HS-LSTM (Z18)	50.0	-	87.8	-
NTI-SLSTM (MY17)	53.1	-	89.3	-
ConTree(Fold) (L17)	52.3	-	89.4	-
BiConTree (TZ17)	53.5	83.5	**90.3**	92.8
RTC + attention	**54.3**	**83.6**	**90.3**	**93.4**

R-Root, P-Phrase.

S13 - Socher et al. (2013);
L15 - Li et al. (2015);
LZ15 - Le and Zuidema (2015);
Z15 - Zhu et al. (2015);
T15 - Tai et al. (2015);
S18 - Shen et al. (2018);
Z18 - Zhang et al. (2018a);
MY17 - Munkhdalai and Yu (2017);
L17 - Looks et al. (2017);
TZ17 Teng and Zhang (2017)

Yuan Zhang, Yue Zhang, Tree Communication Models for Sentiment Analysis, ACL 2019.

Solving problems using graph encoding

- Neural Machine Translation
- Text Generation
- Question Answering
- Information Extraction
- Sentiment
- Social Classification
- Syntactic Parsing
- Semantic Role Labeling
- Word Embedding
- Sentence Representation

Encoding Social Information with GCN

thehill.com *(Center)*

Sen. Mark Warner (D-Va.) on Sunday blasted President Trump for his "inept negotiation" to bring an end to the ongoing partial government shutdown. Warner, the ranking member of the Senate Intelligence Committee, lamented the effect the shutdown has had on hundreds of thousands of federal workers who have been furloughed or forced to work without pay.

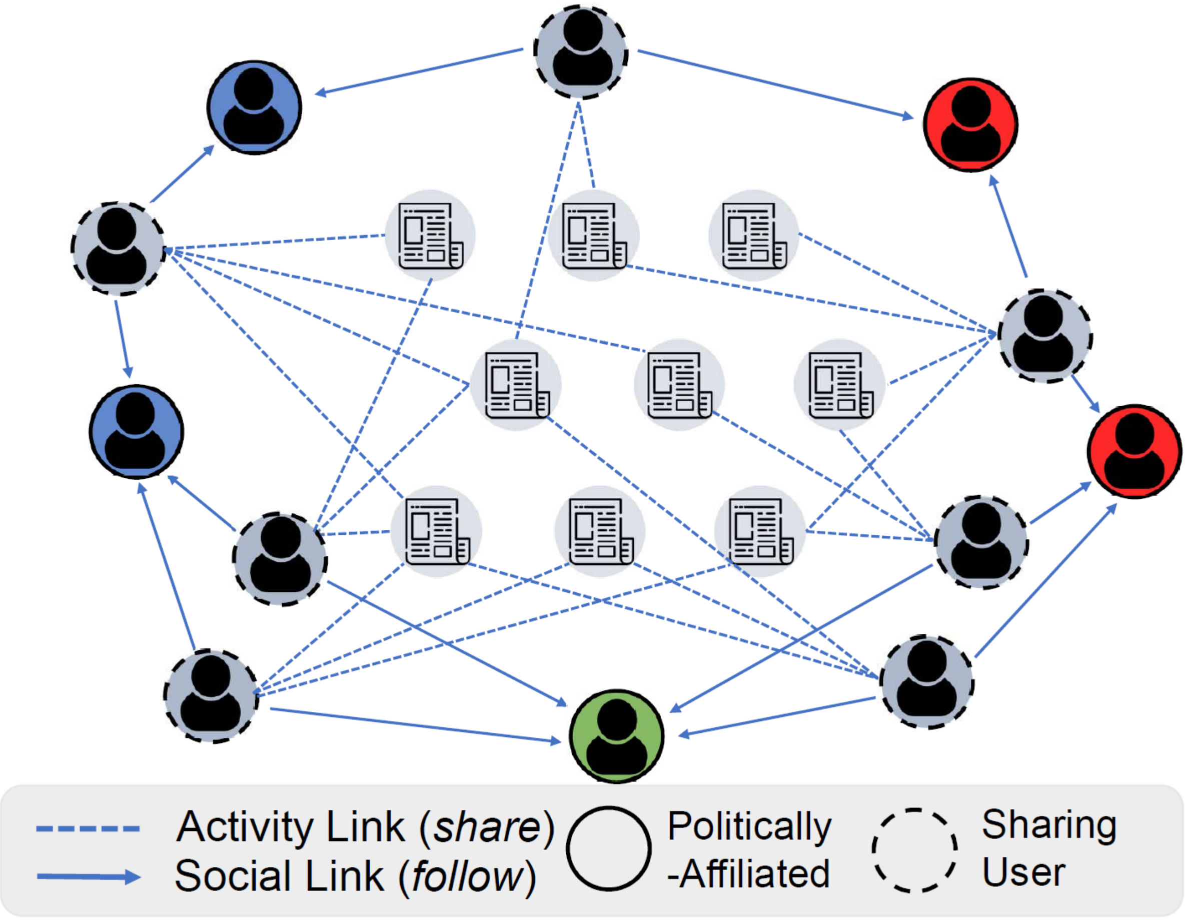
infowars.com *(Right)*

Senator Mark Warner (D-Va.) is being called out on social media for his statement on the partial government shutdown. Warner blamed the "suffering" of federal workers and contractors on President Trump in a Sunday tweet framing Trump as an "inept negotiator". Twitter users pointed out that Democrats are attending a Puerto Rican retreat with over 100 lobbyists and corporate executives.

Chang Li and Dan Goldwasser, Encoding Social Information with Graph Convolutional Networks for Political Perspective Detection in News Media, ACL 2019.

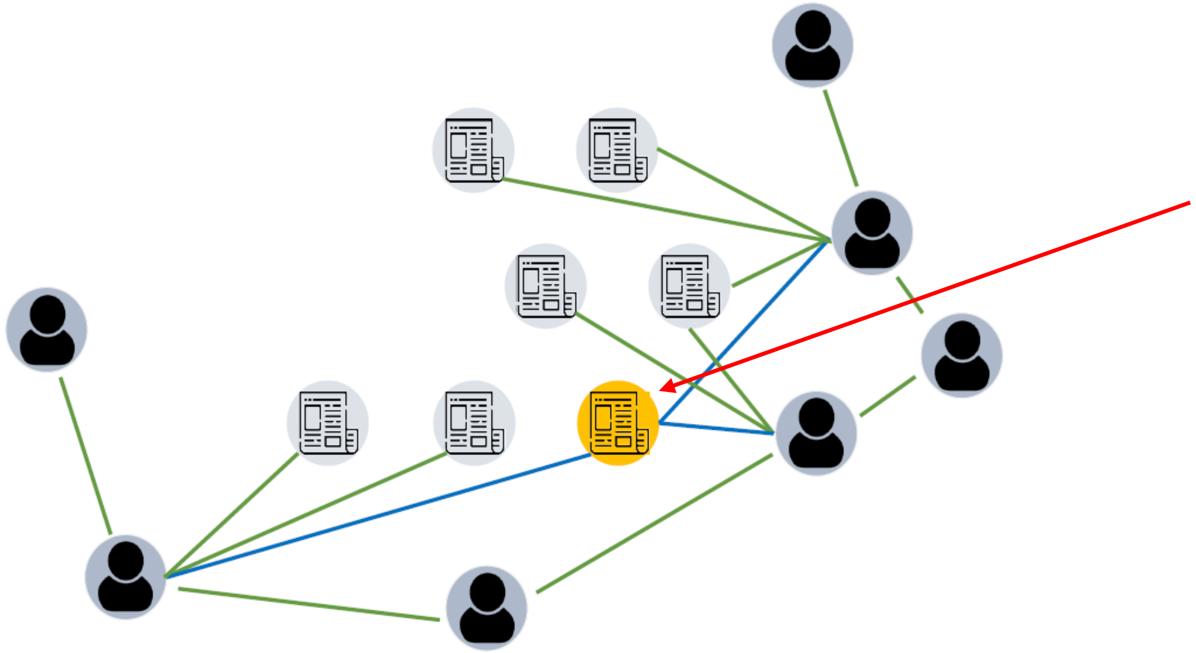
Encoding Social Information with GCN

This paper cast the problem as a 3-class prediction problem, capturing left-leaning bias, right-leaning bias or no bias (center).



Chang Li and Dan Goldwasser, Encoding Social Information with Graph Convolutional Networks for Political Perspective Detection in News Media, ACL 2019.

Encoding Social Information



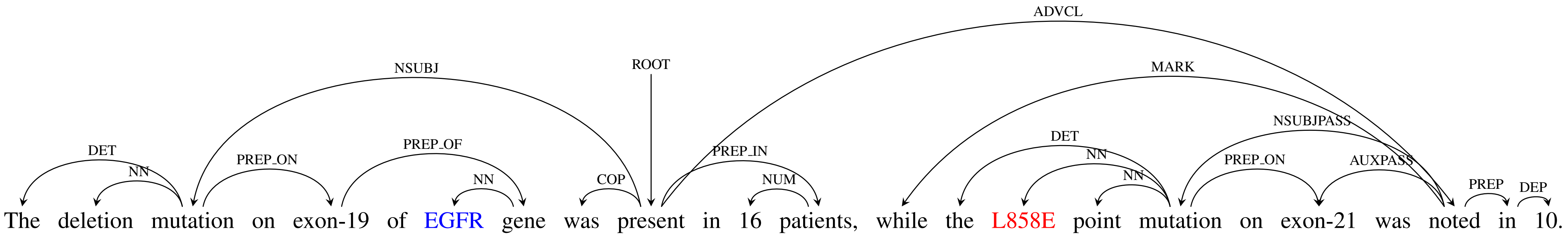
News texts that incorporate social network information using GCN are used for classification.

Chang Li and Dan Goldwasser, Encoding Social Information with Graph Convolutional Networks for Political Perspective Detection in News Media, ACL 2019.

Solving problems using graph encoding

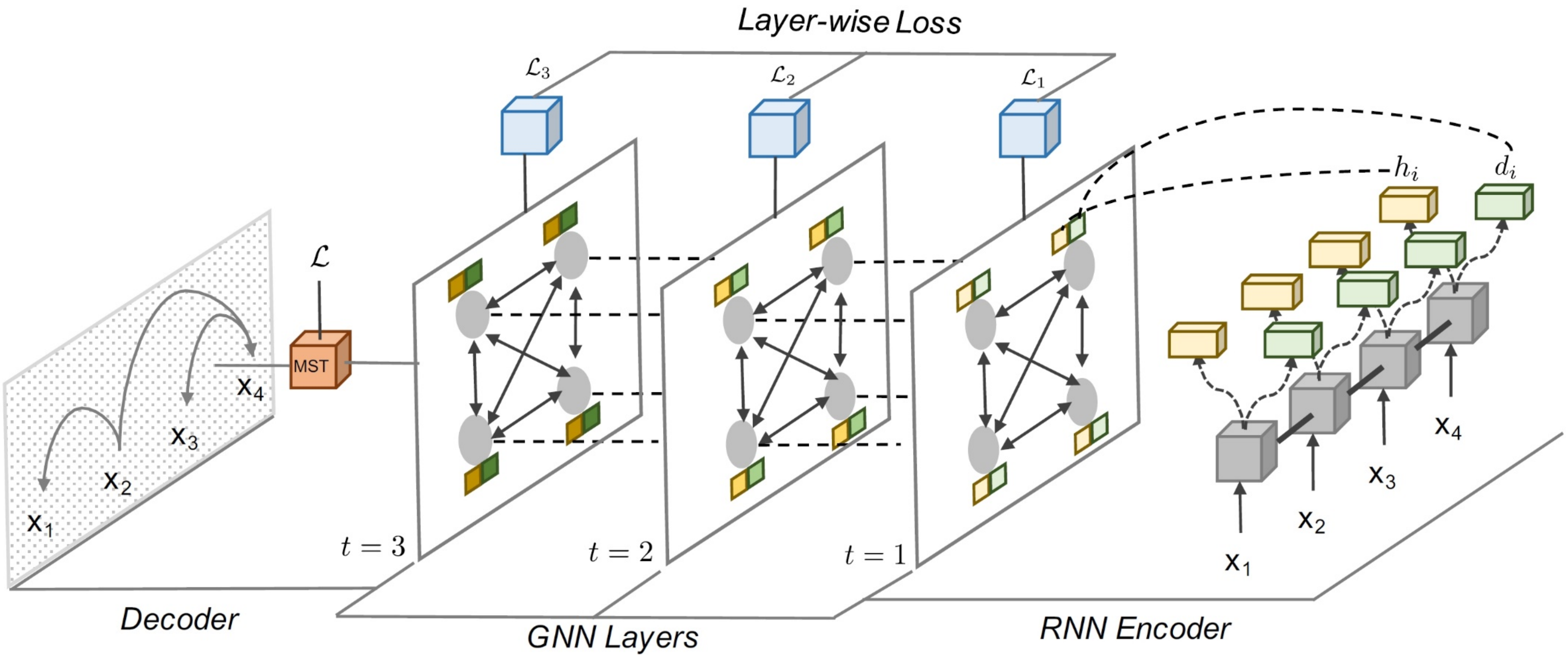
- Neural Machine Translation
- Text Generation
- Question Answering
- Information Extraction
- Sentiment
- Social Classification
- Syntactic Parsing
- Semantic Role Labeling
- Word Embedding
- Sentence Representation

Graph-based Dependency Parsing



Graph-based Dependency Parsing with Graph Neural Networks, Tao ji, Yunabin Wu, and Man Lan ACL 2019.

Parsing with GCNs

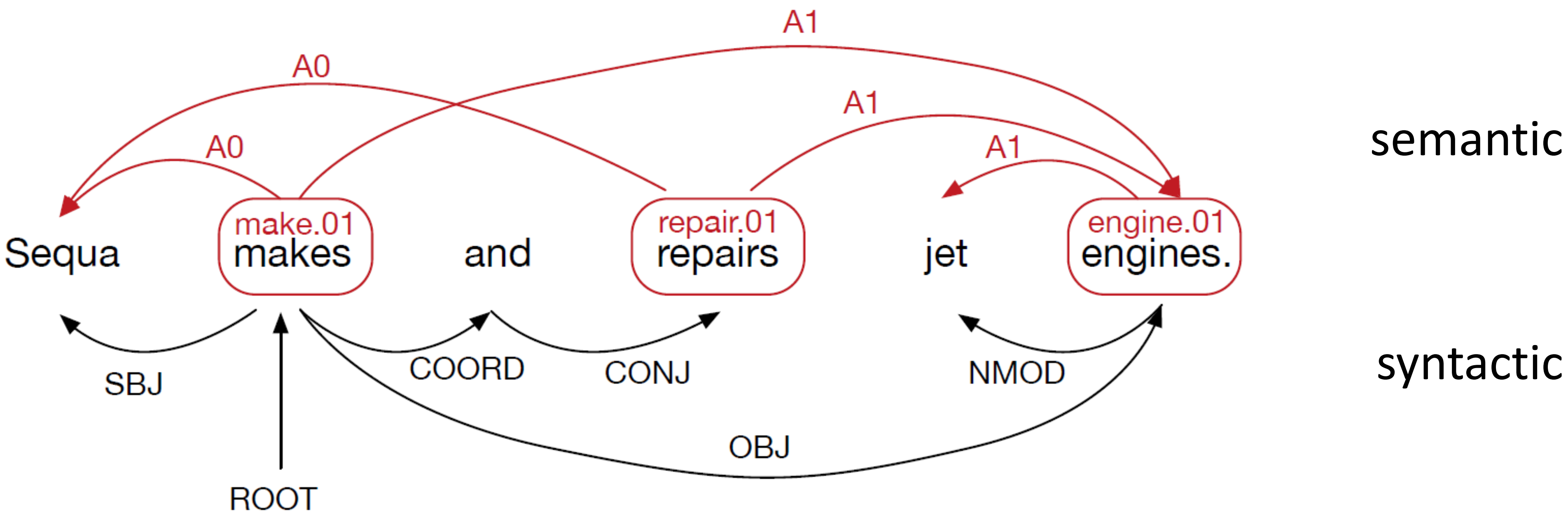


Graph-based Dependency Parsing with Graph Neural Networks, Tao ji, Yunabin Wu, and Man Lan ACL 2019.

Solving problems using graph encoding

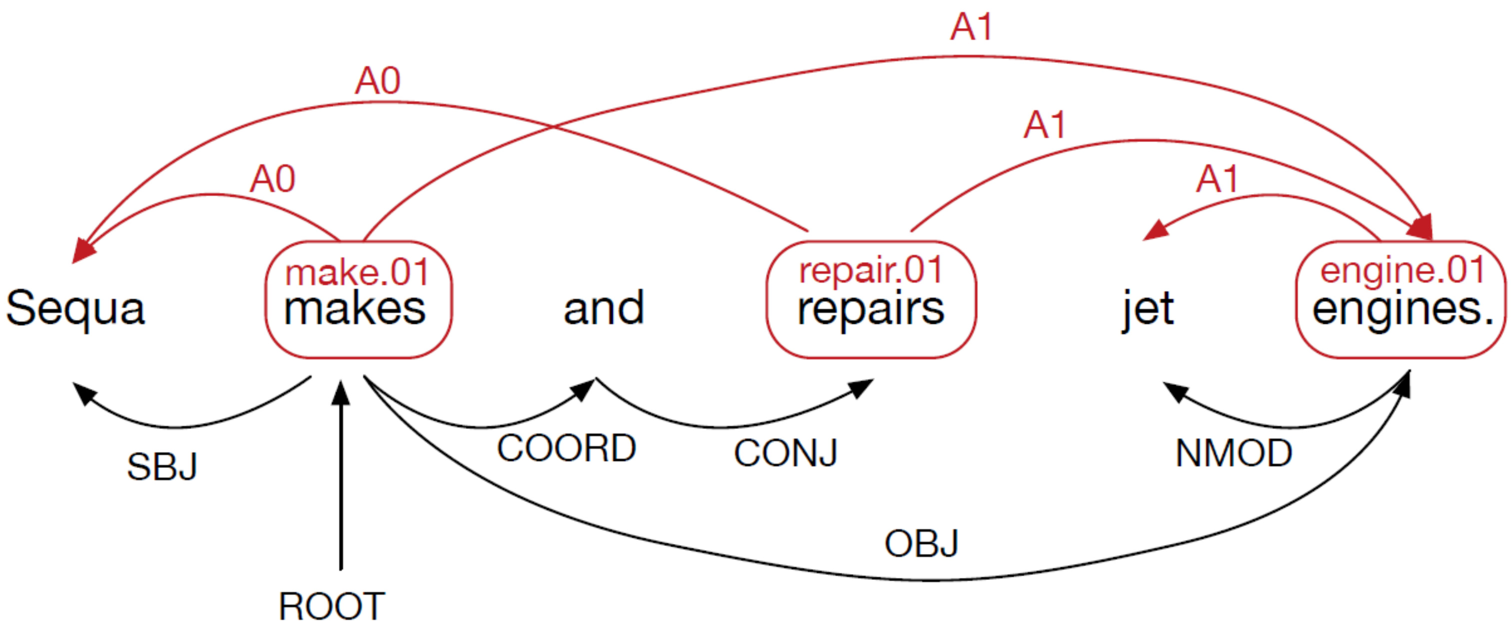
- Neural Machine Translation
- Text Generation
- Question Answering
- Information Extraction
- Sentiment
- Social Classification
- Syntactic Parsing
- Semantic Role Labeling
- Word Embedding
- Sentence Representation

GNNs for Semantic Role Labeling



Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling, Diego Marcheggiani, Ivan Titov, EMNLP 2017.

GNNs for Semantic Role Labeling



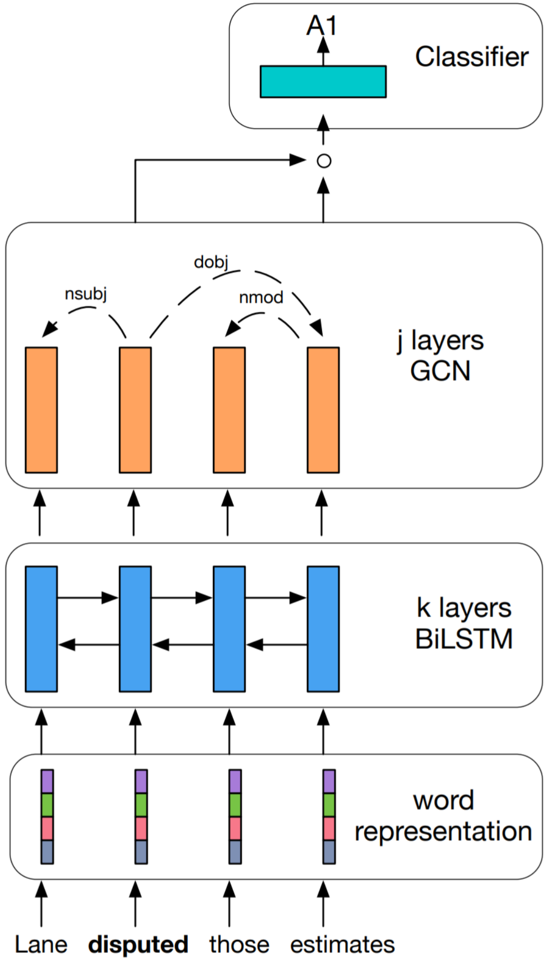
semantic

closely related

syntactic

Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling, Diego Marcheggiani, Ivan Titov, EMNLP 2017.

GNNs for Semantic Role Labeling

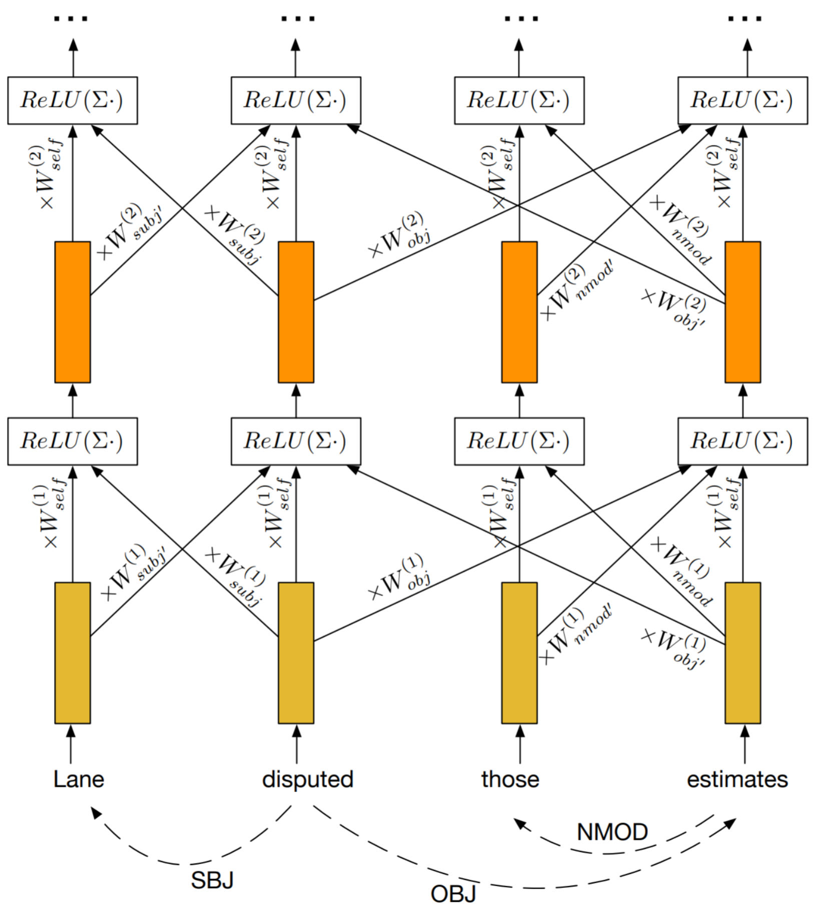


exploit syntactic information when predicting semantics using GCN:

$$h_v = ReLU\left(\sum_{u \in \mathcal{N}(v)} (W x_u + b)\right)$$

Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling, Diego Marcheggiani, Ivan Titov, EMNLP 2017.

GNNs for Semantic Role Labeling



Standard GCN layer

$$h_v^{(k+1)} = ReLU\left(\sum_{u \in \mathcal{N}(v)} W^{(k)} h_u^{(k)} + b^{(k)}\right)$$

Syntactic GCN layer

$$h_v^{(k+1)} = ReLU\left(\sum_{u \in \mathcal{N}(v)} W_{L(u,v)}^{(k)} h_u^{(k)} + b_{L(u,v)}^{(k)}\right)$$

Handle over-parameterized

$$W_{L(u,v)}^{(k)} = V_{dir(u,v)}^{(k)}$$

Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling, Diego Marcheggiani, Ivan Titov, EMNLP 2017.

Experiments

System (English)	P	R	F_1
LSTMs	84.3	81.1	82.7
LSTMs + GCNs (K=1)	85.2	81.6	83.3
LSTMs + GCNs (K=2)	84.1	81.4	82.7
LSTMs + GCNs (K=1), no gates	84.7	81.4	83.0
GCNs (no LSTMs), K=1	79.9	70.4	74.9
GCNs (no LSTMs), K=2	83.4	74.6	78.7
GCNs (no LSTMs), K=3	83.6	75.8	79.5
GCNs (no LSTMs), K=4	82.7	76.0	79.2

English dataset

System (Chinese)	P	R	F_1
LSTMs	78.3	72.3	75.2
LSTMs + GCNs (K=1)	79.9	74.4	77.1
LSTMs + GCNs (K=2)	78.7	74.0	76.2
LSTMs + GCNs (K=1), no gates	78.2	74.8	76.5
GCNs (no LSTMs), K=1	78.7	58.5	67.1
GCNs (no LSTMs), K=2	79.7	62.7	70.1
GCNs (no LSTMs), K=3	76.8	66.8	71.4
GCNs (no LSTMs), K=4	79.1	63.5	70.4

Chinese dataset

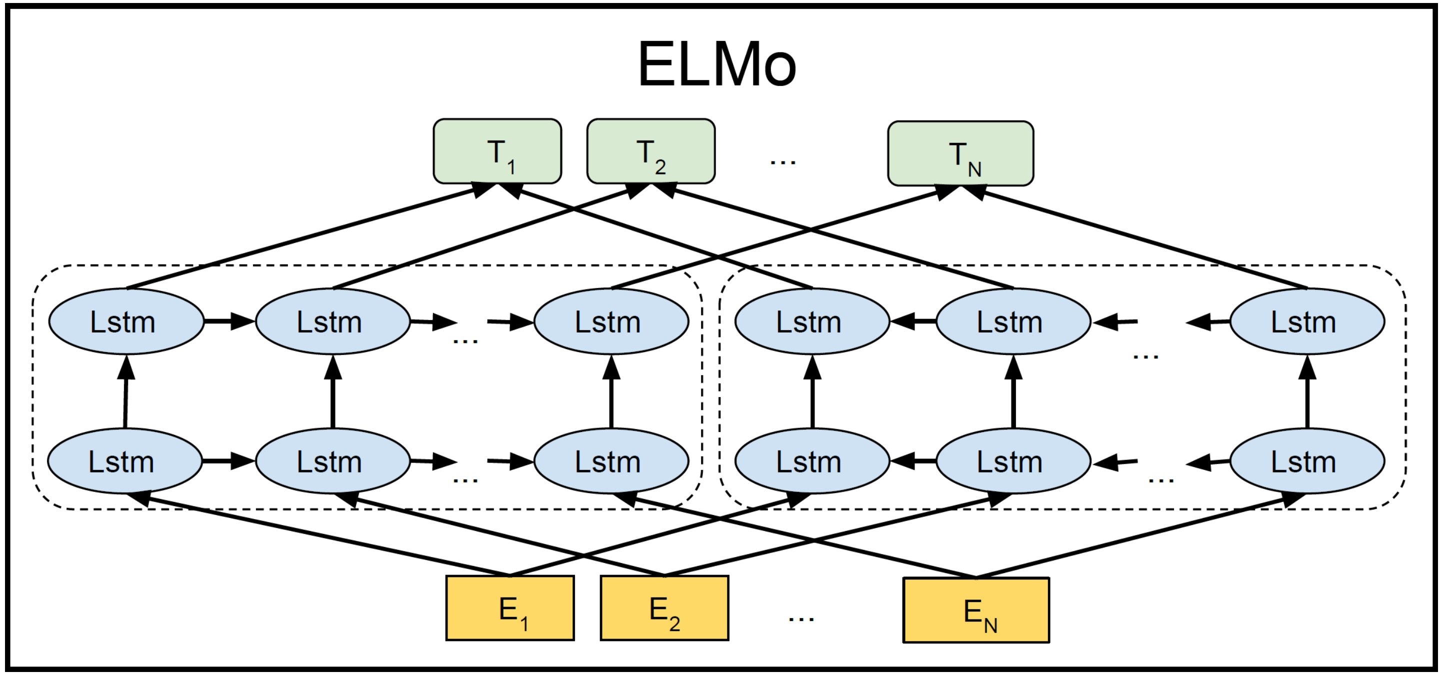
Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling, Diego Marcheggiani, Ivan Titov, EMNLP 2017.

Solving problems using graph encoding

- Neural Machine Translation
- Text Generation
- Question Answering
- Information Extraction
- Sentiment
- Social Classification
- Syntactic Parsing
- Semantic Role Labeling
- Word Embedding
- Sentence Representation

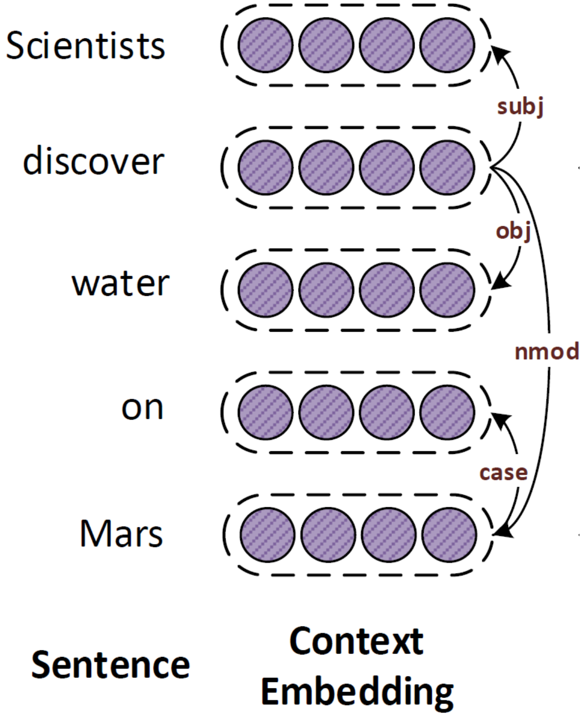
Incorporating Syntactic and Semantic Information in Word Embeddings using GCNs

Most existing word embedding utilize sequential context.



Shikhar Vashishth, Manik Bhandari et al., Incorporating Syntactic and Semantic Information in Word Embeddings using Graph Convolutional Networks, ACL2019.

SynGCN

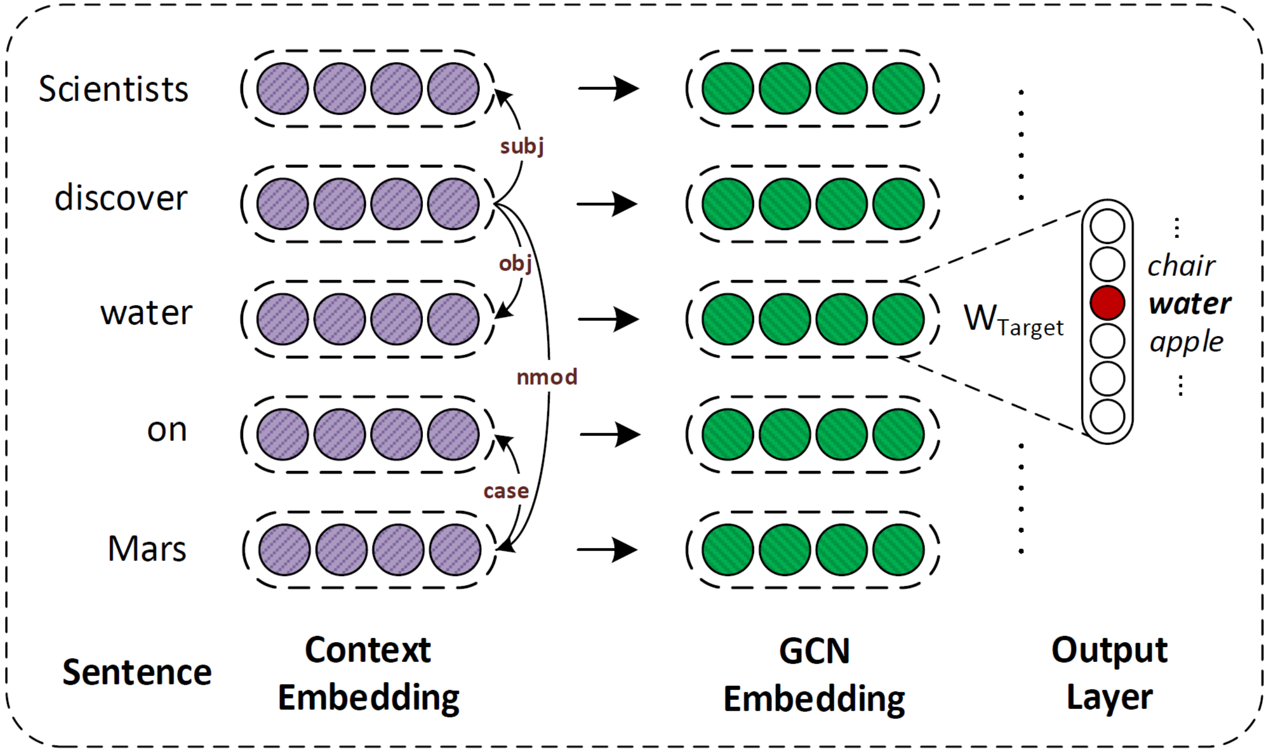


Vocabulary explode:

Scientists
discover
scientists_subj
water_obj
mars_nmod
water
on
Mars
…
…

Shikhar Vashishth, Manik Bhandari et al., Incorporating Syntactic and Semantic Information in Word Embeddings using Graph Convolutional Networks, ACL2019.

SynGCN



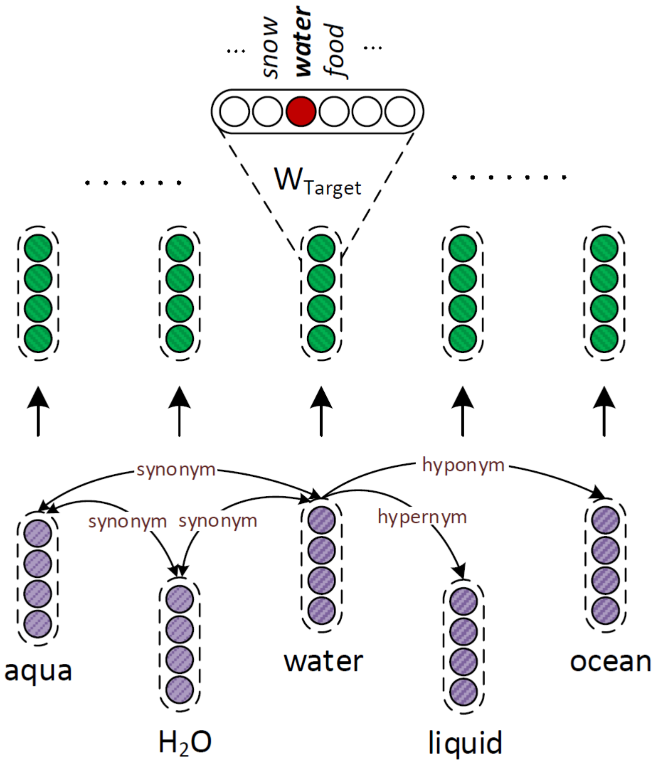
SynGCN (Sentence-level)

$$h_v^{k+1} = f\left(\sum_{u \in \mathcal{N}_+(v)} g_{l_{uv}}^k \times \left(W_{l_{uv}}^k h_u^k + b_{l_{uv}}^k\right)\right)$$

$$g_{l_{uv}}^k = \sigma\left(\hat{W}_{l_{uv}}^k h_u^k + \hat{b}_{l_{uv}}^k\right)$$

Shikhar Vashishth, Manik Bhandari et al., Incorporating Syntactic and Semantic Information in Word Embeddings using Graph Convolutional Networks, ACL2019.

SemGCN



SemGCN (Corpus-level)

$$h_v^{k+1} = f\left(\sum_{u \in \mathcal{N}_+(v)} g_{l_{uv}}^k \times \left(W_{l_{uv}}^k h_u^k + b_{l_{uv}}^k\right)\right)$$

$$g_{l_{uv}}^k = \sigma\left(\hat{W}_{l_{uv}}^k h_u^k + \hat{b}_{l_{uv}}^k\right)$$

Shikhar Vashishth, Manik Bhandari et al., Incorporating Syntactic and Semantic Information in Word Embeddings using Graph Convolutional Networks, ACL2019.

Training word embedding

$$E = \sum_{t=1}^{|V|} \log P(w_t | w_1^t, w_2^t \ldots w_{N_t}^t)$$

$$P(w_t | w_1^t, w_2^t \ldots w_{N_t}^t) = \frac{\exp(v_{w_t}^T h_t)}{\sum_{i=1}^{|V|} \exp(v_{w_i}^T h_t)}$$

$$E = \sum_{t=1}^{|V|} \left(v_{w_t}^T h_t - \log \sum_{i=1}^{|V|} \exp(v_{w_i}^T h_t) \right)$$

Target embedding

Shikhar Vashishth, Manik Bhandari et al., Incorporating Syntactic and Semantic Information in Word Embeddings using Graph Convolutional Networks, ACL2019.

Experiments

	Word Similarity				Concept Categorization				Word Analogy	
Method	WS353S	WS353R	SimLex999	RW	AP	Battig	BLESS	ESSLI	SemEval2012	MSR
Word2vec	71.4	52.6	38.0	30.0	63.2	43.3	77.8	63.0	18.9	44.0
GloVe	69.2	**53.4**	36.7	29.6	58.0	41.3	80.0	59.3	18.7	45.8
Deps	65.7	36.2	39.6	33.0	61.8	41.7	65.9	55.6	22.9	40.3
EXT	69.6	44.9	43.2	18.6	52.6	35.0	65.2	66.7	21.8	18.8
SynGCN	**73.2**	45.7	**45.5**	**33.7**	**69.3**	**45.2**	**85.2**	**70.4**	**23.4**	**52.8**

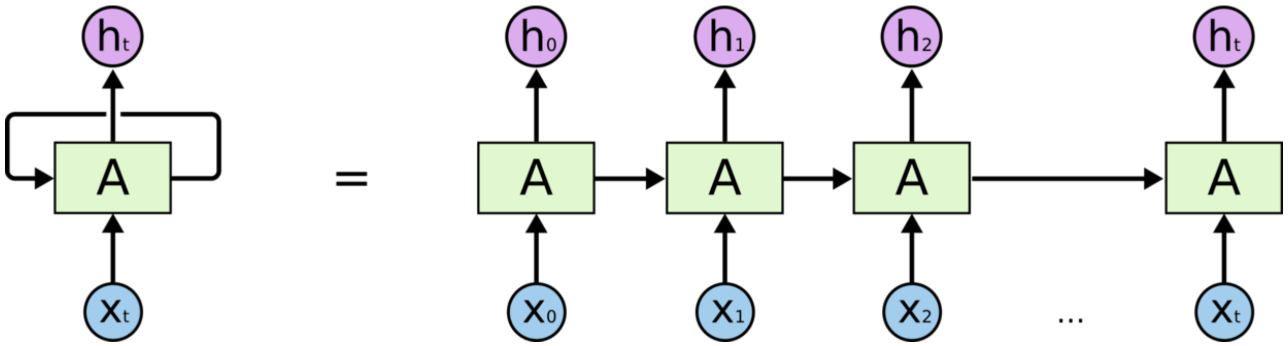
Shikhar Vashishth, Manik Bhandari et al., Incorporating Syntactic and Semantic Information in Word Embeddings using Graph Convolutional Networks, ACL2019.

Solving problems using graph encoding

- Neural Machine Translation
- Text Generation
- Question Answering
- Information Extraction
- Sentiment
- Social Classification
- Syntactic Parsing
- Semantic Role Labeling
- Word Embedding
- Sentence Representation

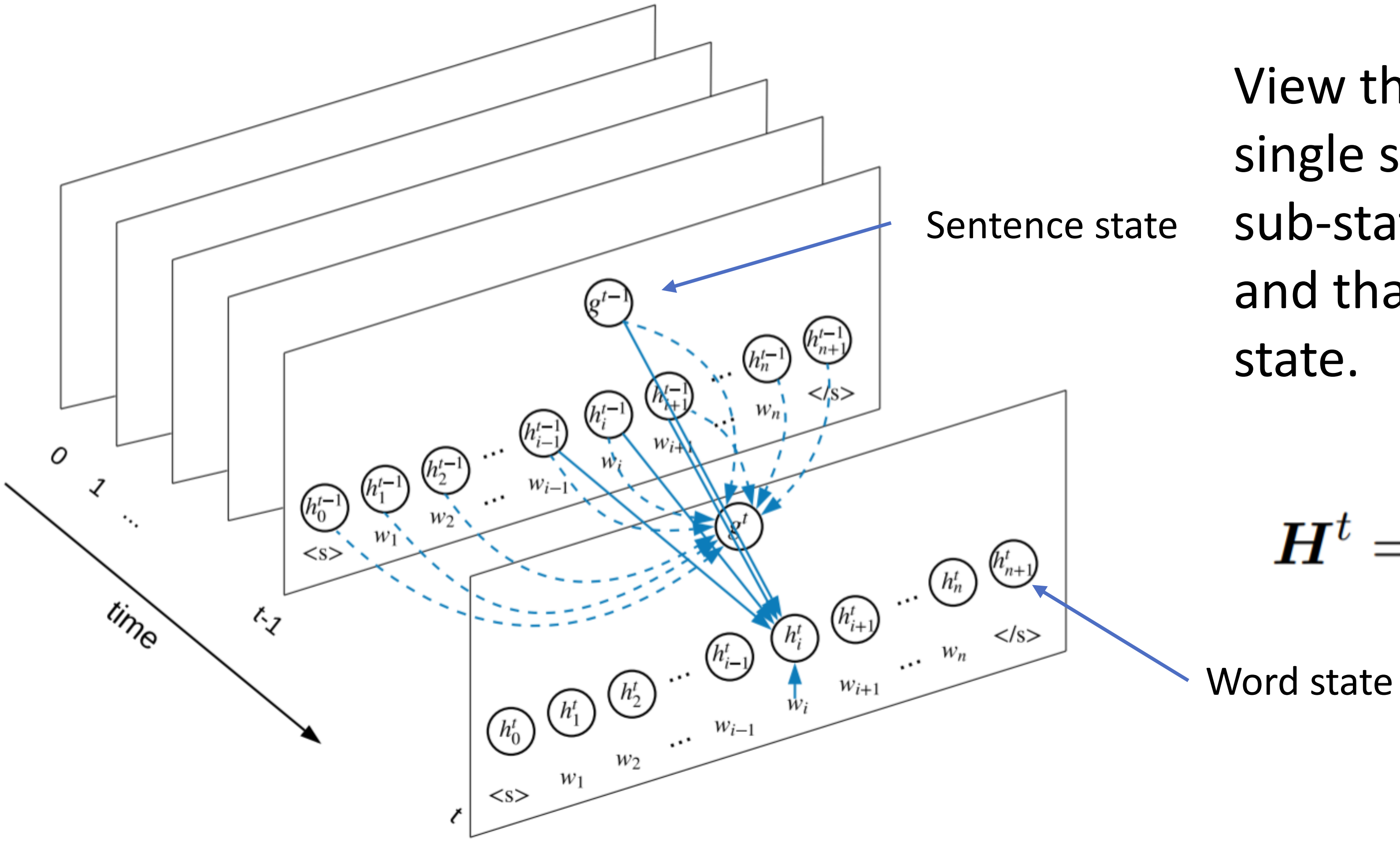
Sentence-state LSTM

Bi-LSTM suffer various limitations due to their sequential nature



Yue Zhang, Qi Liu et al., Sentence-State LSTM for Text Representation, ACL 2018.

Sentence-state LSTM



View the whole sentence as a single state, and we update the sub-states for individual words and than overall sentence-level state.

$$\boldsymbol{H}^t = \langle \boldsymbol{h}_0^t, \boldsymbol{h}_1^t, \ldots, \boldsymbol{h}_{n+1}^t, \boldsymbol{g}^t \rangle,$$

Yue Zhang, Qi Liu et al., Sentence-State LSTM for Text Representation, ACL 2018.

Text classification

Model	Time (s)	Acc	# Param
LSTM	67	80.72	5,977K
BiLSTM	106	81.73	7,059K
2 stacked BiLSTM	207	81.97	9,221K
3 stacked BiLSTM	310	81.53	11,383K
4 stacked BiLSTM	411	81.37	13,546K
S-LSTM	65	82.64*	8,768K
CNN	34	80.35	5,637K
2 stacked CNN	40	80.97	5,717K
3 stacked CNN	47	81.46	5,808K
4 stacked CNN	51	81.39	5,855K
Transformer (N=6)	138	81.03	7,234K
Transformer (N=8)	174	81.86	7,615K
Transformer (N=10)	214	81.63	8,004K
BiLSTM+Attention	126	82.37	7,419K
S-LSTM+Attention	87	83.07*	8,858K

Movie review development results.

Yue Zhang, Qi Liu et al., Sentence-State LSTM for Text Representation, ACL 2018.

Text classification

Model	Accuracy	Train (s)	Test (s)
Socher et al. (2011)	77.70	–	–
Socher et al. (2012)	79.00	–	–
Kim (2014)	81.50	–	–
Qian et al. (2016)	81.50	–	–
BiLSTM	81.61	51	1.62
2 stacked BiLSTM	81.94	98	3.18
3 stacked BiLSTM	81.71	137	4.67
3 stacked CNN	81.59	31	1.04
Transformer (N=8)	81.97	89	2.75
S-LSTM	**82.45***	41	1.53

Test set results on movie review dataset.

Yue Zhang, Qi Liu et al., Sentence-State LSTM for Text Representation, ACL 2018.

Text classification

Dataset	SLSTM	Time (s)	BiLSTM	Time (s)	2 BiLSTM	Time (s)
Camera	**90.02***	50 (2.85)	87.05	115 (8.37)	88.07	221 (16.1)
Video	**86.75***	55 (3.95)	84.73	140 (12.59)	85.23	268 (25.86)
Health	**86.5**	37 (2.17)	85.52	118 (6.38)	85.89	227 (11.16)
Music	**82.04***	52 (3.44)	78.74	185 (12.27)	80.45	268 (23.46)
Kitchen	**84.54***	40 (2.50)	82.22	118 (10.18)	83.77	225 (19.77)
DVD	**85.52***	63 (5.29)	83.71	166 (15.42)	84.77	217 (28.31)
Toys	85.25	39 (2.42)	85.72	119 (7.58)	**85.82**	231 (14.83)
Baby	**86.25***	40 (2.63)	84.51	125 (8.50)	85.45	238 (17.73)
Books	**83.44***	64 (3.64)	82.12	240 (13.59)	82.77	458 (28.82)
IMDB	**87.15***	67 (3.69)	86.02	248 (13.33)	86.55	486 (26.22)
MR	**76.2**	27 (1.25)	75.73	39 (2.27)	75.98	72 (4.63)
Appeal	85.75	35 (2.83)	86.05	119 (11.98)	**86.35***	229 (22.76)
Magazines	**93.75***	51 (2.93)	92.52	214 (11.06)	92.89	417 (22.77)
Electronics	**83.25***	47 (2.55)	82.51	195 (10.14)	82.33	356 (19.77)
Sports	**85.75***	44 (2.64)	84.04	172 (8.64)	84.78	328 (16.34)
Software	**87.75***	54 (2.98)	86.73	245 (12.38)	86.97	459 (24.68)
Average	**85.38***	47.30 (2.98)	84.01	153.48 (10.29)	84.64	282.24 (20.2)

Results on the 16 datasets, time format: train(test)

Yue Zhang, Qi Liu et al., Sentence-State LSTM for Text Representation, ACL 2018.

Sequence labelling

Model	Accuracy	Train (s)	Test (s)
Manning (2011)	97.28	–	–
Collobert et al. (2011)	97.29	–	–
Sun (2014)	97.36	–	–
Søgaard (2011)	97.50	–	–
Huang et al. (2015)	**97.55**	–	–
Ma and Hovy (2016)	**97.55**	–	–
Yang et al. (2017)	**97.55**	–	–
BiLSTM	97.35	254	22.50
2 stacked BiLSTM	97.41	501	43.99
3 stacked BiLSTM	97.40	746	64.96
S-LSTM	**97.55**	237	22.16

Results on PTB (POS tagging).

Yue Zhang, Qi Liu et al., Sentence-State LSTM for Text Representation, ACL 2018.

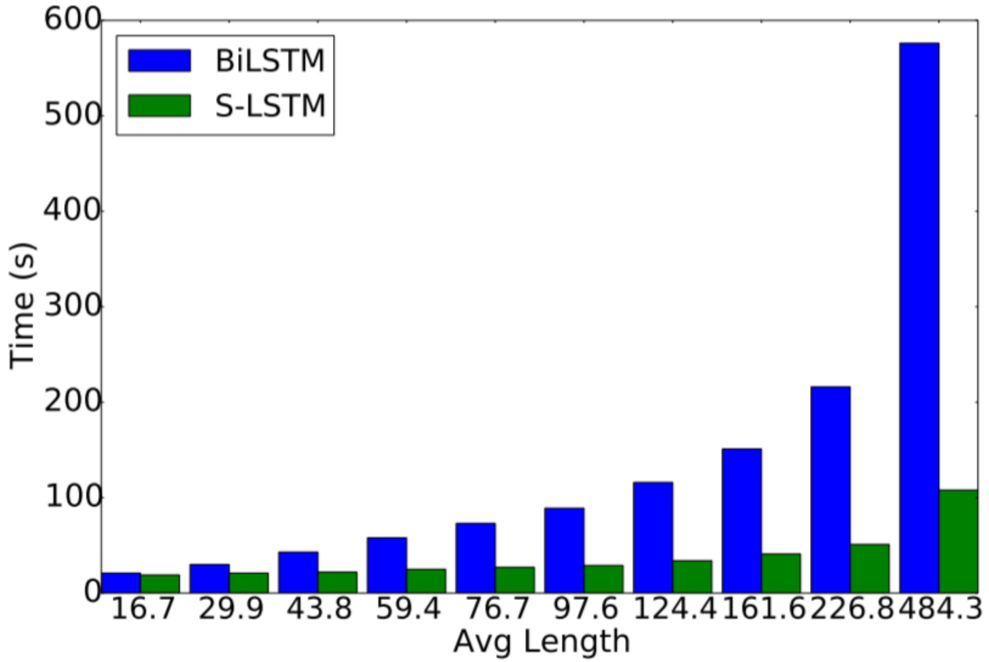
Sequence labelling

Model	F1	Train (s)	Test (s)
Collobert et al. (2011)	89.59	–	–
Passos et al. (2014)	90.90	–	–
Luo et al. (2015)	91.20	–	–
Huang et al. (2015)	90.10	–	–
Lample et al. (2016)	90.94	–	–
Ma and Hovy (2016)	91.21	–	–
Yang et al. (2017)	91.26	–	–
Rei (2017)	86.26	–	–
Peters et al. (2017)	**91.93**	–	–
BiLSTM	90.96	82	9.89
2 stacked BiLSTM	91.02	159	18.88
3 stacked BiLSTM	91.06	235	30.97
S-LSTM	**91.57***	79	9.78

Results on CoNLL03 (NER).

Yue Zhang, Qi Liu et al., Sentence-State LSTM for Text Representation, ACL 2018.

Analysis



Time against sentence length.

Yue Zhang, Qi Liu et al., Sentence-State LSTM for Text Representation, ACL 2018.

Conclusion of this talk

- Graph neural network (GNN) and its applications on several major NLP tasks;

- Thoughts on effective methods.