[EMNLP20]ENT-DESC:
Entity Description Generation by Exploring Knowledge Graph

Liying Cheng,[1,2] Dekun Wu,[3] Lidong Bing,[2] Yan Zhang,[1] Zhanming Jie,[1] Wei Lu,[1] Luo Si[2]

[1] *Singapore University of Technology and Design*
[2] *DAMO Academy, Alibaba Group*
[3] *York University, Canada*

Speaker: 杨晰

xyang41@stu.ecnu.edu.cn

Outline

- Background: KG-to-Text Generation
- Contributions
  - Task Description
  - ENT-DESC Dataset
  - Multi-Graph Convolutional network (MGCN)
- Experiments
- Conclusions

BG: KG-to-Text Generation

- Structured knowledge (e.g. triples about the main entity) ➡️ Natural text (describing the main entity)

BG: KG-to-Text Generation

- Structured knowledge (e.g. triples about the main entity) ➡️ Natural text (describing the main entity)
- Datasets: WebNLG, WIKIBIO, E2E

BG: KG-to-Text Generation

- Structured knowledge (e.g. triples about the main entity) ➡️ Natural text (describing the main entity)
- Datasets: WebNLG, WIKIBIO, E2E

```
<modifiedtripleset>
    <mtriple>Indonesia | leaderName | Jusuf_Kalla</mtriple>
    <mtriple>Bakso | region | Indonesia</mtriple>
    <mtriple>Bakso | ingredient | Noodle</mtriple>
    <mtriple>Bakso | country | Indonesia</mtriple>
</modifiedtripleset>
<lex>
Bakso is a food containing noodles;it is found in Indonesia where Jusuf Kalla is the leader.
</lex>
```

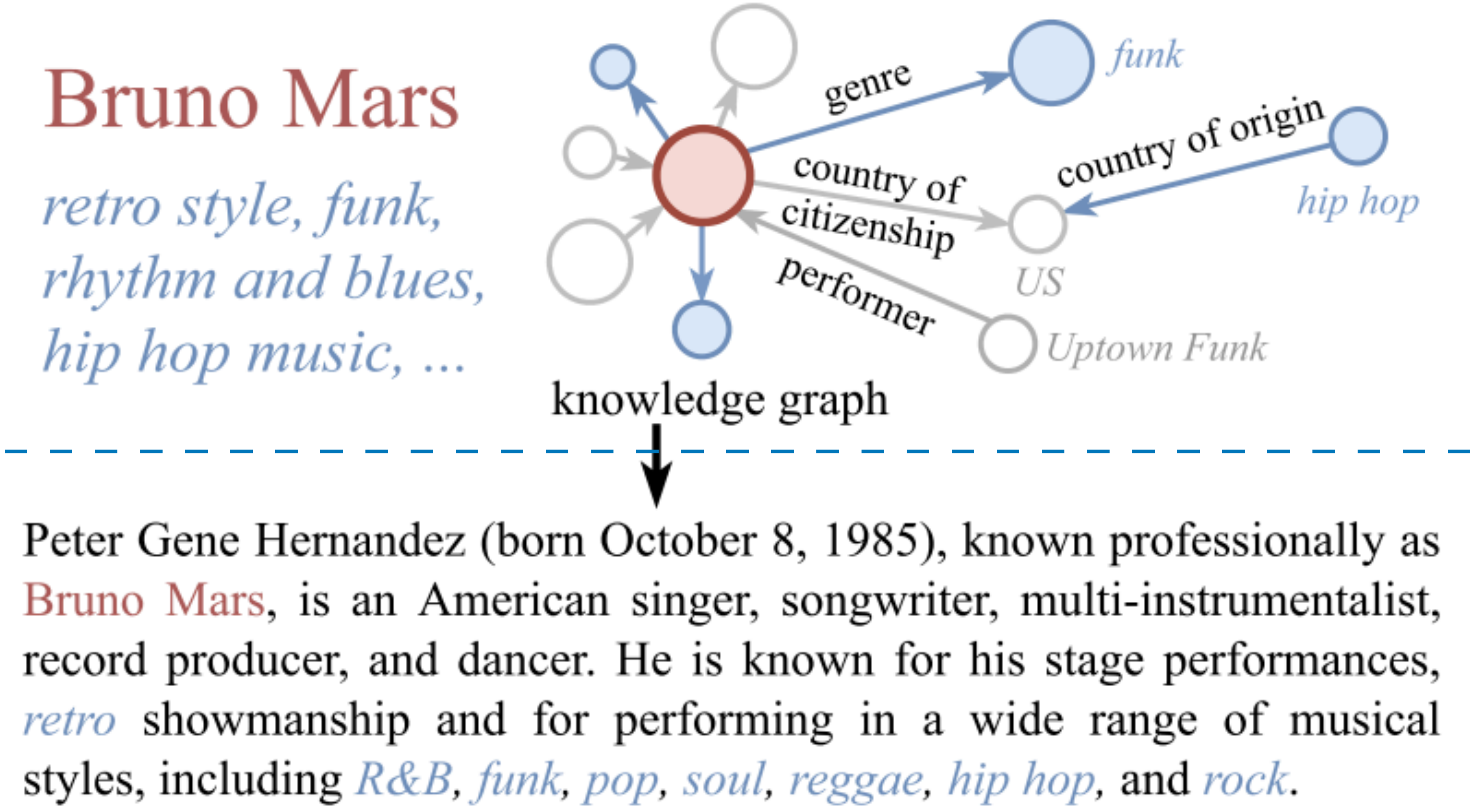
BG: KG-to-Text Generation

- Structured knowledge (e.g. triples about the main entity) ➡️ Natural text (describing the main entity)
- Datasets: WebNLG, WIKIBIO, E2E
- Problems
  - Good alignment between input and output hard to access
  - In practice, knowledge about the main entity may be abundant ➡️ models should have differentiation capability

Contributions

- Tackle a more practical and more challenging task of entity description generation by exploring KG
- Construct a large-scale dataset: ENT-DESC
- Propose a multi-graph structure: MGCN

Task Description



Task: Entity description generation by exploring KG

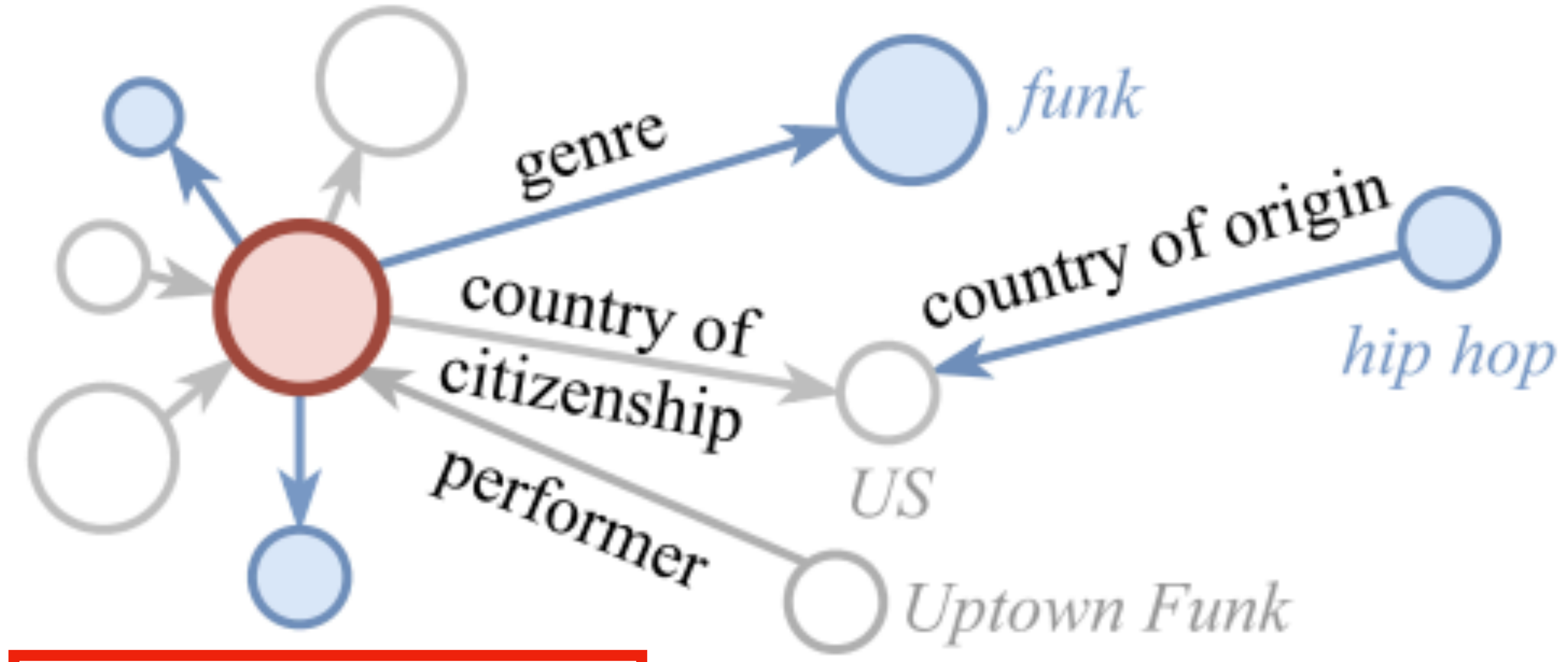
Task Description

Main entity

Topic-related entities

Bruno Mars

retro style, funk, rhythm and blues, hip hop music, ...



knowledge graph

Peter Gene Hernandez (born October 8, 1985), known professionally as Bruno Mars, is an American singer, songwriter, multi-instrumentalist, record producer, and dancer. He is known for his stage performances, *retro* showmanship and for performing in a wide range of musical styles, including *R&B*, *funk*, *pop*, *soul*, *reggae*, *hip hop*, and *rock*.

Task: Entity description generation by exploring KG

Task Description

- Task: entity description generation by exploring KG
- Input
  - A set of entities: $\mathbf{e} = \{\mathrm{E}_1, ..., \mathrm{E}_n\}$
    - Main entity: E_1
    - Topic-related entities: $\mathrm{E}_2, ..., \mathrm{E}_n$
  - KG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, can be written as $\mathcal{G} = \{\langle \mathrm{V}_{\mathrm{S}_1}, \mathrm{P}_1, \mathrm{V}_{\mathrm{O}_1} \rangle, ..., \langle \mathrm{V}_{\mathrm{S}_\mathrm{M}}, \mathrm{P}_\mathrm{M}, \mathrm{V}_{\mathrm{O}_\mathrm{M}} \rangle\}$
- Output
  - Natural language text: $\mathbf{y} = \{y_1, y_2, \cdots, y_T\}$

Dataset: ENT-DESC

❓ How to generate this dataset?

Dataset: ENT-DESC

? How to generate this dataset?

1. Calculate the PageRank score for more than 9.9 million Wikipedia pages
2. Extract the categories from the Wikidata from the top 100k highest scored pages
3. Manually select 90 categories out of the top 200 most frequent ones as the seed categories

Dataset: ENT-DESC

? How to generate this dataset?

1. Calculate the PageRank score for more than 9.9 million Wikipedia pages
2. Extract the categories from the Wikidata from the top 100k highest scored pages
3. Manually select 90 categories out of the top 200 most frequent ones as the seed categories

- Main entities: the entities from these seed categories
- Output text: the first paragraphs of their associated Wikipedia pages
- Topic-related entities: entities with hyperlink in the pages
- KG: gather neighbors of the main entities and 1-hop/2-hop paths in Wikidata

Dataset: ENT-DESC

	WebNLG	AGENDA	E2E	ENT-DESC
# instances	43K	41K	51K	110K
Input vocab	4.4K	54K	120	420K
Output vocab	7.8K	78K	5.2K	248K
# distinct entities	3.1K	297K	77	691K
# distinct relations	358	7	8	957
Avg. # triples per input	3.0	4.4	5.6	27.4
Avg. # words per output	23.7	141.3	20.3	31.0

Table 1: Dataset statistics of WebNLG, AGENDA and our prepared ENT-DESC.

Dataset: ENT-DESC

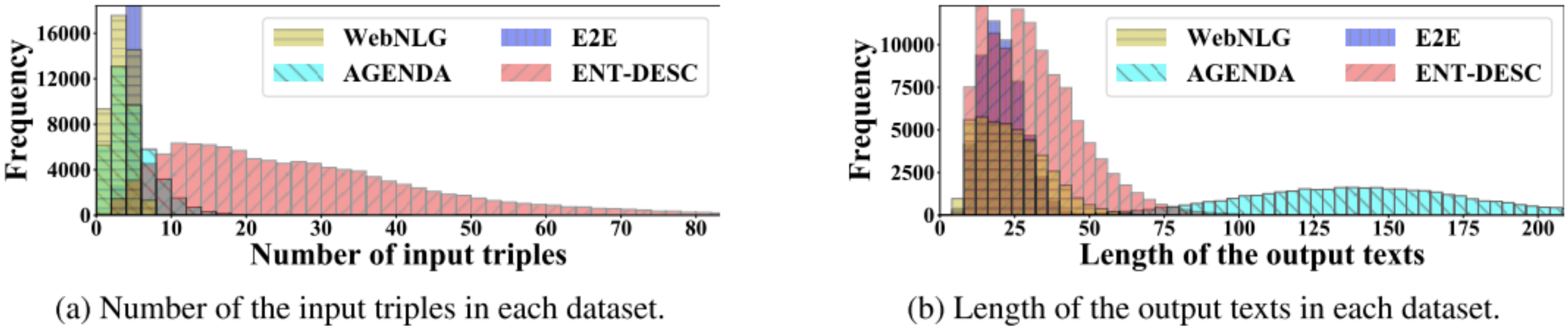
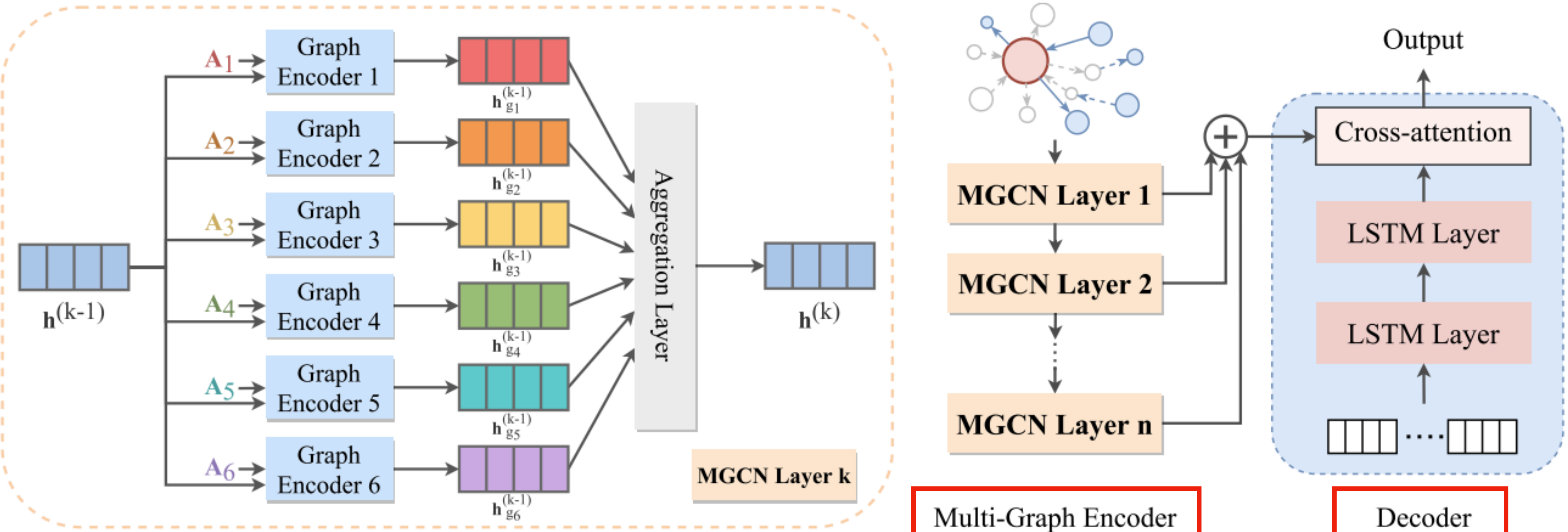


Figure 2: Dataset comparison among WebNLG, AGENDA, E2E and our ENT-DESC.

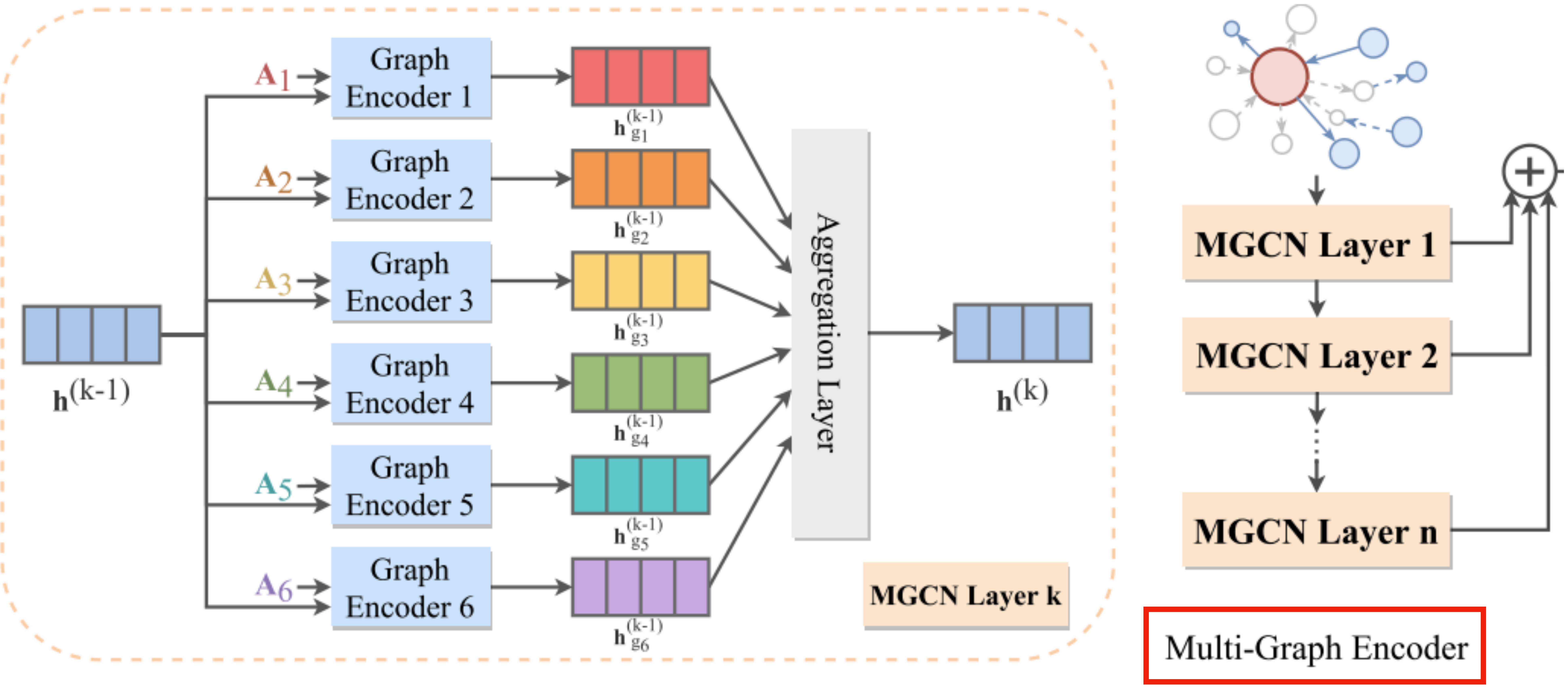
Multi-Graph Convolutional network(MGCN)



Multi-Graph Encoder

Decoder

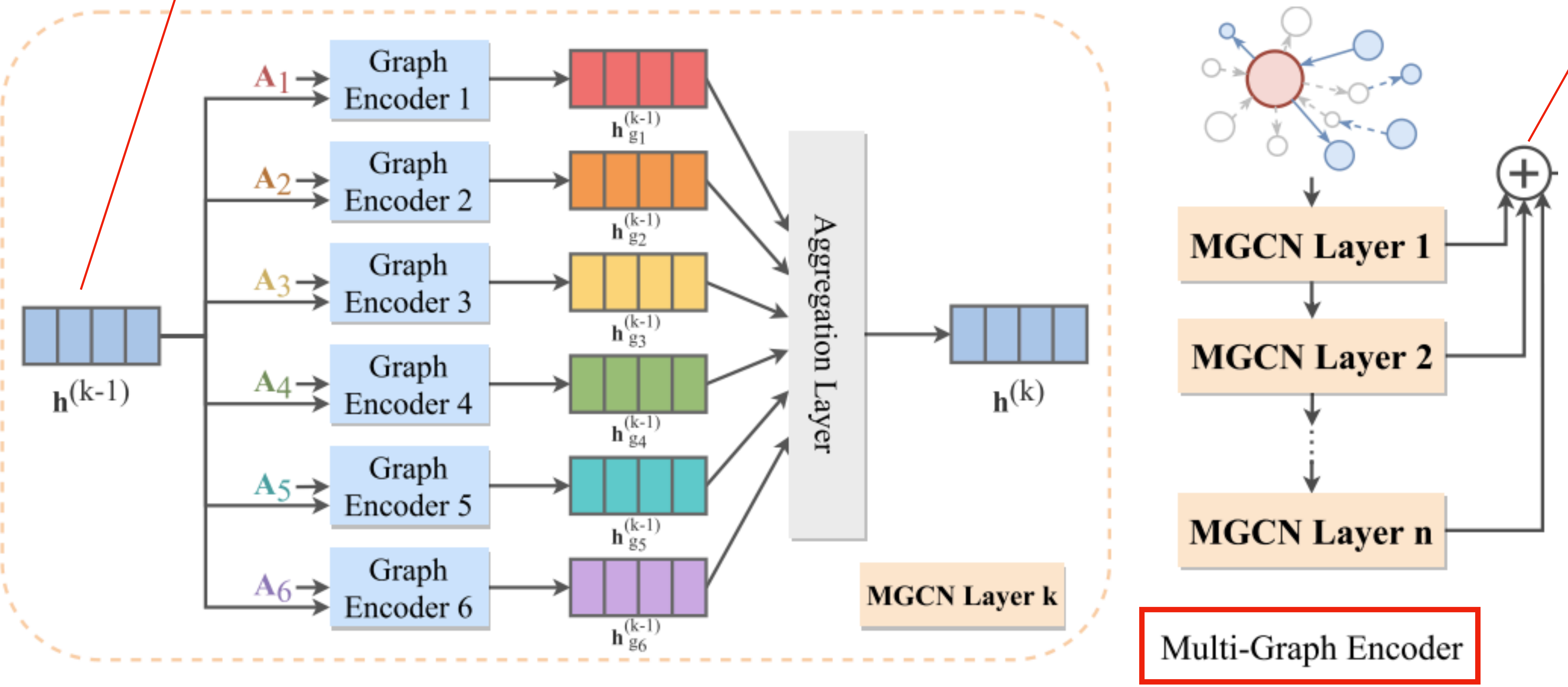
MGCN: Encoder



MGCN: Encoder

nodes representations of the graph

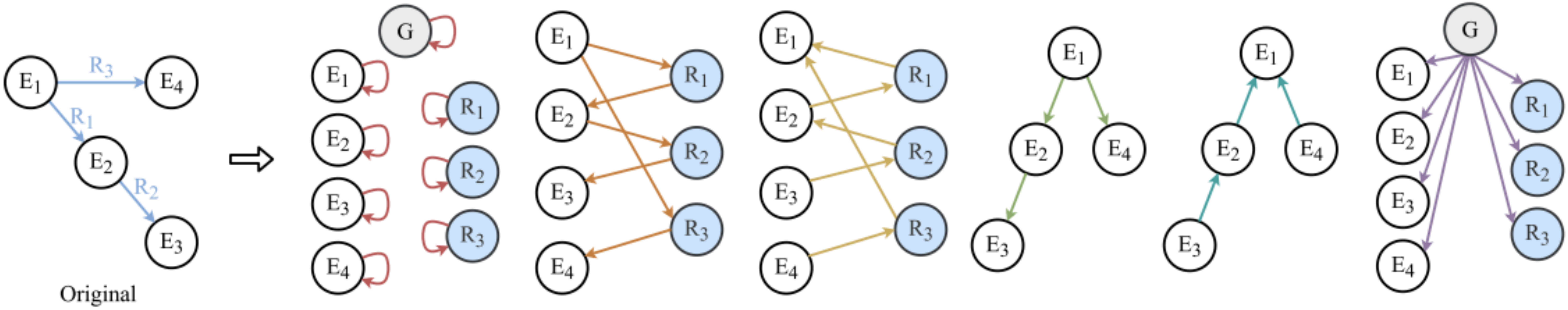
$$\mathbf{h}^{(final)} = [\mathbf{h}^{(1)}, \cdots \mathbf{h}^{(n)}]$$



MGCN: Multi-Graph Transformation

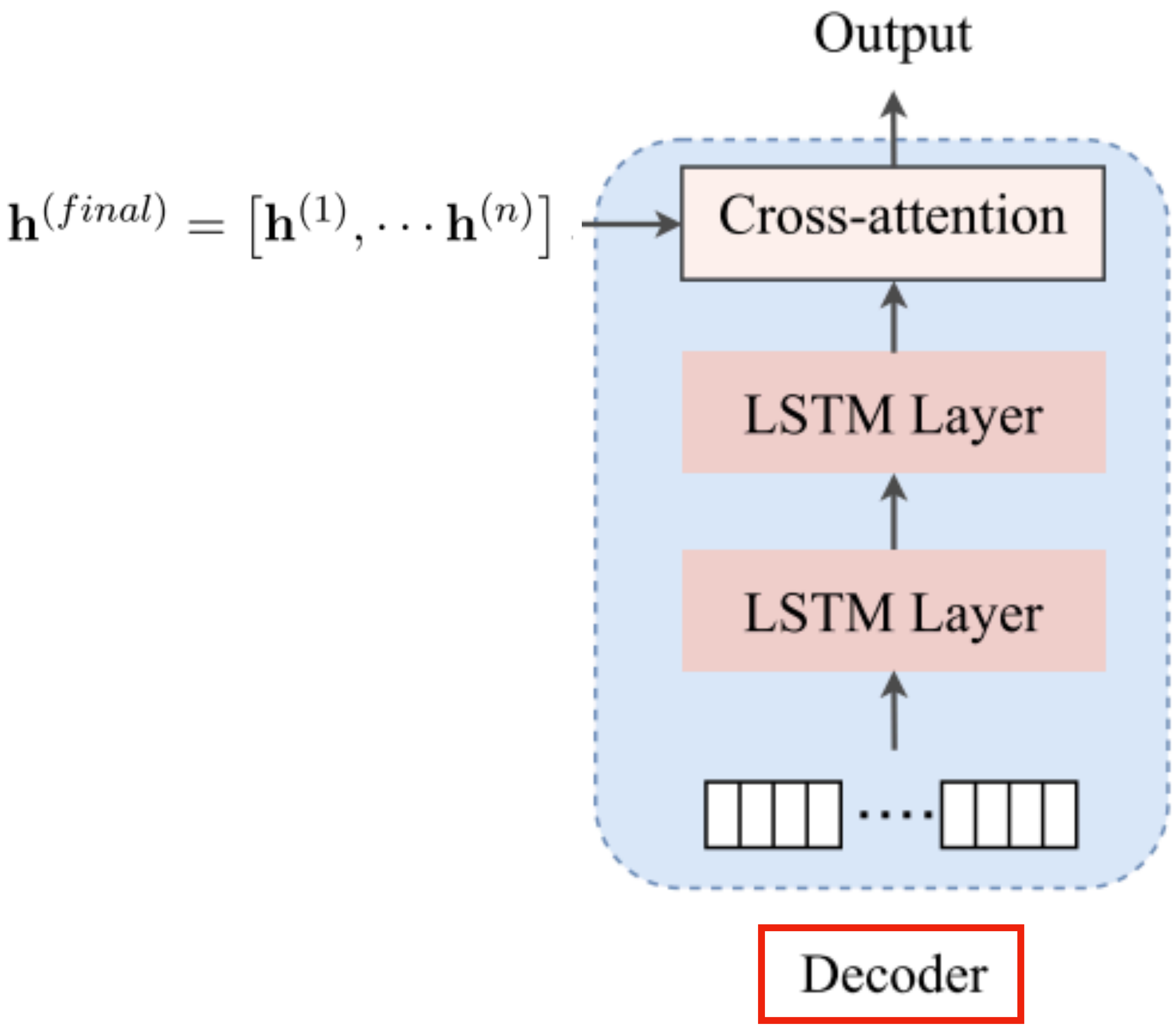
- Reasons
  - using basic graph encoder to encode edges or not: parameter explosion / information loss
  - Levi graph (Bipartite) : no direct entity-to-entity connection

MGCN: Multi-Graph Transformation



KG g1: self g2: default1 g3: reverse1 g4: default2 g5: reverse2 g6: global

MGCN: Decoder



$$\mathcal{L} = -\sum_{t=1}^{T} \log p_{\boldsymbol{\theta}}(y_t | y_1, ..., y_{t-1}, \mathbf{h}^{(final)})$$

Experiments

Models	BLEU	METEOR	TER↓	ROUGE$_1$	ROUGE$_2$	ROUGE$_L$	PARENT
S2S (Bahdanau et al., 2014)	6.8	10.8	80.9	38.1	21.5	40.7	10.0
GraphTransformer (Koncel-Kedziorski et al., 2019)	19.1	16.1	94.5	53.7	37.6	54.3	21.4
GRN (Beck et al., 2018)	24.4	18.9	70.8	54.1	38.3	55.5	21.3
GCN (Marcheggiani and Perez-Beltrachini, 2018)	24.8	19.3	70.4	54.9	39.1	56.2	21.8
DeepGCN (Guo et al., 2019)	24.9	19.3	70.2	55.0	39.3	56.2	21.8
MGCN	**25.7**	**19.8**	**69.3**	**55.8**	**40.0**	**57.0**	**23.5**
MGCN + CNN	**26.4**	**20.4**	69.4	56.4	40.5	**57.4**	**24.2**
MGCN + AVG	26.1	20.2	**69.2**	56.4	40.3	57.3	23.9
MGCN + SUM	**26.4**	20.3	69.8	56.4	**40.6**	**57.4**	23.9
GCN + delex	28.4	22.9	65.9	61.8	45.5	62.1	30.2
MGCN + CNN + delex	29.6	**23.7**	**63.2**	**63.0**	**46.7**	**63.2**	**31.9**
MGCN + SUM + delex	**30.0**	**23.7**	67.4	62.6	46.3	62.7	31.5
The rows below are results of generating from entities only without exploring the KG.							
E2S	23.3	20.4	68.7	58.8	41.9	58.2	27.7
E2S + delex	21.8	20.5	67.5	59.5	39.5	59.2	23.4
E2S-MEF	24.2	21.3	65.8	59.8	43.3	60.0	26.3
E2S-MEF + delex	20.6	20.3	66.5	59.1	40.0	59.3	24.3

Table 2: Main results of models on ENT-DESC dataset. ↓ indicates lower is better.

Experiments

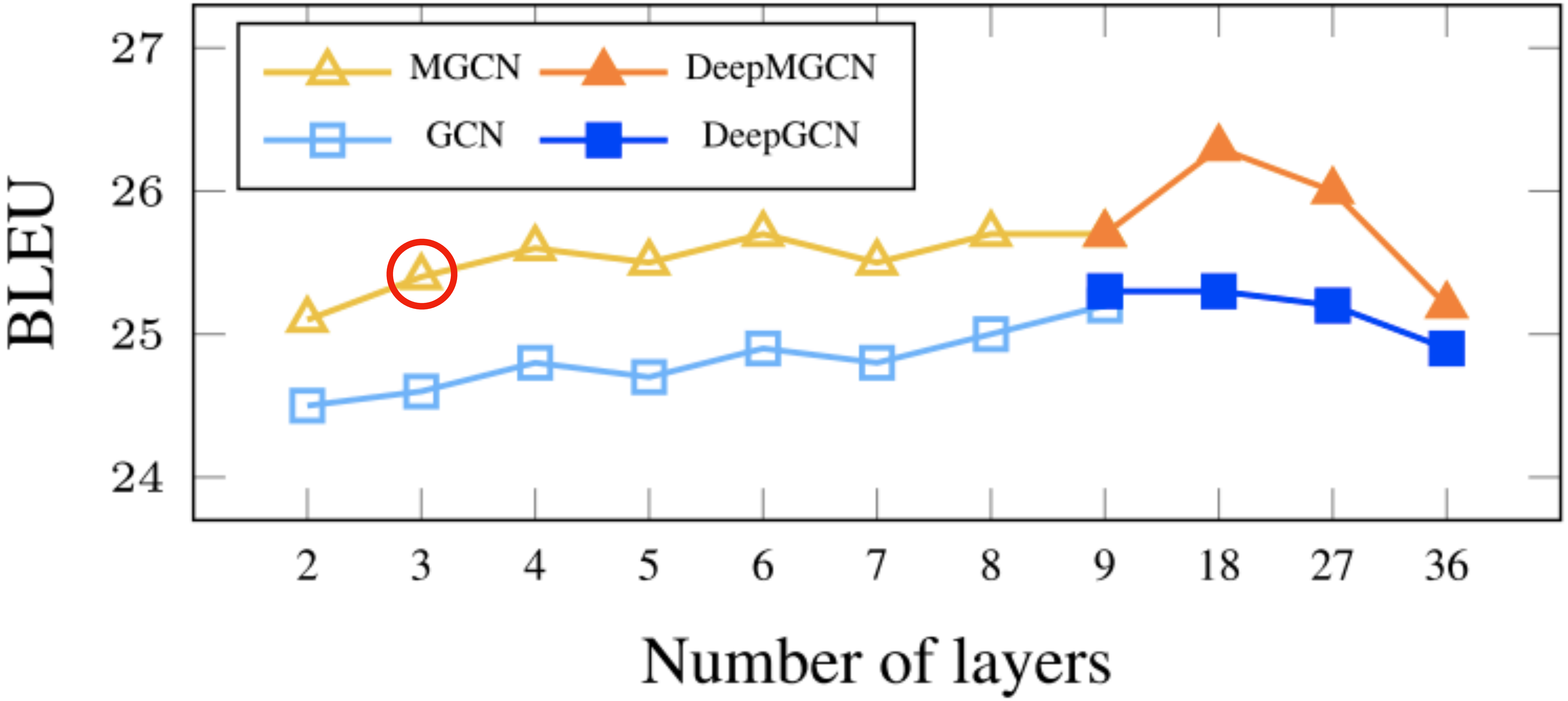


Figure 5: Effect of different numbers of layers.

Experiments

# Input triples	# Instances	GCN	MGCN+SUM	Δ (BLEU)
1 to 10	1,790	19.4	21.3	+1.9
11 to 20	2,999	22.6	24.6	+2.0
21 to 30	2,249	23.2	25.0	+1.8
31 to 50	2,830	31.6	32.8	+1.2
51 to 100	1,213	23.9	24.7	+0.8

Table 3: Effect of different numbers of input triples.

Experiments

Model	**BLEU**	**Δ (BLEU)**
MGCN + SUM	26.4	-
– g_6: *global*	26.0	-0.4
– g_5: *reverse2*	25.8	-0.6
– g_4: *default2*	26.1	-0.3
– g_3: *reverse1*	25.7	-0.7
– g_2: *default1*	26.1	-0.3
MGCN	25.7	-0.7
GCN	24.8	-1.4

Table 4: Results of the ablation study.

Experiments

Gold	The New Jersey Symphony Orchestra is an American symphony orchestra based in the state of New Jersey. The NJSO is the state orchestra of New Jersey, performing concert series in six venues across the state, and is the resident orchestra of the New Jersey Performing Arts Center in Newark, New Jersey.
GCN	The Newark Philharmonic Orchestra is an American orchestra based in Newark, New Jersey, United States.
MGCN +SUM	The New Jersey Symphony Orchestra is an American chamber orchestra based in Newark, New Jersey. The orchestra performs at the Newark Symphony Center at the Newark Symphony Center in Newark, New Jersey.

Table 5: An example of generated sentences.

Additional Experiments

SOTA

Models	BLEU
TILB-SMT (Gardent et al., 2017)	44.28
MELBOURNE (Gardent et al., 2017)	45.13
MGCN	**45.79**
MGCN + CNN	45.83
MGCN + AVG	**46.55**
MGCN + SUM	45.23

Table 6: Results on WebNLG dataset.

Conclusions

- Present a practical task: entity description generation by exploring KG
- Construct a large-scale and challenging dataset ENT-DESC
- Propose the MGCN model and show great effectiveness

[ACL20]Probing Linguistic Features of Sentence-Level Representations in Neural Relation Extraction

Christoph Alt, Aleksandra Gabryszak, Leonhard Hennig

German Research Center for Artificial Intelligence (DFKI)
Speech and Language Technology Lab

Speaker: 杨晰

xyang41@stu.ecnu.edu.cn

Outline

- Motivations
- Research Questions
- Probing Tasks
- Experiments Setup
- Experiments Results
- Conclusions

Motivations

- Majority of RE neural models: Sentence Encoder + Classifier
- Sentence representations are relevant to RE task, but it's unknown which exact properties have been learned
- Probing tasks(i.e. diagnostic classifiers), are a method to analyze the presence of specific information in a model's latent representations

Motivations

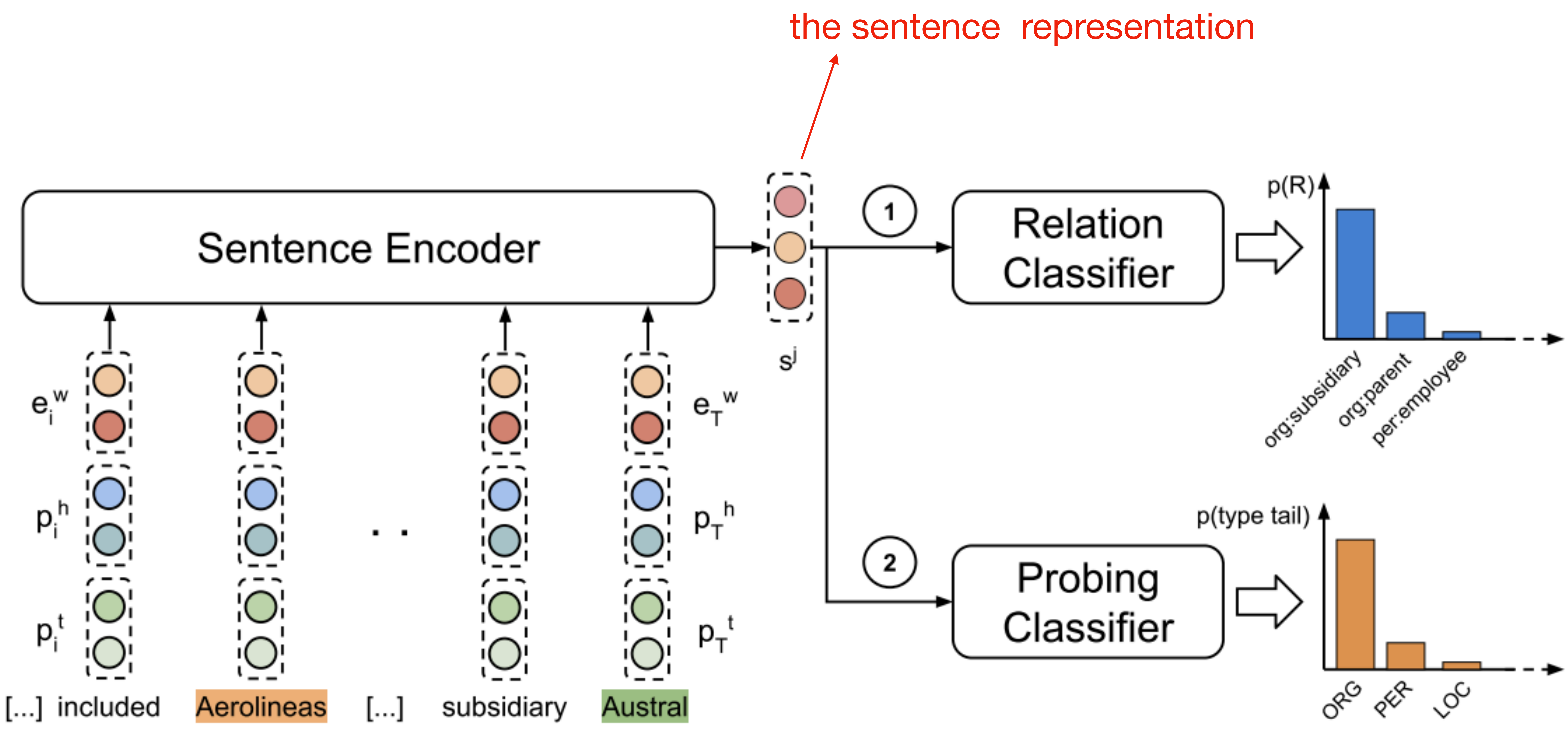
- Majority of RE neural models: Sentence Encoder + Classifier
- Sentence representations are relevant to RE task, but it's unknown which exact properties have been learned
- Probing tasks(i.e. diagnostic classifiers), are a method to analyze the presence of specific information in a model's latent representations

pinpoint the information a given RE model is relying on by probing tasks

Research Questions

- Sentence-level representations in neural RE models
  - 14 probing tasks ➡️ The ability of these representations to encode some well-known linguistic features relevant to RE
  - The effect of the four encoding architectures on the features in these representations
  - The effect of additional input features on the learned representations.

Probing Tasks



Probing Tasks

- Surface information
  - SentLen; ArgDist; EntExist

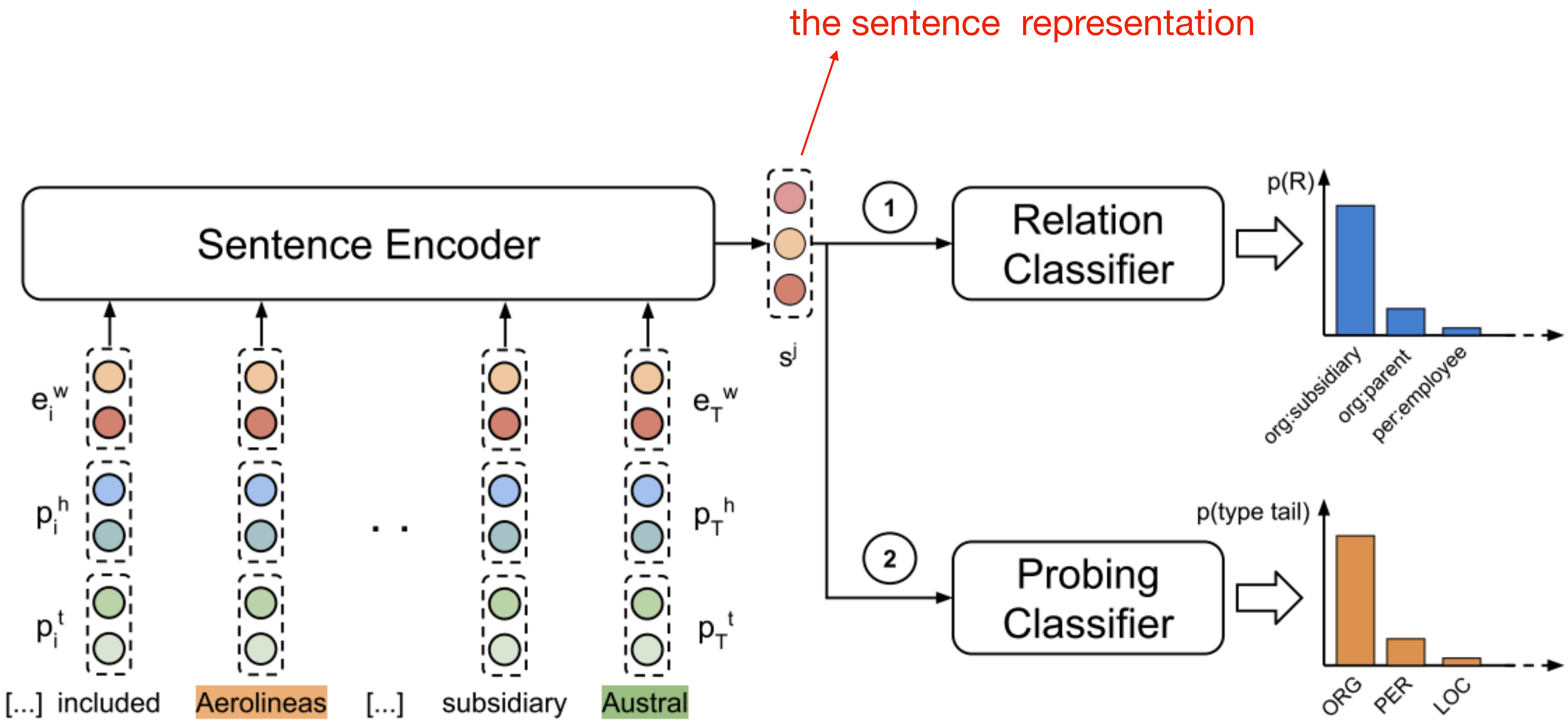
Probing Tasks

- Surface information
  - SentLen; ArgDist; EntExist
- Syntactic information
  - TreeDepth; SDPTreeDepth;
  - ArgOrd
  - PosHeadL; PosHeadR; PosTailL; PosTailR

Probing Tasks

- Surface information
  - SentLen; ArgDist; EntExist
- Syntactic information
  - TreeDepth; SDPTreeDepth;
  - ArgOrd
  - PosHeadL; PosHeadR; PosTailL; PosTailR
- Semantic/Argument information
  - TypeHead; TypeTail
  - GRHead; GRTail (nsubj, nsubjpass, dobj, iobj, other)

Experiments Setup



Experiments Setup

- Sentence Encoders
  - CNN; Bi-LSTM; GCN; Multi-headed self-attention
- Additional Linguistic Knowledge
  - Entity masking: replace entity with its entity type and grammatical role
  - ELMo; BERT

Experiments Setup

Dataset	# Relations	# Examples	Neg. examples
SemEval	19	10,717	17.4%
TACRED	42	106,264	79.5%

Table 1: Comparison of datasets used for evaluation

Experiments Results

- Performance on TACRED

	Type Head	Type Tail	Sent Len	Arg Dist	Arg Ord	Ent Exist	PosL Head	PosR Head	PosL Tail	PosR Tail	Tree Dep	SDP Dep	GR Head	GR Tail	F1 score
Majority vote	66.4	33.5	14.5	14.8	54.7	51.0	22.8	23.0	26.9	20.0	23.7	28.4	58.4	75.2	-
Length	66.4	33.5	**100.0**	13.8	54.8	59.4	18.6	24.7	26.9	20.1	**30.5**	29.6	58.4	75.2	-
ArgDist	66.4	33.5	16.5	**100.0**	54.7	77.5	14.9	23.0	26.9	19.8	23.8	35.3	58.4	75.2	-
BoE	77.7	47.6	61.1	22.6	97.3	66.5	33.7	41.5	32.5	36.3	29.8	31.0	66.3	77.4	39.4
CNN	94.0	85.8	47.6	88.1	98.8	**84.5**	70.7	76.1	84.0	86.5	28.5	44.0	78.0	88.6	55.9
+ ELMo	**97.0**	**90.2**	48.7	91.7	99.1	84.3	**76.1**	**81.2**	**86.6**	**90.1**	28.3	45.0	**82.8**	**91.9**	58.8
+ BERT ↓	95.9	88.8	44.7	46.0	93.8	79.9	64.7	74.4	80.8	88.4	29.4	41.0	77.7	90.0	59.7
+ BERT ↑	96.1	88.8	48.0	43.7	91.9	80.0	56.9	70.3	80.1	87.5	28.0	41.3	75.0	89.6	61.0
CNN ⊗	84.2	60.9	46.4	58.3	94.3	81.5	44.3	50.9	54.4	63.9	27.7	40.0	68.5	82.0	59.5
+ ELMo	82.8	69.8	47.4	75.6	98.1	82.9	54.2	60.2	65.4	77.3	28.7	42.4	71.9	85.0	61.7
+ BERT ↓	87.6	80.3	50.9	29.3	83.2	72.4	39.3	46.1	67.7	80.7	30.1	36.9	67.1	87.4	65.3
+ BERT ↑	87.2	79.3	50.6	25.3	78.3	69.8	39.6	42.9	59.9	77.5	30.3	35.1	65.6	86.9	66.1
Bi-LSTM	93.4	81.2	42.0	47.9	**99.4**	79.2	41.2	50.8	50.6	68.4	28.7	41.7	69.3	85.2	55.3
+ ELMo	96.4	89.6	27.9	47.0	97.9	80.9	47.8	52.5	67.2	72.6	25.2	42.8	72.1	90.0	61.8
+ BERT ↓	96.0	87.3	31.0	45.5	99.1	78.8	46.1	55.6	61.7	71.3	26.6	42.7	72.2	87.7	62.5
+ BERT ↑	96.0	87.7	28.6	45.3	97.7	80.4	48.0	50.9	61.4	67.4	25.1	42.3	70.8	87.0	63.1
Bi-LSTM ⊗	81.9	71.4	27.6	35.6	90.6	73.2	36.1	40.5	59.3	66.4	25.7	38.4	64.6	85.3	62.9
+ ELMo	82.8	50.7	30.6	19.7	73.4	65.0	32.0	35.9	37.9	41.8	28.0	32.2	63.0	79.5	64.1
+ BERT ↓	82.3	77.9	34.1	25.6	87.6	68.4	32.5	36.7	61.5	64.7	27.6	35.1	66.6	86.0	65.4
+ BERT ↑	81.7	79.6	30.2	21.3	81.1	67.0	30.6	33.8	55.9	55.1	27.3	34.2	64.1	84.9	66.1
GCN	93.0	81.9	18.8	35.5	86.0	74.4	48.6	48.8	51.2	52.3	24.0	**49.9**	74.2	85.9	57.4
+ ELMo	96.3	86.2	18.7	29.3	77.5	74.0	50.4	52.0	48.9	51.7	23.2	47.4	77.1	86.9	62.1
+ BERT ↓	96.0	85.2	20.7	31.2	83.6	74.2	48.6	52.4	47.4	50.4	23.9	48.7	74.4	85.3	62.9
+ BERT ↑	96.3	85.7	21.4	32.9	84.3	75.3	50.1	54.6	48.6	52.5	24.5	49.2	76.3	85.8	61.5
GCN ⊗	87.6	67.4	18.1	33.1	81.6	72.8	36.8	51.1	44.8	48.8	24.1	47.3	73.2	83.0	63.7
+ ELMo	92.7	68.6	18.6	26.4	76.8	71.4	41.9	50.4	43.6	45.1	23.8	47.1	76.3	83.9	65.4
+ BERT ↓	93.5	71.5	22.0	33.3	88.5	73.8	44.9	50.6	44.7	47.7	24.4	49.1	72.6	82.3	66.3
+ BERT ↑	93.4	72.0	23.7	33.2	90.4	73.9	42.8	50.1	44.0	48.3	24.9	48.0	72.9	83.0	65.9
S-Att.	89.9	81.8	22.7	32.8	75.7	78.1	34.1	38.9	40.8	44.8	26.1	38.2	60.7	81.1	57.6
+ ELMo	96.6	87.8	24.9	30.6	74.1	79.1	36.0	41.4	39.2	44.1	26.4	37.9	64.1	83.4	64.7
+ BERT ↓	96.2	87.0	25.9	31.4	75.6	76.5	35.3	40.8	39.8	44.4	25.4	39.1	61.8	81.3	63.9
+ BERT ↑	96.5	87.3	26.1	32.6	76.8	78.0	34.7	40.9	40.0	44.0	25.7	38.1	62.2	81.7	63.8
S-Att. ⊗	79.5	56.5	29.0	44.3	91.2	79.5	29.6	43.0	36.1	60.3	26.1	39.6	64.7	79.5	65.9
+ ELMo	78.2	44.4	25.1	31.5	72.3	77.1	31.6	37.5	34.4	34.8	26.2	36.7	62.1	75.9	66.6
+ BERT ↓	82.4	66.9	36.2	33.2	74.9	76.8	32.0	37.6	38.0	41.3	27.4	37.6	63.0	79.8	66.7
+ BERT ↑	80.0	69.0	31.9	32.8	78.6	76.6	30.3	34.2	37.5	39.2	27.0	38.2	60.4	79.9	**66.9**

Experiments Results

- Performance on SemEval

	Type Head	Type Tail	Sent Len	Arg Dist	PosL Head	PosR Head	PosL Tail	PosR Tail	Tree Dep	SDP Dep	GR Head	GR Tail	F1 score
Majority vote	22.0	21.3	25.7	42.1	62.1	39.3	38.3	34.0	25.4	67.2	37.3	80.9	-
Length	25.8	24.7	**100.0**	42.1	62.1	39.1	38.3	46.3	**44.3**	67.2	40.6	80.9	-
ArgDist	23.6	22.3	25.7	**100.0**	62.1	43.7	37.9	35.3	26.2	67.8	45.4	80.9	-
BoE	58.5	58.0	82.4	84.8	65.1	66.1	49.2	72.5	44.1	69.8	65.4	83.6	55.7
CNN	76.1	76.2	34.9	87.5	66.0	85.8	74.2	73.1	34.1	72.1	70.3	89.1	80.2
+ ELMo	81.3	81.8	38.1	88.5	70.0	89.0	79.5	76.4	35.5	71.8	75.1	90.9	84.4
+ BERT ↓	**83.9**	**84.1**	55.9	90.2	74.0	89.3	81.2	**84.6**	41.3	73.1	76.8	90.6	**86.3**
+ BERT ↑	83.4	83.7	54.3	90.4	**74.4**	**89.4**	**82.0**	82.8	42.0	73.0	78.3	90.8	86.0
Bi-LSTM	77.1	77.0	50.5	74.9	63.8	75.9	61.8	68.5	41.3	70.3	69.2	87.7	80.1
+ ELMo	81.5	81.8	41.1	66.6	62.8	71.8	59.3	64.5	37.5	70.1	70.0	87.6	83.7
+ BERT ↓	83.6	83.7	41.8	61.5	62.7	68.9	57.9	63.0	37.1	70.8	67.4	86.7	85.6
+ BERT ↑	82.5	82.8	41.8	66.0	63.1	70.8	58.6	64.3	37.7	71.0	68.9	87.5	85.1
GCN	75.4	75.5	35.0	81.5	68.5	87.5	71.2	55.5	35.5	**80.3**	76.3	91.7	79.6
+ ELMo	80.7	80.8	32.2	68.1	68.3	83.4	65.8	53.2	34.4	75.8	80.0	91.1	84.2
+ BERT ↓	82.5	83.0	42.5	66.5	73.6	84.7	69.2	66.3	38.9	77.2	82.1	91.0	85.7
+ BERT ↑	81.5	81.9	42.7	67.3	73.8	85.1	69.6	67.8	39.6	77.6	**84.2**	**91.9**	84.3
S-Att.	77.4	77.6	34.2	50.0	62.1	56.2	49.8	47.1	35.9	67.9	54.2	84.1	80.2
+ ELMo	80.7	81.3	33.1	46.2	62.0	53.9	49.1	45.7	34.7	68.1	54.9	84.4	83.6
+ BERT ↓	83.4	83.3	31.0	45.3	62.1	51.8	48.4	44.7	33.0	67.8	53.3	83.6	85.6
+ BERT ↑	82.8	82.8	30.6	46.1	62.1	52.7	48.2	44.4	33.6	67.9	54.6	84.1	84.9

Experiments Results

- Encoder Architecture
  - Except SentLen and ArgOrd, all encoders outperforms our baseline: Bag-of-Embeddings (BoE)
  - CNN / Bi-LSTM: perform well on local focus tasks
  - GCN: perform well on tree-related tasks
- Probing tasks
  - Except SentLen and ArgOrd, all encoders outperforms our baseline: Bag-of-Embeddings (BoE)
  - high performance on entity type and Pos tasks

Experiments Results

- Entity masking
  - performance decrease in probing tasks with a focus on argument position and distance (except S-Attn)
  - CNN / Bi-LSTM: decrease most
- Word Representations
  - Adding ELMo/BERT, increase performance on probing tasks with a focus on NE and Pos
  - Encoders with BERT have better RE performance
  - BERT without casing performs >= the cased version, on probing tasks with a focus on NE and Pos

Experiments Results

- Probing vs. Relation Extraction
  - encoders that perform better on probing tasks $\neq$ perform better on the downstream RE task
- Relation Extraction
  - Entity masking works

Conclusions

- Introduced a set of 14 probing tasks to study the linguistic features captured in sentence-level rep encoders trained on RE models
- Self-attentive encoders to be well suited for the RE on sentences of different complexity, though they consistently perform lower on probing tasks
- Future work
  - cover specific linguistic patterns
  - probing ability to specific entity types

Thanks!