Learning Representation Mapping for Relation Detection in Knowledge Base Question Answering

农伟 51194501160

Background

- KBQA: answers questions using an open-domain knowledge base

- KBQA systems:

1. the entity linking
2. the relation detection

Question:	where was **Mark Mifsud** born?
Candidate Relations:	**people.person.place_of_birth**
	people.person.nationality
	people.person.profession
	...
Triple:	<Mark Mifsud, people.person.place_of_birth, Malta>

Background

- **SimpleQuestion**

achieving over 90% accuracy in relation detection

- **Problem**

99% of the test relations exist in the training data

unseen relations never been trained

SimpleQuestion: 1,837 relations

FB2M: 6,700 relations

Representation Adapter

- Representations are fine-tuned on the labeled training data
- Unseen relations will not be updated properly

- Keep the representation unchanged
- **A representation adapter** to bridge the gap between

general purposed representations and task specific representations

Representation Adapter

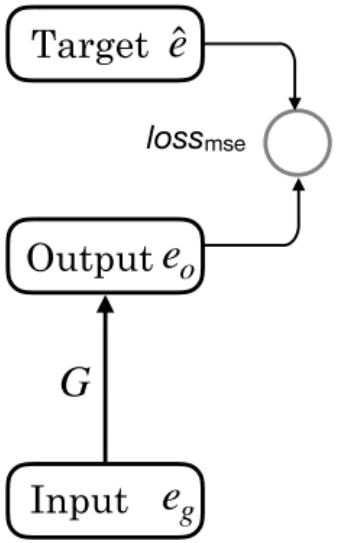
- **Basic Adapter**

train a traditional relation detection model get $\hat{e}$(**seen relations**)

a linear mapping function $G(\cdot)$

$$\mathcal{L}_{\text{adapter}} = \sum_{r \in S} loss(\hat{e}, G(e_{\text{g}})).$$

$$loss_{\text{MSE}}(\hat{e}, G(e_{\text{g}})) = ||\hat{e} - G(e_{\text{g}})||_2^2$$



Representation Adapter

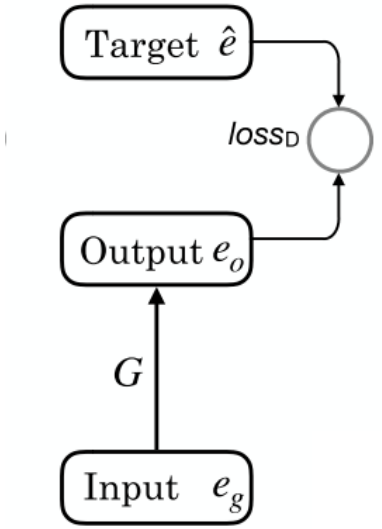
- **Adversarial Adapter**

a discriminator $D(\cdot)$ to discriminate the "real" representation

$G(\cdot)$ is acting as the generator

$$loss_{\mathrm{D}} = \mathbb{E}_{r \in S}[D(G(e_{\mathrm{g}}))] - \mathbb{E}_{r \in S}[D(\hat{e})]$$

$$loss_{\mathrm{G}} = -\mathbb{E}_{r \in S}[D(G(e_{\mathrm{g}}))]$$



Representation Adapter

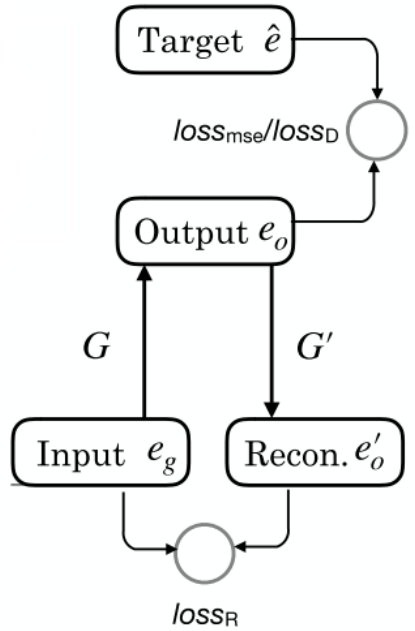
- **Reconstruction Loss**

a reversed adapter $G'(\cdot)$

The reconstruction loss is defined

for both seen and unseen relations

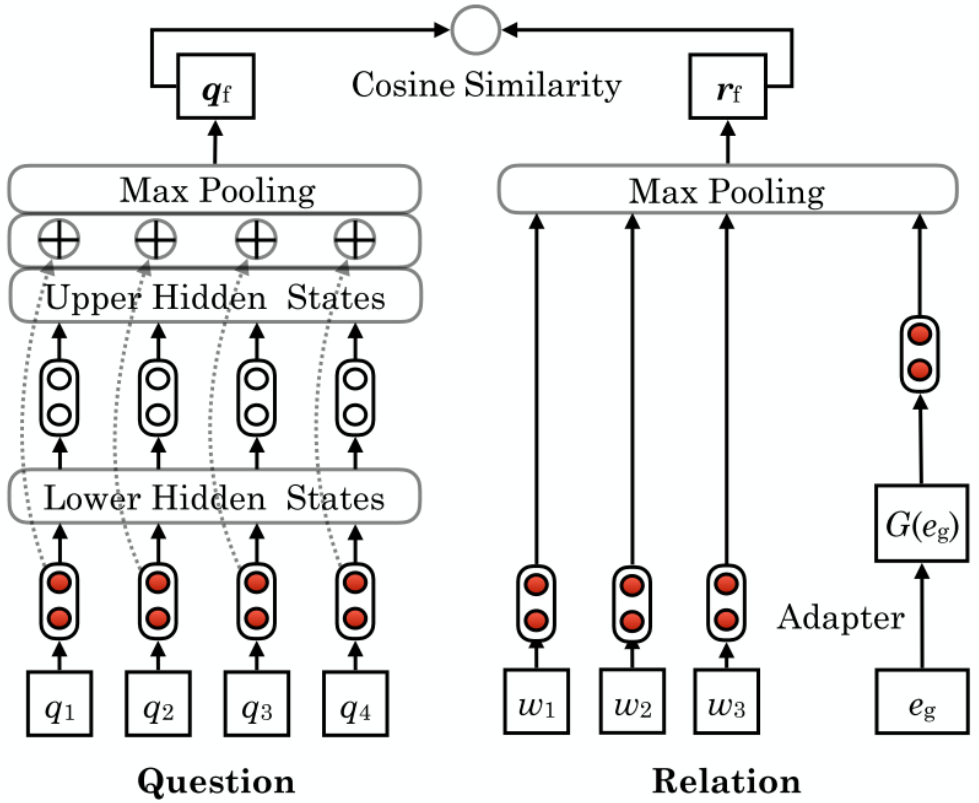
$$loss_{\mathrm{R}} = \sum_{r \in S \cup U} ||G'(G(e_{\mathrm{g}})) - e_{\mathrm{g}}||_2^2$$



Relation Detection with the Adapter

- HR-BiLSTM
- $s(\cdot,\cdot)$ is the cosine distance

$$\mathcal{L}_{\text{rd}} = \sum \max(0, \gamma - s(\mathbf{q}_{\text{f}}, \mathbf{r}_{\text{f}}^{+}) + s(\mathbf{q}_{\text{f}}, \mathbf{r}_{\text{f}}^{-}))$$



SimpleQuestion-Balance (SQB)

- re-organize the SQ dataset

Datasets	SQ	SQB
Train	75,910	75,819
Dev-seen	10,774	5,383
Dev-unseen	71	**5,758**
Test-seen	21,526	10,766
Test-unseen	161	**10,717**

Experiment

#	Model	Micro / Macro Average Accuracy on SQB (%) Test-seen	Test-unseen	All
1	HR-BiLSTM	**93.5±0.6** / 84.7±1.4	33.0±5.7 / 49.3±1.7	63.3±3.6 / 71.2±1.3
2	+ no fine-tune	93.4±0.7 / 83.8±0.7	57.8±9.8 / 60.8±2.0	75.6±5.0 / 75.0±0.6
3	+ no fine-tune + mapping	93.3±0.7 / 84.0±1.6	52.0±7.2 / 60.6±2.1	72.7±3.8 / 75.1±1.3
4	+ Basic-Adapter	92.8±0.7 / 84.1±1.2	76.0±7.5† / 69.5±2.0†	84.5±3.5 / 78.5±1.3
5	+ reconstruction	93.0±0.5 / 84.4±0.8	76.1±7.0† / 70.7±1.8†	84.6±3.3 / 79.2±0.8
6	+ Adversarial-Adapter	92.6±0.9 / **86.4±1.4**	77.1±7.1† / **73.2±2.1**†	**84.9±3.2** / **81.4±1.4**
7	+ reconstruction [Final]	92.4±0.8 / 86.1±0.7	**77.3±7.6**† / 73.0±1.7†	**84.9±3.5** / 81.1±0.8

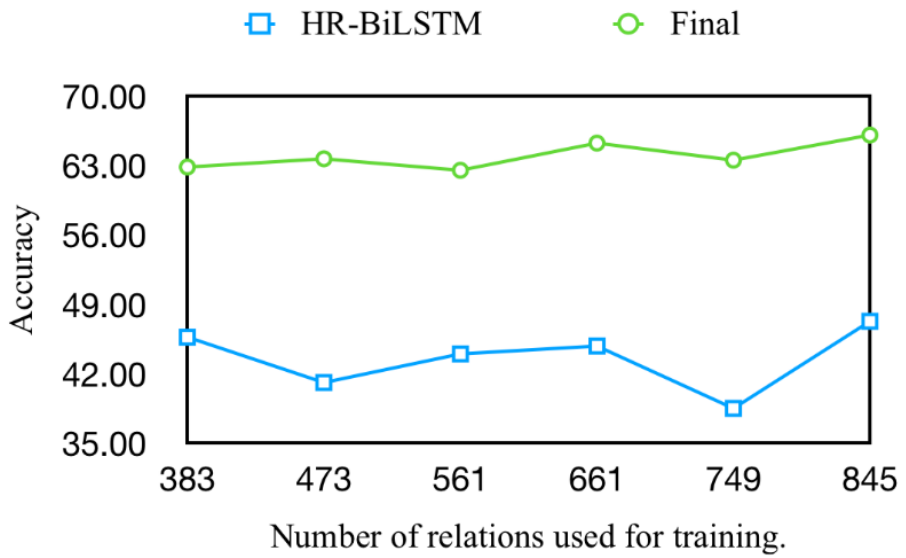
Model	Accuracy (%)
HR-BiLSTM	48.5±3.3
+ no fine-tune	56.4±3.4
Final	**63.7±3.2**

Analysis

- Seen Relation Bias

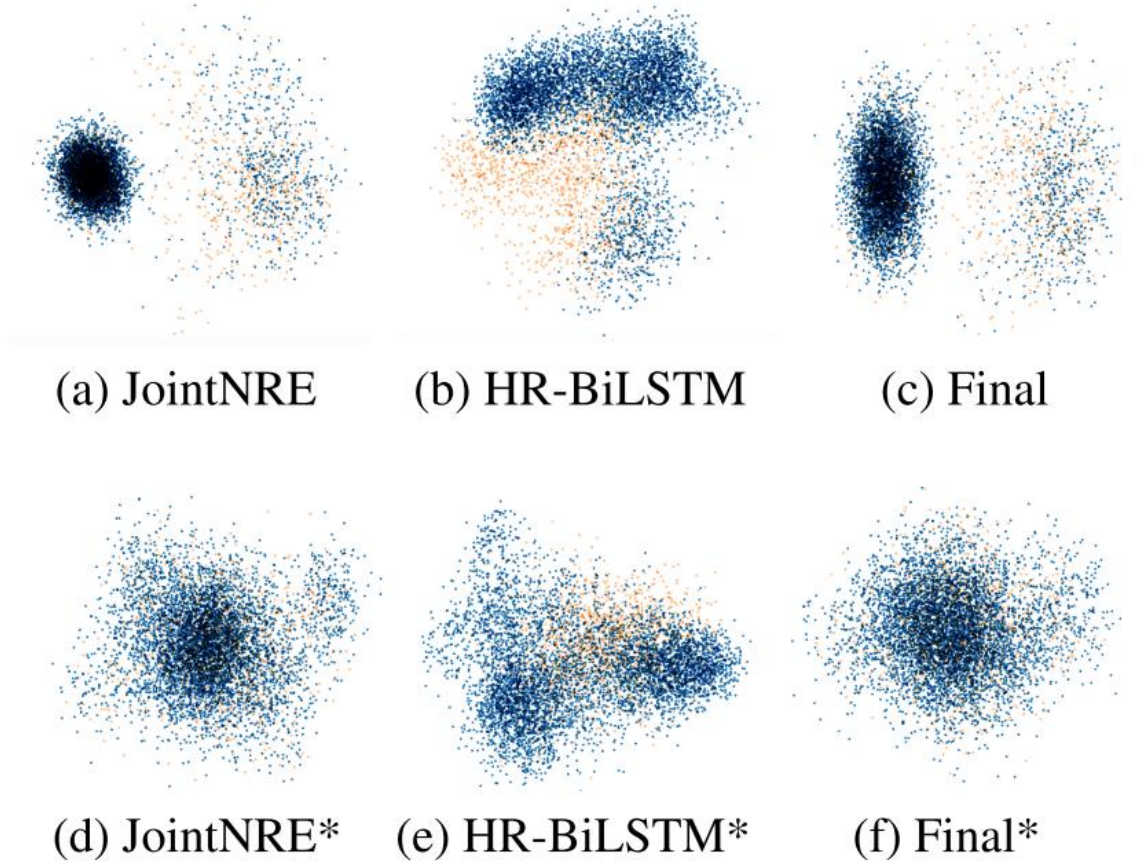
Model	Seen Rate ↓ (%)
HR-BiLSTM	47.2±2.0
+ no fine-tune	34.8±2.3
Final	**21.2±1.7**

- Influence of Number of Relations for Training



Analysis

- Yellow： seen relation
- blue： unseen relation



Analysis

Question 1	who produced recording *Twenty One*
Candidate Relations	**music.recording.producer** music.recording.artist
HR-BiLSTM	music.recording.artist
Final	**music.recording.producer**
Question 2	what is *Tetsuo Ichikawa*'s profession
Candidate Relations	people.person.gender **people.person.profession**
HR-BiLSTM	**people.person.profession**
Final	**people.person.profession**
Question 3	which village is in *Arenac county* ?
Candidate Relations	location.us_county.hud_county_place **location.location.contains**
HR-BiLSTM	location.us_county.hud_county_place
Final	location.us_county.hud_county_place

Conclusion

- We emphasize that for any other tasks which contain a large number of unseen samples, training, fine-tuning the model according to the performance on the seen samples alone is not fair.