Multi-Task Learning for Relation Extraction

Kai Zhou[†], Xiangfeng Luo[†‡*], Hao Wang[†‡], Richard Xu[§]
[‡]Shanghai Institute for Advanced Communication and Data Science
[†]School of Computer Engineering and Science, Shanghai University, Shanghai, China
[§]School of Electrical and Data Engineering, University of Technology Sydney, Ultimo, Australia
{iakzh, Luoxf, wang-hao}@shu.edu.cn, YiDa.Xu@uts.edu.au

Relation Extraction

- **Definition:** RE aims to extract relations between pairs of marked entities in raw texts.
- **Example:** Shanghai University was established in Shanghai in 1922.

→ Extracted relation triples:

  - (Shanghai University, *Founding-year*, 1922)
  - (Shanghai University, *Founding-location*, Shanghai)

- **Applications:** Question answering, Web search, Knowledge Base Population...

Distantly-Supervised Relation Extraction

- **Definition**

"Any sentence that contains a pair of entities that participate in a known Freebase relationship is likely to express that relation in some way."

Train a relation classifier using:

✓ knowledge bases (e.g., Freebase, or the structured information in Wikipedia info-boxes) with a lot of relation triples

✓ unstructured text containing entities

- **Feature**

DSRE is an alternative paradigm combining advantages of both the supervised and unsupervised approaches:

✓ using a probabilistic classifier

✓ large amount of training data

Figure 1 : Wikipedia info-box about Shanghai University

Wrong labeling problem

The distant supervision assumption might lead to wrong labels in the following situations:

- Entities have multiple valid relations in KG

  Trump is the President of United States.

  Trump was born in United States.

- Empty relation between entities in KG.

  About 80 percent of sentences in the common used dataset Riedel2010.

Does

Relation Extraction = Relation Classification

Fit

In

Distant Supervision ?

Motivations

- Semantic and syntactic information like dependency paths and entity types have been found helpful for RE [Mintz et al. 2009] [Zhang et al., 2018] [Vashishth et al., 2018].

- Employing an related task like entity type classification [Liu et al., 2018] in pre-training may improve model's robustness against noisy relation labels in the training procedure.

So, we propose a novel hierarchical model in a multi-task learning fashion for relation extraction.
We choose dependency parsing and entity type classification as auxiliary tasks.

Relation Extraction = Entity Type Classification + Relation Classification + Dependency Parsing

Our methodology

(1)Multi-task Learning Framework

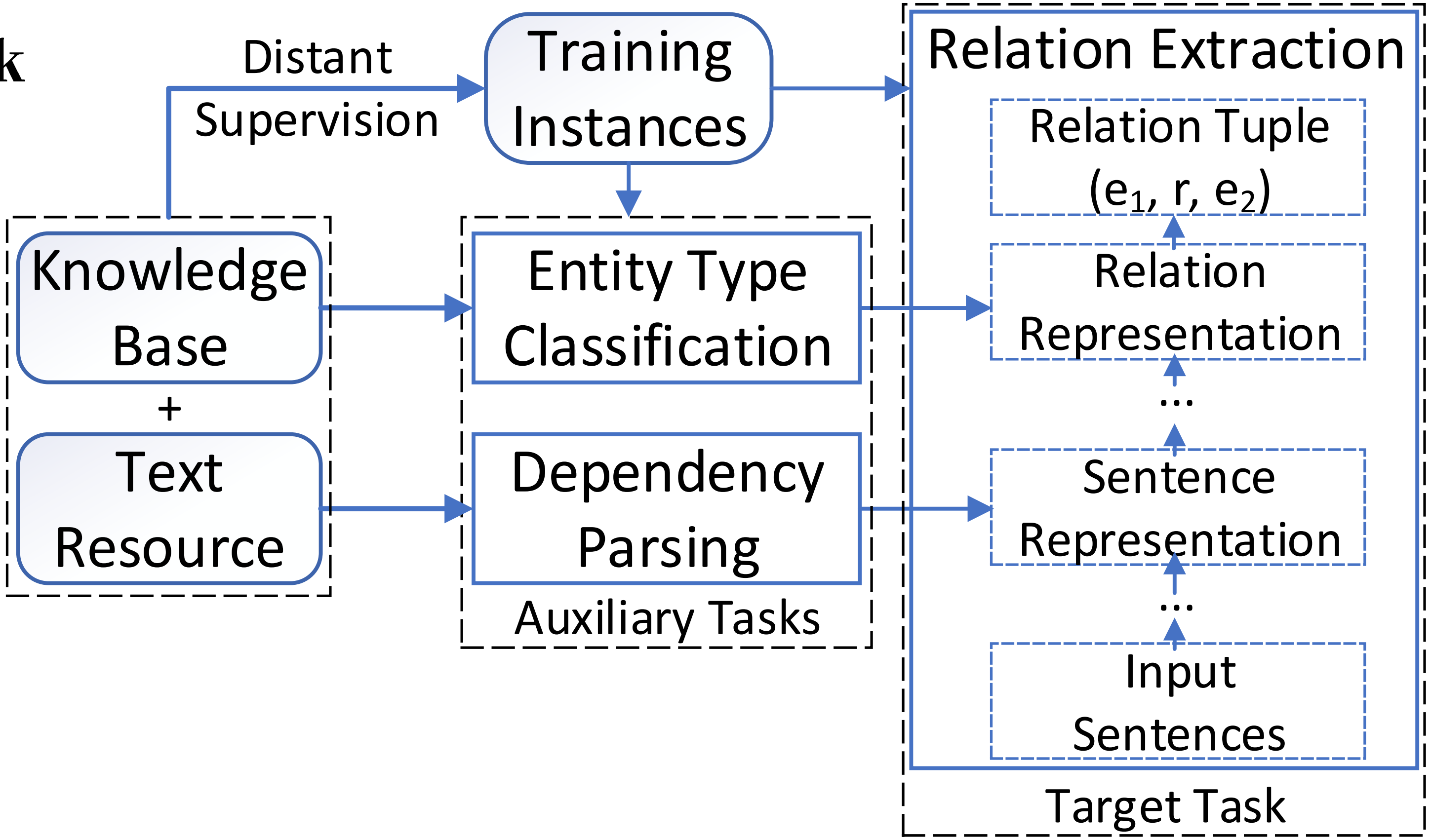
- **Target Task:**

  Relation Extraction

- **Auxiliary Task:**

  ① Entity Type Classification

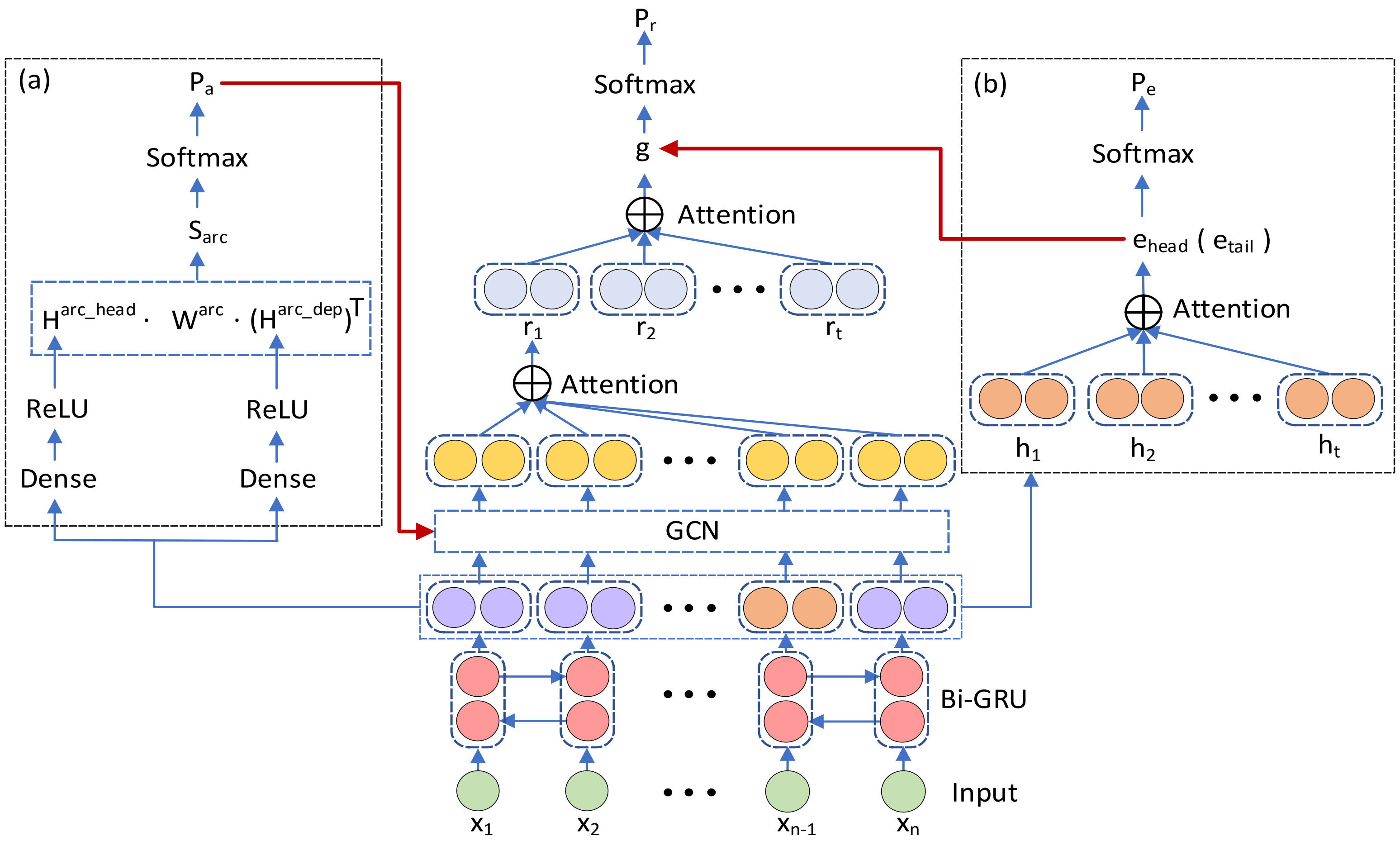
  ② Dependency Parsing



Our methodology

(2)Neural Network Model

- **Pre-training (simultaneously):**
  - Dependency Parsing
  - Entity Type Classification
- **Training (simultaneously):**
  - Dependency Parsing
  - Entity Type Classification
  - Relation Extraction
- **Testing:**
  - Relation Extraction



(a) Dependency parsing ;(b)Entity type classification

Experiments and Results

Dataset : Riedel NYT [Riedel et al., 2010]

- Contain text from the New York Times Annotated Corpus with named entities extracted from the text using the Stanford NER system and automatically linked to entities in the Freebase knowledge base. Pairs of named entities are labelled with relationship types by aligning them against facts in the Freebase knowledge base (distant supervision) .
- We employ Natural Language Toolkit (NLTK) [Steven Bird et al., 2009] to label sentences' dependency arcs.
- We label the entity types using the Fine-Grained Entity Recognizer (FIGER) [Xiao Ling and Daniel S. Weld, 2012] .

Dataset	Split	Sentences	Entity-pairs
Riedel (# Relations:53)	Train Test	570088 172448	291699 96678

Table 1: Details of dataset used.

Experiments and Results

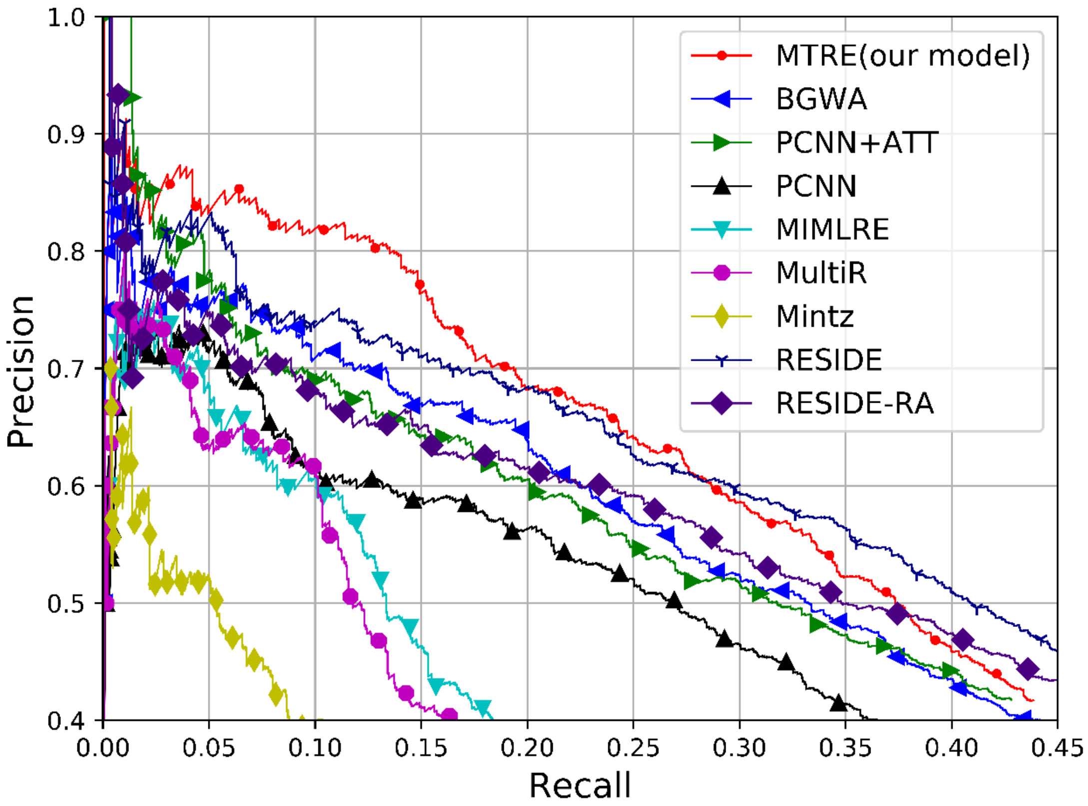


Figure 2: Comparison of Precision-recall curve.

Word embedding	50
Position embedding	5
GRU size	180
Batchsize	200
Dense layer size	256
Projection layer size	128
Loss weight λ_1	0.7
Loss weight λ_2	0.7
GCN size	360
Dropout rate	0.8
Learning rate	0.001
l2 regularizaion	0.0001

Table 2: Details of hyperparameters used.

Ablation Study

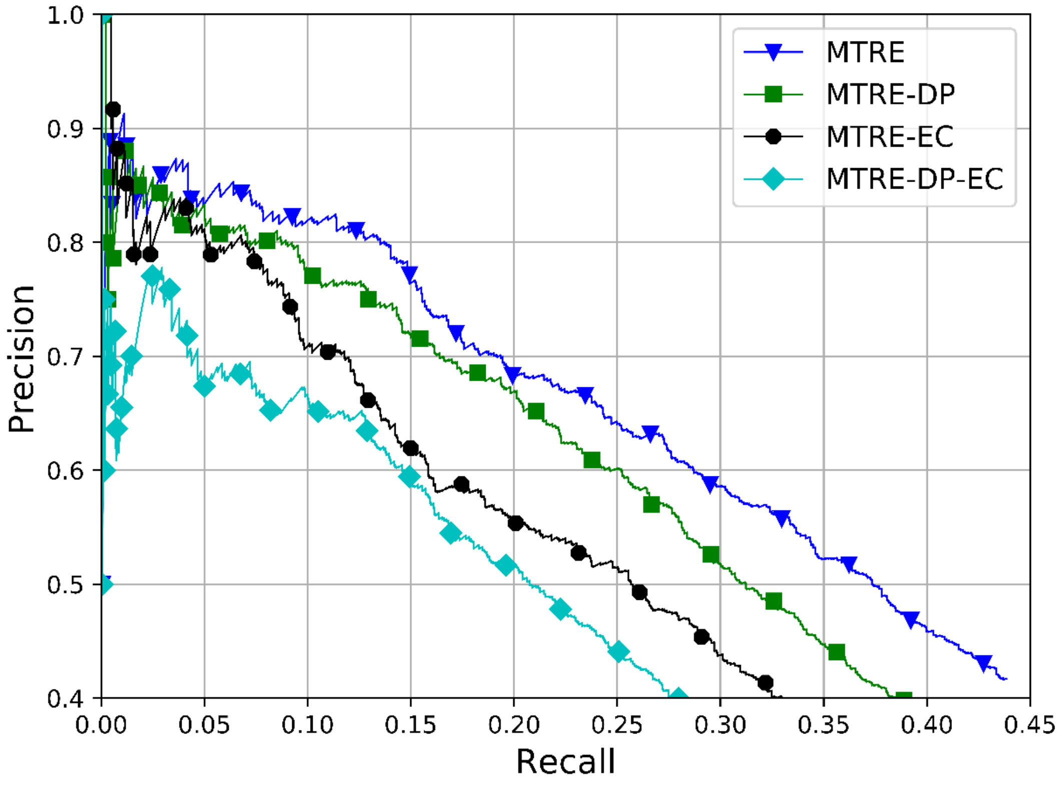


Figure 3: Comparison of different sub-models w/o auxiliary tasks.

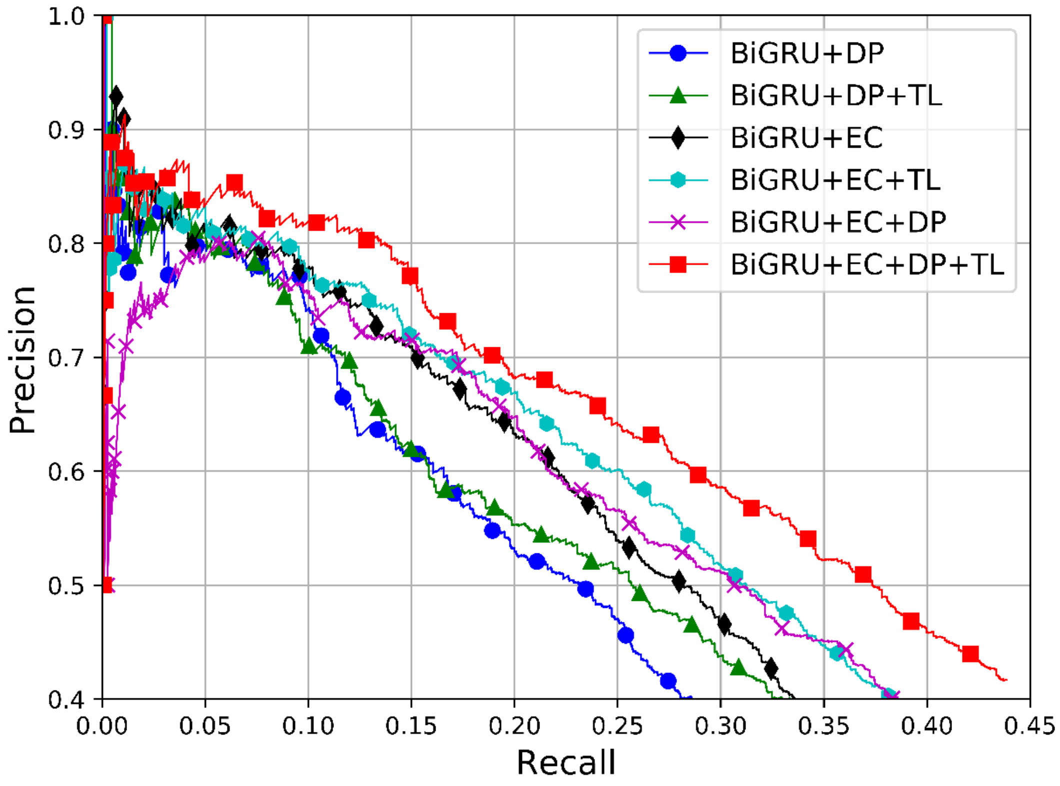


Figure 4: Comparison of different sub-models w/o pretraining.

Conclusion

- Under the multi-task learning framework, we effectively take advantage of related tasks to improve the precision in relation extraction. Leverage supervised signals from related tasks as prior knowledge makes our model more robust against noisy relation labels generated from distant supervision.

- In the future, we would explore more kinds of practical auxiliary tasks and optimization techniques between them because different tasks might conflict with each other and undermine the overall performance.

Thank you for your listening!

Author Email: iakzh@shu.edu.cn

Reference

- M. Mintz, S. Bills, R. Snow, and D. Jurafsky, “Distant supervision for relation extraction without labeled data,” in Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2. Association for Computational Linguistics, 2009, pp. 1003–1011.
- Y. Zhang, P. Qi, and C. D. Manning, “Graph convolution over pruned dependency trees improves relation extraction,” in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 2205–2215.
- S. Vashishth, R. Joshi, S. S. Prayaga, C. Bhattacharyya, and P. Talukdar, “Reside: Improving distantly-supervised neural relation extraction using side information,” arXiv preprint arXiv:1812.04361, 2018.
- T. Liu, X. Zhang, W. Zhou, and W. Jia, “Neural relation extraction via inner-sentence noise reduction and transfer learning,” in EMNLP, 2018.
- S. Riedel, L. Yao, and A. McCallum, “Modeling relations and their mentions without labeled text,” in Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, 2010, pp.148–163.
- Xiao Ling and Daniel S. Weld."Fine-Grained Entity Recognition",in Proceedings of the twenty-sixth aaai conference on artificial intelligence (AAAI), 2012.
- Steven Bird, Ewan Klein, and Edward Loper (2009). Natural Language Processing with Python. O’Reilly Media Inc