Improving Distantly-Supervised Neural Relation Extraction using Side Information

Shikhar Vashishth, Rishabh Joshi, Sai S. Prayaga,
Chiranjib Bhattacharyya, and Partha Talukdar

Indian Institute of Science, Bangalore

Relation Extraction (RE)

- **Definition:** Task of identifying relation between entities in text.

- **Example:** Google was founded in the state of California in 1998.
  - *Founding-year* (Google, 1998)
  - *Founding-location* (Google, California)

- **Applications:** Question answering, Web search, KB Population...

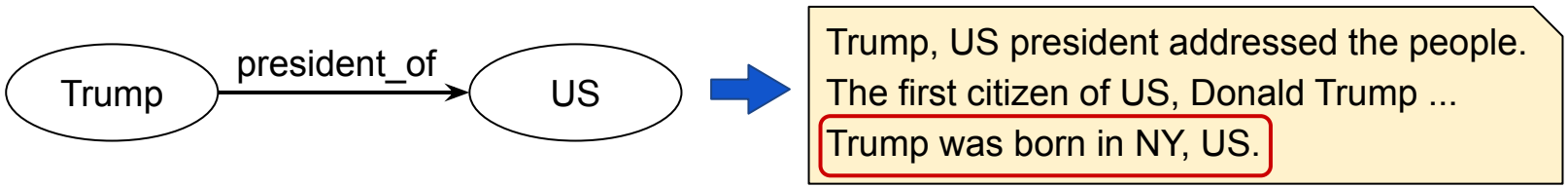
Traditional RE

- **Hand-built** patterns and **Bootstrapping** method
  - Hearst patterns, Snowball (2000) …
- **Supervised** approaches:
  - **Relation detection:** True/False
  - **Relation classification:** employee-of/located-in/...
- **Disadvantages:**
  - Suffers from lack of annotated data

Distant Supervision (DS)

- Alleviates the problem of lack of annotated data.

- **Distant Supervision (DS) assumption:** [Mintz et al., 2009]
  "If two entities have a relationship in a KB, then all sentences mentioning the entities express the same relation"



Multi-instance Multi-label (MIML)

- DS might lead to **Noisy labelled data**

- **MIML:** Relaxed DS assumption [Riedel et al. 2010; Surdeanu et al., 2012]
  - **Does not** transform DS to **traditional supervised RE (sentence-level)**.
  - Allows **multiple relations** to hold **between entities**
  - If a relation holds between entities then **at least one sentence** must support it

Neural Networks for DS

- Earlier techniques utilized **hand-crafted NLP features** which can be **expensive and limiting.**
- Using **NN for feature extraction** helps.
- **PCNN (Piecewise Convolution NN)** [Zeng et a., 2014]
  - Adapts **CNNs** for extracting **sentence features.**
- **PCNN+Attention** [Lin et al., 2016]
  - Uses **attention mechanism** for **aggregating** information.

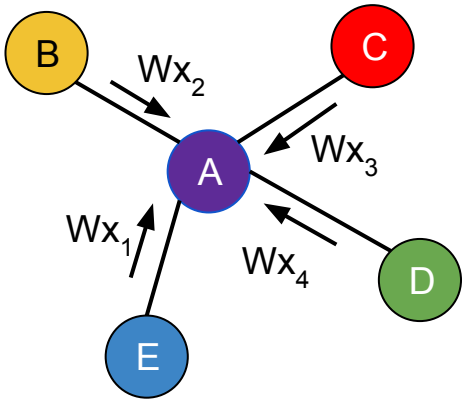
Motivation

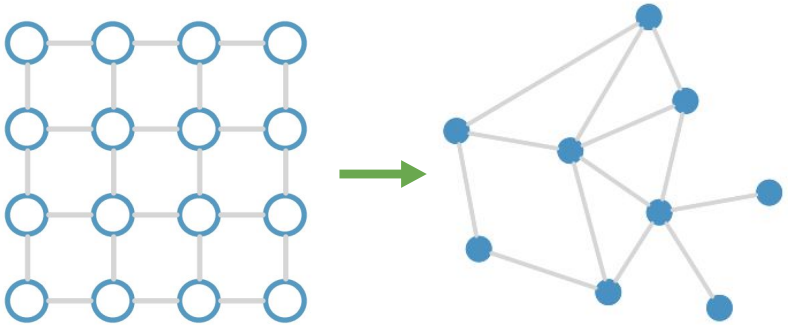
- **KGs** contain **information** which can **improve RE**
  - **Limiting supervision** from KG **to dataset creation**

- **Dependency tree based features** have been found **relevant** for RE [Mintz et al. 2009]
  - Instead of defining **hand-crafted features** can employ **Graph Convolutional Networks (GCNs)**.

Graph Convolutional Networks (GCN)

- **Generalization of CNNs over Graphs.**





$$h_v = f\left(\sum_{u \in \mathcal{N}(v)} (Wx_u + b)\right), \quad \forall v \in \mathcal{V}$$

[Kipf et al., 2016]

Contributions

- Propose **RESIDE**, a novel method which **utilizes additional supervision from KB** in a principled manner for improving distant supervised RE.

- RESIDE uses **GCNs for modeling syntactic information** and performs competitively even with limited side information.

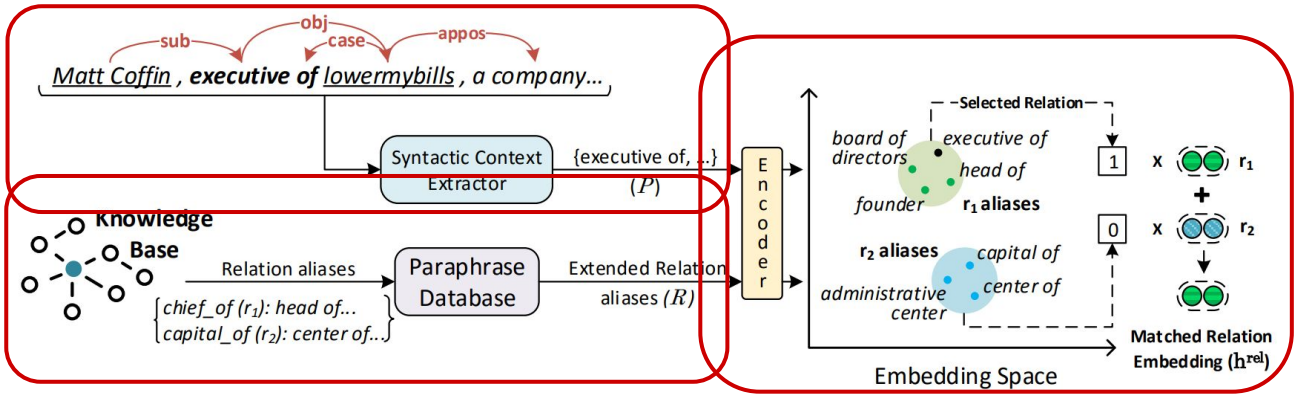
RESIDE

RESIDE: Side Information

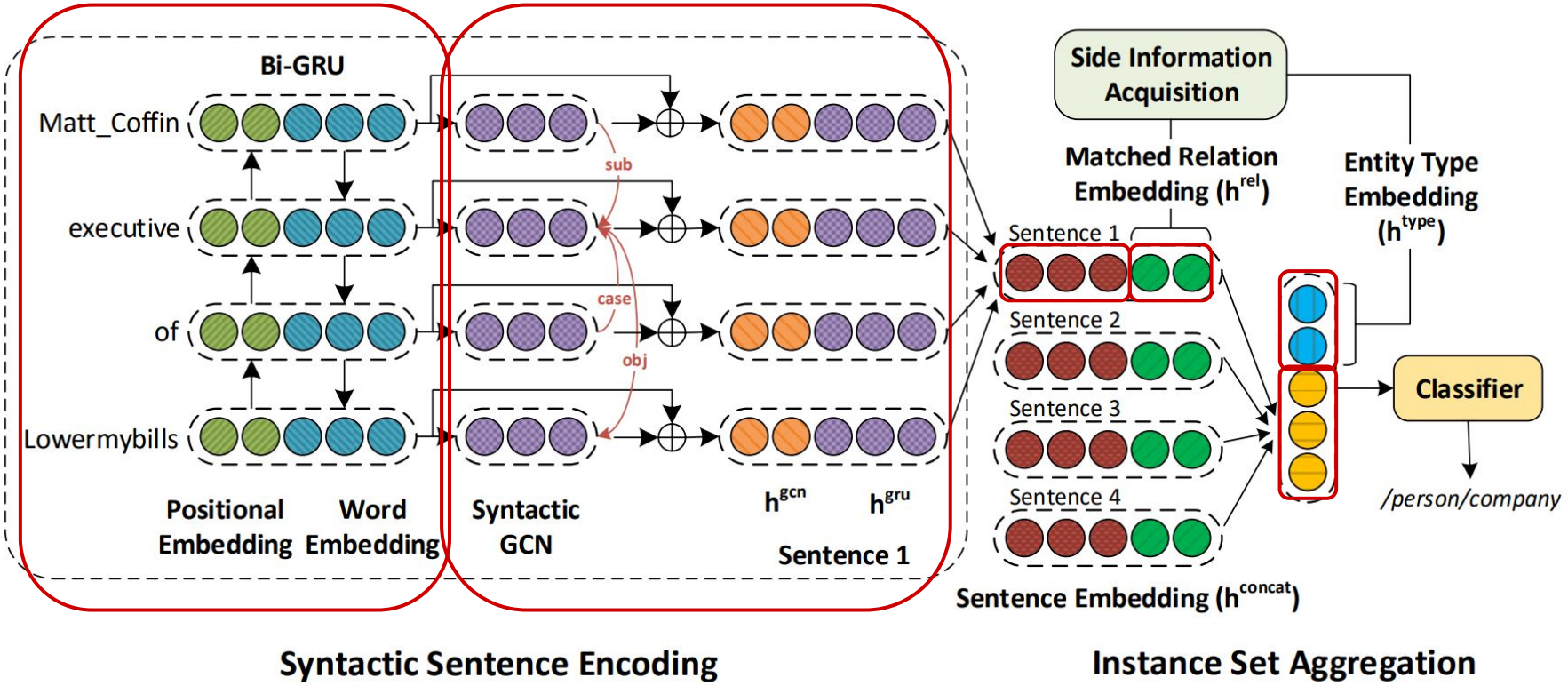
- **Entity Type Information:**
  - All **relations** are **constrained** by the **entity types**
  - ***president_of****(X, Y)* **=>** X = ***Person*** Y = ***Country***
  - Several **KGs** like **Freebase**, **Wikidata** provide entity type info.
  - **Not suitable** as **hard constraints**
    - Not available at **right granularity** or are **noisy**
    - Requires **manual effort** for **each relation**

RESIDE: Side Information

- **Relation Alias Information:**
  - **Utilize relation aliases** provided by KGs.
  - Extract **relation phrases** between **target entities** using **OpenIE** and **dependency parse**
  - **Link extracted phrases** to closest **relations** using **aliases** from KG



RESIDE: Architecture

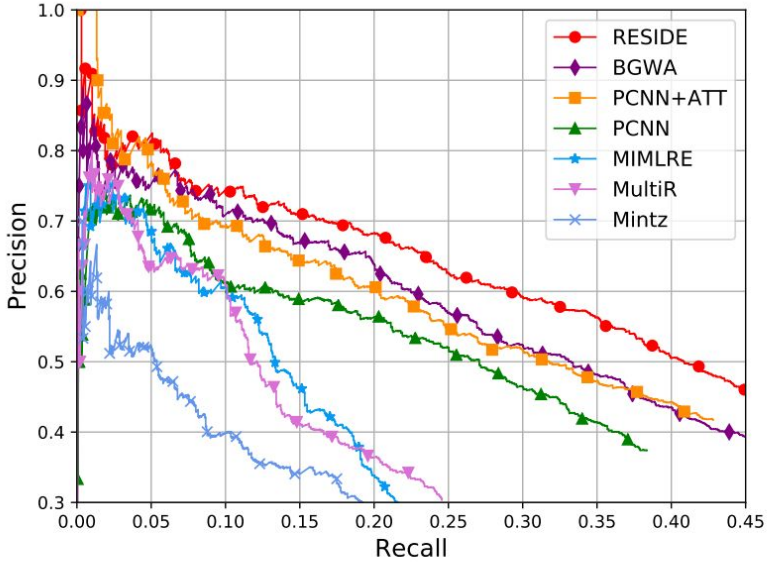


Experiments

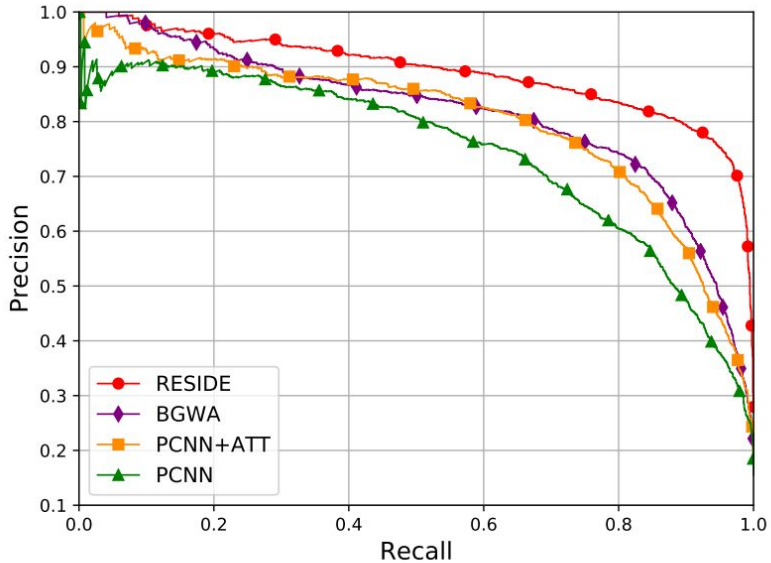
- **Datasets:**
  - **RiedelNYT:** [Riedel et al., 2010]
    - Constructed by aligning Freebase relations with NYT corpus
    - Contains 53 relations
  - **GIDS (Google IISc Distant Supervision):** [Jat et al., 2018]
    - Constructed by extending Google RE corpus
    - Assures at-least-one assumption of MIML paradigm

Results:

Comparison of Precision Recall curves



(a) Riedel dataset

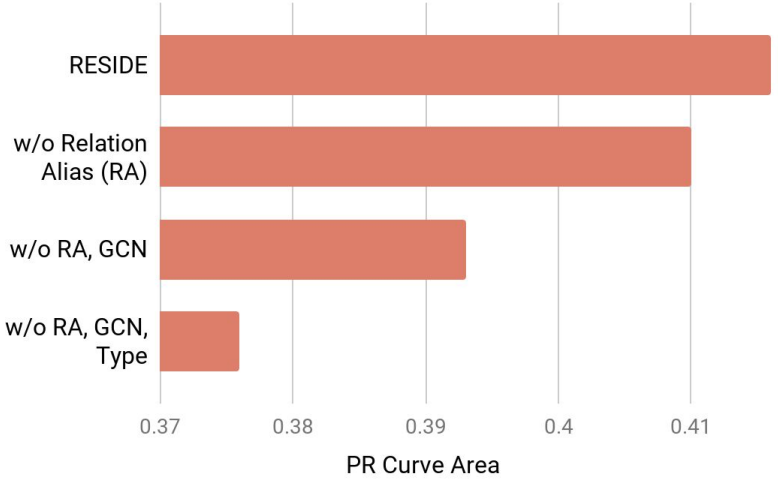


(b) GIDS dataset

RESIDE achieves **higher precision** over the **entire recall range**.

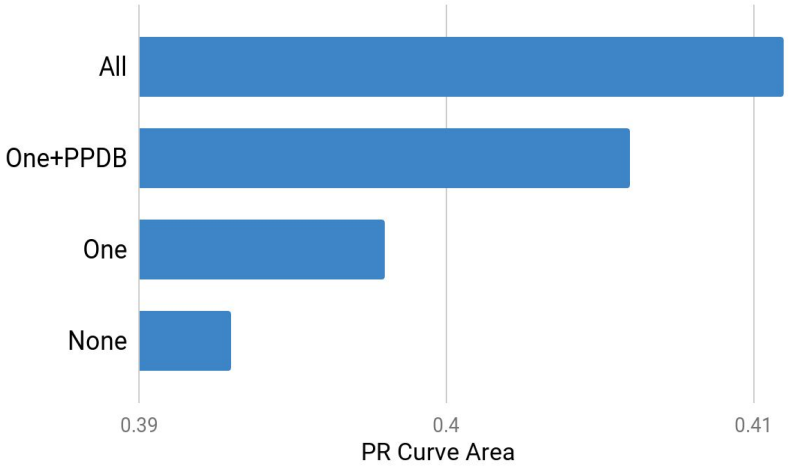
Ablation Study:

- **Comparison of different ablated version of RESIDE**
  - **Cumulatively removing** different side information
  - **Side information helps** improve performance.



Effect of Relation Alias Information

- **Performance on different settings:**
  - **None:** Relation aliases not available
  - **One:** Name of relations used as aliases
  - **One+PPDB:** Relation names extended using Paraphrase Database
  - **All:** Relation aliases from KG



RESIDE performs **comparable** with **limited side information.**

Conclusion

- **Additional supervision** from KG **improves Relation Extraction**
- **Syntactic features** can be efficiently utilized **using Graph Convolutional Networks (GCNs)**
- We propose **RESIDE**, a **novel NN based model** which utilizes the above two ideas for **outperforming existing methods.**

Questions?

Contact email:
shikhar@iisc.ac.in

Source code and data are available
github.com/malllabiisc/RESIDE

- **References:**

  - Vashishth, Shikhar, et al. "Reside: Improving distantly-supervised neural relation extraction using side information." *arXiv preprint arXiv:1812.04361* (2018). https://arxiv.org/abs/1812.04361
  - Lin, Yankai, et al. "Neural relation extraction with selective attention over instances." *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vol. 1. 2016. https://www.aclweb.org/anthology/P16-1200
  - Vashishth, Shikhar, Prince Jain, and Partha Talukdar. "CESI: Canonicalizing Open Knowledge Bases using Embeddings and Side Information." *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2018. https://arxiv.org/abs/1902.00172
  - Jat, Sharmistha, Siddhesh Khandelwal, and Partha Talukdar. "Improving distantly supervised relation extraction using word and entity based attention." *arXiv preprint arXiv:1804.06987*(2018). https://arxiv.org/abs/1804.06987