南京大学万维网软件研究组
The Websoft Research Group, Nanjing University, China

Learning to Exploit Long-term Relational Dependencies in Knowledge Graphs

Lingbing Guo, Zequn Sun, Wei Hu*

Nanjing University, China

* Corresponding author: whu@nju.edu.cn

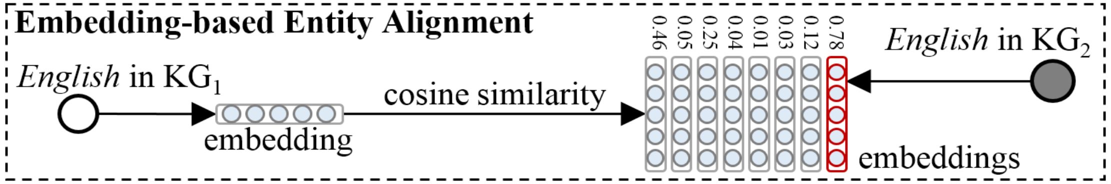
ICML'19, June 9–15, Long Beach, CA, USA

Knowledge graphs

- **Knowledge graphs** (KGs) store a wealth of structured facts about the real world
  - A fact (s, r, o): subject entity, relation, object entity
- KGs are far from complete and two important tasks are proposed

Knowledge graphs

- **Knowledge graphs** (KGs) store a wealth of structured facts about the real world
  - A fact (s, r, o): subject entity, relation, object entity
- KGs are far from complete and two important tasks are proposed
  1. **Entity alignment:** find entities in **different KGs** denoting the same real-world object



  2. **KG completion:** complete missing facts in a **single KG**
     - E.g., predict **?** in (*Tim Berners-Lee*, *employer*, **?**) or (**?**, *employer*, *W3C*)

Challenges

- For KG embedding, existing methods largely focus on learning from **relational triples** of entities

- Triple-level learning has two major limitations
  - **Low expressiveness**
    - Learn entity embeddings from a fairly local view (i.e., 1-hop neighbors)
  - **Inefficient information propagation**
    - Only use triples to deliver semantic information within/across KGs

Learning to exploit long-term relational dependencies

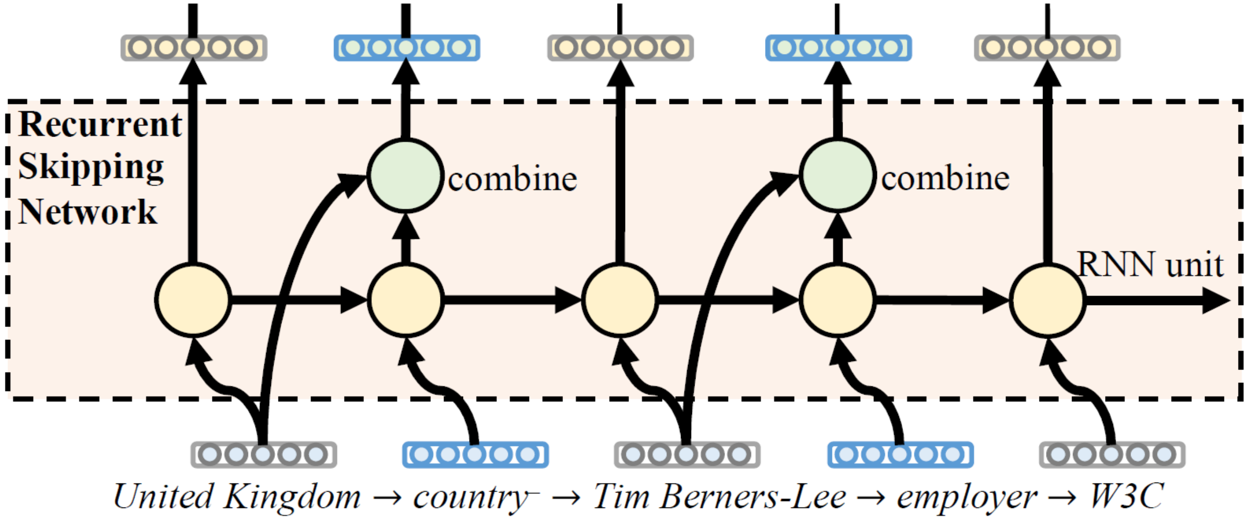
- A relational path is an **entity-relation chain**, where entities and relations appear alternately

$$United\ Kingdom \rightarrow country^{-} \rightarrow Tim\ Berners\text{-}Lee \rightarrow employer \rightarrow W3C$$

- RNNs perform well on sequential data
  - **Limitations** to leverage RNNs to model relational paths
    1. A relational path have two different types: “entity” and “relation”
       - Always appear in an alternating order
    2. A relational path is constituted by triples, but these basic structure units are overlooked by RNNs

Recurrent skipping networks

- A conditional skipping mechanism allows RSNs to **shortcut** the current input entity to let it **directly** participate in predicting its object entity



Tri-gram residual learning

- Residual learning
  - Let $F(\boldsymbol{x})$ be an original mapping, and $H(\boldsymbol{x})$ be the expected mapping
  - Compared to directly optimizing $F(\boldsymbol{x})$ to fit $H(\boldsymbol{x})$,
    it is easier to optimize $F(\boldsymbol{x})$ to fit residual part $H(\boldsymbol{x})$
    - An extreme case, $H(\boldsymbol{x}) = \boldsymbol{x}$

Tri-gram residual learning

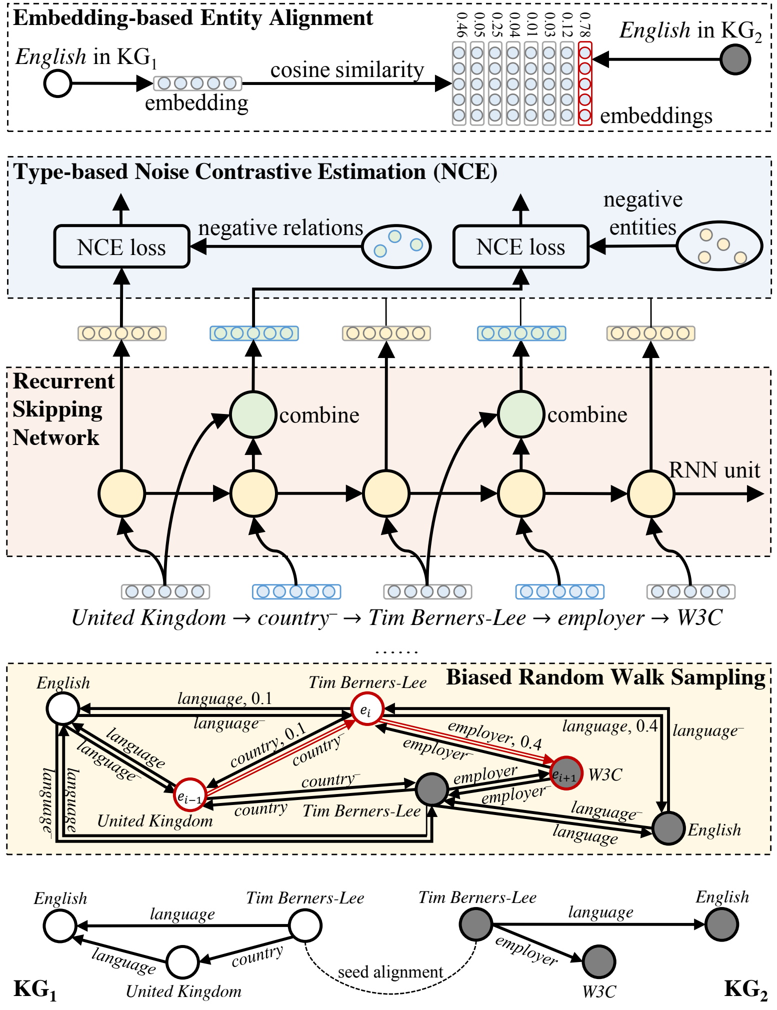
- Residual learning
  - Let $F(\boldsymbol{x})$ be an original mapping, and $H(\boldsymbol{x})$ be the expected mapping
  - Compared to directly optimizing $F(\boldsymbol{x})$ to fit $H(\boldsymbol{x})$, it is easier to optimize $F(\boldsymbol{x})$ to fit residual part $H(\boldsymbol{x})$
    - An extreme case, $H(\boldsymbol{x}) = \boldsymbol{x}$

($United\ Kingdom,\ country^{-},\ Tim\ Berners\text{-}Lee,\ employer,\ W3C$)	
Models	Optimize $F([\cdot], employer)$ as
RNNs	$F([\cdot], employer) \coloneqq W3C$
RRNs	$F([\cdot], employer) \coloneqq W3C - [\cdot]$
RSNs	$F([\cdot], employer) \coloneqq W3C - Tim\ Berners{-}Lee$
$[\cdot]$ denotes context ($United\ Kingdom,\ country^{-},\ Tim\ Berners\text{-}Lee$)	

- **Tri-gram** residual learning
  - $United\ Kingdom \rightarrow country^{-} \rightarrow Tim\ Berners\text{-}Lee \rightarrow employer \rightarrow W3C$
  - Compared to directly learning to predict *W3C* by *employer* and its mixed context, it is easier to learn the residual part between *W3C* and *Tim Berners-Lee*
    - Because they forms a triple, and we should not overlook the triple structure in the paths

Architecture

- An **end-to-end** framework
  1. **Biased random walk sampling**
     - Deep paths carry more relational dependencies than triples
     - Cross-KG paths deliver alignment information between KGs
  2. **Recurrent skipping network**
  3. **Type-based noise contrastive estimation**
     - Evaluate loss in an optimized way



Experiments and results

- Entity alignment results
  - Datasets: normal & dense
  - Performed **best** on all datasets
    - Especially on the normal datasets

Hits@1	DBP-WD	DBP-YG	EN-FR	EN-DE
MTransE	22.3	24.6	25.1	31.2
IPTransE	23.1	22.7	25.5	31.3
JAPE	21.9	23.3	25.6	32.0
BootEA	32.3	31.3	31.3	44.2
GCN-Align	17.7	19.3	15.5	25.3
TransR	5.2	2.9	3.6	5.2
TransD	27.7	17.3	21.1	24.4
ConvE	5.7	11.3	9.4	0.8
RotatE	17.2	15.9	14.5	31.9
RSNs (w/o biases)	**37.2**	**36.5**	**32.4**	**45.7**
RSNs	**38.8**	**40.0**	**34.7**	**48.7**

Experiments and results

- Entity alignment results
  - Datasets: normal & dense
  - Performed **best** on all datasets
    - Especially on the normal datasets

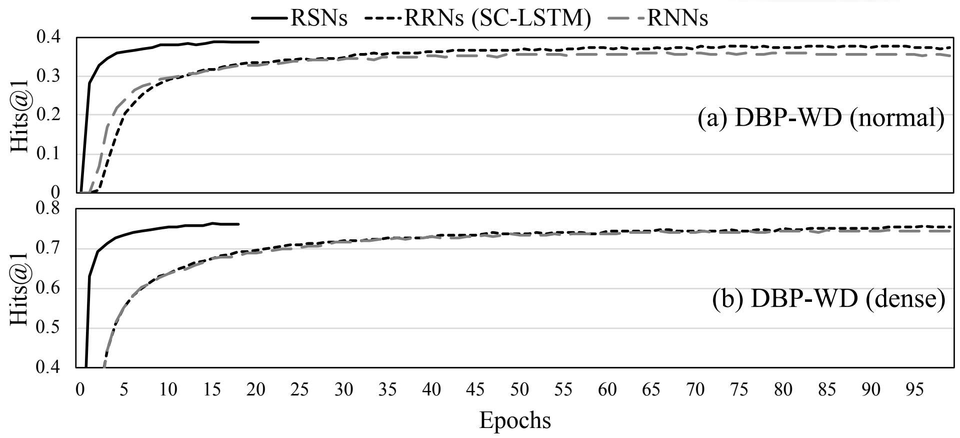
Hits@1	DBP-WD	DBP-YG	EN-FR	EN-DE
MTransE	22.3	24.6	25.1	31.2
IPTransE	23.1	22.7	25.5	31.3
JAPE	21.9	23.3	25.6	32.0
BootEA	32.3	31.3	31.3	44.2
GCN-Align	17.7	19.3	15.5	25.3
TransR	5.2	2.9	3.6	5.2
TransD	27.7	17.3	21.1	24.4
ConvE	5.7	11.3	9.4	0.8
RotatE	17.2	15.9	14.5	31.9
RSNs (w/o biases)	**37.2**	**36.5**	**32.4**	**45.7**
RSNs	**38.8**	**40.0**	**34.7**	**48.7**

- KG completion results
  - Datasets: FB15K, WN18
  - Obtained **comparable** performance
    - Better than all translational models

FB15K	Hits@1	Hits@10	MRR
TransE	30.5	73.7	0.46
TransR	37.7	76.7	0.52
TransD	31.5	69.1	0.44
ComplEx	59.9	84.0	0.69
ConvE	67.0	**87.3**	0.75
RotatE	**74.6**	**88.4**	**0.80**
RSNs (w/o cross-KG biase)	**72.2**	**87.3**	**0.78**

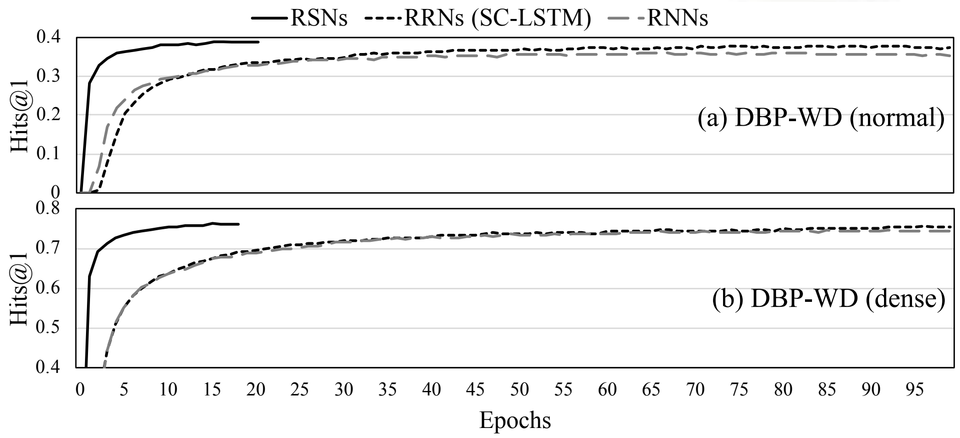
Further analysis

- RSNs vs. RNNs, RRNs [recurrent residual networks]
  - Achieved **better** results with only **1/30** epochs

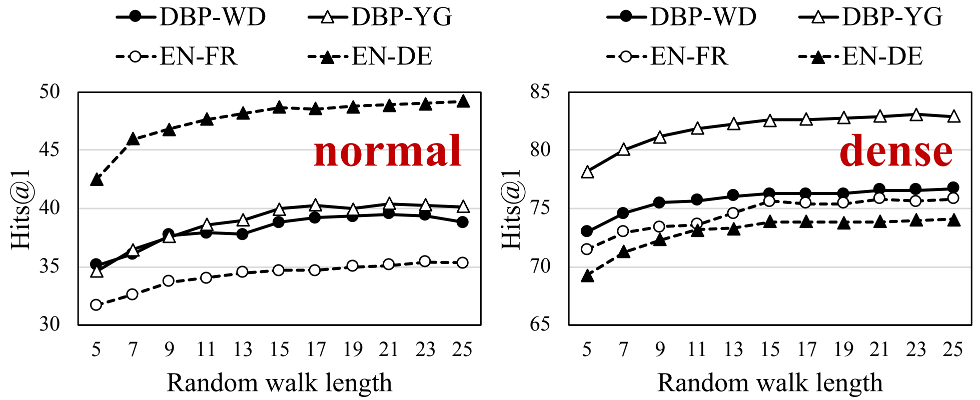


Further analysis

- RSNs vs. RNNs, RRNs [recurrent residual networks]
  - Achieved **better** results with only **1/30** epochs



- Random walk length
  - On all the datasets, increased steadily from length 5 to **15**



Conclusion

- We studied **path-level** KG embedding learning
  1. **RSNs:** sequence models to learn relational paths
  2. **End-to-end framework:** biased random walk sampling + RSNs
  3. **Superior** in entity alignment and **competitive** in KG completion

- Future work
  - **Unified sequence model:** relational paths & textual information

Poster: Tonight, Pacific Ballroom #42

Datasets & source code: https://github.com/nju-websoft/RSN

Acknowledgements:

- National Key R&D Program of China (No. 2018YFB1004300)
- National Natural Science Foundation of China (No. 61872172)
- Key R&D Program of Jiangsu Science and Technology Department (No. BE2018131)