Question Answering by Reasoning Across Documents with Graph Convolutional Networks

Nicola De Cao[1,2], Wilker Aziz[2], and Ivan Titov[1,2]

[1]University of Edinburgh, [2]University of Amsterdam

University of Amsterdam

The University of Edinburgh

Key Ideas

- We do **reading comprehension** on **multiple documents**
- We frame it as an **inference problem on a graph:**
    - mentions of entities are **nodes**
    - **edges** encode relations between different mentions
- **Graph convolutional networks** performs multi-step reasoning
- Our method is scalable and compact, and it achieves **state-of-the-art** results on WikiHop (Welbl et al., 2018)

```
Thorildsplan is a small park in Kristineberg in
Stockholm, named in 1925 after the writer [..]

Stockholm is the capital of Sweden
and the most populous city in [..]

query: country Thorildsplan
candidates: {Denmark, Finland, Sweden, Italy, ...}
answer: Sweden
```

Figure 1. A sample from WikiHop.

Introduction

Data:

- Set of **<documents, question>** pairs
- Automatically constructed from a text corpus (*Wikipedia*) and a knowledge base (*Wikidata*)
- Questions are constructed to encourage **reasoning across documents**

Task:

- multiple choice from a set of **candidate answers** (C)

Approach:

1. explicitly **build a graph** of useful mentions from supporting documents (S) used for reasoning steps
2. **graph convolutional networks (GCN)** are applied to these graphs allowing for information propagation
3. **candidates are scored** through an MLP conditioned on a question representation and a pooling operation

Method

1. Building the graph

Nodes are **mentions** from all supporting documents and are connected with different edge-types/relations:

- Mentions within the **same document** (DOC-BASED)
- Exact matches **across documents** (MATCH)
- **Coreferences** — using an external system (COREF)
- The complement graph (COMPLEMENT)

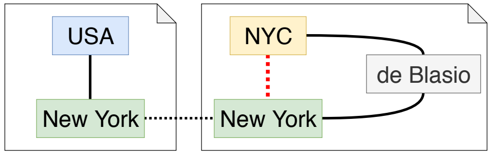

Figure 2. Two supporting documents.

Node embeddings are initialised using ELMo (Peters et al., 2018), a pre-trained language model, and a question representation.

NOTE: no further readers are applied to the text!

2. Entity Relational Graph Convolutional Network

Gated version of **relational-GCNs** (Schlichtkrull et al., 2018).

$$\mathbf{u}_i^{(\ell)} = f_s(\mathbf{h}_i^{(\ell)}) + \frac{1}{|N_i|} \sum_{j \in N_i} \sum_{r \in R_{ij}} f_r(\mathbf{h}_j^{(\ell)}) \;\Big\}$$

The update vector ($\mathbf{u}$) of a node is a function of its neighbours (N) conditioned on the relations between them

$$\mathbf{a}_i^{(\ell)} = \sigma\left(f_a\left([\mathbf{u}_i^{(\ell)}, \mathbf{h}_i^{(\ell)}]\right)\right) \;\Big\}$$

attention gate

$$\mathbf{h}_i^{(\ell+1)} = \phi(\mathbf{u}_i^{(\ell)}) \odot \mathbf{a}_i^{(\ell)} + \mathbf{h}_i^{(\ell)} \odot (1 - \mathbf{a}_i^{(\ell)}) \;\Big\}$$

the new node embedding

3. Candidates scoring

We use the final node embeddings ($\mathbf{h}^{(L)}$) and the question representation ($\mathbf{q}$) to **predict a distribution over candidates.**

$$P(c \mid q, C_q, S_q) \propto \exp\left(\max_{i \in M_c} f_o([\mathbf{q}, \mathbf{h}_i^{(L)}])\right) \;\Big\}$$

A **max-pooling** operation among the mentions (M_c) of the same candidate (c)

Results

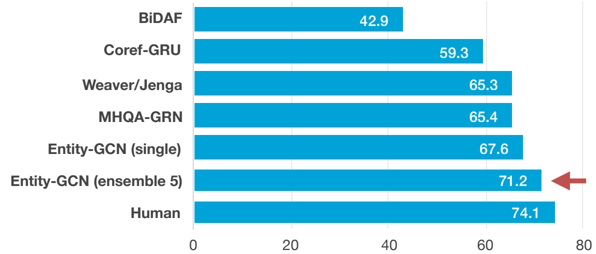

Figure 1. Accuracy on WikiHop closed test set.

Entity-GCN is at least **5 times faster to train than BiDAF**

NOTE: ensemble models add negligible overhead since embeddings are computed only once!

Ablation study

Model	
Full (ensemble)	**68.5**
Full (single)	**65.1±0.11**
GloVe w/o R-GCN	51.2
GloVe w/ R-GCN	59.2
No relation type	62.7
No DOC-BASED	62.9
No MATCH	64.3
No COREF	64.8
No COMPLEMENT	64.1
Induced edges	61.5

Figure 2. Accuracy on WikiHop validation set.

R-GCN is useful

Relations are important

DOC-BASED relations are the most significant

Learning to **predict edges** is hard and an open problem

Contact Information

Nicola De Cao

Ph.D. Candidate at University of Amsterdam

nicola.decao@gmail.com

https://nicola-decao.github.io

https://twitter.com/nicola_decao

References

- Welbl, J., Stenetorp, P., and Riedel, S. Constructing datasets for multi-hop reading comprehension across documents. Transactions of the Association for Computational Linguistics , 6:287–302, 2018
- Schlichtkrull, M., Kipf, T. N., Bloem, P., van den Berg, R., Titov, I., and Welling, M. Modeling relational data with graph convolutional networks. In Gangemi, A., Navigli, R., Vidal, M.-E., Hitzler, P., Troncy, R., Hollink, L., Tordai, A., and Alam, M. (eds.), The Semantic Web, pp. 593–607, Cham, 2018. Springer International Publishing. ISBN 978-3-319-93417-4.
- Dhingra, B., Jin, Q., Yang, Z., Cohen, W., and Salakhutdinov, R. Neural models for reasoning over multiple mentions using coreference. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), pp. 42–48, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.
- Raison, M., Mazare, P.-E., Das, R., and Bordes, A. Weaver: Deep co-encoding of questions and documents for machine reading. In Proceedings of the International Conference on Machine Learning (ICML), 2018.
- Song, L., Wang, Z., Yu, M., Zhang, Y., Florian, R., and Gildea, D. Exploring Graph-structured Passage Representation for Multi-hop Reading Comprehension with Graph Neural Networks. arXiv preprint arXiv:1809.02040, 2018.
- Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics.

Acknowledgements

We would like to thank Johannes Welbl for helping to test our system on WikiHop. This project is supported by SAP Innovation Center Network, ERC Starting Grant BroadSem (678254) and the Dutch Organization for Scientific Research (NWO) VIDI 639.022.518. Wilker Aziz is supported by the Dutch Organisation for Scientific Research (NWO) VICI Grant nr. 277-89-002.