Question Answering by Reasoning Across Documents with Graph Convolutional Networks

Nicola De Cao¹², Wilker Aziz¹², and Ivan Titov¹²
¹University of Edinburgh, ²University of Amsterdam

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 perform multi-step reasoning
- Our method is scalable and compact, and it achieves state-of-the-art results on WikiHop (Welbl et al., 2018)

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
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

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)

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

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

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 Broadcom (642054) and the Dutch Organization for Scientific Research (NWO) VIDI 439.022.518. Wilker Aziz is supported by the Dutch Organization for Scientific Research (NWO) VICI Grant nr. 277.89.002.

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

Processing the Graph
Entity Relational Graph Convolutional Network
Gated version of relational GCNs (Schlichtkrull et al., 2018).

\[ u_i^{(l)} = f_i^{(l)}(h_i^{(l-1)}) + \frac{1}{|N_i|} \sum_{j \in N_i} f_j^{(l)}(h_j^{(l-1)}) \]  \[ a_i^{(l)} = \sigma \left( f_i^{(l)}(u_i^{(l)}, h_i^{(l)}) \right) \]  \[ h_i^{(l+1)} = \theta(u_i^{(l)}) \odot a_i^{(l)} + h_i^{(l)} \odot (1 - a_i^{(l)}) \]

Candidates scoring
We use the final node embeddings (h_i) and the question representation (q) to predict a distribution over candidates.

\[ P(c|q, C_j, S) \propto \exp \left( \max_{i \in M} f_j(q, h_i^{(l)}) \right) \]

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