Reasoning in Multi-hop Reading Comprehension

Yifan Gao

24/4/2019

Reasoning in Multi-hop RC

- Graph-based reasoning (implicit)
  - Question Answering by Reasoning Across Documents with Graph Convolutional Networks, NAACL 2019
  - Exploring Graph-structured Passage Representation for Multi-hop Reading Comprehension with Graph Neural Networks
  - Bi-directional Attention Entity Graph Convolutional Network for Multi-hop Reasoning Question Answering, NAACL 2019 Short
- Path-based reasoning (explicit)
  - Exploiting Explicit Paths for Multi-hop Reading Comprehension

Question Answering by Reasoning Across Documents with Graph Convolutional Networks

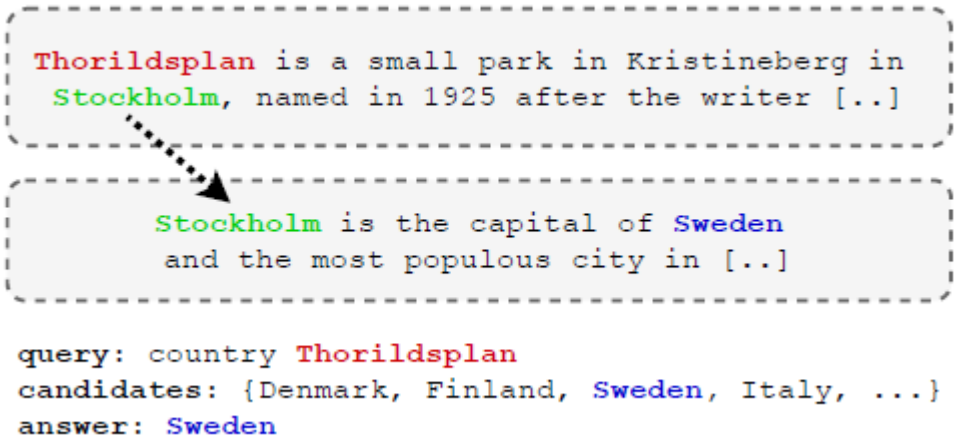
Nicola De Cao
University of Edinburgh
University of Amsterdam
nicola.decao@uva.nl

Wilker Aziz
University of Amsterdam
w.aziz@uva.nl

Ivan Titov
University of Edinburgh
University of Amsterdam
ititov@inf.ed.ac.uk

WIKIHOP QA

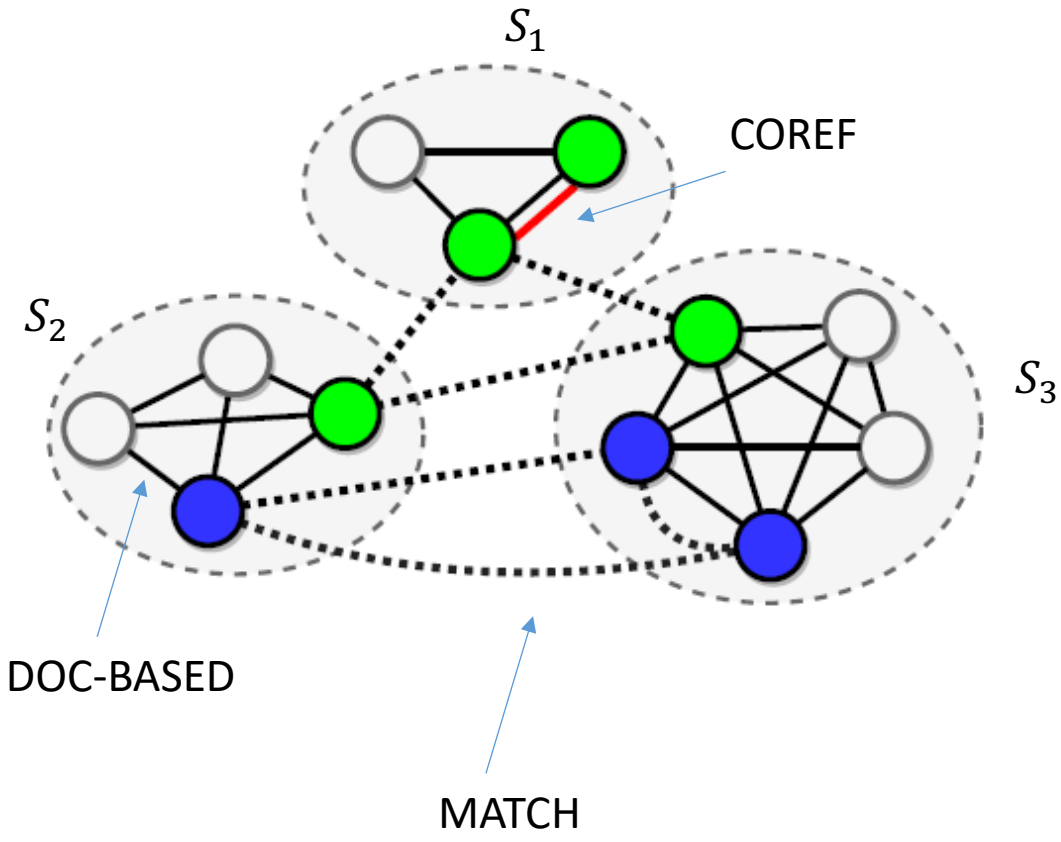
- Answerable using a chain of reasoning crossing document boundaries



- S_q: set of supporting documents (also distractors)
- q: query/question $< s, r, ? >$
- C_q: set of candidate answers, $a^* \in C_q$: correct answer

Construct Entity Graph (offline)

- Node: Identify mentions in S_q of entities of $C_q \cup s$
  - exact match
  - coreference
- Edge
  - DOC-BASED: co-occur in S_i
  - MATCH: identical mentions
  - COREF: coreference chain
  - COMPLEMENT: nodes not connected with any of other relations



Graph Encoding

- Query: Bi-RNN initialized with ELMo
- Nodes: query-dependent mention encodings
  - ELMo representation $\{\mathbf{x}_i\}_{i=1}^{N}$
  - query-dependent $\hat{\mathbf{x}}_i = f_x(\mathbf{q}, \mathbf{x}_i)$

Multi-step Reasoning

- Propagate information through the entity graph
  - relational-GCNs (Schlichtkrull et al., 2018)
  - Initial representation $\{\mathbf{h}_i^{(0)}\}_{i=1}^N$
  - After L layers of R-GCN $\{\mathbf{h}_i^{(L)}\}_{i=1}^N$
- Prediction $$P(c|q, C_q, S_q) \propto \exp\left(\max_{i \in \mathcal{M}_c} f_o([\mathbf{q}, \mathbf{h}_i^{(L)}])\right)$$

Gated R-GCN

- First layer: $\mathbf{h}_i^{(0)} = \hat{\mathbf{x}}_i$
- ith node in l layer:
  - Updated message

$$\mathbf{u}_i^{(\ell)} = f_s(\mathbf{h}_i^{(\ell)}) + \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \sum_{r \in \mathcal{R}_{ij}} f_r(\mathbf{h}_j^{(\ell)})$$

    - f_r: parameterized function specific to an edge type $r \in R$
  - Gating:

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

  - Combination:

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

Exploiting Explicit Paths for Multi-hop Reading Comprehension

Souvik Kundu[+*] and **Tushar Khot**[†] and **Ashish Sabharwal**[†]
[+]National University of Singapore, Singapore
[†] Allen Institute for Artificial Intelligence, Seattle, U.S.A.
souvik@comp.nus.edu.sg, tushark@allenai.org, ashishs@allenai.org

Background

- Graph-based reasoning: implicit, lack of interpretability
- Propose a model extract paths from text and encodes the knowledge

Question: (always breaking my heart, record_label, ?)

Supporting Passages:

(p1) "Always Breaking My Heart" is the second single from Belinda Carlisle's A Woman and a Man album , released in 1996 (see 1996 in music) . It made ...

(p2) A Woman and a Man is the sixth studio album by American singer Belinda Carlisle, released in the United Kingdom on September 23, 1996 by Chrysalis Records (then part of the EMI Group, like Carlisle's former ...

Candidates: chrysalis Records, emi group, virgin records, ...

Answer: chrysalis records

Path:

("Always Breaking My Heart" ... *single from* ... A Woman and a Man)

(A Woman and a Man ... *released* ... *by* ... Chrysalis Records)

- Extract implicit relations
- Compositions of relations

- For example:
  - (x, *single from*, y)
  - (y, *released by*, z)

  - (x, *record_label*, z)

Notations

- WIKIHOP dataset
- Query $< h_e, r, ? >$
- Answer candidates $(\mathrm{c}_1, \mathrm{c}_2, \ldots, \mathrm{c}_\mathrm{N})$
- Paths $P = p_1, \ldots, p_M$
  - p_{kj}: j-th path for candidate c_k, $p_{kj} = \mathrm{h_e} \rightarrow \mathrm{e}_1 \rightarrow \mathrm{c_k}$

Path Extraction

- Find passage p_1 containing h_e
- Find all named entities e_i in the same/subsequent sentence of h_e
- Find e_i in p_2
- Find c_i in p_2
- can be extended into k hops

Question: (always breaking my heart, record_label, ?) h_e

Supporting Passages:

(p1) "Always Breaking My Heart" is the second single from Belinda Carlisle e_1's A Woman and a Man e_2 album , released in 1996 (see 1996 in music) . It made ...

(p2) A Woman and a Man e_2 is the sixth studio album by American singer Belinda Carlisle e_1, released in the United Kingdom on September 23, 1996 by Chrysalis Records c_1 (then part of the EMI Group, like Carlisle's former ...

Candidates: chrysalis Records, emi group, virgin records, ...

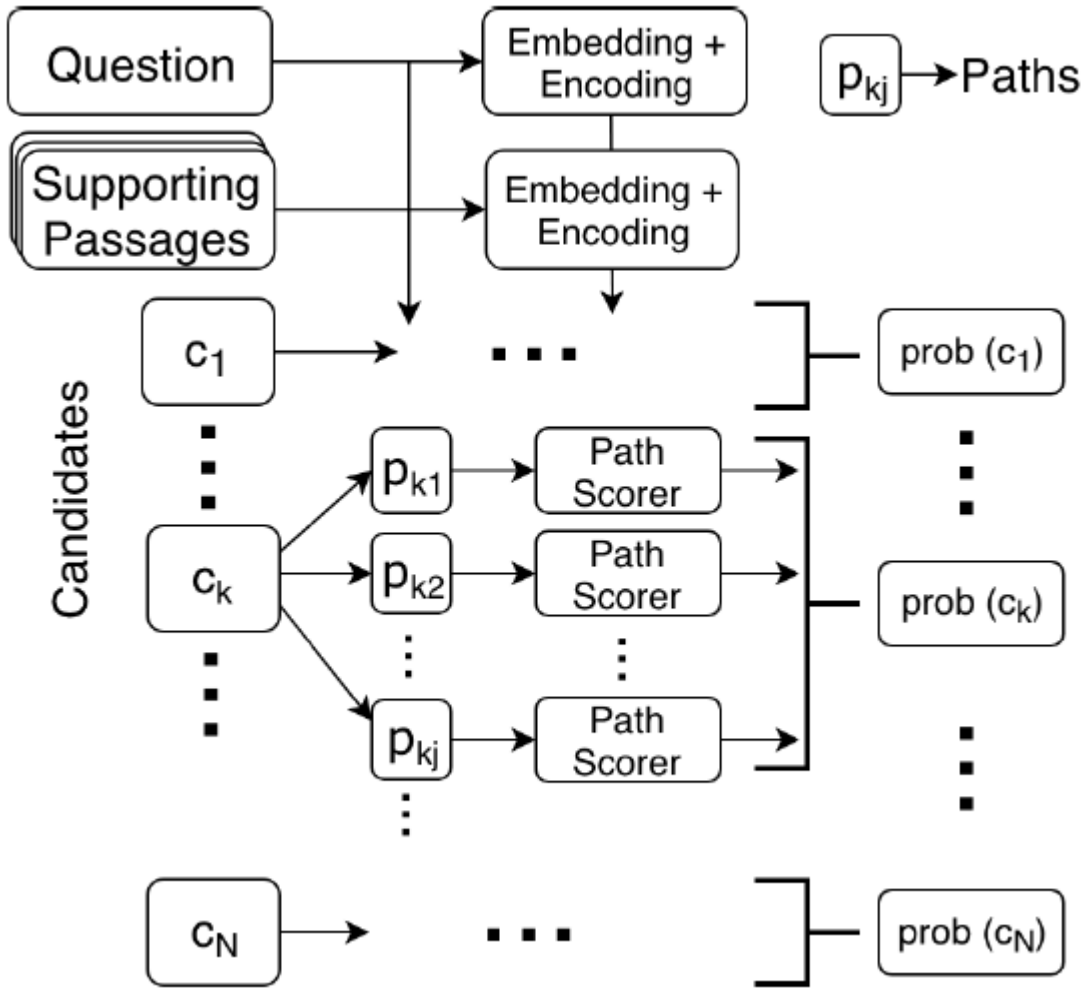
Answer: chrysalis records

Path:

("Always Breaking My Heart" ... *single from* ... A Woman and a Man)

(A Woman and a Man ... *released* ... *by* ... Chrysalis Records)

Path-based Multi-Hop QA Model



$$\text{prob}(c_k) = \sum_j \text{score}(\text{p}_{kj})$$

Path Encoding

- context-based encoding
  - implicitly encode the relation between $(h_e, e_1), (e_1', c_k)$

$$\mathbf{g}_{h_e} = \mathbf{s}_{p_1,i_1} \,||\, \mathbf{s}_{p_1,i_2}$$

$$\mathbf{r}_{h_e,e_1} = \mathrm{FFL}(\mathbf{g}_{h_e}, \mathbf{g}_{e_1})$$

$$\mathbf{x}_{ctx} = \mathrm{FFL}(\mathbf{r}_{h_e,e_1}\ ,\ \mathbf{r}_{e_1\prime,c_k})$$

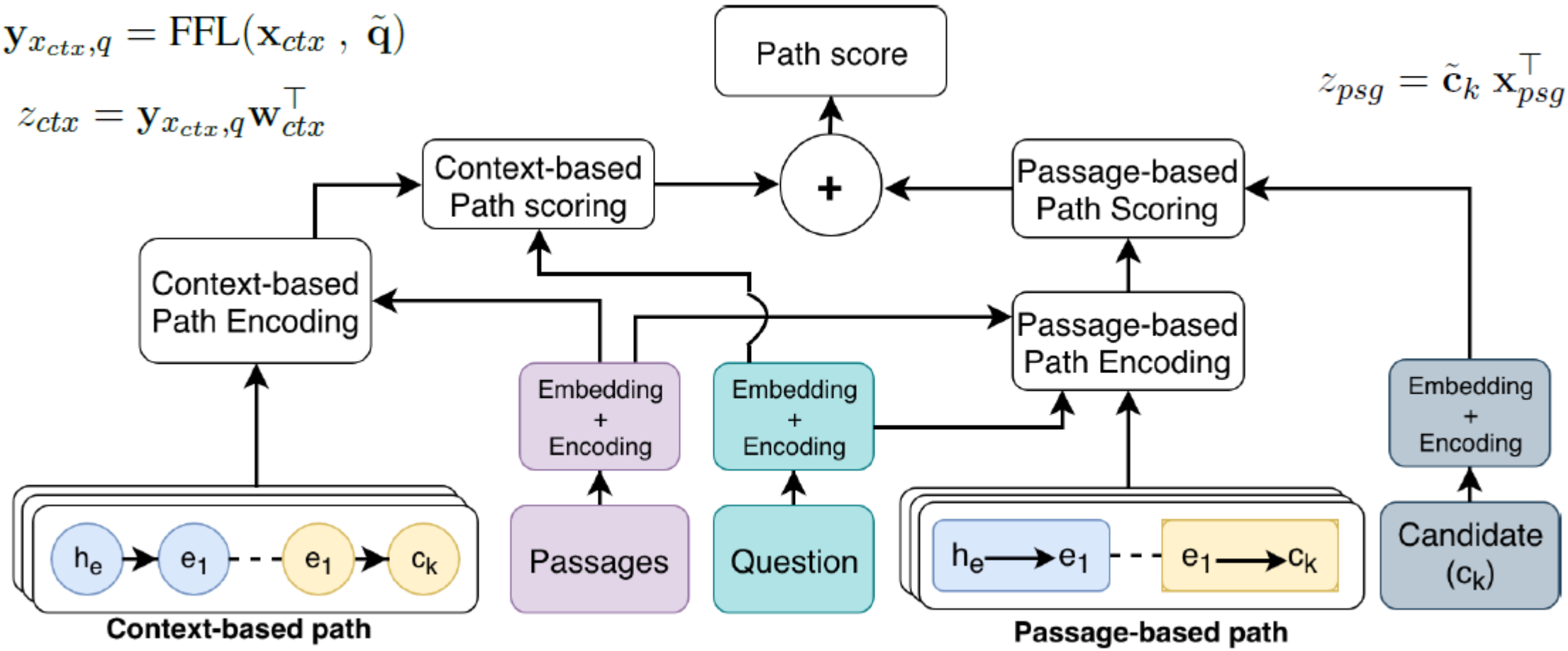
- passage-based
  - use the whole passages to compute the path representation
  - Question-weighted Passage Representation
  - Aggregate Passage Representation

Path Scoring

$$\mathbf{y}_{x_{ctx},q} = \mathrm{FFL}(\mathbf{x}_{ctx}\ ,\ \tilde{\mathbf{q}})$$

$$z_{ctx} = \mathbf{y}_{x_{ctx},q}\mathbf{w}_{ctx}^{\top}$$

$$z_{psg} = \tilde{\mathbf{c}}_k\ \mathbf{x}_{psg}^{\top}$$



Results

Model	Unmasked		Masked	
	Test	Dev	Test	Dev
Human (Welbl et al., 2018)	74.1	–	–	–
FastQA (Welbl et al., 2018)	25.7	–	35.8	–
BiDAF (Welbl et al., 2018)	42.9	–	54.5	–
Coref-GRU (Dhingra et al., 2018)	59.3	56.0	–	–
MHPGM (Bauer et al., 2018)	–	58.2	–	–
Weaver / Jenga (Raison et al., 2018)	65.3	64.1	–	–
MHQA-GRN (Song et al., 2018)	65.4	62.8	–	–
Entity-GCN without coreference (single model)	**67.6**	**64.8**	–	70.5
Entity-GCN with coreference (single model)	**66.4**	**65.3**	–	–
Entity-GCN* (ensemble 5 models)	**71.2**	**68.5**	–	71.6

Model	Accuracy (%)	
	Dev	Test
Welbl et al. (2018)	-	42.9
Dhingra et al. (2018)	56.0	59.3
Song et al. (2018)	62.8	65.4
Cao et al. (2018)	64.8	67.6
Our model	**67.1**	-

Remarks

- Two approaches for reasoning, KB2text, dialogue over KB, graph2seq
  - Path-based
    - Knowledge Aware Conversation Generation with Reasoning on Augmented Graph
    - Step-by-Step Separating Planning from Realization in Neural Data-to-Text Generation, NAACL 2019
  - Graph-based
    - Text Generation from Knowledge Graphs with Graph Transformers, NAACL 2019