No Need to pay Attention: Simple Recurrent Neural Networks Work! (for Answering “Simple” Questions)

Ferhan Ture and Oliver Jojic
Comcast Labs, Washington, DC 20005

EMNLP 2017

Happy Buzaaba
December 2018
University of Tsukuba

Outline

- **Introduction**
- **Related work**
- **Proposed Approach**
- **Experimental results**
- **My results following a similar approach**
- **Conclusion**

Simple question answering:

Questions that require extraction of a single fact from knowledge base.

Example:

where was Baracka Obama born?

Answered by a single fact:

("Baracka Obama"; "Place of birth"; "Hawaii")

Related work

- Memory networks(Bodes et al., 2015)
- Character-level attention-based Encorder-Decorder (Golub & He, 2016)
- Hierarchical Word/char-level Encorder (Lukovinikov et al., 2017)
- Attentive max-pooling (Yin et al., 2016)
- Hierarhical BiLSTM (Yu et al., 2017)

Approach

Decompose the Simple Question answering problem into two subproblems:

- Detecting entities in the question
- Classifying question as one of the relation types in the KB

Approach

Example:

where was Baracka Obama born?

Answered by a single fact:

("Baracka Obama"; "Place of birth"; "Hawaii")

Structured query: {entity: "Baracka Obama", relation: place_of_birth}

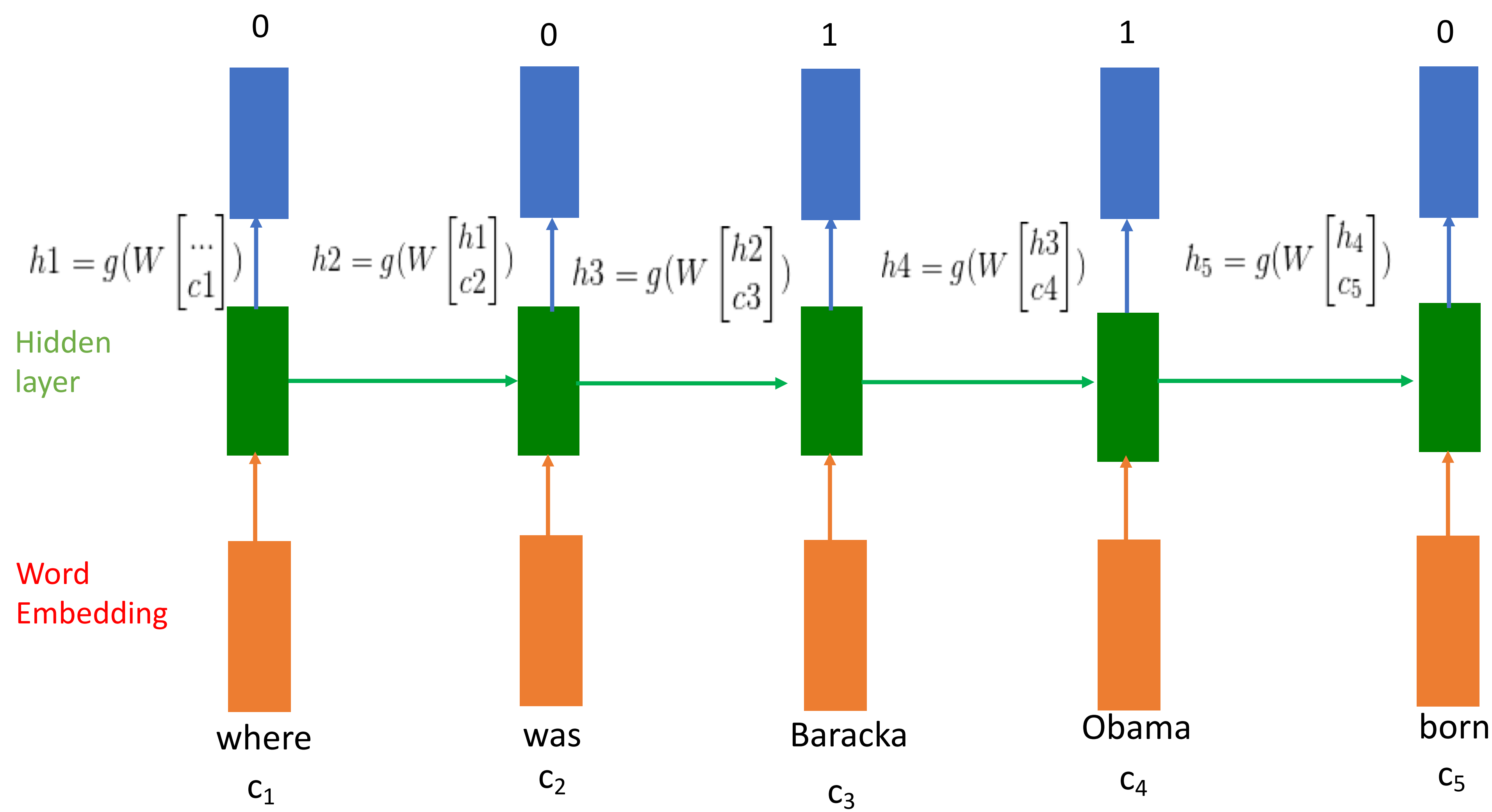
Steps to the approach

- Tag each question word as either entity or not

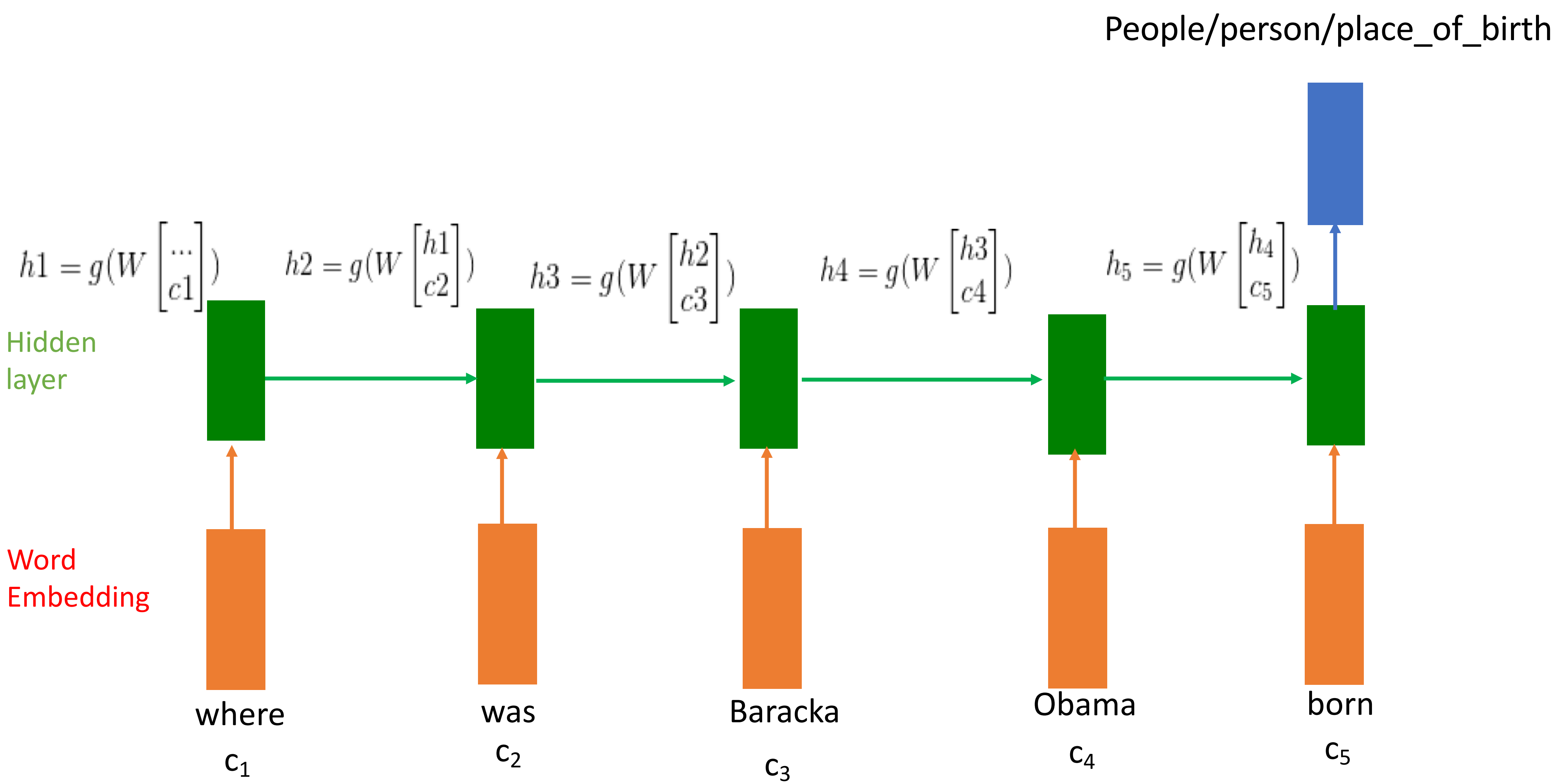
- Classify the question into one of the KB relation types

NB: Both are modeled using a standard RNN architecture

Entity detection



Relation Prediction



Inferred structured query

entity phrase extraction

where	was	Baracka	Obama	born
0	0	1	1	0

[0 0 1 1 0]

Single entity extraction

"Baracka Obama"

Output of relation prediction

People/person/place_of_birth

Inferred structured query becomes {entityText:"Baracka Obama", relation: People/person/place_of_birth}

Entity linking

- Need to link to the actual node in the knowledge graph

  Baraka Obama **fb:m.07f3jg, fb:m.03hjy39, fb:m.040qyz …**

Creating indexes:

- **Names index:** maps all freebase MID's to the names file (**Dai et al., 2016**)
- **Inverted entity Index**: maps n_grams of an entity for **n** ∈ {1, 2, 3 } to entity MID
- **Reachability index**: maps each entity node in freebase to all reachable nodes

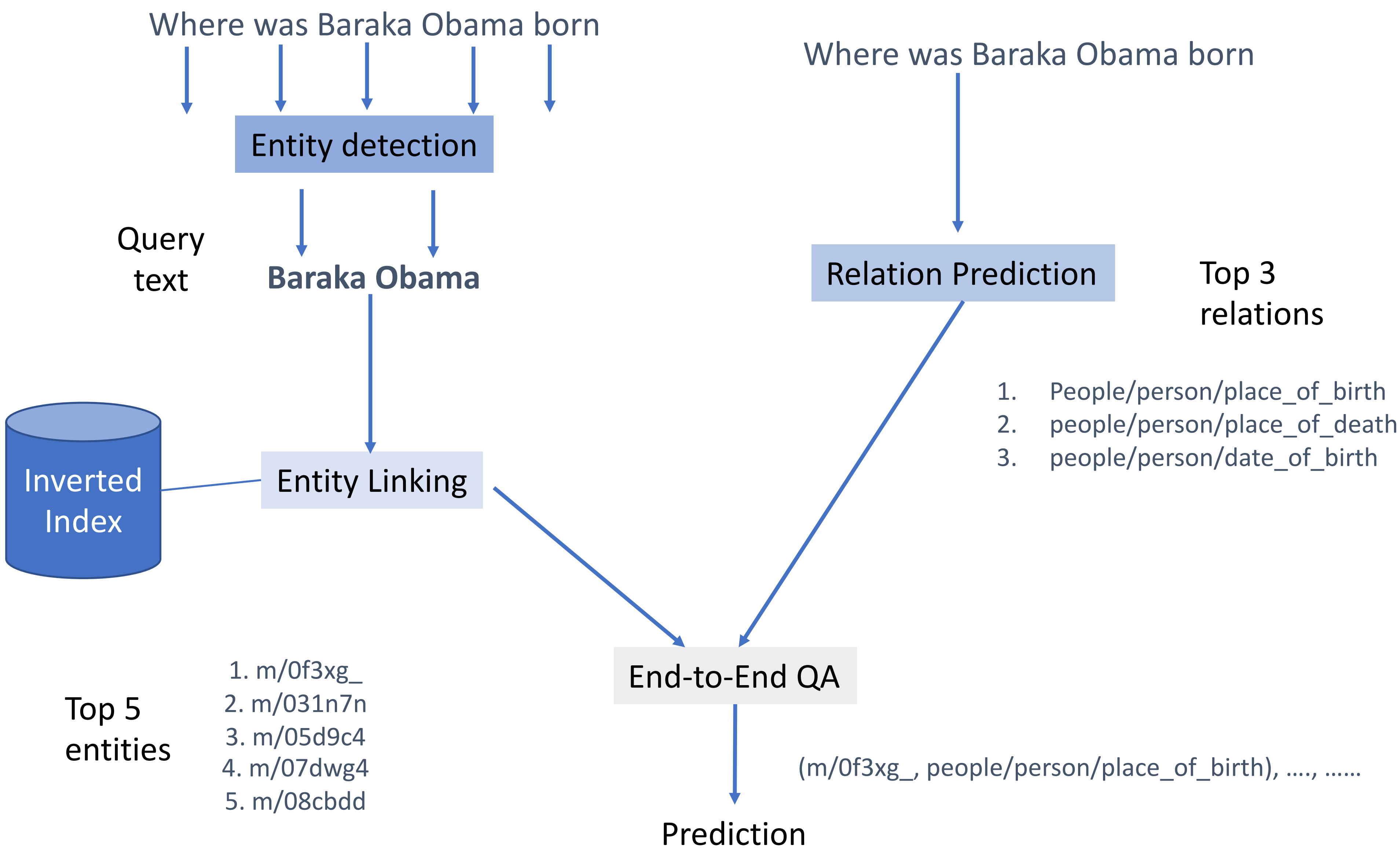
Entity Inverted Index

Example: e_i : "Baracka Obama"

alias:
- 2-unigrams ("baracka", "obama'')
- 1-bigram ("baraka obama")

- I_{entity} ("barack") $\Longrightarrow$ { node: e_i score: TF-IDF ("baracka", "baracka obama") }
- I_{entity} ("baracka obama") $\Longrightarrow$ { node: e_i score: TF-IDF ("baracka obama", "baracka Obama") }

End-to-end.



Where was Baraka Obama born
Entity detection
Query text
Baraka Obama
Inverted Index
Entity Linking
Top 5 entities
1. m/0f3xg_
2. m/031n7n
3. m/05d9c4
4. m/07dwg4
5. m/08cbdd
Where was Baraka Obama born
Relation Prediction
Top 3 relations
1. People/person/place_of_birth
2. people/person/place_of_death
3. people/person/date_of_birth
End-to-End QA
(m/0f3xg_, people/person/place_of_birth), …., ……
Prediction

Experimental Results

Model	P@1
Memory Network (2015)	63.9
Char-level CNN (2016)	70.9
Attentive Max-pooling (2016)	76.4
RNN-QA (best models)	**88.3**
naive ED	58.9
naive RP	4.1
naive ED and RP	3.7

Table 1: Top-1 accuracy on test portion of simples questions, Ablation study on last three rows

Error	**Count**
Correct	220
Incorrect entity	16
Incorrect relation	42
Not first-order questions	17
Total Latency	76±16 ms

Table 2: Evaluation of RNN-QA on real questions from Comcast X1 platform

Results following a similar approach

R@N	BiLSTM	
	Val	Test
1	0.677	0.66
5	0.825	0.81
20	0.887	0.88
50	0.91	0.9

Table 3: Results for entity linking using LSTM entity detection model

Model	Val R@1	Test R@1
BiLSTM	81.75	81.28
CNN	82.87	81.92

Table 4: Results for relation prediction using different models

Results following a similar approach

Entity	Relation	Acc.
BiLSTM	BiLSTM	74.59
BiLSTM	CNN	74.63
Previous work		
Bodes et al. (2015)		62.7
Golub and He (2016)		70.9
Lukovnikov et al. (2017)		71.2
Dai et al. (2016)		75.7
Yin et al. (2016)		76.4
Ture and Jojic (2017)		**88.3**

Table 5: Accuracy on test set with different model combinations compared with some previous work

Conclusion

- We describe a simple yet effective approach for QA, focusing on simplequestions. Although we understand the benefit of exploring task-agnostic approaches that aim to capture semantics in amore general way eg (Kumar et al., 2015) it is also important to acknowledge that there is no “one-size-fits-all” solution as of yet.
- While an ablation study revealed the importance of both entity detection and relation prediction, we are hoping to further study the degree of which improvements in either component affect QA accuracy
- Even though deep learning has opened the potential for more generic solutions, we believe that taking advantage of problem structure yields a more accurate and efficient solution.