On Incorporating Structural Information to Improve Dialogue Response Generation

Nikita Moghe[ǂ], Priyesh Vijayan[ᵻ], Balaraman Ravindran[*+], Mitesh M. Khapra[*+]
*Indian Institute of Technology Madras
+ Robert Bosch Centre for Data Science and Artificial Intelligence, IIT Madras
ǂUniversity of Edinburgh ᵻMcGill University & MILA
{cs16s016, priyesh, ravi, miteshk}@cse.iitm.ac.in

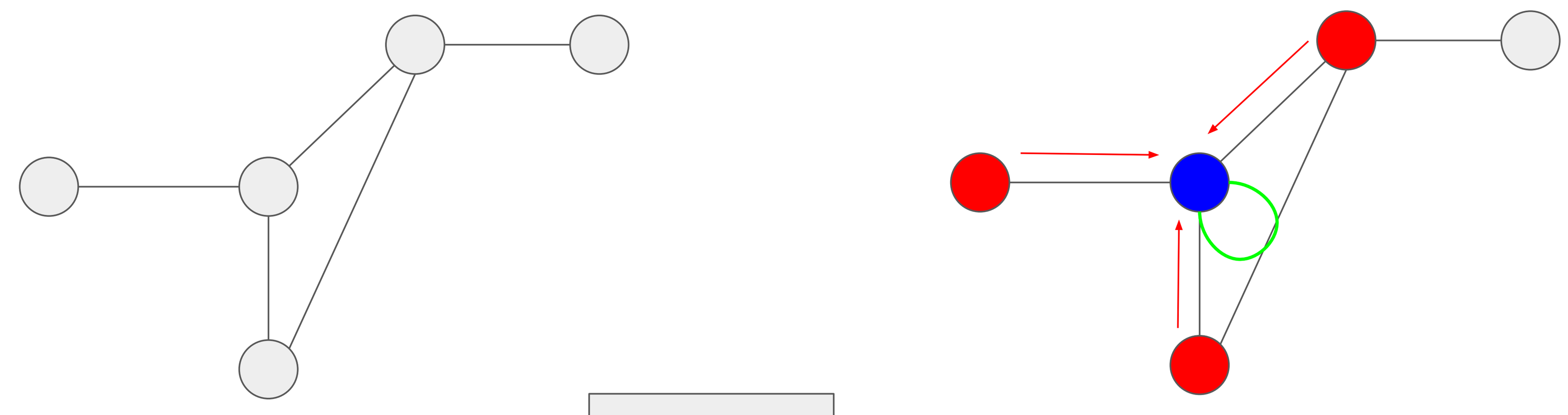
This work was done by Nikita and Priyesh at IIT Madras.

Motivation

- **Structural** information is **ubiquitous** in natural language
  Dependency parses, constituency graphs, co-reference graphs
- But **RNNs** process text sequentially and have **trouble** in learning **structure-sensitive dependencies** without explicit supervision.
- **Deep contextualized** word representations **hope** to capture structural properties **implicitly** by training on large amount of data.

Can we *explicitly* incorporate structural information to improve these neural architectures?

Graph Convolutional Network (GCN)



Undirected Graph

Node information

$$h_i = ReLU\left(W_0 h_i + \sum_{u \in N(i)} W_1 h_u\right)$$

Neighbourhood information

Background Aware Conversation Systems

Task: Given conversation history and associated background knowledge, generate a response

Dataset: Holl-E Domain: Movies Language: English Stats: ~9k chats , ~90k utterances, ~9k resources
Every alternate response is formed by copying words from the resource with appropriate prefixes/ suffixes

Speaker 1: Yes very true, this is a real rags to riches story. Russell Crowe was excellent as usual

Speaker 2: Russell Crowe owns the character of James Bradock. He's a good fighter turned hack. Injury, bad luck and the Depression sends him down the drain.

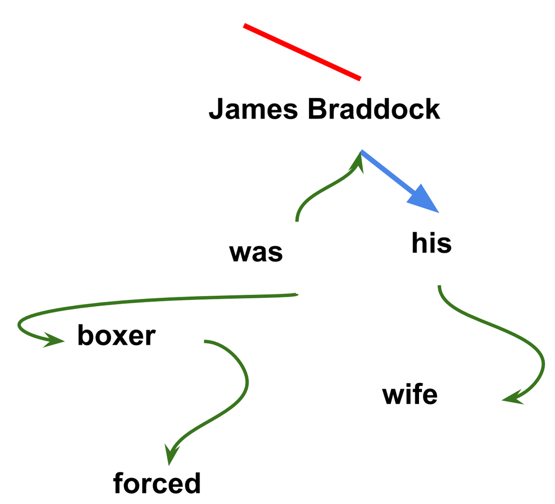
Speaker 1: Totally! Oh by the way do you remember his wife ... how she wished he would stop

Speaker 2: Yes! **His wife Mae had prayed for years that he would quit boxing, before becoming permanently injured.**

Conversation

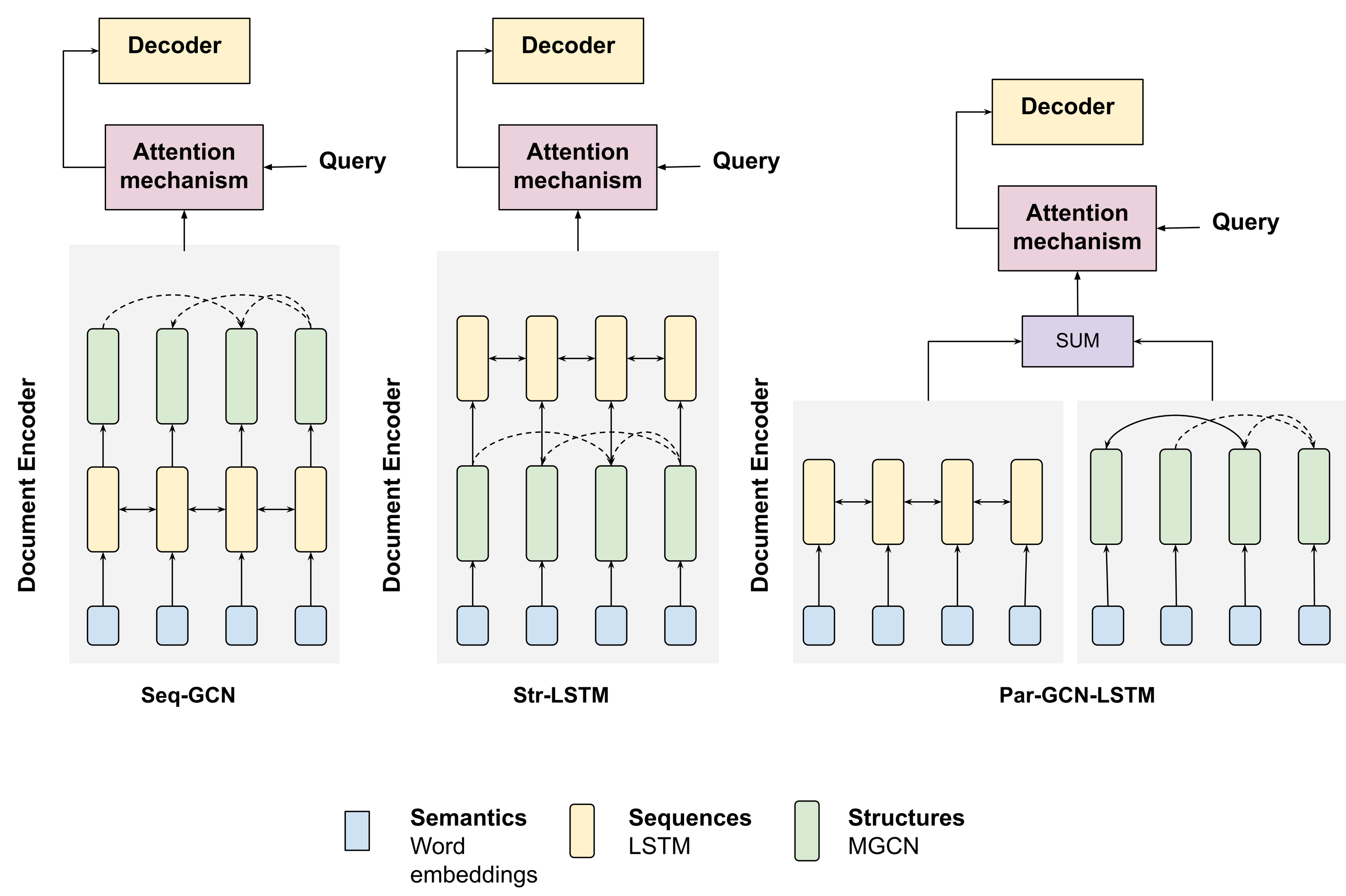
Russel Crowe plays the lead. At this point James Braddock was a boxer, who was forced to retired from the ring after breaking his hand in his last fight. His wife Mae had prayed for years that he would quit boxing, before becoming permanently injured. To support his family, Braddock works as a laborer at the docks, but he still has a dream to box.
.

Resource



Dependency graph
Entity graph
Co-reference graph

Semantics-Sequences-Structures (SSS) Framework



Multi Graph GCN: $h^{k+1}_i = ReLU\left(W^k_0 h^k_i + \sum_{N \in G} \sum_{u \in N(i)} \left(W^k_{dir(i,u)} h^k_u + b^k_{L(i,u)}\right)\right)$

Results

	Model	BLEU	ROUGE		
			1	2	L
	HRED	05.23	24.55	07.61	18.87
	GTTP	13.92	30.32	17.78	25.67
	BiDAF	16.79	26.73	18.82	23.58
GloVe	Sem	04.40	29.72	11.72	22.99
	Sem + Seq	14.83	36.17	24.84	31.07
	Sem + Seq + Str	18.96	38.61	26.92	33.77
ELMo	Sem	14.36	32.04	18.75	26.71
	Sem + Seq	14.61	35.54	24.58	30.71
	Sem + Seq + Str	19.32	39.65	27.37	34.86
BERT	Sem	11.26	33.86	16.73	26.44
	Sem + Seq	18.49	37.85	25.32	32.58
	Sem + Seq + Str	**22.78**	**40.09**	**27.83**	**35.20**

Table 1: Results of automatic evaluation. Our proposed framework outperforms the baseline methods.

Semantics	Seq-GCN		Str-LSTM		Parl-GCN-LSTM	
	BLEU	ROUGE-L	BLEU	ROUGE-L	BLEU	ROUGE-L
GloVe	15.61	31.68	18.96	33.77	17.10	32.20
ELMo	18.44	33.05	19.32	34.86	16.35	32.12
BERT	20.43	34.85	22.78	35.20	21.32	34.87

Table 2: Performance of different hybrid architectures to combine structural information with sequence information.

Qualitative Evaluation of 100 chats:
A/B testing of SSS models against GTTP baseline. Best SSS models chosen across different graph combinations, different contextual and structural infusion methods, and three M-GCN layers

SSS	Win	Loss	Same	None
GloVe	24	17	47	12
ELMo	22	23	41	14
BERT	29	25	29	17

Analysis

- Explicit incorporation of structural information along with semantic and sequential information improves over standard architectures.

- Architectures with a sequence layer at the top are best suited for span-based copy generation task.

- Graphs must have some linguistic property. Random graphs don't work!

- Deep contextualized representation based models also benefit from addition of explicit structural information.

- SSS (GloVe) is competitive and has lesser memory footprint than deep contextualized architectures that capture structural information implicitly.

Conclusion

We believe that the analysis presented in this work would serve as a blueprint for analysing future work on GCNs ensuring that the gains reported are robust and evaluated across different configurations.

References

[1] *Semi-supervised classification with Graph Convolutional Network*, Thomas N Kipf and Max Welling. In International Conference on Learning Representations (ICLR) 2016

[2] *Towards Exploiting Background Knowledge for Building Conversation Systems*, Nikita Moghe, Siddhartha Arora, Suman Banerjee, and Mitesh M. Khapra. In proceedings of Empirical Methods in Natural Language Processing (EMNLP) 2018