Graph-to-Sequence Learning using Gated Graph Neural Networks

Daniel Beck, Gholamreza Haffari, Trevor Cohn

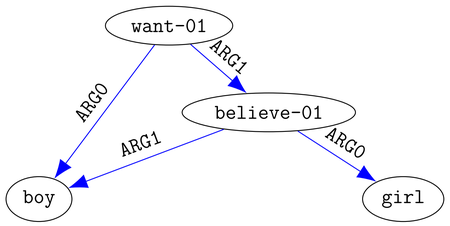
d.beck@unimelb.edu.au, gholamreza.haffari@monash.edu, t.cohn@unimelb.edu.au

Code available at: github.com/beckdaniel/acl2018_graph2seq

Introduction

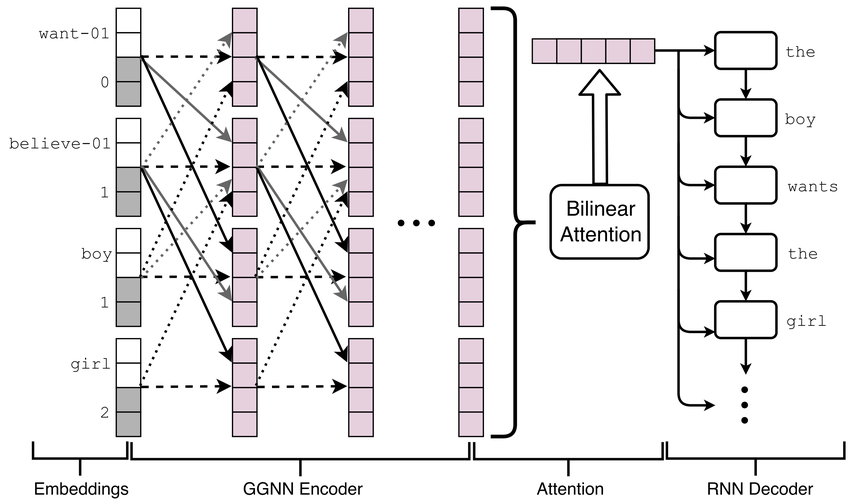
Goal: generate natural language sentences from *graphs*.

- Generation from structured representations.
- Neural Machine Translation with source syntax information.



⇒ "The boy wants the girl to believe him."

- Previous work linearise graphs and apply off-the-shelf sequence-to-sequence networks.
- **Our approach:** replace the sequential encoder with a *Gated Graph Neural Network* [Li et al., ICLR 2016].



Acknowledgements

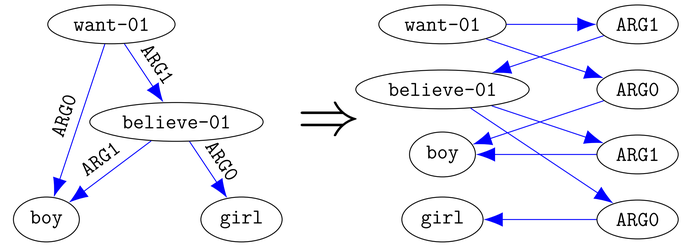
This work was supported by the Australian Research Council (DP160102686) and by the 2017 JSALT workshop, hosted at Carnegie Mellon University and sponsored by Johns Hopkins University with unrestricted gifts from Amazon, Apple, Facebook, Google, and Microsoft. The authors would also like to thank Joost Bastings for sharing the data from his paper's experiments.

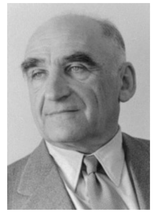
Graph Transformations

Standard GGNN caveats:

- Parameters increase quadratically with number of edge labels.
- Does not generate hidden vectors for edges.
- Does not encode sequential information.

Levi graphs [Levi,1942]

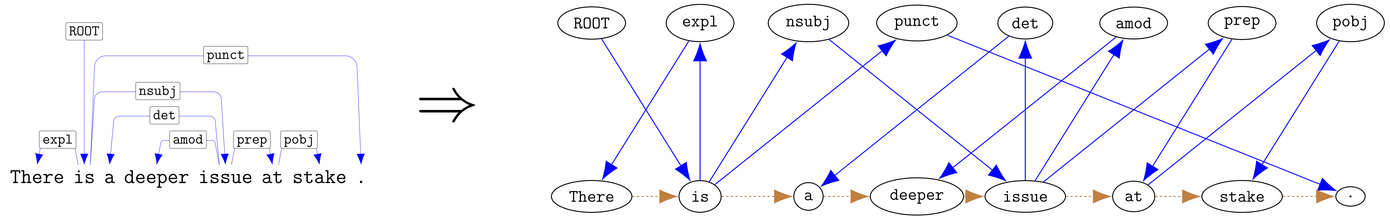




Friedrich Wilhelm Levi

(source: University of Leipzig)

For sequential information we add extra edges to the Levi graph.



Translation: *"Doch steht etwas Grundlegenderes auf dem Spiel ."*

Experiments and Discussion

Data and preprocessing

AMR generation: LDC2017T10 with default splits. Graph simplification and anonymisation [Konstas et al., ACL 2017]

Syntax-based NMT: same settings as in [Bastings et al. EMNLP 2017]. News Commentary V11, English-German and English-Czech. Source side parsed using SyntaxNet and target side segmented using Byte-Pair Encoding.

Models

s2s: sequence-to-sequence baseline.

g2s: our graph-to-sequence model.

g2s+: as above but with extra sequential connections.

Main conclusions

- g2s consistently outperforms s2s in AMR generation
- For NMT, performance drops for standard g2s but g2s+ outperforms the baselines. *Sequential biases added as a graph transformation: no RNNs required in the encoder.*

	BLEU	CHRF++	#params
AMR Generation			
s2s	21.7	49.1	28.4M
g2s	**23.3**	**50.4**	28.3M
NMT English-German			
PB-SMT	12.8	**43.2**	–
s2s	15.5	40.8	41.4M
g2s	15.2	41.4	40.8M
g2s+	**16.7**	42.4	41.2M
NMT English-Czech			
PB-SMT	8.6	**36.4**	–
s2s	8.9	33.8	39.1M
g2s	8.7	32.3	38.4M
g2s+	**9.8**	33.3	38.8M