Paper Reading

黄轩成

2020/07/02

GNN for NMT

Paper Reading #1

- EMNLP 2017

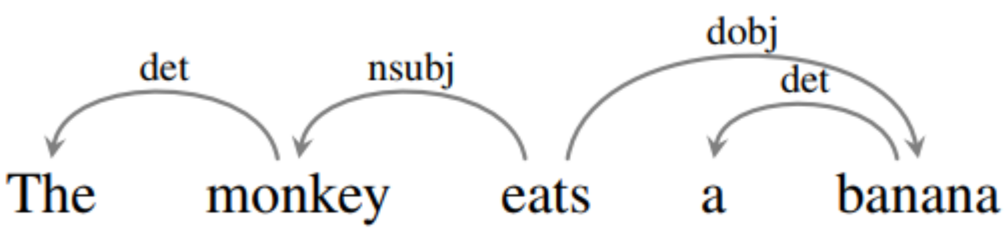
Graph Convolutional Encoders for Syntax-aware Neural Machine Translation

Joost Bastings[1] **Ivan Titov**[1,2] **Wilker Aziz**[1]
Diego Marcheggiani[1] **Khalil Sima'an**[1]
[1]ILLC, University of Amsterdam [2]ILCC, University of Edinburgh
{bastings,titov,w.aziz,marcheggiani,k.simaan}@uva.nl

Goal

- Goal: incorporating syntactic information into NMT.

- Example of a dependency tree:



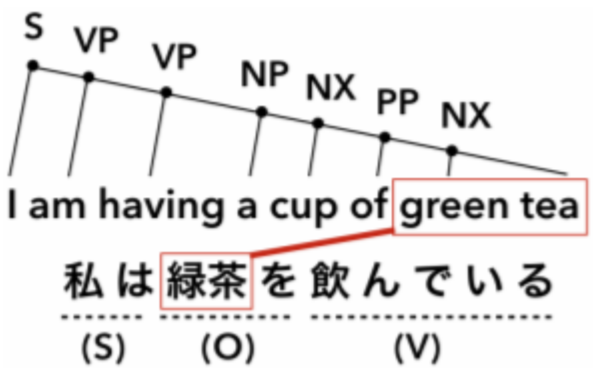
(Bastings et al., 2017)

Previous Work

- How to incorporate syntactic information?
  - Implicit:
    - Multi-task: learning to parse and translate simultaneously (Eriguchi et al., 2017).

Previous Work

- How to incorporate syntactic information?
  - Implicit:
    - Multi-task: learning to parse and translate simultaneously (Eriguchi et al., 2017).
  - Explicit but too restrictive:
    - Learning representations of linguistic phrases (Eriguchi et al., 2016).



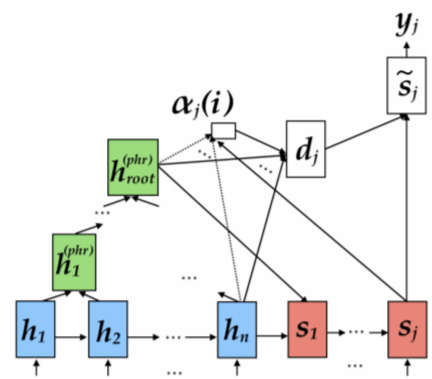


Figure 3: Proposed model: Tree-to-sequence attentional NMT model.

(Eriguchi et al., 2016)

Previous Work

- GNN is already used in other NLP task to incorporate syntactic information.
  - Semantic Role Labeling (Marcheggiani and Titov, 2017).

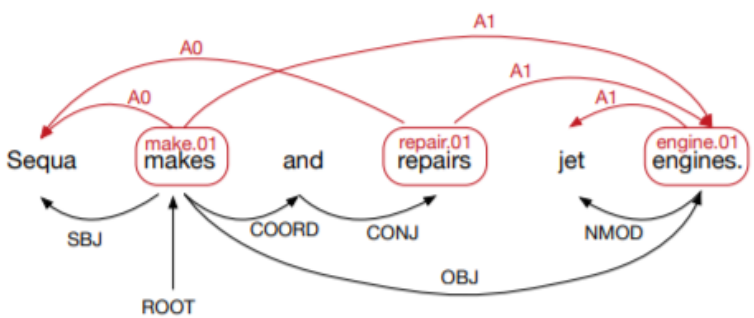
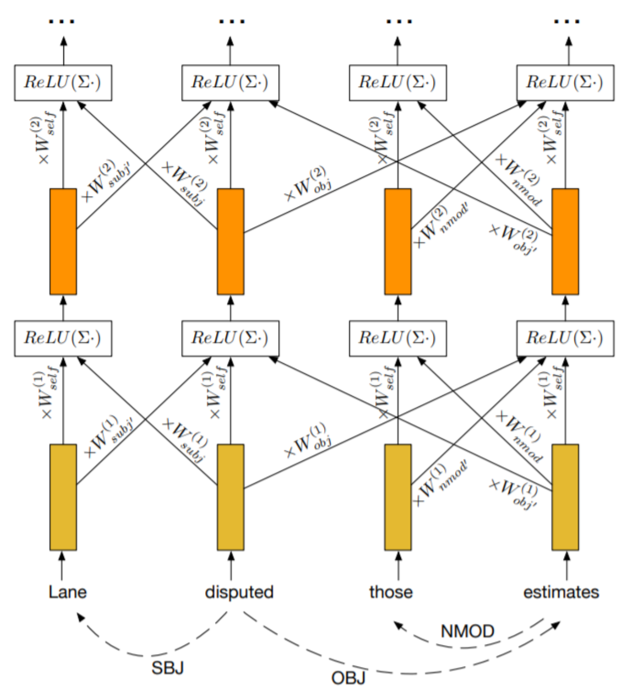


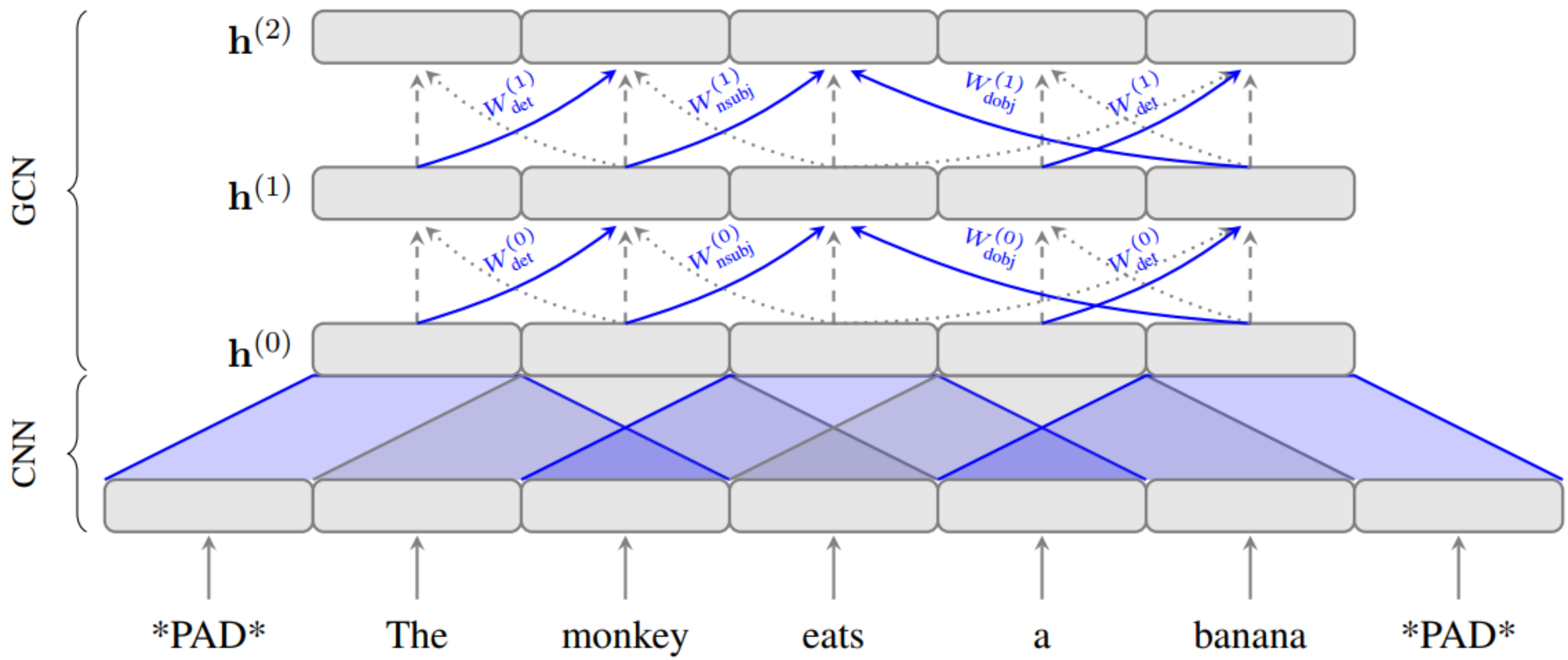
Figure 1: An example sentence annotated with semantic (top) and syntactic dependencies (bottom).



(Marcheggiani and Titov, 2017)

Syntactic GCN (this work)

- A two-layer syntactic GCN on top of a convolutional encoder.



(Bastings et al., 2017)

GCN

- Simple recursive computation:

$$\mathbf{h}_v^{(j+1)} = \rho\left(\sum_{u \in \mathcal{N}(v)} W^{(j)} \mathbf{h}_u^{(j)} + \mathbf{b}^{(j)}\right)$$

where $\mathcal{N}(v)$ is the set of neighbors of v.

Directionality and Labels

- Direction-specific weight matrix.

$$\mathbf{h}_v^{(j+1)} = \rho\left(\sum_{u \in \mathcal{N}(v)} W_{\mathrm{dir}(u,v)}^{(j)} \, \mathbf{h}_u^{(j)} + \mathbf{b}_{\mathrm{dir}(u,v)}^{(j)} \right)$$

- Label-specific weight matrix.

$$\mathbf{h}_v^{(j+1)} = \rho\left(\sum_{u \in \mathcal{N}(v)} W_{\mathrm{lab}(u,v)}^{(j)} \, \mathbf{h}_u^{(j)} + \mathbf{b}_{\mathrm{lab}(u,v)}^{(j)} \right)$$

Edge-wise Gating

- Edge-wise gating:

$$g_{u,v}^{(j)} = \sigma\left(\mathbf{h}_u^{(j)} \cdot \hat{\mathbf{w}}_{\mathrm{dir}(u,v)}^{(j)} + \hat{b}_{\mathrm{lab}(u,v)}^{(j)}\right)$$

$$\mathbf{h}_v^{(j+1)} = \rho\Big(\sum_{u \in \mathcal{N}(v)} g_{u,v}^{(j)} \big(W_{\mathrm{dir}(u,v)}^{(j)}\, \mathbf{h}_u^{(j)} + \mathbf{b}_{\mathrm{lab}(u,v)}^{(j)}\big)\Big)$$

which can down-weight the contribution of individual edges.

Reordering Experiment

- Reordering artificial sequences: provide an intuition for the capabilities of GCNs.
- Randomly permute tokens in sentences, pointing every token to its original predecessor with a label.
- Point every token to an arbitrary position in the sequence with a label from a distinct set of labels.

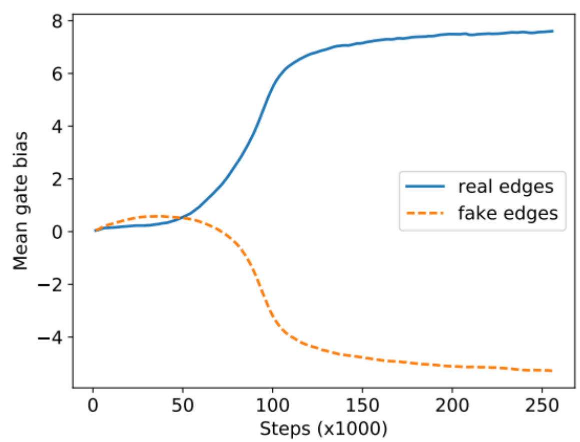


Figure 3: Mean values of gate bias terms for real (useful) labels and for fake (non useful) labels suggest the GCN learns to distinguish them.

(Bastings et al., 2017)

Main Results

- Even though RNN can already capture syntactic information, GCN is effective.

	Kendall	$BLEU_1$	$BLEU_4$
BoW	0.3352	40.6	9.5
+ GCN	0.3520	44.9	12.2
CNN	0.3601	42.8	12.6
+ GCN	0.3777	44.7	13.7
BiRNN	0.3984	45.2	14.9
+ GCN	0.4089	47.5	16.1
BiRNN (full)	0.5440	53.0	23.3
+ GCN	0.5555	54.6	23.9

Table 3: Test results for English-German.

(Bastings et al., 2017)

Effect of Sentence Length

- GCN is effective for all sentence lengths.

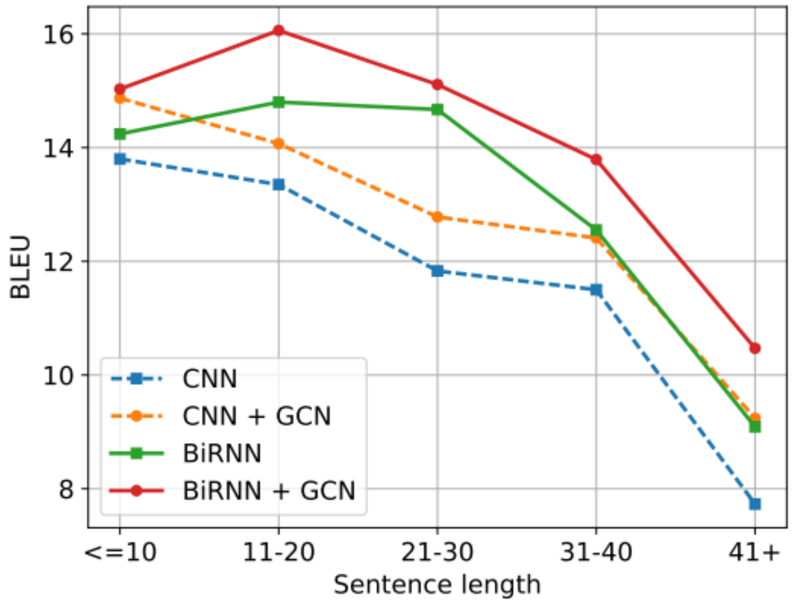


Figure 4: Validation BLEU per sentence length.

(Bastings et al., 2017)

Summary

- Their approach provides the encoder with access to rich syntactic information but let it decide which aspects of syntax are beneficial for MT, without placing rigid constraints on the interaction between syntax and the translation task.

Influence

- Citation statistics：

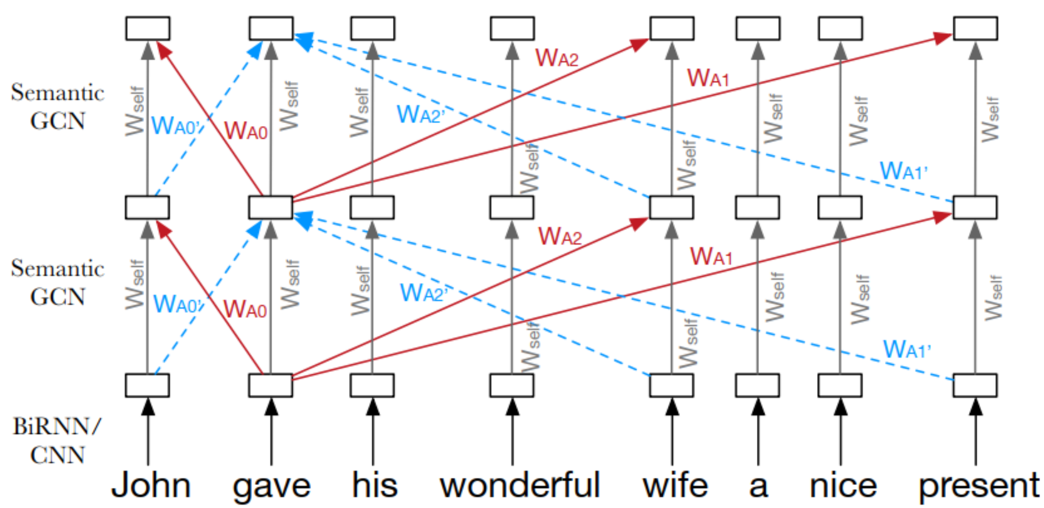
- Why is it so influential?
  - 处于GNN+NLP浪潮的前奏

Follow-up Researches

- NAACL 2018 short

Exploiting Semantics in Neural Machine Translation with Graph Convolutional Networks

Diego Marcheggiani[1,2] **Joost Bastings**[2] **Ivan Titov**[1,2]
[1]ILCC, School of Informatics, University of Edinburgh
[2]ILLC, University of Amsterdam



(Marcheggiani et al., 2018)

Follow-up Researches

- GNN + other NLP tasks
  - Text Classification
  - Relation Extraction
  - AMR-to-Text
  - Reading Comprehension
  - Question Answering
  - …

Follow-up Researches

- Other GNN algorithm + NLP
  - Gated Graph Neural Networks (GGNN)
  - Graph Attention Networks (GAT)
  - Graph Transformer
    - Structure-Aware Self-Attention

Thanks for listening