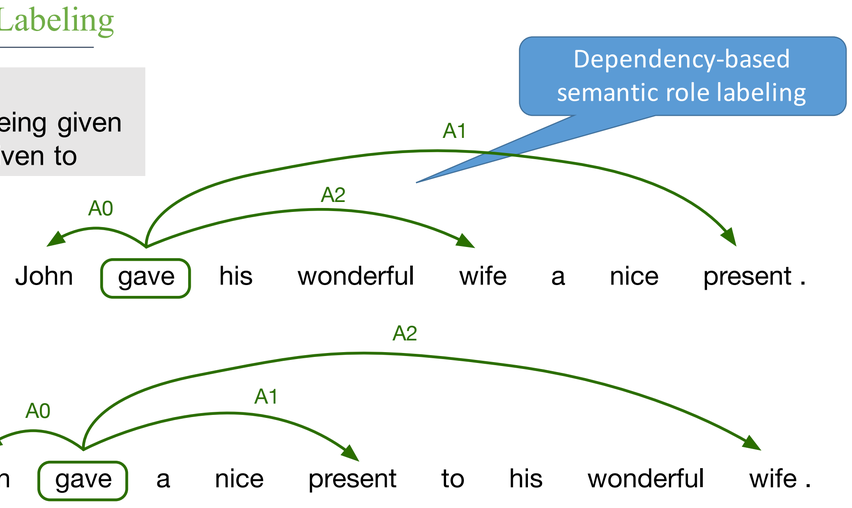
Exploiting Semantics in Neural Machine Translation with Graph Convolutional Networks

Diego Marcheggiani[1,2], Jasmijn Bastings[2], Ivan Titov[1,2]
{marcheggiani, bastings}@uva.nl, ititov@inf.ed.ac.uk
[1]ILCC, School of Informatics, University of Edinburgh
[2]ILLC, University of Amsterdam

- We use graph convolutional networks (GCNs) to inject semantic structures into neural machine translation.
- The semantic-aware model outperforms the linguistic-agnostic one (EN-DE WMT16).
- Syntax and semantics are complementary.

Semantic Role Labeling

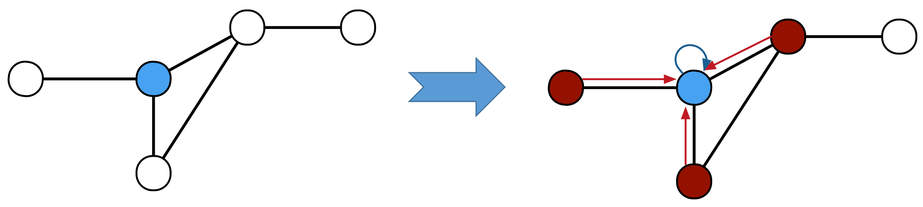
- A0: the giver
- A1: the thing being given
- A2: an entity given to



- Predicate-argument structures define an abstraction which helps to generalize over different surface realizations of the same underlying "meaning".
- We explicitly encode semantic role labeling (SRL) graphs using GCNs.

Graph Convolutional Networks

[Kipf and Welling, ICLR 2016]

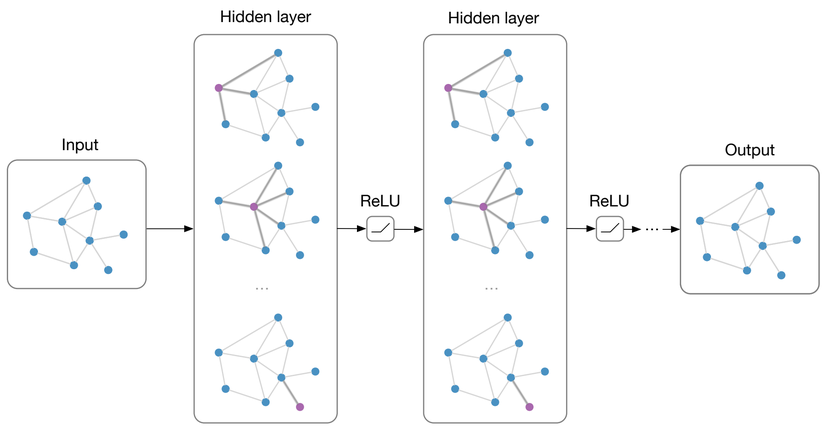


- Nodes are represented as vectors
- Messages are summed up following the graph structure
- Same W for every edge but self loop

$$\mathbf{h}_i^{(l+1)} = \sigma(\mathbf{W}_0^{(l)}\mathbf{h}_i^{(l)} + \sum_{j \in \mathcal{N}_i} \mathbf{W}_1^{(l)}\mathbf{h}_j^{(l)})$$

Self loop

Node neighborhood



Semantic GCN

Marcheggiani and Titov (2017)
Bastings et al. (2017)

- GCNs are used to encode syntactic dependencies for SRL and NMT.
- SRL graphs have directed and labeled edges.
- Message passed through different edges have different meaning.
- Parameterize each message according to syntactic label and edge direction.

$$h_v^{(k+1)} = ReLU\left(\sum_{u \in \mathcal{N}(v)} W_{L(u,v)}^{(k)} h_u^{(k)} + b_{L(u,v)}^{(k)}\right)$$

Self loop is included

Returns label and direction indices

Edge-wise gates

- Not all edges are equally important
- Gates decide the importance of each edge based on edge labels and node representations

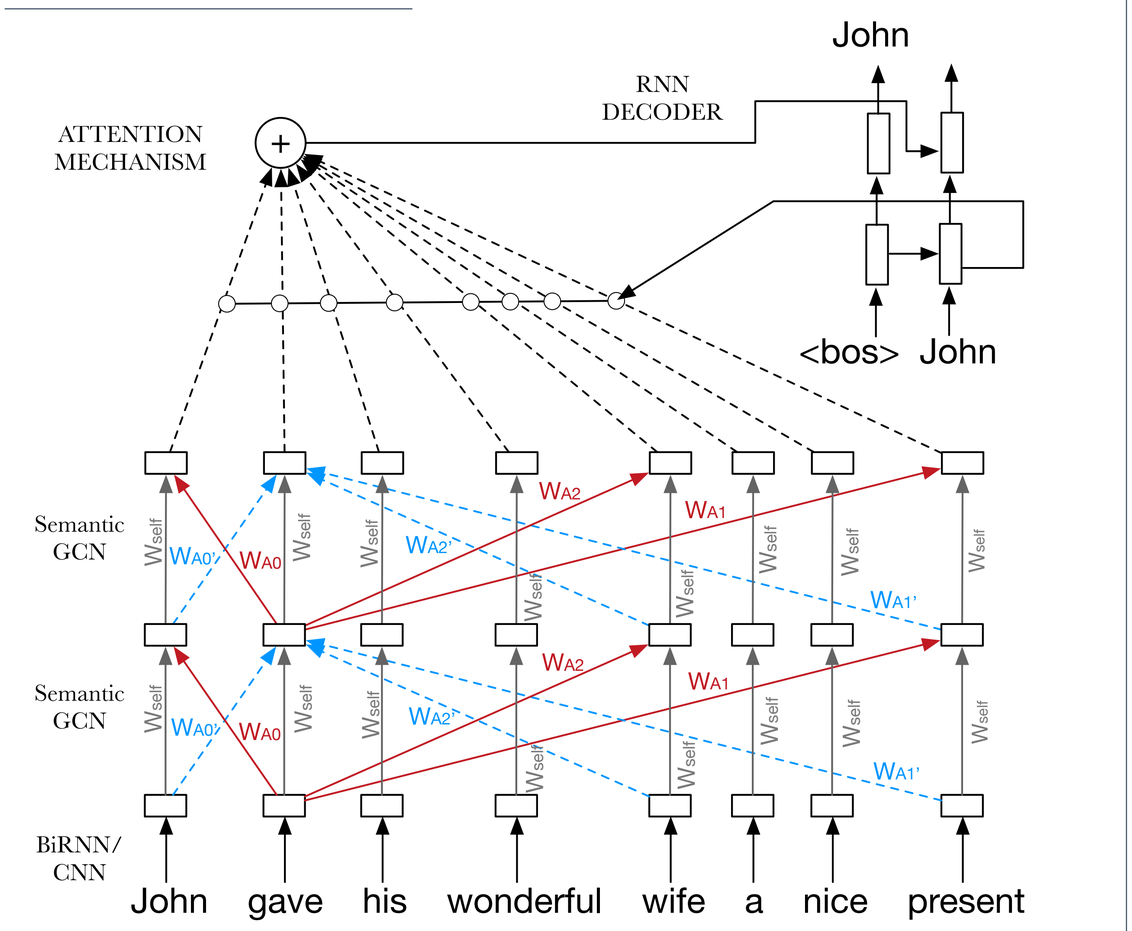
$$g_{u,v}^{(k)} = \sigma\left(h_u^{(k)} \cdot \hat{v}_{L(u,v)}^{(k)} + \hat{b}_{L(u,v)}^{(k)}\right)$$

Scalar gate

Sigmoid function

Resulting update: $h_v^{(k+1)} = ReLU(\sum_{u \in \mathcal{N}(v)} g_{v,u}^{(k)}(W_{L(u,v)}^{(k)} h_u^{(k)} + b_{L(u,v)}^{(k)}))$

GCN Encoder-Decoder Model

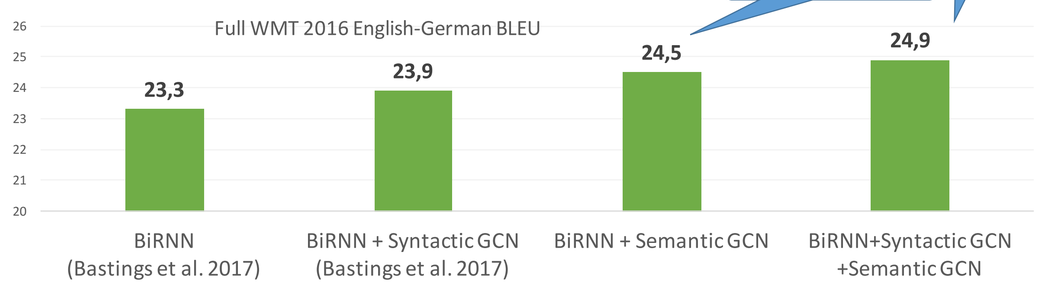


Experiments

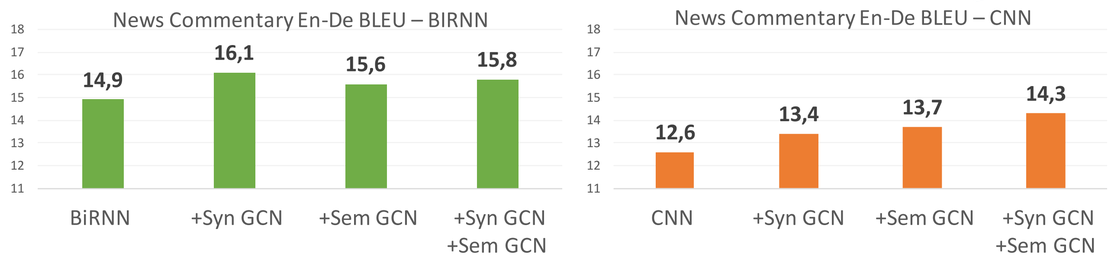
Data:
- Full WMT16 En-De (~4.5 million sentence pairs)
- News Commentary En-De (~226.000 sentence pairs)

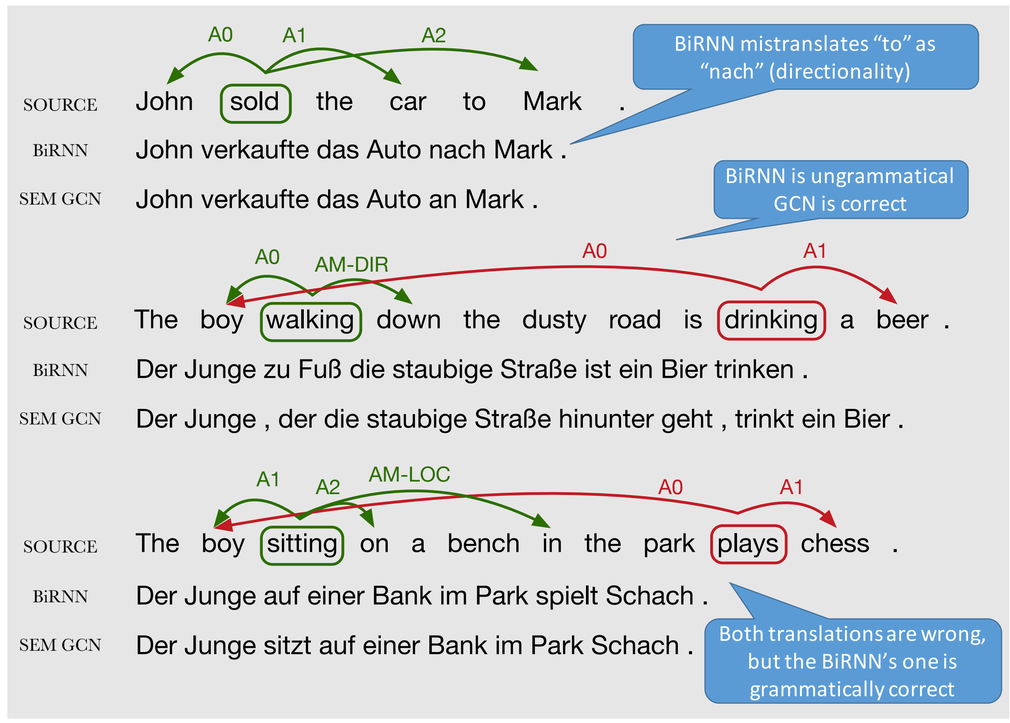
Syntax and semantics are complementary

Semantics is helpful



Weak encoders benefit more from linguistic information





References

- Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. ICLR 2016
- Diego Marcheggiani and Ivan Titov. Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling. EMNLP 2017
- Jasmijn Bastings, Ivan Titov, Wilker Aziz, Diego Marcheggiani, Khalil Sima'an. Graph Convolutional Encoders for Syntax-aware Neural Machine Translation. EMNLP 2017
- Diego Marcheggiani, Anton Frolov, Ivan Titov. A Simple and Accurate Syntax-Agnostic Neural Model for Dependency-based Semantic Role Labeling. CoNLL 2017

Acknowledgments

We thank Stella Frank and Wilker Aziz for their suggestions and comments.
The project was supported by ERC StG BroadSem 678254 grant, NWO VIDI 639.022.518 grant.
We thank NVIDIA for donating the GPUs used for this research.