Sequence-to-sequence models

used not for machine translation

Murat Apishev
Katya Artemova

Computational Pragmatics Lab, HSE

October 8, 2019

Today

Natural language understanding

Two tasks (intent detection and slot filling): identify speaker's intent and extract semantic constituents from the natural language query

Sentence	first	class	fares	from	boston	to	denver
Slots	B-class_type	I-class_type	O	O	B-fromloc	O	B-toloc
Intent	airfare						

Figure: ATIS corpus sample with intent and slot annotation

- Intent detection is a classification task
- Slot filling is a sequence labelling task

NLU datasets: ATIS [1], Snips [2]

Joint intent detection and slot filling [3]

1. The encoder models is a biLSTM
2. The decoder is a unidirectional LSTM
3. At each step the decoder state s_i is: $s_i = f(s_{i-1}, y_{i-1}, h_i, c_i)$, where $c_i = \sum_j^T \alpha_{i,j} h_j$,
$$\alpha_{i,j} = \frac{\exp(e_{i,j})}{\sum_k^T \exp(e_{i,k})},$$
$$e_{i,k} = g(s_{i-1}, h_k)$$

The inputs are explicitly aligned. Costs from both decoders are back-propagated to the encoder.

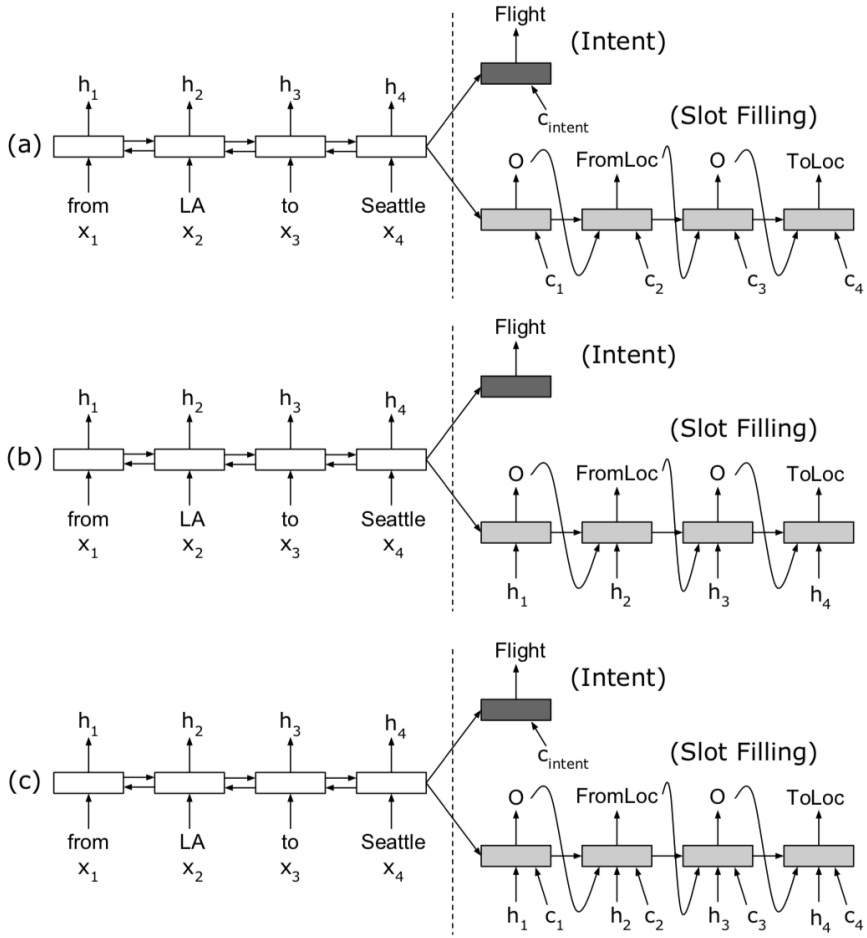


Figure: Encoder-decoder models

Joint intent detection and slot filling [3]

- BiLSTM reads the source sequence
- forward RNN models slot label dependencies
- the hidden state h_i at each step is a concatenation of the forward state fh_i and backward state bh_i
- the hidden state is h_i combined with the context vector c_i
- c_i is calculated as a weighted average of $h = (h_1, ..., h_T)$

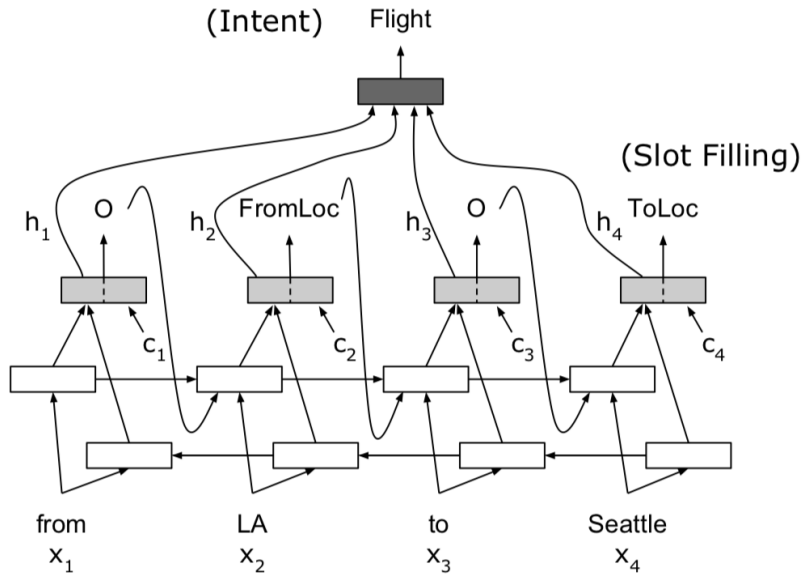


Figure: RNN-based model

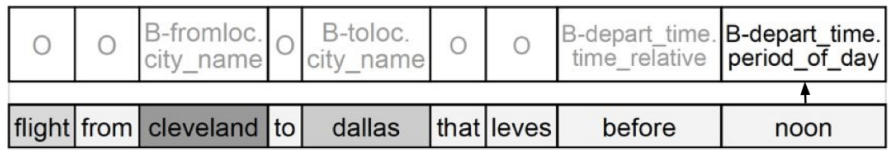
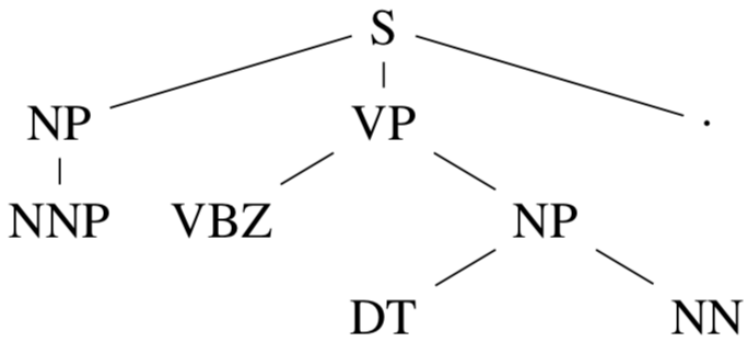


Figure: Attention weights

Today

Grammar as a Foreign Language [4]

John has a dog . →



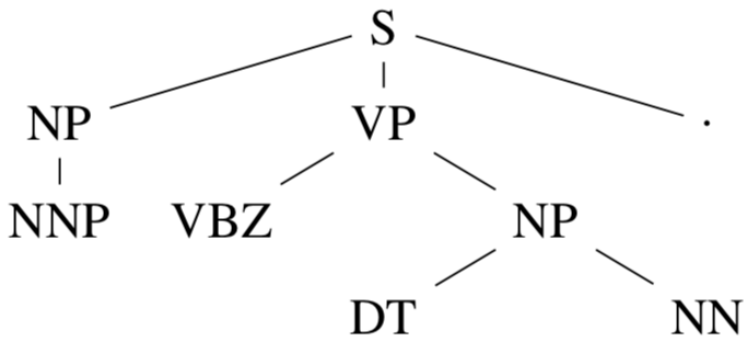
John has a dog . → (S (NP NNP)$_{\text{NP}}$ (VP VBZ (NP DT NN)$_{\text{NP}}$)$_{\text{VP}}$.)$_{\text{S}}$

Figure: Example parsing task and its linearization

Today

Grammar as a Foreign Language [4]

John has a dog . →



John has a dog . → (S (NP NNP)$_{\text{NP}}$ (VP VBZ (NP DT NN)$_{\text{NP}}$)$_{\text{VP}}$.)$_{\text{S}}$

Figure: Example parsing task and its linearization

Grammar as a Foreign Language [4]

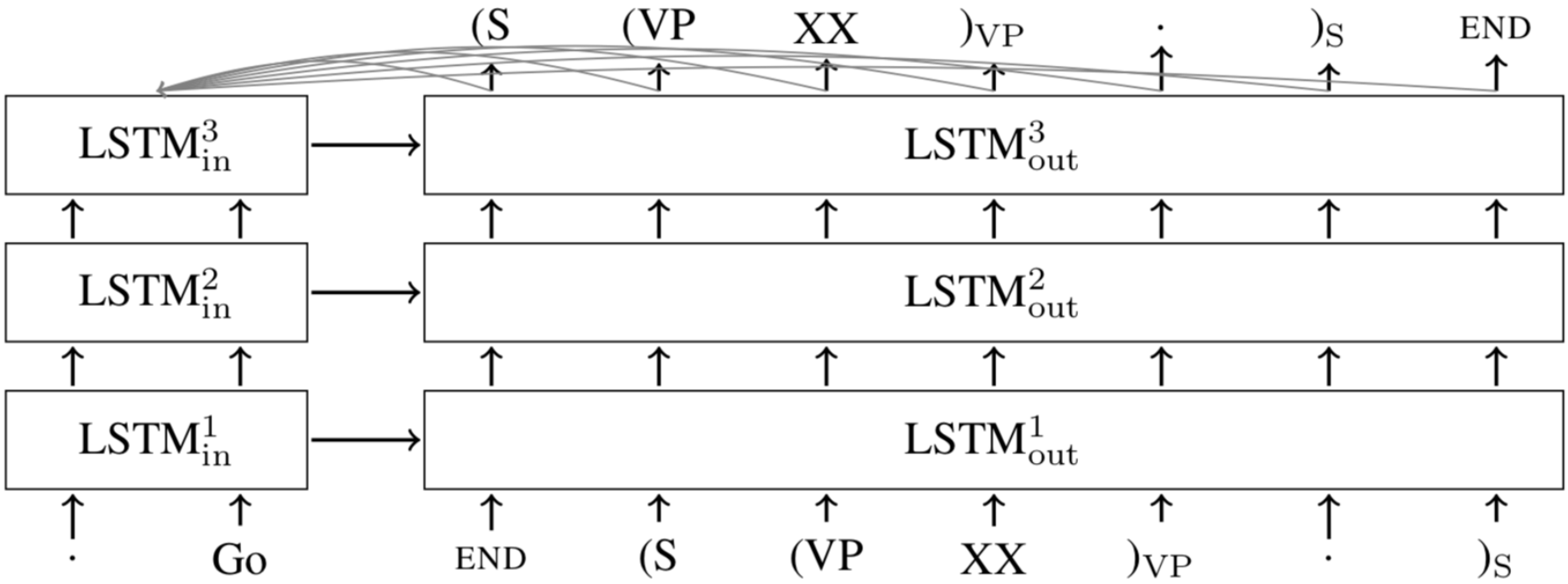


Figure: LSTM+attention encoder-decoder model for parsing

Grammar as a Foreign Language [4]

An important extension of the sequence-to-sequence model is by adding an attention mechanism. We adapted the attention model from [2] which, to produce each output symbol B_t, uses an attention mechanism over the encoder LSTM states. Similar to our sequence-to-sequence model described in the previous section, we use two separate LSTMs (one to encode the sequence of input words A_i, and another one to produce or decode the output symbols B_i). Recall that the encoder hidden states are denoted $(h_1, \ldots, h_{T_A})$ and we denote the hidden states of the decoder by $(d_1, \ldots, d_{T_B}) := (h_{T_A+1}, \ldots, h_{T_A+T_B})$.

To compute the attention vector at each output time t over the input words $(1, \ldots, T_A)$ we define:

$$\begin{aligned} u_i^t &= v^T \tanh(W_1' h_i + W_2' d_t) \\ a_i^t &= \text{softmax}(u_i^t) \\ d_t' &= \sum_{i=1}^{T_A} a_i^t h_i \end{aligned}$$

The vector v and matrices W_1', W_2' are learnable parameters of the model. The vector u^t has length T_A and its i-th item contains a score of how much attention should be put on the i-th hidden encoder state h_i. These scores are normalized by softmax to create the attention mask a^t over encoder hidden states. In all our experiments, we use the same hidden dimensionality (256) at the encoder and the decoder, so v is a vector and W_1' and W_2' are square matrices. Lastly, we concatenate d_t' with d_t, which becomes the new hidden state from which we make predictions, and which is fed to the next time step in our recurrent model.

Reference I

C. T. Hemphill, J. J. Godfrey, and G. R. Doddington, "The atis spoken language systems pilot corpus," in *Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, June 24-27, 1990*, 1990.

A. Coucke, A. Saade, A. Ball, T. Bluche, A. Caulier, D. Leroy, C. Doumouro, T. Gisselbrecht, F. Caltagirone, T. Lavril, M. Primet, and J. Dureau, "Snips voice platform: An embedded spoken language understanding system for private-by-design voice interfaces," *CoRR*, vol. abs/1805.10190, 2018. arXiv: 1805.10190. [Online]. Available: http://arxiv.org/abs/1805.10190.

B. Liu and I. Lane, "Attention-based recurrent neural network models for joint intent detection and slot filling," *arXiv preprint arXiv:1609.01454*, 2016.

Reference II

O. Vinyals, Ł. Kaiser, T. Koo, S. Petrov, I. Sutskever, and G. Hinton, "Grammar as a foreign language," in *Advances in neural information processing systems*, 2015, pp. 2773–2781.