Is It Worth the Attention?

A Comparative Evaluation of Attention Layers for Argument Unit Segmentation

Maximilian Spliethöver
Jonas Klaff
Hendrik Heuer

GENERAL

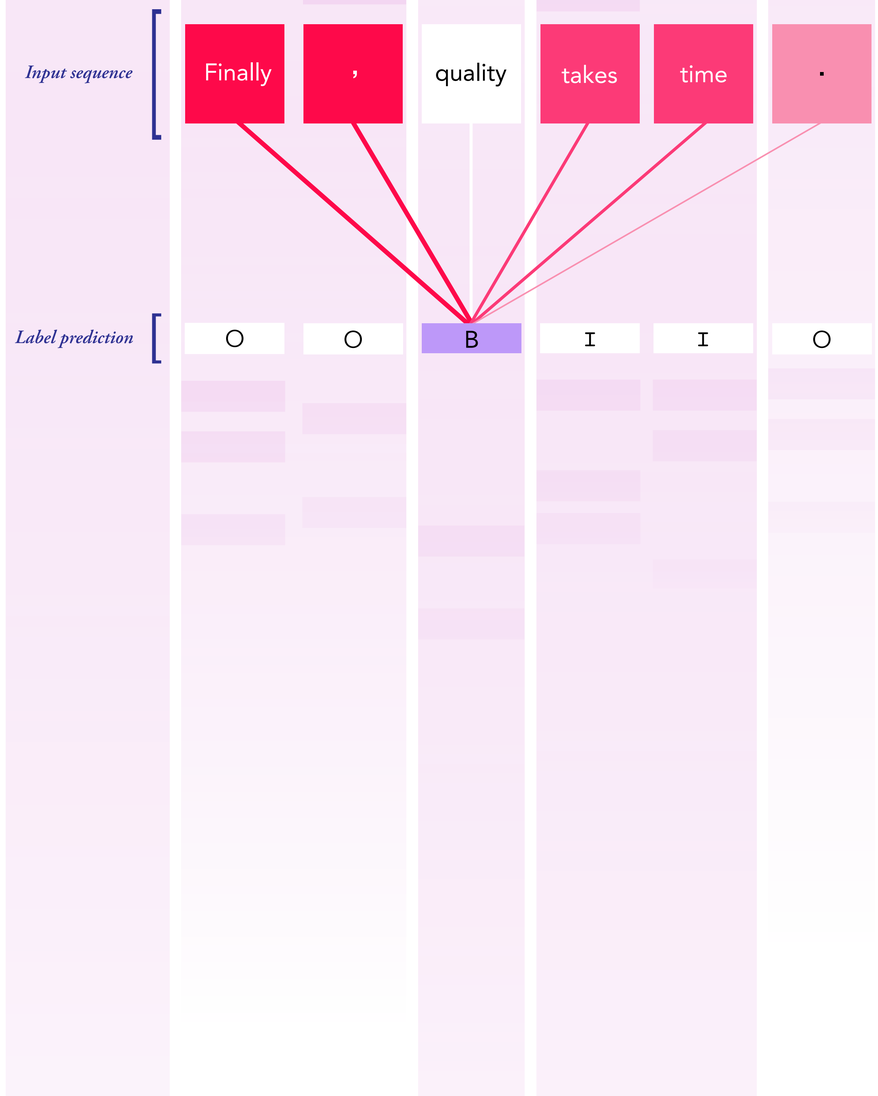
The general approach tackles the unit segmentation problem for monologic written texts. It is based on an approach by Ajjour et. al [1], with the aim to enhance the performance, as well as simplifying the preprocessing by using different input features in addition to the added attention layer. The concept of adding attention to the architecture is derived from recent advances in Neural Machine Translation [2,3]. The performance is evaluated on the "Argument annotated essays" corpus compiled by Stab, et al. [6].

BERT

An attention-based pre-trained language model [5], derived from the Encoder part of the Transformer [2] architecture. Embeddings are generated by feeding one sentence into the language model and extracting the representations for the first subword token from the last four layers of the pre-trained BERT model via the Flair library. The subtoken embedding is then used as representations for the whole token.

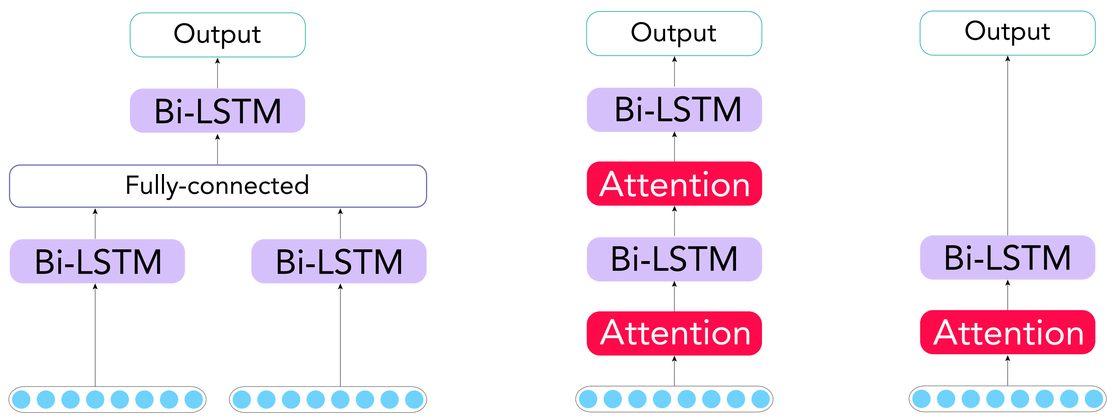
FLAIR

A pre-trained character-level Bi-LSTM [4] to predict the most probable next character for a given input sequence. Embeddings are generated by feeding one sentence into the language model, extracting the corresponding weights via the Flair library and stack them with GloVe300 embeddings.



- *calculates a relevance score for each part of an input sequence*
- *self-attention: with respect to the complete input vector of a single token*
- *multi-head attention: with respect to distinct sub-spaces of the token's input vector*

ARCHITECTURES



Baseline | *Baseline with attention* | *Single BiLSTM*

RESULTS

Embedding	Baseline weighted F1	Baseline with attention weighted F1	Single BiLSTM weighted F1
GloVE	0.86	**0.85** 0.67	**0.86** **0.84**
BERT	0.83	0.68 **0.68**	0.86 0.83
Flair	**0.87**	0.67 0.67	0.86 0.81

input attention
error attention

REFERENCES

[1] Y. Ajjour, W.-F. Chen, J. Kiesel, H. Wachsmuth, and B. Stein, "Unit Segmentation of Argumentative Texts," in Proceedings of the 4th Workshop on Argument Mining, Copenhagen, Denmark, 2017, pp. 118–128.

[2] A. Vaswani et al., "Attention Is All You Need," arXiv:1706.03762 [cs], Jun. 2017.

[3] Y. Wu et al., "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation," arXiv:1609.08144 [cs], Sep. 2016.

[4] A. Akbik, D. Blythe, and R. Vollgraf, "Contextual String Embeddings for Sequence Labeling," p. 12.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv:1810.04805 [cs], Oct. 2018.

[6] Stab, Christian, and Iryna Gurevych. "Parsing Argumentation Structures in Persuasive Essays." Computational Linguistics 43, no. 3 (June 9, 2017): 619–59. https://doi.org/10.1162/COLI_a_00295.