An explainability analysis of a sentiment prediction task using a transformer-based attention filter

Neşet Özkan Tan (**speaker**), Joshua Bensemann, Diana Benavides-Prado, Yang Chen, Mark Gahegan, Lia Lee, Alex Yuxuan Peng, Patricia Riddle, Michael Witbrock

November 17, 2021

STRONG AI LAB

Introduction

Skim Reading as a Human

- The reader does not attend to every word fully.
- Learns to skip words of lower importance.
- About 40% of the words can be skipped without substantial loss of understanding [Hahn and Keller, 2018].

Skim Reading as a Machine

- [Yu et al., 2017] found that the model could skip over several words at a time and still be as accurate or more accurate than the non-skipped models.
- [Hahn and Keller, 2018] showed that you could model the skipping processes using human eye movements and achieve the same result for sentiment analysis tasks.

The Transformer [Vaswani et al., 2017]

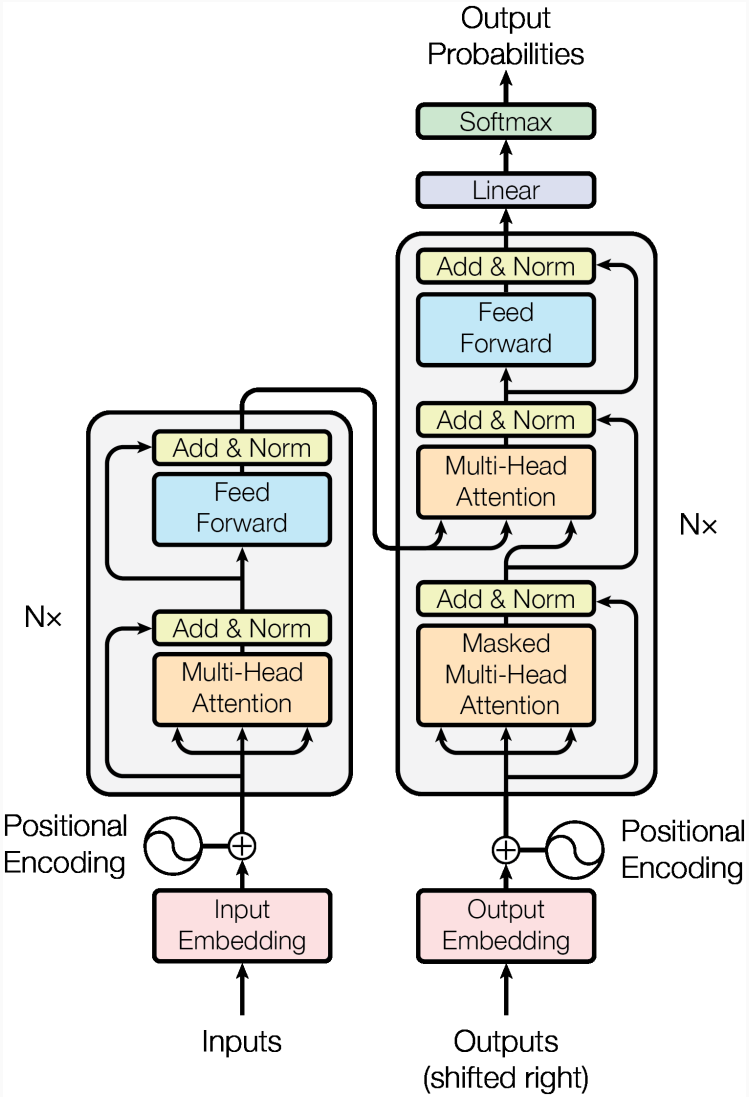

Figure 1: The transformer architecture in [Vaswani et al., 2017]

Transformers Inputs

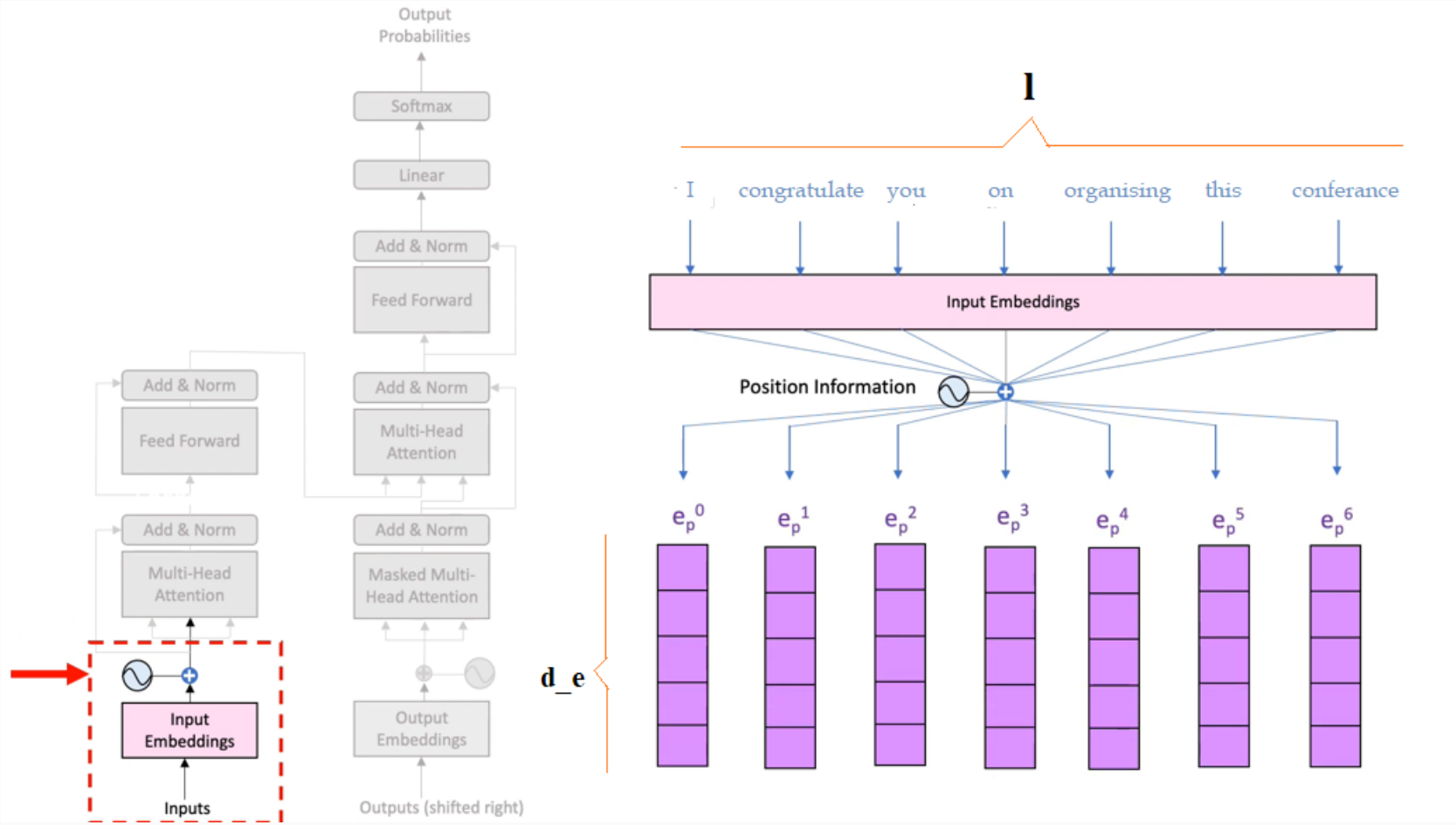

Figure 2: The input of the transformer.

Self-Attention

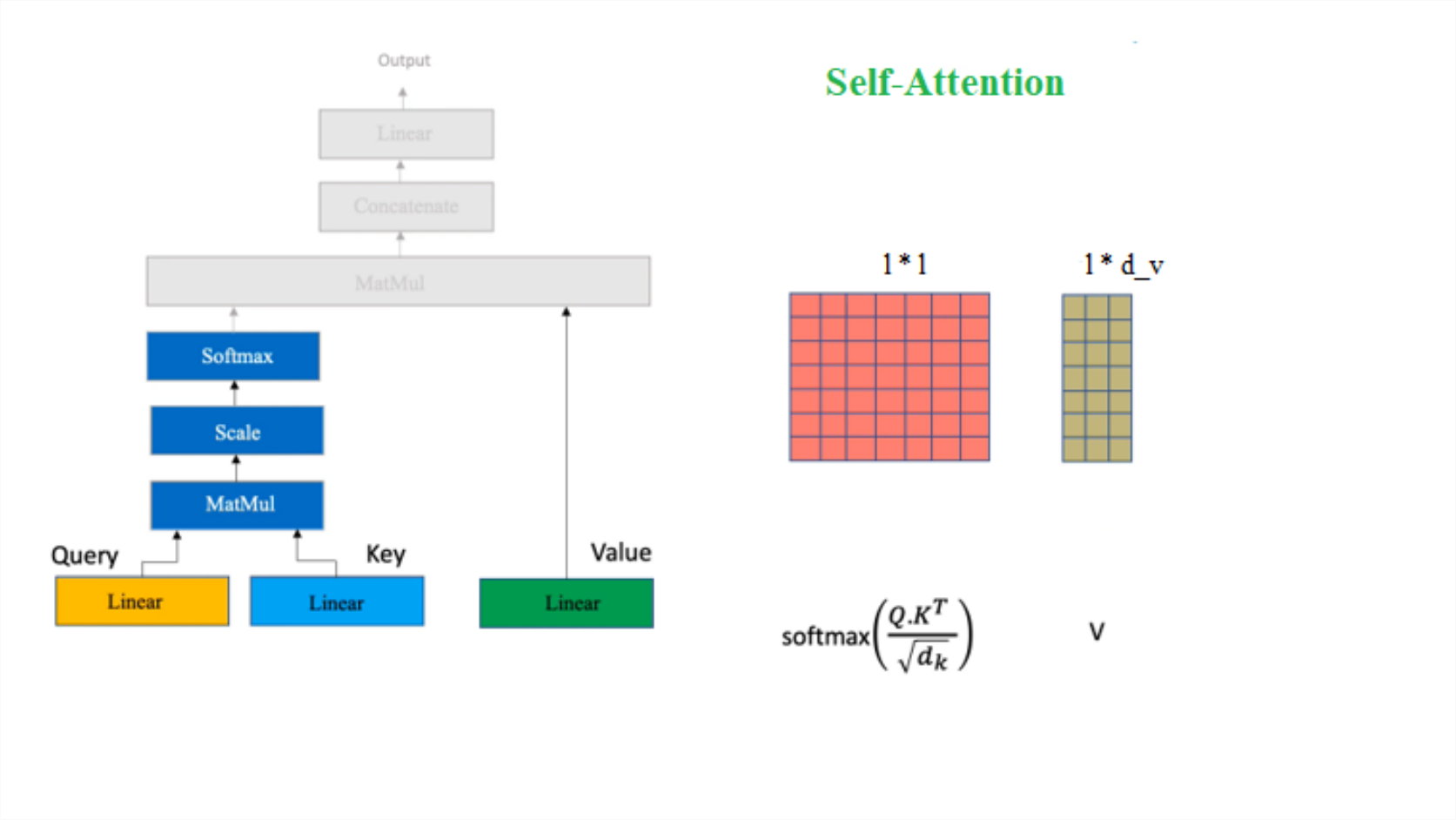

Figure 3: Self-Attention

Experiments

Model: BERT [Devlin et al., 2018]

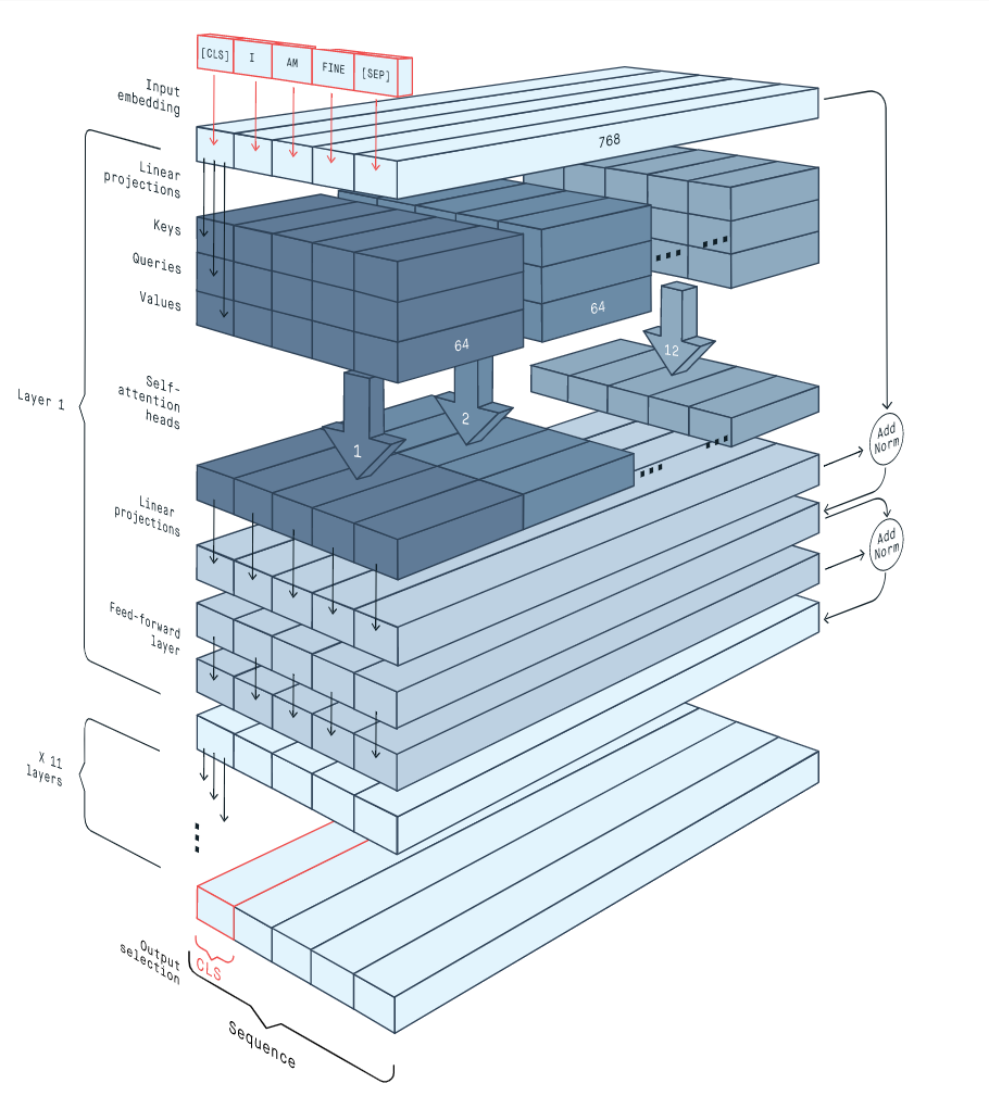

Figure 4: BERT-base architecture.(image source:peltarion.com)

Data : IMDB [Maas et al., 2011]

Text

The movie Titanic makes it much more than just a " night to remember. " It re writes a tragic history event that will always be talked about and will never been forgotten. Why so criticised? I have no idea. Could / will they ever make a movie like Titanic that is so moving and touching every time you watch it. Could they ever replace such an epic masterpiece? It will be almost impossible. The director no doubt had the major impact on the film. A simple disaster film (boring to watch) converted to an unbelievable romance. Yes, I'm not the Romance type either, but that should not bother you, because you will never see a romance like this. Guaranteed! Everything to the amazing effects, to the music, to the sublime acting. The movie creates an amazing visual and a wonderful feeling. Everything looks very real and live. The legend herself " TITANIC " is shown brilliantly in all classes, too looks, too accommodation. The acting was the real effect. DiCaprio and Winslet are simply the best at playing their roles. No one could have done better. They are partly the reason why the film is so great. I guess it 'not too much to talk about. The plot is simple, the acting is brilliant, based on a true story, probably more than half of the consumers that watch the film will share tears, thanks to unimaginable ending which can never be forgotten. Well, if you haven't seen this film, you're missing out on something Historical, and a film to idolise for Hollywood. Could it get better? No. Not at all. The most moving film of all time, don't listen to people, see for yourself then you will understand. A landmark. (don't be surprised if you cry too)

Label

'positive' : 1

Figure 5: A Data point.

Experiments

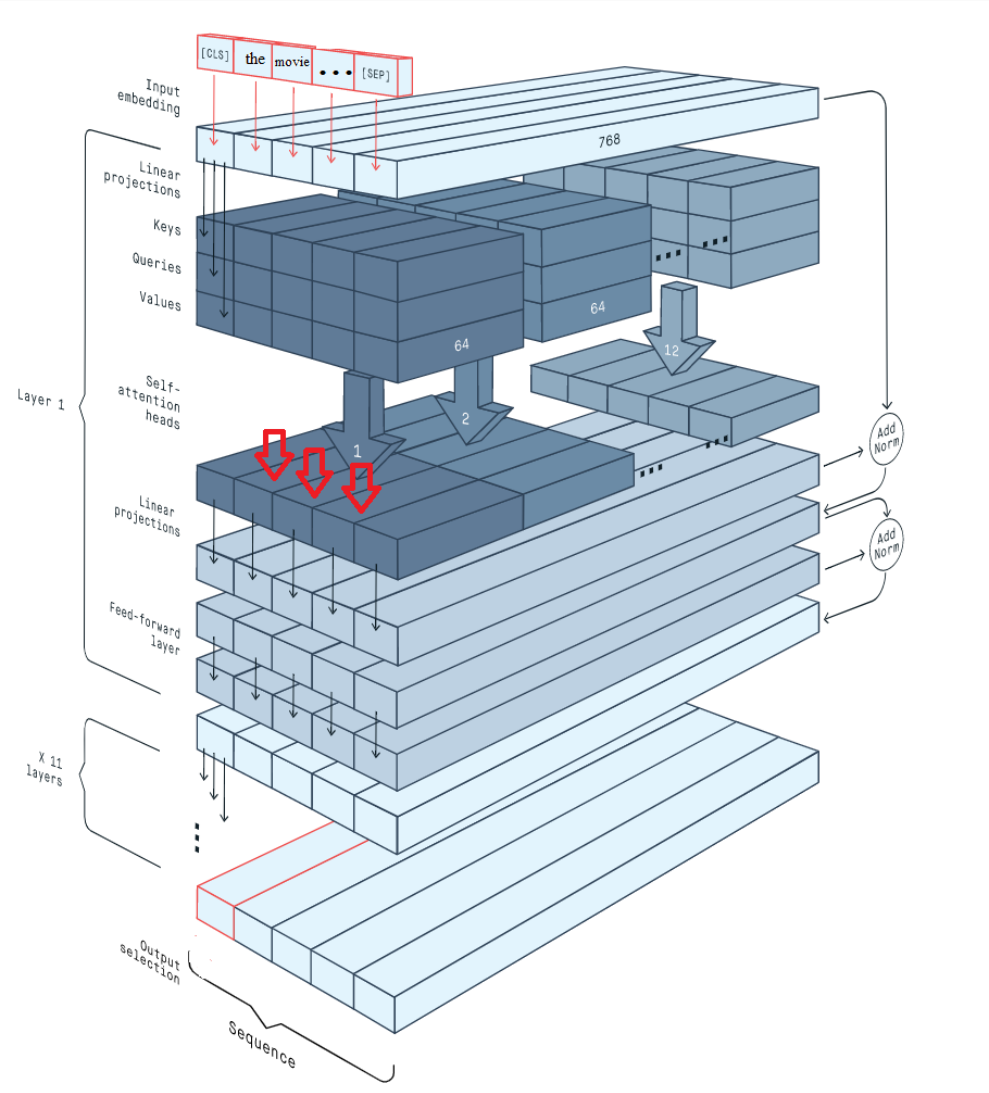

Figure 6: Sum of attention scores of each token from 1st and 12th layers.

Experiment1

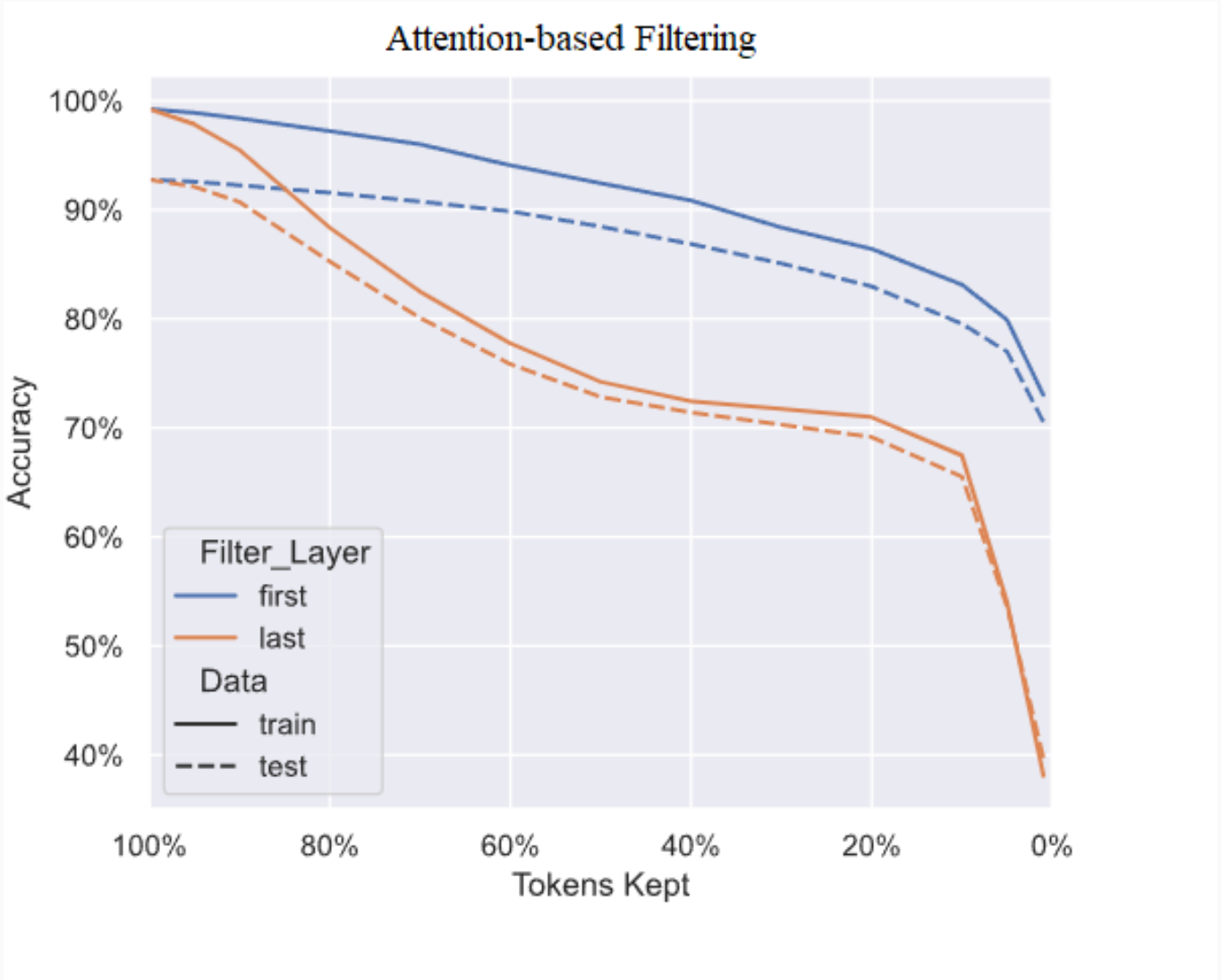

Figure 7: Sentiment prediction accuracy of the text when only top x percentile tokens are kept.

Experiment2

The movie Titanic makes it much more than just a " night to remember. " It re writes a tragic history event that will always be talked about and will never been forgotten. Why so criticised? I have no idea. Could / will they ever make a movie like Titanic that is so moving and touching every time you watch it. Could they ever replace such an epic masterpiece? It will be almost impossible. The director no doubt had the major impact on the film. A simple disaster film (boring to watch) converted to an unbelievable romance. Yes, I'm not the Romance type either, but that should not bother you, because you will never see a romance like this. Guaranteed! Everything to the amazing effects, to the music, to the sublime acting. The movie creates an amazing visual and a wonderful feeling. Everything looks very real and live. The legend herself " TITANIC " is shown brilliantly in all classes, too looks, too accommodation. The acting was the real effect. DiCaprio and Winslet are simply the best at playing their roles. No one could have done better. They are partly the reason why the film is so great. I guess it 'not too much to talk about. The plot is simple, the acting is brilliant, based on a true story, probably more than half of the consumers that watch the film will share tears, thanks to unimaginable ending which can never be forgotten. Well, if you haven't seen this film, you're missing out on something Historical, and a film to idolise for Hollywood. Could it get better? No. Not at all. The most moving film of all time, don't listen to people, see for yourself then you will understand. A landmark. (don't be surprised if you cry too)

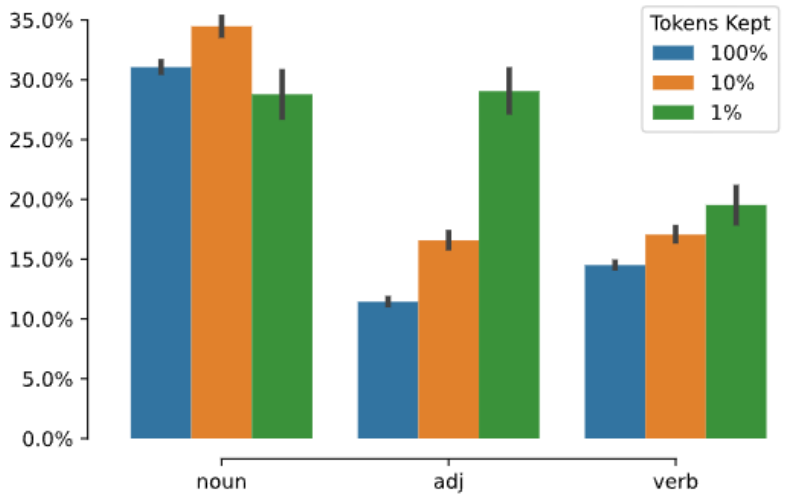

Figure 8: The top 10% of tokens based on the first layer are colored, where greens are adjectives. The second figure shows how POS changes during filtering.

Conclusion and Future Work

Summary

- We show that BERT's first layer attention can be used as a filter that gives a remarkably effective sequence selection for the sentiment analysis task.
- We show that the distribution of the parts of speech chosen by the filter changes as the number of filtered tokens increases. We also show that adjectives are the most persistent parts of speech in the filtering progress for the sentiment analysis task.

Summary and Future Work

- How much content may be needed for a sentiment analysis task.
- As future work, we intend to benefit from similar inductive biases in order to reduce the long-distance-dependency costs for other downstream tasks.

Thank you!

Questions?

References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Michael Hahn and Frank Keller. Modeling task effects in human reading with neural attention. *arXiv preprint arXiv:1808.00054*, 2018.

Andrew L Maas, Raymond E Daly, Peter T Pham, Dan Huang, Andrew Y Ng, and Christopher Potts. Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies. *Learning Word Vectors for Sentiment Analysis*, pages 142–150, 2011.

References ii

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

Adams Wei Yu, Hongrae Lee, and Quoc V Le. Learning to skim text. *arXiv preprint arXiv:1704.06877*, 2017.