PSY9511: Seminar 8

Sequence modelling (with an emphasis on language)

Esten H. Leonardsen

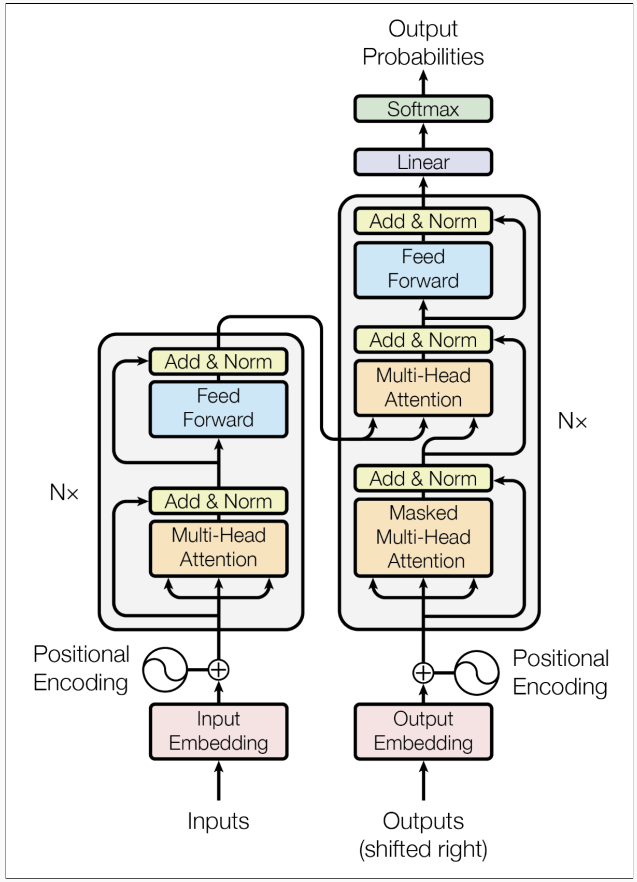
13.05.24

UNIVERSITETET
I OSLO

Transformers

UNIVERSITETET
I OSLO

Transformers



Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30

Transformers: Attention

The movie was great, the actors were_____

Transformers: Attention

The movie was great, the actors were _____

Transformers: Attention

The movie was great, the actors were _____

Transformers: Attention

The movie was great, the actors were _____

Transformers: Attention

The movie was great, the actors were _____

Transformers: Attention

The movie was great, the actors were _____

Transformers: Attention

The movie was great, , the actors were _____

Transformers: Attention

The movie was great, the actors were _____

Transformers: Attention

The movie was great, the actors were _____

Transformers: Attention

The movie was great, we saw it at the new Cinema in the city center, right down by the restaurant where we went for my birthday that one year, the one where the clown was inside the cake, the actors were _____

Transformers: Attention

The movie was great, we saw it at the new Cinema in the city center, right down by the restaurant where we went for my birthday that one year, the one where the clown was inside the cake, the actors were _____

Transformers: Attention

The movie was great, we saw it at the new Cinema in the city center, right down by the restaurant where we went for my birthday that one year, the one where the clown was inside the cake, the actors were _____

Transformers: Attention

The movie was great, the actors were_____

Transformers: Attention

The movie was great, the actors were _____

[8 7 9 6 0 8 4 10] → ?

Transformers: Attention

The movie was great, the actors were _____

$$[8 \quad 7 \quad 9 \quad 6 \quad 0 \quad 8 \quad 4 \quad 10] \rightarrow ?$$

$$[0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0]$$

Transformers: Attention

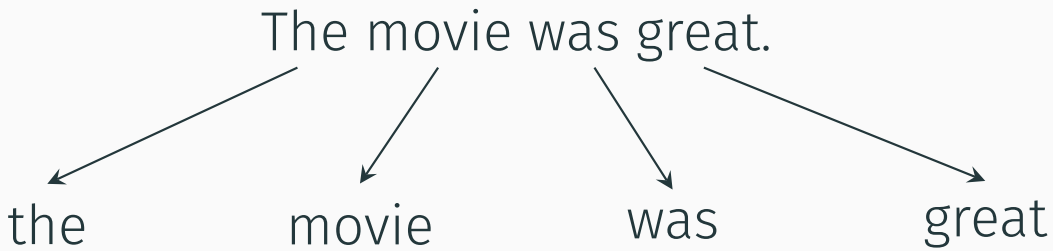
The movie was great, the actors were _____

$$
\begin{array}{llllllll}
[8 & 7 & 9 & 6 & 0 & 8 & 4 & 10] \rightarrow ? \\
 & & & \times & & & & \\
[0 & 0 & 0 & 1 & 0 & 0 & 0 & 0] \\
 & & & = & & & & \\
[0 & 0 & 0 & 6 & 0 & 0 & 0 & 0]
\end{array}
$$

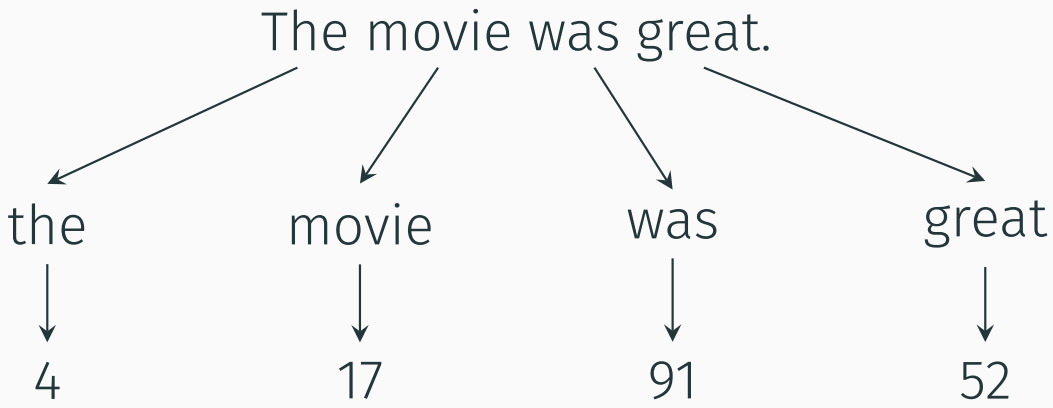
Transformers: Positional encoding

The movie was great.

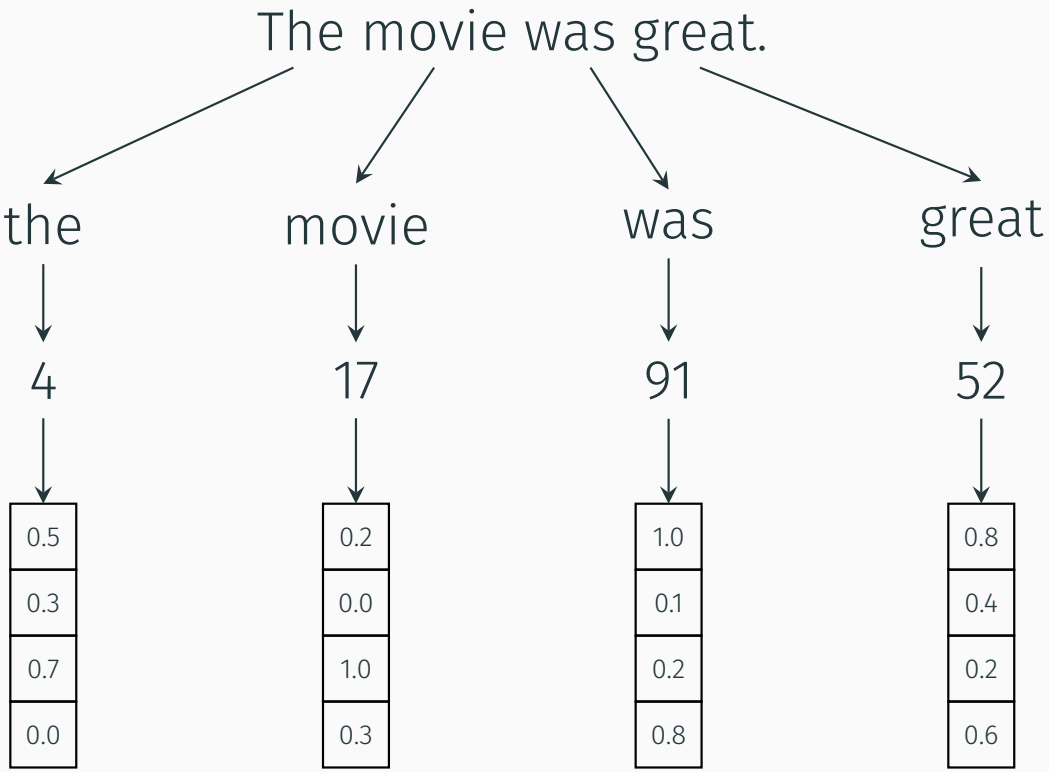
Transformers: Positional encoding



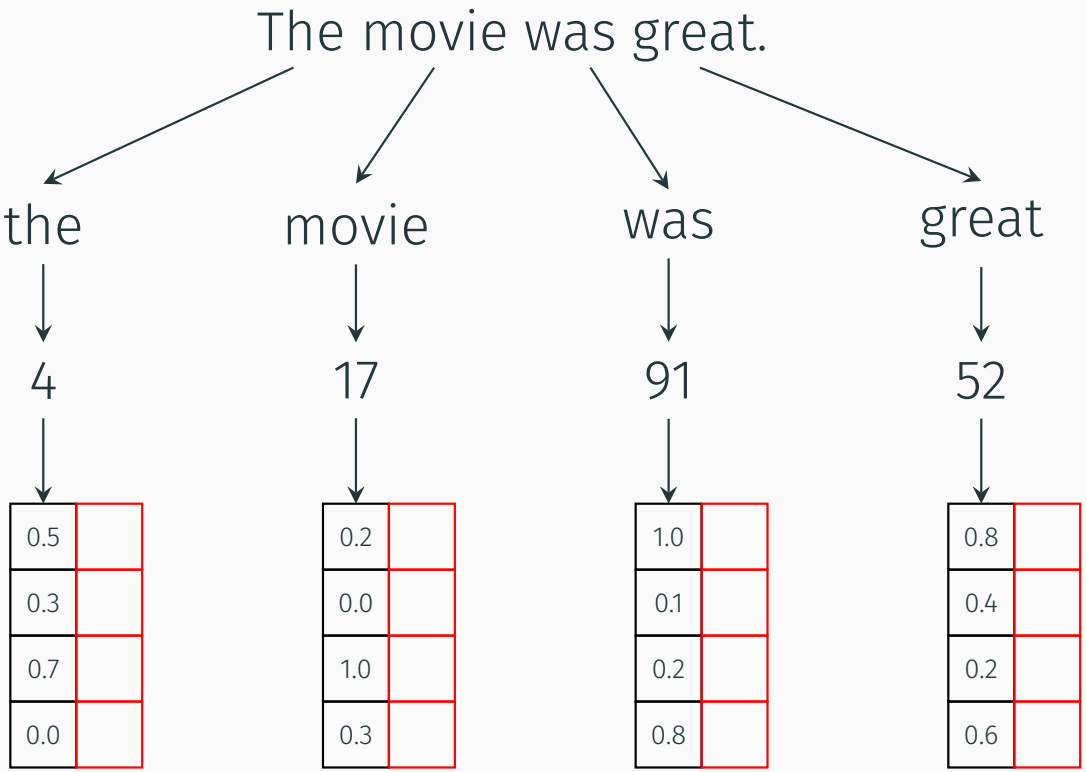
Transformers: Positional encoding



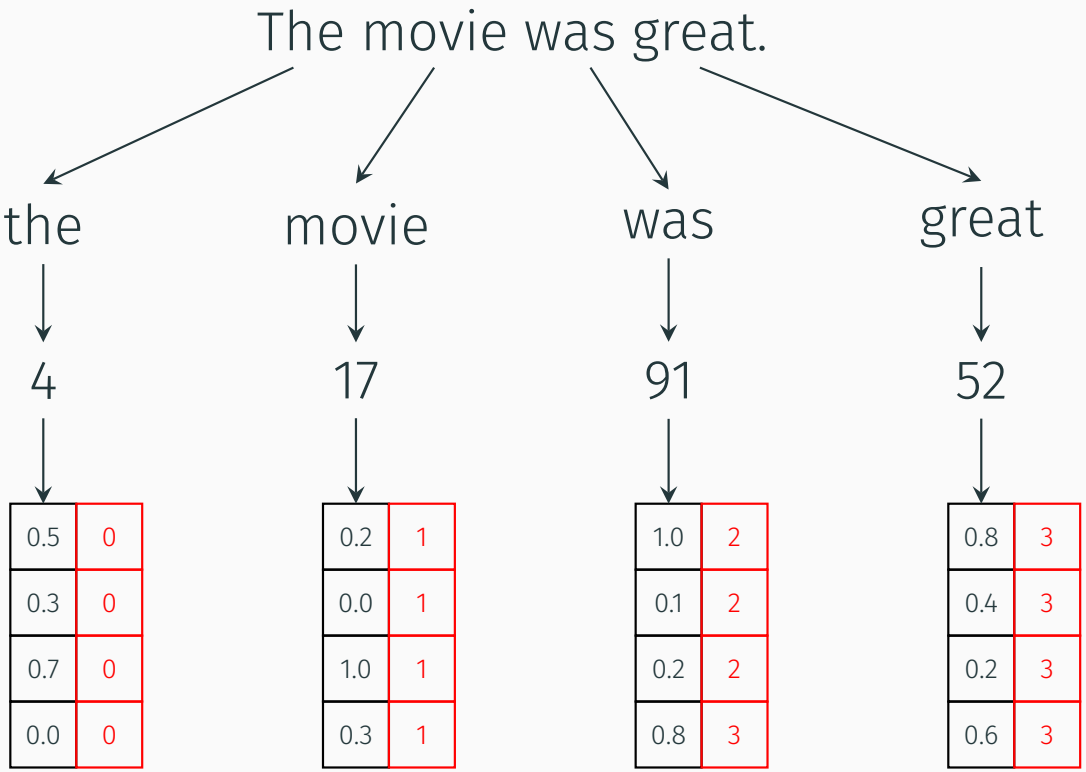
Transformers: Positional encoding



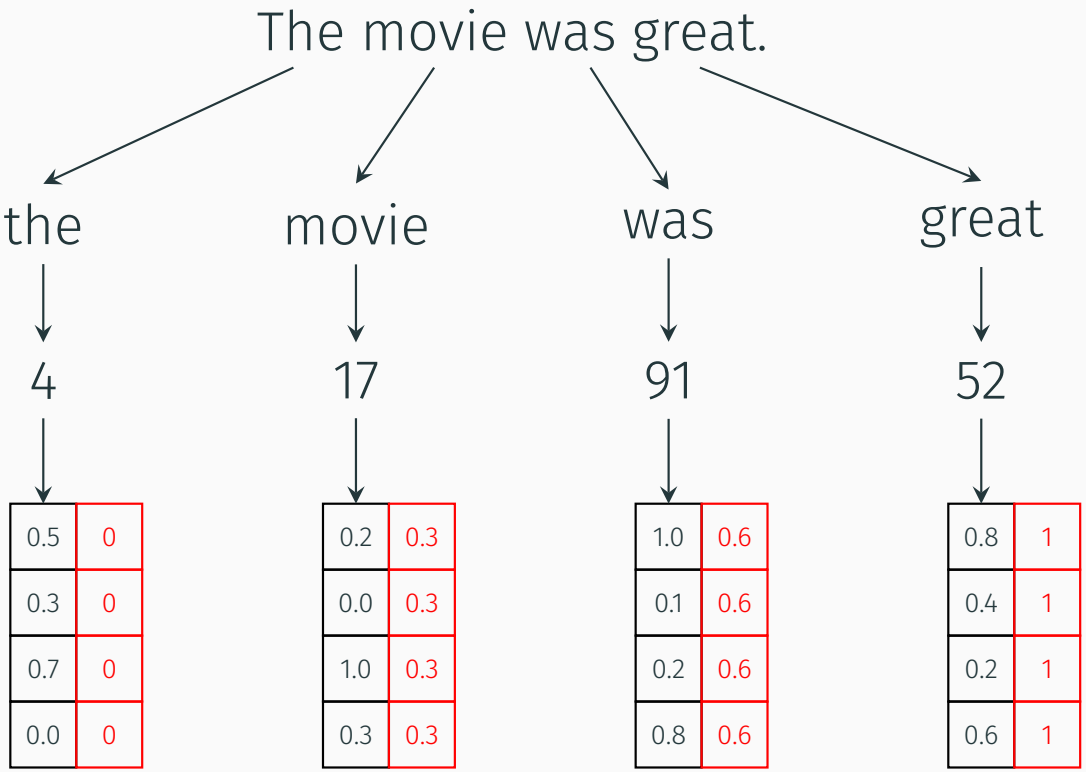
Transformers: Positional encoding



Transformers: Positional encoding



Transformers: Positional encoding



Transformers: Embedding

Transformers: Demo

https://huggingface.co/docs/transformers/model_doc/llama2

Transformers: Demo

http://localhost:8888/notebooks/notebooks/GPT%20Embedding%20demo.ipynb

Transformers

Transformers: Revolutionized language modelling by combining feed forward neural networks with multihead attention and positional endcodings (and infinite data and compute)

- Main advantage: Outperforms everything else for almost all language modelling tasks
- Main disadvantage: Can either be used locally, which is fidgety and requires a good computer, or via an API, which is costly and gives others access to your data