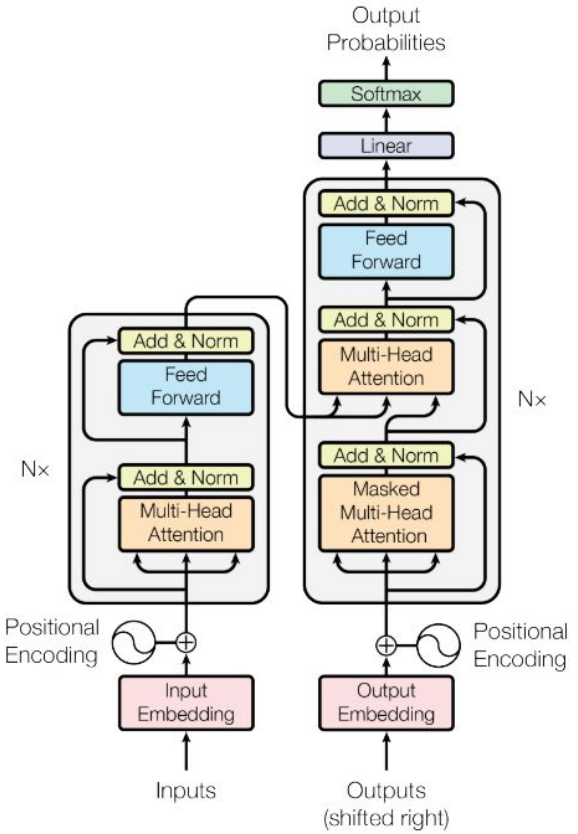
Transformers in NLP

-MSDS 631 Final Project

Janson Tao, Yueling Wu, Dawn(Shuyan) Li

Our Goals

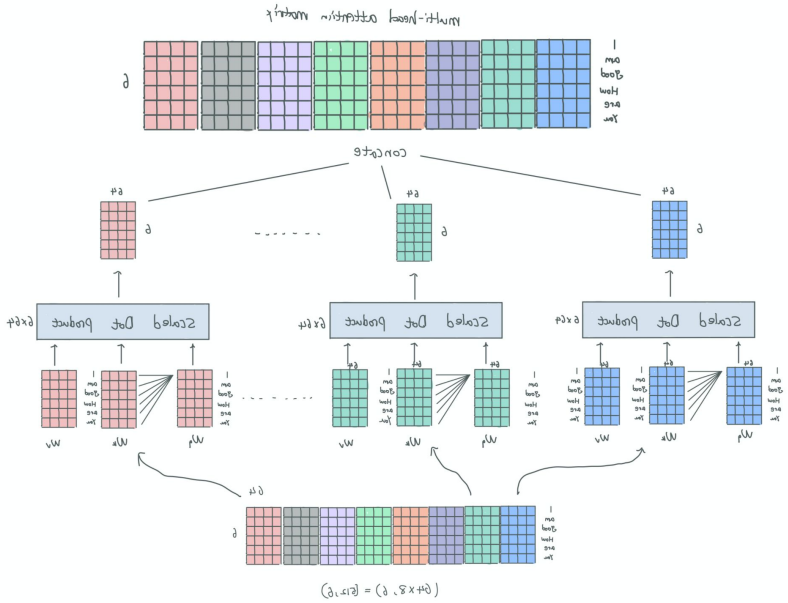
- Implement Transformer Model from Scratch
- Detailed Step by Step Instruction



Our Methods

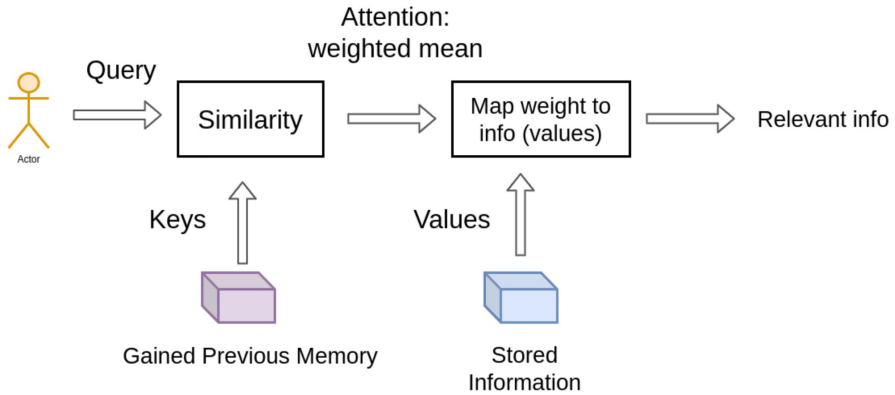
- Input Embedding Trick
- Positional Embedding
- Attention and self-attention Mechanism
- Add & Norm Block
- Feed Forward Block

What went wrong?



- Chunk matrix (dim = 512) into 8 parts and perform linear transformation on each individual part will take the model forever for training.
- **Solution**: Expand the the last dimension into the 4th dimension (10x faster).
- **eg.** *batch_size, sent_len, d_model) => (batch_size, sent_len, h * d_k)*

What went right?



- **Query** - asking for information;
- **Key** - saying that it has some information;
- **Value** - giving the information.

Many resource on web don't explain why the model uses Query, Keys and Values.

What's next

- Further improve model efficiency, reduce training time.
- Implement on real-life data, adjust parameters to compare the model performance.