
Combining Contextual Words Embeddings and Knowledge Graph Embeddings

IDMC, University of Lorraine

18/12/2020

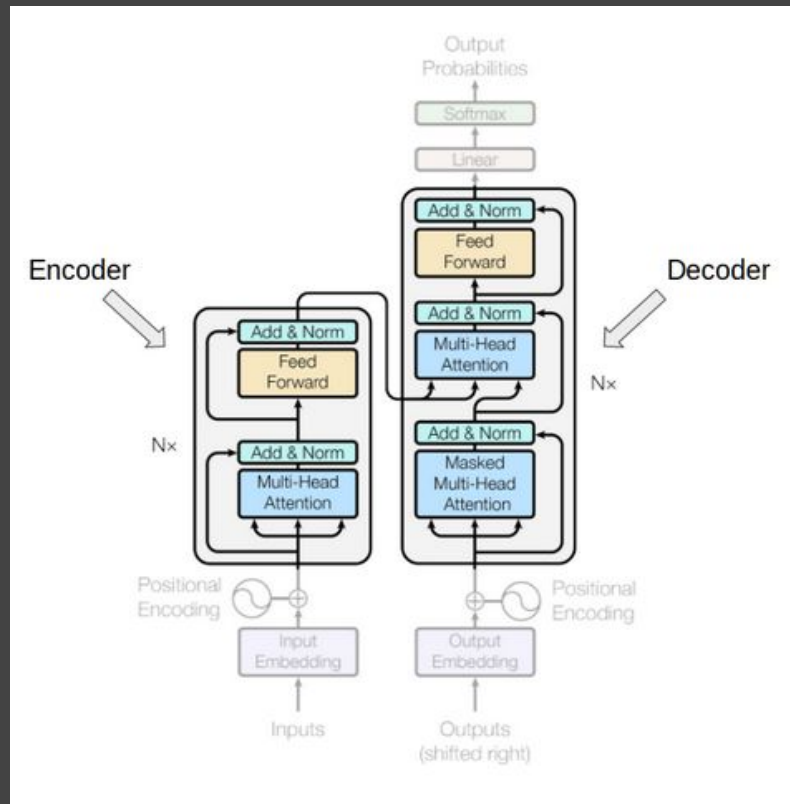
Outline

1. Background
2. Recap
3. Results

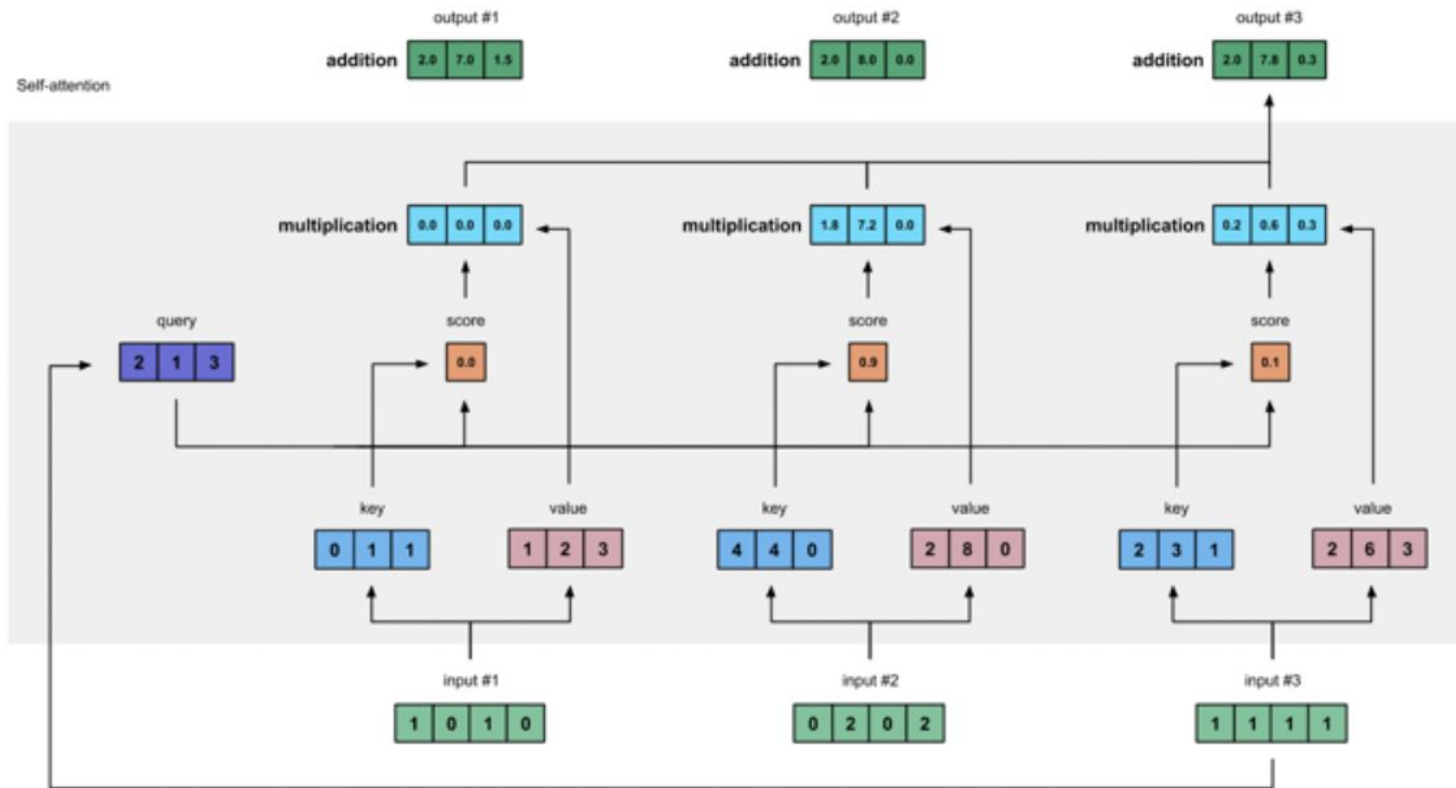
Background

What is Transformer ?

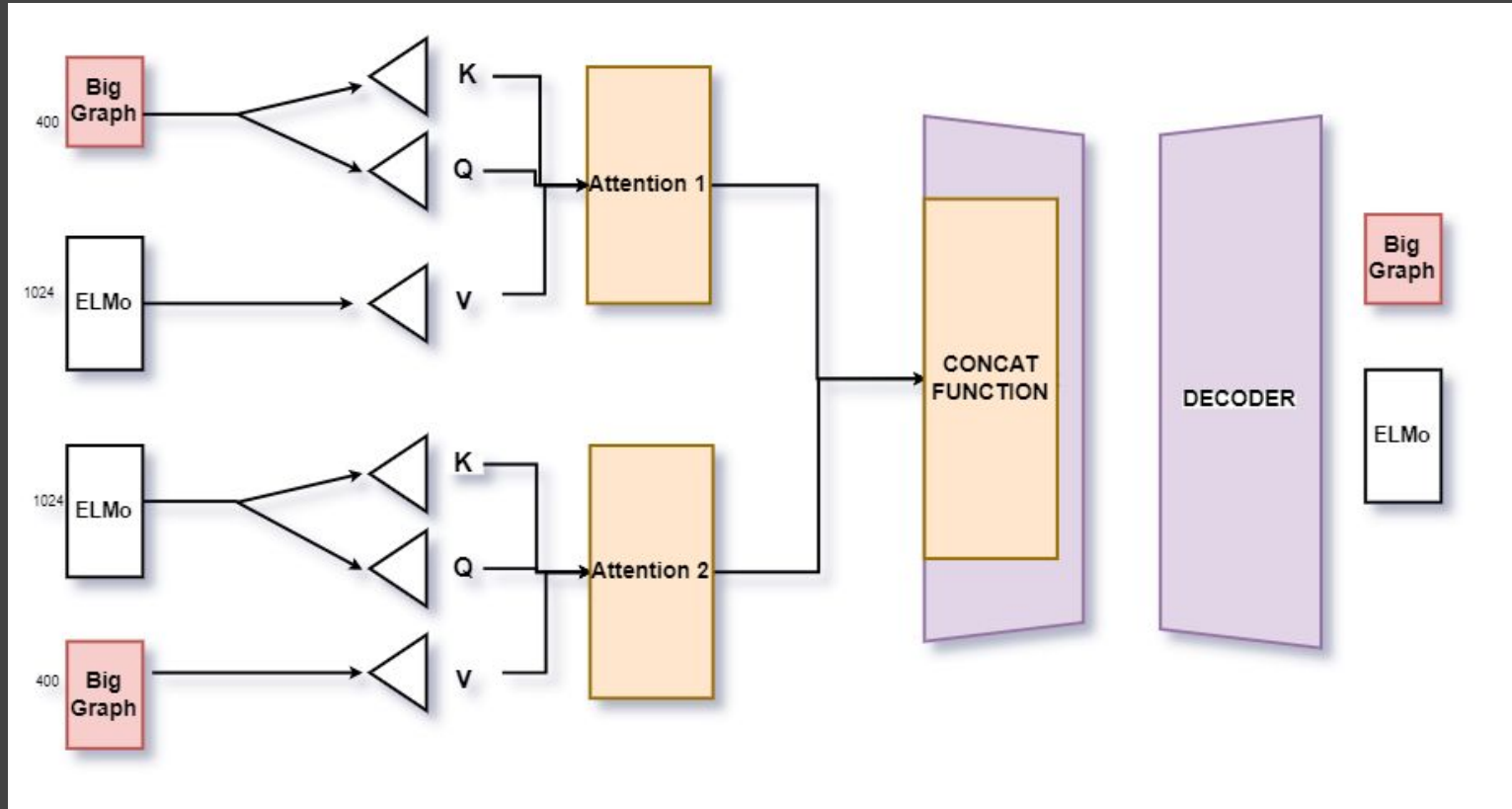
- + **Transformer:** The first transduction model relying entirely on self-attention to compute representations of its input and output
- + The decoder have an additional layer of multi-head attention => focus on appropriate part of the input sequence



INSIDE A MULTI-HEAD SELF-ATTENTION

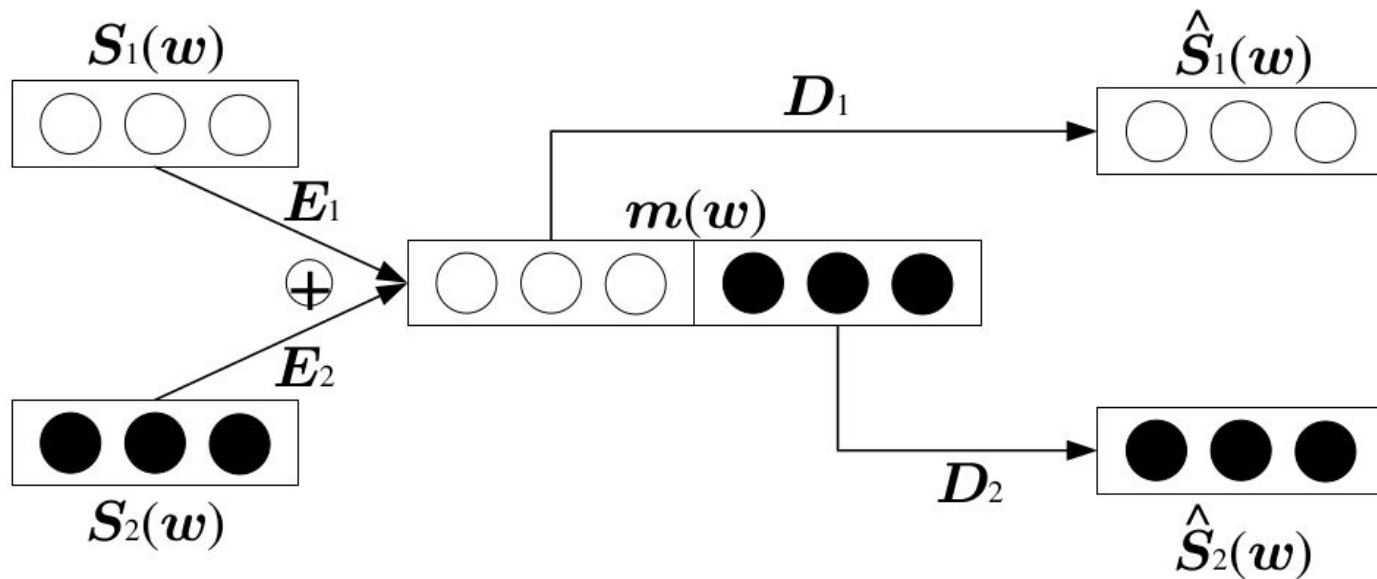


“Parallel” Multi Head Self Attention

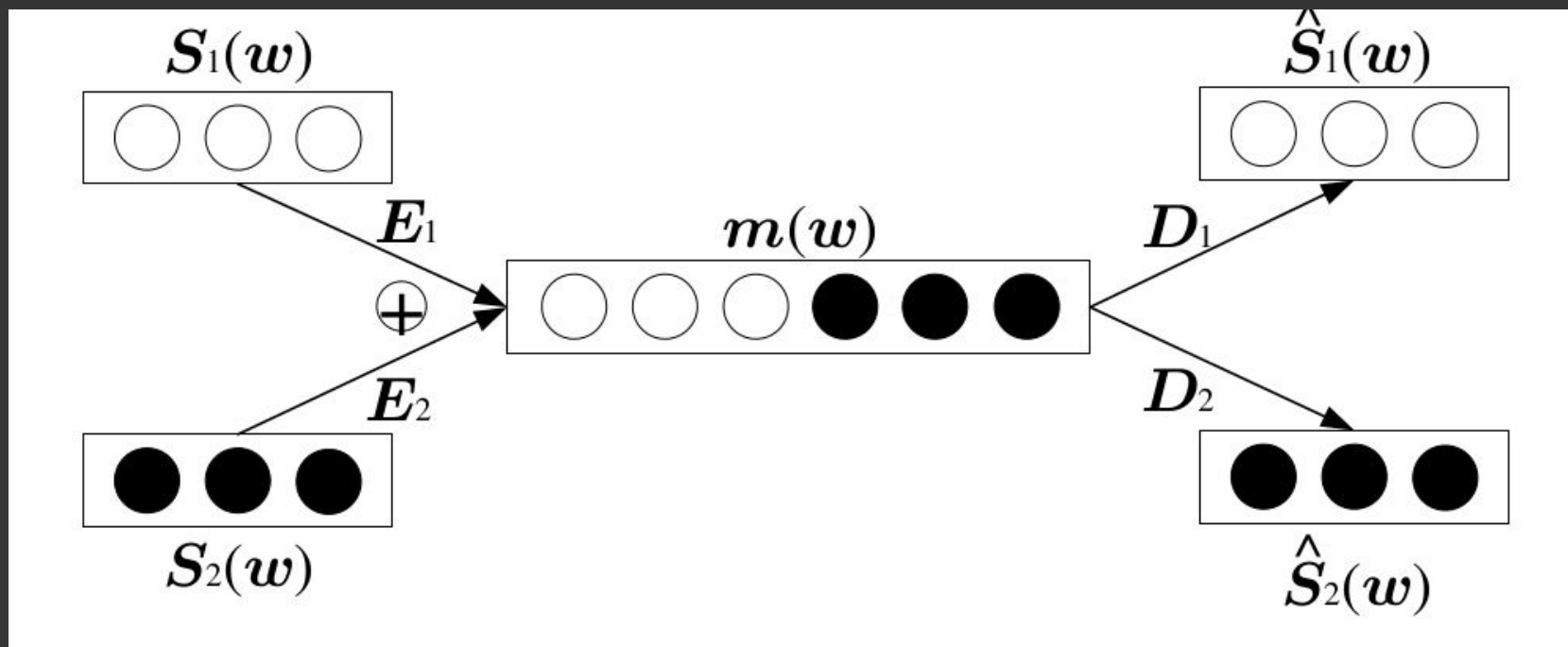


Recap

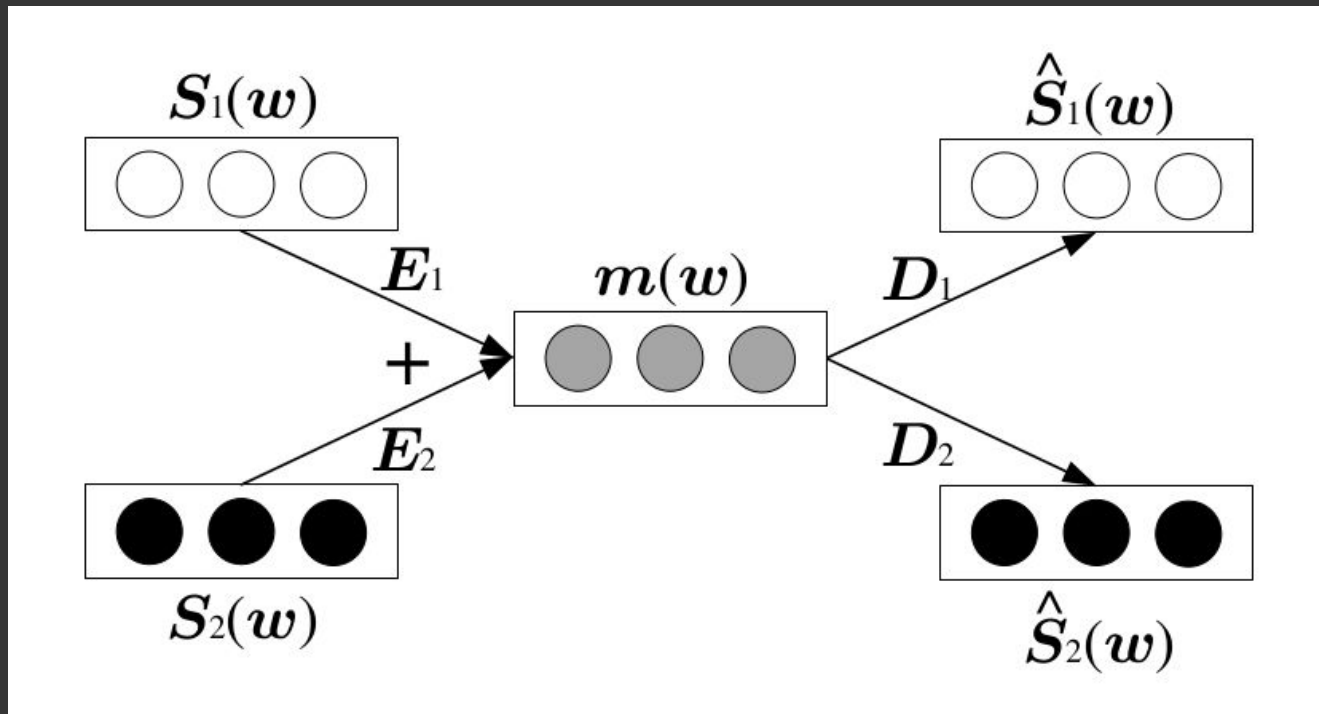
Decoupled Autoencoded Meta-Embedding (DAEME)



Concatenated Autoencoded Meta-Embedding (CAEME)

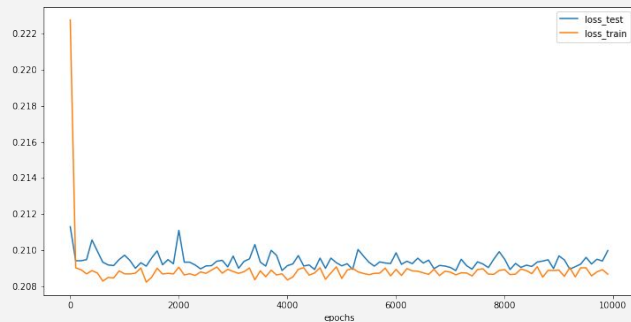


Averaged Autoencoded Meta-Embedding (AAEME)

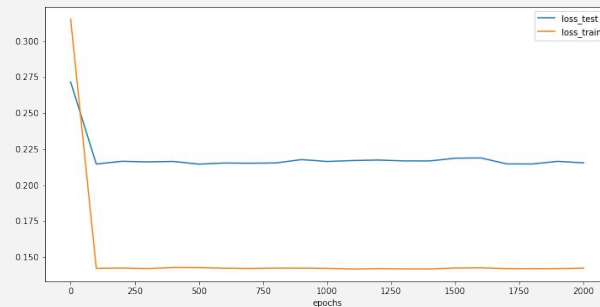


Training

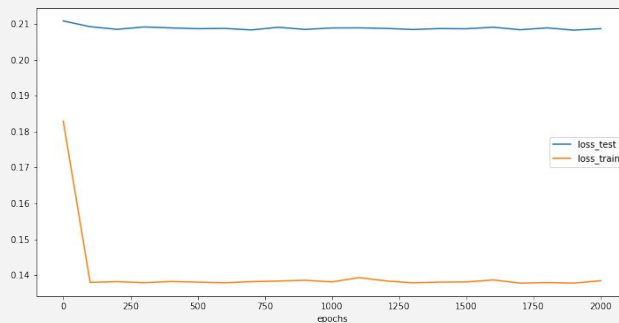
CAEME - Number of Epochs vs Train Loss, Test Loss



AAEME - Number of Epochs vs Train Loss, Test Loss



DAEME - Number of Epochs vs Train Loss, Test Loss



Results

Entity Typing

Model	MAP@k=10	Precision@k=10 (mean \pm std)
Previous Work		
Contextual Embeddings (1)	0.631	0.449 \pm 0.271
KG Embeddings (2)	0.825	0.528 \pm 0.269
Concatenation (1) + (2)	<u>0.828</u>	0.527 \pm 0.268
Attention Based Auto Encoder Model		
DAEME (400)	0.602	0.461 \pm 0.283
CAEME (400)	0.775	0.514 \pm 0.271
AAEME (400)	0.277	0.283 \pm 0.271

Relation Prediction

Model	MRR	Precision (MAP@k=1)
Previous Work		
Contextual Embeddings (1)	0.750	0.554
KG Embeddings (2)	0.817	0.663
Concatenation (1) + (2)	0.738	0.533
Attention Based Auto Encoder Model		
DAEME (400)	<u>0.829</u>	0.687
CAEME (400)	0.816	0.666
AAEME (200)	0.820	0.671
AAEME (400)	0.809	0.653

Conclucion

- **Poor** performance in **Entity Typing**.
- **Great** performance in **Relation Typing**.
- Able to **solve the pertaining problem**, but created a new!

Next Steps!

Hyperparameter
Tuning

Try to improve
the performance
on Entity Typing

Use individual
attention instead
of merging.

Thank you!