Attention is All You Need

Bill Watson

S&P Global

November 22, 2019

Recap: Encoder-Decoder Models

How can we improve this model?

(Naive) Attention Mechanisms

Mean Attention

Location Based: Laplace

Location Based: Gaussian

(Complex) Attention Mechanisms

What are Context Vectors?

Bahdanau Attention

Luong Attention

More Madness

Fine-Grained Attention

Multi-Headed Attention

Self Attention

Practical Considerations

Masking

Which one to use?

Mixing and Matching

Tools, References, and Further Reading

Refrences & Further Reading

- Machine Learning: A Probabilistic Perspective by Kevin Murphy