Introduction to Seq2Seq Modeling

Bill Watson

S&P Global

November 22, 2019

What are Sequence to Sequence Models?

Common Applications for Seq2Seq Models

Encoder-Decoder Archectecture

Overview: Encoders and Decoders

What makes an Encoder?

What makes a Decoder?

Word Embeddings: Practical Considerations

Recap: Word Embeddings

Recap: Pretrained Word Embeddings

DIY Embeddings

Special Tokens for Sequence Modeling

Mangaging the Vocab Size

Morphology, Compounding, and Transliteration

Handling Numbers

Factored Decomposition

Backoff

Character Models

BPE Subwords

Modeling Recurrent Relations

Recap: Recurrent Layers

Vanilla RNNs

Long Short Term Memory (LSTM)

Gated Recurrent Units (GRU)

Aside: The Influence of Padding in RNNs

Right to Left Sequence Modeling

Bidirectional Sequence Modeling

Putting together an Encoder

The Components of an Encoder

Putting together an Decoder

The Components of a Decoder

Connecting the Encoder

Decoder Inference: Making Predictions

Teacher Forcing

Training Considerations

Increasing Throughput through Batching

Label Smoothing

Masked Loss

Decoding: Making better Translations

Beam Search

Monte Carlo Beam Search

Ensembling

Reranking

Tools, References, and Further Reading

Refrences & Further Reading

- Machine Learning: A Probabilistic Perspective by Kevin Murphy