Imperial College London

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

Chameleon Text: Exploring ways to increase variety in artificial data

Author:
Thien P. Nguyen

Supervisor:
Lucia Specia

Submitted in partial fulfillment of the requirements for the MSc degree in Computing (Artificial Intelligence) of Imperial College London

May 2019

Chapter 1

Literature Survey

1.1 Background

1.1.1 Neural Networks

Neural Networks are non-linear statistical models that generate complex relationships between input and output vectors. Note that the input and output vectors are of a fixed dimension, which becomes a problem for our task at hand.

1.1.2 Backpropagation

1.1.3 Autoencoders

Autoencoders are a specialised form of MLPs where the model attempts to recreate the inputs on the output. Autoencoders typically have a neural network layer in the model where its dimension is smaller than the input space, therefore representing a dimensionality reduction in the data. Autoencoders are composed of two different networks, an encoder α and a decoder β, such that $\alpha : X \to F$ and $\beta : F \to X$. The two networks are trained together in a manner that allows them to preserve the input as much as possible.
Increasing the robustness of Autoencoders typically includes the introduction of noise (that the Autoencoder would attempt to ignore - this particular model is described as a Denoising Autoencoder), and experimentation of
Autoencoders are popularised through their use in Machine Translation, Word Embeddings, and document clustering.

1.1.4 Recurrent Neural Networks

Recurrent neural networks (CNNs) are a particular class of neural networks such that the outputs are not necessarily restricted and discrete (as opposed to the MLP). CNNs essentially operate over a sequence of vectors, making them popular in contempoary NLP problems. Given a sequence of inputs $(x_1, ..., x_T)$, a standard RNN computes a sequence of outputs $(y_1, ..., y_T)$ by iterating the following equation:

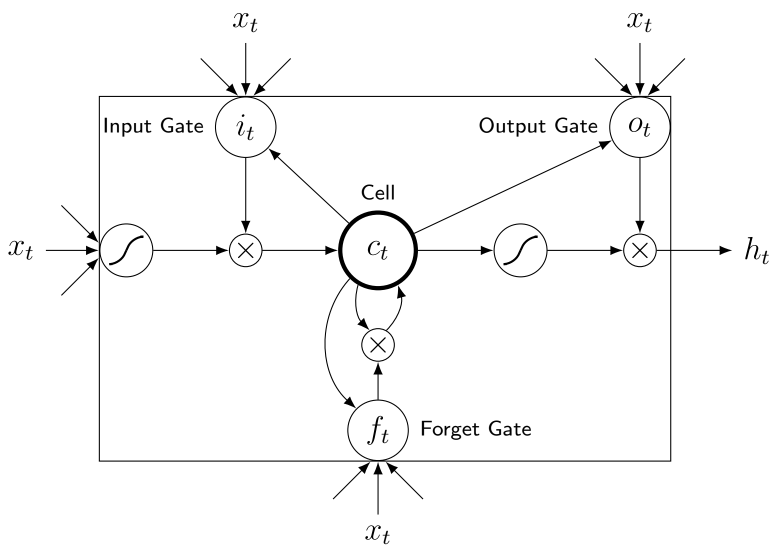

Figure 1.1: A diagram of the LSTM model.

LSTM

LSTM (Long Short Term Memory) is a type of Recurrent Neural Network that attempts to retain information based on the previous inputs through the introduction of gated architectures.

Gated Recurrent Networks

GRUs are used to solve the common issue with LSTMs where the training time was relatively slow due to the number of derivatives necessary to compute. Within the GRU architecture, a feature to retain the previous weights remain, but there exists an direct path to the input data from the output, allowing a reduction in training time.

1.2 Related Work

1.2.1 Seq2Seq

Seq2Seq, introduced by **?** is a modern interpretation of the encoder model, by providing an attention mechanism. The attention mechanism looks at all of the inputs from the hidden states of the encoders so far.

Traditionally, Seq2seq would be the most common method of tackling this particular problem, but it also presents problems that make it suboptimal. Firstly, it's discrete nature suggests that it is prone to noise in a similar fashion to how linear regression is not necessarily optimal as opposed to

1.2.2 Transformers

1.2.3 Variational Autoencoders

Variational Autoencoders (VAEs) introduce a constraint on the encoding network that forces the model to generate latent vectors that roughly follow a gaussian distribution. The architecture of a VAE looks relatively different:

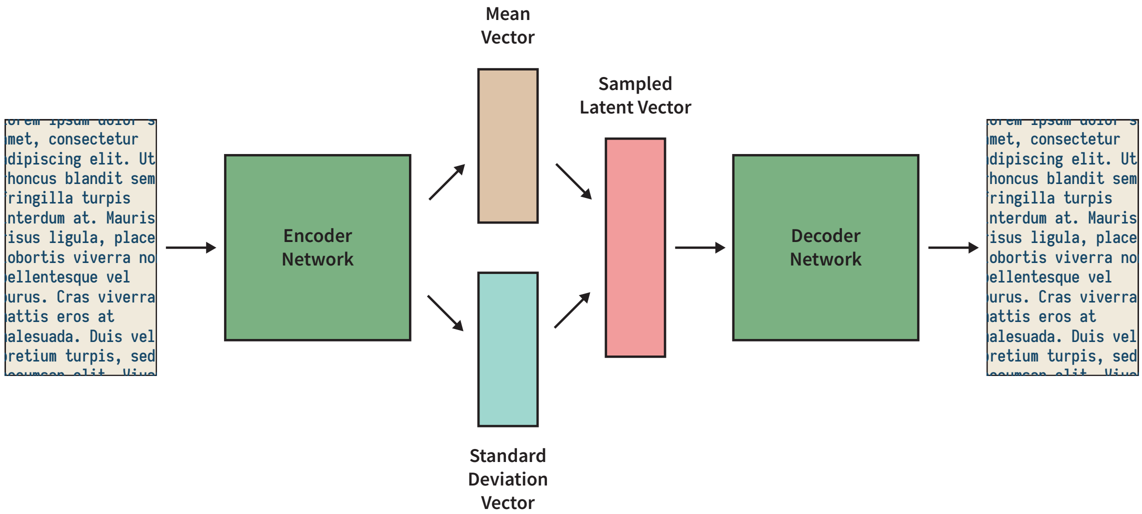

Figure 1.2: A simplified model diagram for a variational autoencoder, which takes as input some text, and it's predicted output being the same text as the input.

Due to the stochastic nature of the network, The loss function used for VAEs is specifically the Kullback–Leibler (KL) Divergence. KL measures the relative difference of two probability distributions.

$$D_{KL}(P||Q) = \sum_{x \subset X} P(x) \cdot log(\frac{P(x)}{Q(x)})$$

In other words, it is the expectation of the logarithmic difference between the probabilities P and Q, where the expectation of P is understood.

1.2.4 Conditional Variational Autoencoders

Introduced by
However, it was found by **?** that since the responses were generated from the same latent variable, it was not necessarily sufficent to model high-variability in responses.

1.2.5 Variational Autoregressive Decoders

Introduced by **?**, Variational Autoregressive Decoders (VADs) attempt to circumvent the sampling problem introduced from CVAEs by introducing latent variables in a autoregressive Decoder. At different time-steps, this allows the decoder to produce a multimodal distribution of text sequences, allowing a variety of responses to be produced.