Machine Translation

Rajat Nagpal
M.Tech(SE)
rajatnagpal@iisc.ac.in

Github Link to this Report

Abstract

Machine Translation is based on naive encoder decoder approach. Different Global Attention Mechanisms are used with seq2seq Models to get better Machine Translation.

1 Introduction

A sequence to sequence model aims to map a fixed length input with a fixed length output where the length of the input and output may differ. Seq to Seq Models are used with Attention to get the corresponding word embedding vectors from the encoder to get the desired output of decoder. In this task, we have used two Attention Mechanisms, Dot product Attention and Additive Attention. In attention mechanism, what we do is that we pay attention to certain specific features for the output to be calculated. We have experimented with applying different Attention Mechanisms to get the English to German Translation and English to Hindi Translation. The accuracy of the Model is calculated by BLUE Score.

2 Encoder

Encoder is basically a cell which takes one input sample and converts it to a proper vector that contains all the needed information for the decoder to decode. It is a stack of RNN or LSTM or GRU units which generates the needed input vector for the decoder.

Output layer of the Encoder goes in to the decoder and acts as its first hidden layer.

3 Decoder

Decoder is basically a cell which takes input from the output encoded vector and update its hidden state vectors to get the decoded output. At the decoder output, we get the translated vector. Decoder also consist of stack of GRUs or LSTMs. We update the hidden states of the decoder and Attention Weights are introduced in the decoder to get the context of the input sequence.

4 Attention Models

Following attention function variants calculate unnormalized alignment scores between hidden states, current and previous, for the prediction.

4.1 Additive Attention

It uses one hidden layer feed-forward network to calculate the following attention function, given by:

$$\text{score}(\boldsymbol{s}_t, \boldsymbol{h}_i) = \mathbf{v}_a^\top \tanh(\mathbf{W}_a[\boldsymbol{s}_t; \boldsymbol{h}_i])$$

where v_a and W_a are parameters of score attention function learned. This variant of Attention modelling is slow and space inefficient but works well for larger dimension matrices.

4.2 Scaled Dot Product Attention

$$\text{score}(\boldsymbol{s}_t, \boldsymbol{h}_i) = \frac{\boldsymbol{s}_t^\top \boldsymbol{h}_i}{\sqrt{n}}$$

n is used as the scaling factor incorporated to the basic Dot product attention variant. It denotes dimension of the source hidden state. This variant uses highly optimized matrix multiplication. Also, it solves the problem of softmax function approaching regions of extremely small gradients,

Input File	Target File	Attention Model	BLEU Score
europarl-v7.de-en.de	`europarl-v7.de-en.en`	`Additive`	`0.13`
europarl-v7.de-en.de	`europarl-v7.de-en.en`	`Multiplicative`	`0.16`
europarl-v7.de-en.de	`europarl-v7.de-en.en`	`Scaled Dot Product`	`0.07`

Table 1: **BLUE Score** using different **Attention Models**

when the dot product grows in basic Dot product attention variant, using this scaling factor.

4.3 Multiplicative Attention

It is a more efficient variant over Additive attention model, given by:

$$\text{score}(\boldsymbol{s}_t, \boldsymbol{h}_i) = \boldsymbol{s}_t^\top \mathbf{W}_a \boldsymbol{h}_i$$

Matrix multiplication are performed here, where W_a denotes the weight matrix which is to be trained in attention layer. This make it faster and space efficient.

4.4 Key-Value Attention

In this attention variant, the trainable matrix of the source hidden states is split into two sections. First section calculates the weights of the distribution, which is referred to as keys. Second section is used for context representation on the hidden source side, referred to as value. This makes up a key-value Attention variant.

5 BLEU Score

BLEU Score is used to calculate the accuracy of the translates text. In BLEU score calculation, the output text and the text computed by our model is given and based on the differences, the BLEU Score is calculated. BLEU Score of 1 corresponds to maximum accuracy and the BLEU Score of 0 corresponds to minimum accuracy.

6 Translations

Following are some translations we got for different attention models :

6.1 Additive Attention

Source: thirdly , not all tax competition is harmful.
Target: drittens ist nicht alle steuern nicht schadlich .

Source: this is why , mr barroso , you must respect the institutions .
Target: deshalb mussen sie die institutionen respektieren .

Source: we unk treat monetary stability as an end in itself , although no one should ignore its importance .
Target: wir nicht als preisstabilitat nicht als selbstzweck , als auch wenn man seine bedeutung nicht ignorieren .

Source: creating conditions enabling disabled people to participate fully in society is also very high on our liberal agenda .
Target: die behinderten menschen in der gesellschaft zu erreichen , ist auch auf der liberalen agenda zu erreichen .

Source: it will be overcome simply by adhering to a basic principle , the principle of solidarity .
Target: es wird einfach nur durch ein grundprinzip , das prinzip der solidaritat .

6.2 Multiplicative Attention

Source: i know that time is up but i wish to make two short points .
Target: schlie lich wei ich , dass ich zeit , aber ich mochte zwei kurze anmerkungen machen .

Source: the agricultural proposals , as represented in agenda , are not satisfactory , particularly from an irish perspective .
Target: wie sie in der agenda , nicht zufriedenstellend , und zwar aus der sicht der irischen sicht nicht zufriedenstellend

Source: thirdly , not all tax competition is harmful .
Target: drittens , da alle unk nicht unk .

Source: this is why , mr barroso , you must respect the institutions .
Target: deshalb mussen sie , herr barroso , die institutionen respektieren .

Source: we unk treat monetary stability as an end in itself , although no one should ignore its importance .
Target: nicht als selbstzweck betrachten , auch wenn niemand seine bedeutung nicht au er acht gelassen werden sollte .

6.3 Scaled Dot Product Attention

Source: i know that time is up but i wish to make two short points .
Target: schlie lich ich zeit , aber ich mochte zwar aus der sicht anmerkungen machen .

Source: the agricultural proposals , as represented in agenda , are not satisfactory , particularly from an irish perspective .
Target: zwei kurze wie sie in der zufriedenstellend , und der irischen sicht nicht zufriedenstellend

Source: thirdly , not all tax competition is harmful .
Target: drittens , da alle unk agenda , nicht unk .

Source: this is why , mr barroso , you must respect the institutions .
Target: deshalb , herr barroso , die auch wenn institutionen respektieren .

Source: we unk treat monetary stability as an end in itself , although no one should ignore its importance .
Target: nicht als mussen sie selbstzweck betrachten , niemand seine bedeutung nicht au er acht gelassen werden sollte .

References

1. Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014).

2. Tensorflow Tutorial on Machine Translation using Attention Model