Natural Language Understanding - E1246
Assignment 2
Sequence to Sequence model using Attention for Machine Translation

Suman Gupta
Computer Science and Automation
Sr No. - 15788
Indian Institute of Science
sumangupta@iisc.ac.in

1 Introduction

In this assignment, we have implemented seq2seq encoder decoder model with attention for machine translation task from English to German and English to Hindi language.

2 Solution Sketch

We had to implement four kinds of attention mechanism, out of which I could complete only three, namely additive, scaled dot product and multiplicative attention.

2.1 Hyperparameters of the model

Batch size = 1
The hidden layer size for both encoder and decoder = 256
Learning rate = 0.001
No of layers in encoder and decoder = 1

2.2 Approach

1. Data Preporcessing - Preprocessing for Hindi corpus was slightly different from that of German because in Hindi, unicode to ascii conversion was not done.

2. Vocabulary creation and tokens generation

3. Source and target language pair generation

4. For every source - target pair,

   (a) Pass the input sequence to the encoder
   (b) Generate and store the encoder outputs for every time step
   (c) Pass the encoder hidden state to the decoder as input
   (d) Calculate attention based on the type of attention specified and take a linear combination of attention along with decoder hidden state to generate the output of the decoder.
   (e) Based on the output probability distribution, generate the predicted word.
   (f) Continue the whole process until either end of sequence token is generated or maximum length is reached.

5. Output: Optimized weights for the encoder and decoder

3 Results

Model trained on 40000 examples from common crawl data set for English to German and hindencorp05 data set for English to Hindi translation.
Loss criterion = NLL loss

3.1 English to German

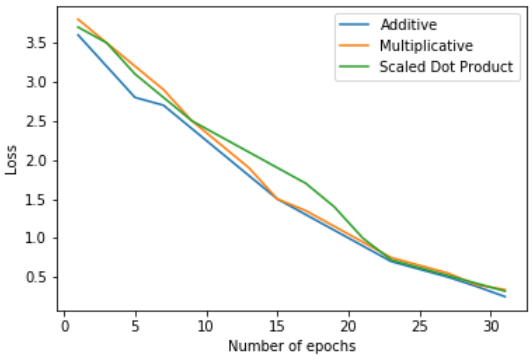
Objective function value at the end of training

1. Additive : 0.256742

2. Multiplicative : 0.349814

3. Scaled dot product : 0.327951

Plot of no of epochs vs loss function for English to German translation task

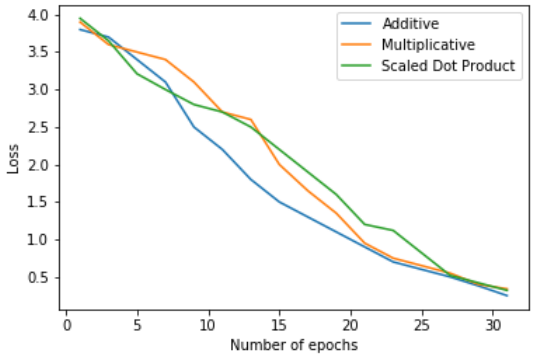


3.2 English to Hindi

Objective function value at the end of 30 epochs

1. Additive : 0.3182
2. Multiplicative : 0.3498
3. Scaled dot product : 0.3141

Plot of no of epochs vs loss function for English to Hindi translation task



3.3 BLEU Score

BLEU score calculated for the newstest2014 dataset

English to German

1. Additive : 0.0831
2. Multiplicative : 0.0927
3. Scaled dot product : 0.124

English to Hindi

1. Additive : 0.149
2. Multiplicative : 0.0965
3. Scaled dot product : 0.138