**Project Report – Sequence2Sequence**

## **Title: Sequence to Sequence Learning with Neural Networks**

## **Reference:**

Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." *Advances in neural information processing systems*. 2014.

## **Motivation:**

This paper refers to the problem of machine translation, i.e. translating sentences from one language to another using machine learning, and in this case specifically using a multilayered LSTM neural network.

## **Short description:**

The algorithm was designed to receive a sequence of words in some language, and output the translated sequence in a target language. Each word was embedded to a 1000 dimensional vector, before passing as an input to the network. The network consisted of 4 layers – 2 layers for encoding the input sequence, and 2 layers decoding the output sequence. Also, the authors stated that reversing the order of the input had an extremely positive influence on the results, so we did the same.

The article tested translated from English to French, and we decided to translate from English to Hebrew instead.

## **Architecture:**

As stated above, this paper used a 4 layers LSTM network, with a 1000 dimensional embedding in order to translate a sentence from one language to another. This architecture can be examined in the following diagram:

***INSERT MODEL DIAGRAM***

**Hyper-parameters:**

|  |  |  |
| --- | --- | --- |
|  | **Paper Original Values** | **Our Implementation Values** |
| **Initial learning rate** | 0.7 | 0.7 |
| **learning rate decline** | After 5 epochs, begin halving the learning rate every half epoch. | After 5 epochs, begin halving the learning rate every epoch. |
| **Non-linear functions** | ??? | ??? |
| **Loss function** | Categorical cross entropy | Categorical cross entropy |
| **# of epochs** | 7.5 | 25 |
| **Batch size** | 128 | 128 |
| **Embedding size** | 1000 | 64 |
| **Parameters initialization** | [-0.08 , 0.08] | [-0.08 , 0.08] |
| **Validation while training** | None | 20% of the training data was used for validation |
| **Optimization Function** | SGD | Adam |

## **Dataset:**

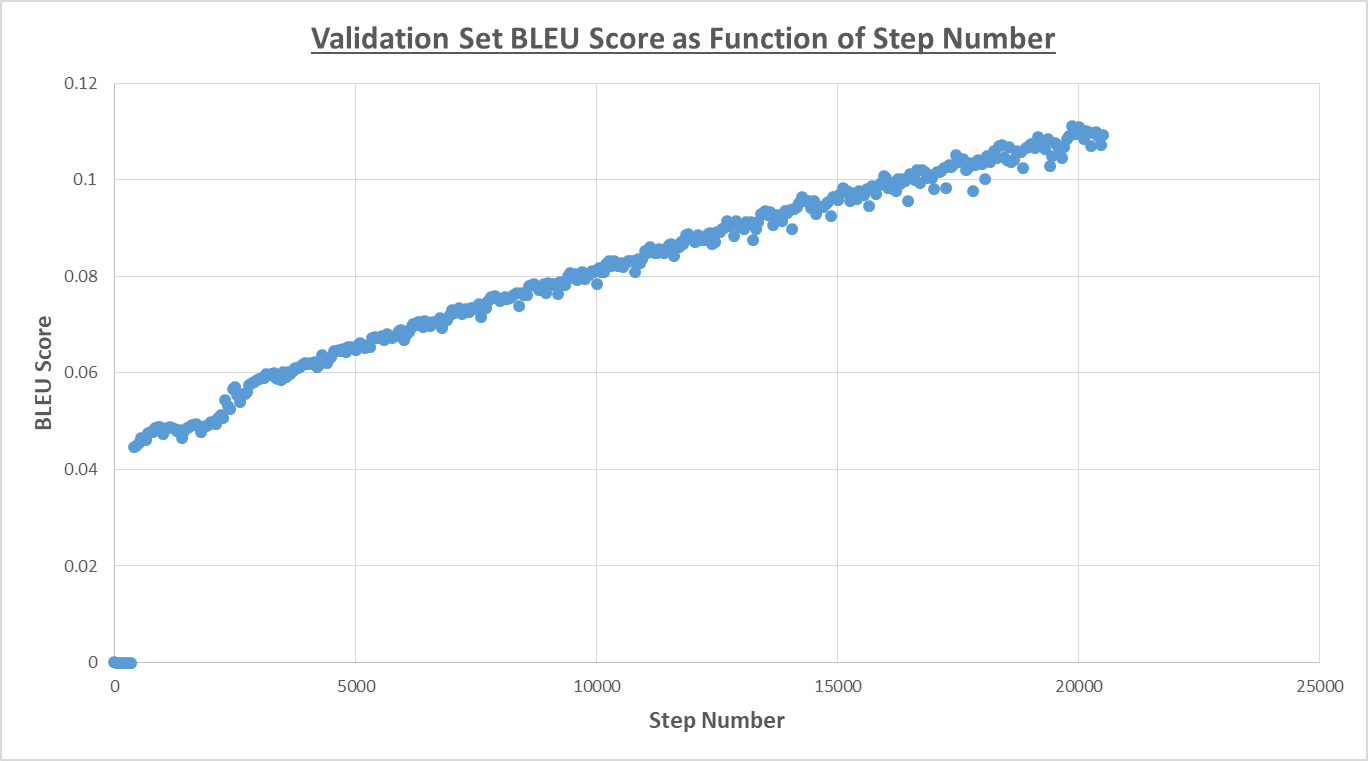
As stated above, the paper translated sentences from English to French, but we decided to translate from English to Hebrew instead, so there are many differences between the data that was used:

|  |  |  |
| --- | --- | --- |
|  | **Paper Original Values** | **Our Implementation Values** |
| **Data Set** | WMT’14 English to French | English to Hebrew from https://tatoeba.org/eng |
| **Size** | 12 million sentences | 160,951 sentences |
| **Source Vocabulary size** | 160,000 of the most frequent words | 20,550 words |
| **Target Vocabulary size** | 160,000 of the most frequent words | 55,736 words |

## **Results:**

* **Final results on test set-** The metric that was used to evaluate the results was BLEU. The value on the test set was *0.098*.
* **Training/validation set loss and BLEU score:**



****

* **Test Set Translation Examples:**

|  |  |  |
| --- | --- | --- |
| **Original Sentence** | **Our Model Translation** | **True Translation** |
| how far does this go ? | כמה זה לעשות ? ? | כמה רחוק זה מגיע ? |
| layla is love for fadil was starting to become an obsession . | כל היה דבר היא היא מכלבים לבן . | האהבה של לילה לפדיל התחילה להיות כפייתית . |
| did you know there was a secret passage hidden behind the bookcase ? | האם עדיין פעם דולר על החיות ? שלי | ידעת שיש מעבר סודי מאחורי כוננית הספרים ? |
| a woman is virtue is everything . | ציפור שייך מזה זה . . | סגולת האשה היא חזות הכל . |
| tom is a policeman . | תום הוא . | תום שוטר . |
| tom just could not resist teasing mary about the way she was dressed . | תום לא לא הצליח לתת להתעלם ממרי מרי כדי את מרי . מרי . . | תום פשוט לא יכול היה לעמוד בפני הפיתוי להקניט את מרי על אופן לבושה . |
| i am still not entirely comfortable with this . | אני עדיין לא עובד בבית טוב . זה . | אני עדיין לא מרגיש לגמרי בנוח עם זה . |

## **Results analysis:**

In general our model did not perform very well, as the maximum BLEU score for a sentence in the test set was *0.643*, which compared to the original article is very low. This is relatively expected, as we had significantly less data.

We thought that our model might have better results for shorter sentences, but that wasn’t the case. This might be because of differences between the languages – a short sentence in English can be long in Hebrew and vice versa. Secondly, even though English’s origin is German, and French is a Latin language, it is known that English was highly influenced by Latin [1] and thus have some basic similarities that English and Hebrew don’t.

## **Conclusions:**

Despite the relatively low results, we believe that this model was able to capture some of the connections between the languages, as seen in the translation examples above, and simply didn’t have enough data to learn from.

As stated above, there were 2 main aspects that we hypothesize that gave as these results – the size of the data set and the use of Hebrew as a target language instead of French. We believe that training on more data, and tweaking the model (perhaps don’t reverse the input) will retain better results.

## **References:**

1. Baugh, A.C. and T. Cable, *A history of the English language*. 1993: Routledge.

**נשאר לעשות:**

* לייצר דיאגרמה של המודל
* לעבור על הדו"ח
* לכתוב הערות על הקוד
* לכתוב readMe על איך משתמשים בקוד (ראה הוראות)