

Final Project: Machine Translation

Leart Krasniqi
Prof. Sable
Spring 2020

1 Project Overview

The goal of this project was to create a machine translation system. The system utilizes a recurrent neural network (RNN) which was created and trained using the TensorFlow library in Python. We also used the Natural Language Toolkit (NLTK) to evaluate our trained network. Although the system is general, we decided to focus on translating Italian text into English text (in order to put our four-year old AP Italian skills to the test).

2 System Architecture

The RNN architecture is a sequence to sequence model, which uses an encoder and a decoder. The encoder accepts an input sequence of one-hot vectors (representing the Italian text) and generates a corresponding sequence of contextualized representations (also known as a context vector). This context vector is then fed into the decoder which generates a sequence of hidden states, from which a translation can be made. Figure 1 provides a schematic representation of the architecture.

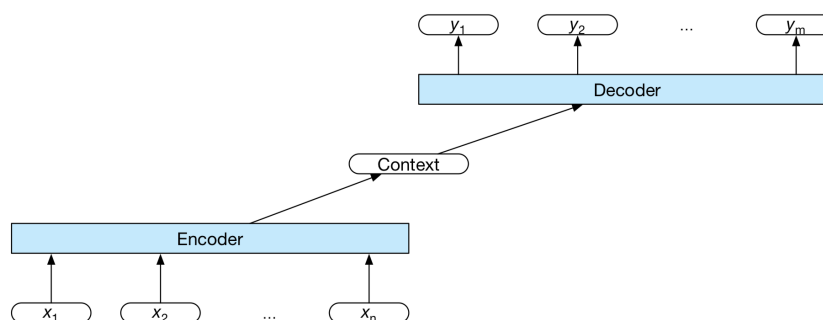


Figure 1: Encoder Decoder Model [1]

Both the encoder and decoder use gated recurrent units (GRUs) rather than LSTMs in order to improve efficiency. Additionally, the decoder uses an attention mechanism in order to improve the system performance by taking into account information from the entire encoder state [3].

3 Performance Evaluation

The system was trained with a corpus containing Italian sentences and their respective English translations, obtained from an online database [2]. A random set of 80,000 sentence pairs were used

to train the model, and 20,000 were used for testing (this split was done using the `train_test_split()` function from `scikit-learn`). Training was performed with 10 epochs and a batch size of 64. The cross entropy loss was calculated for each epoch (using a built-in TensorFlow function). A summary of the training results is shown in Table 1.

| Epoch | Time (s) | Loss |
|-------|----------|--------|
| 1 | 1940 | 1.9645 |
| 2 | 11584 | 1.0529 |
| 3 | 1764 | 0.7651 |
| 4 | 1773 | 0.6710 |
| 5 | 1770 | 0.6353 |
| 6 | 7230 | 0.6173 |
| 7 | 35801 | 0.6088 |
| 8 | 10005 | 0.6029 |
| 9 | 2239 | 0.5983 |
| 10 | 2678 | 0.5960 |

Table 1: Training Summary

When the model finished training, we tested its performance on the 20,000 test pairs. The Bilingual Evaluation Understudy (BLEU) score was used as the metric to determine the effectiveness of the system. The BLEU score is a common metric used in machine translation which compares the translated sentence to a list reference sentences and assigns the translation a score between 0 and 1 (a score of 1 indicates that the translation is identical to at least one of the references) [4]. The model trained on 80,000 sentence pairs achieved a corpus BLEU score of 0.6681 when tested against the 20,000 other pairs. Additionally, we performed some general “human evaluation” where the program was run with user input and the results were judged based on our own knowledge of the correct translation. A sample run of the system is shown in Figure 2.

```
Type in your Italian sentence (or type 'quit' to exit): mi piace mangiare

Input: <s> mi piace mangiare <e>
Predicted Translation: i like to eat . <e>

Type in your Italian sentence (or type 'quit' to exit): Sei perfetto.

Input: <s> sei perfetto . <e>
Predicted Translation: you re perfect . <e>

Type in your Italian sentence (or type 'quit' to exit): Non lo so!

Input: <s> non lo so ! <e>
Predicted Translation: i don t know that ! <e>

Type in your Italian sentence (or type 'quit' to exit): quit
Translation session ended
```

Figure 2: Translation of User-Supplied Sentences

References

- [1] URL: <https://web.stanford.edu/~jurafsky/slp3/10.pdf>.
- [2] URL: <http://www.manythings.org/anki/>.
- [3] Jan K Chorowski et al. “Attention-Based Models for Speech Recognition”. In: *Advances in Neural Information Processing Systems 28*. Ed. by C. Cortes et al. Curran Associates, Inc., 2015, pp. 577–585. URL: <http://papers.nips.cc/paper/5847-attention-based-models-for-speech-recognition.pdf>.
- [4] Kishore Papineni et al. “BLEU: a Method for Automatic Evaluation of Machine Translation”. In: (Oct. 2002). DOI: 10.3115/1073083.1073135.