From Shakespearian to modern English
Machine Learning for Natural Language Processing 2021

Elise Barrere
ENSAE
elise.barrere@ensae.fr

Vincent Delale
ENSAE
vincent.delale@ensae.fr

Abstract

In this project, we have attempted to translate Shakespearian English into modern English and to generate Shakesperian text.

1 Problem Framing

For a 21st century reader, the English of classical texts can be very difficult to understand. For this reason, we have decided to make the translation of Shakespeare's English into modern English the subject of this project. This project is really translation and not style transfer because we don't want to have rhymes as a result, even though Shakespeare's work is written in this format.

We scraped the website www.sparknotes.com to get our data. The code we used to scrape the website is available in our github repository[1] for information purposes. We scraped 4 entire books : Hamlet, Othello, Macbeth and Othello. Visualization of these books is provided in Appendix A and other results are available in the Google Colab file[2]

Several difficulties arose during the course of the project. In particular, our data was translated line by line (in the theatrical sense) and not sentence by sentence. Indeed, even though in total, this dataset contains $81,101$ words, they are only separated into about $4,000$ lines.

In order to remedy to this problem, we used data that we found on github[3] (Xu et al., 2012), which was also taken in part from Spark Notes.This dataset has about $20,000$ lines.

We also briefly explore another branch of Natural Language Processing by training a generator capable of making up Shakesperian sentences.

[1] https://github.com/vdelale/NLP-ENSAE
[2] https://colab.research.google.com/drive/1nRjlgT-8ymx_jswrBvaGmJ9Jj27d6bKH?usp=sharing
[3] https://github.com/cocoxu/Shakespeare

2 Experiments Protocol

We chose to implement our model using Pytorch.

2.1 Formatting the data

First, we formatted the data. As it was extracted from a serious website, it was rather clean. We just needed to remove the comments added by the translators to help understanding. When using the second dataset we found on github this step was unnecessary.

2.2 Translation

We implemented a sequence to sequence model, by reusing the basis of what we saw in the course, during the TP5. This is a summary of the steps taken :

1. **Tokenization** We decided to use the TREEBANKTOKENIZER available in the package NLTK. We then added Start of Sequence and End of Sequence tokens to each line.

2. **Vocabulary** We followed the architecture of the TP5 to create dictionaries of terms.

3. **Building the model** A seq2seq model has a defined Encoder-Decoder architecture. We defined 3 classes : the encoder, the decoder and a global class that contains methods to do the training and the evaluating. After each pass, the loss between the expected result and the result is computed. In our case, we computed the Cross Entropy loss and we used the Adam optimizer with 0.001 as learning rate.

We will discuss the results obtained in part 3.

2.3 Generation

We did not implement a generating model from scratch, we used the powerful transformer architecture and took advantage of generation pretrained model. GPT-2 (Radford et al., 2019)

Original	Translation	Model Translation
Come, let's make haste; she'll soon be back again.	Come on, let's hurry. She'll be back again soon.	come on , come on the ships .
How does my wife?	How is my wife?	what ' you ?
Macduff is missing, and your noble son.	Macduff is missing, and so is your noble son	i ' , , , , , , , ,

Table 1: Example of results of the translation model

is a pretrained model on English language that was trained to do a variety of tasks such as language modeling, summarizing or even reading comprehension. Even though this project is entitled "From Shakespearian to modern English" it seemed more interesting to generate Shakesperian English.

3 Results

3.1 Translation

To evaluate translation, we chose to compute the BLEU score (Papineni et al., 2002; Post, 2018) and the BERT score (Zhang et al., 2019). BLEU score relies solely on the data provided to computed its score, whereas the BERT-score using the knowledge of BERT. We use the BLEU score for information purposes only, because it takes a lot of time to compute it.

3.1.1 Books dataset

Quantitative analysis We can see that the highest achieving model only has a BLEU score of 50, which is ow, as the highest possible is 100. The poor score of our model is confirmed by an analysis done by hand. The results of the gridsearch we did to find optimal parameters are in Figure 9.

Qualitative analysis As we can see in the table 1, the model does not make much sense, it seems that the beginning of the input sentences are better understood than the ending.

3.1.2 Lines dataset

The results of the model trained on the lines dataset are a bit better, especially qualitatively. There is an example of a selection of outputs in the Table 2 in Appendix B. Once again, the short sentences are better translated than the long ones.

3.2 Generative model

It is really hard to assess the results of a generative model. The sentences produced by this algorithm are really convincing. Here is an example *"Amen to her. She will be punished, her husband"* or *"Sir, the Duke of Cornwall needs to get back to his castle."* or *"Were they as popular then when I lived in the city? Do they with smaller audience?"*. Only grammar errors and nonsense between phrases reveal the trick.

However, there is no apparent link to the inputs of modern English sentences.

In order to look for overfitting, or to check if the model did not reuse lines, we checked whether the generated sentences were found in our corpus, and it was never the case.

4 Discussion/Conclusion

The results of the translation are really disappointing. It is mostly due to the fact that we did a simple model, because we preferred to implement a model that we understood on the main task, rather than a powerful model that is more expensive to train like a transformer. Moreover, the machine reduction task is very complex and the style transfer task is much less explored than other more classical NLP tasks such as classification, so we had fewer resources.

Several paths are to be explored to improve our model, such as enlarging the database, or such as a better splitting of the replicas. Some long sentences can be divided in 2 or 3, or even by sentences, but the translation on which we based our dataset is not done sentence by sentence. And of course to use a much more complex model. We tried some methods to reduce overfitting like adding dropout and earlystopping but we didn't have conclusive results, however these methods can bring very good results.

The model we used for the generation part of the project is very powerful, therefore our seemingly good results are not surprising. To achieve better results, we could study more thoroughly the outputs generated and look for similarities within our corpus.

References

Kishore Papineni, Salim Roukos, Todd Ward, and Wei jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. pages 311–318.

Wei Xu, Alan Ritter, Bill Dolan, Ralph Grishman, and Colin Cherry. 2012. Paraphrasing for style. In *COLING*, pages 2899–2914.

Matt Post. 2018. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Belgium, Brussels. Association for Computational Linguistics.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with BERT. *CoRR*, abs/1904.09675.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.

A Exploring the books

B Results

Original	Translation	Model Translation
Be ruled.	Calm down.	quiet .
How but well, sir?	Beautifully, sir, beautifully!	but sir , sir ?
She's an excellent sweet lady, and, out of all suspicion, she is virtuous.	She's an excellent, sweet woman, and there's no doubt that she is virtuous.	she ' s true , and it ' s true , it ' s true , and it ' s true .
My hour is almost come When I to sulfurous and tormenting flames Must render up myself.	The hour has almost come when I have to return to the horrible flames of purgatory.	i ' m well to with these under and with her , and with her , and with our place and said with her himself

Table 2: Example of results of the translation model, line dataset

Figure 1: Keywords in Hamlet (original text)

Figure 2: Keywords in Hamlet (modern translation)

Figure 3: Keywords in Macbeth (original text)

Figure 4: Keywords in Macbeth (modern translation)

Figure 5: Keywords in King Lear (original text)

Figure 6: Keywords in King Lear (modern translation)

Figure 7: Keywords in Othello (original text)

Figure 8: Keywords in Othello (modern translation)

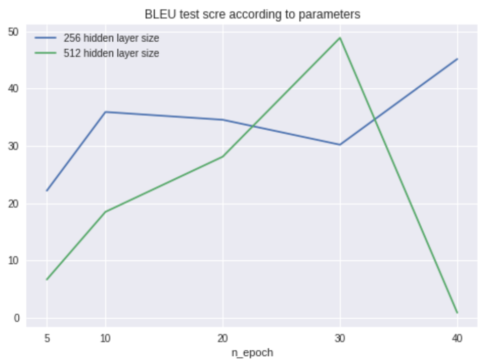

Figure 9: BLEU test score according to parameters on books dataset

Input	Generation
What? Are you going to start eating worms and flies?.	Ma'am, my lord's coming.¡/p¿ ¡s¿You are the one who has to bring this woman here, sir
Come on, I'll escort you to dinner.	When I thought it was a trick of my wit, as it is presently, To come back. Now, my lord.
Rest easy in the meantime.	The sad things from the city. They have for ever been a source of all joy.
Jove is the head of the gods in Roman mythology.Jove's lover.	Pardon the villain, madam but you'll not believe what I have seen tonight

Table 3: Example of generated sentences