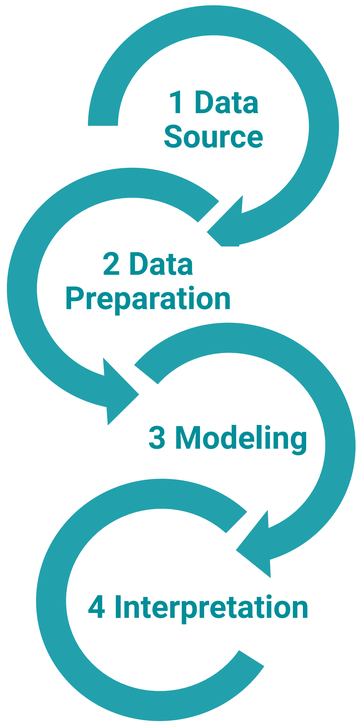
Machine Translation of Shakespearean English

By Qidu(Quentin) Fu, Haotian Shen, Yixing (Leo) Zhu, Kiley Jolicoeur, Ed Cogan
Under Prof. Lu Xiao

Overview

Variations in text styles are frequently based on the content targeting specific contexts, audiences, or purposes (Jhmatani et al., 2017). In this project, we use a multi-head attention model to translate Shakespearean text data from Shakespeare's Early Modern English (late 16th to early 17th century) to Modern English, using concepts of machine translation and style transfer. This model is effective in the translation, defined as "chang[ing] the stylistic properties of the text while maintaining its independent content of style" (Hu, 2022, p.14), achieving an accuracy of .94. The applications of this model are twofold:
1) the model can be in production enhancing translation efficiency and accuracy, and reducing manual labor cost.
2) the model may contribute to the existing methods in text style automating transformations.



Data Source

The text data are from Jhamtani's (2017) study consisting of parallel text data in Modern English and Early Modern Shakespearean English. The data are preprocessed and fitted by the model, and their results are interpreted accordingly.

Data Preparation

- **Tokenization**:
  The first step is to tokenize both the original Early Modern English Shakespeare text (source) and the translated Shakespearean text in Modern English (target). The text data is tokenized into integer sequences, which can be fed into a neural network model in the next step.

- **Embedding**:
  For this project, an embedding technique of the combination of token embedding and positional coding is used for representing words in the transformer model. This technique helps the neural network model understand the semantics of words and their positions in the input sequences better.

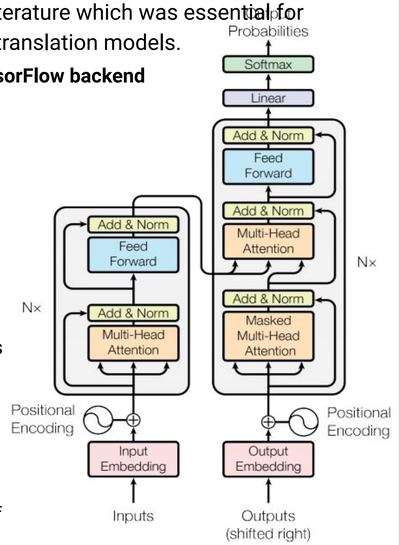
Data Modeling

Data modeling involves representing and structuring linguistic data in a way that allows computer systems to understand and process and "understand" natural language. With this in mind, the basis of modeling that was done was to be able to make predictions and other language related tasks.

The attempt for this study is to transform raw data that was written in Early Modern English text into Modern English Text. Translating between historical languages is a challenging task, and we found that the availability of parallel texts with regards to Shakespeare for training is somewhat limited. It took a deep dive into historical linguistics and literature which was essential for ensuring the accuracy and fidelity of the translation models.

Transformer model using Keras with the TensorFlow backend

A Transformer model is a type of deep learning architecture designed for sequence-to-sequence tasks, such as machine translation. This type of model was initially Introduced in the paper "Attention is All You Need" by Vaswani et al.; Transformers are a significant advancement in the field of sequential data modeling, particularly for applications like natural language processing. There is a level of efficiency that translates into how the data is processed at the processor level. This is achieved by replacing recurrence with attention, allowing computations to occur simultaneously. Additionally, Transformers operate without assumptions about temporal/spatial relationships in the data, making them versatile for processing sets of objects.



Vaswani et al. (2017)

Interpretation

A Transformer model trained with Keras and the TensorFlow backend for transforming Early Modern English text into Modern English text involves understanding how the model processes linguistic nuances and temporal shifts. This can be used to gain insights into the model's focus on specific words or phrases during the translation process.

- The interpretation of a Transformer model involves analyzing how the model captures and represents linguistic features at different layers.
- This involves studying the hierarchical encoding of information, which aids in comprehending how the model handles the evolution of language from Early Modern English to Modern English.

In essence, interpreting the Transformer model for this specific task involves unraveling its attention mechanisms and examining the learned representations to decipher the intricate process of language transformation across time periods. Using Keras with the TensorFlow backend involves gaining insights into the model's decisions and understanding its inner workings. This process included techniques like attention visualization, which allowed the ability to observe which parts of the input sequence the model focuses on during prediction. By analyzing attention weights, it can interpret the importance assigned to different tokens in the input.

Conclusion

Using Shakespearean text, we successfully built a Sequence-to-Sequence Transformer model to perform the translation task from Early Modern English into Modern English and achieved an accuracy of 0.94, showcasing its capability to understand and bridge the English language between two distinct eras. This translation model could be beneficial for education, research, literature analysis, etcs, and, most importantly, makes Shakespearean literature more accessible for people struggling to read but interested in Shakespearean literature.

References

1. Jhamtani, H., Gangal, V., Hovy, E., & Nyberg, E. (2017). Shakespearizing modern language using copy-enriched sequence-to-sequence models. *arXiv preprint* arXiv:1707.01161.
2. Hu, Z., Lee, R. K. W., Aggarwal, C. C., & Zhang, A. (2022). Text style transfer: A review and experimental evaluation. ACM SIGKDD Explorations Newsletter, 24(1), 14-45.
3. TensorFlow. (2023, November 16). Neural Machine Translation with a Transformer and Keras. Retrieved from www.tensorflow.org/text/tutorials/transformer
4. Vaswani et al., (2017). The Transformer-model architecture. 31st Conference on Neural Information Processing Systems (NIPS2017), https://arxiv.org/pdf/1706.03762.pdf. Accessed November 21, 2023.
5. Hong, Tao. "Coherent Wave Dynamics and Language Generation of a Generative Pre-trained Transformer" 8 May 2023, arXiv:2305.05061, Cornell University, https://arxiv.org/abs/2305.05061. Accessed 10 Nov. 2023.
6. Jin, D., Jin, Z., Hu, Z., Vechtomova, O., Mihalcea, R. (2022). Deep Learning for Text Style Transfer: A Survey. Computational Linguistics 48 (1): 155–205. doi: https://doi.org/10.1162/coli_a_00426
7. Troiano, E., Veluthar ambath, A., Klinger, R. (2023). From theories on styles to their transfer in text: Bridging the gap with a hierarchical survey. Natural Language Engineering, 29(4), 849-908. doi:10.1017/S1351324922000407
8. Early modern English: grammar, pronunciation, and spelling. Oxford English Dictionary. https://www.oed.com/discover/early-modern-english-spelling-grammar-and-pronunciation/ Accessed 27 November 2023.
9. Domingo, M., Casacuberta, F. (2023). Interactive machine translation for the language modernization and spelling normalization of historical documents. Pattern Anal Applic 26, 1601–1614 (2023). https://doi.org/10.1007/s10044-023-01164-w