Text Style Transfer with BERT: From Shakespeare to Modern English

Motivation

In this project, our task is to paraphrase language to a particular writing style. We are going to use the plays of Williams Shakespeare and corresponding modern translation as a testbed.

Paraphrasing language can be useful in many ways:

- For academics, it can help non-native speakers to use more appropriate writing style, for example, writing academic papers.
- For a historical analysis, to analyze words in ancient books, it can help people understand text meaning and change of linguistics over time.
- For social and media analysis, we can compare the difference between how people use language now and ten years ago. Also, people of different generations are likely to have different ways of word expression.

In this project, we pay attention to the style of early modern English and modern English.

Background and Rough Plan

Previous works in this field achieve transformation by using Phrase-based translation(Xu et al.,2012), or by inserting related adjectives and adverbs(Saha Roy et al., 2015), or by using sequence-to-sequence models(Harsh Jhamtani et al.,2017). These efforts have shown some performance in the existing evaluation standards, but there is still room for improvement.

In this project, we plan to transfer Shakespearean style English to modern English, like paraphrasing lines from Romeo and Juliet “hast thou slain tybalt” to modern English “ have you killed tybalt”. Unlike traditional style transfer, stylistic paraphrasing in our task is more challenging yet more interesting because one of the styles uses the contemporary language while the other one uses Early Modern English from 16th century. As a newly emerged pre-trained model, Bert realized a breakthrough in sentence representation learning and the distributional relations between sentences learned by the model can be directly applied in various areas of NLP. The methods are simple and straightforward, and the effect is generally better. Therefore, we try to apply the BERT model as both encoder and decoder to text style transforming on sentence-level and explore its effect and impact.

We formulate this problem under the supervised machine learning domain and will fine-tune BERT using a parallel corpus with Shakespeare’s plays and their translation in modern English. we will use the same configuration for both the encoder and decoder. We also plan to delve deeper into the words generated by BERT and analyze the pattern after adding different tags like 'in the 16th century' or 'in the 21st century' to the end of one sentence. We will compare our model with three models COPY, SIMPLES2S and STAT and one baseline model As-it-is (Harsh Jhamtani et al.,2017) using BLEU to measure the closeness between the system prediction and the human annotation. The baseline model achieves a BLEU of 21+.

Data and tools

There are some dataset and tools that are necessary and some may be helpful. As we want to do the text style transformation from Shakespeare to modern English, the data of paraphrases between them are needed to train the model and evaluate the results. The main dataset is the line-by-line modern paraphrases for 16 Shakespeare’s plays from the educational site http://nfs.sparknotes.com. It is completed by Xu et al. (2014; 2012) and is available on github https://github.com/cocoxu/Shakespeare/tree/master/data/align/plays/merged.

We will use the sentences from the first 14 plays as the training dataset, those from Twelfth Night as validation dataset, and those from Romeo and Julie as the test dataset. Secondly, in order to compare the model-transformed results with the given transform, we will use BLEU to do the automatic evaluation of semantic adequacy. Thirdly, while training the baseline model, we use publicly available open-source toolkit MOSES (Koehn et al., 2007) and GIZA++ word aligner (Och, 2003), which is the same as contents in Xu et al., 2012. Moreover, Jhamtani et al. used a mixture model of pointer network and LSTM to transform Modern English text to Shakespearean style English, their codes are available on https://github.com/harsh19/Shakespearizing-Modern-English. Even though we will use BERT to do the transformation, these codes may be helpful to guide us to go through the process.

Collaboration statement

Motivation: Yan Li/ Background and Rough Plan: Yuyue Hua, Hui Cai/ Data and tools: Jiyang Ge

Reference

1. Xu, W., Ritter, A., Dolan, W.B., Grishman, R., & Cherry, C. (2012). Paraphrasing for Style. *COLING*.
2. Jhamtani, H., Gangal, V., Hovy, E.H., & Nyberg, E. (2017). Shakespearizing Modern Language Using Copy-Enriched Sequence-to-Sequence Models. *ArXiv, abs/1707.01161*.
3. Dai, N., Liang, J., Qiu, X., & Huang, X. (2019). Style Transformer: Unpaired Text Style Transfer without Disentangled Latent Representation. *ACL*.
4. Shen, T., Lei, T., Barzilay, R., & Jaakkola, T.S. (2017). Style Transfer from Non-Parallel Text by Cross-Alignment. *NIPS*.
5. Li, Z., Ding, X.L., & Liu, T. (2019). Story Ending Prediction by Transferable BERT. *ArXiv, abs/1905.07504*.
6. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding.
7. Kishore Papineni, Salim Roukos, Todd Ward, and Wei Jing Zhu. (2002). Bleu: a method for automatic evaluation of machine translation.