Annotated Bibliography: Part 1

COSI 137
Prof. Ben Wellner
Yonglin Wang
Due 3.16.2021

Research Topic Outline

This annotated bibliography aims to provide preliminary information for my final paper on text style encoding and extraction. In addition, it is hoped that the information on current advancement in style extraction and style encoding will be able to inform my capstone project, which involves the implementation of a formality classifier and, potentially, a formality re-styler. This section will provide a general summary of the papers discussed in detail in the following sections.

There has been growing body of research in text style transfer aiming to train models that modify specific attributes of input text (e.g., sentiment or formality) while preserving the remaining content (Riley et al., 2020). To stay relevant to our topic on Information Extraction, the tentative scope falls around summarizing and comparing the existing approaches to encode text styles, especially formality, using one of the following types of corpus: 1) aligned parallel training data, commonly found in supervised approaches for text style transfers, 2) non-parallel, labeled training data, commonly found in unsupervised approaches for text style transfers, and 3) non-parallel, label-free training dataset.

Overall, the questions I hope to answer through the literature search are:

- What types of datasets are used?
- How, if any, is the style represented? Is content also extracted as a style-independent representation?
- What tasks are used to train the model? Reconstruction and/or back-translation (what are these tasks anyways)?
- What are the inputs, sentences or sentences with contexts? How are the input texts processed and vectorized?
- Is the style-transfer model an encoder-decoder model? Does the model use RNN or Transformer? Is there a discriminator (as in a GAN model) involved? Where is style extracted?
- How much control do we have over which aspect of style we want to extract/encode?

Summary

This section will briefly summarize some of the technical themes in the papers examined in detail in the following sections.

Many papers, especially the latest ones, address the issue of having limited amount of labeled data. Initially, most work in style transfer used parallel dataset to train a sequence-to-sequence model in a supervised fashion. Such systems are straightforward to design in terms of task definition. However, the problem would be the extremely insufficient amount of training data due to expensive human rewriting.

In light of this, researchers have proposed tasks to allow the incorporation of label-free text. For example, (Riley et al., 2020) proposed the use of reconstruction task, where the input is corrupted with random words being substituted with other words, and the task of the encoder-decoder system is to reconstruct the original input sequence. Another example is defining a back translation task for non-parallel data (Cheng et al., 2020). Through introducing unsupervised objectives to the loss function, the model will be able to leverage unlabeled data in training. Pretrained embeddings and/or models are also used to provide extra information for training (e.g. John et al., 2018; Riley et al., 2020).

Regarding the executability in terms of the final project, in the papers listed below, one major trend in extracting style embeddings is that such extractor is usually trained jointly with an end-to-end system (e.g. Riley et al., 2020; Xu et al., 2020); the end-to-end systems proposed below tend to involve fine-tuning large pretrained networks such as T5 or GPT-2. Therefore, it seems less realistic, within the scope and timeline of our final project, to train a model and obtain the output of the style extractor during inference time.

Taking this fact into account, another interesting technique that we might be able explore is the framework of disentangled representation learning, where, for style transfer tasks, the input will usually be mapped to two separate latent spaces: one for content, and the other for style. In (John et al., 2018), the authors specifically examined the t-SNE plot of their VAE latent embeddings and showed that the negative sentiment and positive sentiment samples were clearly separated in the style space, but not so in the content space, suggesting that their model had worked as expected. This seems like an interesting direction that may be possible for us to pursue in the final project.

TextSETTR: Label-Free Text Style Extraction and Tunable Targeted Restyling

Formal Citation

(Riley et al., 2020) (see last section for full formal bibliography citation)

Summary

This work is the one of the latest in label-free style transfer. Some fundamental assumptions of this work are: 1) large pretrained models (such as T5) already contain textual style features, so it is possible to finetune them for style-transfer instead of training a model from scratch, and 2) style is “slow-moving” and remains similar between two adjacent sentences, and therefore we can extract a “style vector” from the first one and use it to rewrite the second during training.

Question(s) about this work

How are the reconstruction tasks defined here? And relatedly, what are the main tasks used to train the model?

Commentary on Relevance

This work provides a neural approach to extracting text styles on label-free dataset. Its training is trained similar to T5, using a corrupted/noised version of the target (a reconstruction task). A "style extractor" is jointly trained.

Additional Reference from this work

(Lample et al., 2018)

- The original model that this work claims to be based on, with the key improvement that this work uses unlabelled dataset instead of labelled ones.

Contextual Text Style Transfer

Formal Citation

(Cheng et al., 2020)

Summary

This work preserves semantic meaning and contextual consistency of the target sentence through encoding input sentences and their surrounding contexts separately and adding in a pre-trained style and coherence classifier. It also attempts to solve the lack of parallel data for transfer via reconstruction and back-translation losses. In addition, this work proposes two new parallel corpora for style transfer, one for formality and the other for offensiveness.

Question(s) about this work

How does this approach's model architecture differ from that of TextSETTR (Riley et al., 2020)? Can we directly compare their performance?

Commentary on Relevance

This work also uses context as input, but with an emphasis on maintaining the consistency of the target and context, rather than using context as style cue. This paper is also probably one of the most comprehensive in the line of research for style transfer, in terms of model comparison and the number of questions they attempt to raise and solve.

Additional Reference from this work

All works mentioned right above section 5.2 seem to be good places to figure out the background of this task.

On Variational Learning of Controllable Representations for Text without Supervision

Formal Citation

(Xu et al., 2020)

Summary

This work offers a modified version of sequence Variational Autoencoder (VAE), which yields the first success in unsupervised learning of controllable representations for text. Essentially, the modified architecture learns a predetermined number of basis vectors, and then an algorithm determines the corresponding vector for each style label. The model is trained on unlabeled non-parallel text data and uses GloVe embeddings for each input word token.

Question(s) about this work

See comments from Riley below: what advantage, if any, does this approach offer compared to Riley et al., 2020? Is it computationally more efficient? Does it require smaller amount of training data? Or does it take even shorter to train a model?

Commentary on Relevance

Though also using a non-parallel unlabeled corpus, this work takes a different approach from that of Riley et al.'s (2020), according to whom the key difference of this work is that “the number of latent factors must be chosen ahead of time, and this limits the number of attributes that may be controlled.” Additionally, somewhat analogous to topic modeling, Riley and colleagues also noted that we do not know and cannot control which basis correspond to which target attribute (e.g., dialect, politeness, etc.).

Additional Reference from this work

(He et al., 2019, p. 20): architecture this model is based on, provided baselines on text data.

(Higgins et al., 2017): architecture this model is based on, originally for CV only.

Style Transformer: Unpaired Text Style Transfer without Disentangled Latent Representation

Formal Citation

(Dai et al., 2019)

Summary

This work uses labeled non-parallel corpora. Prior to this work, neural models in text style transfer use an encoder to map the text into a style-independent latent vector representation.

This work claims to be the first one that applies Transformer architecture style transfer task. It's also noteworthy that it does not use disentangled latent representations as its predecessors.

Question(s) about this work

Is this work different from (Riley et al., 2020; Xu et al., 2020) in that it doesn't use a style vector? Conversely, the other two papers mention the use of constructing style vectors, but did they use a latent style-independent representation?

Commentary on Relevance

Figure 1 of their work gives a good comparison of disentangled style transfer models and their style transformer. Their work interestingly seems to not require a content/semantic vector, while still requiring a "style control variable" as part of the input.

Additional Reference from this work

In their related work, they mentioned that many previous works, except for (Lample et al., 2018), did not use attention mechanisms to capture the long-term dependency.

References

Cheng, Y., Gan, Z., Zhang, Y., Elachqar, O., Li, D., & Liu, J. (2020). Contextual Text Style Transfer. *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2915–2924. https://doi.org/10.18653/v1/2020.findings-emnlp.263

Dai, N., Liang, J., Qiu, X., & Huang, X. (2019). Style Transformer: Unpaired Text Style Transfer without Disentangled Latent Representation. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 5997–6007. https://doi.org/10.18653/v1/P19-1601

He, J., Spokoyny, D., Neubig, G., & Berg-Kirkpatrick, T. (2019). Lagging Inference Networks and Posterior Collapse in Variational Autoencoders. *ArXiv:1901.05534 [Cs, Stat]*. http://arxiv.org/abs/1901.05534

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., & Lerchner, A. (2017). *β-VAE: LEARNING BASIC VISUAL CONCEPTS WITH A CONSTRAINED VARIATIONAL FRAMEWORK*. 13.

John, V., Mou, L., Bahuleyan, H., & Vechtomova, O. (2018). Disentangled Representation Learning for Non-Parallel Text Style Transfer. *ArXiv:1808.04339 [Cs]*. http://arxiv.org/abs/1808.04339

Lample, G., Subramanian, S., Smith, E., Denoyer, L., Ranzato, M., & Boureau, Y.-L. (2018, September 27). *Multiple-Attribute Text Rewriting*. International Conference on Learning Representations. https://openreview.net/forum?id=H1g2NhC5KQ

Riley, P., Constant, N., Guo, M., Kumar, G., Uthus, D., & Parekh, Z. (2020). TextSETTR: Label-Free Text Style Extraction and Tunable Targeted Restyling. *ArXiv:2010.03802 [Cs]*. http://arxiv.org/abs/2010.03802

Xu, P., Cheung, J. C. K., & Cao, Y. (2020). On Variational Learning of Controllable Representations for Text without Supervision. *ArXiv:1905.11975 [Cs]*. http://arxiv.org/abs/1905.11975