\section{Related work}

In this section, we summarize the related work about the text style transfer, dual learning, data augmentation and stylized dialogue generation.

\subsection{Text style transfer}

In recent years, text style transfer has garnered significant attention within the research community.

The primary objective of this task is to develop models capable of altering a given text's stylistic features while maintaining its core meaning, enabling the automatic rewriting of textual content into various styles, such as formal or informal, sentimental or objective.

Existing text style transfer includes two categories data settings, parallel supervised, non-parallel supervised.

The text style transfer models are trained with parallel style texts (i.e. informal-formal) in parallel supervised.

Non-parallel supervised is a scenario close to the real world, without any matching text to transfer the style of the text.

Some early works try supervised approaches, where a work is the transfer between modern English and Shakespearean style sentences~\cite{Xu2012ParaphrasingFS, Jhamtani2017ShakespearizingML}. \citet{Jhamtani2017ShakespearizingML} utilized a Seq2Seq model trained on a parallel corpus to translate modern English phrases into Shakespearean English. Nevertheless, more and more recent research begins to pay attention to the unsupervised approach due to the lack of parallel corpora~\cite{Zhang2018LearningSM, Wu2019AHR}. In general, there are three typical approaches. The first is to learn disentangled representations of content and style. Shen et al.~\cite{Shen2017StyleTF} learned a shared representation of context across different styles and then leveraged the refined alignment of latent representations for style transfer. Fu et al.~\cite{Fu2018StyleTI} utilize a multi-decoder with a style embedding model to learn separate content representations, where style representations are obtained via adversarial networks. Another is to generate stylized texts with back translation~\cite{Zhang2018StyleTA, Lample2019MultipleAttributeTR}, which produces pseudo parallel data to jointly train the forward and backward model. The third is based on the template, which tries to replace original stylistic words with the target stylistic words~\cite{Li2018DeleteRG, Wu2019AHR}.

\subsection{Controllable text generation}

Controllable text generation is a field of research that aims to generate text that adheres to certain constraints, which include various aspects of generated content~\cite{Logeswaran2018ContentPT}, topic~\cite{Dathathri2019PlugAP,Khalifa2020ADA,Wang2019TopicGuidedVA}, emotion~\cite{Liu2018AnER,Zhang2017BuildingEC} and so on.

One popular approach to constrained text generation is through the use of supervised methods. In supervised learning, a model is trained on a labeled dataset of input-output pairs, and the goal is to learn a function that maps input to output. In the case of text style transfer, the input might be a piece of text in one style, and the output might be the same text re-written in a different style. Some popular supervised methods for constrained text generation include sequence-to-sequence models. For instance, \cite{Gupta2017ADG} proposed a combination of variational autoencoders with LSTM models to generate paraphrases.~\cite{keskar2019ctrl} introduced CTRL, which is a transformer-based language model that allows for fine-grained control over the generation process through conditioning on a variety of inputs. What’s more, a simple plug-and-play architecture were invented, thus made it easier to guide the generation process.~\cite{dathathri2019plug} Moreover, the above approaches rely on an adequate parallel supervised corpus, which is hard to obtain in real-world application scenarios.

Another approach to constrained text generation is through the use of unsupervised methods, which do not rely on labeled data~\cite{Xu2019UnsupervisedCT}. One popular unsupervised method for constrained text generation is style transfer via back-translation, which involves first translating the input text into the target style, and then back-translating it into the original language. These methods can effectively get rid of the reliance on supervised datasets but remain difficult to control and incorporate generative constraints.

\subsection{Back translation}

Back translation is to train a target-to-source seq2seq model for producing source sentences, and then construct pseudo parallel datasets.

The core idea behind back translation is to first translate a given text from the source language to a target language, and then translate the resulting text back to the source language. This process typically results in a paraphrased version of the original text, which can be utilized to improve the performance of NLP models by introducing diversity in the training data and increasing their robustness.

This approach is widely used in various tasks. In machine translation, additional parallel training data were provided through using back translation on monolingual training data ~\cite{sennrich2015improving, Artetxe2018UnsupervisedNM}. While in text style transfer, model based on this mechanism produces better generations ~\cite{Subramanian2018MultipleAttributeTS}. When it comes to dialogue generation, back translation can effectively incorporate non-conversational text with conversational text,thus produce more diverse responses. ~\cite{Su2020DiversifyingDG, zheng2021stylized}.

Similar concurrent work is \textbf{Dual learning}~\cite{He2016DualLF}, which is first proposed for neural machine translation, involves two tasks with an immediate reward. Dual learning has proved to be very effective in neural machine translation and image translation~\cite{Lin2019ImagetoImageTW}. Our work is also based on the idea of them. However, most of the recent works have focused on two domain mapping. We employ dual learning to work on a three-domain text related problem, then the contents of non-conversational text can be effectively utilized to enrich the dialogue generation.

\subsection{Dialogue generation}

A dialogue system is a system that interacts with human in natural language, one of its main goal is to pass the turing test. Before neural age, methods to create diaogues include rule-based~\cite{weizenbaum1966eliza} as well as learning-based~\cite{litman2000njfun}. However, it is costly to write rules for every types of dialogue, and the learning-based method is quite inmature because of the irrelevant responses it often produces.

In recent years, as neural networks are developing rapidly, retrieval based dialogue systems and generation-based methods become two options for building chit-chat systems. The former one retrieve a proper response from existing texts on the internet. To be more specific, ~\cite{ji2014information} divide the system into three stages including retrieval, matching and ranking. And deep neural networks are essensial in ranking matching responses, such can be viewed as a response selection task.~\cite{tao2021survey}. Compared with retrieval-based dialogue systems, generation-based method creates new responses. It is especially helpful when the former system fails to find a proper answer. Starting from ~\cite{sutskever2014sequence}, various sequence to sequence networks are provided. Then ~\cite{bahdanau2014neural} add context attention mechanism to enhance performance. At last, Transformer model is introduced to the task and become the most popular framework because of its ability to capture long-range correlations and merits of both parallel training and dynamic context window modeling.~\cite{vaswani2017attention}

\subsection{Stylized dialogue generation}

Stylized dialogue generation refers to generate a response in the target style. This task has become a hot spot of research due to its wide applications in online chatbots and customer services. ~\cite

{Zhou2018EmotionalCM} propose an Emotional Chatting Machine (ECM) model for the emotional response. An intuitive solution generates a response with a pre-trained style language model~\cite{Yang2020StyleDGPTSR}, which builds a response generation model STYLEDGPT on top of a pre-trained language model DialoGPT~\cite{Zhang2020DialoGPTLG} and devise both a word-level loss and a sentence-level loss to fine-tune the DialoGPT towards the target style. Niu and Bansal~\cite{Niu2018PoliteDG} propose three weakly supervised models that can generate diverse, polite (or rude) dialogue responses without parallel data. Some work bridge conversation modeling and non-parallel style transfer by sharing a structured latent space~\cite{Gao2019StructuringLS}. ~\cite{li2021stylized}improved to work through introducing a more lightweighted deep neural model.Recently, some work iteratively tuned to find the optimal mapping relation between conversational context and non-conversational utterances via iterative back translation~\cite{Su2020DiversifyingDG, Zheng2020StylizedDR}. These works have achieved good performance, but we have achieved better diversity and efficiency through an auxiliary domain.

Controllable text generation部分

@article{keskar2019ctrl,

title={Ctrl: A conditional transformer language model for controllable generation},

author={Keskar, Nitish Shirish and McCann, Bryan and Varshney, Lav R and Xiong, Caiming and Socher, Richard},

journal={arXiv preprint arXiv:1909.05858},

year={2019}

}

@article{dathathri2019plug,

title={Plug and play language models: A simple approach to controlled text generation},

author={Dathathri, Sumanth and Madotto, Andrea and Lan, Janice and Hung, Jane and Frank, Eric and Molino, Piero and Yosinski, Jason and Liu, Rosanne},

journal={arXiv preprint arXiv:1912.02164},

year={2019}

}

Back translation部分

@article{sennrich2015improving,

title={Improving neural machine translation models with monolingual data},

author={Sennrich, Rico and Haddow, Barry and Birch, Alexandra},

journal={arXiv preprint arXiv:1511.06709},

year={2015}

}

@article{subramanian2018multiple,

title={Multiple-attribute text style transfer},

author={Subramanian, Sandeep and Lample, Guillaume and Smith, Eric Michael and Denoyer, Ludovic and Ranzato, Marc'Aurelio and Boureau, Y-Lan},

journal={arXiv preprint arXiv:1811.00552},

year={2018}

}

@article{su2020diversifying,

title={Diversifying dialogue generation with non-conversational text},

author={Su, Hui and Shen, Xiaoyu and Zhao, Sanqiang and Zhou, Xiao and Hu, Pengwei and Zhong, Randy and Niu, Cheng and Zhou, Jie},

journal={arXiv preprint arXiv:2005.04346},

year={2020}

}

@inproceedings{zheng2021stylized,

title={Stylized dialogue response generation using stylized unpaired texts},

author={Zheng, Yinhe and Chen, Zikai and Zhang, Rongsheng and Huang, Shilei and Mao, Xiaoxi and Huang, Minlie},

booktitle={Proceedings of the AAAI Conference on Artificial Intelligence},

volume={35},

number={16},

pages={14558--14567},

year={2021}

Dialogue generation部分

@article{weizenbaum1966eliza,

title={ELIZA—a computer program for the study of natural language communication between man and machine},

author={Weizenbaum, Joseph},

journal={Communications of the ACM},

volume={9},

number={1},

pages={36--45},

year={1966},

publisher={ACM New York, NY, USA}

}

@inproceedings{litman2000njfun,

title={NJFun-a reinforcement learning spoken dialogue system},

author={Litman, Diane and Singh, Satinder and Kearns, Michael S and Walker, Marilyn},

booktitle={ANLP-NAACL 2000 Workshop: Conversational Systems},

year={2000}

}

@article{ji2014information,

title={An information retrieval approach to short text conversation},

author={Ji, Zongcheng and Lu, Zhengdong and Li, Hang},

journal={arXiv preprint arXiv:1408.6988},

year={2014}

}

@inproceedings{tao2021survey,

title={A Survey on Response Selection for Retrieval-based Dialogues.},

author={Tao, Chongyang and Feng, Jiazhan and Yan, Rui and Wu, Wei and Jiang, Daxin},

booktitle={IJCAI},

pages={4619--4626},

year={2021}

}

@article{vaswani2017attention,

title={Attention is all you need},

author={Vaswani, Ashish and Shazeer, Noam and Parmar, Niki and Uszkoreit, Jakob and Jones, Llion and Gomez, Aidan N and Kaiser, {\L}ukasz and Polosukhin, Illia},

journal={Advances in neural information processing systems},

volume={30},

year={2017}

}

@article{sutskever2014sequence,

title={Sequence to sequence learning with neural networks},

author={Sutskever, Ilya and Vinyals, Oriol and Le, Quoc V},

journal={Advances in neural information processing systems},

volume={27},

year={2014}

}

@article{bahdanau2014neural,

title={Neural machine translation by jointly learning to align and translate},

author={Bahdanau, Dzmitry and Cho, Kyunghyun and Bengio, Yoshua},

journal={arXiv preprint arXiv:1409.0473},

year={2014}

}

Stylized dialogue generation部分

@article{li2021stylized,

title={Stylized dialogue generation with multi-pass dual learning},

author={Li, Jinpeng and Xia, Yingce and Yan, Rui and Sun, Hongda and Zhao, Dongyan and Liu, Tie-Yan},

journal={Advances in Neural Information Processing Systems},

volume={34},

pages={28470--28481},

year={2021}

}