M.J. Hornos et al. (Eds.)

doi:10.3233/AISE240025

Large Language Models for Text Style Transfer: Exploratory Analysis of Prompting and Knowledge Augmentation Techniques

Martina TOSHEVSKA [a,1] and Sonja GIEVSKA [a]
[a] *Ss. Cyril and Methodius University*
Faculty of Computer Science and Engineering
Skopje, North Macedonia

Abstract. Large language models have gained extensive research interest in the past few years. They have demonstrated remarkable ability to process and generate human-like text, and have improved performances on various natural language processing tasks. This paper is focused on the prompting techniques and knowledge augmentation techniques for text style transfer tasks. Text style transfer involves the transformation of a given sentence in a stylistically different manner while preserving its original meaning. It requires models to understand and manipulate different aspects such as politeness, formality, and sentiment. This paper provides an overview of several methods for prompting large language models for text style transfer and presents an overview of several methods for knowledge augmentation with a discussion about potential use for text style transfer. Preliminary results on formality transfer using the T5 model are presented to evaluate prompting and knowledge augmentation techniques. The results show that using knowledge augmentation techniques improves the performance compared to models without augmentation, while zero-shot prompting techniques are less effective. This emphasizes the necessity of fine-tuning and incorporating knowledge augmentation for enhanced model performance.

Keywords. text style transfer, large language models, prompting, knowledge augmentation

1. Introduction

Recently, Large Language Models (LLMs) have shown promising performances on a wide range of natural language processing (NLP) tasks. LLMs are built following the Transformer [1] architecture. These models utilize pre-training on large corpora that enable learning the language semantics and patterns. The pre-training phase includes predicting the next word for a particular sequence [2,3], or masked language modeling and next sentence prediction [4]. Pre-trained models could be fine-tuned on downstream tasks using task-specific datasets that are usually smaller than the datasets used for pre-

[1] Corresponding author: Martina Toshevska, martina.toshevska@finki.ukim.mk

training. The fine-tuning phase allows the model to learn task-specific language patterns and semantics. Since it was found that increasing the model or data size improves the performances on downstream tasks, there have been a variety of LLMs proposed in the past few years: T5 [5], FLAN-T5 [6], PaLM [7], GPT-3 [8], GPT-4 [9], Falcon [10], LaMDA [11], LLaMA [12], LLaMA-2 [13], and many others.

Generative Pre-trained Transformer (**GPT-3**) [8] is a language model with 175B parameters that introduced the technique of in-context-learning i.e. learning from a few or no samples. **GPT-4** [9] is an extended version able to process multimodal input data. Text-To-Text Transfer Transformer (**T5**) [5] aims to unify all NLP tasks and treat them as a text-to-text tasks. T5 achieved state-of-the-art results on multiple NLP tasks at the time it was proposed. Fine-tuning Language Models [6] was designed to enhance and improve the effectiveness of zero-shot learning (**FLAN-T5** enhanced and improved the effectiveness of zero-shot learning of the T5 model). Pathways Language Model (**PaLM**) [7] is a 540B parameter model that outperformed the fine-tuned state-of-the-art models on several multi-step reasoning tasks. **Falcon** [10] is a model with 40B parameters that was trained with multilingual data. Language Models for Dialog Applications (**LaMDA**) [11] are a family of models that were designed for dialog applications and have up to 137B parameters. Large Language Model Meta AI (**LLaMA**) [12] and the succeeding **LLaMA-2** [13] are a collection of several models with number of parameters ranging from 7B to 70B that achieved superior performances in instruction following tasks. Considering the popularity of the LLMs, a lot of survey papers emerged. For a more comprehensive overview and analysis of these LLMs, one can refer to [14], [15], and/or [16].

This paper is focused on the applications of LLMs for text style transfer. Text style transfer is the task of rewriting a sentence in a different style while preserving its content. It involves generating a new sentence with the same explicit meaning that is stylistically different from the original one. The term style encompasses diverse properties such as the individual style of the author, politeness, formality, sentiment, and various other styles. Text style transfer has been utilized to adjust, modify, or adapt the manner in which a sentence is expressed. Altering the expressed emotions in a sentence is known as *sentiment transfer* while adjusting the politeness or formality is associated with *politeness transfer* and *formality transfer*, respectively. Rewriting text to align with an individual author's writing style, such as Shakespeare or Taylor Swift, falls under *personal style transfer*. Transforming a sentence in a way that is more comprehensible for non-experts in a particular field (e.g. medical experts vs. layman) is referred to as *expertise style transfer*. One possible use case of text style transfer is in the realm of social media platforms. By using text style transfer on social media, platforms can enhance user experiences and prevent potential miscommunications. Sentiment style transfer could be utilized to modify the emotional tone of user-generated content to be more positive. Similarly, politeness and formality transfer can customize communication to align with the anticipation of particular social media communities.

Text style transfer methods could be generally categorized into two groups: (1) *supervised* that rely upon parallel data, and (2) *unsupervised* that work with non-parallel data. Supervised methods [17,18] are based on the sequence-to-sequence architecture that was proposed for machine translation [19]. Due to the limited availability of parallel data for the task, unsupervised methods [20,21,22,23,24] have gained more attention. Creating pseudo-parallel data for supervised training on non-parallel datasets has also been explored [25]. For a broader perspective of text style transfer tasks and methods, one can refer to [26] and/or [27]. More recent methods proposed in the field explore the usage of prompting techniques for LLMs for text style transfer, which is one of the two directions covered in this paper. The second direction covers the knowledge augmentation techniques which, to the best of our knowledge, have not been explored for text style transfer yet. We believe it is a research direction that is worth exploring and that could potentially improve the existing methods in the field.

The rest of the paper is organized as follows. In Section 2 an overview of methods for prompting LLMs is presented and several applications on the text style transfer task are described. In Section 3 various methods for augmenting LLMs with knowledge are described and potential applications for text style transfer are discussed. In Section 4 preliminary results for prompting and knowledge augmentation for text style transfer are presented. Section 5 concludes the paper.

2. Overview of Prompting Techniques for Text Style Transfer

GPT-3 [8] demonstrated better performance than the state-of-the-art models at the time for benchmark datasets without fine-tuning. Being given only a few examples as a demonstration of the task at inference time without updating the weights, GPT-3 was able to successfully solve various tasks such as machine translation and question answering. Since then, prompting techniques for LLMs gained an extensive research interest. Prompting LLMs is the technique of providing specific input instructions to obtain the desired output by guiding the model, and without learning additional data apart from the pre-training data. For zero-shot prompting [28], no examples of the task are given, while for one-shot and few-shot prompting [29] one or n examples are provided as part of the prompt, respectively. LLM prompting techniques have been explored for various NLP tasks. The following paragraphs analyze several research papers that explore prompting techniques for text style transfer.

Augmented Zero-Shot Learning [30] is a few-shot prompting technique that performs multitask text style transfer using a single set of exemplars. An LLM was prompted with samples of several sentence rewriting options instead of only one. The LLMs used for the experiments were GPT-3 [8] and LaMDA [1] with a non-embedding parameter count of 137B [11]. The models were evaluated on two text style transfer tasks: formality transfer and sentiment transfer. On these tasks, the models achieved high accuracy and low perplexity in comparison with several baseline models. The BLEU [31] scores were low because of the tendency of LLMs to add additional information to the generated sentences. Several model sizes were evaluated indicating that enhancing the model size improves the performance. The LLMs were also evaluated on 6 non-standard text style transfer tasks such as ”more descriptive” or ”more melodramatic”, for which human evaluation was performed. The generated outputs received nearly as high ratings as the human-written ground truth sentences.

Prompt-and-Rerank [32] is a method for arbitrary text style transfer with zero-shot and few-shot prompting that applies a ranking method to choose the best generated sentences. For each source sentence, k candidate outputs were generated and then ranked according to a joint score calculated using textual similarity, transfer strength, and fluency. The candidate with the highest score was chosen as an output sentence in the target style. The re-ranking method improves the style accuracy and often improves the sBLEU and fluency scores. Four GPT-2 [3] models with varying sizes were assessed on five text style transfer tasks: sentiment transfer, Shakespearean style transfer, formality transfer, grammar error correction, and symbol to natural language translation. The method showed competitive performances with prior methods on the tasks and obtained better sBLEU and accuracy scores than some of the settings evaluated with the augmented zero-shot learning method [30]. The experimental results demonstrated that larger models often perform better than smaller models. The GPT-2-Small model consistently achieved high sBLEU scores and low accuracy scores indicating that it often copied long sections of the input without changing the style.

Augmented zero-shot learning method [30] explored only a *vanilla* prompt design that specified the target style in the second half of the prompt. Prompt-and-Rerank [32] explored three additional prompt designs: *contrastive* and two versions of a *negation* prompt. The contrastive prompt provided information about the source style and created a contrast between the source and the target styles. Negation prompts specify the source style as a negation of the target style and vice versa. The evaluation results showed that the contrastive prompt achieved the best accuracy scores suggesting that this type of prompt improves style transfer quality. The choice of the delimiter type for separating the samples in the prompt had a large impact on the performance with the best performances achieved when curly brackets $\{\cdot\}$, square brackets $[\cdot]$, parentheses $(\cdot)$, and quotes "$\cdot$" were used.

Prompt-Based Editing [33] method transformed the text style transfer generation task into a style classification task. Given a candidate sentence, the goal was to obtain a classification probability with the GPT-J-6B. The classification probability was combined with GPT-2 [3] score for fluency and RoBERTa [34] score for semantic similarity to obtain a joint style score. The steepest-ascent hill climbing (SAHC) algorithm [35] was applied for local search using editing operations (insertion, deletion, and replacement). For each editing position every editing operation was performed and the candidate sentence with the highest score was selected. The average edit distance was 2.9 and 4.7 steps for sentiment transfer and formality transfer, respectively. This method achieved better performances than the Prompt-and-Rerank method [32], and a better balance between content preservation and style transfer strength compared to the augmented zero-shot learning method [30]. The improvement was smaller for the formality transfer task because it is more challenging than the sentiment transfer task.

3. Overview of Knowledge Augmentation of LLMs for Text Style Transfer

Knowledge augmentation has been explored for LLMs in two directions [36]: querying LLMs as knowledge bases (KBs) and augmenting LLMs with knowledge. Querying LLMs as KBs aims to retrieve relevant knowledge learned from the LLM. This task was explored in various cases including casting the knowledge contained within language models into a knowledge graph [37] or fine-tuning to answer questions without access to any external context or knowledge [38]. Many NLP tasks are knowledge-intensive i.e. they require access to external knowledge sources.

To the best of our knowledge, by the time of writing this paper, there is a limited number of research work that explores knowledge augmentation for text style transfer tasks. In what follows, we describe a few directions that could be viewed as a kind of knowledge augmentation: style markers[2] of the sentence from the set of sentences with the target style that is the most similar to the input sentence, and information retrieved from a style memory for the target style. **SMAE** [39] was proposed for sentiment transfer. This model used additional components called memories for each sentiment (positive and negative) to learn and store information about the target sentiment. The memories were used to extract sentiment information from the memory of the target sentiment that was fed into the model as additional input information. The evaluation results showed that including sentiment memories leads to an improvement of 62.56% on style strength suggesting that the sentiment memories are key components to ensure successful sentiment transfer. Considering the improvements made to the SMAE model, sentiment memories could emerge as a potential method for knowledge augmentation. It could be extended to diverse text style transfer tasks assuming these memories are developed using appropriate data that represent the key aspects of the particular style.

DeleteAndRetrieve [20] and **G-GST** [21] utilize style markers of the retrieved most similar sentence from the corpus of sentences with the target style. These models were also primarily proposed for sentiment transfer. They consist of three components: (1) *delete* that deletes the style markers from the input sentence, (2) *retrieve* responsible for retrieving the most similar sentence from the corpus with the sentences in the target style, and (3) *generate* that generates the output sentence in the target style. DeleteAndRetrieve used (1) TF-IDF weighted word overlap and (2) Euclidean distance of the content embeddings to retrieve the most similar sentences, while G-GST used cosine similarity between (1) TF-IDF weighted representation of the sentences, (2) averaged GloVe over words, and (3) Universal Sentence Encoder representation. DeleteAndRetrieve achieved the best performance on the task and a good balance between fluency, content preservation, and style strength. G-GST achieved worse performances possibly due to a weak retrieve mechanism. The *retrieve* component could be considered as a knowledge augmentation component that enhances the generation process with additional knowledge about the target style.

4. Exploratory Experiments into Prompting and Knowledge Augmentation for Formality Style Transfer

Prompting and knowledge augmentation techniques have been evaluated on the formality transfer task using the GYAFC [18] dataset. The dataset is composed of 165,030 parallel sentences in informal and formal styles. The goal was to transform an informal sentence into a formal sentence. In what follows our preliminary results of employing the T5-small and T5-base models [5] are presented. A total of five experiments have been performed for both T5-small and T5-base models:

- **Naive**: the input sentence is copied as an output sentence.
- **Retrieve**: a sentence from the set of sentences in the target style that is most similar to the input sentence is retrieved as an output sentence.
- **Fine-tune**: the model is fine-tuned in a standard way for a text generation task.
- **Zero-shot prompting**: output sentences are generated using the following prompt *"Rewrite from <input_style> style to <output_style> style: <input_sentence>."*.

[2] Style markers are words that have the most discriminative power for determining the style of a sentence.

- **Fine-tune with knowledge augmentation**: before being fed to the model, the input sentence is modified in the following format *"<input_sentence> | <knowledge>"* where *knowledge* is the sentence from the set of sentences in the target style that is most similar to the input sentence.

Following the prior studies on the text style transfer task, evaluation experiments have been performed across three dimensions: (1) content preservation, (2) style transfer strength, and (3) fluency. BLEU [31] was computed to measure semantic content preservation. Following the evaluation procedure of the Prompt-and-Rerank [32] method, self-BLEU (sBLEU) and reference-BLEU (rBLEU) were computed. Self-BLEU measures the degree to which the model directly copies the input sentence, while reference-BLEU measures the distance from the ground-truth references. To determine whether the generated sentences correspond to the target style, accuracy was calculated with a pre-trained DistilRoBERTa model on the task of formality detection as a percentage of the generated sentences that were labeled with the target style by the model. The perplexity of the generated sentences was calculated with a pre-trained GPT-2 [3] language model to measure their fluency.

PyTorch implementation of the models available in the HuggingFace Transformers library[3] and evaluation metrics available in the HuggingFace Evaluate library[4] have been used. All models have been trained with AdamW optimizer with a learning rate of 0.0001, weight decay of 0.0005, 25 epochs, and batch size of 32.

Table 1 summarizes the evaluation results for the experiments. The models fine-tuned with knowledge augmentation (T5-small KA and T5-base KA) achieved the best overall performance for all evaluation metrics compared with the standard fine-tuning and zero-shot prompting. The higher accuracy indicates the superior ability to generate sentences that are more consistent with the target style, while the lower perplexity indicates higher fluency and coherence of generated sentences for both models. T5-small KA and T5-base KA models exhibit higher rBLEU scores and lower sBLEU scores compared to their counterparts without knowledge augmentation indicating that the output sentences are closer to the ground-truth references in terms of n-gram overlap. The T5-base model achieved slightly better evaluation scores than the T5-small model which confirms the statement that increasing the model size improves the performances [30]. Compared with the two baseline approaches (Naive and Retrieve), this approach achieved better overall results than the naive approach, while the accuracy is lower and the perplexity is higher than the retrieve approach. The latter is expected since the retrieve approach does not generate new sentences and the retrieved ones are from the set of sentences in the target style. The zero-shot and fine-tune approaches obtain low rBLEU and high sBLEU suggesting that with these approaches the T5 model is prone to copying parts of the input sentences.

5. Conclusion

The recent advances with the LLMs in many NLP tasks have been the primary motivation for writing this brief survey. Utilizing LLMs improved performance on many tasks

[3]https://huggingface.co/docs/transformers/en/index, last visited: 15.02.2024
[4]https://huggingface.co/docs/evaluate/en/index, last visited: 15.02.2024

Table 1. Evaluation results for the formality transfer experiments with the T5-small and T5-base models on the GYAFC dataset. rBLEU - reference-BLEU, sBLEU - self-BLEU, PPL - perplexity.

	rBLEU	sBLEU	Accuracy	PPL
Naive	21.4	100	17.2	284
Retrieve	**48.2**	32.2	**93.1**	**85**
T5-small zero-shot	11.5	47.1	23.3	369
T5-base zero-shot	0.6	**2.6**	1.2	462
T5-small fine-tune	17.4	53.0	53.4	477
T5-base fine-tune	24.6	51.2	81.6	119
T5-small KA	40.2	**28.0**	81.5	163
T5-base KA	**44.9**	31.5	**90.3**	**90**

showcasing their capacity to comprehend and generate human-like text. Since GPT-3 demonstrated better performances than the state-of-the-art models at the time it was proposed without fine-tuning and being given only a few or no samples, a wide new research direction of prompting was opened. On the other hand, knowledge augmentation techniques have also gained attention in enhancing the capabilities of LLMs. This paper was focused on the application of LLMs for the task of text style transfer with a specific interest in prompting and knowledge augmentation techniques. Preliminary experimental results demonstrate that knowledge augmentation techniques contribute positively to model performance, as evidenced by improvements in all evaluation metrics for both T5-small and T5-base models compared to their non-augmented counterparts. Zero-shot prompting techniques exhibit limited effectiveness, with lower scores across all evaluation metrics, highlighting the importance of fine-tuning and knowledge augmentation. The results presented in this paper are preliminary from our initial experiments to support and validate the ideas that were discussed. We plan on extending the experiments with various LLMs for several other text style transfer tasks including politeness transfer, sentiment style transfer, Shakespearean style transfer, and others.

References

[1] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in neural information processing systems. 2017;30.

[2] Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. 2018.

[3] Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. OpenAI blog. 2019;1(8):9.

[4] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); 2019. p. 4171-86.

[5] Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. The Journal of Machine Learning Research. 2020;21(1):5485-551.

[6] Chung HW, Hou L, Longpre S, Zoph B, Tay Y, Fedus W, et al. Scaling instruction-finetuned language models. CoRR. 2022.

[7] Chowdhery A, Narang S, Devlin J, Bosma M, Mishra G, Roberts A, et al. Palm: Scaling language modeling with pathways. Journal of Machine Learning Research. 2023;24(240):1-113.

[8] Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. 2020.

[9] OpenAI. GPT-4 Technical Report. CoRR abs/230308774. 2023.

[10] Almazrouei E, Alobeidli H, Alshamsi A, Cappelli A, Cojocaru R, Debbah M, et al. Falcon-40B: an open large language model with state-of-the-art performance. Findings of the Association for Computational Linguistics: ACL. 2023;2023:10755-73.

[11] Thoppilan R, De Freitas D, Hall J, Shazeer N, Kulshreshtha A, Cheng HT, et al. Lamda: Language models for dialog applications. arXiv preprint arXiv:220108239. 2022.

[12] Touvron H, Lavril T, Izacard G, Martinet X, Lachaux MA, Lacroix T, et al. Llama: Open and efficient foundation language models. arXiv preprint arXiv:230213971. 2023.

[13] Touvron H, Martin L, Stone K, Albert P, Almahairi A, Babaei Y, et al. Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:230709288. 2023.

[14] Hadi MU, Qureshi R, Shah A, Irfan M, Zafar A, Shaikh MB, et al. Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects. Authorea Preprints. 2023.

[15] Zhao WX, Zhou K, Li J, Tang T, Wang X, Hou Y, et al. A survey of large language models. CoRR. 2023.

[16] Guo Z, Jin R, Liu C, Huang Y, Shi D, Yu L, et al. Evaluating large language models: A comprehensive survey. arXiv preprint arXiv:231019736. 2023.

[17] Jhamtani H, Gangal V, Hovy E, Nyberg E. Shakespearizing Modern Language Using Copy-Enriched Sequence to Sequence Models. In: Proceedings of the Workshop on Stylistic Variation; 2017. p. 10-9.

[18] Rao S, Tetreault J. Dear Sir or Madam, May I Introduce the GYAFC Dataset: Corpus, Benchmarks and Metrics for Formality Style Transfer. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers); 2018. p. 129-40.

[19] Cho K, van Merriënboer B, Bahdanau D, Bengio Y. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. In: Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation; 2014. p. 103-11.

[20] Li J, Jia R, He H, Liang P. Delete, Retrieve, Generate: a Simple Approach to Sentiment and Style Transfer. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers); 2018. p. 1865-74.

[21] Sudhakar A, Upadhyay B, Maheswaran A. "Transforming" Delete, Retrieve, Generate Approach for Controlled Text Style Transfer. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); 2019. p. 3269-79.

[22] Xu R, Ge T, Wei F. Formality Style Transfer with Hybrid Textual Annotations. CoRR. 2019;abs/1903.06353. Available from: http://arxiv.org/abs/1903.06353.

[23] Zhu X, Guan J, Huang M, Liu J. StoryTrans: Non-Parallel Story Author-Style Transfer with Discourse Representations and Content Enhancing. CoRR. 2022.

[24] Yang Y, Zhao Z, Liu Q. MSSRNet: Manipulating Sequential Style Representation for Unsupervised Text Style Transfer. CoRR. 2023.

[25] Madaan A, Setlur A, Parekh T, Póczos B, Neubig G, Yang Y, et al. Politeness Transfer: A Tag and Generate Approach. In: Jurafsky D, Chai J, Schluter N, Tetreault JR, editors. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020. Association for Computational Linguistics; 2020. p. 1869-81. Available from: https://www.aclweb.org/anthology/2020.acl-main.169/.

[26] Jin D, Jin Z, Hu Z, Vechtomova O, Mihalcea R. Deep learning for text style transfer: A survey. Computational Linguistics. 2022;48(1):155-205.

[27] Toshevska M, Gievska S. A review of text style transfer using deep learning. IEEE Transactions on Artificial Intelligence. 2021.

[28] Wei J, Bosma M, Zhao V, Guu K, Yu AW, Lester B, et al. Finetuned Language Models are Zero-Shot Learners. In: International Conference on Learning Representations; 2021. .

[29] Schick T, Schütze H. Few-shot text generation with natural language instructions. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing; 2021. p. 390-402.

[30] Reif E, Ippolito D, Yuan A, Coenen A, Callison-Burch C, Wei J. A Recipe for Arbitrary Text Style Transfer with Large Language Models. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers); 2022. p. 837-48.

[31] Papineni K, Roukos S, Ward T, Zhu WJ. Bleu: a method for automatic evaluation of machine translation.

In: Proceedings of the 40th annual meeting of the Association for Computational Linguistics; 2002. p. 311-8.
[32] Suzgun M, Melas-Kyriazi L, Jurafsky D. Prompt-and-Rerank: A Method for Zero-Shot and Few-Shot Arbitrary Textual Style Transfer with Small Language Models. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing; 2022. p. 2195-222.
[33] Luo G, Han YT, Mou L, Firdaus M. Prompt-Based Editing for Text Style Transfer. CoRR. 2023.
[34] Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach. 2019.
[35] Russell SJ, Norvig P. Artificial intelligence a modern approach. London; 2010.
[36] Zhen C, Shang Y, Liu X, Li Y, Chen Y, Zhang D. A Survey on Knowledge-Enhanced Pre-trained Language Models. arXiv e-prints. 2022:arXiv-2212.
[37] Wang C, Liu X, Song D. Language Models are Open Knowledge Graphs. 2020.
[38] Roberts A, Raffel C, Shazeer N. How Much Knowledge Can You Pack Into the Parameters of a Language Model? In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP); 2020. p. 5418-26.
[39] Zhang Y, Xu J, Yang P, Sun X. Learning Sentiment Memories for Sentiment Modification without Parallel Data. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing; 2018. p. 1103-8.