氏　　名	Xu Sheng（徐　勝）
博士の専攻分野の名称	博士（工学）
学 位 記 番 号	医工農博甲第１６４号
学位授与年月日	令和７年３月１９日
学位授与の要件	学位規則第４条第１項該当
専　　攻　　名	工学専攻　システム統合工学コース
学 位 論 文 題 目	A Study on Text Style Transfer via Pre-trained Language Models　(事前学習に基づく言語モデルによるテキストスタイル変換に関する研究)
論 文 審 査 委 員	主査　教　授　福本　文代
	教　授　鈴木　良弥
	教　授　郷　健太郎
	教　授　西崎　博光
	教　授　張本　鉄雄
	准教授　李　吉屹

学位論文内容の要旨

This thesis focuses on five PLM-based approaches for the Text Style Transfer (TST) task. In any language, text serves as a medium for recording and transmitting information. From a macro perspective, text can be categorized into style and style-independent content. Content typically refers to the core message we intend to convey. At the same time, style pertains to how the content is expressed, such as whether the tone is positive/negative or formal/informal. Therefore, all TST tasks can be interpreted as transforming an input text with an original style into an output text that preserves content while adopting a target style. TST can include a range of sub-tasks, such as Positive Text Reframing (PTR), Sentiment Style Transfer (SST), and Formality Style Transfer (FST), depending on the type of style pair involved. Building upon a review and analysis of the existing research in TST, this thesis proposes two methods for fine-tuning and three studies based on prompting techniques.

Chapter 1 introduces the definition of text and TST, surveys three main approaches in previous research, and illustrate the applications of TST. To demonstrate the innovation

and motivation of our five research works, we first, in Chapter 2, review and summarize the previous approaches based on pre-trained language models (PLMs) to TST tasks before presenting each of our works in detail.

Following the decomposition of the PTR task, we introduce a novel fine-tuning method and a data augmentation strategy for auxiliary tasks in Chapter 3. Unlike previous methods that directly perform end-to-end fine-tuning of PLMs using parallel datasets to model the PTR task, we first analyze the PTR task and decompose the transformation process into two sub-tasks: paraphrase generation and style transfer. Furthermore, to enhance the model's performance in these sub-tasks, we constructed two pseudo-datasets using existing paraphrasing and sentiment datasets for further fine-tuning of the PLMs.

While decomposing the PTR task at the task level provides the aforementioned advantages, the overall performance of the model may be influenced by the quality of the two pseudo-datasets. If the data used for training the auxiliary tasks is of low quality, it may adversely affect the PLM's performance on the PTR task. To this end, in Chapter 4, we propose a new disentangling training objective based on fine-tuning to mitigate potential errors introduced by pseudo-parallel data augmentation and to rely only on the PPF dataset. Furthermore, to enable the model to learn the more precise and fine-grained data features implicitly embedded in the PTR sentence pares, this approach uses contrastive learning to constrain the hidden space of the Transformer model, effectively separating style representations from content hidden states, and controlling the generation of sentences with the target style representations. Through experiments, we found that our fine-tuning method improved the style transfer strength of the model on two different PLMs (BART and T5). Compared to the baseline, the model was able to generate more fluent sentences.

While the large language models (LLMs) are inherently a special case of PLM, their superior generalization ability offers significant advantages compared to fine-tuning large models, which are more computationally expensive. In contrast, prompting-based methods can efficiently improve overall model performance. Moreover, since these methods do not rely on parallel datasets, LLMs based on prompting can be applied to a wide range of downstream tasks. Therefore, in Chapters 5 through 7, we explore three different prompting strategies in TST sub-tasks such as SST, and FST.

We first analyze the characteristics of the SST task and find that for the same model or prompting pipeline the difficulty of transferring varies along the diversity of input sentences in Chapter 5. Especially, we define an intuitive method to assess the transfer difficulty of a specific case. For the more challenging transfer cases, we decompose the language model's operations into two steps, reduction and synthesis (RS). To implement our idea, we propose a novel Plug-and-Play strategy. The experiment results on two popular SST datasets, Yelp and Amazon, demonstrate that our RS pipeline improves the style transfer strength of the baseline model in more complex transfer cases.

In Chapter 6, we refocus on the prompt template and aim to enhance the baseline's performance on the SST task by improving the quality of the prompt templates. To this end, based on a review of previous prompting strategies, I explored the usage of the aspect-based sentiment analysis (ABSA) model for constructing dynamic prompt templates. This strategy mitigates the limitations of static prompts, which often fail to adapt accurately and effectively to diverse input cases. Similarly, we also investigated a new variant of dynamic prompting, in which the ABSA model is applied to the self-refinement algorithm to dynamically construct refinement prompting templates. In contrast to the use of self-refine in Chapter 5, the feedback is predicted by the ABSA model rather than generated by the LLM based on the examples provided in the prompt.

The approaches introduced from Chapter 3 to 6 focus on a specific TST sub-task related to sentiment. To fully harness the generalization capabilities of LLMs, we further investigated the performance of LLM-based prompting methods across four different TST sub-tasks in Chapter 7. Based on two fundamental perspectives, disentanglement, and entanglement, we utilize a chain-of-thought (CoT) approach to directly decouple or couple the content and style components of the input sentence at the linguistic level. We then design two prompting methods to implement the disentangle and entangle pipelines. Through experiments across seven datasets for TST subtasks, we analyze the performance of these strategies both individually and in combination and demonstrate their effectiveness.

Lastly, in Chapter 8, we present a comprehensive summary of the five research studies discussed above. Through analysis from three distinct perspectives (dataset, model, and evaluation), we systematically assess their strengths, weaknesses, contributions, and

limitations, while identifying promising directions for future research.

論文審査結果の要旨

令和 7 年 2 月 4 日 17:00 より，論文題目：A Study on Text Style Transfer via Pre-trained Language Models (事前学習に基づく言語モデルによるテキストスタイル変換に関する研究) に関する学位論文審査，続く 18:00 より同博士論文に関する最終試験を実施した.

Sentiment style transfer (SST)，すなわちテキストの文体を変換・生成するタスク，及びそのサブタスクに相当する Positive text reframing (PTR) task，すなわち入力されたテキストの意味を保持したまま，肯定的なテキストを生成するタスクは，近年 AI 生成の進展により自然言語処理の応用分野の一つとして注目を集めている. 従来の手法の多くは，事前学習済の言語モデル（Pre-training language model）の優れた表現能力を利用し，このモデルをファインチューニング（fine-tuning）することにより高精度なベースラインが達成できている. しかし，多様な文脈をどのようにとらえ，流暢なテキストを生成するかは，依然として未解決の問題となっている. 特に，事前学習で用いるデータ量が少ない場合，この問題はさらに深刻である. 本研究は PTR，及び SST に着目し，与えられたテキスト（入力文）から，Entanglement，すなわち直接出力文を生成するアプローチと Disentanglement であるテキストの意味とスタイルを分割した後，生成すべき文の内容と出力スタイルとを統合するアプローチに着目し，5 つの手法を提案している.

先ずはじめに事前学習で用いるデータ量が少量であるという問題に着目し，2 種の拡張戦略を提案することにより，疑似的な訓練データセットを作成した. さらにこれらデータに対し，マルチタスク学習を適用しテキストの意味とスタイルを独自に学習後，統合する disentanglement 手法を提案することにより流暢なテキストを生成することに成功している. PTR のベンチマークデータセットを用いた実験では，事前学習モデルである BART，及び T5 を用い，6 種類の評価尺度で評価した結果，ベースラインよりも上回っていることが確認できた. しかし，本手法は，大量の疑似的な訓練データを必要とすること，さらに本手法で提案している 2 種の拡張戦略はいずれも hidden-space での正規化が行われていないことから，精度面でベースラインを上回っているものの，依然課題が残されている. そこで新たに対照学習を導入することによりこの問題を解決する 2 つ目の手法を提案している.

上記で提案している 2 つの手法は，fine-tunning において，高品質な入力文と正解出力文とのペアを必要とすることから，様々なジャンル，及び TST に適用することは難しい. そこでタスクを TST に拡張し，新たに 3 つの prompting 手法を提案している. 具体的には，テキストの意味とスタイルを分割する disentanglement 手法において新しい PTR の枠組みである

Reduction-and-Synthesis 手法を開発している．

Reduction-and-synthesis に関する一つ目の手法は，多様な表現でポジティブな印象を与えるテキストを生成するために，Plug-and-Play 手法による生成手法である．具体的には，極性判定に着目し，判定が正解であるものは入力文を直接変換する．一方，極性が誤っている，すなわち変換が困難な入力文に対しては，テキストの肯定や否定といった極性スタイルとテキスト内容を分離した上で，テキスト内容を多様な表現で言い換える手法を提案している．Yelp dataset を用いた実験の結果，提案手法が評価尺度のうちスタイル，及び流暢さにおいて高い精度が得られている反面，内容を評価する BLEU スコアにおいてはベースラインを上回ることができなかった．理由としてベースライン手法は，入力原文をコピーし出力としているものが多いため，高い BLEU スコアが得られているためである．一方，人手による評価を実施した結果，本手法が内容，スタイル，流暢さのいずれにおいてもベースラインを上回る結果が得られている．

上記で述べた提案手法は 2 つの推論により最終的な文を生成しているため，コスト面，及びハルスネーションについて十分な対応ができていないことが問題として残されていた．そこで Reduction-and-synthesis に関する二つ目の手法として，Static Prompt と Dynamic Prompt を定義し，入力文によりこれらを適用することにより文を生成する手法を提案している．定量的な実験の結果，スタイル変換については Dynamic Prompt，文内容の保持と出力文の流暢性については Static Prompt が安定してよい結果が得られることを確認している．

本研究における 5 つ目の手法は，本研究のコアである Entanglement と Disentanglement 手法の各々において，Chain-of-thought 手法により生成までの過程を prompt として随時入力することにより LLM の生成能力を高める手法である．7 種類のデータ，及び 6 種類の LLM を用いた定量的な実験の結果，データ，及び LLM の種類によらず，Entanglement に Chain-of-thought を統合した手法はスタイルと流暢さにおいて優れていること，Disantanglement に Chain-of-thought を統合した手法は，BLEU において高い精度が得られると結論づけている．

博士論文で提案しているこれら 5 つの手法は別紙 3 で記載されている論文として，プログラムコードと共に公開されている．

公聴会では，各手法に関する確認の他，1. 同じデータと LLM を用いた 5 つの提案手法の比較は可能か，2. LLM 単体と比較した場合，どの程度精度の向上が見込めるか，3. 多くの手法は，流暢さは高いが文の保持は低い，あるいはその逆の結果が得られている．その理由，及び共に高い精度を得るための工夫はあるか，などの質疑が行われ，これらについ

て議論した.

続く最終審査では，1. TST について，例えば肯定/否定であればその度合い，すなわちどの程度高い(低い)肯定であるか，そこまで生成することは可能か，さらに LLM の prompt を利用することにより，極性の程度を含め出力することに関してどう考えるか，2. 最後の手法（5 つめの手法)の実験に関する計算量，及びその高速化に関してアイデアはあるか，3. 今後の課題，特に評価手法，及びスタイルの定義についてどのように考えるか，などについて議論が行われた. 最後に，博士論文に関する査読者からの種々のコメントを反映し，最終論文として提出することを確認した.

PTR 含む TST に関する問題に対し，5 つの手法を提案していること，それぞれにおいて定量的な実験を実施し，評価尺度，及び人手による評価から結果のエラー解析を詳細に実施していること，及びこれらを 4 編の国際会議論文として公表，プログラムコードを公開していることから，質・量において本学の博士として相応しいと判断し，論文審査，及び最終審査を合格とした.