CLSEG: CONTRASTIVE LEARNING OF STORY ENDING GENERATION

Yuqiang Xie, Yue Hu, Luxi Xing, Yunpeng Li , Wei Peng , Ping Guo*

Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

Introduction

Interesting Problem

The story ending generation methods based on Pre-trained Language Model and the MLE training objective lack the support of causal, temporal, and sentiment knowledge, resulting in poor consistency between generated story ending and the story context.

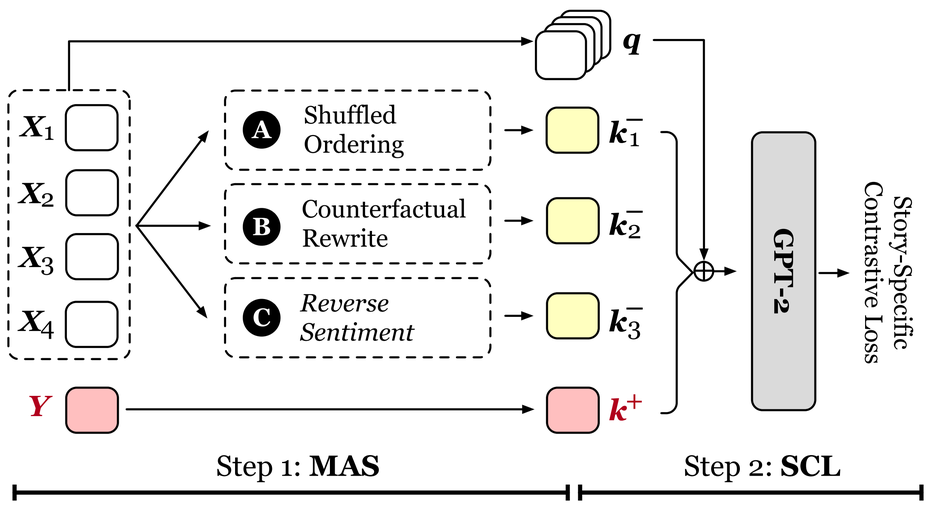
Two crucial challenges in CLSEG:

A. How to generate high-quality wrong endings inconsistent with story contexts?
B. How to design a contrastive training strategy adapted for story ending generation?

Contributions:

I. The **Multi-Aspect negative sampling** is performed from the perspectives of causality, temporal, and sentiment;
II. A **Story-level Contrastive Learning Objective** is designed, which enables the model to compare multi-source knowledge during the training process and improves consistency of the generated story ending;
III. Automatic and manual evaluations demonstrate that CLSEG outperforms baselines and can generate story endings with stronger consistency and rationality.

Overview of CLSEG

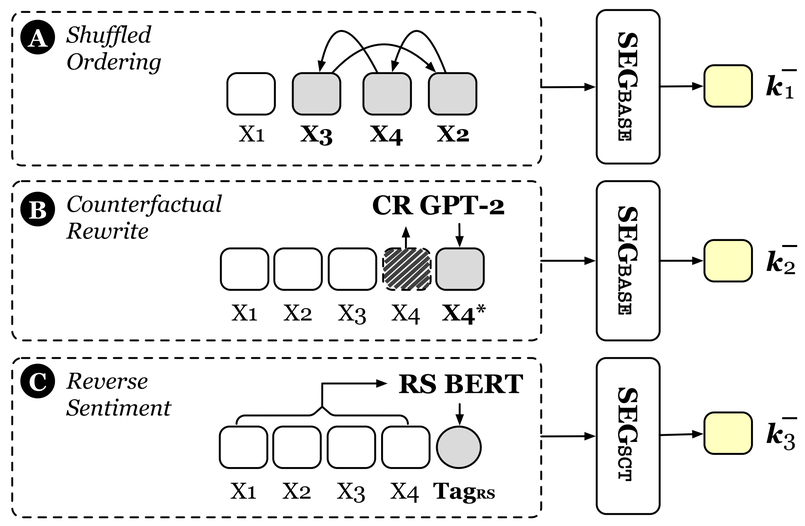


We propose a novel Contrastive Learning framework for Story Ending Generation (CLSEG)†, which has two steps: multi-aspect sampling and story-specific contrastive learning.

Details of CLSEG

Multi-Aspect Sampling

generate high-quality wrong endings



order of events is different -> inconsistent ending

$$\mathcal{Y}_{\text{SO}}^{-} = \text{SEG}_{\text{BASE}}(\textit{shuffled}(\{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_m\}))$$

what if event A turns B -> inconsistent ending

$$\mathcal{Y}_{\text{CR}}^{-} = \text{SEG}_{\text{BASE}}(\{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_m^{*}\})$$

change the sentiment orientation -> inconsistent ending

$$\mathcal{Y}_{\text{RS}}^{-} = \text{SEG}_{\text{SCT}}(\{\mathcal{X}_1, \mathcal{X}_2, ..., \mathcal{X}_m, \text{Tag}_{\text{RS}}\})$$

Notes:

- **SEG$_{\text{BASE}}$:** GPT-2 fine-tuned on **ROCStories** corpus
- **SEG$_{\text{SCT}}$:** GPT-2 fine-tuned on selected stories of **SCT**, where wrong ending has reversed sentiment to the story context
- **CR GPT-2:** GPT-2 fine-tuned on counterfactual story corpus, **TIMETRAVEL**
- **RS BERT**: BERT$_{\text{LARGE}}$ fine-tuned on **SST-2**

Story-level Contrastive Learning Objective

adapted for story ending generation

$$\mathcal{L}_{\text{SCL}}^{t}\left(p_\theta\left(\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{X}}\right), \{\boldsymbol{\mathcal{Y}}_i^{-}\}_{i=1}^{3}\right) = -\log p_\theta\left(\boldsymbol{k}_t^{+} \mid \boldsymbol{k}_{<t}^{+}, \boldsymbol{q}\right)$$
$$-\alpha \cdot \frac{1}{N}\sum_{i}^{N}\log\left(1 - p_\theta\left(\boldsymbol{k}_{i,t}^{-} \mid \boldsymbol{k}_{i,<t}^{-}, \boldsymbol{q}\right)\right)$$

where $\boldsymbol{\mathcal{X}}$ denoted as $\boldsymbol{q}$
$\{\boldsymbol{\mathcal{Y}}_{\text{SO}}^{-}, \boldsymbol{\mathcal{Y}}_{\text{CR}}^{-}, \boldsymbol{\mathcal{Y}}_{\text{RS}}^{-}\}$ denoted as $\{\boldsymbol{k}_i^{-}\}_{i=1}^{3} = \{\boldsymbol{k}_1^{-}, \boldsymbol{k}_2^{-}, \boldsymbol{k}_3^{-}\}$ for SCL

Experiment & Analysis

Dataset: ROCStories (Train 78,530/Dev 9,816/Test 9,816)

Baselines

1. **GPT-2(PT):** vanilla GPT-2.
2. **GPT-2(FT):** vanilla GPT-2 fine-tuned on the ROCStories corpus with LM objective.
3. **GPT-2(GCL):** This model is pre-trained similar to our model, only the negative sampling is the general noisy ways: random shuffle/drop/replace tokens.

Human Evaluation

Metrics

1. Content Quality: whether the generated story ending is fluent and coherent.
2. Content Rationality: whether the story endings are consistent with the story context.

Results

Model	Quality ↑		Consistent ↑		
	Fluency	Coherence	Order	Causal	Sentiment
GPT-2(PT)	1.87	1.58	1.22	1.07	1.19
GPT-2(FT)	**2.24**	2.13	1.47	1.67	1.68
GPT-2(GCL)	2.08	2.07	1.62	1.73	1.63
CLSEG	2.20	**2.43**	**2.08**	**2.19**	**2.02**

Table 2: Manual Evaluation in terms of quality and rationality about the generated story endings.

Main Results

Models	BLEU ↑	R-1-P ↑	R-1-R ↑	R-1-F1 ↑	R-2-P ↑	R-2-R ↑	R-2-F1 ↑	R-L-P ↑	R-L-R ↑	R-L-F1 ↑	Meteor ↑
GPT-2(PT)	1.14	13.99	**15.99**	13.56	1.34	**2.51**	1.68	13.17	**15.89**	12.65	10.36
GPT-2(FT)	**2.57**	15.15	12.55	13.30	**2.12**	1.82	1.87	13.87	11.45	12.14	10.48
GPT-2(GCL)	1.71	14.81	14.66	14.21	1.82	2.13	1.89	13.67	12.56	13.12	10.54
CLSEG	1.97	**15.68**	14.73	**14.63**	1.87	2.15	**1.91**	**14.44**	13.61	**13.48**	**10.73**

Table 1: The results of automatic evaluation considering common-used metrics. ↑ indicates the higher, the better.

Case Study

Context	Yesterday, I needed to buy a belt so I went to the mall. After choosing which store to shop from, I quickly got out of my car. Upon going inside there was an associate to help me. We talked about what i wanted and she led me to the section.
Golden	I picked out a great belt and left feeling excellent!
GPT-2(PT)	I was so out of my car I was about to buy a belt.
GPT-2(FT)	I was so quickly and quickly, i was so out of my car.
GPT-2(GCL)	I was so so so ...
CLSEG	After I got my belt, I went to the store.

Table 4: Generated story endings by different models.

Ablation Study

Model	BLEU ↑	R-1-R ↑	R-2-R ↑	R-L-R ↑	Meteor ↑
CLSEG	1.97	**14.73**	**2.15**	**13.61**	**10.73**
only SO	1.94	13.77	1.96	12.78	10.16
only CR	2.13	13.71	1.96	12.66	10.20
only RS	**2.33**	14.24	2.03	13.02	10.64

Table 3: Ablation study of **CLSEG**.

Error Analysis

Error Type	Story Context + Ending
Repetition	Morgan enjoyed long *walks* on *the beach*. ... Morgan decided to *propose* to her boyfriend. *The walk to the beach and the propose.*
Conflicting	Frank ... Since Frank was already a bit drunk, he could not drive. *Frank had to drive to his date.*
Ambigious	Sunny enjoyed going to the beach. As she stepped out of her car, ... Sunny got back into her car and heading towards the mall. *She was going to the park.*

Table 5: Typical errors generated by **CLSEG**. *Italic* words denote the error generated story endings.

Acknowledgments

We thank all anonymous reviewers for their constructive comments. This work is supported by the National Natural Science Foundation of China (No.62006222 and No.U21B2009).

University of Chinese Academy of Sciences 中国科学院大学 INSTITUTE OF INFORMATION ENGINEERING,CAS 中国科学院信息工程研究所 icassp 2022 Singapore

E-Mail: xieyuqiang@iie.ac.cn **Homepage:** https://yuqiangx.github.io