

FAIRYTAILOR:

A Multimodal Generative Framework for Storytelling

Contents

1. Introduction
2. Datasets
3. System Architecture
4. FairyTailor Demo
5. Experiment
6. Conclusion

1. Introduction

Introduction

- Automated Story Generation
 - 다른 자연어 생성보다 어려움
 - Why? 개방적, 장기 기억 유지, 주제와 어조를 모두 고려해야함
- Multimodal Content
 - 오디오, 비디오나 이미지가 텍스트와 같이 구성된 것
 - 어린 독자에게 더욱 매력적인 결과 생성 가능
- Interactive Writing Platform
 - 사용자-플랫폼 상호작용
 - 최종 편집은 사용자가



FAIRYTAILOR

Introduction



FAIRY TAIL

[Shuffle Story](#)



tab for Autocomplete (2 sec)



High-Quality Autocomplete (5 sec)

Image Style



The Magician

Once upon a time, there lived a great man who was great in influence. He ruled over a country which boasted the great treasures which lie hidden in every part of it. His subjects were called the people of Magic, and were always in quest of new wonders, such as the Winged Dragon and the Dragon's Tongue.



2. Datasets

Datasets

- Text Dataset
 - Reddit Writing Prompts (open-source)
 - Project Gutenberg (hand-picked)
 - 전처리
 - 최대 단어 수 1,000개
 - 특수 문자 및 기호 삭제
 - 불쾌감, 부정적 감정 포함된 단어 삭제
 - 프롬프트와 스토리 사이의 end-of-sentence 추가하여 사용
- Image Dataset
 - Unsplash
 - 이미지의 개체와 풍경의 feature가 다양
 - Flickr30k
 - 초기 모델에 이미지 검색으로 활용

3. System Architecture

System Architecture

Output

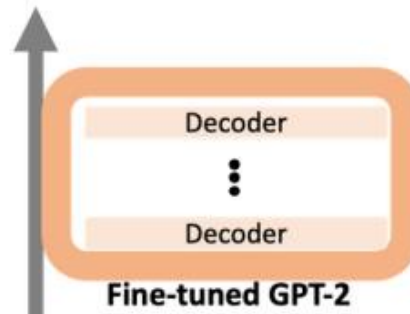
You may think that I may have an
Improper view of my fellow men, but
I can see only one point. A man's
heart changes when a man comes to
accept his object as more important
than his hopes and hopes change his
heart.

(3) Images coherency re-ranking + Style transfer

(2) Image retrieval by caption-generated text cosine similarity

Coherency | Readability | Sentiment | Diversity | Simplicity | Tale Like

(1) Per sentence Rule Based Re-ranking



Fine-tuned GPT-2

Input

The Myth of the Objective \n



System Architecture

- Readability (가독성)

$$readability = 0.5 * word_chars + sent_words \quad (2)$$

- 텍스트가 얼마나 복잡한지 추정
- word_chars, sent_words는 각각 단어 수와 문장 수가 0인 경우, -10을 부여
- 문장 당 단어 수에 더 높은 순위를 부여하기 위해 word_chars에 0.5를 곱함

- Sentiment (감정)

- SentiWordnet을 사용해 감정 점수를 계산
- 긍정적 감정에 더 높은 순위를 부여

System Architecture

- Diversity (다양성)

$$diversity = \frac{\text{len}(\text{set}(filtered_words))}{\text{len}(filtered_words)} \quad (3)$$

- 총 단어 수에서 고유한 단어의 비율을 계산
- `filtered_words`는 stop words (ex; at, in, is)와 구두점을 제외한 단어 토큰
- 필터링 된 단어가 없다면 점수는 0

- Simplicity (단순성)

$$simplicity = \text{len}(\text{set}(filtered_words) \cap freq_words) \quad (4)$$

- 주어진 텍스트에서 스토리와 유사한 특징적인 단어의 비율을 계산
- `freq_words`는 Gutenberg 동화 말뭉치에서 가장 빈번한 단어 7%의 단어 토큰

System Architecture

- Coherency (일관성)
 - 첫 번째 문장과 비교하여 스토리 문장 내에서 Latent Semantic Analysis 유사성을 계산
- Tale like (동화 유사성)
 - 사전 설정된 GPT-2와 fine-tuned GPT-2의 생성 텍스트의 예측 점수 사이의 KL 발산 손실을 계산
 - 텍스트가 fine-tuned 분포와 더 유사하다면 좋은 것
 - 점수가 높을 수록 좋음

4. FairyTailor Demo

<https://fairytailor.org>

FairyTailor Demo

- Autocomplete
 - Fine-tuned 모델에서 생성된 3개의 문장을 즉시 반환
 - 빠르고 간단함
- High-Quality Autocomplete
 - 10개의 문장을 생성
 - 그 후, 앞서 설명한 점수로 순위를 다시 매긴 후, 상위 3개 반환
 - 기본 생성보다 오래 걸리지만 더 품질 좋음
- Human vs AI edits
 - 사용자는 생성된 문장 및 이미지를 원하는 대로 추가, 삭제, 편집 가능
 - 생성된 데이터는 텍스트와 다르게 표시

FairyTailor Demo



FAIRY TAILOR

[Shuffle Story](#)



[tab](#) for Autocomplete (2 sec)



High-Quality Autocomplete (5 sec)

Image Style



The Magician

Once upon a time, there lived a great man who was great in influence. He ruled over a country which boasted the great treasures which lie hidden in every part of it. His subjects were called the people of Magic, and were always in quest of new wonders, such as the Winged Dragon and the Dragon's Tongue.



FairyTailor Demo

WHY THE CHASM DOES NOT WANT TO SAVE THE LAD: Because no Man can cure the Curse

by JASON STIPEKNER THE SHADOW UNDER THE SEA THERE appeared a white dragon with scales like polished teeth

As we have seen that the spells by which the Ocean sends forth all that it calls upon must be a power of thought,



(a) Autocomplete

In the spring and summer you can feel the ocean washing off your shoes. We know now that the spell of the gods is

As the water began to flood the walls, Harry put out his hand to stop her and was drowned. Then he raised

When the spell was over, the two men went into the ocean together. There, under the surface, they saw an immense



(b) High-Quality Autocomplete

5. Experiment

Experiment

- Qualitative Evaluation

- 스토리 생성 전문가와 함께 데모에 대한 피드백을 수집
- + 13명의 추가 인원으로 human evaluation 진행

1. Email Address*
2. Have you written stories before? If yes, elaborate on the intended audience and the stories' structure*
3. Paste the URL of your story (created after pressing "submit story" at the bottom)*
4. Do you agree with the following statement?* (Choose one of strongly Disagree, Somewhat Disagree, Neither Agree nor Disagree, Somewhat Agree, Strongly Agree).Purdy et al. [2018]
 - Autocompletes exhibit CORRECT GRAMMAR
 - Autocompletes occur in a PLAUSIBLE ORDER
 - Autocompletes MAKE SENSE given sentences before and after them.
 - Autocompletes AVOID REPETITION
 - Autocompletes use INTERESTING LANGUAGE
 - This story is of HIGH QUALITY.
 - This story is ENJOYABLE.
 - This story follows ONE OVERALL THEME.
5. When prompted, how often did you decline to use the suggested autocompletions?*
6. Did you use autocomplete (Tab) or High-Quality autocomplete (shift + Tab)?*
7. Please elaborate on your choice above*
8. Did you use images? Why?*
9. What did you like?*
10. What not so much?*
11. Other comments/ suggestions?

Results

- Demo Interface
 - 전반적으로 참가자들은 데모에 대해 긍정적
 - 자동 완성된 문장으로 자동 완성을 다시 하는 것이 매력적
- Autocomplete vs. High-Quality Autocomplere
 - 두 가지 모두 사용한 참가자는 high-quality를 활용하는 것이 더 우수하다고 함
 - 퀄리티가 좋아 더 오랜 시간이 걸려도 기다릴 의향이 있음
- Multimodal Framework Completions
 - 제출된 스토리의 1/3은 스토리와 관련이 없거나 자신의 스타일과 맞지 않아 이미지를 사용하지 않음
 - 텍스트 또한, 어조나 어휘가 사용자의 스타일과 맞지 않으면 사용하지 않음

Results

- Published Stories
 - 공개된 co-created 스토리의 평균 점수는
 - clarity: 3.4, coherence: 3.5, creativity: 3.7
 - 참가자 대부분이 창의성과 동화 같은 특성을 강조함

Score Auto-Generated Story

Clarity

☆☆☆☆☆

Coherence

☆☆☆☆☆

Creativity

☆☆☆☆☆

Any Comments, Questions or Suggestions

0 / 150

☐ Do you agree to submit your feedback and publish your story?

SUBMIT & SHARE STORY

6. Conclusion

Conclusion

- Strengths

- 사용 용이성이 좋고 플랫폼 디자인이 좋음
- 사용자(작가)에게 스토리 작성을 안내하는데 있어 도움이 됨

- Weaknesses

- 일부 사용자는 스타일이나 내용이 스토리에 맞지 않아 이미지를 사용하지 않음
- 이미지 검색은 이미지 데이터셋으로 제한되어 모든 스토리에 적합한 이미지를 찾을 수 없음
- 기본 자동 생성을 이용하는 경우, 자동 완성 사용 거부율이 50%-75%로 높음
- 그러나, 고품질 자동 생성의 경우는 0-25%
- 이는, 콘텐츠 품질에 시간을 할애하는 것이 중요하고 앞으로는 더 빠른 고품질 모델을 연구할 것