

학습 데이터 용량 및 반복 학습 횟수에 따른 이미지 기반 GPT 문장생성 및 성능 분석

이동희^o, 최봉준^{*}

^o동서대학교 소프트웨어학과,

^{*}동서대학교 소프트웨어학과

e-mail: dukeldh1128@gmail.com^o, bongjun.choi@dongseo.ac.kr^{*}

Analyze GPT sentence generation performance based on Image by training data capacity and number of iterations

Dong-Hee Lee^o, Bong-Jun Choi^{*}

^oDivision of Software, Dongseo University,

^{*}Division of Software, Dongseo University

● 요약 ●

현재 많은 사람이 GPT를 통해 다양한 활동 및 연구를 진행하고 있다. 사람들은 GPT를 통해 문장생성 시 문장에 대한 정확도를 중요하게 생각한다. 하지만 용도에 따라 GPT를 통해 생성하는 문장의 문체와 같은 표현방식이 다르다. 그래서 생성된 문장이 유의미한 문장이라는 것에 판단이 매우 주관적이기 때문에 수치적 평가가 어렵다. 본 논문에서는 자연어처리 모델이 생성한 문장의 유의미함을 판단하기 위해 각 모델을 학습하는 데이터 용량과 반복 학습의 횟수에 따른 결과물을 비교하였다. 본 연구에서는 Fine-Tuning을 통해 총 4개의 GPT 모델을 구축하였다. 각 모델로 생성 문장을 BLEU 평가지표를 통해 평가한 결과 본 연구에 BLEU 모델은 부적합하다는 결과를 도출하였다. 이를 해결하기 위해 본 연구에서는 생성된 모델을 평가하고 자 설문지를 만들어 평가를 진행하였다. 그 결과 사람에게 긍정적인 평가를 받는 결과를 얻을 수 있었다.

키워드: GPT, Fine-Tuning, 데이터 용량(data capacity), 모델평가(evaluating models), BLEU

I. Introduction

최근 GPT를 활용한 많은 연구 활동이 활발하다. 보편적으로 GPT는 ChatGPT와 같이 상용화된 서비스를 통해 게시물을 작성하고 과제와 업무에도 많이 활용한다. 하지만 한글 GPT 사용 시 문장이 부자연스럽거나 문장의 맥락이 맞지 않아 수정하여 사용하는 문제가 발생하고 있다. 이 문제를 보완하기 위해 GPT 모델을 학습하여 사람들에게 제공할 수 있다. 하지만 학습된 모델이 생성한 문장이 유의미한지 판단할 수 있는 평가가 필요하다.

본 논문은 GPT의 성능을 파악하기 위해 데이터양과 반복 학습 횟수에 따른 GPT 모델을 생성하여 모델별 문장을 비교하고 분석한다.

n-gram을 기반으로 일치하는 n-gram 개수, 문장 길이, 보정 항목, 가중치 등을 종합적으로 고려하여 score를 측정한다.

2. 이미지 객체 탐지 기반 문장 생성

사용자에게 게시물에 사용할 이미지를 삽입하면 이미지 Object Detection을 통해 객체 키워드를 추출하고 해당 키워드로 문장을 생성하는 연구이다.

II. Preliminaries

1. BLEU

기계 번역된 문장과 사람이 작성한 기대 번역 문장 사이의 일치 정도를 비교하여 성능을 측정하는 방법이다. BLEU는 n개의 연속적인

III. The Proposed Scheme

본 연구는 II-2 연구의 이미지 기반으로 생성한 문장 성능을 파악하기 위한 연구를 수행한다.

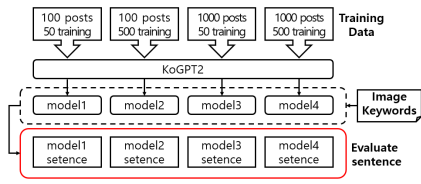


Fig. 1. System Concept

1. Create Model

본 연구에서는 KoGPT2 모델에 데이터 용량 및 반복 학습 횟수의 영향력을 파악하기 위해 비교문장을 생성할 모델이 필요하다. 학습을 위해 100, 1,000개의 게시물 데이터를 각 50, 500회 반복 학습을 통해 [100개_50회], [100개_500회], [1,000개_50회], [1,000개_500회] 총 4개의 모델을 생성한다.

2. Generate Sentences

본 연구에서 ①~④ 모델을 생성하였다. 각 모델은 이미지에서 추출한[사람] 키워드로 문장을 생성한다.

Table 1. Generated Sentence

models	result
①	(a) “사람 카페 인테리어 액자 감성까지 너무 귀여웠어요. 사장님이랑 대화 하면서 커피 한 잔 하기 딱 좋아 보이는 곳이에요~“
②	(b) “사람인데, 생각보다 너무 귀엽고 좋아서 리뷰 쓰기로 함!! 더빌리지샵 카페로 가기 위해서는 이 계단을 따라 2층으로 올라가면 된다.”
③	(c) “사람 많아서 커피 나오는데 정말저어어는 기 있는 원두라고 하더라고요 그걸 사용해서 저는 얼그레이 티라미수를 먹었습니다.”
④	(d) “사람 많아서 커피 나오는데 정말저어어말 오래걸리는 점을잊지말기! 그리고 꼭 먹어봐야 할 BEST 추천! : 크루아상/소금빵/”

3. Evaluate Sentences(BLEU)

본 연구에서 첫 번째 문장 평가 방법은 BLEU이다. 활동 2에서 생성된 문장을 동일한 키워드 3개로 비교 문장(Reference)을 만들어 각각 비교하였다.

Table 2. BLEU Reference value and score

Reference value	Score
(1) “사람이 진짜 많았어요.디피되어 있는 곳으로 가는 길에는 벽면에 책장이 있고, 책장 안...”	0.0
(2) “사람이 너무 많아 정신 없었어요.보통 4시 이후부터 좀 널널한 편인 듯 해요.주말에 ...“	0.0
(3) “사람들의 평에 의하면 원두가 맛있고종류가 다양해서 매력이 있는것 같았다.실제로 방 ...“	0.0

4. Evaluate Sentences(Questionnaire Surveys)

본 연구에서 두 번째 문장 평가 방법인 설문 조사는 47명의 사람에게 진행되었다. 조사 방식은 [키워드:키워드]로 생성된 문장을 보여주고 문장에 대해서 3개의 평가 항목을 제시하여 진행하였다.

Table 3. Questionnaire Surveys result

항목 \ 모델	①	②	③	④
즉시 사용 가능	6	2	3	24
일부 수정 필요	23	22	11	16
사용 불가	18	23	33	7

IV. Conclusions

본 연구를 통해 데이터 용량 및 반복 학습 횟수에 따른 생성 문장을 분석하였다. 활동 3, 4의 결과를 통해서 BLEU를 활용한 평가는 부적합하다는 결론을 얻었다. 활동 4에서 모델 ①,②,③의 결과물 차이는 크게 없었다. 모델 ④의 결과에서 10배 차이의 데이터양과 학습 횟수가 유의미한 결과를 생성한다는 것을 파악할 수 있었다. 하지만 학습 데이터와 무관한 키워드로 문장생성 시 모델 ④ 역시 무의미한 문장을 생성하였다. 이미지에서 학습 데이터와 관련 있는 키워드를 추출할 수 있다면 추후 연구를 통해 더욱 자연스러운 문장생성이 가능할 것으로 기대된다.

ACKNOWLEDGEMENT

본 연구는 2023년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음 (2019-0-01817)

REFERENCES

[1] Gyu-Hyeon Park, and Hee-Yun Kwon. "Shopping Mall Review Generator using KoGPT2." KSCI, Vol. 30, No. 1, pp.31-32, 2002.

[2] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu, "BLEU: a Method for Automatic Evaluation of Machine Translation" Computational Linguistics (ACL), pp.311-318, 2002.

[3] Sangun Park. "A Survey on Deep Learning-based Pre-Trained Language Models." The Korean Journal of BigData, Vol. 7 No. 2, 11-29. 2022.