제어 가능한 한국어 자연어 생성 모델

김은총 정민수 전현규 정윤경
성균관 대학교
prokkec@naver.com pisces03@g.skku.edu hkjeon13@gmail.com aimecca@skku.edu

Korean Controllable Generation Model

Kim Eunchong Jung Minsu Jeon Heonku Cheong Youngyung
Sungkyunkwan University

요 약

본 논문에서는 한국어 controllable generation model을 연구하고자 한다. controllable generation model이란 텍스트 생성을 제어할 수 있는 조건부 자연어 생성 모델이다. 이 분야는 이미 많은 연구가 이루어졌으며, 본 연구에서는 PPLM[1] 논문을 바탕으로 한국어 controllable generation model을 구현하였다. 우선 3가지의 영역, 스포츠, 정치, 과학 분야에 관련된 위키백과 데이터를 활용하여 키워드를 선별하였다. 그 후, 키워드를 활용하여 의도한 카테고리의 방향으로 문장을 출력하는 모델을 완성했다. 검증 단계에서는 LSA 알고리즘을 통한 주제 분석을 진행하여 원하는 방향으로 출력되었는지를 확인하였다.

1. 서 론

텍스트 생성은 주어진 입력 데이터를 기반으로 특정 의도에 맞게 텍스트를 생성하는 것을 목표로 하는 과제를 말한다. 근래에는 attention에 기반을 둔 transformer 계열의 모델[2]이 많이 활용하여 텍스트 생성 과제를 해결하려는 연구가 증가하고 있다.

단순히 텍스트를 생성하는 단계에서 발전하여 원하는 주제, 혹은 스토리의 흐름을 직접 제어하려는 시도가 늘고 있다. 어떤 하나의 토큰 x가 나올 확률 $p(x)$을 학습하는 것이 아닌, 의도된 특성 a와 관련된 텍스트를 출력하고자 하는 $p(x|a)$을 학습하는 것이 controllable generation model이 지향하는 바이다. 다시 말하자면, 이 모델은 자연스러운 문장의 출력을 넘어 사람이 직접 의도를 가지고 출력을 제어할 수 있게 하는 것을 목표로 한다.

최근에는 한국어 모델을 만들려는 시도가 늘고 있다. KoBERT[1], KoGPT2[2]에 이어 최근에는 40GB 이상의 한국어 텍스트를 학습한 KoBART[3] 모델까지 공개된 바가 있다. 본 연구에서는 KoGPT2에 기반한 한국어 controllable generation model을 구현하였다. 이는 우리가 조사한 바에 의하면 시도된 적 없으며, Plug and Play Language Model(PPLM)[1]의 구조를 참고하여 구현했다. 필요한 데이터를 마련하기 위해 위키백과를 활용하여 키워드를 도출하였다. 이 키워드를 바탕으로 controllable generation model을 구현한 후, 이 모델의 성능 검증단계에서 LSA 알고리즘[3]으로 원하는 주제가 추출되는지를 분석하였다.

1) https://github.com/SKTBrain/KoBERT
2) https://github.com/SKT-AI/KoGPT2
3) https://github.com/SKT-AI/KoBART

2. 관련 연구

기존에 controllable generation model을 만들기 위한 많은 노력이 있었다.

먼저 LSTM을 활용한 모델들이 등장했다. [4] 논문에서는 storytelling 분야에서 BiLSTM을 사용하여 4개의 문장으로 구성된 데이터에 결말을 창작하려는 시도를 하였다. 이때 RAKE algorithm을 활용해 happy ending과 sad ending, 두 가지 버전의 결말을 제어할 수 있는 모델을 구현했다. 마찬가지로 Bi-LSTM을 활용하여 Analyzer에서 추출되는 제어 요소를 Decoder에 첨가해 감정의 깊이에 따른 문장을 출력하도록 학습하는 시도 역시 존재했다 [5]. 또한 앞선 방식뿐만 아니라, 문장의 길이, 동사, Frame Semantics 등 다양한 요소들을 통해 출력을 제어하기도 하였다[6].

LSTM보다 파라미터가 상대적으로 많은 transformer 기반 모델에서도 이러한 시도는 계속되었다. [7] 논문에서는 약 16억 개의 거대한 파라미터에 추가적으로 Control code c를 활용하여 같은 단어로 시작하더라도 c에 따라 출력문이 달라지는 모델을 구현하였다. 본 논문에서 구조를 참고한 PPLM[1]에서는 특정 카테고리에 해당하는 단어들을 활용하여 출력문의 주제를 제어하였다. 이 논문은 무엇보다도 더 이상의 학습을 진행하지 않고도 원하는 주제에 따른 결과물을 출력할 수 있다는 강점이 있다. 따라서 과도한 컴퓨팅 파워를 요구하지 않기에, 우리는 이 논문에서 제시한 모델의 구조를 활용하였다.

* 이 논문은 2019년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원(No. 2019R1A2C1006316)과 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2019-0-00421, 인공지능대학원지원).

3. 방법

3.1 데이터

본 연구에서는 한국어 위키피디아 데이터를 활용하였는데, 현재 기준으로 가장 최근에 올라와 있는 2021년 4월20일 데이터[5]를 사용하였다. 이 중 서로 중복될 일이 적은 카테고리 정치(4,648개), 과학(5,242), 스포츠(6,490개)에 해당하는 문서를 활용하였다.

3.2 모델

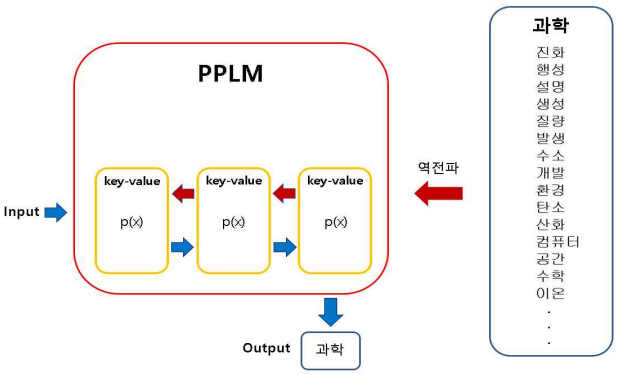

그림 1 PPLM 작동 방식

PPLM[1]은 Uber에서 개발한 controllable generation model이다. PPLM은 학습된 언어 모델에 원하는 키워드에 관련된 단어들을 입력받아 출력을 제어한다. 이 모델의 가장 큰 강점은 모델을 더 이상 학습시키지 않고도 원하는 주제로 텍스트를 생성한다는 것이다. 이에 대한 방법은 다음과 같다. 우선 입력으로 시작 문구를 제시한다. 이때 모델 내부에서는 하나의 카테고리 속 단어들을 바탕으로 역전파가 진행된다. 이 역전파는 모델의 가중치를 업데이트하지 않으며 각 attention layer의 내부 key-value 값에만 변화를 준다. 마지막에 이를 순전파함으로써 원하는 주제에 속하는 토큰을 출력하게 된다. 이때 결과물은 단순히 제시하는 카테고리 속 단어들만을 출력하는 것이 아니라, 의도한 주제로의 방향성을 띠게 된다.

[1]에서는 GPT-2[8] 모델을 활용하여 PPLM을 구현하였기에, 본 연구에서도 KoGPT-2[6] 모델을 사용하였다. 이 모델은 12-layer, 768-hidden, 12-heads, 117M 파라미터를 가진 gpt2 base 모델과 동일하며 약 20GB의 2천5백만 이상의 문장으로 학습되었다.

3.3 실험 방법

KR-WordRank[7]에서 제공하는 한국어 키워드 추출기를 활용하여 정치, 과학, 스포츠에 해당하는 위키백과 문서들을 모아 각각 키워드 분석을 진행한다. 그리고 카테고리별로 명사 40개를 추출한 후, 이를 사용하여 PPLM 모델에서 문장을 출력시킨다. 이때 20종류의 시작 문구를 입력하여 각각 20개, 총 400개의 문장이 출력된다. 총 3개의 카테고리에서 이 과정이 반복되며, 출력되는 결과물로 평가를 진행한다. 평가에는 부가적인 데이터 없이 문서 내부의 토픽을 효율적으로 도출하는 LSA(잠재 의미 분석) 알고리즘[3]을 활용한다. 각 카테고리별로 5개의 토픽을 도출하도록 설정한 후 이를 바탕으로 의도한 카테고리에 속하는 문장이 출력되었는지를 평가한다.

4. 실험

4.1 결과

시작 문구	제어적용 여부	Output
내가 생각하는	X	내가 생각하는 그 분은 그 분이 아니다"라고 말했다. 이 같은 현상은 지난해 9월 한 차례 나타난 바 있다. 이후 2년 가까이 이어져 온 것이다. 이에 대해 한 업계 관계자는 "경기회복이 지연되고 있는 상황에서 부동산시장이 실수요
	O (스포츠)	내가 생각하는 최고의 선수"라며 "이번 시즌에도 정규리그 1경기 출장정지를 당했지만, 챔피언결정전은 경험 많은 우리 팀의 챔피언이 될 것으로 믿는다. 특히 그는 지난 시즌 K리그 득점왕(2골)과 득점왕을 동시에 차지하는 등
	O (정치)	내가 생각하는 건 이런 거였으면 하는 거다"라며 "내가 어떤 일을 하고, 어떤 정책을 펴나, 이 정책이 옳다고 하는 사람이 과연 국가발전을 이끌어 갈 수 있는지 국민이 판단하실 몫이라고 생각해서 정치개혁을 해야 하는데 왜 이렇게 자꾸 정치 개혁 논의를
	O (과학)	내가 생각하는 '세상을 바꾸다'는 어떤 것을 의미하는 '창조성'은 창조적인 물질과 그 결합을 통해 만들어진다. 이 때문에, 그 당시, 이 모든 것이 하나의 거대한 물결에 의해 형성되었다. 그 결과, 그것은
결론을 말하자면,	X	결론을 말하자면, 이 책의 저자가 '인간의 본질에 대한 성찰'을 하고 있다고 본다. 그러나 이는 어디까지나 이 책의 마지막 장에서 언급될 것이다. 이 책에서 '인간이 지닌 본질'에 관해 서술하는 것은 '인간에게
	O (스포츠)	결론을 말하자면, "한국에서 1승이 더 중요하다"는 게 그의 생각이다. 또 "그 동안 우리가 많은 것을 해왔고 앞으로 더욱 잘 될 것"이라면서 "그러나 우리 선수들이 그라운드를 떠난 이후에도 좋은 경기를 할 수 있는 팀,
	O (정치)	결론을 말하자면, 우리는 그 문제를 어떻게 하면 정치에서 정당과 의회로 분리하는 것이 바람직한지에 대해 논의해야 한다. 또한, 이 모든 것을 가능하게 하기 위해서, 어떤 형태의 정치체제 또는 어떤 형태의 연방을 만드는 것은 반드시 필요한 것이라고 생각했다.
	O (과학)	결론을 말하자면, '그들은' 어떤 특정한 생물학적 작용에 대한 이론이다. 그러나 그 결과 인간의 두뇌가 더 복잡한 신경계를 통해 정보를 저장함으로써 인지기능을 작동시킬 수 있는 영역이 점차 확장되고, 이에 따라 인지기능과 관련된 뇌의 구조와 기능의

표 1 각 카테고리 별 출력 결과물 예시

총 20종류의 시작 문구와 3가지의 카테고리를 활용하여 1,200개의 문장을 출력하였다. 결과에 대한 일부 예시는 표 1과 같다.

5) https://dumps.wikimedia.org/kowiki/
6) https://github.com/SKT-AI/KoGPT2
7) https://github.com/lovit/KR-WordRank

4.2 평가

스포츠	
Topic1	('좋은', 0.19803), ('리그', 0.182), ('시즌', 0.16902), ('많은', 0.16182), ('원하는', 0.15915)
Topic2	('위안화', 0.50678), ('중국', 0.44519), ('가치', 0.23837), ('들어', 0.21091), ('정부의', 0.18799)
Topic3	('리그', 0.35839), ('시즌', 0.18391), ('경기', 0.15842), ('위안화', 0.12616), ('우승을', 0.11577)
Topic4	('같은', 0.24687), ('서울', 0.2314), ('평소에도', 0.2039), ('말해왔다', 0.2039), ('오늘의', 0.14949)
Topic5	('리그', 0.31228), ('않는다', 0.18842), ('생각하지', 0.18671), ('이쯤', 0.15707), ('되면', 0.15707)
정치	
Topic1	('박근혜', 0.19239), ('정치', 0.16831), ('일을', 0.1609), ('하는', 0.15676), ('같은', 0.15034)
Topic2	('위안화', 0.44702), ('중국', 0.30772), ('들어', 0.26264), ('절하와', 0.25118), ('관련해', 0.23983)
Topic3	('일을', 0.33978), ('하는', 0.22367), ('생각하지', 0.2031), ('않는다', 0.2031), ('생각을', 0.1767)
Topic4	('후보', 0.32082), ('적지', 0.3027), ('이러하다', 0.21722), ('반론은', 0.16393), ('중에는', 0.14942)
Topic5	('서울', 0.29338), ('일정은', 0.20931), ('오늘의', 0.20931), ('제회', 0.19406), ('오후', 0.18273)
과학	
Topic1	('같은', 0.21084), ('하는', 0.18508), ('새로운', 0.13922), ('않는다', 0.13732), ('많은', 0.13322)
Topic2	('화학', 0.25113), ('분자', 0.21734), ('질량', 0.1434), ('새로운', 0.13864), ('정리하자면', 0.13327)
Topic3	('않는다', 0.2756), ('생각하지', 0.27198), ('그래', 0.13971), ('화학', 0.12906), ('분자', 0.12553)
Topic4	('화학', 0.2946), ('분자', 0.26259), ('생각하지', 0.20156), ('않는다', 0.19727), ('질량', 0.18094)
Topic5	('일정은', 0.32015), ('오늘의', 0.32015), ('않는다', 0.24327), ('생각하지', 0.2385), ('서울', 0.12239)

표 2 LSA 토픽 추출

표 2, 3, 4를 통해 카테고리별로 추출된 토픽들을 볼 수 있다. 모든 토픽이 카테고리와 강한 연관성을 띄지는 않지만, 일부 토픽에서는 직접 연관 있는 단어들이 추출되었다. 스포츠 카테고리에서는 '리그', '시즌', '경기', '우승'과 같은 스포츠와 밀접한 단어들이 토픽으로 도출되었다. 정치 카테고리에서는 '박근혜', '정치', '중국', '후보', '반론은'과 같은 관련성 높은 단어들이 도출되었다. 마지막으로 과학 카테고리에서는 '화학', '분자', '질량'같은 과학 용어들이 토픽으로 도출되었음을 알 수 있다.

5. 결론 및 향후 연구

본 논문에서는 KoGPT2과 PPLM을 활용한 한국어 controllable generation model에 대해 살펴보았다. 또한 이를 통해 생성되는 문장들을 평가하였다.

모든 결과물에서 의미 있는 결과를 내보이지는 않았지만, LSA를 통한 토픽 분석 결과를 통해 원하는 주제에 맞게 제어된 문장들의 존재를 간접적으로 파악할 수 있었다.

향후 연구로는 큰 규모의 모델을 활용하여 더욱 정교한 controllable generation model을 구현할 수 있기를 기대한다. 더 나아가 단순히 하나의 주제로 제어된 텍스트 생성을 넘어 구체적이고 다양한 조건을 제어할 수 있게 되기를 기대한다.

참고문헌

[1] Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. Plug and play language models: a simple approach to controlled text generation. arXiv preprint arXiv:1912.02164. 2019.

[2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob, Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Advances in Neural Information Processing Systems, pages 6000-6010, 2017.

[3] Peter W Foltz. Latent semantic analysis for text-based research. Behavior Research Methods, Instruments, & Computers, 28(2):197-202, 1996.

[4] Nanyun Peng, Marjan Ghazvininejad, Jonathan May, and Kevin Knight. Towards Controllable Story Generation, Association for Computational Linguistics, 2018.

[5] Fuli Luo, Damai Dai, Pengcheng Yang, Tianyu Liu, Baobao Chang, Zhifang Sui and Xu Sun. Learning to Control the Fine-grained Sentiment for Story Ending Generation, Learning to Control the Fine-grained Sentiment for Story Ending Generation. 2019.

[6] Lifu Tu, Xiaoan Ding, Dong Yu and Kevin Gimpel. Generating Diverse Story Continuations with Controllable Semantics, Workshop on Neural Generation and Translation, 2019.

[7] Nitish Shirish Keskar, Bryan McCann, Lav R. Varshney, Caiming Xiong, and Richard Socher. Ctrl: A conditional transformer language model for controllable generation, arXiv preprint arXiv:1909.05858, 2019.

[8] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. OpenAI Blog, 1(8), 2019.