

Sequence-to-Sequence 모델을 이용한 신문기사의 감성 댓글 자동 생성*

박천용⁰, 박요한, 정혜지, 김지원, 최용석, 이공주
충남대학교

sdpcy0520@gmail.com, happy005012@naver.com, hyeji6138@naver.com,
top9076@cnu.ac.kr, yseokchoi@cnu.ac.kr, kjoolee@cnu.ac.kr

Automatic Generation of Emotional Comments on News-Articles using Sequence-to-Sequence Model

Chun-Young Park⁰, Yo-Han Park, Hye-Ji Jeong, Ji-Won Kim, Yong-Seok Choi, Kong-Joo Lee
Chungnam National University

요 약

본 논문은 신문기사의 감성 댓글을 생성하기 위한 시스템을 제시한다. 감성을 고려한 댓글 생성을 위해 기존의 Sequence-to-Sequence 모델을 사용하여 긍정, 부정, 비속어 포함, 비속어 미포함 유형의 4개의 감성 모델을 구축한다. 하나의 신문 기사에는 다양한 댓글이 달려있지만 감성 사전과 비속어 사전을 활용하여 하나의 댓글만 선별하여 사용한다. 분류한 댓글을 통해 4개의 모델을 학습하고 감성 유형에 맞는 댓글을 생성한다.

주제어: 감성, 댓글, Sequence-to-sequence

1. 서론

딥 러닝 기법이 발달하면서 자연어 처리 분야 또한 눈에 띄게 많은 발전을 이루었다. 그 중, Sequence-to-Sequence(seq2seq)모델은 기계번역 분야에서 뛰어난 성능을 보이고 있다[1]. 그밖에도 QA 시스템에서 질문에 대한 응답, 문서에 대한 질문, 신문기사의 제목, 프로그램 코드에 대한 주석을 생성하는 데에도 다양하게 활용된다[2-6]. 영어권에서는 seq2seq 모델을 이용하여 감성에 따른 응답을 생성하는 챗봇[7]에 관한 연구가 있었으나 국내에서는 아직 감성을 고려한 생성에 관한 연구가 많지 않았다. 본 논문에서는 seq2seq 모델을 이용하여 감성이 표현된 문장을 생성해 보고자 한다.

모델의 학습 데이터로는 인터넷 기사와 댓글을 수집하여 사용한다. 왜냐하면 댓글은 “좋다”, “싫다”, “잘했다”, “못했다” 등과 같이 글쓴이의 감성을 포함하고 있는 경우가 많기 때문이다. 또한 기사와 댓글의 양이 많고 수집도 용이하기 때문에 대량의 데이터 수집도 가능하다. 자동으로 수집한 댓글을 감성사전을 이용하여 긍정이나 부정의 의미를 가진 댓글로 분류한다. 또한 비속어 사전을 이용하여 비속어 포함 댓글과 그렇지 않은 댓글로도 분류하여 모델의 학습 데이터로 사용해 본다.

본 연구에서는 긍정, 부정, 비속어 포함, 비속어 미포함,

4개의 모델을 각각 구축하여 감성이 표현되어 있는 댓글을 자동 생성해 보고자 한다.

2. 댓글 자동 생성

2.1 전체 시스템

감성을 고려한 댓글을 자동 생성하기 위해 그림 1과 같이 네 가지 감성 유형의 모델을 사용한다. 각 모델의 입력은 기사 제목과 첫 문장이며, 출력은 입력을 토대로 감성 유형에 맞게 생성된 댓글이다.

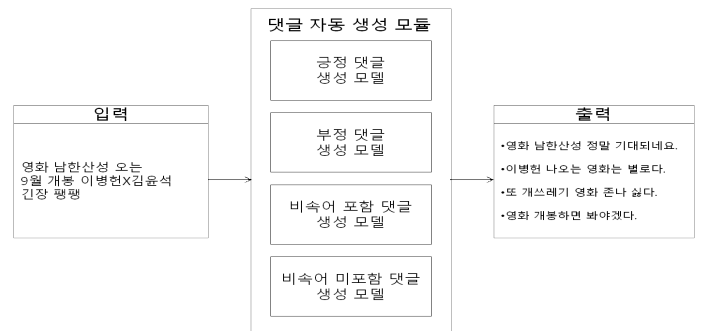


그림 1. 감성을 고려한 댓글 생성 모델

* 이 논문은 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No.NRF-2015R1C1A2A01051685)

2.2 Seq2seq 모델

본 연구에서는 댓글 생성을 위해 seq2seq 모델을 사용한다. 그림 2는 입력단에 “특이한 고양이 있다”가 들어갔을 때 “참 좋은 소식이네요”라는 댓글을 생성하는 예시이다. 인코더의 입력은 형태소의 임베딩 벡터를 사용한다. 그리고 인코더의 출력은 입력 문장의 의미가 포함된 벡터이다. 디코더는 인코더가 출력한 벡터와 Bahdanau[8]의 주의집중(attention) 알고리즘을 바탕으로 계산한 가중치를 입력받는다. 디코더의 출력은 Softmax 함수를 사용하여 가장 높은 확률의 단어를 선택한다.

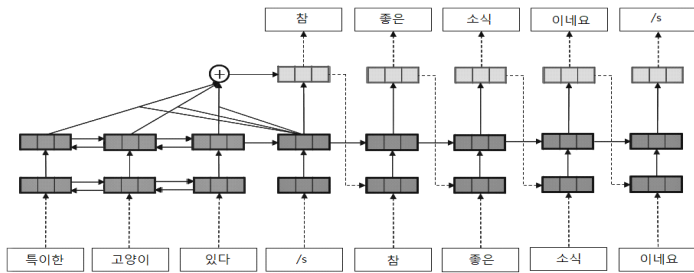


그림 2. Sequence-to-sequence Model

2.3 데이터 수집 및 처리

seq2seq 모델을 이용하기 위해서는 대량의 데이터가 필요하다. 따라서 다음(Daum) 신문기사²⁾를 수집하였다. 신문기사는 독자의 이해를 돕기 위해 대체적으로 두괄식 구조를 채택하고 있다[9]. 그러므로 본 연구에서는 신문기사의 전문이 아닌 제목과 첫 문장만을 모델의 입력으로 사용한다. 표 1은 수집한 데이터의 개수와 평균 어절 수이다.

표 1. 분야별 신문기사의 수

분야	기사	댓글	단위 : 개	
			기사 ¹⁾ 평균 어절	댓글 평균 어절
it	6,079	137,544	28.78	13.16
문화	8,129	282,506	24.72	13.32
연예	43,928	675,755	23.64	9.78
스포츠	13,326	2,192,384	22.35	13.47
경제	6,848	273,429	30.03	14.80
사회	12,126	756,754	30.81	12.62
기타	49,517	12,233,498	31.41	14.47
총합	139,953	16,551,870	27.39	13.08

본 논문은 감성사전[10]과 비속어사전[10]을 이용하여 댓글을 긍정, 부정, 비속어 포함, 비속어 미포함 댓글로 분

류하였다. 감성사전에서 1,647개의 긍정 표현과 4,573개의 부정 표현을 사용하였고 비속어 사전에서 409개의 비속어 단어를 사용하였다. 표 2는 긍정, 부정, 비속어 단어들의 예시이다.

표 2. 유형별 단어 예시

유형	예시
긍정	가슴 뚫리는 기분, 딱 좋다, 감동 받다, 걱정 말아요, 가치 있다, 너무 보기 좋다, 매우 좋다, 보기 좋다
부정	거부감 들다, 골치 아프다, 너무 후회, 답답해 보이다, 미심쩍다, 분노하다, 마음에 안 들다, 눈꼴사납다
비속어	시발, 병신, 놈, 전나, 쪽발이, 개새끼, 지랄, 쥐뿔

하나의 신문기사에는 평균 118개 이상의 댓글이 달려 있다. 본 연구에서는 각 유형에 맞게 선별된 댓글과 그 댓글이 달린 신문기사를 학습데이터로 사용한다.

댓글을 선별할 때에는 감성사전과 비속어 사전에서 추출한 단어가 댓글에 포함되어 있는지 확인한다. 그 중 댓글의 어절 수가 14 어절 이하이고, 기사의 제목과 첫 줄에서 사용된 단어가 가장 많이 포함된 댓글을 최종적으로 선택한다. 그 후 문장의 모든 특수기호를 제거하고, 숫자는 "<NUM>"로, 공백은 "<SP>"로 정규화 한다. twitter-korean-text[11]를 이용해 형태소 분석을 하고 입력단의 조사만 제거한다. 표 3은 이와 같이 추출한 유형별 댓글 수이다. 각 유형별로 선별된 댓글 중 100개는 평가 데이터이고 그 외의 댓글은 모두 학습데이터이다.

표 3. 유형별 선별된 댓글 수

유형	단위 : 개	
	선별된 댓글 수	
긍정	51,472	
부정	64,240	
비속어 포함	70,043	
비속어 미포함	101,383	

3. 실험 환경 및 결과

3.1 실험환경

모델을 학습할 때 사용한 파라미터는 표 4와 같다. 네 가지의 모델은 동일한 파라미터를 사용하였다. seq2seq 모델은 OpenNMT³⁾를 활용하여 학습 시킨다.

1) 기사의 평균어절 수는 제목+첫 문장의 평균 어절 수이다.

2) <http://media.daum.net>

3) <http://opennmt.net/>

표 6. 평가항목-2 기준

유형	입력 (제목+신문기사)	출력 (댓글)
입력 문장의 키워드가 출력 문장에 포함된 경우	고개 숙인 종근당 회장 운전기사 폭언 ...	종근당 불매운동 해야 한다
입력 문장과 관련된 어휘가 출력된 경우	타격의 팀 기아 타이거즈와 SK 와이번스 가 역대급 명승부를 펼쳤다	김성근 선수 수고 많았습니다 내년엔 좋은 결과 있길 바랍니다
문맥상 의미가 적합한 경우	300mm 폭우 피해 눈덩이 사망 4명 실종 2명 이재민 517명 ...	삼가 고인 의 명복 을 빕니다 잊지 않겠습니다
두 문장의 관련성이 없는 경우	호텔에 등 달고 변기 뜯고 천상천하 유아독존	서세원이 더 나쁘다는 말이 아니라 이게 무슨 의미가 있다는 것 자체가 한심하다

표 4. 모델에 사용한 파라미터

Parameter	Value
Word embedding size	256
RNN size	512
Encoder/Decoder layer	2
Encoder/Decoder type	GRU
Dropout	0.3
Optimizer	Adam
Learning rate	0.001
Epoch	200

3.2 실험 평가 기준

seq2seq 모델을 이용한 대부분의 생성 결과는 정답과 동일한 표현이 얼마나 발생했는지를 평가의 기준으로 사용한다[1,4,12]. 그러나 본 연구의 경우, 정답으로 간주할 수 있는 댓글이 각 신문기사 당 118개 이상이기 때문에 정답 댓글과의 정량적인 비교가 어렵다. 그러므로 정량적인 평가보다는 정성적인 평가를 수행해 보고자 한다.

각 모델의 대한 평가는 다음의 3가지 기준을 사용한다. 평가 항목-1은 출력된 문장의 완성도와 의미를 평가한다. 댓글의 경우 표 5와 같이 여러 유형의 문장이 존재하기 때문에 이를 고려하여 평가한다. 문장이 온전하며 의미전달이 명확할 때 최고점을 부여한다. 문장이 이루어지지 않거나 명사로 끝나는 경우라도 의미가 전달된다면 높은 점수를 부여한다. 의미가 명확히 전달되지 않으면 낮은 점수를 부여한다.

표 5. 인터넷 기사의 댓글 유형

유형	예시
문장이 온전한 경우	쉬고 싶은 숲 너무 보기 좋습니다.
문장이 이루어지지 않는 경우	목숨 걸고 갈 만큼 가치 있는지
문장이 명사로 끝나는 경우	읽어보고 왜 인기 있는지 이해가 안 되었던 책

평가항목-2는 입력 문장과 출력 문장 사이의 의미 관련성을 평가한다. 입력 문장의 주요 키워드가 출력 문장에

포함되어 있으면 높은 점수로 평가한다. 그 외에 관련 어휘가 포함되었을 경우와 문맥상의 의미가 적합할 경우에는 그보다 낮은 점수로 평가한다. 입력문장과 출력문장이 연관성이 없을 경우 낮은 점수로 평가한다. 평가예시는 표 6와 같다.

평가항목-3은 출력된 문장이 감성유형에 적합한지를 평가한다. 긍정 댓글 생성기의 경우 출력된 문장이 긍정 표현에 가까운 지를 평가한다. 부정 댓글 생성기의 경우 출력 문장이 부정표현에 가까운 지를 평가한다. 비속어 포함 모델 및 비속어 미포함 모델의 경우는 평가항목-3은 평가하지 않았다.

평가의 공정성을 확보하기 위해 한 모델 당 평가자 2명씩 총 8명이 평가를 실시하였다. 평가할 댓글은 100개의 신문기사에 대해서 모델이 생성한 댓글과 실제 기사의 댓글을 섞어 총 200개의 댓글을 함께 평가한다. 각 항목의 점수는 1-5점 범위이다. 평가자는 어떤 결과가 실제 사람이 작성한 댓글인지 시스템이 자동 생성한 댓글인지 모른 채 평가를 수행하도록 하였다.

3.3 실험 결과 및 분석

표 7은 실제 인터넷 기사 댓글, 표 8은 댓글 자동 생성 모듈이 생성한 댓글에 대한 각 평가항목별 평균점수이다. 예상했던 것과 같이 실제 인터넷 기사 댓글이 거의 모든 항목에서 좋은 점수를 받았다.

감성적합성의 부문에서는 모델이 생성한 댓글의 점수가 더 높았다. 부정 유형의 경우, 모델이 생성한 댓글이 “짜증난다.”, “개편이다.” 등 의미가 명확한 단어가 인터넷 기사의 댓글에 비해 많이 나타났기 때문인 것으로 판단된다. 긍정 유형의 경우 실제 인터넷 기사의 댓글에서

표 9. 각 모델의 좋은 예와 나쁜 예

	유형	입력 (제목+신문기사)	출력 (댓글)
좋은 예	(1) 긍정	TV 예능 프로그램 동상이몽 통해 기존 샤프 이미지 벗고 부드러운 면서도 속정 깊은 아저씨로 새롭게 인식되고 있는 이계명 성남시장 침체 된 한국 관광 활성화 나섰다	우와 좋은 모습 보기 좋아요
	(2) 부정	이철성 음주 사고 은폐 민정 수석실 알고도 넘어갔다 청와대 민정 수석실 이철성 경찰청장 후보자 과거 음주운전 사고 당시 신문 속여 징계 받지 않았다는 사실 알고도 이를 문제 삼지 않은 것으로 드러났다	진짜 짜증난다 진짜 나라꼴이 개판이다 이 나라의 미래는 암울하다
	(3) 비속어 포함	박주영 리더 대신 그림자를 택하다 홍명보호의 만행 격인 박주영이 리더 대신 조력자 역할을 선택했다	홍명보 같은 놈들은 뭐나 양심도 없는 놈들이네
	(4) 비속어 미포함	귀국한 어보 박수치는 문 대통령 내외 배제만 기자 미국 방문을 마친 문재인 대통령이 2일 오후 서울공항에 도착해 함께 전용기편으로 들어온 조선훈종어보 문정왕후 어보를 보며 박수치고 있다	문재인 대통령님 감사드립니다
나쁜 예	(5) 긍정	옥에 티 얼룩진 흰 운동화 하얗게 만들려면 김대리가 생활 속 꿀팁을 전합니다	저런 사람들이 살아있는 사람들은 알고 있는거 아닌가 뭘 의미가 있나
	(6) 부정	관광버스 참사 속에서 가장 먼저 탈출한 운전기사 지난 13일 경부고속도로 언양 분기점 인근을 달리던 전세버스에 화재 사고가 발생한 직후 운전기사 이 모씨가 가장 먼저 버스에서 탈출했다는 경찰 조사 결과가 나왔다	제목 좀 내려라 너무 비싸다 좀 내려라
	(7) 비속어 포함	MLB 류현진 다음등판 장담 못해 발똥 생각보다 나빠 호주 개막시리즈에서 기본 좋은 첫 승을 신고했던 류현진이 다음 이어지는 본토 개막시리즈 등판을 장담할 수 없게 됐다	기자놈아 전범기가 아니라 전범기가 아니라 전범기가 아니라 전범기가 아니라 쪽마리 색히들아
	(8) 비속어 미포함	김상곤 송영무 엄호 조대엽은 땀띠름 낙마사유는 없어 강병철 한지훈 기자 더불어민주당은 2일 김상곤 송영무 후보자를 엄호하면서 조대엽 후보자에 대해서는 여론을 주시했다	노회찬 의원님 존경합니다 고인의 명복을 빕니다

표 7. 실제 인터넷 기사 댓글

	평가항목-1 문장완성도	평가항목-2 의미관련성	평가항목-3 감성적합성
긍정	3.78	3.945	3.02
부정	4.82	3.88	4.13
비속어 포함	4.64	4.205	-
비속어 미포함	4.58	4.25	-

표 8. 감성 모델이 생성한 댓글

	평가항목-1 문장완성도	평가항목-2 의미관련성	평가항목-3 감성적합성
긍정	3.37	2.205	3.585
부정	3.615	3.1	4.265
비속어 포함	3.525	3.08	-
비속어 미포함	3.3	3.85	-

긍정적인 단어가 나타나더라도 반어적인 의미로 사용된 경우가 있기 때문이다.

표 9는 각 모델이 생성한 문장 중 좋은 예와 나쁜 예이다. (5)와 (6)의 경우 입력과 관련이 없는 댓글이 생성되었다. (7)의 경우 seq2seq의 반복 생성 문제를 살펴볼 수 있었다. (8)의 예에서는 같은 분야지만 입력에서 거론되지 않은 이름이 댓글에서 언급된 것을 살펴볼 수 있다.

각 모델의 출력에서 사용된 단어와 실제 댓글에서 사용된 고유 단어 개수와 평균 어절 수는 표 10과 같다. 출력 결과의 고유한 단어 개수가 실제 댓글의 단어 개수보

표 10. 댓글에 쓰인 고유 단어 개수와 평균 어절 수

	실제 댓글		모델이 생성한 댓글	
	단어 수	평균 어절	단어 수	평균 어절
긍정	941	8.49	287	6.67
부정	1,014	9.08	321	8.57
비속어	1,041	9.05	285	7.93
비속어 미포함	1,007	8.9	375	9.76

다 적다. 본 연구에서는 다양한 표현을 생성할 수 있기를 기대하였으나 실제 출력에서는 단어의 사용이 한정적인 것을 확인하였다.

4. 결론

본 논문에서는 감성을 고려한 댓글을 자동으로 생성하였다. 이를 위해 신문 기사를 수집하고 감성 사전과 비속어 사전을 토대로 댓글을 분류하였다. seq2seq 모델을 이용하여 긍정, 부정, 비속어 포함, 비속어 미포함 모델을 구성하였다. 실험의 결과로 기사 내용에 따라 감성이 잘 나타나는 문장 생성이 가능하다. 그러나 모델이 자동 생성한 댓글은 실제 인터넷 기사의 댓글에 비해 맞춤법과 띄어쓰기가 더 정확하였다. 실제 인터넷 기사의 댓글처럼 생성하기 위해서는 추가적인 연구가 필요하다. 또한 모델이 자동 생성한 댓글은 신문기사 내용과 의미관련성이 적은 경우가 많았고, 학습한 단어 수에 비해 다양한 표현들이 댓글에 많이 나타나지 않았다. 향후 기사 분야를 좁혀 학습시킬 경우 의미관련성이 향상될 것으로 기

대된다.

참고문헌

- [1] SUTSKEVER, Ilya; VINYALS, Oriol; LE, Quoc V. Sequence to sequence learning with neural networks. In: *Advances in neural information processing systems*. 2014. p. 3104–3112.
- [2] HE, Shizhu, et al. Generating Natural Answers by Incorporating Copying and Retrieving Mechanisms in Sequence-to-Sequence Learning. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2017. p. 199–208.
- [3] SERBAN, Iulian Vlad, et al. Multiresolution Recurrent Neural Networks: An Application to Dialogue Response Generation. In: *AAAI*. 2017. p. 3288–3294.
- [4] YUAN, Xingdi, et al. Machine Comprehension by Text-to-Text Neural Question Generation. *arXiv preprint arXiv:1705.02012*, 2017.
- [5] Lei Xu, Ziyun Wang, Ayana, Maosong Sun. Topic Sensitive Neural Headline Generation. 2016.
- [6] HU, Xing, et al. CodeSum: Translate Program Language to Natural Language. *arXiv preprint arXiv:1708.01837*, 2017.
- [7] ZHOU, Hao, et al. Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory. *arXiv preprint arXiv:1704.01074*, 2017.
- [8] BAHDANAU, Dzmitry; CHO, Kyunghyun; BENGIO, Yoshua. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.
- [9] 박재영. 뉴스 평가 지수 개발을 위한 신문 1면 머리기사 분석. *한국의 뉴스 미디어*, 2006, 147–220.
- [10] CHO, Young Hwan; LEE, Kong Joo. Automatic affect recognition using natural language processing techniques and manually built affect lexicon. *IEICE transactions on information and systems*, 2006, 89.12: 2964–2971.
- [11] PARK, Eunjeong L.; CHO, Sungzoon. KoNLPy: Korean natural language processing in Python. In: *Proceedings of the 26th Annual Conference on Human & Cognitive Language Technology*. 2014. p. 133–36.
- [12] BAHDANAU, Dzmitry; CHO, Kyunghyun; BENGIO, Yoshua. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.