복사 방법 및 검색 방법을 이용한 종단형 생성 기반 질의응답 채팅 시스템

김시형○, 김학수, 권오욱*, 김영길*
강원대학교 컴퓨터정보통신공학과, 한국전자통신연구원*
{sureear,nlpdrkim}@kangwon.ac.kr, {ohwoog*,kimyk*}@etri.re.kr

End-to-End Generative Question-Answering Chat System Using Copying and Retrieving Mechanisms

Sihyung Kim○, HarkSoo Kim, Oh-Woog Kwon*, Young-Gil Kim*
Kangwon National University Computer and Communication Engineering
Electronics and Telecommunications Research Institute*

요 약

채팅 시스템은 기계와 사람이 서로 의사소통 하는 시스템이다. 의사소통 과정에서 질문을 하고 질문에 대한 답변을 하는 질의응답 형태의 의사소통이 상당히 많다. 그러나 기존 생성 기반 채팅 시스템에서 자주 사용되는 Sequence-to-sequence모델은 질문에 대한 답변보다는 좀 더 일반적인 문장을 생성하는 경우가 대부분이다. 이러한 문제를 해결하기 위해 본 논문에서는 복사 방법과 검색 방법을 이용한 생성 기반 질의응답 채팅 시스템을 제안한다. 템플릿 기반으로 구축한 데이터를 통한 실험에서 제안 시스템은 복사 방법만 이용한 질의응답 시스템 보다 45.6% 높은 정확도를 보였다.

주제어: 복사 방법, 검색 방법, 생성 기반 질의응답 채팅 시스템

1. 서론

채팅 시스템(Chatting System)은 기계가 사람의 말을 적절하게 이해하여 답변을 하는 시스템이다. 사람의 대화에서는 질문을 하고 질문에 대한 답변을 하는 질의응답 형태의 대화가 상당히 많다. 예를 들어 “버락 오바마”에 대한 대화 중, “그런데 버락 오바마가 어디 출신이지?”라는 질문이 등장 할 수 있다. 이에 대한 정답은 “호놀룰루”로, 사람은 “버락 오바마는 호놀룰루 출신이야”와 같이 자연스러운 문장을 만들어 답변한다. 이 과정을 시스템이 수행하기 위해서는 먼저 질문에 해당하는 주어(Subject)와 술어(Predicate)를 찾아 목적어(Object)를 찾아야 한다. 그 이후 찾은 목적어를 정해진 형식의 템플릿(Template)에 맞추어 문장으로 변환하여 답변한다. 그러나 이 방법은 정해진 답변밖에 생성할 수 없는 문제점이 있다. 본 논문에서는 이러한 문제를 해결하기 위하여 질의에 대한 정답만 출력 하는 것이 아니라 답변에 주어의 다른 목적어 정보를 추가하여 보다 자연스러운 문장을 생성하는 종단형(End-to-end) 생성 기반 질의응답 시스템을 제안한다.

2. 관련 연구

두 개의 Recurrent Neural Network(RNN)을 사용하는 Sequence-to-sequence모델[1]은 입력 열을 인코딩 하여 출력 열을 디코딩 하는 모델로써 다양한 자연어 처리 연구에 활용 되고 있다. 번역 분야에서 주의 집중(attention)을 기반으로 한 Encoder-Decoder 모델은 디코더가 인코더에서 좀 더 중요한 정보를 선택할 수 있게 함으로써 디코더의 성능을 향상시켰으며[2], 생성 채팅 모델에서도 주로 사용되고 있다[3]. 복사 방법(Copying mechanism)은 디코더를 변형한 모델로써 기존의 주의 집중 방법에 더하여 입력을 복사하여 디코더의 출력으로 활용하는 연구가 진행 되었다[4]. 본 논문과 가장 연관이 깊은 [5]는 검색 방법(Retrieving mechanism)을 통한 생성 채팅 모델로써 복사 방법에 지식베이스를 검색하여 목적어를 찾아내고, 복사 방법과 유사하게 목적어를 디코더의 출력으로 활용하였다. 본 논문에서는 Decoder의 출력에 복사 방법과 검색 방법을 적용하는 종단형 생성 기반 질의응답 채팅 시스템을 제안한다.

3. 생성 채팅 모델

3.1 복사 방법을 활용한 생성 채팅 모델

그림 1은 복사 방법을 적용한 주의 집중 기반 Encoder-Decoder 모델이다. 복사 방법은 주의 집중 기반 Encoder-Decoder 모델에서 입력 어휘를 출력으로 복사되도록 하는 방법이다.

입력 열 $X=\{x_1,x_2,...,x_i\}$을 인코딩하기 위해 각각의 입력을 Bi-directional-LSTM[6]을 사용하여 다음과 같은 수식으로 인코딩 한다.

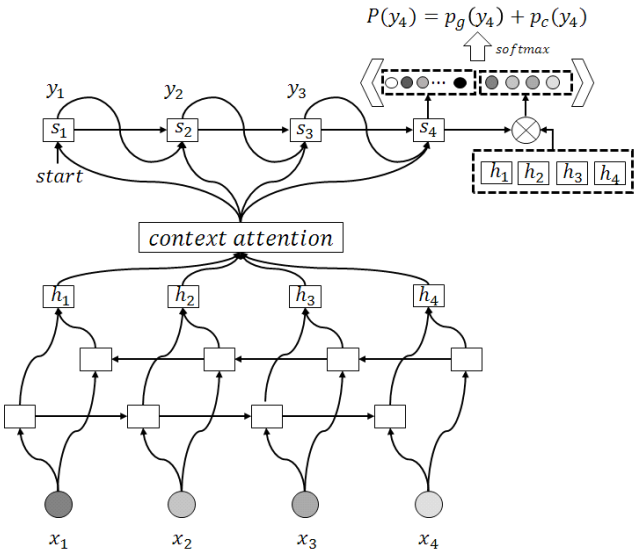

그림 1 복사 방법이 적용된 Encoder-Decoder 모델

$$
\begin{aligned}
h_{f,i} &= LSTM(x_i, h_{f,i-1}) \\
h_{b,i} &= LSTM(x_i, h_{b,i+1}) \\
h_i &= [h_{f,i}; h_{b,i}]
\end{aligned} \tag{1}
$$

식 1에서 $h_{f,i}$는 정방향의 은닉 계층이고, $b_{f,i}$는 역방향의 은닉 계층이며, [;]는 결합(concatenate)을 의미한다. 인코딩 한 은닉 계층들을 사용하여 다음 수식을 사용하여 고정된 차원의 문맥 벡터(context attention)를 생성한다.

$$
\begin{aligned}
e_{ij} &= f(s_{i-1}, h_j) \\
\alpha_{ij} &= \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})} \\
c_i &= \sum_{j=1}^{T_x} \alpha_{ij} h_j \\
s_i &= f(s_{i-1}, c_i)
\end{aligned} \tag{2}
$$

식 2에서 e_{ij}은 계산된 i번째 출력과 디코더 입력의 j번째 입력을 신경망에 적용한 값, α_{ij}는 e_{ij}를 확률로 표현한 주의 집중 가중치이다. T_x는 인코더의 길이, c_i는 주의 집중 가중치와 입력을 통해 생성된 문맥 벡터, s_i는 디코더의 은닉 계층, f는 비선형함수(Nonlinear function)이다. 복사 방법의 디코더의 출력은 다음과 같이 정의된다.

$$
\begin{aligned}
&P(y_i | s_i, c_i, y_{i-1}, h_{T_x}) = \\
&p_g(y_i | s_i, c_i, y_{i-1}, h_{T_x}) + p_c(y_i | s_i, c_i, y_{i-1}, h_{T_x})
\end{aligned} \tag{3}
$$

식 3에서 p_g는 생성 모드(Mode)의 확률을 의미하고, p_c는 복사 모드의 확률을 의미하고 h_{T_x}는 인코더의 은닉 계층을 의미한다. 두 확률은 각각 다음 수식과 같이 계산된다.

$$
\begin{aligned}
p_g(y_i | s_i, c_i, y_{i-1}, h_{T_x}) &= \frac{1}{Z} exp(\psi_g(y_i)), \\
& y_i \in V \\
p_c(y_i | s_i, c_i, y_{i-1}, h_{T_x}) &= \frac{1}{Z} \sum_{j: x_j = y_i} \exp(\psi_c(x_j)), \\
& y_i \in X, y_i \notin V
\end{aligned} \tag{4}
$$

식 4에서 V는 디코더에서 출력 될 수 있는 어휘 사전이고, Z는 생성 모드와 복사 모드가 공유하는 정규화 항(Normalization term) 이다. ψ_g는 생성 모드의 점수(Score)이고, ψ_c는 복사 모드의 점수이다. ψ_g와 ψ_c는 다음과 같이 계산된다.

$$
\begin{aligned}
\psi_g(y_i = v_k) &= \nu_k^T W_g * s_i + b_g \\
\psi_c(y_i = x_j) &= \sigma(h_j^T * W_c) s_i
\end{aligned} \tag{5}
$$

식 5에서 ν_k는 V에서 v_k의 벡터를 가리키는 One-hot이고, W_g는 생성 모드가 사용하는 가중치(Weight) 행렬이다. h_j는 입력 x_j의 은닉 계층이고, W_c는 복사 모드가 사용하는 가중치이고, σ는 비선형함수로 본 논문에서는 $Tanh$를 사용한다. 입력 어휘가 출력 어휘 사전에 같은 단어로 존재하면 생성 점수와 복사 점수를 합산하여 계산한다.

3.2 검색 방법을 활용한 생성 채팅 모델

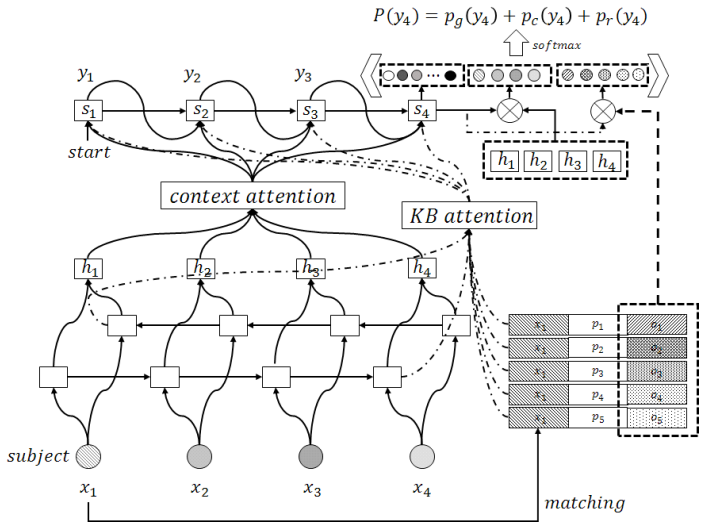

그림 2 복사 및 검색 방법이 적용된 Encoder-Decoder 모델

그림 2은 복사 방법 모델에 검색 방법을 적용한 모델이다. 검색 방법은 주어에 해당하는 술어와 목적어의 후보들을 검색하고 지식베이스 벡터(KB attention)를 생성한 이후 복사 방법과 비슷한 방법으로 목적어를 출력으

로 사용하는 방법이다.

검색 방법은 3.1절의 복사 방법의 입력 열 X에서 주어를 찾는다. 찾은 주어에 해당하는 술어 $P=\{p_1,p_2,...,p_l\}$와 목적어 $O=\{o_1,o_2,...,o_l\}$를 주어 x_s과 결합하여 트리플 임베딩 $k_s=[x_m;p_n;o_n]$를 생성한다. $k_s, h_{b,1}, h_{f,i}$를 사용하여 3.1절의 문맥 벡터 생성 방법과 같은 방법으로 지식베이스 벡터 kb_i를 생성한다. 검색 방법의 디코더의 출력은 다음과 같이 정의 된다.

$$\begin{aligned} &P(y_i|s_i,c_i,y_{i-1},h_{T_x},kb_i)= \\ &p_g(y_i|s_i,c_i,y_{i-1},h_{T_x},kb_i)+ \\ &p_c(y_i|s_i,c_i,y_{i-1},h_{T_x},kb_i)+ \\ &p_r(y_i|s_i,c_i,y_{i-1},h_{T_x},kb_i) \end{aligned} \quad (6)$$

식 6에서 p_r은 검색 모드의 확률을 의미한다. 세 확률은 각각 다음 수식과 같이 계산된다.

$$\begin{aligned} p_g(y_i|s_i,c_i,y_{i-1},h_{T_x},kb_i) &= \frac{1}{Z}exp(\psi_g(y_i)), \\ y_i &\in V \end{aligned}$$

$$\begin{aligned} p_c(y_i|s_i,c_i,y_{i-1},h_{T_x},kb_i) &= \frac{1}{Z}\sum_{j:x_j=y_i}\exp(\psi_c(x_j)), \\ y_i &\in X, y_i \notin V, y_i \notin O \end{aligned} \quad (7)$$

$$\begin{aligned} p_r(y_i|s_i,c_i,y_{i-1},h_{T_x},kb_i) &= \frac{1}{Z}\sum_{j:o_v=y_i}\exp(\psi_r(o_v)), \\ y_i &\notin X, y_i \notin V, y_i \in O \end{aligned}$$

식 7에서 Z는 생성 모드와 복사 모드와 검색 모드가 공유하는 정규화 항이고, ψ_r는 검색 모드의 점수이다. ψ_r는 다음과 같이 계산된다.

$$\psi_r(y_i=o_v)=\sigma(v_o^{T*}W_{kb})s_i \quad (8)$$

식 8에서 v_o는 V에서 o_v의 벡터를 가리키는 one-hot 벡터이고, W_{kb}는 검색 모드의 가중치이다. 목적어 어휘가 출력 어휘에 존재하면 검색 점수와 생성 점수를 합산하여 계산한다.

4. 실험 및 평가

4.1 실험 준비

본 논문에서는 복사 및 검색 방법을 실험하기 위해 [7]에서 사용한 템플릿 확장 방법을 사용하여 데이터를 구축하였다. 템플릿 확장 방법을 사용하기 위한 트리플 데이터는 DBPedia 2016-04[8]의 데이터 중 한국어 관계 트리플 정보(주어, 술어, 목적어)로 이루어진 데이터를 사용하였다. 표 1은 본 논문에서 사용한 Predicate의 종류와 개수이다. 확장 결과 총 1,050,575개의 데이터를 얻을 수 있었고 약 90%인 945,517개를 학습 데이터로 사용하고 나머지 데이터 105,058개를 실험 데이터로 사용하였다.

실험에 사용한 파라미터는 각 LSTM의 은닉 크기를 256, 단어의 차원 크기는 50, 배치 크기는 256으로 설정하였다. 검색 방법에서 검색을 하는 최대 길이는 5로 설정하고 주어는 문장에서 이미 찾았다는 가정하에 실험을 진행하였다. 실험의 입력과 출력 단위는 [7]에서 사용한 의사 형태소로 진행하였다. 주어와 목적어가 의사 형태소로 분리되는 것을 방지하기 위해 개체(Entity)를 하나의 단위로 묶어 사용 하였다.

표 1 Predicate의 종류와 개수

Predicate	데이터 개수
birthplace	25662
occupation	15508
nationality	12660
deathplace	4113
spouse	1674
developer	1448
place	1169
award	1003
parent	925
child	645
capital	533

4.2 실험 평가

실험 평가를 위해 예측한 문장에서 정답 목적어 존재 여부를 자동으로 체크하여 이에 대한 정확도(Accuracy)를 측정하였다.

표 2 복사 방법과 검색 방법의 성능

	Copying	Copying+Retrieving
정확도	37.8%	83.4%

표 2는 복사 방법 모델과 복사 방법에 검색 방법을 결합한 모델의 성능이다. 복사 방법 모델의 성능은 37.8%를 보여준 반면 복사 방법에 검색 방법을 결합한 모델의 성능은 83.4%로 복사 방법 모델보다 45.6% 높은 성능을 보여주었다. 이는 검색 방법이 검색된 목적어들의 후보를 바로 출력 디코더에 사용함으로써 예측 문장에 목적어가 등장할 확률을 상승시켰다고 판단된다.

4.3 결과 예시 및 분석

표 3은 복사 방법과 검색 방법을 결합한 모델의 실험 결과 예시이다. 주어는 굵은 글씨로, 목적어는 밑줄로 표현하였다. ID(1,4)는 정답 데이터와 일치된 예측 결과를 보여준다. ID(2)는 “생각”을 “만들”로 잘못 예측하였다. 이는 템플릿에 “생각 했어”와 “만들 었어”가 존재하는데, 가중치가 “만들”에 더 집중 된 것으로

표 3 복사 방법과 검색 방법을 결합한 모델의 결과 생성 예시

ID	Question	Gold	Predict
1	**요코미치 다카히로**가 태어난 곳은 ?	홋카이도에서 태어났습니다	홋카이도에서 태어났습니다
2	**모질라 애플리케이션 스위트**를 누가 만들었니 ?	모질라 재단이 생각 했어	모질라 재단이 만들 했어
3	**마르뱅 마르탱**은 어떤 나라 사람 인지 ?	**마르뱅 마르탱**은 France국적이지	**마르뱅 마르탱**는 France 국적이지
4	**어니스트 월턴**의 수상 내역 알려 줘	노벨 물리학상을 받았어~	노벨 물리학상을 받았어~
5	**왕다레이**의 국가는?	**왕다레이**는 차이나 국가를 가지고 있어	차이나는 차이나 국적를 가지고 있어
6	**교황 클레멘스 11세**가 죽은곳은?	이탈리아 입니다	이탈리아 에서
7	**오우라 가네다케**의 고향 은 어디인가요?	사쓰마 국에서 태어났습니다	사쓰마 국입니다 태어났습니다
8	**박병규**의 국가는?	**박병규**는 South Korea 국적이지	**박용택**는 South Korea 국적이지

보인다. 이와 비슷한 오류로 ID(7)이 있다. ID(3)은 자연스럽지 않은 조사가 예측되었는데, 이는 개체를 입력 단위로 분리하여 입력한 것이 아니라, 하나의 단위로 입력하였기 때문에 개체와 단어가 같은 공간에 매핑되지 않는 문제로 보인다. ID(5,8)은 복사 방법의 점수와 검색 방법의 점수가 상반되어 생긴 에러이다. ID(6)은 문장이 완전히 생성 되지 않은 문제로, 이는 디코더가 생성한 전체 문장을 고려하는 Search방법을 적용해야 할 것으로 보인다.

5. 결론

본 논문에서는 질의에 해당하는 목적어를 찾기 위한 방법으로 복사 방법과 검색 방법을 결합한 질의응답 채팅시스템을 제안하였다. 기존의 복사 방법에 검색 방법을 더하여 목적어가 출력에 더 잘 표현 될 수 있게 하였다. 향후 연구로 학습된 지식 임베딩을 직접 입력하는 연구를 진행할 예정이다.

감사의 글

이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임. (R0126-15-1117, 언어학습을 위한 자유발화형 음성대화처리 원천기술 개발)

참고문헌

[1] O. Vinyals and Q. Le, A neural conversational model. *arXiv preprint arXiv:1506.05869*, 2015.

[2] D. Bahdanau, K. Cho and Y. Bengio, Neural machine translation by jointly learning to align and translate, *arXiv preprint arXiv:1409.0473*, 2014.

[3] J. Li, M. Galley, C. Brockett, J. Gao and B. Dolan, A diversity-promoting objective function for neural conversation models, *arXiv preprint arXiv:1510.03055*, 2015.

[4] J. Gu, Z. Lu, H. Li and V. O. Li, Incorporating copying mechanism in sequence-to-sequence learning, *arXiv preprint arXiv:1603.06393*, 2016.

[5] S. He, C. Liu, K. Liu and J. Zhao, Generating Natural Answers by Incorporating Copying and Retrieving Mechanisms in Sequence-to-Sequence Learning. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vol. 1, pp. 199-208, 2017.

[6] M. Schuster and K. K. Paliwal, Bidirectional recurrent neural networks, *IEEE Transactions on Signal Processing,* 45.11, pp. 2673-2681, 1997.

[7] 김시형, 김학수, “의사 형태소 단위 채팅 시스템”, *제 28회 한글 및 한국어 정보처리 학술대회 논문집*, pp. 263-267, 2016.

[8] http://wiki.dbpedia.org/dbpedia-version-2016-04