Dual WGAN 기반 페르소나 Multi-Turn 챗봇

오신혁○, 김진태[1], 김학수, 이정엄*, 김선아*, 박영민*, 노명호*
강원대학교 컴퓨터정보통신공학과, 현대자동차 로보틱스팀*
osh7605@kangwon.ac.kr, kjt1505@ncsoft.com, nlpdrkim@kangwon.ac.kr,
{lee.jeongeom, seona, pymnlp, myungho.noh}@hyundai.com

Personalized Multi-Turn Chatbot Based on Dual WGAN

Shinhyeok Oh○, JinTae Kim, Harksoo Kim,
Jeong-Eom Lee*, Seona Kim*, Youngmin Park*, Myungho Noh*
Kangwon National University Computer and Communication Engineering
Robotics Team, Hyundai Motor Company*

요 약

챗봇은 사람과 컴퓨터가 자연어로 대화를 주고받는 시스템을 말한다. 최근 챗봇에 대한 연구가 활발해지면서 단순히 기계적인 응답보다 사용자가 원하는 개인 특성이 반영된 챗봇에 대한 연구도 많아지고 있다. 기존 연구는 하나의 벡터를 사용하여 한 가지 형태의 페르소나 정보를 모델에 반영했다. 하지만, 페르소나는 한 가지 형태로 정의할 수 없어서 챗봇 모델에 페르소나 정보를 다양한 형태로 반영시키는 연구가 필요하다. 따라서, 본 논문은 최신 생성 기반 Multi-Turn 챗봇 시스템을 기반으로 챗봇이 다양한 형태로 페르소나를 반영하게 하는 방법을 제안한다.

주제어: 챗봇, 페르소나, WGAN

1. 서론

챗봇은 사람과 컴퓨터가 자연어로 대화하는 시스템을 말한다. 최근 휴대폰 메시지 앱에도 챗봇을 포함하고 있는 만큼 챗봇 연구는 활발히 진행되고 있다. 페르소나를 반영한 챗봇이란 개인 특성(Persona)이 반영된 응답을 하는 챗봇을 말한다. 사용자가 원하는 개인 특성을 가진 챗봇과 대화를 하게 되면 더욱 친근한 대화가 가능해지기 때문에 기존의 기계적인 응답 대신 페르소나를 반영한 응답을 하는 챗봇 연구가 필요하다. 따라서 본 논문은 생성 기반 Multi-Turn 챗봇에 페르소나를 반영시킨 페르소나 Multi-Turn 챗봇 시스템을 제안한다.

2. 관련 연구

기존의 sequence-to-sequence 모델을 기반으로 한 챗봇 연구는 짧고 일반적인 응답을 하거나, 의미상으로 잘못된 응답을 하는 문제가 있다[1-2]. 위의 문제점을 해결하기 위해 VAE(Variational Auto-Encoder)[3]를 사용하면 잠재변수(Latent Variable)를 통해 짧고 일반적인 응답을 피하고, 다양한 응답을 생성할 수 있다[4-5]. 하지만, VAE를 사용한 모델은 디코더가 잠재 변수를 무시하고 표준 정규 분포로 잠재 변수를 단순화하는 문제가 있다[6-7]. 이 문제는 적대적 학습 방법(Generative Adversarial Networks : GAN)[7]을 통해 잠재 변수 공간을 학습하여 부분적으로 해결되었다[8]. 하지만, 자연어와 같은 이산 데이터를 적대적으로 학습시키는 것은 어렵다[9-10]. 특히, 판별 모델(discriminant model)을 학습할 때 디코더의 각 단계에서 단어 확률 분포를 계산해야 하는데 미분을 할 수 없는 문제가 있다. 위 문제를 해결하기 위해 [11]은 학습 시 디코더가 생성한 응답 벡터(Response Vector)를 사용하여 미분 불가능에 대한 문제를 해결한 새로운 학습 방법을 제안했다.

생성 기반 챗봇에 페르소나를 반영하는 연구는 디코더가 단어를 생성할 때 하나의 벡터를 사용하여 페르소나 정보를 반영했다[12-13]. 하지만, 개인의 특성을 담고 있는 페르소나는 하나의 벡터와 같이 한 가지 형태로 정의할 수 없어서 챗봇 모델에 페르소나 정보를 다양한 형태로 반영시키는 연구가 필요하다. 따라서, 본 논문에서는 적대적 학습 방법을 통해 다양한 응답을 생성할 수 있는 [11]의 생성 모델을 기반으로 챗봇이 응답을 생성할 때 특정 페르소나를 반영하여 응답을 생성할 수 있도록 구현했으며, [14]의 사전학습 방법을 통해 소량의 페르소나 말뭉치에 대해 효과적인 모델을 제안한다.

3. Dual WGAN 기반 페르소나 Multi-Turn 챗봇

그림 1은 본 논문에서 제안하는 Dual WGAN 기반 페르소나 Multi-Turn 챗봇이다. 제안 모델은 [11]의 Dual WGAN 모델과 페르소나 주의 집중 모델로 구성된다. Dual WGAN 모델은 입력 인코더(Query Encoder), 입력-응답 매퍼(Query-to-Response Mapper; 이하 QR Mapper), 응답 문장 매퍼(Response-to-Response Mapper; 이하 RR Mapper)로 구성되고, 페르소나 주의 집중 모델은 성격에 대한 문장 형태의 페르소나를 반영하는 성격 기반 주의 집중 모듈(Style Attention), 응답에 직접 나타날 수 있

1) current address, NC Soft

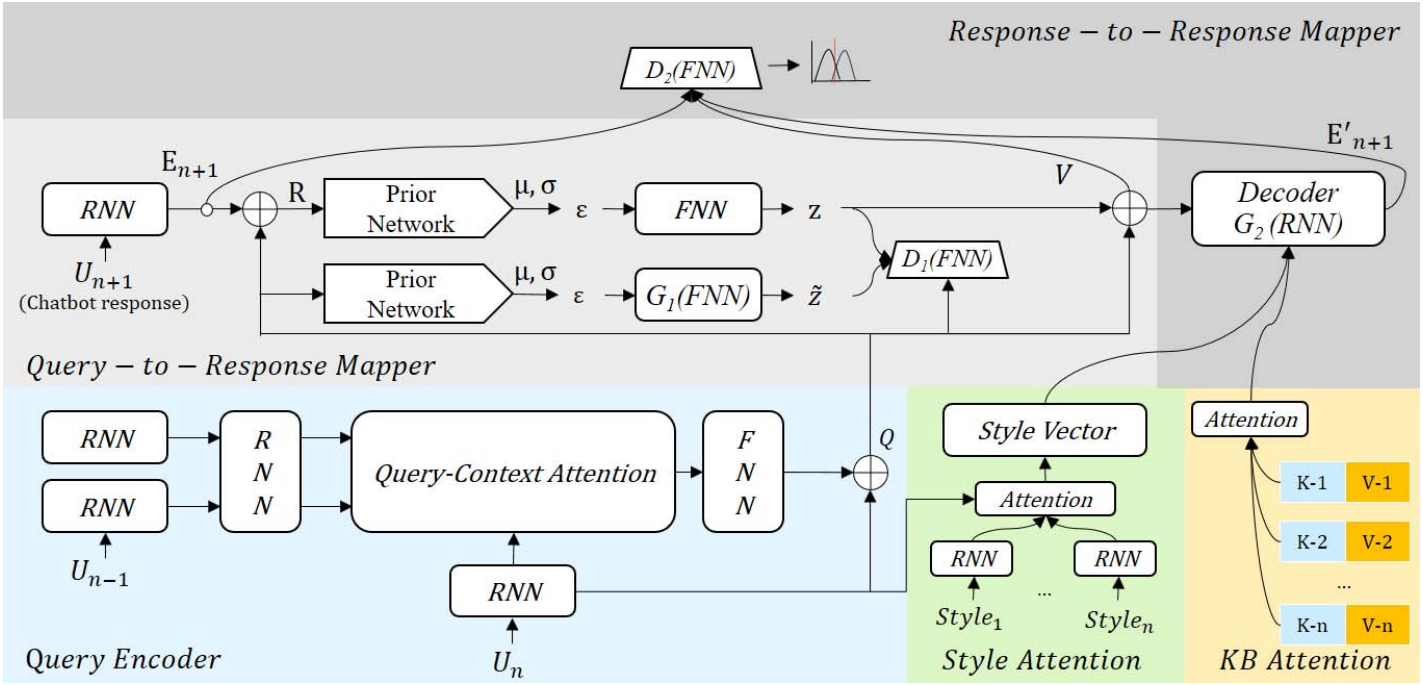

그림 1 페르소나 주의 집중 방법을 적용한 Dual WGAN 기반 Multi-Turn 챗봇 구조도

는 단어(word) 또는 구(phrase) 형태의 지식을 Key-Value 형태로 반영하는 지식 기반 주의 집중 모듈(KB Attention; Knowledge Based Attention)의 두 가지로 구성된다.

입력 인코더는 순환 신경망(Recurrent Neural Networks : RNN)과 주의 집중(Scaled Dot-Product Attention Mechanism)[15]을 통해 사용자의 현재 발화와 k개의 이전 발화를 인코딩한 입력 벡터(Q; Query Vector)를 반환한다. 성격 기반 주의 집중 모듈은 챗봇의 성격을 반영하는 n개의 문장을 각각 순환 신경망의 입력으로 사용하고, 각 순환 신경망의 출력(output)에 대해 주의 집중 방법[16]을 사용해 성격 벡터(Style Vector)를 반환한다. 지식 기반 주의 집중 모듈은 답변에 직접 포함될 수 있는 페르소나를 Key-Value 형태로 반영하는 모듈이다. Key에는 사는 곳, 취미 등 도메인 정보를 담고 있고, Value는 춘천, 영화감상 등 Key에 해당하는 내용에 대한 정보를 담고 있다. 디코더의 은닉 상태 정보를 기준으로 Key 정보들의 주의 집중 분포를 구하고, 최종 주의 집중 벡터(Attention Vector)는 Value의 벡터로 생성한다. 입력-응답 매퍼는 순환 신경망으로 인코딩된 벡터(E_{n+1})와 입력 벡터를 연결(concatenation)한 응답 벡터(R; Response Vector)와 질의 벡터를 유사하게 만들도록 학습한다. 응답 문장 매퍼는 순환 신경망으로 인코딩된 벡터(E_{n+1})와 적대적 학습 방법으로 디코딩된 응답 벡터(E'_{n+1})가 유사해지도록 학습한다.

3.1. 입력 인코더(Query Encoder)

입력 인코더는 GRU(Gated Recurrent Unit)[17]으로 이루어진 순환 신경망이며, 현재 사용자 발화(U_n)와 이전 발화(U_{n-k}, ... , U_{n-1})로 구성된 대화를 입력받아 인코딩한다.

$$E_i = GRU(U_i) \quad (1)$$

식 (1)에서, E_i는 순환 신경망에 의해 인코딩된 발화 벡터이다. 그리고 식 (2)와 같이 순환 신경망을 사용하여 문맥 정보를 각각 인코딩된 발화에 반영한다.

$$\tilde{E}_i = GRU(E_i) \quad (2)$$

식 (2)에서, $\tilde{E}_i$는 순환 신경망에서 i번째 단계의 출력 벡터이다. 현재 발화와 이전 발화 간의 문맥적인 연관성을 강하게 반영하기 위해 식 (3)과 같이 주의 집중 점수를 계산한다.

$$a_i = \frac{1}{z}\exp\left(\frac{\tilde{E}_i \circ E_n}{\sqrt{d}}\right), where\ i \neq n \quad (3)$$

여기서 z와 d는 각각 정규화 인자(Normalization Factor), 인코딩된 벡터의 크기를 의미한다. 이후, 식 (3)의 a_i를 이용하여 그림 1의 Query-Context Attention(이하, QC Attention) 벡터 A를 계산한다. 벡터 A는 식 (4)에 표현된 것처럼 응답을 생성하기 위해 고려해야 하는 문맥 벡터를 의미한다.

$$A = \sum_{i=1}^{k} a_i \tilde{E}_i \quad (4)$$

마지막으로, 입력 인코더는 식 (5)와 같이 인코딩된 현재 사용자 발화와 QC Attention 벡터를 연결한 벡터 Q를 반환한다.

$$Q = E_n \oplus A \quad (5)$$

3.2. 성격 기반 주의 집중 모듈(Style Attention)

성격을 나타내는 문장($style_1$, ... , $style_n$)들이 입력되면 Style Attention은 순환 신경망을 사용하여 각 문장을 인코딩한다. 인코딩된 문장들과 현재 발화 간에 주의 집중을 계산하여[16] 그림 1의 Style Vector를 반환한다.

3.3. 지식 기반 주의 집중 모듈(KB Attention)

KB Attention에서는 Key 단어의 벡터(K_1, ... , K_n)와 Value 단어의 벡터(V_1, ... , V_n)를 연결한 형식으로 입력된다. 그 후 디코더의 상태(State)를 기준으로 Key 단어 벡터의 주의 집중[18] 분포를 구하고, 반환할 최종 Attention Vector는 Value 단어의 벡터로 생성한다.

3.4. 입력 응답 매퍼(QR Mapper)

QR Mapper는 Query Vector(Q)를 Response Vector(R)에 매핑하는 역할을 한다. 매핑 성능을 향상시키기 위해 WAE(Wasserstein auto-encoder) 모델[8]을 사용했다. WAE 모델은 Wasserstein GAN(WGAN)[19]을 사용하여 Generator를 최적화한다. QR Mapper는 응답 발화를 나타내는 U_{n+1}과 Q를 연결하여 순환 신경망을 사용하여 인코딩한다. 이후, 식 (6)과 같이 응답 벡터 R을 생성한다.

$$R = Q \oplus E_{n+1} \quad (6)$$

이후 QR Mapper는 R과 Q 각각에 완전 연결 신경망(Fully-connected Neural Networks : FNN)을 사용한 후(그림 1의 Prior Network) 가우시안 잡음(Gaussian Noise; ϵ)을 추가한다. 가우시안 잡음은 완전 연결 신경망(그림 1의 FNN)을 통해 각각 잠재 변수(Latent Variable) z와 $\tilde{z}$로 변환된다. 그리고 그림 1의 판별기($D_1(FNN)$)에서는 z와 $\tilde{z}$ 중 어떤 것이 실제 z인지 구별하도록 학습한다. 이후 학습 과정에서는 Q와 z를 연결하여 그림 1의 $G_2(RNN)$에 입력한다. 추론 과정에서는 Q와 $\tilde{z}$를 연결하여 $G_2(RNN)$에 입력한다.

3.5. 응답 문장 매퍼(RR Mapper)

RR Mapper는 E'_{n+1}을 E_{n+1}에 매핑하는 역할을 한다. 매핑 성능을 향상시키기 위해 WAE 모델[8]을 다시 사용한다. 적대적 학습 방법을 통해 E'_{n+1}이 E_{n+1}과 유사해지도록 학습한다. 학습 시 E'_{n+1}과 E_{n+1} 각각 그림 1의 V와 연결하여 판별기($D_2(FNN)$)로 입력한다.

4. 실험 및 평가

4.1. 데이터 및 실험 방법

본 논문에서는 [14]의 사전학습 방법 적용 유무에 대한 실험을 진행했다. 실험 데이터는 동일한 질문에 대한 가상의 인물 2명의 답변 쌍 데이터 1,365쌍을 사용했다. 가상 인물에 따른 페르소나는 개인의 인적사항을 포함한 Key-Value 쌍과 성격 등을 포함한 다수의 문장으로 이루어져 있다. 실험은 전체 데이터를 학습 데이터 1,245개, 검증 데이터 60개, 평가 데이터 60개로 나누어 진행했다. 사전학습에는 구어체 말뭉치 40,000개를 사용했다. 사전학습 시 [14]의 학습 방법을 적용하여 오토인코더(Autoencoder) 방식으로 학습했다. 실험에 사용한 임베딩은 미등록어에 강건한 [20]의 복합 표현 단위를 사용했다. 복합 표현 단위 임베딩이란 형태소 단위 임베딩을 기본으로 사용하되 열린 단어(고유 명사 및 미등록어 형태소)에 대해 [21]의 합성곱 신경망을 음절 단위로 적용하여 형태소 단위 임베딩을 보완하는 역할을 한다. 실험의 평가는 정성평가로 진행했다. 정성평가는 입력으로 문장 20개를 사용하여 가상의 인물 X(30대 여성), Y(10대 남성)에게 질문했을 때 나올 응답과 일반적으로 나올 응답 3가지를 생성해 총 60개의 답변 데이터로 평가를 진행했다. 평가는 문법, 의미, 페르소나 반영 여부의 3가지에 대해 진행했다. 문법 평가는 문법에 부합하면 1점, 부합하지 않으면 0점으로 선택하게 했으며 의미 평가는 입력 문장에 대해 의미상 적합한 답변을 제공했으면 1점, 모호한 답변을 제공했으면 0.5점, 의미상 맞지 않는 답변을 제공했으면 0점으로 선택하게 했다. 페르소나 반영 여부 평가는 평가자(A, B)에게 가상 인물 X, Y의 프로필을 제공하고, 생성한 답변이 각각의 가상 인물이 답할만한 내용이면 1점, 모호하면 0.5점, 전혀 맞지 않는 답변을 제공했으면 0점으로 선택하게 했다.

4.2. 사전학습 적용 실험

모델	평가자	문법 (0/1)	의미 (0/0.5/1)	페르소나 (0/0.5/1)
대화학습	A	0.683	0.275	0.600
	B	0.750	0.458	0.938
	평균	0.7165	0.3665	0.7690
사전학습 + 대화학습	A	0.917	0.458	0.563
	B	0.950	0.550	0.988
	평균	**0.9335**	**0.5040**	**0.7755**

표 1 사전학습 적용에 따른 정성평가 결과

표 1은 사전학습 적용에 따른 정성평가 결과이다. “대화학습” 모델은 1,245개의 페르소나 반영 데이터로만 학습한 모델을 의미하고, “사전학습” 모델은 [14]의 사전학습 방법을 사용하여 구어체 말뭉치 40,000개의 데이터로 사전학습을 진행한 모델을 의미한다. 대화학습만 진행한 모델보다 사전학습을 진행한 후 대화학습을 진행한 모델이 높은 점수를 보였다.

4.3. 사전학습 적용 실험 예제

질문		#t #v <sp> 자주 <sp> 보 아 ?
대화학습	일반	지금 <SP> 저 와 <SP> 보 고 <SP> 관심 <SP> 많이 <SP> 알리 는 <SP> 명 도 <SP> 없 어요 ?
	X	우리나라 <SP> 피겨 <SP> 스케이팅 <SP> 역사 를 <SP> 새로 <SP> 쓰 ㄴ <SP> 대단 하 ㄴ <SP> 선수 이 죠 .
	Y	그럼요 . <SP> 로봇 이 라 는 <SP> 말 이 <SP> 있 던데 <SP> #에 #너 #자 #자 #저 <SP> 배터리 <SP> 쓰 ㄴ다고 <SP> 소문나 았 어요 . <SP> 히히 .
사전학습 + 대화학습	일반	날씨 가 <SP> 정말 <SP> 어떻 ㄴ가요 ?
	X	매일 <SP> 밤 <SP> 영상 을 <SP> 챙기 어 <SP> 보 ㄴ답니다 .
	Y	네 . <SP> 많이 <SP> 보 았 어요 .

표 2 사전학습 적용에 따른 실험 예제

표 2는 사전학습 적용에 따른 실험 예제이다. #이 붙은 내용은 음절로 생성한 부분이다. “일반”은 페르소나를 반영하지 않은 결과를 의미한다. “X”와 “Y”는 각각 30대 여성과 10대 남성의 페르소나를 적용한 모델을 의미한다. 표 2에 따르면 사전학습까지 적용한 대화학습이 사전학습을 적용하지 않은 대화학습보다 문법과 의미적으로 올바른 응답을 생성하는 것을 확인할 수 있다.

5. 결론 및 향후 연구

본 논문은 Dual WGAN 기반 Multi-Turn 챗봇에서 페르소나를 반영한 응답을 생성하는 방법을 제안했다. 향후 연구로 디코더를 양방향으로 학습시키는 방법[22]을 제안 모델에 적용해보고, 디코더 각 step 생성 시 출력값과 임베딩 점수를 측정하여 생성하는 방법[23]을 적용해 볼 예정이다.

감사의 글

본 연구는 현대자동차 산학연구용역 과제의 지원을 받아 수행되었음.

참고문헌

[1] J. Li, M. Galley, C. Brockett, J. Gao, B. Dolan, “A Diversity-Promoting Objective Function for Neural Conversation Models”, arXiv preprint arXiv:1510.03055, 2015.

[2] S. Sato, N. Yoshinaga, M. Toyoda, M. Kitsuregawa, “Modeling Situations in Neural Chatbots”, In Proceedings of the ACL 2017, Student Research Workshop, pp. 120-127, 2017.7.

[3] D. P. Kingma, M. Welling, “Auto-encoding variational bayes”, arXiv preprint arXiv:1312.6114, 2013.

[4] T. Zhao, R. Zhao, M. Eskenazi, “Learning Discourse-level Diversity for Neural Dialog Models using Conditional Variational Autoencoders”, In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Volume 1, pp. 654-664, 2017.7.

[5] X. Shen, H. Su, S. Niu, V. Demberg, “Improving Variational Encoder-Decoders in Dialogue Generation”, In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, 2018.4.

[6] P. Goyal, Z. Hu, X. Liang, C. Wang, E.P. Xing, “Nonparametric Variational Auto-Encoders for Hierarchical Representation Learning”, In Proceedings of the IEEE International Conference on Computer Vision, pp. 5094-5102, 2017.10.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, “Generative adversarial nets”, In Advances in neural information processing systems, pp. 2672-2680, 2014.

[8] X. Gu, K. Cho, J.W. Ha, S. Kim, “DialogWAE: Multimodal Response Generation with Conditional Wasserstein Auto-Encoder”, arXiv:1805.12352, 2018.

[9] J. Li, W. Monroe, T. Shi, S. Jean, A. Ritter, D. Jurafsky, “Adversarial Learning for Neural Dialogue Generation”, arXiv:1701.06547, 2017.

[10] L. Yu, W. Zhang, J. Wang, Y. Yu, “Seqgan: Sequence Generative Adversarial Nets with Policy Gradient”, In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, 2017.2.

[11] J. Kim, S. Oh, O.-W. Kwon, H. Kim, “Multi-Turn Chatbot Based on Query-Context Attentions and Dual Wasserstein Generative Adversarial Networks”, Appl. Sci. 3908, 2019.9.

[12] Jiwei Li, Michel Galley, Chris Brockett, “A Persona-Based Neural Conversation Model”, arXiv preprint arXiv:1603.06155v2, 2016.

[13] 오신혁, 김진태, 박영민, 김선아, 이정엄, 김학수, “생성 기반 챗봇 시스템을 위한 페르소나 반영 방

법”, 2019 한국컴퓨터종합학술대회, pp. 1761-1763, 2019.06.
[14] K. Song, X. Tan, T. Qin, J.Lu, & T.Y. Liu, “Mass: Masked sequence to sequence pre-training for language generation”, arXiv preprint arXiv:1905.02450, 2019.
[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, AN. Gomez, Ł. Kaiser, I. Polosukhin. “Attention is All You Need”, In Advances in neural information processing systems, pp. 5998-6008, 2017.
[16] D. Bahdanau, K. Cho, Y. Bengio, “Neural machine translation by jointly learning to align and translate”, arXiv preprint arXiv:1409.0473, 2014.
[17] K. Cho, B.V. Merriënboer, D. Bahdanau, Y. Bengio, “On the properties of neural machine translation: encoder-decoder approaches”, arXiv preprint arXiv:1409.1259, 2014.
[18] M.T. Luong, H. Pham, C.D. Manning, “Effective approaches to attention-based neural machine translation”, arXiv preprint arXiv:1508.04025, 2015.
[19] M. Arjovsky, S. Chintala, L. Bottou, “Wasserstein GAN”, arXiv preprint arXiv:1701.07875, 2017.
[20] 김진태, 이현구, 김학수, “소량의 대화 말뭉치에서 학습 가능한 효과적인 생성 기반 챗봇 모델”, 정보과학회논문지 제46권 제3호, pp. 246-252, 2019.
[21] Y. Kim, Y. Jernite, D. Songtag, and A. M. Rush, "Character-Aware Neural Language Models," Proc. of AAAI, pp. 2741-2749, 2016.
[22] L. Zhou, J. Zhang, C. Zong, “Synchronous bidirectional neural machine translation”, Transactions of the Association for Computational Linguistics, 7, 91-105, 2019.
[23] S. Ma, X. Sun, W. Li, S. Li, W. Li, X. Ren, “Query and output: Generating words by querying distributed word representations for paraphrase generation”, arXiv preprint arXiv:1803.01465, 2018.