
| | |
|--------------------|--|
| 저자 (Authors) | 김학수 Harksoo Kim |
| 출처 (Source) | 정보과학회지 35(8) , 2017.8, 9-18(10 pages) Communications of the Korean Institute of Information Scientists and Engineers 35(8) , 2017.8, 9-18(10 pages) |
| 발행처 (Publisher) | 한국정보과학회 KOREA INFORMATION SCIENCE SOCIETY |
| URL | http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07225985 |
| APA Style | 김학수 (2017). 인공지능 음성언어 비서 시스템의 자연언어처리 기술들. 정보과학회지, 35(8), 9-18 |
| 이용정보 (Accessed) | 상명대학교 59.30.116.*** 2021/03/07 20:51 (KST) |

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

인공지능 음성언어 비서 시스템의 자연언어처리 기술들

강원대학교 ■ 김학수*

1. 서 론

애플(Apple)의 시리(Siri), 아마존(Amazon)의 알렉사(Alexa)로 대표되는 인공지능 음성언어 비서 시스템은 ‘에코(Echo)’, ‘기가지니(GiGA Genie)’, ‘누구(Nugu)’와 같은 디지털 음성 보조 기기(digital voice assistant device)와 결합되면서 시장 규모가 급격히 증가되고 있다. 시장 분석 기관인 Strategy Analytics는 디지털 음성 보조 기기의 전세계 판매량이 올해 290만대에서 2020년에는 1,510만대 규모의 시장으로 커질 것으로 전망하고 있다([그림 1] 참조)[1].

인공지능 음성언어 비서 시스템의 핵심 기능은 사용자의 의도를 파악하여 적절한 기능을 수행하거나 응답을 생성하는 것이라고 할 수 있으며, 이러한 기능을 수행하기 위해서는 다양한 자연어처리 기술이 요구된다.

<표 1>에서 보듯이 음성언어 비서 시스템의 대화 기능을 유형별로 분류해 보면 신변잡기 채팅 대화, 1문1답 질의응답 대화, 목적지향 대화로 나눌 수 있다. 채팅 대화를 처리하기 위해서는 유사 질의 검색과 언어모델 기반 응답 생성과 같은 언어처리 기술이

표 1 음성언어 비서 시스템의 대화 예제

| 화자 | 발화 | 유형 |
|-----|----------------------------|---------|
| 사용자 | 안녕 | 채팅 대화 |
| 시스템 | 반갑습니다. | |
| 사용자 | 오늘 날씨 좀 알려줘 | 질의응답 대화 |
| 시스템 | 오늘 기온은 29.5도로 다소 더울 예정입니다. | |
| 사용자 | 내일 일정 등록해줘 | 목적지향 대화 |
| 시스템 | 몇 시로 등록할까요? | |
| 사용자 | 오전 11시 | |
| 시스템 | 등록할 내용을 말씀해 주세요. | |

필요하며, 질의응답 대화를 처리하기 위해서는 질의유형 분류, 개체명 인식 및 엔터티 링킹(entity linking), SPO(Subject-Predicate-Object) 트리플 인식과 같은 언어처리 기술이 필요하다. 또한 목적지향 대화를 처리하기 위해서는 담화 분석과 대화 모델링 기술이 필요하다. 다시 말해서 음성언어 비서 시스템은 거의 모든 자연어처리 기술을 총망라하여 사용한다고 할 수 있다. 본 논문에서는 음성언어 비서 시스템 구현에 필요한 자연어처리 기술들 중에서 반드시 필요하면서도 접근 방법이 서로 대비되는 기술을 이해하기 쉽게 개념적으로 소개한다.

2. 음성언어 비서 시스템을 위한 자연어처리 기술

2.1 음성인식 후처리

음성언어 비서 시스템 기술은 화자의 말을 정확히 인식하는 데서 시작한다. 음성인식기의 성능이 꾸준히 좋아지고 있지만 영역이 정해지지 않은 자유 발화에 대해서는 여전히 많은 인식 오류율을 보이고 있다. 이러한 음성인식 오류는 자연어처리 모듈의 성능하락에 직접적인 영향을 끼친다. [그림 2]는 음성인식 단계에서 발생한 띄어쓰기 오류와 철자 오류가 형태소 분석 결과에 전파된 예를 보여준다.

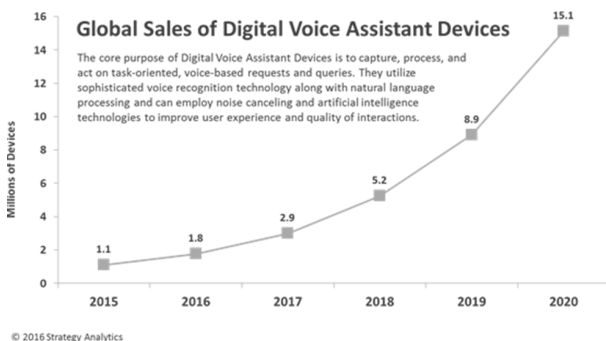


그림 1 디지털 음성 보조 기기의 전세계 판매량 예측

* 종신회원

† 이 논문은 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2016R1A2B4007732). 또한 엔씨소프트 산학연구과제의 지원을 받아 수행된 연구임(과제명: 지능형 채팅 시스템 개발).

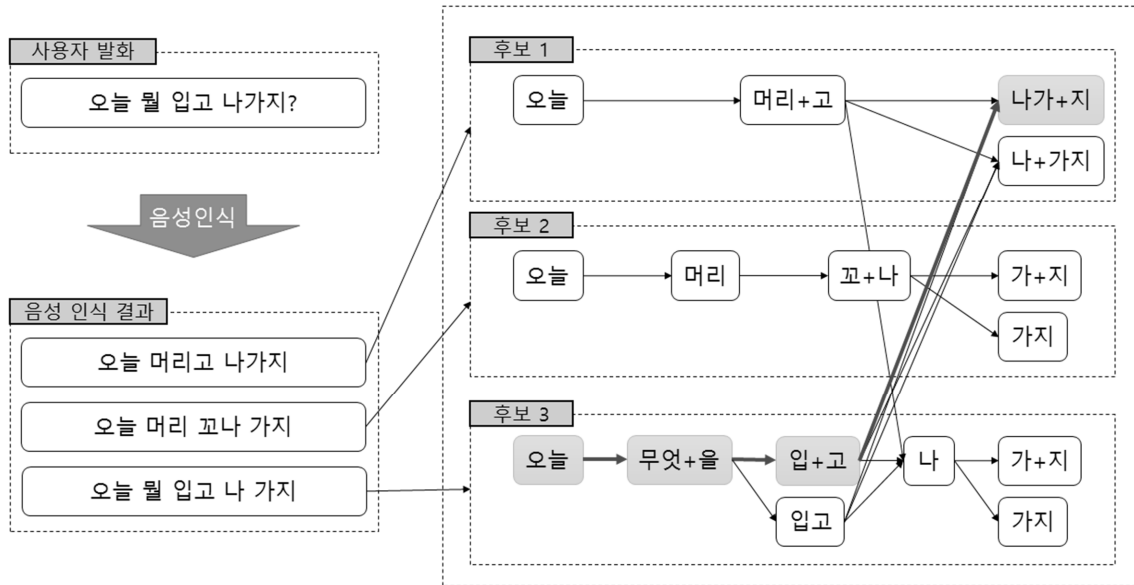


그림 2 음성인식 오류에 따른 형태소 분석 결과

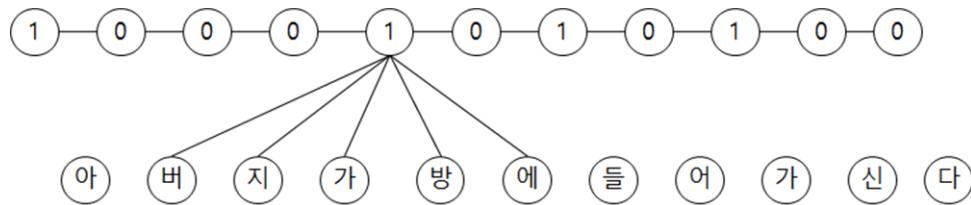


그림 3 순차 레이블링 모델을 이용한 띄어쓰기 오류 보정

음성인식 오류를 보정하기 위한 후처리 연구는 크게 띄어쓰기 오류 보정, 철자 오류 보정, 음성인식 결과 재순위로 나뉜다.

띄어쓰기 오류 보정 모델은 음성인식 결과에서 모든 띄어쓰기를 삭제한 후, 기계학습 모델을 이용하여 띄어 쓸 위치를 결정하는 순차 레이블링(sequence labeling) 방법을 주로 사용한다[2,3].

[그림 3]에서 보는 것과 같이 순차 레이블링 방법은 문장의 처음부터 끝까지 음절과 음절 사이에 띄어쓰기 여부를 판단하여 표지(그림에서는 0 또는 1)를 부착한다. 표지 결정을 위해서는 현 위치를 기준으로 앞뒤 n 개의 음절을 자질로 사용하며, 기계학습 방법으로는 양방향(처음부터 현재위치와 마지막부터 현재 위치까지) 전이 정보(transition information)를 잘 반영하는 것으로 알려져 있는 CRFs(Conditional Random Fields)나 Structural SVM(Support Vector Machine)이 주로 사용된다.

철자오류 보정 모델은 사전이나 통계 정보를 주로 이용한다[4,5,6,7,8,9]. 사전을 사용한 철자 오류 보정은 주로 영어권 국가들이 사용하는 방법으로 모든 입력문장의 각 단어들을 일반 사전에서 검색한 후, 일반 사전에

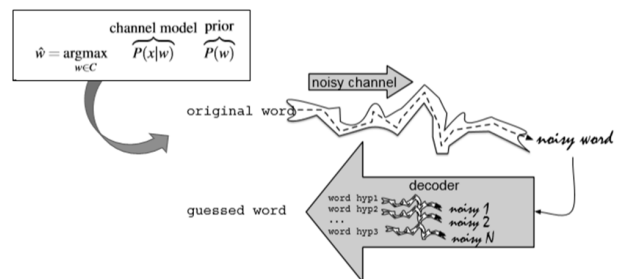


그림 4 노이즈 채널 모델을 이용한 철자 오류 보정

존재하지 않는 단어는 철자 오류라고 가정한다. 철자 오류로 가정된 문자열은 편집거리(edit-distance)나 메타폰(metaphone) 알고리즘 등을 사용하여 해당 단어와 거리가 가까운 사전 단어로 대체된다[6]. 이러한 사전 기반의 방법은 모든 철자 오류 후보 단어들을 미리 구축해야 하는 단점이 있으며, 용언의 활용이 많은 한국어의 경우에는 어휘의 기본형을 찾는 데 따르는 비용이 클 뿐만 아니라 잘못된 기본형을 찾을 가능성이 높다는 단점이 있다. 이러한 단점을 극복하기 위해 통계 정보에 기반한 확률 모델들이 제안되었다. [그림 4]는 대표적인 확률 기반 철자 오류 보정 방법인 노이즈 채널(noise channel) 모델을 보여준다[10].

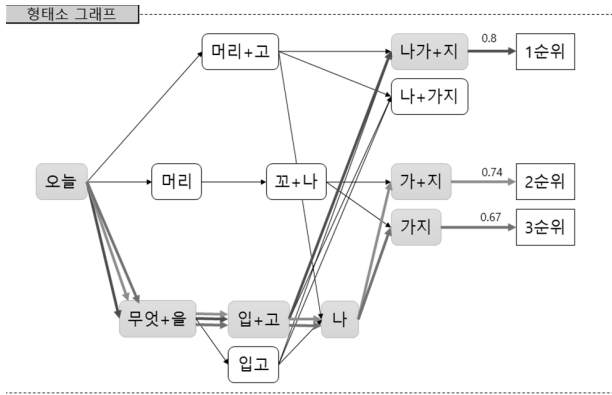


그림 5 형태소 그래프 기반 음성인식 결과 재순위화

[그림 4]에서 보듯이 노이즈 채널 모델은 입력된 노이즈(철자 오류)를 생성할 확률을 최대화하는 단어를 찾는 생성 모델의 일종이다. [그림 4]에서 $P(X|W)$ 는 올바른 단어와 오류 단어 사이의 음절(또는 알파벳) 단위 편집 거리를 기반으로 계산된다. 예를 들어, 올바른 단어 'actress'로부터 오류 단어 'acress'가 생성될 확률 $P(acress|actress)$ 는 $P(c|ct)$ (='ct'에서 'c'가 삭제되거나 'ct'가 'c'로 대체될 확률)를 이용하여 계산된다.

음성인식 결과 재순위화 모델은 1순위 음성인식 결과를 바탕으로 오류를 보정하는 기존 연구의 한계를 극복하기 위해서 상위 n 개의 음성인식 결과로부터 올바른 형태소열들의 결합을 찾고자하는 방법이다. [그림 5]는 상위 n 개의 음성인식 결과에 대한 형태소 그래프로부터 올바른 형태소열을 찾아내는 재순위화 모델의 개념도를 보여준다.

[그림 5]에서 보듯이 음성인식 결과 재순위화 모델은 후보 형태소들의 집합으로부터 올바른 형태소열을 생성하는 문제이며, CNN(Convolutional Neural Network)을 이용하여 이러한 문제를 풀려는 연구가 시도되고 있다[11].

2.2 개체명 인식과 엔터티 링킹

음성언어 비서 시스템의 핵심 서비스 중에 하나는 사용자의 검색 의도를 파악하여 정답을 제시해 주는 질의응답 모델이라고 할 수 있다. 질의응답 모델을 구현하기 위해서는 질의에 포함된 개체명을 추출하여 범주를 할당하는 개체명 인식 기술과 해당 개체명을 지식베이스 엔터티와 연결하는 엔터티 링킹 기술이 필요하다. 예를 들어, "빌 클린턴의 딸의 배우자는?"이라는 질의가 입력되면 '빌 클린턴'이 'PERSON'에 해당하며, 그것은 지식베이스의 'Bill_Clinton'에 해당한다는 것을 결정해야 지식베이스로부터 정답을 찾아낼 수 있다.

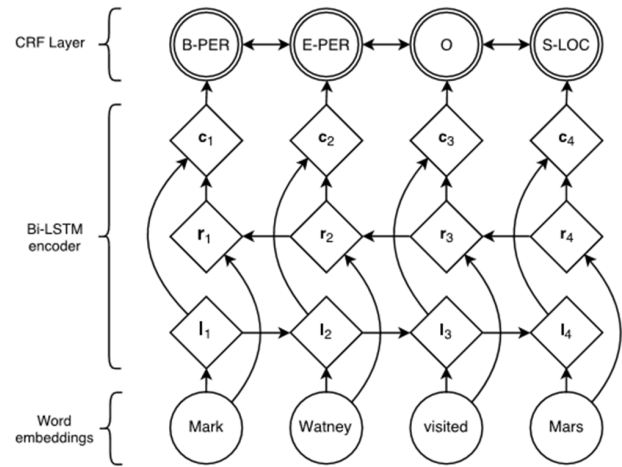


그림 6 Bi-LSTM-CRF 기반 개체명 인식[15]

개체명 인식 모델은 순차 레이블링 문제의 하나로 다루어져 왔으며, 띄어쓰기 오류 보정 모델과 마찬가지로 CRFs나 Structural SVM 기반의 다양한 연구들이 진행되었다[12,13,14]. 최근에는 자질 엔지니어링(feature engineering; 효과적인 자질을 추출하고 선택하는 작업)의 부담을 줄이기 위해서 양방향성 Bi-LSTM (Bidirectional Long Short Term Memory)이나 Bi-GRU (Bidirectional Gated Recurrent Units)와 같은 딥뉴럴넷(deep neural network)에 CRFs를 결합한 연구들이 활발히 진행되고 있다[15]. [그림 6]은 Bi-LSTM-CRF를 이용하여 개체명을 인식하는 모델의 개념도를 보여준다.

[그림 6]에서 'B-PER', 'E-PER', 'O', 'S-LOC'은 각각 인명의 시작, 인명의 끝, 개체명 아님, 독립단어로 이루어진 장소명을 나타내는 표지이다. 이와 같이 개체명 표지는 일반적으로 개체명 경계를 나타내는 표지(예: B(경계 시작), I(경계 내부), E(경계 끝), O(경계 외부), S(경계 시작 및 끝) 등)와 범주를 나타내는 표지(예: PER(인명), LOC(지명), ORG(기관명) 등)가 결합된 형태를 취한다. [그림 6]에서 보는 것과 같이 딥뉴럴넷 기반 개체명 인식 모델의 입력은 각 단어의 임베딩 벡터(Word2Vec과 같은 것으로 학습된 n 차원 벡터)가 사용되며, 최근에는 성능 향상을 위해서 품사 분포나 사전 참조 정보(gazetteer look-up information) 등을 추가적으로 사용하기도 한다.

엔터티 링킹 모델은 사전 기반 방법, 알고리즘 기반 방법, 기계학습 기반 방법 등이 존재한다. 사전 기반 방법은 각 엔터티에 연결되는 다양한 이형태들을 사전에 등록한 후 단순 검색을 통해서 링킹을 하는 것으로 구현하기 쉽다는 장점이 있다. 그러나 모든 이형태를 사전에 등록하는 것이 불가능하기 때문에 음절 수준의 편집 거리(edit distance)나 발음 유사성 계산을

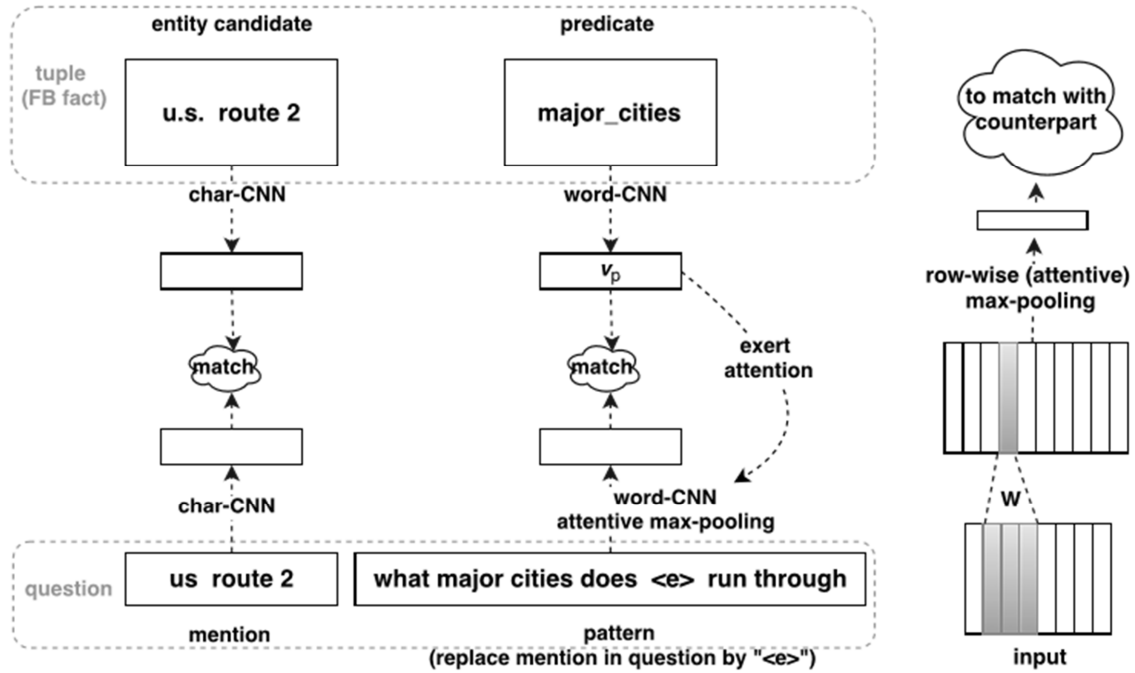


그림 7 문자 단위 CNN을 이용한 단어 추상화와 엔터티 링크 방법[16]

위한 Soundex 알고리즘을 활용하는 방법이 제안되었다. 최근에는 [그림 7]과 같이 음절 단위 임베딩을 통해 단어들을 추상화하고 이들 사이의 유사도 비교를 통해 이형태를 찾는 기계학습 기반 방법들이 주로 사용된다[16].

2.3 담화 분석 및 대화 모델

음성언어 비서 시스템은 대화를 통해 사용자의 의도를 파악하고 영역 지식을 활용하여 적절한 응답을 생성하는 대화시스템의 일종이다. 대화시스템이 적절한 응답을 생성하기 위해서 화자의 의도를 파악하는 과정을 담화 분석(discourse analysis)이라고 하며, 영역 지식을 활용하여 어떤 응답을 생성할지를 결정하는 과정을 대화 모델링(dialogue modeling)이라고 한다.

특정 영역의 대화에서 발화에 포함된 사용자의 의도를 dialogue act라고 하며, dialogue act는 영역 독립적으로 정의된 발화 행위를 나타내는 화행(speech act)과 특정 영역에 의존적으로 정의된 세부적 발화 의미를 나타내는 서술자(predicator; main act 또는 concept sequence라고도 함)로 구성된다. 즉, 화행은 ‘ask-ref(WH-질문), ask-if(YN-질문), response(대답), inform(정보제공), accept(호응)’ 등과 같이 영역 독립적으로 정의된 의미 범주를 말하며, 서술자는 ‘like(좋아하다), hate(싫어하다), set-alarm(알람을 맞추다)’ 등과 같이 대화 영역에 맞춰 정의된 서술어의 의미 범주를 말한다. 담화 분석은 수식 (1)과 같이 화자의 발

화에 내포된 dialogue act를 찾아내는 과정이라고 말할 수 있다.

$$DA(D) \stackrel{def}{=} \arg \max_{S_{1,n}, P_{1,n}} P(S_{1,n}, P_{1,n} | U_{1,n}) \quad (1)$$

수식 (1)에서 $U_{1,n}$, $S_{1,n}$, $P_{1,n}$ 은 n 개의 발화로 이루어진 대화에서의 발화열, 화행열, 서술자열을 각각 의미한다. 수식 (1)은 체인 규칙(chain rule)에 의해서 수식 (2)와 같이 재정의될 수 있다.

$$DA(D) \stackrel{def}{=} \arg \max_{S_{1,n}, P_{1,n}} P(S_{1,n} | U_{1,n}) P(P_{1,n} | U_{1,n}, S_{1,n}) \quad (2)$$

수식 (2)에 1차 마코프 가정(Markov assumption)과 독립 가정(independent assumption)을 적용하면 수식 (3)이 된다.

$$DA(D) \stackrel{def}{\approx} \arg \max_{S_{1,n}, P_{1,n}} \prod_{i=1}^n P(S_i | U_i) P(S_i | S_{i-1}) P(P_i | U_i, S_i) P(P_i | P_{i-1}) \quad (3)$$

담화 분석에 관한 기존의 많은 연구들은 수식 (3)을 기반으로 하고 있으며, 최적의 화행과 서술자를 찾기 위해서 MEMM(Maximum Entropy Markov Model)이나 CRFs와 같은 통계 기반 기계학습 모델들을 사용되어 왔다[17,18]. 최근에는 CNN이나 LSTM과 같은

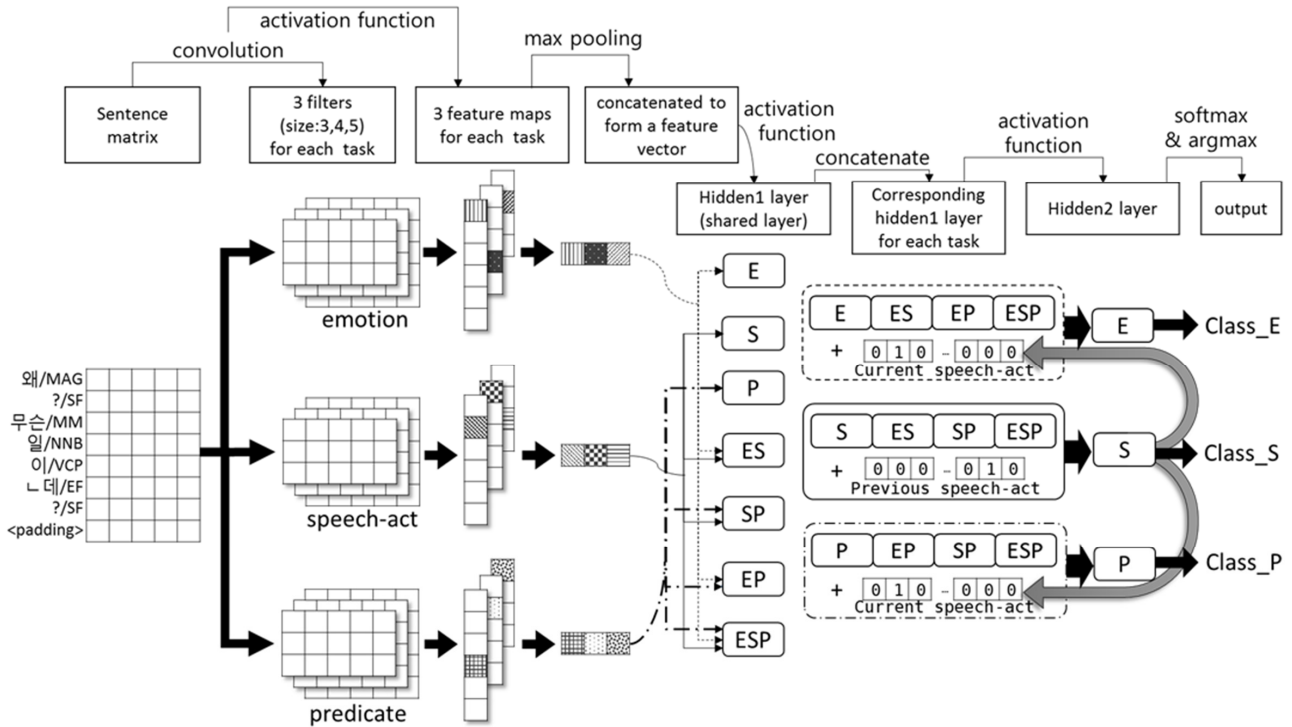


그림 8 CNN을 이용한 화행, 서술자, 감정 통합 분석 모델

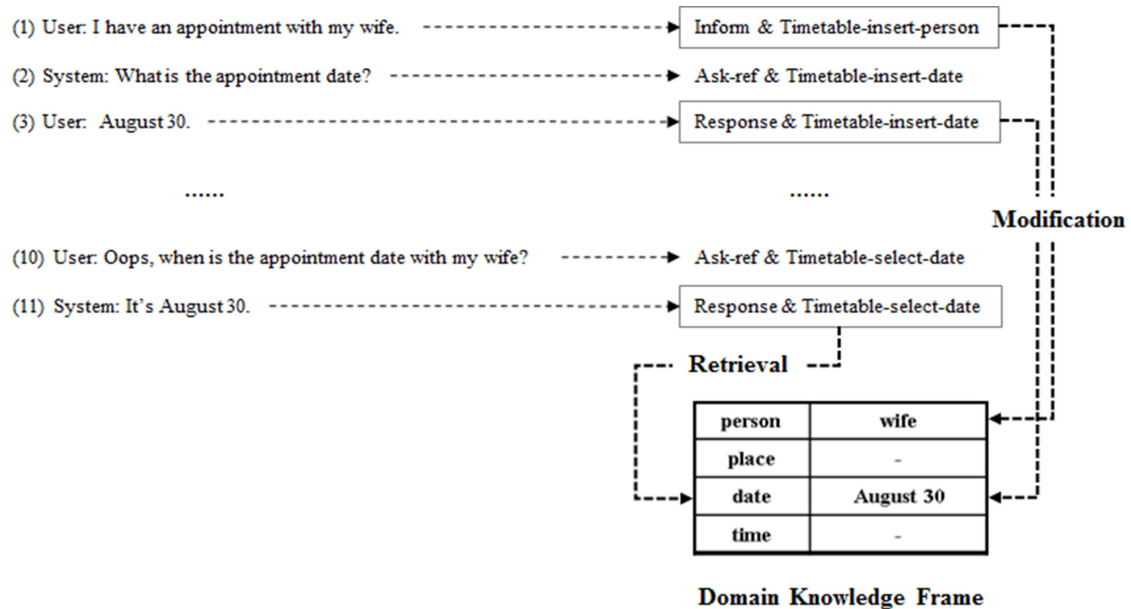


그림 9 프레임 기반 영역 지식 모델

답뉴럴넷을 이용하여 수식 (3)을 구현한 모델들이 제안되고 있다. [그림 8]은 CNN을 이용하여 화행, 서술자, 감정을 동시에 분석하는 모델의 한 예이다[17,18].

[그림 8]에 제시된 모델은 계산을 단순화시키기 위해서 수식 (6)의 $P(P_i|U_i, S_i)$ 를 $P(P_i|U_i)$ 로 변형하지 않고, 상호 의존적인 부분을 공유층(shared layer)를 통해 결합함으로써 성능 향상을 꾀하였다.

대화 모델(dialogue model)은 영역 지식을 관리하는 영역 지식 모델(domain knowledge model)과 대화 흐름을 제어하는 담화 모델(discourse model)로 구성된다. 영역 지식 모델은 시스템이 사용자에게 응답을 제공하기 위해서 참조해야 하는 지식을 관리하는 역할을 하며, 플랜 기반 모델(plan-based model; domain plan model)과 프레임 기반 모델(frame-based model) 등이 대표적이다.

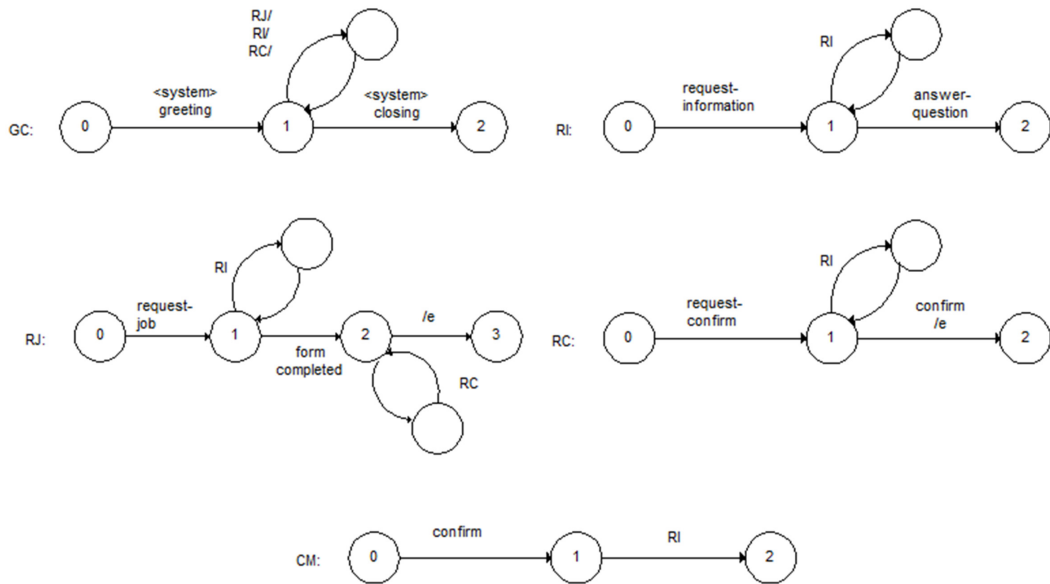


그림 10 대화 상태 전이망의 예

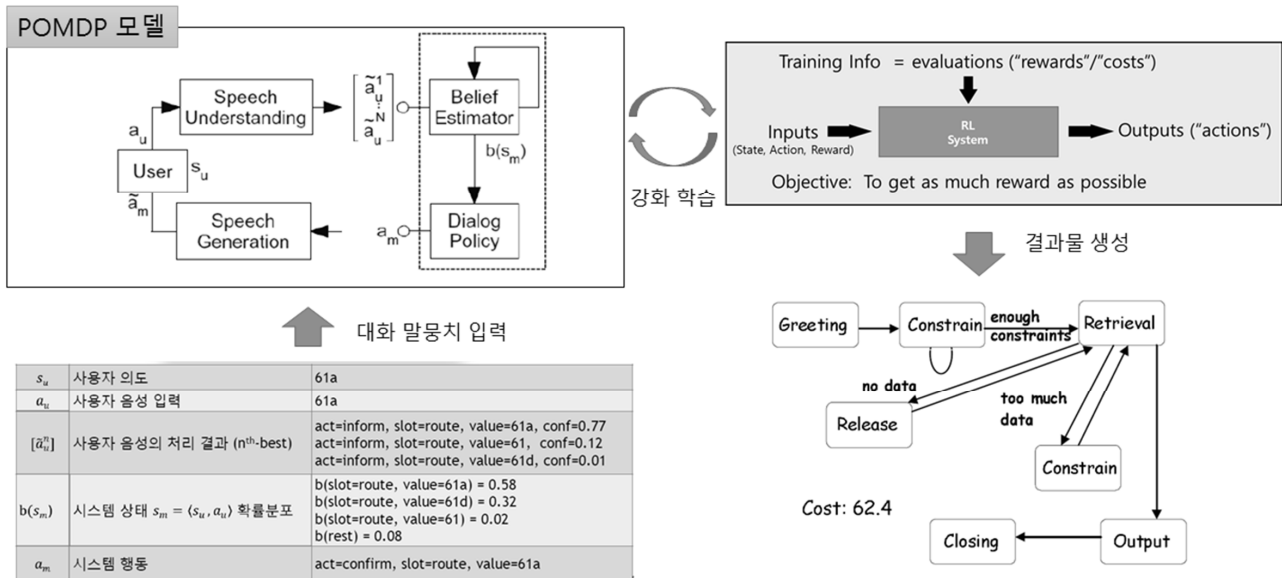


그림 11 POMDP 기반 대화 모델

상용 대화 시스템에서는 유지 보수의 어려움 때문에 플랜 기반 모델은 많이 사용되지 않으며 주로 [그림 9]와 같은 프레임 기반 모델이 사용된다.

[그림 9]에서 보는 것과 같이 프레임 기반 모델에서 시스템과 사용자가 공유하는 영역 지식은 프레임 형태로 표현되며, 시스템은 프레임의 각 슬롯(slot)에 값을 삽입, 삭제, 변경, 검색하면서 응답에 필요한 지식을 관리한다. 대화 모델은 사용자의 의도와 영역 지식을 참고하여 시스템의 의도를 결정하는 역할을 하며, 플랜 기반 모델(plan-based model; discourse plan model), 상태 전이망 모델(state transition network model), 확률 기

반 모델 등이 대표적이다. 상태 전이망 모델은 [그림 10]과 같이 미리 정의된 전이망을 참조하여 시스템의 의도를 결정적으로 선택하는 모델이다.

이러한 상태 전이망 모델은 구현이 용이하다는 장점이 있지만 예측치 못한 사용자의 의도에 유연하게 대처하지 못한다는 단점이 있다. 이러한 문제를 해결하기 위해서 최근에는 대용량의 대화 말뭉치로부터 강화학습(reinforcement learning)을 통해 시스템의 대화 상태를 확률적으로 결정하는 POMDP(Partially Observable Markov Decision Process)나 NNMDP(Neural Network와 같은 모델들이 연구되고 있다. [그림 11]은 POMDP 기반 통계적 대화

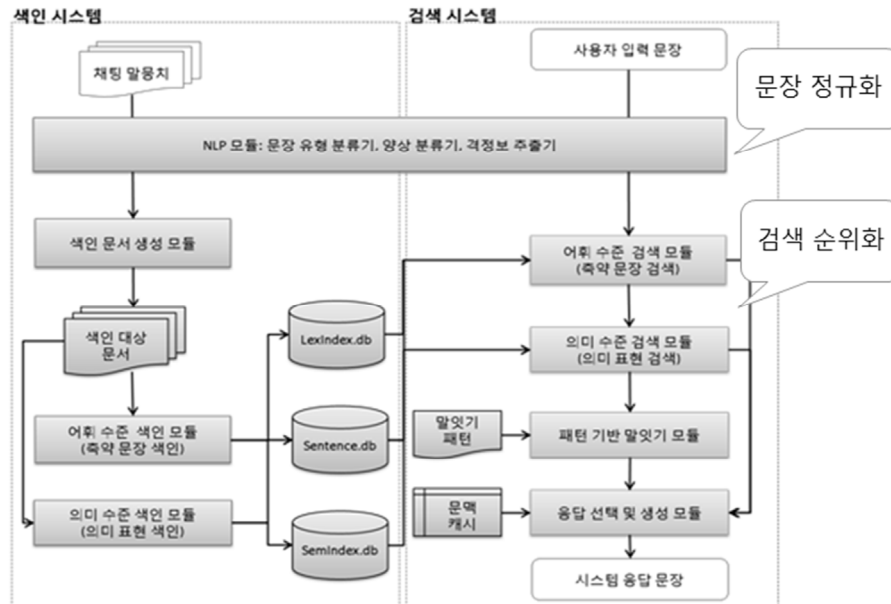


그림 12 검색 기반 채팅 모델 구조도

모델링 과정을 개념적으로 표현한 것이다[19,20].

[그림 11]과 같은 확률 기반 모델은 음성 인식 오류를 포함한 예기치 않은 사용자 입력에 대해서 부분적으로 관찰된 정보를 바탕으로 시스템의 의도를 결정할 수 있다는 장점이 있지만 학습을 위해서 매우 많은 담화 정보 부착 말뭉치를 필요로 한다는 문제를 안고 있다. 예를 들어, [그림 11]의 결과물로 제시된 상태 전이 모델은 약 71만개의 항공 예약 관련 대화 말뭉치로부터 학습된 것이다.

2.4 채팅 모델

채팅 모델은 크게 검색 기반 방법과 생성 기반 방법으로 나뉜다[21]. 검색 기반 채팅 모델은 사용자가 입력한 질의를 미리 수집해 놓은 질의-응답 데이터베이스에 있는 질의와 비교하여 가장 유사한 질의를 찾고, 찾아진 질의와 연결되어 있는 응답을 시스템의 답변으로 제시하는 방법이다[22]. [그림 12]는 검색 기반 방법의 일반적인 처리 과정을 보여준다.

[그림 12]에서 보듯이 검색 기반 방법에서 가장 핵심이 되는 것은 문장 정규화와 검색 순위화이다. 문장 정규화 기술은 검색의 커버리지(coverage)를 높이기 위해서 형태적으로나 구문적으로 다소 상이한 문장이라도 동일한 색인어로 변환하는 기술을 말한다. 예를 들어, “(1) 무한도전 좋아하니? (2) 무한 도전 좋아하니? (3) “무도 좋아해? (4) 무한도전 좋아해.”라는 4개의 문장이 있을 때, 체언(명사류)과 용언(동사류)을 색인어로 추출하는 일반적인 방법을 사용하면 “(1) [무한도전 좋아하], (2) [무한 도전 좋아하], (3) [무도 좋

아하], (4) [무한도전 좋아하]”가 색인어 집합으로 추출된다. 이렇게 추출된 색인어를 바탕으로 검색을 수행하면 질문에 해당하는 (1)번과 응답에 해당하는 (4)번이 가장 높은 순위가 되며, 동일한 의미인 (2)번과 (4)번은 ‘좋아하’라는 용언만 매칭(matching)되게 되어 낮은 순위에 머무르게 된다. 이러한 문제를 해결하기 위해서는 다음과 같은 정규화 처리가 필요하다.

- (1) 개체명 인식이나 명사구 추출을 통한 색인 단위 결정: ‘무한도전’, ‘무한 도전’을 하나의 색인 단위로 묶어주는 과정이 필요하다.
- (2) 유의어 사전이나 온톨로지를 이용하여 동일 의미의 색인 단위에 의미 표지 부여: 유의어에 해당하는 ‘무한도전’, ‘무한 도전’, ‘무도’의 대표어를 찾고 ‘@무한도전’과 같은 정규화된 의미 표지를 부착하는 과정이 필요하다.
- (3) 문형(sentence type), 시제(tense), 양상(modal) 분석을 통한 보조 색인 단위 추출: 체언과 용언이 동일하더라도 문형(평서문, 명령문, 의문문 등), 시제(현재, 과거, 미래 등), 양상(추측, 가능, 확신, 희망 등)[23]에 따라 문장의 의미가 달라지므로 이러한 정보를 색인화하는 과정이 필요하다. 상기 예에서 (1), (2), (3)번 문장에는 ‘#의문문’과 같은 색인어가, (4)번 문장에는 ‘#평서문’과 같은 색인어가 추가되어야 한다.

이러한 정규화 과정은 어휘와 품사 태그(tag)가 포함된 어휘-구문 패턴(LSP; Lexico-Syntactic Pattern) 매칭을 통해서 이루어진다. 상기한 예에 정규화를 적용

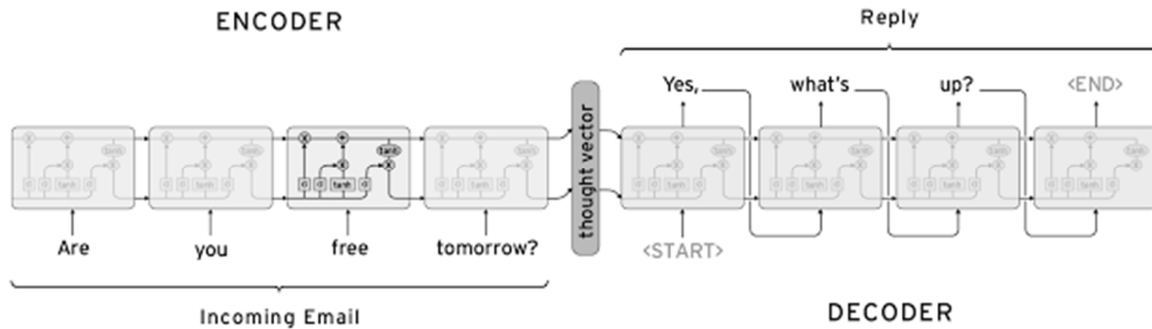


그림 13 Sequence-to-sequence 인공지능 기반 채팅 모델[21]

하면 (1), (2), (3)번 문장은 ‘[@무한도전 좋아하 #의문문]’와 같이 동일한 색인어 집합으로 변환된다. 검색 순위화 기술은 불리언(boolean) 모델에 기반 한 것과 유사도 측정 모델에 기반 한 것으로 나뉜다. 불리언 모델에 기반 한 방법은 색인어들 사이의 불리언 AND 오퍼레이션(operation)을 통해서 매칭된 것을 찾고 그 중에 1개를 임의로 선택한 후, 해당 질의와 연결되어 있는 응답을 결과로 제시한다. 이러한 불리언 모델은 AND 오퍼레이션을 사용하기 때문에 부분 매칭된 비슷한 의미의 문장을 찾을 수 없으므로 재현율(recall rate)이 낮다는 단점이 있다. 그러나 잘못된 응답 결과를 제시할 가능성이 매우 낮아지기 때문에 정확률(precision)이 높다는 장점이 있다. 유사도 측정 모델은 벡터 공간 모델(vector space model), 2-포아송 모델(2-poisson model), 언어 모델(language model)과 같은 랭킹(ranking) 모델을 사용하여 질의-응답 데이터베이스 내의 질의를 순위화하고 1순위 결과의 응답을 제시한다. 이러한 유사도 측정 모델은 불리언 모델과 비교하여 재현율을 높일 수 있지만 유사도에 따라 응답의 신뢰도가 달라진다는 단점이 있다. 이러한 문제를 해결하기 위해서는 유사도의 임계치를 실험을 통해서 세밀하게 설정해야 한다. 최근에는 딥뉴럴넷과 같은 기계학습 모델을 이용하여 검색 결과를 응답으로 내보낼지 여부를 판단하는 분류 기법을 사용하기도 한다[21].

생성 기반 채팅 모델은 검색 기반 채팅 모델이 미리 정의되어 있는 질의에 대해서 정해진 응답만을 선택할 수밖에 없다는 단점을 극복하기 위해서 제시된 방법이다. [그림 13]은 구글(google)에서 제안한 대표적인 생성 기반 채팅 모델을 보여준다.

[그림 13]의 sequence-to-sequence 인공지능망은 원래 기계번역(machine translation)을 위해 개발된 것으로 입력 문장(원시 문장)을 LSTM이나 GRU와 같은 RNN(Recurrent Neural Network)에 넣어서 인코딩

(encoding)하여 추상화하고, 추상화된 결과를 RNN의 입력으로 사용하여 첫 단어부터 마지막 단어까지 하나 하나 디코딩(decoding)(일반적으로 언어모델링이라고 함)하여 출력 문장(번역 문장)을 완성한다. [그림 13]에서 보듯이 구글의 모델은 기존 sequence-to-sequence 모델의 입력을 사용자 질의로, 출력을 시스템의 응답으로 변환한 것이다. [그림 13]에서 사용자의 질의는 단어로 분리되어 sequence-to-sequence 모델의 입력으로 사용되며, 시스템의 응답은 단어들의 열 형태로 출력된다. 이러한 sequence-to-sequence 모델을 한국어에 적용하기 위해서는 다음과 같은 사항들이 고려되어야 한다.

- (1) 입력 단위 결정: 영어의 경우에 일반적으로 어절을 입력 단위로 사용하지만 한국어의 경우에 어절은 여러 개의 단어(형태소)로 구성되기 때문에 어절을 입력단위로 사용할 경우에 영어보다 훨씬 많은 양의 학습데이터가 필요하다. 그러므로 한국어의 경우에 형태소 분석을 수행하여 어절을 형태소로 분리한 후, 형태소와 품사의 쌍을 입력 단위로 사용한다.
- (2) 출력 단위 결정: 영어의 경우에 입력 단위와 마찬가지로 어절을 출력 단위로 사용한다. 한국어의 경우에는 입력 단위에서 나타나는 문제와 마찬가지로 어절을 출력 단위로 사용할 수 없기 때문에 형태소를 출력 단위로 사용한다. 그러나 형태소를 출력 단위로 사용할 경우에 활용과 첨언이 빈번한 한국어의 특성으로 인해 형태소열을 결합하여 어절을 구성하는 합성 과정이 필요하다. 예를 들어, ‘아름답/형용사 + 어/어미’로 출력된 형태소열을 ‘아름다워’로 합성해야 한다. 이러한 번거로움을 해결하기 위해서 활용 형태를 유지하면서 형태소를 분리하고 품사를 부착한 의사 형태소를 출력단위로 사용하기도 한다[24]. 예를 들어, ‘아름다워’에 대한 형태소 분석 결과인 ‘아름답/형용사 + 어/어미’를

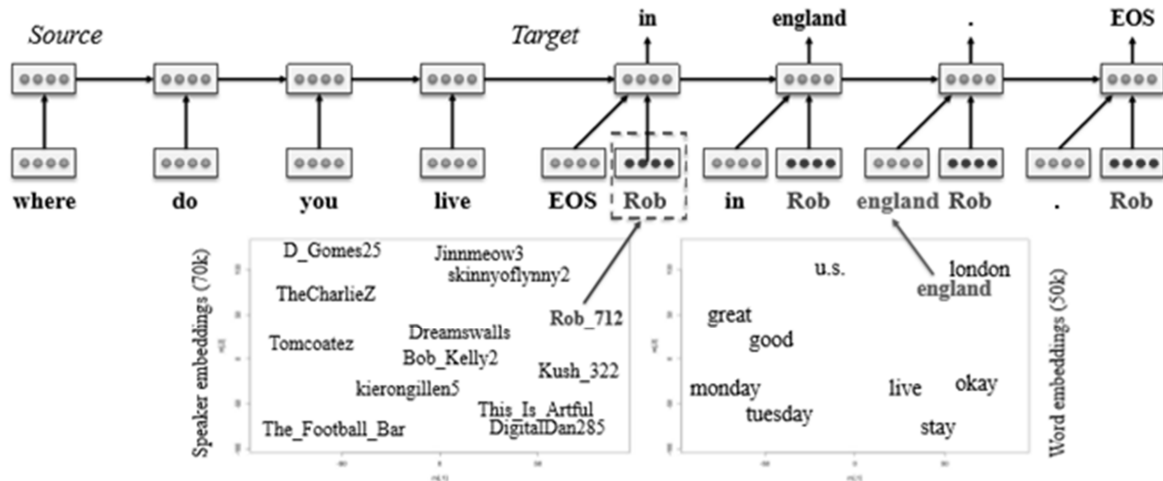


그림 14 Persona를 반영한 생성 기반 채팅 모델

출력 단위로 사용하지 않고 ‘아름다/형용사 + 워/어미’와 같은 의사형태소를 출력 단위로 사용하는 것이다.

생성 기반 채팅 모델은 미리 준비되지 않은 문장을 생성하여 응답을 제시할 수 있다는 장점이 있다. 그러나 예상치 못한 잘못된 응답을 생성하는 경우가 종종 있으며, 학습을 위해서 천만 쌍 이상의 말뭉치가 필요하다는 단점이 있다. 또한 언어모델링에 기초한 sequence-to-sequence 모델의 특성상 짧은 응답을 선호하는 경향성(확률의 곱을 기초로 하기 때문에 짧은 문장이 확률값이 높아지는 경향성)으로 인해 “I don’t know”와 같은 응답 회피 문장이나 “Aha”, “Uhm”과 같은 호응 문장이 생성되는 경우가 많다. 최근에는 sequence-to-sequence 모델의 손실 함수(loss function)을 기존의 크로스엔트로피(cross-entropy)에서 상호정보(mutual information)로 변경함으로써 이러한 경향성을 완화하기 위한 연구가 진행되고 있다[25]. 또한 [그림 14]와 같이 사용자 맞춤형 채팅 문장 생성을 위해서 사용자의 질의 로그(log) 또는 사용자 그룹의 질의 로그를 사용자 정보와 함께 학습에 반영하여 어투와 응답 내용에 변화를 피하고자 하는 연구가 진행 중이다[26].

5. 결 론

본 논문에서는 인공지능 음성언어 비서 시스템 개발에 필요한 핵심 자연어처리 기술인 음성언어 후처리, 개체명 인식 및 엔터티 링킹, 담화 분석 및 대화 모델, 그리고 채팅 모델에 대해서 살펴보았다. 음성언어 후처리 모델을 소개하는 부분에서는 띄어쓰기 오류 보정이 순차 레이블링 모델로, 철자 오류 보정은

노이즈 채널 모델로, 음성인식 결과 재순위화는 그래프 탐색 모델로 다뤄질 수 있음을 제시하였다. 개체명 인식과 담화 분석 모델을 소개하는 부분에서는 대표적인 딥뉴럴넷 기반 기계학습 모델과 최신 연구동향을 살펴보았다. 마지막으로 채팅 모델을 소개하는 부분에서는 그 동안 많이 다뤄지지 못한 검색 기반 채팅 모델의 구현 이슈(issue)들을 살펴보고, 딥뉴럴넷을 이용한 최신 생성 기반 방법들을 소개하였다.

참고문헌

- [1] 조성선, 김용균, “플랫폼으로 진화하는 인공지능(AI) 음성비서”, IITP ICT Spot Issue, Vol. 2, 2017.
- [2] D. Lee, H. Rim, and D. Yook, “Automatic Word Spacing Using Probabilistic Models Based on Character N-grams”, IEEE Intelligent Systems, Vol. 22 (1), pp. 28 - 35, 2007.
- [3] Y. Song and H. Kim, “Automatic Word Spacing of Erroneous Sentences in Mobile Devices with Limited Hardware Resources”, Information Processing and Management, Vol. 47 (4), pp. 518-527, 2011.
- [4] 노형중, 차정원, 이근배, “띄어쓰기 및 철자 오류 동시 교정을 위한 통계적 모델”, 정보과학회논문지: 소프트웨어 및 응용, 34권 제2호, pp. 131-139, 2007.
- [5] S. Jacquemont, F. Jacquenet, and M. Sebban, “Correct your text with Google”, Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence, pp. 170-176, 2007.
- [6] J. Schaback, “Multi-Level Feature Extraction for Spelling Correction”, Proceedings of IJCAI-2007 Workshop on Analytics for Noisy Unstructured Text Data, pp. 78-86, 2007.

- [7] E. Brill and R. C. Moore, "An Improved Error Model for Noisy Channel Spelling Correction", Proceedings of the 38th Annual Meeting of the ACL, pp. 286-293, 2000.
- [8] 강승식, 장두성, "SMS 변형된 문자열의 자동 오류 교정 시스템", 정보과학회논문지: 소프트웨어 및 응용, 제35권 제6호, pp. 386-391, 2008.
- [9] 이세희, 김학수, "교정사전과 신문기사 말뭉치를 이용한 한국어 철자 오류 교정 모델", 정보처리학회 논문지, 16권 5호, pp. 427-434, 2009.
- [10] D. Jurafsky and J. H. Martin, "Spelling Correction and the Noisy Channel", Speech and Language Processing (Draft of November 7), 2016.
- [11] 박진우, 이현구, 김학수, "Sequence-to-sequence 기반 다중 발화 후보를 이용한 형태소 분석기", 한국컴퓨터종합학술대회(발표예정), 2017.
- [12] C. Lee, Y. Hwang, H. Oh, S. Lim, J. Heo, C. Lee, H. Kim, J. Wang and M. Jang, "Fine-Grained Named Entity Recognition using Conditional Random Fields for Question Answering," Proceedings of AIRS 2006, pp. 581-587, 2006.
- [13] 이창기, 장명길, "Structural SVMs 및 Pegasos 알고리즘을 이용한 한국어 개체명 인식", 인지과학, 21권 4호, pp. 655~667, 2010.
- [14] S. Lee, Y. Song, M. Choi, and Harksoo Kim, "Bagging-Based Active Learning Model for Named Entity Recognition with Distant Supervision", Proceedings of BigComp 2016, pp. 321-324, 2016.
- [15] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural Architectures for Named Entity Recognition", Proceedings of NAACL 2016, 2006.
- [16] W. Yin, M. Yu, B. Xiang, B. Zhou, H. Schütze, "Simple Question Answering by Attentive Convolutional Neural Network", Proceedings of COLING 2016, 2016.
- [17] 김민경, 김학수, "Long Short-Term Memory를 이용한 통합 대화 분석", 제28회 한글 및 한국어 정보처리 학술대회, pp. 119-121, 2016.
- [18] M. Kim and H. Kim, "Integrated Neural Network Model for Identifying Speech Acts, Predicators, and Sentiments of Dialogue Utterances", Kangwon National University, Technical Report (KNU NLP Lab.), 2017.
- [19] D. Kim, J. Choi, K.-E. Kim, J. Lee, and J. Sohn, "Engineering Statistical Dialog State Trackers: A Case Study on DSTC", Proceedings of the SIGDIAL 2013 Conference, pp. 462-466. 2013.
- [20] Y. Jang, J. Ham, B.-J. Lee, Y. Chang, and K.-E. Kim, "Neural Dialog State Tracker for Large Ontologies by Attention Mechanism", Proceedings of IEEE Workshop on Spoken Language Technology, 2016.
- [21] WildML, <http://www.wildml.com>, Visited in 21/05/2017.
- [22] 전원표, 송영길, 김학수, "채팅 시스템 구현을 위한 3단계 문장 검색 방법", 한국마린엔지니어링학회지, 37권, 2호, pp. 205-212, 2013.
- [23] 안동연, "Corpus를 기반으로 하는 한국어 슬어의 양상 생성", KAIST 박사학위논문, 1995.
- [24] 김시형, 김학수, "의사 형태소 단위 채팅 시스템", 제28회 한글 및 한국어 정보처리 학술대회, pp. 263-267, 2016.
- [25] J. Li, M. G., C. Brockett, J. Gao, and B. Dolan, "A Diversity-Promoting Objective Function for Neural Conversation Models", Proceedings of NAACL 2016, 2016.
- [26] J. Li, M. Galley, C. Brockett, G. P. Spithourakis, J. Gao, and B. Dolan, "A Persona-Based Neural Conversation Model", Proceedings of ACL 2016. 2016.

약 력



김 학 수

1996 건국대학교 전자계산학과 졸업(학사)

1998 서강대학교 컴퓨터학과 졸업(석사)

2003 서강대학교 컴퓨터학과 졸업(박사)

2004 University of Massachusetts, Amherst 박사후 연구과정

2005 한국전자통신연구원 선임연구원

2006~2008 강원대학교 컴퓨터정보통신공학전공 전임강사

2008~2012 강원대학교 컴퓨터정보통신공학전공 조교수

2012~2017 강원대학교 컴퓨터정보통신공학전공 부교수

2017~현재 강원대학교 컴퓨터정보통신공학전공 교수

관심 분야: 자연어처리, 대화모델링, 정보검색, 정보추출, 질의응답 시스템

Email : nlpdrkim@kangwon.ac.kr