MovieDic 말뭉치를 이용한 대화 참여 모델의 구성

구상준○, 유환조, 이근배
포항공과대학교
giantpanda@postech.ac.kr, hwanjoyu@postech.ac.kr, gblee@postech.ac.kr

Construction of Dialog Engagement Model using MovieDic Corpus

Sangjun Koo○, Hwanjo Yu, Gary Geunbae Lee
Pohang University of Science and Technology

요 약

다중 화자 대화 시스템에서, 시스템의 입장에서 어느 시점에 참여해야하는지를 아는 것은 중요하다. 이러한 참여 모델을 구축함에 있어서 본 연구에서는 다수의 화자가 대화에 참여하는 영화 대본으로 구축된 MovieDic 말뭉치를 사용하였다. 구축에 필요한 자질로써 의문사, 호칭, 명사, 어휘 등을 사용하였고, 훈련 알고리즘으로는 Maximum Entropy Classifier를 사용하였다. 실험 결과 53.34%의 정확도를 기록하였으며, 맥락 자질의 추가로 정확도 개선을 기대할 수 있다.

주제어: 다중 화자 대화 시스템, MovieDic 말뭉치, 대화 참여 모델

1. 서론

음성 대화 시스템 (Spoken Dialog System, SDS)은 사용자로 하여금 시스템과 말소리로 교신할 수 있는 인터페이스를 의미한다. 음성 대화 시스템은 기존의 문자 기반 사용자 인터페이스 (Character User Interface, CUI)나 그래픽 기반 사용자 인터페이스 (Graphic User interface, GUI)에 익숙하지 않은 사용자들에게 효과적인 대안으로써 각종 어플리케이션에서 활용되고 있다. 음성 대화 시스템에서의 대두되고 있는 과제 중 하나는 다수의 화자의 음성 대화를 처리하는 것이다. 이는 다중 화자 의도 파악, 대화 참여 행동의 인식, 음성 데이터로부터의 패턴 인식 등의 세부 과제로 나뉜다[1, 2].

이 중 대화 참여 행동 인식 모듈은 다수의 화자 중에서 현재 대화 맥락에 해당되는 발화를 포착하는 것을 그 목표로 하는 모듈로, 이 모듈을 통해 다중 화자 시스템은 특정 시점에서 맥락에 관련 있는 특정 화자(대화 참여자)에게 적절한 응답을 형성할 수 있다.

대화 참여 행동 인식 모듈을 훈련함에 있어서 필요한 것은 대화 참여 행동 인식에 대한 자질(Feature)과 그 자질들에 대해서 각 화자가 실제로 대화에 참여하는 시점을 나타낸 라벨(Label)이 표기된 입력 데이터 셋이다. 하지만 실제로 이러한 데이터를 표집 하는 것에는 어려움이 있다. 그 이유로 기존의 대화 시스템이 주로 단일 화자를 대상으로 구현되어 있어서 이러한 대화 시스템으로부터 다중 화자 대화 로그를 채록하기 어렵다는 점을 들 수 있다.

아울러, 기존의 대화 시스템은 목적 기반 대화 시스템으로 다중 화자가 참여하는 채팅(Chatting)시스템의 개발에 적합하지 않다. 이는 목적 기반 대화 시스템에서 표집한 말뭉치의 양상이 획일적이기 때문이다.

본 논문에서는 이 목적을 위한 대화 로그의 대안으로 MovieDic 말뭉치[3]를 이용하여 다중 화자 환경에서의 대화 참여 모델을 구성하는 방법에 대해서 논하고자 한다. MovieDic 말뭉치는 인터넷 영화 대본 저장소 (IMSDB)[4]의 대본을 채록하여, 구성해놓은 것으로 한 시나리오에서 2명이상의 화자가 등장하며, 구어체로 구성이 되어있어서 달성하고자 하는 목표에 부합한다는 강력한 장점을 가지고 있다. MovieDic 말뭉치에 대한 상세한 설명은 2장에서 다루고자 한다.

본 논문에서 다루는 대화 참여 모델의 구성은 세 부분으로 나뉜다: (1) 대본의 화자들에 대해서 임의의 화자를 선택하여 시스템 화자로 치환하는 부분 (2) 시스템 화자의 선행 발화에 있어서의 자질을 추출하는 부분 (3) 자질을 기계학습 하는 부분. 이들 부분에 대한 논의는 3장에서 다루고자 한다.

2. MovieDic 말뭉치

MovieDic 말뭉치는 싱가포르의 A-star 산하의 Human Language Technology Institute for Infocomm Research에서 제작되었다. 기존의 대화 처리 시스템 제작을 위한 말뭉치는 목적 지향적, 특정 도메인 한정적으로 제작되었으며 이는 채팅 대화 시스템의 제작에 적합하지 않았다[3]. MovieDic 말뭉치는 영화의 대본의 구성이 일상적인 대화 양상을 보여준다는 것에 착안하여 구축되었다.

OLEG # I want to document my trip to America.
IMMIGRATION OFFICER # Next. Could I see your documents, please?
EMIL # Yes sir.
IMMIGRATION OFFICER # What is your intended purpose of your visit to the United States?
EMIL # Two weeks holiday.
IMMIGRATION OFFICER # How much money are you carrying with you?
EMIL # I have five-hundred dollars.
IMMIGRATION OFFICER # Can you show me? Sir, no cameras in the FIS area!
IMMIGRATION OFFICER # Is he with you? Are you travelling together?

그림 1. MovieDic 말뭉치 예시. 15 Minutes 시나리오의 한 장면.

말뭉치는 다음과 같이 구성된다(그림 1):

1. 시나리오 번호 (그림 1에서는 생략됨)
2. 영화 제목 (15 Minutes, 그림 1에서는 생략됨)
3. 배역 (그림 1에서의 OLEG, EMIL 등)
4. 대사 (그림 1에서 # 이후 부분, 상대 배역에 대한 호칭은 <name-other>, 자신 배역에 대한 호칭은 <name-self>등으로 치환됨)

MovieDic 말뭉치의 개괄 통계는 다음과 같다[3](표 1).

표 1. MovieDic 말뭉치의 통계 요약

항목	값
표집된 대본 수	911
처리된 대본 수	753
총 대화 시나리오 수	132,229
총 화자 턴 수	764,146
영화당 시나리오 수 평균	175.60
영화당 턴 수 평균	1,014.80
시나리오당 턴 수 평균	5.78

표집된 대본 전체에 대해 대화 참여자 수는 1인에서 7인이었으며(최빈값: 2인), 이중 1인 (독백)의 형태는 본 논문의 처리 과정에서 제외되었다.

3. 대화 참여 모델의 구성과정

논의를 전개함에 앞서서 해결하고자 하는 문제에 대해 간략히 설명한다. 대화 참여 모델은 각 턴의 끝에 있어서 시스템의 참여 여부를 결정하는 모델로, 시스템이 언제 화자들의 대화에 '참여' 할 수 있는지를 결정하는 모델이다. 이는 이전 발화 s_0, s_1, … s_{t-1}가 주어졌을 때 특정 시점 t 에서 시스템의 대화참여 여부 X_t의 조건부 확률을 구하는 것과 같다(수식 1).

$$P(X_t|s_1,s_2,...s_{t-1}) \quad (1)$$

문제 해결을 보다 간략하기 위해서 차후 논의에서 1st order 마르코프 가설 (Markov Assumption)을 적용하고자 한다(수식 2). 비록 연쇄적인 대화의 맥락 양상에서 마르코프 가설을 적용하는 것이 전체 모델의 기술 성능 저하를 가져올 수 있으나, 이 가설의 도입을 통하여 우리는 ' 직전 턴의 발화' 에 대해 한정하여 문제 공간을 축소시킬 수 있으며, 궁극적으로는 계산복잡도를 줄일 수 있다.

$$P(X_t|s_1,s_2,...s_{t-1}) \approx P(X_t|s_{t-1}) \quad (2)$$

3.1. 시스템 화자의 치환

각 시나리오에서 임의의 화자를 선택하여 이를 '시스템 화자' 로 간주할 수 있다 (그림 2). 즉, 대본대로 대화가 진행된다고 가정했을 때, 한 화자의 발화는 어떤 가상의 음성 대화 시스템에 의해서 발화되었다고 가정하는 것으로, 구성 알고리즘의 목표는 이 가상 화자의 대화 참여 모델을 모사하는 것이다.

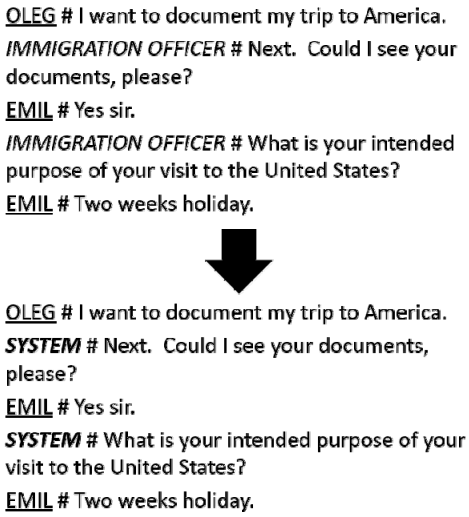

그림 2. 시스템 화자의 치환 예: 화자 Immigration Officer 가 시스템 화자로 선택/치환됨.

3.2. 선행 화자 발화의 자질 추출

앞선 논의에 따라, 대화 참여 모델은 수식 2로 표현되며, 발화 변수 s는 곧 자질 $f_1,f_2,...,f_n$으로 나타난다. 본 연구에서 사용한 자질은 의문사 자질, 호칭 자질, 명사 자질, 어휘 자질로 구성된다 (표 2).

표 2. 화자의 발화에 대한 자질 추출

의문사 자질	설명
W-all	의문사의 존재 여부
W-Why	문장에 Why가 있는지 여부
W-When	문장에 When이 있는지 여부
W-Where	문장에 Where가 있는지 여부
W-Who	문장에 Who가 있는지 여부
W-Which	문장에 Which이 있는지 여부
W-How	문장에 How가 있는지 여부
호칭 자질	설명
Pronoun-all-X	대명사 X의 존재 여부
Pronoun-you	문장에 you가 있는지 여부
Name-other	상대 화자 호칭이 있는지 여부
명사 자질	설명
Noun-all-X	명사 X의 존재 여부
어휘 자질	설명
Bigram-<s1,s2>	Bigram 자질

의문사 자질의 경우, 직전의문문 및 간접 의문문 이후에 화자의 대화 참여가 이루어질 확률이 높다는 직관에서 설계되었다. 호칭 자질의 경우, 상대방을 지칭하는 대명사 및 호칭이 있을 경우 대화 참여가 이루어질 확률이 높다는 점에서 비롯되었으며, 명사나 어휘 자질의 경우,

특정 문구 (예를 들어 aren't you? 등의 부가의문문) 혹은 특정 고유명사 등을 명시했을 때, 그 다음 턴에 대화 참여가 이루어질 수 있다는 점에서 설계되었다.

3.2. 자질을 통한 모델의 학습

모델의 학습을 위해서 선행되어야 하는 것은 결과의 라벨링이다. 전 턴의 화자의 발화에 대해서 다음 턴 시스템이 발화하는 경우를 "참여(Join)", 그렇지 않은 경우를 "대기(Wait)"으로 보았다. 따라서 풀고자 하는 문제는 참여와 대기를 값으로 가지는 확률 변수에 대한 분류 문제로 변환된다.

개별 분류기를 훈련하기 위한 알고리즘은 여러 가지가 있으며, 대표적인 것은 Naive Bayesian Classifier, Support Vector Machine[5]와 Maximum Entropy Classifier[6]을 들 수 있다. 본 논문에서는 Maximum Entropy Classifier를 사용하였다.

4. 실험 결과

실험은 전체 Movie Dic 말뭉치에서 임의로 뽑은 60,000 시나리오에 대해서 훈련 데이터와 실험 데이터를 각각 5:1로 나눈 데이터로 진행되었다. 총 141,190 라벨링 데이터에 대해서 훈련하였으며 자질의 수는 409,007개였다. 총 34,019 라벨링 데이터에 대해서 실제 라벨링과 분류기를 통해 도출된 라벨링을 비교하였다. 실험 결과 Accuracy가 53.34%를 기록하였다 (표 3).

표 3. 분류 실험 결과

	실제 라벨(Join)	실제 라벨(Wait)
분류라벨 (Join)	6,042	7,969
분류라벨 (Wait)	7,903	12,105

5. 결과 분석 및 결론

실험 결과는 무작위로 참여했을 때 정확도인 50%에 비해 6.7%의 향상이 있었다. 다만, 성능 향상폭이 미비하였는데 이는 맥락 자질을 추가하지 않았기 때문으로 분석된다. 대화의 진행에 있어서 대화의 돌출도(Salience)는 대화가 얼마나 지속되는지에 따라 감소하게 되고, 따라서 대화 턴이 넘어가게 되는 효과를 가져오게 된다. 동어 반복 / 대화의 진행 턴 수 등의 한정된 턴 제한에 있어서의 맥락 자질의 추가를 통해 실험 성능의 개선을 기대할 수 있다.

추후 연구로써는 상술한 맥락 자질의 추가와 자질들의 추가/변경/삭제를 통한 성능 추이 분석 및 채팅 대화 말뭉치에서 사용될 수 있는 자질의 개발 등을 들 수 있다.

Acknowledgement

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 Grand ICT연구센터지원사업의 연구결과로 수행되었음 (IITP-2016-R6812-16-0001)

참고문헌

[1] Oriol Vinyals, and Rich Caruana. Learning speaker, addressee and overlap detection models from multimodal streams. Proceedings of the 14th ACM International Conference on Multimodal Interaction. ICMI, pp. 417-424, 2012.

[2] Dan Bohus, and Eric Horvitz. Models for multiparty engagement in open-world dialog, Proceedings of the SIGDIAL 2009 Conference. ACL, 2009.

[3] Banchs, Rafael E. Movie-DiC: a movie dialogue corpus for research and development. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Vol. 2. ACL, 2012.

[4] IMSDB, http://imsdb.com, 2016.8월 접속됨.

[5] Cortes, Corinna, and Vladimir Vapnik. Support-vector networks. Machine learning 20.3: pp. 273-297. 1995.

[6] McCallum, Andrew, Dayne Freitag, and Fernando CN Pereira. Maximum Entropy Markov Models for Information Extraction and Segmentation. Vol. 17. ICML, 2000.