

신경망을 이용한 영역 행위 예측*

이 현 정[†] 김 학 수[†] 서 정 연
서강대학교 강원대학교 서강대학교

목적 지향 대화에서 사용자의 의도는 화행과 개념열의 쌍으로 구성된 영역행위로 표현될 수 있다. 사용자 발화에 대한 영역행위 예측은 음성 인식 오류를 보정하는데 유용하며, 시스템 발화에 대한 영역행위 예측은 유연한 응답 생성에 유용하다. 본 논문에서는 신경망을 이용하여 영역행위를 예측하는 모델을 제안한다. 제안 모델은 대화 이력 벡터와 현재 영역행위를 신경망의 입력으로 사용하여 다음 영역행위를 예측한다. 실험 결과, 제안 모델은 화행 예측과 개념열 예측에서 각각 80.02%, 82.09%의 정확률을 보였다.

주제어 : 영역 행위 예측, 화행 예측, 개념열 예측, 신경망

* 이 연구는 과학기술부 지원으로 수행하는 뇌신경정보학연구사업의 지원을 받았습니다.
또한 김학수의 이 연구는 부분적으로 강원대학교 정보통신연구소의 지원을 받았습니다.

† 제1저자: 이현정, 서강대학교 컴퓨터공학과 자연어처리연구실

E-mail: juvenile@sogang.ac.kr

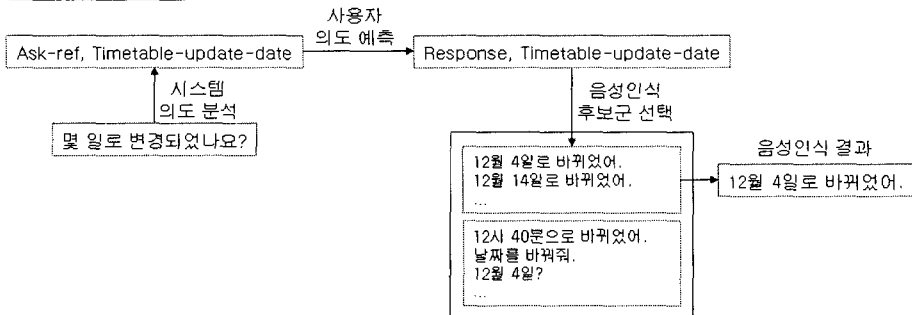
† 교신저자: 김학수, 강원대학교 IT학부(대학) 컴퓨터정보통신공학전공

E-mail: nlpdrkim@kangwon.ac.kr

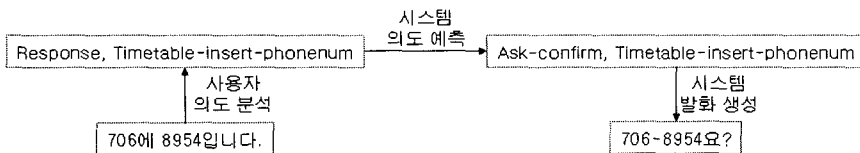
서론

대화시스템은 인간과 컴퓨터가 자연어 대화를 통하여 상호작용할 수 있도록 인간의 언어를 이해하고 적절한 응답을 찾아 제시해 주는 지능형 소프트웨어를 말한다. 이러한 대화시스템을 구현하기 위해서는 발화에 숨겨진 화자의 의도를 찾아내는 의도 분석 기술과 적절한 시스템 응답을 만들어내기 위한 의도 예측 기술이 모두 필요하다. 그러나 화자의 의도와 관련된 기존의 많은 연구들은 추론 규칙 [1][3][8]이나 통계 기반의 기계학습[4][6][10][11][12]을 이용하여 입력된 발화를 특정한 형태의 문자열로 일반화하는 분석기술에만 초점을 맞추어 왔다. 그러므로 의도예측에 관한 연구[9]는 다양한 활용 가능성에도 불구하고 폭넓게 이루어지고 있지 않다. <그림 1>에서 (1)은 사용자 의도예측 연구가 음성 인식 후보들의 탐색 공간을 줄이기 위한 오류 보정용 후처리 엔진으로 활용될 수 있다는 것을 보여준다. 그리고 (2)는 시스템 의도 예측 연구가 유연한 응답 생성을 위한 의미구조 생성 엔진으로 활용될 수 있다는 것을 보여준다.

(1) 사용자 의도 예측 예제



(2) 시스템 의도 예측 예제



<그림 1> 의도 예측 모델의 활용 예

신경망(neural network)은 일반적으로 디지털 컴퓨터가 할 수 있는 모든 것을 모델링할 수 있으며, 특히 규칙들이 쉽게 적용될 수 없는 벡터 공간에서의 매핑(mapping) 문제들을 푸는데 유용하다. 이러한 장점에도 불구하고 입력 형태가 고정된 길이의 벡터로 표현되어야 하고, 벡터의 차수가 너무 클 경우에 학습 시간이 오래 걸린다는 단점 때문에 자연어처리 분야에서 활발히 사용되지 못하였다[5]. 신경망이 가지는 장점을 수용하고 단점을 극복하기 위하여 본 논문에서는 대화 이력(dialogue history)을 짧은 길이의 벡터로 표현한 후, 신경망의 입력으로 사용하여 다음 발화 의도를 예측하는 모델을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 의도 예측에 필요한 자질들을 알아보고 신경망을 이용하여 의도를 예측하는 모델을 제안한다. 3장에서는 실험 데이터를 설명하고 실험 결과를 분석한다. 마지막으로 4장에서 결론을 내린다.

신경망을 이용한 영역 행위 예측

화행과 개념열

목적 지향 대화에서 각 발화에 내포된 화자의 의도는 <표 1>에서 보는 것과

<표 1> 발화와 영역 행위의 예

발화 (U: 사용자, S:시스템)	영역행위 (화행, 개념열)
(1) U: 안녕	Greeting, NULL
(2) S: 무엇을 도와드릴까요?	Opening, NULL
(3) U: 내일 일정 좀 알려줘	Request, Timetable-select
(4) S: 오전11시에 홍길동님과 약속이 있습니다.	Response, Timetable-select
(5) U: 약속이 변경되었어.	Inform, Timetable-update
(6) S: 어떤 내용이 변경되었나요?	Ask-ref, Timetable-update
(7) U: 날짜가 바뀌었어.	Response, Timetable-update-date
(8) S: 언제로 바뀌었나요?	Ask-ref, Timetable-update-date
(9) U: 12월 5일이야.	Response, Timetable-update-date

같이 포괄적인 담화 의도를 나타내는 화행(speech act)과 영역 종속적인 세부 의미를 나타내는 개념열(concept sequence)의 쌍으로 구성된 영역행위(domain action)로 일반화될 수 있다[7].

본 논문에서는 이현정(1997)[13]을 기초로 영역 독립적인 화행을 정의하고, 영역 의존적인 의미들은 개념열을 통하여 보충하는 방법으로 화자의 의도를 일반화한

〈표 2〉 화행의 종류

화행	설명
Greeting	대화 서두의 인사말
Expressive	대화 후미의 인사말
Opening	실제 대화의 시작
Ask-ref	WH-질문
Ask-if	YN-질문
Response	응답
Request	행위를 요청
Ask-confirm	이전 발화의 확인
Confirm	확인 발화의 응답
Inform	정보 제공
Accept	호응

〈표 3〉 개념열의 구성

테이블명	연산자명	필드명
timetable	insert	agent
alarmtable	delete	date
	select	day_of_week
	update	time
		person
		place
		content
		field

다. <표 2>는 본 논문에서 정의하여 사용하는 화행을 보여준다.

<표 3>은 일정관리 영역에서 데이터베이스 연산을 모델링하여 정의한 개념열의 구성을 보여준다. <표 3>과 같이 본 논문에서는 2가지 테이블, 4가지 연산자, 8가지 필드를 3층 구조의 개념열 부착 방법[5]에 따라 53개를 정의하여 사용한다.

영역행위 예측 모델

n 개의 발화로 구성된 대화에서 i 번째까지의 발화가 입력되었을 때, 다음 발화에 대한 영역행위 예측 모델은 화행과 개념열이 서로 독립이라고 가정하면 수식 (1)과 같이 근사된다.

$$P(DA_{i+1}|U_{1,i}) \approx P(SA_{i+1}|U_{1,i})P(CS_{i+1}|U_{1,i}) \quad (1)$$

수식 (1)에서 $U_{1,i}$ 는 대화의 시작에서 i 번째까지의 발화를 의미하며, SA_{i+1} 과 CS_{i+1} 는 각각 $i+1$ 번째 발화에 대한 화행과 개념열을 의미한다.

수식 (1)에서 $U_{1,i}$ 를 직접 계산하는 것은 현실적으로 불가능하다. 그러므로 $U_{1,i}$ 가 i 번째 발화까지의 대화 이력(dialogue history) $DH_{1,i}$ 와 i 번째 영역행위 DA_i 로 일반화될 수 있다고 가정하면 수식 (2)와 같다.

$$P(DA_{i+1}|U_{1,i}) \approx P(SA_{i+1}|DH_{1,i}, DA_i)P(CS_{i+1}|DH_{1,i}, DA_i) \quad (2)$$

수식 (2)에서 대화 이력 $DH_{1,i}$ 는 대화에 참여한 두 화자가 i 번째 발화까지의 상호작용을 통하여 공유한 영역 지식으로 유한상태 모델(finite-state model), 프레임 기반 모델(frame-based model)[2], 계획기반 모델(plan-based model)[8] 등으로 표현될 수 있다. 본 논문에서는 일정 수준의 대화 유연성을 보장하면서 구현하기 쉬운 프레임기반 모델의 영역 프레임(domain frame) 정보를 대화 이력으로 사용한다. 즉, 영역 프레임에 어떤 슬롯(slot)이 채워져 있고, 어떤 슬롯이 비었는가하는 정보를 대화에 참여한 두 화자가 서로 공유하는 지식으로 간주한다.

수식 (2)에서 다음 화행과 개념열이 각각 현재 화행과 개념열에 의해서만 영향을 받는다고 가정하면 수식 (3)과 같다.

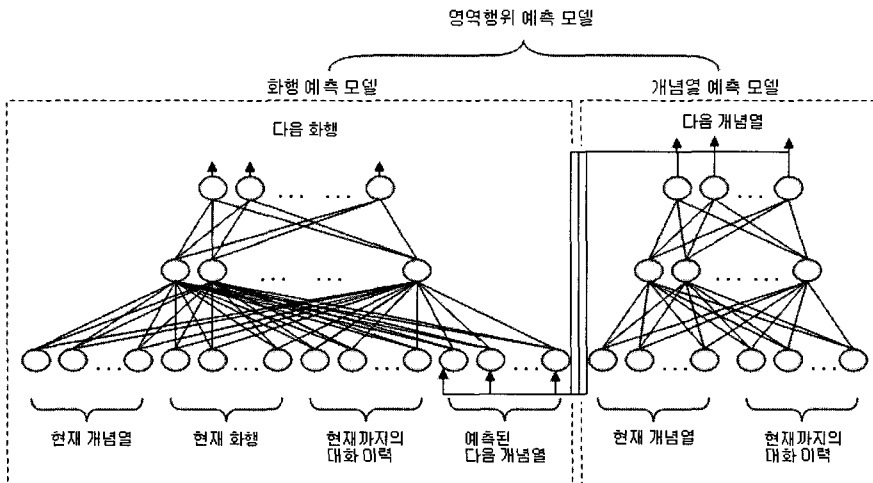
$$P(DA_{i+1}|U_{1,i}) \approx P(SA_{i+1}|DH_{1,i}, SA_i)P(CS_{i+1}|DH_{1,i}, CS_i) \quad (3)$$

수식 (3)에서 개념열 예측 모델 $P(CS_{i+1}|DH_{1,i}, CS_i)$ 를 먼저 적용한 후, 화행 예측 모델을 계산한다고 가정하면 수식 (4)와 같이 예측된 개념열 CS'_{i+1} 를 화행 예측에 사용할 수 있다.

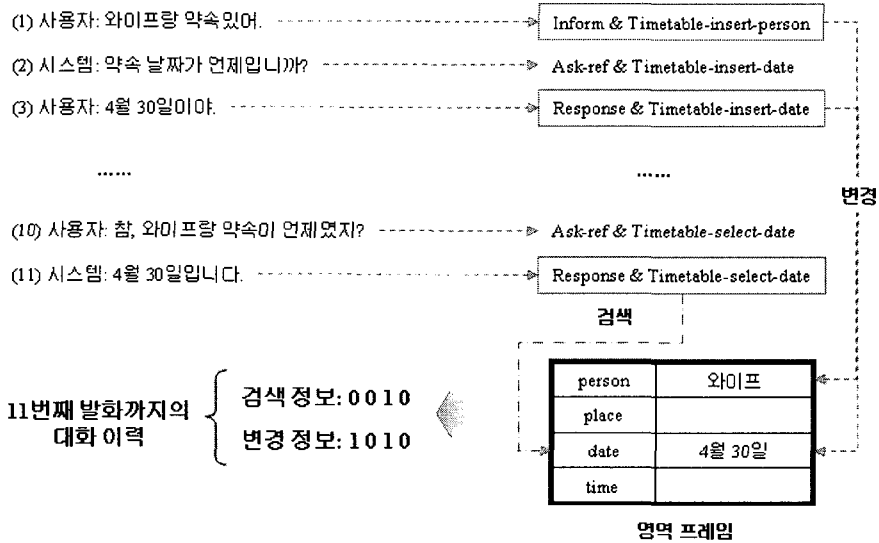
$$P(DA_{i+1}|U_{1,i}) \approx P(CS'_{i+1}|DH_{1,i}, CS_i)P(SA_{i+1}|DH_{1,i}, SA_i, CS'_{i+1}) \quad (4)$$

최종적으로 완성된 영역 행위 예측 모델인 수식 (4)를 다중 퍼셉트론(multi-layer perceptron)으로 표현하면 <그림 2>와 같다.

<그림 2>에서 '현재까지의 대화 이력'은 영역 프레임의 검색 정보와 변경 정보를 바이너리(binary)로 표현한 것이다. 검색 정보는 영역 프레임에서 검색된 슬롯은 '1'로, 그렇지 않은 슬롯은 '0'으로 표현한 벡터이다. 변경 정보는 삽입, 삭제, 변경



<그림 2> 신경망을 이용한 영역행위 예측 모델



<그림 3> 대화 이력 구성 예제

이 이루어진 슬롯은 '1'로, 그렇지 않은 슬롯은 '0'으로 표현한 벡터이다. <그림 3>은 대화가 진행되면서 영역 프레임이 어떻게 변경되는지를 보여준다.

<그림 3>에서 보는 것과 같이 (1) 발화에 의해서 영역 프레임의 'person' 슬롯에는 '와이프'라는 값이 채워진다. 그리고 (3)번 발화에 의해서 'date' 슬롯에 '4월 30일'이라는 값이 채워진다. 마지막으로 (11)번 발화에 의해서는 'date' 슬롯에 채워져 있는 값이 검색된다. 이렇게 11개의 발화가 끝나고 난 후의 대화 이력은 검색 정보 ($person:0, place:0, date:1, time:0$)과 변경 정보 ($person:1, place:0, date:1, time:0$)으로 표현되어 신경망의 입력으로 사용된다.

실 험

실험 데이터

영역행위 예측 모델의 성능을 실험하기 위하여 일정관리 영역에서 Wizard-Of-Oz

〈표 4〉 화행과 개념열이 부착된 말뭉치의 예

태그	값	태그	값
/ID/	4-5	/ID/	4-7
/SP/	User	/SP/	User
/UT/	약속 시간이 몇 시지?	/UT/	장소는 어디야?
/SA/	Ask-ref	/SA/	Ask-ref
/CS/	Timetable-select-time	/CS/	Timetable-select-place
/ID/	4-6	/ID/	4-8
/SP/	System	/SP/	System
/UT/	11시 30분입니다.	/UT/	코엑스홀입니다.
/SA/	Response	/SA/	Response
/CS/	Timetable-select-time	/CS/	Timetable-select-place

방식으로 시뮬레이션(simulation)한 대화 말뭉치를 수집한 후, 수동으로 화행과 개념열을 부착하였다. 수집된 말뭉치는 일정 추가, 삭제, 변경과 관련된 내용을 포함하고 있다. 말뭉치의 구성은 956개의 대화(21,336개의 발화)로 구성되며, 대화당 평균 발화의 수는 22.32개이다. <표 4>는 화행과 개념열이 부착된 말뭉치의 일부를 보여준다.

<표 4>에서 '/ID/'는 발화의 고유 번호로서 값은 대화 번호와 발화 번호로 구성된다. '4-5'는 4번 대화의 5번째 발화를 의미한다. '/SP/'는 발화한 화자의 종류를 의미하며, 값은 'User'와 'System'으로 구성된다. '/UT/'는 실제 발화이고, '/SA/'는 화행, '/CS/'는 개념열을 나타낸다. 화행과 개념열 부착은 자연어처리를 전공하는 5명의 석사과정 학생이 수행하였으며, 일관성을 위하여 1명의 박사과정 학생이 최종 검토하였다.

실험 데이터는 856개의 대화로 이루어진 학습 데이터와 100개의 대화로 이루어진 평가 데이터로 구분하였다. 그리고 학습량에 따른 성능을 평가하기 위하여 학습 데이터를 100개 대화 단위로 8등분하였다. 신경망의 학습은 역전파 알고리즘(back propagation algorithm)을 이용하였으며, 학습율(learning rate)은 0.2, 학습시간은 200 epoch로 설정하였다.

실험 결과

제안 모델의 성능을 상대 평가하기 위하여 동일한 환경(동일한 학습데이터와 평가 데이터)에서 Reithinger(1995)[9]와 비교하였다. 수식 (5)는 Reithinger(1995)에서 제안한 확률식이다.

$$P(SA_n | SA_{n-1}, SA_{n-2}) = q_1 f(SA_n) + q_2 f(SA_n | SA_{n-1}) + q_3 f(SA_n | SA_{n-1}, SA_{n-2}) \quad (4)$$

수식 (5)에서 f 는 해당 화행의 상대 빈도수를 의미하는 기호이며, q 는 합이 1이 되는 가중치 인자이다. 수식 (5)에서 보는 것과 같이 Reithinger(1995)는 화행 예측에 대한 연구만을 진행하였기 때문에 개념열 예측에 대한 비교가 불가능하다. 그러므로 본 논문에서는 수식 (5)를 개념열 예측에 그대로 적용한 모델을 만들고 정확률을 계산하여 비교하였다. <표 5>는 학습량에 따른 시스템 발화에 대한 영역행위 예측 정확률을 보여준다. <표 5>에서 보는 것과 같이 제안한 영역행위 예측 모델은 300개의 대화를 학습 말뭉치로 사용했을 때 가장 높은 정확률을 보였으며, 개념열 예측 정확률이 화행 예측 정확률보다 높았다. 그리고 모든 실험에서 Reithinger

<표 5> 시스템 발화에 대한 영역행위 예측 정확률

학습 말뭉치의 크기(대화)	화행 예측 정확률(%)		개념열 예측 정확률(%)	
	제안 모델	Reithinger(1995)	제안 모델	Reithinger(1995)
100	78.76	51.75	79.68	53.63
200	78.23	61.41	81.29	56.13
300	80.02	42.70	82.09	59.27
400	79.21	42.70	82.01	64.73
500	79.30	42.08	82.01	64.19
600	79.75	42.70	81.74	67.14
700	79.30	42.08	82.18	65.71
856	79.84	42.70	81.65	66.97

(1995)보다 높은 정확률을 보였다.

실험 결과를 분석한 후, 두 가지 사실을 알 수 있었다. 첫 번째, 53개의 범주 중에 1개를 선택하는 개념열 예측 모델의 정확률이 그보다 적은 수인 11개의 범주 중에 1개를 선택하는 화행 예측 모델의 정확률보다 높은 이유는 응용 영역이 가지는 특수성 때문인 것으로 생각된다. 일정 관리 영역과 같은 프레임 기반의 목적 지향 대화에서 대화 참여자의 주요 역할은 자신이 알지 못하는 정보를 얻기 위하여 비어있는 슬롯을 채우는 것이기 때문에 다음 발화의 개념열은 현재까지의 영역 프레임 정보에 매우 많은 영향을 받는다. 그러므로 개념열 예측에 사용된 대화 이력 정보가 다음 개념열을 예측하는데 매우 중요한 단서를 제공한 것으로 보인다. 두 번째, 화행 예측 정확률이 낮은 이유는 개념열 예측 정확률에 영향을 받은 것이었다. 제안된 화행 예측 모델은 입력 자질로써 개념열 예측 결과를 이용하는데, 개념열 예측 결과가 잘못될 경우에 에러가 파급되어 잘못된 화행을 예측하는 경우가 많이 관찰되었다. 입력 자질로써 올바른 개념열 예측 결과를 사용한 경우 95%의 정확률을 얻을 수 있었다.

제안 모델에 대한 실험은 시스템 발화에 대한 영역 행위 예측을 위주로 진행하였다. 그 이유는 제안 모델의 개발 목적이 자연어 대화 시스템을 구현하던 중 시스템 발화의 유연성을 주기 위한 것이었고, 사용자 의도 예측 실험에 적합한 음성 인식 시스템을 찾을 수가 없었기 때문이었다. 향후에는 실제 음성 인식 시스템과 연동하여 사용자 발화에 대한 다양한 형태의 의도 예측 실험을 진행할 예정이다.

결 론

본 논문에서는 음성인식 오류의 보정과 유연한 발화 생성 등에 활용할 수 있는 신경망 기반의 영역행위 예측 모델을 제안하였다. 신경망이 가지는 벡터 공간에서의 매핑 능력을 살리면서 입력 패턴의 형태적 제한에 따른 단점을 극복하기 위하여, 제안 모델은 영역 프레임과 영역 행위를 이용하여 대화 이력과 발화 내용을 근사하였다. 실험 결과, 제안 모델은 300개의 대화를 학습 말뭉치로 사용하였을 경우에 화행 예측과 개념열 예측에서 각각 80.02%, 82.09%의 정확률을 보였다. 제안

모델을 사용자 의도 예측에 사용한다면 음성인식의 오류를 보완할 수 있으며, 시스템 의도 예측에 사용한다면 유연한 발화 생성에 도움을 줄 수 있을 것으로 생각된다.

참고문헌

- [1] Caberry, S. (1989), A pragmatics-based approach to ellipsis resolution, *Computational Linguistics*, 15-2, 75-96.
- [2] Goddeau, D., Meng, H., Polifroni, J., Seneff, S. and Busayapongchai, S. (1996), A form-based dialogue manager for spoken language applications, *Proceedings of International Conference on Spoken Language Processing*, pp. 701-704.
- [3] Lambert, L. and Caberry, S. (1991), A tripartite plan-based model of dialogue, *Proceedings of ACL 1991*, 47-54.
- [4] Langley, C. (2002), Analysis for speech translation using grammar-based parsing and automatic classification, *Proceedings of the ACL Student Research Workshop*.
- [5] Lee, H., Kim, H., and Seo, J. (2006), Efficient domain action classification using neural networks, *Lecture Notes in Computer Science*, vol. 4233, pp. 150-158.
- [6] Lee, S. and Seo, J. (2001), An analysis of Korean speech act using hidden Markov model with decision trees, *Proceedings of ICCPOL 2001*, 397-400.
- [7] Levin, L., Langley, C., Lavie, A., Gates, D. (2003), Wallace, D., and Peterson, K., Domain specific speech acts for spoken language translation, *Proceedings of 4th SIGdial Workshop on Discourse and Dialogue*.
- [8] Litman, D. J. and Allen, J. F. (1987), A plan recognition model for subdialogues in conversations, *Cognitive Science*, 11, 163-200.
- [9] Reithinger, N. (1995), Some Experiments in Speech Act Prediction, *Proceedings of Empirical Methods in Discourse Interpretation and Generation(1995 AAAI Symposium)*.
- [10] Samuel, K., Caberry, S., and Vijay-Shanker, K. (1998), Computing dialogue acts from features with transform-based learning, *Proceedings of the AAAI Spring Symposium*,

90-97.

- [11] 김경선, 서정연 (2003), 자질 선택 기법을 이용한 한국어 화행 결정, 한국정보과학회 논문지, 30-3, 278-284.
- [12] 은종민, 이성욱, 서정연 (2005), 지지벡터기계(support vector machines)를 이용한 한국어 화행분석, 한국정보처리학회 논문지, 12B-3, 365-368.
- [13] 이현정 (1997), 한국어 대화체 문장의 화행 분석, 석사학위논문, 서강대학교.

1 차원고접수 : 2007. 3. 13

최종게재승인 : 2007. 5. 22

(Abstract)

Prediction of Domain Action Using a Neural Network

Hyunjung Lee

Harksoo Kim

Jungyun Seo

Sogang University

Kangwon National University

Sogang University

In a goal-oriented dialogue, speakers' intentions can be represented by domain actions that consist of pairs of a speech act and a concept sequence. The domain action prediction of user's utterance is useful to correct some errors that occur in a speech recognition process, and the domain action prediction of system's utterance is useful to generate flexible responses. In this paper, we propose a model to predict a domain action of the next utterance using a neural network. The proposed model predicts the next domain action by using a dialogue history vector and a current domain action as inputs of the neural network. In the experiment, the proposed model showed the precision of 80.02% in speech act prediction and the precision of 82.09% in concept sequence prediction.

Keywords : domain action prediction, speech act prediction, concept sequence prediction, neural network