열차 예약을 위한 POMDP 기반의 대화 관리 시스템

성주원○, 은지현, 김현정, 장두성
KT 미래기술연구소
{ jwsung, jh06, hyunj, dschang }@kt.com

POMDP based Dialogue Management System for Train Reservation Service

Joo Won Sung○ , Jihyun Eun, Hyunjeong Kim, Du-Seong Chang
Future Technology Laboratory, KT

요 약

본 연구에서는 열차 예약 영역에 통계적 대화형 인터페이스를 도입하여 보다 자연스럽고 오류에 강인한 서비스 제공의 가능성을 검토하였다. 훈련용 코퍼스를 기반으로 사용자 및 시스템 행동 유형, 상태 변이 확률을 추출하여 정책을 도출하고, 성능분석용 코퍼스 기반 사용자 모델로 그 성능을 실험하였다. 방대한 시나리오의 반영을 위해 대량의 코퍼스 수집이 필요한 예제 기반 대화 정책, 혹은 인식기에 의한 오류나 노이즈를 고려하지 않음으로써 현실의 불확실성을 자연스럽게 반영하지 못하는 MDP 대화 정책에 비해 POMDP 정책은 효율적이고 빠른 훈련 알고리즘을 지속적으로 개선시켜 나간다면 적은 노력과 비용으로 효율적이고 강인한 대화 서비스의 제공이 가능할 것으로 기대된다.

1. 서 론

대화 시스템은 자연어를 기반으로 사용자와 지속적으로 대화를 주고 받으며 질의에 대한 적절한 정보를 제공하는 등 가장 자연스럽고 효과적으로 사용자와 컴퓨터간의 커뮤니케이션을 제공하는 수단으로 활용될 수 있다. 급변하는 통신 환경과 고객의 요구에 부합하기 위한 다양한 서비스 영역에서 보다 자연스럽고 효과적인 대화시스템을 구현하기 위한 다양한 연구가 진행되고 있다. 특히 예약 서비스 등 문제 해결 영역(problem solving domain)에서는 구체적이고 목적 지향적인 대화를 통해 사용자가 원하는 정보 및 서비스의 제공으로 보다 높은 만족도를 제공할 수 있다.

최근에는 대화 관리 정책의 최적화 문제에서 사용자의 모델을 구축하고 그로부터 사용자의 행동을 시뮬레이션한 후, 이를 부분 관찰 마르코프 의사결정 과정(Partially Observable Markov Decision Process, 이하 POMDP)으로 모델링하여 대화 시스템의 훈련 및 평가에 적용하는 연구가 활발히 진행되고 있다. POMDP 모델은 인식 오류에 강인하고 보다 신뢰도 높은 대화 정책을 제공할 수 있는 기법으로 최근 그에 대한 고찰의 깊이와 범위를 넓히고 있으나, 아직 실제 상용 서비스와 연관된 대용량 도메인에서의 설계와 적용 및 실험은 매우 부족한 실정이다.

본 논문에서는 열차 예약 서비스에 적용하기 위해 POMDP에 기반하여 대화 추론 및 응답을 제공하는 대화관리시스템을 설계하였다. 2장에서는 통계적 기반 대화 추론 방법론 및 효율적인 정책 도출을 위해 본 연구에서 적용한 기법을 간략히 설명한다. 3장에서는 열차 예약 영역에 적용하기 위한 사용자의 행동 유형 및 시스템 응답 모델, 수집된 코퍼스를 활용한 상태 변이 확률 분포의 도출 과정을 설명하고, 언어이해와 대화추론, 응답생성으로 이어지는 열차 예약 대화관리시스템의 전체 구조를 간략히 기술한다. 아울러, 통계적 기법을 적용한 대화관리자의 성능을 측정하기 위해 실시한 실험의 결과를 정리하였다. 마지막으로 4장에서 결론을 도출하고 향후 연구 방향을 논하도록 한다.

2. 통계적(POMDP) 기반 대화 추론

2.1 부분 관찰 마르코프 의사 결정

대화의 문제를 유한 개수의 상태와 그들 간의 천이 확률로 모델링하고 각 상태에 따라서 미리 학습된 정책에 따라 행동을 취하도록 하는 마르코프 의사 결정 모델(MDP)은 현재의 상태가 정확하게 인식되었다고 가정하고 인식기에 의한 오류나 노이즈를 고려하지 않음으로써 현실의 불확실성을 자연스럽게 반영하지 못하고, 실제로 텍스트 기반의 자연어를 이해하거나 음성을 인식하는 과정의 오류 상황을 적절히 제어하지 못하고 대화 관리 성능이 크게 떨어질 수 있는 단점이 있다.

최근 들어, POMDP에 기반한 대화 관리 시스템이 양자화된 신뢰도 구간 대신 연속적인 신뢰도를 환경 변수로 사용하고 최적의 신뢰도 경계를 찾을 수 있는 방식으로 주목받고 있다.

POMDP 모델은 실제 상태를 정확히 알 수 없는 상황에서 관측치(observation)를 통해 실제 상황을 예측하도록 확률적으로 모델링하고 강화학습(reinforcement learning)을 통해 시스템 행동에 대한 보상치를 장기적으로 최대화할 수 있는 정책을 도출하는 방식으로 다음과 같이 {S,A,T,R,O,Z} 모델로 정의될 수 있다.

- S : 대화의 실제 상태(state) s의 집합
- A : 시스템이 취할 수 있는 행동(action) a의 집합
- T : 상태 s의 변이 확률 $P(s'|s,a)$의 분포
- R : 시스템의 행동에 따른 보상값의 기대치 $r(s,a)$의 확률 분포
- O : 관측치 o의 집합

- Z : 관측치의 확률 $P(o'|s',a)$의 분포

현재의 시점 t에서 대화가 특정 상태 s에 있으리라고 예측되는(belief state) 확률을 $b_t(s)$ 라고 할 때, 현재 시점에서 시스템 행동 a를 취함으로써 얻을 수 있는 즉각적인 보상의 기댓값은 다음으로 산출된다.

$$\rho(b,a_t)=\sum_{s\in S} b_t(s)r(s,a_t)$$

대화추론 정책 훈련의 궁극적인 목적은 다음의 식으로 표현되는 시간의 흐름에 따른 누적 보상을 최대화 할 수 있도록 최적의 시스템 행동을 도출하는 것이다.

$$\sum_{t=0}^{\infty}\gamma^t\rho(b,a_t)=\sum_{t=0}^{\infty}\gamma^t\sum_{s\in S} b_t(s)r(s,a_t)$$

여기서 γ 는 할인 상수(discount factor)로서 미래에 받게 될 보상을 현재 상태의 가치로 조정해 준다.($0\le\gamma\le 1$)

POMDP는 음성 인식의 오류나 자연어 처리의 오류에 강인하고, 사용자 의도 분석 문제를 자연스럽게 모델링할 수 있으며, 대화 플로우를 대화 설계자가 절차적으로 기술할 필요가 없다는 장점이 있다. 가상의 유저를 통한 훈련 에피소드를 다량으로 생성하는 방법으로 대화관리 시스템(특히 대량의 코퍼스를 필요로 하는 예제 기반 대화 관리자)이 일반적으로 겪게 되는 수집된 코퍼스의 절대적인 부족을 보완할 수 있고, 실제 코퍼스에 포함되지 않은 가능한 모든 경우의 공간을 탐색하고 훈련할 수 있도록 하여 주어진 대화 정책을 개선시키고 보다 우수한 정책을 찾아낼 수 있다.

2.2 정책 도출 기법 연구

그러나, 실용적인 측면에서 볼 때 POMDP 모델은 대화 관리 문제를 직관적으로 표현할 수 있음에도 불구하고, 모델을 기반으로 최적의 대화 관리 정책을 계산하고자 할 때의 복잡도가 매우 높으며, 특히 대용량의 대화 도메인에 적용할 때 빠르고 효율적인 알고리즘이 필수적으로 요구된다.

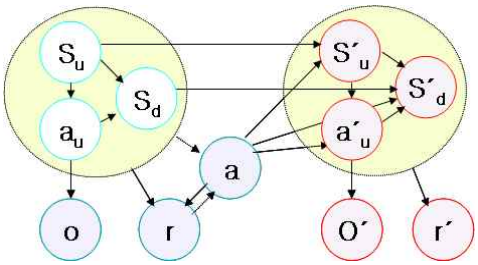
보다 직관적으로 POMDP 문제를 표현하고 풀기 위해 대화의 진행 상태(state)를 세부 구성 요소로 나누고 대수결정도(ADD: Algebraic Decision Diagram)를 활용하여 밀집된 구조로 모델링하는 기법(Factored POMDP)과 PBVI(Point-based Value Iteration) 방식이 제시되었다[5].

최근에는 방대한 문제의 상태 공간을 효율적으로 탐색해 나가기 위해, 상위임계치 및 하위임계치를 모두 계산하여 해를 개선해 나가는 휴리스틱한 근사화 기법(Hueristic Search Value Iteration)이 다양하게 연구되고 있다[1][6][7].

상태 s 를 유저의 목표(s_u),유저의 발화(a_u),대화의 현재 상태(s_d)의 구성요소로 팩토링할 경우, 위의 목적함수의 인자인 믿음 상태 $b(s)$는 대화가 진행됨에 따라 다음과 같이 갱신된다.(k : normalization factor)

$$b'(s_u',s_d',a_u')=k\bullet p(o'|a_u')p(a_u'|s_u',a)\bullet\sum_{s_u\in S_u}p(s_u'|s_u,a)\bullet\sum_{s_d\in S_d}p(s_d'|a_u',s_d,a)\bullet\sum_{a_u\in A_u}b(s_u,s_d,a_u)$$

본 연구에서는 반복적으로 상태 공간에서 최적해를 찾아 이동하는 과정에서 모든 상태와 관측치로의 변이를 명시적으로 계산하는 대신 상위임계치(upper bound)와 하위임계치(lower bound)를 산출하여 해를 개선시켜 나가는 기법을 활용하여 현실적인 대용량 영역인 열차 예약 서비스의 정책 도출 문제에 적용하였다[1].



[그림1] POMDP state transition

3. 열차 예약 대화관리시스템 설계

통계적 기반 열차 예약 대화관리시스템을 설계하기 위하여, 가상의 유저의 행동과 시스템의 응답 유형, 그 변이 확률을 수집된 코퍼스로부터 추출하여 상징적 언어로 모델링하였다. 시스템의 행동 유형별로 적절한 보상 값을 부여한 후 궁극적 보상을 최대화할 수 있는 정책을 훈련시켰으며, 그 결과를 성능평가용 코퍼스를 활용하여 성능을 분석하였다.

3.1 유저 모델링

열차 예약 서비스에 대한 실제 사용자와 시스템간의 대화를 가정하고 WoZ(Wizard-of-Oz)방식으로 수집된 대화 코퍼스에서 순수하게 "예약" 을 주요 목적으로 하는 코퍼스 2179 쌍을 추출하여 1918쌍(약 88%)은 훈련용으로 261쌍(약 12%)은 테스트용으로 활용하였다. 열차의 예약에 필요한 필수 슬롯은 출발지,목적지 등 8개로 가정하였으며, 이들 슬롯이 채워졌는지의 여부를 그 상태 값으로 가진다. 전체 대화의 상태를 표현하는 변수 turn 은 목적 달성에 이르기 위한 현재의 진행 상태 정도를 나타낸다.

[표1] 상태 변수(Internal variable)

변수	상태값
출발지, 목적지, 출발일, 출발시각, 열차유형, 객실유형, 승객 유형, 예약 매수	u(unknown) k(known)

전체대화상태(turn)	n(unknown) v(not confirmed) c(confirmed) r(reserved)

사용자의 발화(User Act)의 유형은 훈련용 코퍼스로부터 추출된 64개로 분류하였다. 발화의 형태는 화행(Speech Act)과 발화에 포함된 해당 슬롯의 조합으로 표현되며, 관측치(observation) 변수는 실제 사용자 발화 유형에 어느 상태로도 인식되지 못한 오류 상태(error)를 추가하였다.

[표2] 사용자 발화(User Act) 유형

유형분류	유형(총 64)
Reserve	50
Cancel	10
Agree	1
Reject	1
Repeat	1
Bye	1

3.2 시스템 응답 설계

시스템의 응답은 Greet, 사용자로부터 아직 인식되지 못한 슬롯 정보를 요구하는 Specify류, 인식된 슬롯을 확인하는 ConfirmD, 최종 예약 여부를 다시 확인하는 ConfirmQ, 실제 철도청 예약 서비스를 실행하는 Operate 등의 시스템화행과 해당 슬롯의 조합으로 25개의 응답을 표현하였다.

[표3] 시스템 응답(System Act) 유형

유형분류	유형 (총 25)
Greet	1
Specify	21
ConfirmD	1
ConfirmQ	1
Operate	1

각 행동 유형별로 유저 측면의 만족도 및 효율적으로 목적 지향적 대화를 주도할 수 있는지의 여부에 따라 적절한 보상 값을 부여하였다.

[그림2]는 Specify류의 시스템 응답 a에 대해 보상값을 부여하는 절차를 기술한 것으로, 시스템 응답 a가 사용자 입장에서는 이미 인식되었다고 생각되는(known) 슬롯에 대한 정보를 요구할 때에는 +10 의 패널티를 주고, 인식되지 못한(unknown) 슬롯에 대한 정보 요구 시 적절한 행동으로 보아 -10 의 보상을 주도록 한다. 또한 열차 유형이 정해져야 가능한 좌석 유형을 알 수 있는 경우 등, 슬롯 별로 채워지는 우선 순위를 부여하기 위해 a에서 요청하는 슬롯 보다 우선 순위가 높은 슬롯이 하나라도 덜 채워졌을 때에는 +2의 패널티를 주었다.

ConfirmD와 ConfirmQ는 하나라도 정보가 채워지지 않은 상태에서 사용자의 확인을 요청하는 경우 최대의 패널티 +100 을 주고, 전체 대화 상태(turn)의 미확인, 확인, 예약완료일 각각의 경우에 적합하게 보상을 추가로 받는다. [그림3]은 ConfirmD 에 대한 보상 값의 부여 과정을 기술한 것으로, ConfirmD 는 모든 슬롯이 인식되고 전체 대화의 상태가 unknown(n) 상태일 때 가장 보상을 많이 받을 수 있도록 한다.

규칙에 어긋나는 값을 발화하거나 잔여석 조회 결과의 부재, 예약 과정에서의 오류 상황 등에 대한 시스템의 응답은 통계적 추론을 거치지 않고 규칙에 기반한(Rule-based) 응답 추론을 함께 활용하도록 하여 대화의 효율성과 품질의 제고를 도모하였다.

```
Function reward_SPECIFY(a)
  cost := defaultreward
  PRI := min_{s∈a} priority(s)
  for i:=1 to coreslot# do
   if  s_i ∉ a then
    if state(s_i) = unknown and priority(s_i) >= PRI
        cost += 2
   else then
      if state(s_i) = known  cost += 10
      else cost -= 10
   end if
  end for
  return cost
end Function
```

[그림2] 시스템 응답별 보상 모델링(Specify류)

```
Function reward_CONFIRMD
  cost := defaultreward
  for i:=1 to coreslot# do
    if state(s_i) = unknown cost += 100
    else cost -= 10
  end for
  switch turn
      case n : cost -= 10
      case v : cost += 20
      case c : cost += 20
      case r : cost += 20
  end switch
  return cost
end Function
```

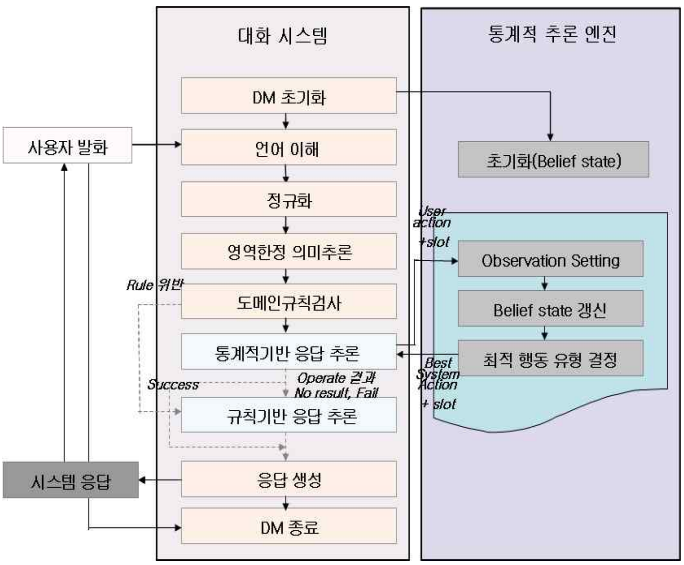
[그림3] 시스템 응답별 보상 모델링(ConfirmD)

3.3 확률 변이

대화의 초기 상태(initial belief) 확률 분포, 현재 상태와 주어진 시스템 응답에 따른 다음 상태로의 변이 확률(transition probability)은 주어진 훈련용 코퍼스로부터 추출하여 적용하였으며, 할인상수 (discount factor)는 0.95 를 적용하였다.
언어 이해 모듈에 의한 평균적 인식 오류율은 0.05 로 가정하였다.

3.4 열차 예약 대화관리시스템 구성

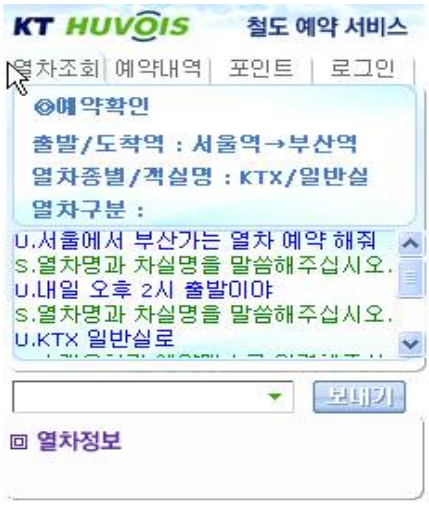
열차 예약을 위한 대화관리시스템은 사용자의 발화를 이해하기 위한 언어이해 모듈 ,대화 추론 모듈, 응답 생성 모듈로 구성된다.



[그림4] 열차 예약 대화관리시스템 구성도

언어 이해 모듈은 조건부랜덤필드(Conditional Random Field)를 적용하여 사용자의 입력에 대한 의미구조를 추출하고, 어휘 정규화를 수행한다. 대화추론 모듈은 일반적인 대화규칙이나, 영역 한정 규칙에 의한 제약 사항을 먼저 확인하고, 제약 위반 발생 시 규칙기반 응답 추론을 따르며, 정상적인 발화에 대해서는 통계적 기반 응답 추론을 수행한다. 관찰된 사용자의 발화에 따라 현재의 믿음 상태 (belief state)를 갱신하고, 미리 학습된 POMDP기반 정책 파일을 참조하여 현재 상태에서의 최적의 시스템 행동 유형을 도출한다.
마지막으로 응답생성모듈에서 추론된 시스템의 행위와 관련된 발화 템플릿으로부터 사용자에게 제공할 자연어 응답을 생성하게 된다. POMDP 모델에서 시스템 행동 유형별 발화 템플릿은 미리 정의된 XML형태로 제공된다.

시간과 공간에 제약받지 않는 열차 예약 서비스의 제공을 위하여 Web 및 PDA 상에서 사용자 인터페이스를 구현하여 사용자 편의의 극대화를 도모하였다.



[그림5] 열차 예약(PDA) 서비스 화면

3.5 성능분석

제안된 통계적 대화관리시스템의 성능을 측정해 보기 위하여 테스트용 대화 코퍼스 261쌍을 활용하여 유저 모델을 구축하고, 훈련에 쓰인 유저 모델과 대비하여 3000 번의 가상 대화 시뮬레이션을 각각 시행하였다.

[표5] 성능 측정 결과

목표차이값	항 목	훈련용 코퍼스	성능분석용 코퍼스
20	대화성공횟수	3000	2997
	평균대화길이	7.72	8.99
	표준편차(보상값)	75	173
200	대화성공횟수	2997	2423
	평균대화길이	9.23	10.86
	표준편차(보상값)	248	384

시뮬레이션 결과 훈련에 쓰인 동일한 유저 모델을 적용할 경우 매우 높은 목표 달성율을 보였으며, 성능분석용 모델에서도 비교적 높은 성공률을 보였다. 상위임계치와 하위임계치간 차이에 대한 목표차이값(goal width)을 적정 수준 이하로 작게 주고 훈련을 시킨 경우, 성능분석용 코퍼스에 대한 성공률이 훈련용 코퍼스와 거의 유사한 성능을 보였다. 즉, 탐색의 과정에서 상하위 임계치의 차이에 대한 목표치를 적게 줄수록 훈련에 소요되는 시간 및 메모리가 많이 소요되지만, 도출된 정책의 성능을 높일 수 있으며, 본 실험의 경우 목표차이값을 20 혹은 그 이하로 두고 훈련을 시킬 경우 훈련에 쓰인 코퍼스뿐만 아니라 테스트용 코퍼스에서도 유사한 성능을 보일 수 있는 강인하고 신뢰성 있는 정책을 얻어낼 수 있었다.
실제 사용자의 행동 유형에 대한 보다 다양하고 정밀한 관찰과, 음성이해모듈의 오류를 포함한 양질의 코퍼스 수집을 통해 지속적으로 모델을 개선하고 정책 훈련을 시행하면 보다 지능화되고 신뢰도 높은 통계적 대화 추론 정책을 얻을 수 있을 것으로 기대된다.

4. 결 론

본 연구에서는 열차 예약 영역에 통계적 대화형 인터페이스를 도입하여 보다 자연스럽고 오류에 강인한 서비스 제공의 가능성을 검토하였다. 훈련용 코퍼스를 기반으로 사용자 및 시스템 행동 유형, 상태 변이 확률을 추출하여 정책을 도출하고, 성능분석용 코퍼스 기반 사용자 모델로 그 성능을 실험하였다. 방대한 시나리오의 반영을 위해 대량의 코퍼스 수집이 필요한 예제 기반 대화 정책, 혹은 인식기에 의한 오류나 노이즈를 고려하지 않음으로써 현실의 불확실성을 자연스럽게 반영하지 못하는 MDP 대화 정책에 비해, POMDP 정책은 효율적이고 빠른 훈련 알고리즘을 지속적으로 개선시켜 나간다면 적은 노력과 비용으로 효율적이고 강인한 대화 서비스의 제공이 가능할 것으로 기대된다.

향후에는 서로 다른 목적 영역간의 상태 공간의 믿음치의 상속과 그들 간 변이 확률을 모델링함으로써 대용량 도메인에서의 POMDP 의 효율적인 적용 방안을 검토하고, 다양한 실험을 통해 도출된 정책의 품질 측정 및 성능 분석을 추진함으로써 대화시스템의 신뢰도 및 안정성을 높이기 위한 방안을 지속적으로 연구해 나갈 예정이다. 특히, 관련된 슬롯의 수와 가능한 값이 일정 수준 이상 증가할 경우 학습에 필요한 메모리와 시간이 기하급수적으로 증가하여 현실적으로 제어하기 힘든 단점을 보완하기 위하여, 훈련 알고리즘의 보완이 진행되고 있다.

또한, 음성이해모듈의 신뢰도 값 자체를 믿음치의 갱신에 직접 적용하여 실제 대화에서의 관측치의 신뢰도를 즉각적으로 반영한 통계적인 상태 변이를 적용하는 작업이 진행 중이다.

열차 예약 시스템에서의 안정화 작업 이후에는 다양한 목적 지향적 영역의 서비스에 확대 적용하여 보다 안정적인고 효율적인 대화관리시스템을 개발하고자 한다.

참고 문헌

[1] Hyeong Seop Sim, Kee-Eung Kim, et al, “Symbolic Heuristic Search Value Iteration for Factored POMDPs“, AAAI, p1008~1093, 2008

[2] 최준기, 은지현, 장두성, 김현정, 구명완, “마르코프 의사결정 과정에 기반한 대화 관리자 설계”, 대한음성과학회 추계학술대회 논문집, 2006

[3] 은지현, 최준기, 장두성, 김현정, 구명완, “마르코프 의사결정 과정에 기반한 대화 관리 시스템”, HCI, p475~480, 2007

[4] 장두성, “담화인터페이스개요”, 음향학회, Tutorial, 2007

[5] Williams J.D,et al, “Factored Partially Observable Markov Decision Processes“, In proceedings of IJCAI-2005 Workshop on Knowledge and Reasoning in Practical Dialogue Systems, 2005

[6] Smith T. and Simmons R, “Point-based POMDP algorithm: Improved analysis and implementation“, In Proceedings of UAI, 2005

[7] Jost Schatzmann, et al, “A survey of statistical user simulation techniques for reinforecement learning of dialogue management strategies“, The knowledge Engineering Review Vo.21:2, p97~126, 2006