

## Conditional Random Fields를 이용한 영역 행위 분류 모델\*

김 학 수<sup>†</sup>

강원대학교

목적 지향 대화에서 사용자의 의도는 화행과 개념열의 쌍으로 구성된 영역 행위로 표현될 수 있다. 그러므로 지능적인 대화 시스템을 구성하기 위해서는 영역 행위를 정확히 파악하는 것이 매우 중요하다. 본 논문에서는 CRFs (Conditional Random Fields)를 이용하여 화행과 개념열을 동시에 결정하는 통계 모델을 제안한다. 편향 학습 문제를 피하기 위하여 제안한 모델은 어휘와 품사 같은 낮은 수준의 언어 자질을 입력 자질로 사용하며, 카이 제곱 통계량을 이용하여 불필요한 자질들을 제거한다. 일정 관리 영역에서 실험을 수행한 결과, 제안한 모델은 화행 분류 정확률에서 93.0%, 개념열 분류 정확률에서 90.2%의 좋은 성능을 보였다.

주제어 : CRFs, 영역 행위 분류, 화행 분류, 개념열 분류

---

\* 이 연구(논문)은 산업자원부 지원으로 수행하는 21세기 프론티어 연구개발사업(인간기능 생활지원 지능로봇 기술개발사업)의 일환으로 수행되었습니다. 또한 부분적으로 강원대학교 정보통신연구소의 지원을 받았습니다.

† 교신저자 : 김학수, 강원대학교 IT학부(대학) 컴퓨터정보통신공학전공  
연구세부분야 : 한국어정보처리, E-mail : nldrkim@kangwon.ac.kr

## 서 론

목적 지향 대화는 목적 지향 발화들의 열로 구성되며, 각 발화에 내포된 사용자의 의도는 영역 행위(domain action)라는 의미 형태로 일반화될 수 있다[11]. 그러므로 지능적인 대화 시스템을 구현하기 위해서는 영역 행위를 빠르고 정확하게 분석하는 일이 선행되어야 한다. 영역 행위는 <표 1>에서 보는 것과 같이 발화에 표현된 영역 독립적인 일반 의도를 나타내는 화행(speech act)과 영역 종속적인 세부 의미를 나타내는 개념열(concept sequence)로 구성되며, 각각은 문맥에 의존하기 때문에 표층적인 발화 형태만으로 추론하는 것은 매우 어렵다[11]. 예를 들어, <표 1>의 발화 (9)를 표층적으로만 분석했을 경우에 영역 행위는 'inform & timetable-select-date' 또는 'response & timetable-update-date'가 될 수 있다. 이러한 모호성을 해결하기 위해서는 발화 (9)의 문맥을 고려해야 한다. 이 예제의 경우에 바로 이전의 발화 (8)을 고려하면 올바른 영역 행위인 'response & timetable-update-date'를 선택할 수 있다.

사용자 의도 파악에 대한 기존의 연구들은 계획 추론 모델(plan inference model)을 위한 레서피(recipe)나 영역 의존적인 지식에 기초해 왔다[1,8,12]. 그러나 이러한 영

〈표 1〉 발화와 영역 행위의 예

발화 (U: 사용자, S:시스템)	영역 행위 (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

역 의존적인 지식을 이용한 연구들은 응용 영역을 확장하거나 변경할 때마다 대용량의 지식들을 수동으로 다시 구축해야 한다는 단점이 있다. 최근에는 이러한 단점을 극복하기 위하여 다양한 기계 학습 모델들이 이용되고 있다[9,10,14]. 기계 학습 모델들은 대용량의 학습 데이터로부터 자동으로 추출된 정보를 이용하여 의도 범주에 사용자의 발화들을 효과적으로 할당한다. 그러나 기계 학습 모델들은 시스템 성능이 학습에 사용된 자질에 따라 크게 영향을 받는다는 단점이 있다. 즉, 입력 자질이 일정 수준이상의 오류를 포함하고 있다면 편향 학습(biased learning)에 의하여 잘못된 결과를 초래할 가능성이 많다. 또한, 선행된 기계 학습 모델들 [9,10,14]은 영역 독립적인 화행을 정의하고 인식하는 문제만을 다루고 있기 때문에 대화 시스템 구현 시에 필요한 영역 의존적인 세부 의미를 얻기 위해서는 별도의 분석 과정이 필요하다는 단점이 있다. 이러한 문제를 해결하기 위하여 본 논문에서는 화행과 개념열을 동시에 결정하는 기계 학습 기반의 통계 모델을 제안한다. 기계 학습 방법으로는 개체명 인식, 정보 추출과 같은 자연어 처리 분야에서 좋은 성능을 보이고 있는 CRFs(Conditional Random Fields)[7,13]를 이용한다. 그리고 편향 학습 문제를 줄이고 영역 변경의 편의성을 높이기 위하여 어휘나 품사와 같은 낮은 수준의 언어적 자질만을 사용한다. 또한 노이즈(noise) 자질에 따른 학습 시간의 증가와 성능의 저하를 막기 위하여  $\chi^2$  통계량[15]을 기반으로 정보량이 많은 자질만을 선택적으로 사용한다.

본 논문의 구성은 다음과 같다. 2장에서 CRFs를 이용한 영역 행위 분류 모델을 제안한다. 3장에서는 실험 데이터와 실험 결과를 분석한다. 마지막으로 4장에서 결론을 내린다.

## 영역 행위 분류

### 화행과 개념열

지금까지 대화 시스템 구현에 사용된 화행들은 합의된 표준 없이 영역에 따라

〈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

매우 주관적으로 정의되어 왔다. 화행이 영역 의존적으로 정의되면 사용자의 의도를 영역에 따라 정밀하게 분석할 수 있다는 장점이 있지만, 응용 영역을 확장하거나 변경하기 어려워진다는 단점이 있다. 이러한 문제를 줄이기 위하여 본 논문에서는 영역 독립적으로 화행을 정의하고, 영역 의존적인 의미들은 개념열을 통하여 보충하는 방법을 사용한다. <표 2>는 본 논문에서 정의한 화행을 보여준다.

목적 지향 대화는 상대방으로부터 특정 정보를 얻거나 전달하기 위하여 진행되며, 기존의 많은 대화 시스템들은 데이터베이스 오퍼레이션(database operation)을 이용하여 이러한 현상들을 모델링(modeling)하였다[4,6]. 본 논문에서는 목적 지향 대화에 나타나는 발화의 의미들이 데이터베이스 오퍼레이션과 유사성을 가진다는 가정 하에 <표 3>과 같이 3층 구조의 개념열 부착 방법을 제안한다. 3층 구조의 개념열 부착 방법은 각 발화에 적합한 테이블명(table name)과 오퍼레이션명(operation name), 필드명(field name)을 선택한 후, 그것들을 '-'로 연결하여 개념열을 표현하는 것이다.

예를 들어, <표 1>의 발화 (7)에 부착된 개념열이 의미하는 것은 *timetable*이라는 테이블의 *date* 필드를 *update*하는 것과 연관된 의미를 담고 있는 발화라는 뜻이다. 본

〈표 3〉 개념열 부착 방법

테이블명	오퍼레이션명	필드명
Timetable		Agent
	Insert	Date
	Delete	Day_of_week
	Select	Time
	Update	Person
		Place

논문에서는 실험을 위하여 3층 구조 개념열 부착 방법에 따라 일정관리 영역에서 53개의 개념열을 정의하였다.

문맥에 많이 의존하는 화행과는 다르게 적은 수의 개념열은 LSP(lexico-syntactic pattern)와 같은 규칙을 통하여 발화로부터 쉽게 얻을 수 있다. 그러나 응용 영역의 복잡도가 증가하게 되면 정의해야 하는 개념열의 수가 늘어나게 되고, 그에 따라 LSP 구축에 필요한 자원의 크기나 규칙의 수, 규칙의 순서 등에 많은 영향을 미치게 된다. 수동으로 구축해야 하는 규칙의 높은 복잡도는 영역 확장성을 떨어뜨리는 원인이 되므로 본 논문에서는 통계 기반의 방법을 이용하여 개념열을 결정하는 모델을 제안한다.

### CRFs를 이용한 영역 행위 분류 모델

사용자의 발화로부터 영역 행위를 찾기 위하여 본 논문에서는 수식 (1)과 같이 SACM(화행 분류 모델; Speech Act Classification Model)과 CSCM(개념열 분류 모델; Concept Sequence Classification Model)이라는 2가지 통계 모델을 제안한다.

$$\begin{aligned}
 SA(U_{1:n}) &\approx \arg \max_{S_{1:n}} \prod_{i=1}^n P(U_i | S_i) P(S_i | S_{1:i-1}) \\
 CS(U_{1:n}) &\approx \arg \max_{C_{1:n}} \prod_{i=1}^n P(U_i | C_i) P(C_i | C_{1:i-1})
 \end{aligned}
 \tag{1}$$

수식 (1)에서 첫 번째 것이 SACM이고, 두 번째 것이 CSCM이다.  $U_i, S_i, C_i$ 는  $n$ 개의 발화로 구성된 대화  $U_{1,n}$ 에서  $i$ 번째 발화,  $i$ 번째 화행,  $i$ 번째 개념열을 의미한다.  $P(U_i|S_i)$ 와  $U_i$ 와  $S_i$  사이의 관계를 나타내며,  $P(U_i|C_i)$ 는  $U_i$ 와  $C_i$  사이의 관계를 나타낸다.

화자는 개인의 언어적 습관에 따라 동일한 의미의 문장을 단어의 순서를 바꾸거나 불필요한 단어를 생략하는 방법 등을 통하여 다양한 형태로 표현하므로  $P(U_i|S_i)$ 나  $P(U_i|C_i)$ 를 직접적으로 계산하는 것은 매우 힘들다. 그러므로 본 논문에서는 발화가 문장 자질들의 집합에 의해서 근사화될 수 있다고 가정하고 수식 (1)을 수식 (2)와 같이 고쳐 쓴다.

$$SA(U_{1,n}) \approx \arg \max_{S_{1,n}} \prod_{i=1}^n P(F_i | S_i) P(S_i | S_{i-1})$$

$$CS(U_{1,n}) \approx \arg \max_{C_{1,n}} \prod_{i=1}^n P(F_i | C_i) P(C_i | C_{i-1})$$
(2)

수식 (2)에서  $F_i$ 가  $i$ 번째 발화의 문장 자질 집합으로서 어휘 자질과 품사 자질이라는 두 종류의 자질들로 구성된다[16]. 어휘 자질은 품사가 부착된 어휘를 말하며, 품사 자질은 품사 바이그램(bigram)을 말한다. 예를 들어, ‘안녕하세요?’라는 발화의 형태소 분석 결과가 ‘안녕/ncp+하/xsp+세요/ef+ ?/sf’라고 했을 때, 어휘 자질은 안녕/ncp, 하/xsp, 세요/ef, ?/sf이고, 품사 자질은 ncp-xsp, xsp-ef, ef-sf이다.

어휘 자질과 품사 자질의 추출이 끝나면 본 논문에서는  $\chi^2$  통계량을 이용하여 정보량이 적은 자질들을 제거한다.  $\chi^2$  통계량은 수식 (3)에서 보듯이 자질  $f$ 와 범주  $c$  사이의 독립성의 결여 정도를 나타낸다[15]. 본 논문에서 범주  $c$ 는 화행 또는 개념열을 의미한다.

$$\chi^2(f, c) = \frac{(A+B+C+D) \times (AD-CB)^2}{(A+C) \times (B+D) \times (A+B) \times (C+D)}$$
(3)

수식 (3)에서  $A$ 는  $c$ 에 속해 있는 발화 중에서  $f$ 를 포함하는 발화 수,  $B$ 는  $c$  이외의 범주에 속해 있는 발화 중에서  $f$ 를 포함하는 발화 수,  $C$ 는  $c$ 에 속해 있는 발화

중에서  $f$ 를 포함하지 않는 발화 수,  $D$ 는  $c$  이외의 범주에 속해 있는 발화 중에서  $f$ 를 포함하지 않는 발화 수를 의미한다.  $\chi^2$  통계량을 각 자질마다 계산을 한 후에, 정보량이 적은 것들을 제거하기 위하여 수식 (4)에서 보듯이 자질-범주 쌍의 최대  $\chi^2$  통계량을 계산하고 그것에 따라 상위  $n$ 개의 자질을 선택한다.

$$\chi^2_{\max}(f) = \max_{i=1}^m \{\chi^2(f, c_i)\} \quad (4)$$

<표 4>는 SACM과 CSCM에서 추출된 일정관리 영역에서의 문장 자질들을 보여 준다.

<표 4> 문장 자질의 예

SACM	
문장 자질	$\chi^2$ 통계량
고맙/pa	8052.532914
안녕/ncn	6564.320496
말씀/ncn	4568.785303
ma-ef	4449.282909
수고/ncp	4275.294269
px-ef	4157.455616
아니/ma	4062.648729
CSCM	
문장 자질	$\chi^2$ 통계량
취소/ncp	16523.85656
다시/ma	11840.10938
알람/ncn	10592.86434
설정/ncp	10101.11744
오/pv	9520.337957
원하/pv	9433.568085
요일/ncn	9225.014189

수식 (1)에서  $P(S_i | S_{1,i-1})$ 는 이전의 모든 화행들  $S_{1,i-1}$ 과 현재 화행  $S_i$  사이의 관계를 나타내며,  $P(C_i | C_{1,i-1})$ 는 이전의 모든 개념열들  $C_{1,i-1}$ 과 현재 개념열  $C_i$  사이의 관계를 나타낸다. 그러나 모든 이전 화행들이나 개념열들을 문맥 정보로써 고려하는 것은 불가능하기 때문에 본 논문에서는 1차 마코프(Markov) 가정을 통하여 수식 (2)를 수식 (5)와 같이 고쳐 쓴다.

$$\begin{aligned}
 SA(U_{1,n}) &\approx \arg \max_{S_{1,n}} \prod_{i=1}^n P(F_i | S_i) P(S_i | S_{i-1}) \\
 CS(U_{1,n}) &\approx \arg \max_{C_{1,n}} \prod_{i=1}^n P(F_i | C_i) P(C_i | C_{i-1})
 \end{aligned}
 \tag{5}$$

수식 (5)에서  $S_{i-1}$ 과  $C_{i-1}$ 는 각각  $i-1$ 번째 화행과 개념열을 의미하며, 본 논문에서는 문맥 자질이라고 부른다.

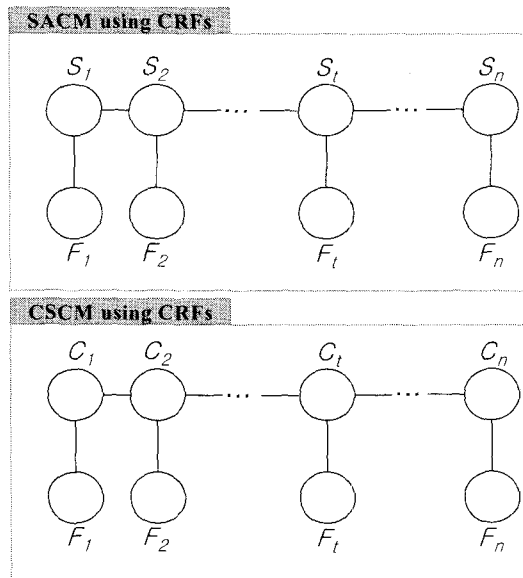
CRFs는 다양한 입력 노드의 값이 주어졌을 때 지정된 출력 노드의 조건부 확률값을 계산하기 위한 무방향성 그래프 모델로서 출력 노드가 선형 체인형태로 연결되었을 경우에 1차 마코프 모델의 형태를 따른다. CRFs의 주요 장점은 HMM(Hidden Markov Model)의 단점인 독립 가정을 완화시키는 효과가 있다는 것과 MEMM(Maximum Entropy Markov Model)의 단점인 레이블 편향 문제(label bias problem)를 극복할 수 있다는 것이다[7,13]. 본 논문에서는 1차 마코프 모델의 형태를 따르는 수식 (5)가 CRFs와 유사하다는 사실에 기초하여 수식 (6)과 같이 조건부 확률값을 계산한다.

$$\begin{aligned}
 P_{CRF}(S_{1,n} | F_{1,n}) &= \frac{1}{Z(F)} \exp\left(\sum_{i=1}^n \sum_k \lambda_k f_k(S_{i-1}, S_i, F_i^k, t)\right) \\
 P_{CRF}(C_{1,n} | F_{1,n}) &= \frac{1}{Z(F)} \exp\left(\sum_{i=1}^n \sum_k \lambda_k f_k(C_{i-1}, C_i, F_i^k, t)\right)
 \end{aligned}
 \tag{6}$$

수식 (6)에서  $S_{1,n}$ 과  $C_{1,n}$ 은  $n$ 개의 발화로 구성된 대화에 나타난 화행들과 개념열들의 순서를 가진 집합이다.

(그림 1)에서 보듯이 CRFs에서  $S_{1,n}$ 과  $C_{1,n}$ 은 관찰된 입력 데이터열(observed input data sequence)인 각 발화의 문장 자질들  $F_t$ 에 대한 출력 상태열(output





(그림 1) 영역 행위 분류를 위한 CRFs 모델

state sequences)을 의미한다.  $Z(F)$ 는 모든 상태열에 대한 정규화 요소이다.  $f_k(S_{t-1}, S_t, F_t^k, t)$ 와  $f_k(C_{t-1}, C_t, F_t^k, t)$ 는 화행 분류와 개념열 분류를 위한 자질 함수들로서 해당 자질이 나타나면 1의 값을 가지고 그렇지 않으면 0의 값을 가진다. 예를 들어, <표 1>의 발화 (9)에 적용되는 2번째 자질이  $\text{월}/nbn$ 이라고 가정을 한다면 자질 함수  $f_2(\text{ask-ref}, \text{response}, \text{월}/nbn, 9)$ 는 1의 값을 가진다.  $\lambda_k$ 는 각 자질 함수의 학습된 가중치를 의미한다.

## 실 험

### 실험 데이터

실험을 위하여 일정관리 영역에서 Wizard-Of-Oz 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. 훈련 데이터와 실험 데이터의 비율은 4:1로 나누었으며, 5배 교차검증(5-fold cross validation)을 수행하였다. CRFs의 내부 설정 인자로 추정알고리즘은 L-BFGS를 이용하였으며[3], 희소 데이터 문제를 위한 평탄화 요소는 Gaussian Prior를 이용하였다[7]. Gaussian Prior의 값은 10으로 설정하였으며, 훈련 반복 회수는 30으로 설정하였다.

### 실험 결과

영역 행위 분류 모델에 대한 실험에 앞서 최적의 문장 자질 수를 얻기 위한 실험을 수행하였다. 100개 단위로 증가하며 다양한 문장 자질 수에 대한 실험을 수행한 결과, SACM과 CSCM 모두 800개 근처에서 좋은 결과를 얻었다.

<표 5>는 동일한 훈련 데이터와 실험 데이터를 이용하여 화행과 개념열을 분류한 결과를 보여준다. 각 발화에 대해 화행과 개념열을 모두 맞춘 정확률은 88.1%였다.

<표 5>에서 Kim-2003[5]은 신경망을 이용하여 화행을 분류하는 모델이고, Choi-2005[2]는 MEMM을 이용하여 화행을 분류하는 모델이다. 제안 모델과 Kim-2003, Choi-2005는 서로 다른 입력 자질을 사용하기 때문에 직접적으로 비교하는 것은 힘들다. 그러나 <표 5>에서 보는 것과 같이 SACM은 이전의 두 모델보다 월등히 좋은 결과를 보였다. 또한 제안 시스템은 개념열 분류도 동일한 구조 내에서 수행할 수 있다는 장점이 있다. Kim-2003, Choi-2005와 제안 모델 사이의 화행 분

<표 5> 영역 행위 분류 정확률

모델	화행 분류 평균 정확률(%)	개념열 분류 평균 정확률(%)
Kim-2003	85.5	-
Choi-2005	88.6	-
En-2005	90.2	
제안 모델	93.0	90.2

류 성능 차이를 분석한 결과, 표층 발화를 구문 자질로 일반화하는 과정에서 많은 정보들이 생략된 것이 성능 차이의 가장 큰 원인임을 알 수 있었다. 이 분석 결과는 어느 수준의 언어 정보를 자질로 사용하느냐 하는 것 보다는 얼마나 많은 정보량을 가지고 있는 자질을 사용하느냐 하는 것이 화행 분류 성능에 영향을 미친다는 것을 간접적으로 보여준다고 할 수 있다. 성능 차이 분석 과정에서 Kim-2003과 Choi-2005에서 사용한 구문 자질은 반자동 방법에 따라 오류를 정제한 것으로 제안 모델과의 성능 차이가 편향 학습 문제에 기인한 것은 아니라는 사실도 알 수 있었다. En-2005[17]은 제안 모델과 동일한 입력 자질(문장 자질과 문맥 자질)을 SVM(Support Vector Machine)의 입력으로 사용하여 화행을 분류하는 모델이다. [17]에 따르면 문장 자질과 문맥 자질에 구문 자질을 추가하였을 경우에 화행 분류 성능이 향상되는 것으로 알려져 있다. 그러나 [17]에서 사용한 구문 자질은 Kim-2003과 Choi-2005에서 사용한 것과 동일한 방법에 따라 반자동으로 정제한 것으로, 이는 본 논문의 목적 중에 하나인 하위 수준의 자질을 사용하여 편향 학습 문제를 줄이고자 하는 것에 위배되기 때문에 구문 자질을 이용한 실험은 생략하였다. [표 5]에서 보듯이 제안 모델은 화행 분류 정확률에서 En-2005보다 3.2% 좋은 성능을 보였다. 그러나 이 실험 결과는 자연어처리 모든 영역에서 CRFs가 SVM보다 우수하다는 것을 보여주는 것은 아니며, 단지 화행 분류의 경우에 CRFs가 더 적당할 수 있다는 가능성을 보여주는 것이다.

잘못된 영역 행위 분류를 수행한 유형을 살펴본 결과, 많은 오류들이 문맥 정보를 잘못 사용하여 생긴 것이었다. 제안된 모델들은 문맥 정보로써 바로 이전의 화행이나 개념열을 이용한다. 그러나 부대화가 삽입된 경우에 대화는 계층적 담화구조를 형성하며, 바로 이전의 화행이나 개념열보다는 계층적으로 인접한 것들이 올바른 문맥이 된다. 이런 문제를 해결하기 위해서는 담화 구조 분석 모델을 구현하고 제안된 모델들과 통합하여 올바른 문맥을 자질로 사용해야 한다.

## 결 론

본 논문에서는 사용자의 발화 의도를 분류하는 모델을 제안하였다. 효과적인 발

화 의도 분류를 위하여 영역 독립적인 화행과 영역 의존적인 개념열의 쌍으로 구성된 영역 행위를 정의하였다. 그리고 화행 분류 모델과 개념열 분류 모델을 제안하고 CRFs를 이용하여 구현하였다. 실험 결과에 따르면 화행 분류와 개념열 분류 성능이 각각 93.0%, 90.2%로 매우 우수함을 알 수 있었다.

## 참고문헌

- [1] Caberry, S. (1989), *A pragmatics-based approach to ellipsis resolution*, Computational Linguistics, 15-2, 75-96.
- [2] Choi, W., Kim, H., and Seo, J. (2005), *An Integrated dialogue analysis model for determining speech acts and discourse structures*, IEICE Information and Systems, E88D-1, 150-157.
- [3] Fei, S. and Pereira, F. (2003), "Shallow parsing with conditional random fields," in *Proceedings of HLT & NAACL 2003*.
- [4] Goddeau, D., Meng, H., Polifroni, J., Seneff, S., and Busayapongchai, S. (1996), "A form-based dialogue manager for spoken language applications," in *Proceedings of International Conference on Spoken Language Processing*, 701-704.
- [5] Kim, H. and Seo, J. (2003), *An efficient trigram model for speech act analysis in small training corpus*, Journal of Cognitive Science, 4-1, 107-120.
- [6] Kim, H., Seon, C., and Seo, J. (2005), *A dialogue-based information retrieval assistant using shallow NLP techniques in online sales domains*, IEICE Information and Systems, E88D-5, 801-808.
- [7] Lafferty, J., McCallum, A., and Pereira, F. (2001), "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of ICML 2001*.
- [8] Lambert, L. and Caberry, S. (1991), "A tripartite plan-based model of dialogue," in *Proceedings of ACL 1991*, 47-54.
- [9] Langley, C. (2002), "Analysis for speech translation using grammar-based parsing and

- automatic classification,” in *Proceedings of the ACL Student Research Workshop*.
- [10] Lee, S. and Seo, J. (2001), “An analysis of Korean speech act using hidden Markov model with decision trees,” in *Proceedings of ICCPOL 2001*, 397-400.
- [11] Levin, L., Langley, C., Lavie, A., Gates, D. (2003), Wallace, D., and Peterson, K., “Domain specific speech acts for spoken language translation,” in *Proceedings of 4th SIGdial Workshop on Discourse and Dialogue*.
- [12] Litman, D. J. and Allen, J. F. (1987), *A plan recognition model for subdialogues in conversations*, Cognitive Science, 11, 163-200.
- [13] Pinto, D., McCallum, A., Wei, X., and Croft, W. B. (2003), “Table extraction using conditional random fields,” in *Proceedings of SIGIR 2003*.
- [14] Samuel, K., Caberry, S., and Vijay-Shanker, K. (1998), “Computing dialogue acts from features with transform-based learning,” in *Proceedings of the AAAI Spring Symposium*, 90-97.
- [15] Yang, Y. and Pedersen, J. O. (1997), “A comparative study on feature selection in text categorization,” in *Proceedings of ICML 1997*.
- [16] 김경선, 서정연 (2003), 자질 선택 기법을 이용한 한국어 화행 결정, 한국정보과학회 논문지, 30-3, 278-284.
- [17] 은종민, 이성욱, 서정연 (2005), 지지벡터기계(support vector machines)를 이용한 한국어 화행분석, 한국정보처리학회 논문지, 12B-3, 365-368.

1 차원고접수 : 2006. 8. 21

2 차원고접수 : 2006. 12. 12

최종게재승인 : 2007. 2. 15

*(Abstract)*

## A Domain Action Classification Model Using Conditional Random Fields

Harksoo Kim

Kangwon National 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. Therefore, if we plan to implement an intelligent dialogue system, it is very important to correctly infer the domain actions from surface utterances. In this paper, we propose a statistical model to determine speech acts and concept sequences using conditional random fields at the same time. To avoid biased learning problems, the proposed model uses low-level linguistic features such as lexicals and parts-of-speech. Then, it filters out uninformative features using the chi-square statistic. In the experiments in a schedule arrangement domain, the proposed system showed good performances (the precision of 93.0% on speech act classification and the precision of 90.2% on concept sequence classification).

*Keywords : CRFs(Conditional Random Fields), domain action classification, speech act classification, concept sequence classification*