3

음성 언어 이해를 위한 기계학습

포항공과대학교 ▌정민우 · 이근배*

1. 서 론

최근 음성 대화, 비전, 멀티모달 인터페이스와 같은 차세대 지능형 인터페이스에 대한 수요가 증가하면서 이와 관련된 많은 연구가 진행되고 있다. 이 중에서 음성 대화 인터페이스는 자연어 대화를 통해 인간-기계 사이의 상호 작용을 이끌어내는 것으로 사람의 말을 기계가 이해하고 적절한 응답을 다시 사람에게 전달하는 것을 목적으로 한다[1].

음성 대화 인터페이스를 구현하기 위해서는 사람의 언어인 자연어를 기계가 이해할 수 있는 형식 언어 혹은 구조적 데이터로 변환하는 과정, 즉 음성 언어를 이해하는 과정이 필요하다. 즉, **음성 언어 이해란 사용자(사람)의 발화로부터 유의한 정보를 구분/추출하여 이를 구조화 시키는 문제로 요약 될 수 있다.**

음성 언어 이해에 대한 연구는1990년도 초반 미국의 DARPA기관의 지원으로 ATIS(Air Travel Information Service) 프로젝트와 Communicator 프로젝트를 중심으로 진행되어 왔다[2,3]. 초기의 음성 언어 이해 시스템은 서비스 개발 전문가가 작성한 특정한 문법과 어휘 사전을 이용하는 규칙 기반의 방법을 이용하였다. 하지만 이는 규칙 작성 비용이 크고 다른 서비스 영역으로 확장이 어려운 단점이 있다. **규칙 기반 방법의 대안으로 대량의 데이터를 이용한 통계적 음성 언어 이해를 시도되었으며, 최근 연구는 기계학습 방법론을 이용한 음성 언어 이해 방법론 개발에 초점을 맞추고 있다.**

본 고에서는 음성 언어 이해를 위한 기계학습 방법론에 대하여 소개하고, 최근의 연구 동향과 관련된 이슈들을 살펴보고자 한다.

2. 통계적 음성 언어 이해

2.1 의미 구조 표현

음성 언어 이해는 사용자가 자연어로 발화한 문장을 입력으로 받아 구조적 정보를 생성하는 과정이다. 생성된 정보는 형식 언어 혹은 구조적 데이터(XML)로 표현된다. 예를 들어 사용자가 항공 예약을 위해 시스템에 전화를 걸어 "Show me flights from Seoul to New York" 라고 얘기하면 시스템은 이를 다음과 같은 XML로 분석한다.

```
〈frame name= 'FLIGHT' type=void〉
  〈slot type= 'DA' 〉ShowFlight〈/slot〉
  〈slot type= 'NE' name= 'City' 〉
    〈slot name= 'DCity' 〉Seoul〈/slot〉
    〈slot name= 'ACity' 〉New York〈/slot〉
  〈/slot〉
〈/frame〉
```

XML로 표현된 정보는 크게 frame, slot, value로 구분된다. 이러한 표현 방식을 격틀(Case Frame)이라고 하며 자연어의 의미를 구조화하기 위한 표현 방식 중 하나이다[4]. 항공 예약에 대한 위의 예는 FLIGHT 격틀로 정의 할 수 있다. 각 slot(XML 태그)은 메타 데이터에 해당하며 의미를 나타나는 클래스가 되며, value는 그에 해당하는 실제 값이다. 사용자 발화의 의도를 나타내는 화행(Dialog Act; DA) slot은 ShowFlight라는 value를 가지며, 의미를 갖는 고유 명사에 해당하는 개체명(Named Entity; NE)은 City slot으로 표현된다. City slot은 출발지인 DCity와 도착지 ACity로 구분되어 계층적 구조를 가지며 이에 해당하는 value는 각각 Seoul과 New York이 된다.

결국 **음성 언어 이해는 주어진 발화로부터 화행 분류와 개체명 추출, 두 문제로 분할된다.** 다시 말해, 음성 인식기를 통해 텍스트 형태로 변환된 사용자의 발화 x는, 개체명 추출, 화행 분류 단계를 거쳐, 개체명 y과 화행 z을 생성하여 대화 관리 시스템에 전달한다. 이 전달된 정보들을 바탕으로 대화 시스템이 운영된다 (그림 1).

* 종신회원

† 이 연구는 21세기 프론티어 연구 개발사업(인간기능 생활지원 지능로봇 기술개사업)의 일환으로 수행 되었습니다.

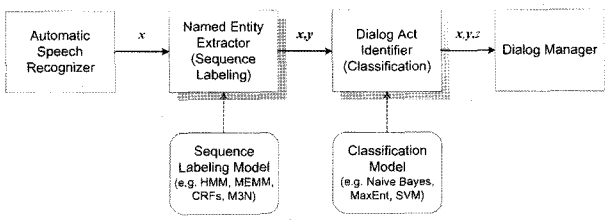


그림 1 음성 대화 시스템과 음성 언어 이해

2.2 통계적 방법

음성 언어 이해 문제를 수학적으로 표현하면 다음과 같다. y와 z는 x가 주어졌을 때 서로 조건부 독립이라고 가정한다. 결국 **음성 언어 이해의 개체명, 화행 분류 문제는 확률변수 x에서 각각의 이산확률변수 y, z로 매핑되는 함수 f: x→y와 함수 g: x→z를 찾는 문제가 된다. 이는 기계학습 분야의 용어에 따라 분류(Classification) 문제로 정의한다.**

함수 f와 g는 다른 특성을 지닌다. z는 스칼라 변수이지만 y는 벡터 변수이다(본 고에서는 벡터를 볼드체로 표기하기로 한다). 함수 f는 일반적인 분류 문제인 g와는 다르게 벡터와 벡터에 대한 함수 즉 구조적 분류 문제이다[5]. 이 문제는 입력 벡터 x와 출력 벡터 y사이에 특정한 구조를 가정한다. **기계학습 분야에서는 구조적 분류(Structured Classification), 구조적 학습(Structured Learning), 관계 학습(Relational Learning) 등의 여러 가지 용어로 불린다[5,6,7].** 본 고에서는 구조적 분류로 통일하여 부르기로 한다.

확률변수 x, y 사이에 구조는 일반적으로 서열(Sequence), 트리(Tree), 이분 그래프(Bipartite Graph) 등의 구조가 이용될 수 있다[7]. 만약 아무런 구조가 없다고 가정하면(독립 구조라고 부르기로 한다) 이 문제는 일반적인 분류 문제 g와 동일하다. 본 고에서는 개체명 인식 문제를 서열 구조로, 화행 분류 문제를 독립 구조로 가정한다.

서열 구조를 이용한 구조적 분류 방식을 서열 레이블링(Sequence Labeling)이라고 부른다. 여기에서 개체명 인식 문제는 관측 확률변수 x와 이전 개체명 $y_{1:t-1}$가 주어졌을 때 y_t를 예측하는 문제가 된다. 서열 레이블링 알고리즘을 적용하기 위해서는 계층적 구조를 갖는 XML 형식을 다음과 같은 서열 형태로 변환되어야만 한다.

Show/O me/O flights/O from/O Seoul/Departure.City-B to/O New/Arrival.City-B York/Arrival.City-I

이때 하나 이상의 단어가 모여 하나의 개체명을 이룰 수도 있는데(New York) 이를 위해 B/I/O 인코딩 방법을 이용한다. 각 클래스는 X-B, X-I, O와 같은 형태로 확장 되며, X-B는 클래스 X의 시작, X-I는 클래스 X가 계속됨, O는 어떠한 개체명도 아님을 나타낸다. 또한 계층적 구조가 평탄화되어 Departure.City와 Arrival.City는 다른 클래스로 구분된다.

2.3 탐색 공간

구조적 분류 문제는 모든 가능한 이산확률변수의 조합으로부터 가장 최적의 결과를 찾는 문제이다. 이것은 다음과 같이 탐색 공간 S에서 최적의 경로를 탐색하는 문제로 해석할 수 있다.

$$\hat{y} = \arg\max_{y \in S} f(x, y, w)$$

이때 w는 경로 평가 함수 f의 파라미터이다. 독립 분류를 다루는 함수 g는 f의 부분 집합으로 보고 본 고에서는 f 를 구하는 문제에 초점을 맞추기로 한다.

이제 구조적 분류 문제는 주어진 탐색 공간 S에서 평가 함수 f를 이용하여 최적 경로를 찾는 문제가 된다. 최적 경로를 찾는 방법은 동적 프로그래밍을 기반으로 한 탐색 알고리즘으로 효율적으로 구현할 수 있다. 효율적인 탐색을 위해 Viterbi, Beam, A*, Max-Sum 알고리즘 등의 다양한 탐색 기법들이 사용될 수 있다[5].

본 절에서는 개체명 인식을 위한 서열 레이블링 문제에서의 Viterbi 탐색 기법에 대해 간략히 설명한다. 1차 마르코프 모델을 가정하면 y_t 변수는 바로 이전 y_{t-1} 변수에만 영향을 받는다. 가능한 모든 이산 변수 y_t의 클래스의 수를 |y|라고 정의하고 서열의 길이를 T라고 하면 탐색 공간 S는 |y|*T의 격자로 표현된다. 그림 2는 격자로 표현된 서열 구조 탐색 공간을 나타낸다. 이때 구현상 편의를 위해 격자의 시작과 끝 부분에 더미 노드를 추가한다. 각 전이에 대한 평가 값은 행렬 Ψ의 요소인 $\Psi_t(j,i)$에 저장된다. $\Psi_t(j,i)$는 t 시간 상태에서 i번째 노드에서 j번째 노드로의 전이에 대한 평가 값를 저장한다(확률 모델에서는 Ψ 행렬은 확률 값이 가지며, 일반적으로는 비 정규화된 평가 값을 가진다).

$$\Psi_t(j,i) \equiv Score(y_t = j, y_{t-1} = i, x_t = x)$$

Viterbi 탐색은 평가 누적 값이 최대인 경로를 찾는 동적 프로그래밍 알고리즘이다. |y|*T 크기의 행렬 δ, ξ는 각각 t시간 상태에서 최대의 누적 값과 그에 해당하는 인덱스를 저장한다.

$$\delta_t(j) = \max_{i \in y} \Psi_t(j,i)\delta_{t-1}(i)$$

$$\xi_t(j) = \arg\max_{i \in y} \Psi_t(j,i)\delta_{t-1}(i)$$

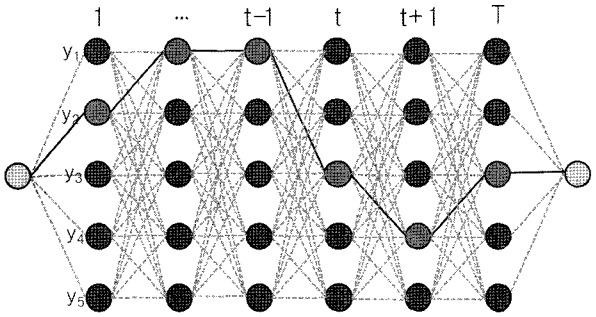


그림 2 서열 구조 탐색 공간

순환 함수인 δ, ξ를 이용하여 마지막 T+1 노드까지 계산을 하면, 인덱스 행렬 ξ를 역탐색하는 방법으로 최적의 경로를 선택하게 된다.

3. 구조적 분류 기법

음성 언어 이해를 위한 구조적 분류 기법은 기본적으로 교사 학습에 기반을 두고 있다. 구조적 분류 모델을 학습하기 위해 다음과 같은 학습 데이터 D가 주어진다.

$$D = \{x^{(n)}, y^{(n)} \mid 1 \le n \le N\}$$

이때 데이터는 독립 동일 분포(independent identically distribution; i.i.d)에서 관측된다고 가정한다. **음성 언어 이해를 위한 기계학습 알고리즘의 목표는 주어진 데이터 D를 이용하여 예측 오류를 최소화하는 파라미터 벡터 w를 구하는 것이 된다.** 본 절에서는 소개하는 구조적 분류 기법들은 모두 이러한 가정을 바탕으로 한다.

3.1 은닉 마르코프 모델

은닉 마르코프 모델(hidden Markov model; HMM)은 1970년대 이후 음성 인식 분야에서 널리 사용되던 방법[8]으로, 결합 확률인 $p_w(x,y)$을 평가 함수 f로 사용한다. 결합 확률 $p_w(x,y)$는 베이스 정리에 의해 다음과 같이 분해된다.

$$p_w(x,y) = \frac{p_w(x \mid y) p_w(y)}{\sum_y p_w(x \mid y) p_w(y)}$$

$p_w(x|y)$를 관측 확률 (observation probability), $p_w(y)$를 전이 확률(transition probability)이라고 부른다. 분모 항은 정규화 항으로 탐색 문제에서는 필요치 않다. 1차 HMM을 이용한 탐색은 다음과 같은 식이 된다.

$$\hat{y} = \arg\max_{y \in S} \prod_{t=1}^{T} p_w(x_t \mid y_t) p_w(y_t \mid y_{t-1})$$

HMM 학습 과정은 각각의 확률 $p_w(x_t \mid y_t)$, $p_w(y_t)$을 최대 우도 추정(Maximum Likelihood Estimation)을 사용하여 구하는 것을 말한다. 주어진 데이터 D에서 x, y 모두 관측 가능하므로 학습 과정은 단순한 출현 빈도를 세는 Count 함수로 간단히 구할 수 있다.

$$p_w(x_t \mid y_t) = \frac{count(x_t, y_t \mid D)}{count(y_t \mid D)}$$

$$p_w(y_t \mid y_{t-1}) = \frac{count(y_t, y_{t-1} \mid D)}{count(y_{t-1} \mid D)}$$

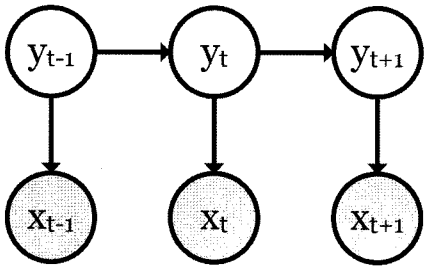
HMM은 음성 인식, 품사 부착, 정보 추출, 염기 서열 분석 등의 많은 문제에서 널리 사용된 모델이다. 하지만 결합 확률 분포를 모형화 하는 것은 복잡한 관계 정보 및 특징 정보를 혼합하여 사용하는데 어려움이 따른다[6]. 따라서 음성 언어 이해와 같이 다양한 정보를 이용해야 하는 분류 문제에서는 좋은 성능을 보여주지는 못한다.

3.2 최대 엔트로피 기반 모델

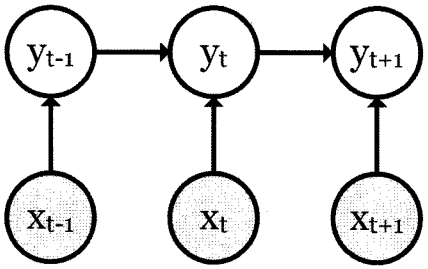
분류 문제를 확률적으로 모델링 하려고 하면, 결합 확률보다는 조건부 확률 $p_w(y \mid x)$를 직접 다루는 것이 좋은 해결책이 된다[6]. 이러한 방법을 Discriminative Model이라고 하며 HMM과 같은 결합 확률을 이용하는 방법인 Generative Model과 구분 된다.

최대 엔트로피(Maximum Entropy) 모델은 많은 분류 문제에 적용된 대표적인 조건부 확률 모델이다(통계학 분야에서는 로지스틱 회귀분석으로 알려져 있다). **최대 엔트로피 모델은 독립 분류 문제만을 다루며 이를 구조적 분류 모델로 확장시킨 첫 시도는 최대 엔트로피 마르코프 모델(Maximum Entropy Markov Model; MEMM)이다[9].** MEMM은 $y_{1:t-1}$정보를 y_t 예측에 필요한 관측 벡터 x에 더하여 일반적인 최대 엔트로피와 같은 학습 방법을 이용하였다. 이는 효율적이긴 하나 미 관측 데이터에 대해 Label Bias문제를 야기시키는 단점이 있다[10]. 이러한 문제점을 해소하기 위해 조건부 임의 필드(Conditional Random Field; CRF) 모델이 제안되었고 많은 구조적 분류 문제에 적용되었다. 그림 3은 각 확률적 서열 구조 모델들을 그래프 모형으로 비교한 것이다[10].

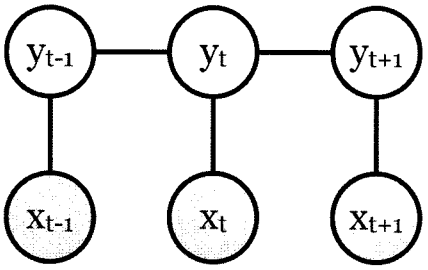
CRF 모델이 다른 두 모델과 다른 점은 방향성이 없는 그래프를 이용한다는 점이다. CRF에서는 조건부 확률 $p_w(y \mid x)$을 Gibbs 분포라고 가정하여 다음과 같은 식을 유도한다.



(a) HMM



(b) MEMM



(c) CRF

그림 3 확률 서열 모델

$$p_w(y \mid x) = \frac{e^{\sum_t w^T \phi(x_t, y_t)}}{\sum_y e^{\sum_t w^T \phi(x_t, y_t)}}$$

이때 w는 파라미터 벡터, ϕ는 특징 함수 벡터이며 분모 항은 전역 정규화 항이다. CRF 모델은 다양한 특징(Feature)들을 결합해 사용하는 것이 가능하다. 또한 탐색 가능한 모든 경로 S에 대해 전역적인 정규화 과정을 거친다는 특징이 있다. CRF에서는 이러한 전역 정규화 과정을 통해 Label Bias 문제를 해소하고 있다.

CRF의 학습은 최대 우도 추정을 이용한다. 조건부 로그 우도 함수 LL(w)는 다음과 같다.

$$LL(w) = \sum_n \log p_w(y^{(n)} \mid x^{(n)}) =$$
$$\sum_n \sum_t w^T \phi(x_t^{(n)}, y_t^{(n)}) - \sum_n \log \sum_y e^{\sum_t w^T \phi(x_t^{(n)}, y_t^{(n)})}$$

이때 LL(w)는 Convex함수가 되어 유일해가 존재한다. 파라미터 w는 Generalized Iterative Scaling, Improved Iterative Sailing, Conjugate Gradient, Quasi-Newton 알고리즘 등과 같은 최적화 방법을 이용하여 구할 수 있다[11]. 대부분의 최적화 알고리즘은 w를 구하기 위해 Gradient을 사용한다.

$$\nabla LL(w) = \sum_{n,t} \phi(x_t^{(n)}, y_t^{(n)}) - \sum_{n,t,y} p_w(y_t^{(n)} \mid x_t^{(n)}) \phi(x_t^{(n)}, y_t^{(n)})$$

Gradient를 구하기 위해서는 $p_w(y_t \mid x_t)$ 항을 구해야만 한다. 서열 구조에서는 Sum-Product혹은 Forward-Backward 알고리즘을 이용하여 효율적으로 계산이 가능하다[10].

3.3 마진 기반 모델

본 절에서는 앞서 소개된 확률 기반 모델들과는 다른 마진(Margin) 기반의 모델에 대하여 소개한다. 마진 기반의 모델에서는 확률 함수 p_w 대신 평가 함수 Score를 사용한다.

$$\hat{y} = \arg\max_{y \in S} Score(x, y) = \arg\max_{y \in S} w^T \phi(x, y)$$

이때 Score함수를 w와 ϕ에 대한 선형 모델로 가정한다. 이때 w는 데이터가 분포된 함수 공간에서 분류를 결정하는 Separating Hyperplane으로 해석된다. 마진 기반의 모델에서는 분류 문제를 위해서 정답 클래스 y와 정답이 아닌 모든 다른 클래스에 대해 다음과 같은 식을 항상 만족하는 Score 함수를 찾는 것이 목표가 된다.

$$Score(x, y) - Score(x, \bar{y}) > 0, \forall \bar{y} \neq y$$

이러한 선형 모델을 학습하는 대표적인 알고리즘은 Perceptron 이 있다[12]. y가 (+1,-1)의 이진 값을 가진다고 가정하고, y값을 결정하는 결정 경계(Decision Boundary)를

$$y = \begin{cases} +1 & w^T \phi \geq 0 \\ -1 & w^T \phi < 0 \end{cases}$$

라고 정의하면, 손실(loss) 함수 l(w)는 다음과 같이 정의 할 수 있다.

$$l(w) = -\sum_{y \neq \hat{y}} w^T \phi(x) y$$

Perceptron 학습은 각 학습 데이터에 대해 평가를 수행한 후 손실이 감소되는 방향으로 파라미터 벡터 w를 수정하는 방법이다.

$$w^{(i+1)} = w^{(i)} + \nabla l(w^{(i)})$$

```
Algorithm StructuredPerceptron
Input: Training examples (x1:N,y1:N)
1: w ← 0
2: for i=1...I do
3:     for n=1...N do
4:         yn* ← argmaxy wTφ(xn,yn)
5:         if yn* ≠ yn then
6:             w ← w + φ(xn,yn) - φ(xn,yn*)
7:         end if
8:     end for
9: end for
10: return w
```

그림 4 구조적 분류를 위한 Perceptron 학습

Perceptron 알고리즘은 기본적으로 이진 분류 문제만을 다루며, 이를 K개의 클래스를 다루는 구조적 분류 문제로 확장한 알고리즘은 그림 4와 같다[5,12].

또 다른 마진 기반의 모델로서 최대 마진 마르코프 모델이 있다. 최대 마진 마르코프 모델(Maximum Margin Markov Model; M3N)은 지지 벡터 기계(Support Vector Machine; SVM)을 확장하여 구조적 분류 문제에 적용하였다[7]. M3N은 다음과 같은 최적화 문제로 구조적 분류 문제를 다룬다.

$$\min_w \frac{1}{2}\|w\|^2 + C\sum_{n=1}^{N}\sum_{\hat{y}}\xi_{n,\hat{y}}$$

$$s.t \quad w^T\phi(x_n,y_n) - w^T\phi(x_n,\hat{y}) \geq l(x_n,y_n,\hat{y}) - \xi_{n,\hat{y}}$$

$$\xi_{n,\hat{y}} \geq 0$$

M3N은 최적화 문제에서 제약 조건 항에 손실 항을 첨가함으로써 구조적 분류 문제를 효과적으로 모델링 하였다. 즉, M3N은 다음 식과 같은 손실 첨가 탐색(loss-augmented search)을 수행한다.

$$\hat{y} = \arg\max_{y\in Y} w^T\phi(x,y) + l(x,y,\hat{y})$$

지금까지 대표적인 구조적 분류 기법에 대해 소개하였다. 소개된 방법 외에도 현재에도 구조적 분류를 위한 많은 기계학습 방법들이 연구 개발되고 있다. 좀 더 다양한 기법에 대해서 알아보길 원하는 독자는 NIPS, ICML등의 학회를 참고하길 바란다.

4. 음성 언어 이해 연구 이슈

본 절에서는 기계학습을 이용한 음성 언어 이해의 최근 연구 이슈에 대해 간략히 소개한다.

4.1 음성 인식 오류에 강건한 모델

음성 언어 이해는 음성 인식기의 결과를 입력으로 받는 것을 가정하고 있다. 따라서 음성 인식기의 오류는 음성 언어 이해 성능 저하에 영향을 준다. 이러한 인식 오류에 강인한 시스템을 구성하는 것이 최근 연구 이슈 중 하나이다.

최근의 시도는 음성 인식기의 다양한 정보를 이용하는 방향으로 연구가 진행되고 있다[13,14]. 입력 변수 x가 단순한 단어열이 아닌 그래프, 격자, 문장 리스트등의 더 많은 정보를 가지는 형태가 된다. 이러한 확장된 입력에 따라 탐색 공간이 지수적으로 증가하게 되므로 이에 따라 효율적인 탐색 알고리즘에 대한 연구가 요구된다.

4.2 전역 관계 정보 이용

1차 마르코프 가정에 의한 서열 레이블링 모델을 이용할 때 문제점은 전역적인 관계 정보가 누락된다는 것이다. 이러한 전역 관계(Global Dependency or Long-distance Dependency) 문제는 특히 계층적 구조를 단순한 서열 구조로 변환시켰을 때 문제가 된다. 이를 해결하기 위한 방법으로는 계층적 구조를 도입하거나[15], 전역 정보를 자동으로 추출하는 방법[16] 등이 이용될 수 있다.

4.3 결합 모델을 이용한 성능 향상

본 고에서는 음성 언어 이해 문제를 개체명 인식 문제와 화행 분류 문제로 나누어 다루었다. 하지만 이 두 문제가 서로 독립이라는 가정은 더 이상 성립하지 않는 경우가 있을 수 있다. 기계학습 측면에서 볼 때 함수 f:x→y와 함수 g:x→z는 각각 (x,y)와 (x,z)에 대해서만 모델링 한다. 이때 (y,z) 사이의 구조를 추가하여 (x,y,z)을 동시에 모델링 하는 함수 h:x→y,z를 구성하는 것이 가능하다[17]. 이러한 모델은 결합 모델이라고 부르며 최근에는 이러한 복잡한 관계 분석을 통한 성능 향상에 대한 연구가 진행되고 있다.

4.4 데이터 부족 문제 및 영역 적응

통계적 기계학습 방법, 특히 교사 학습에 기반한 방법은 다량의 데이터를 필요로 한다. 하지만 음성 언어 이해는 특정한 서비스 영역에 대한 음성 대화 데이터를 수집하기가 어려우므로 데이터 부족 문제가 발생하게 된다. 이를 해소하기 위해 반교사 학습 방법[18]을 이용하거나 선험 지식인 규칙을 추가하는 방법을 이용할 수 있다. 이뿐만 아니라 구축된 모델을 다른 영역으로 쉽게 적응시키는 방법 또한 최근의 연구 이슈이다.

5. 결 론

본 고에서는 음성 언어 이해를 위한 기계학습 방

법을 소개하였다. **음성 언어 이해 문제는 구조적 분류 문제로 접근할 수 있으며, 이를 위한 다양한 기법들이 연구되고 있다. 구조적 분류 모델과 이를 응용한 연구는 기계학습 분야에서 최근 들어 활발히 연구된 주제이다.** 음성 언어 이해뿐만이 아닌 다양한 응용 분야에 구조적 분류 알고리즘이 적용되어 효과를 이룰 것으로 기대해 본다.

참고문헌

[1] Allen, J., Byron, D., Dzikovska, M., Ferguson, G., Galescu, L., and Stent, A., Toward conversational human-computer interaction, AI Magazine, 22(4) : 27-37, 2001.

[2] Price, P.J., Evaluation of spoken language systems: the ATIS domain. DARPA Workshop on Speech and Natural Language, 1990.

[3] Walker, M., Rudnicky, A., Prasad, R., Aberdeen, J., Bratt, E., Garofolo, J., Hastie, H., Le, A., Pellom, B., Potamianos, A., Passonneau, R., Roukos, S., Sanders, G., Seneff, S., Stallard, D., Darpa communicator: cross-system results for the 2001. ICSLP, 2002.

[4] Fillmore, C. J., The case for case. In Bach and Harms (Ed.): Universals in Linguistic Theory. New York: Holt, Rinehart, and Winston, 1-88, 1968.

[5] Daumé III, H., Practical structured learning techniques for natural language processing. Ph.D Thesis, USC, 2006.

[6] Sutton, C., McCallum, A., An introduction to conditional random fields for relational learning. In Getoor, L., Taskar, B. (Eds.), Introduction to Statistical Relational Learning. MIT Press, 2006.

[7] Taskar, B., Julien, S. L., Jordan, M. I., Structured Prediction, Dual Extragradient and Bregman Projections. JMLR 7(Jul) : 1627-1653, 2006.

[8] Rabiner, L. R., A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE 77(2), 257-286, 1989.

[9] McCallum, A., Freitag, D., and Pereira, F., Maximum Entropy Markov Models for Information Extraction and Segmentation. ICML, 2000.

[10] Lafferty, J., McCallum, A., Pereira, F., Conditional random fields: Probabilistic models for segmenting and labeling sequence data. ICML. pp. 282-289. 2001.

[11] Nocedal, J., Wright, S. J., Numerical Optimization. Springer, New York, 1999.

[12] Collins, M. Discriminative Training Methods for Hidden Markov Models: Theory and Experiments with Perceptron Algorithms. EMNLP, 2002.

[13] Hakkani-Tür, D., Béchet, F., Riccardi, G., and Tur, G., Beyond ASR 1-Best: Using Word Confusion Networks in Spoken Language Understanding. Computer Speech and Language, 20(4) : 495-514, 2006.

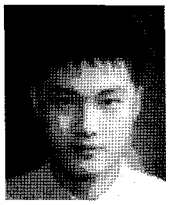
[14] Jeong, M., Eun, J., Jung, S., and Lee, G.G. An error-corrective language-model adaptation for automatic speech recognition, Eurospeech, 2005.

[15] He, Y., and Young, S., Semantic Processing using the Hidden Vector State Model, Computer Speech and Language, 19(1) : 85-106, 2005.

[16] Jeong, M., Lee, G. G., Exploiting non-local features for spoken language understanding. COLING/ACL, 2006.

[17] Jeong, M., and Lee, G.G., Jointly Predicting Dialog Act and Named Entity for Spoken Language Understanding, IEEE/ACL Workshop on SLT, 2006.

[18] Tur, G., Hakkani-Tur, D., Schapire, R. E., Combining active and semi-supervised learning for spoken language understanding. Speech Communication, 45(2) : 171-186, 2005.



정 민 우

1999.3～2003.2 전북대학교 컴퓨터공학과 학사
2003.3～2005.2 포항공과대학교 컴퓨터공학과 석사
2005.3～현재 포항공과대학교 컴퓨터공학과 박사과정
관심분야 : Language Understanding and Modeling, Structured Prediction, Machine Learning for NLP
E-mail : stardust@postech.ac.kr

이 근 배

1984 서울대학교 컴퓨터공학과 학사
1986 서울대학교 컴퓨터공학과 석사
1991 UCLA 컴퓨터학과 박사
1991.3～1991.9 UCLA 연구원
1991～1996 포항공과대학교 조교수
1997～2003 포항공과대학교 부교수
2000～2001 Stanford CSLI 연구원
2004～현재 포항공과대학교 정교수
관심분야 : Natural Language Processing, Speech Recognition and Synthesis, Statistical Machine Translation, Question Answering and Web/Text Mining, Spoken Dialog System
E-mail : gblee@postech.ac.kr

KCC 2007(한국컴퓨터종합학술대회)

❑ 일　　자 : 2007년 6월 25일~27일
❑ 장　　소 : 무주리조트
❑ 내　　용 : 학술발표 등
❑ 주　　최 : 한국정보과학회
❑ 상세안내 : http://www.kiss.or.kr/conference02/index.asp