

# HMM 에 기반한 연속음성인식에서의 형태소 그래프 생성

최준기 이근배 이종혁  
포항공과대학교 전자계산학과

## Morpheme Graph Generation with HMM based Continuous Speech Recognition

Joonki Choi Geunbae Lee Jong-Hyeok Lee  
Dept. of Computer Science and Engineering  
Pohang University of Science and Technology  
e-mail : nuno@postech.ac.kr

### 요 약

본 논문에서는 형태소 그래프를 정의하고 이를 한국어 연속 음성 인식의 결과로서 사용함과 동시에 한국어의 자연어 처리를 위한 지식 표현 방법으로 사용한다. 또한 형태소 그래프를 연속 음성 인식과정에서 효율적으로 생성하는 알고리즘으로서 Tree-Trellis 탐색 알고리즘을 소개한다. 한국어 연속 음성 인식기는 HMM 인식기를 사용하며 탐색 알고리즘 또한 HMM 음소 인식기의 사용을 전제로 한다. 실험 DB로는 한국과학기술원 통신연구실에서 제작한 3000 단어급의 무역상담관련 DB를 사용하였다.

### 1. 서론

최근의 음성 인식 기술은 짧은 키워드와 짧은 키워드들의 조합으로 GUI와 특정 어플리케이션 프로그램의 음성 명령이나 음성 입력으로 사용할 수 있도록 한다[1][2]. 짧게 발음되는 키워드보다 진보된 형태의 인터페이스인 연속 음성과 인간의 자연스러운 대화형 발화를 인식하기 위해서는 인간의 언어에 대한 기술, 즉 자연어 처리 기술이 음성 인식 기술에 도입이 되어야 한다.

자연어 처리 기술이 연속 음성 인식에 도입되는 기존의 방법들은 연속 음성 인식의 후처리, 즉 문장의 재구성 과정에서 Bigram이나 Trigram 등의 자연어 지식이 적용되는 방법, 다수의 문장 인식 후보를 출력하고 그 문장 인식 후보들의 개수를 줄이는 과정에서 자연어 처리 기술이 적용되는 방법 등이 있다. 그러나 위와 같은 방법들은 해당 코퍼스에만 국한되고 일반적이지 못하며 음성 인식과 자연어 처리가 밀접하게 결합(strongly coupled)되어 있지 못하다는 단점을 가지고 있다.

음성 한국어의 자연어 처리를 위해서는 형태소 분석이 선행되어야 한다. 따라서 연속 음성 인식과 형태소 분석이 결합되는 것이 바람직하다. HMM을 이용한 대용량 어휘 연속 음성 인식에

서는 문장의 재구성 과정, 즉 HMM의 확률에 기반해서 lexicon들을 탐색하는 과정에서 형태소 분석이 이루어질 수 있으며 품사 태깅기술을 이용해서 연속 음성 인식의 후처리에 이용되는 언어 모델을 만들어 줄 수 있다. 위 방법을 사용하면 연속 음성 인식과 자연어 처리 기술이 밀접하게 결합된다.

본 논문에서는 연속 음성 인식에서 범용의 한국어의 형태소 접속 검사를 수행하는 탐색 알고리즘을 제시하고 연속 음성 인식의 출력 결과로서 형태소 그래프를 정의한다. 그리고 형태소 그래프를 생성하기 위한 한국어 형태소 및 음운 접속 검사와 음소열 형태소 사전과 음운 접속 검사 테이블에 대해 설명한다. 형태소 그래프는 N-Best list에 비해 메모리를 적게 차지하고 구문 분석이나 의미 분석과 같은 상위의 자연어 처리의 입력으로 바로 사용이 가능하다.

본 논문에서 제안한 형태소 그래프를 출력하는 연속 음성 인식기는 동시 통역기와 같은 어플리케이션에 응용이 가능하다. 실제로 연속 음성 인식단계에서 인식되는 단어에 Class label을 붙여서 동시 통역기의 효율을 높이고자 하는 시도가 있다[3].

본 논문의 구성은 다음과 같다. 2장에서 기본적인 개념들에 대하여 설명하고 3장에서는 시스템의 구성과 알고리즘에 대해 설명하고 4장에서 구

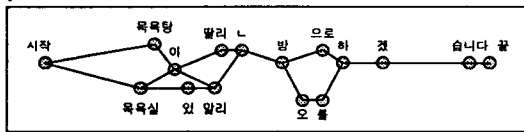
현 결과 및 결론을 맺는다.

## 2. 기본 개념 소개

본 장에서는 본 논문에서 사용되는 몇 가지의 개념들의 정의를 내리고 설명한다.

### 2.1 형태소 그래프의 정의

형태소 그래프는 Vertex가 형태소이며 Edge가 형태소 간의 Connectivity를 나타내는 그래프이다. 형태소 그래프의 예는 아래 [그림 1]과 같다



[그림 1] 형태소 그래프의 예

형태소 그래프의 Edge의 길이는 특별히 시간축의 정보를 가지고 있지 않으며 Vertex의 위치도 형태소 lexicon에 대한 정보를 가지고 있지 않다. 형태소 그래프는 [4]에서 사용된 것과 동일하며 Segmentation 정보 및 각 형태소의 태그가 포함되게 된다. 형태소 그래프는 인도-유럽어 언어에서 사용되는 단어 그래프와 유사하며 한국어에 적합한 개량형이라고 생각할 수 있다.

### 2.2 음소열 형태소 사전

한국어의 특징상 한 형태소가 여러 발음을 가지고, 반대로 한 발음이 여러 형태소를 나타내는 경우가 있다. 이런 현상을 고려하기 위해서 형태소와 형태소의 여러 가지 발음을 코딩한 것이 음소열 형태소 사전이다. [그림 2]는 음소열 형태소 사전의 한 부분이다. [7]

품사태그<원형>(변이형)	발음	접속정보
eGSm<겠>(겠)	/ㄱ ㅈ ㅌ/	<ㄱ   ㅈ: ㅌ >
eGSm<겠>(겠)	/ㄱ ㅈ ㅌ/	<ㄱ   ㅈ: ㅌ >
eGSm<겠>(겠)	/ㄱ ㅈ ㅌ/	<ㄱ=ㄱ   ㅈ >
eGSm<겠>(겠)	/ㄱ ㅈ ㄴ/	<ㄱ   ㅈ: ㄴ >

[그림 2] 음소열 형태소 사전

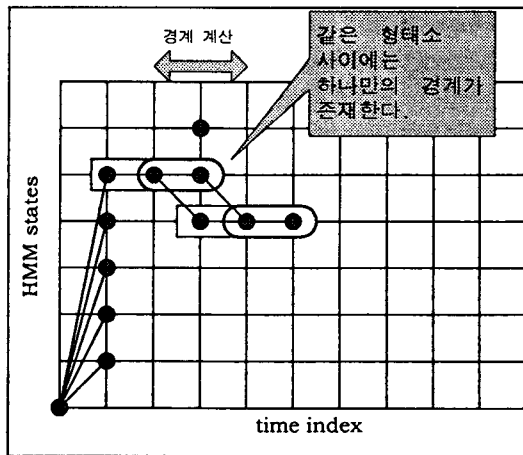
음소열 형태소 사전은 먼저 해당 코퍼스로부터 자동으로 추출되고, 몇 가지의 규칙에 의해서 튜닝된다.

### 2.3 형태소 접속 검사

본 논문에서 사용된 HMM 연속 음성 인식기는 HMM Trie 사전의 state들을 탐색해가면서 형태소를 추출한다. 그리고 추출되는 형태소들 간의 접속 검사를 행하여 언어모델로서 사용하고 동시에 한국어의 형태소 분석을 수행한다. [그림 2]에 나타난 바와 같이 HMM Trie Lexicon 사전에는 형태소 표제어와 각 형태소의 태그도 같이 명시되어 있다. 형태소의 태그는 [4]을 따르며 음운 접속 검사는 한국어 표준 발음법[10]에 기본을 둔다. 한국어 표준 발음법에 기반한 음운 접속 검사 테이블을 만들고 탐색 과정에서 두 음소의 접속 여부를 검사한다. 이때 음소열 형태소 사전의 음운 접속 정보가 사용된다.

### 2.4 Morph Pair Approximation

연속 음성은 음소나 형태소 candidate들간의 경계를 정하기가 어렵다. 형태소들이 연속해서 발음되기 때문에 두 형태소 간의 경계가 모호해지기 때문이다.[8] 탐색 과정에서 형태소 candidate 간의 경계가 정확하게 정해지지 않으면 형태소 그래프의 같은 Partial-path를 중복해서 계산하게 된다. 형태소 그래프의 Partial-path를 중복 계산하게 되면 인식 시간이 길어지고 형태소 그래프의 크기도 커진다[5]. 이런 중복 계산을 피하기 위하여 Word Pair Approximation이 적용된다. 형태소의 경우에도 Word Pair Approximation이 적용 가능한데 이것을 Morph Pair Approximation이라 한다. 즉 하나의 형태소 candidate가 time t에 끝났을 경우, 이전의 형태소 candidate와의 경계는 보다 이전의 Partial-Path에 종속되지 않는다. 따라서 같은 형태소 pair끼리는 고정된 하나의 경계만을 가진다.



[그림 3] Morph pair approximation

[그림 3]에서 보는 것처럼 한 형태소 candidate

와 이전의 형태소 후보와의 경계는 단 하나로만 정해지게 된다.

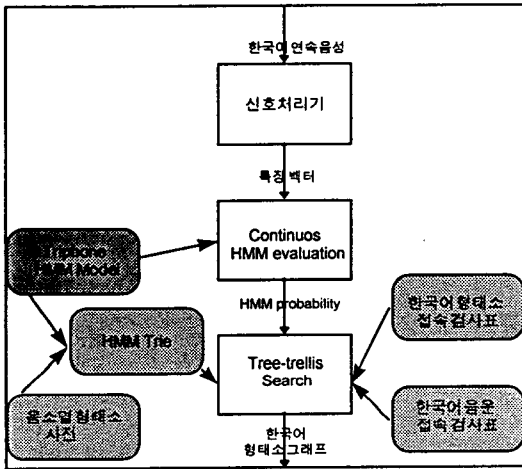
### 3. 시스템 설명

#### 3.1 전체 시스템의 개요

시스템의 기본적인 구조는 [그림 4]와 같다. 각각의 세부 시스템에 대해 살펴본다.

##### 3.1.1 신호 처리기

한국어 연속 음성을 입력으로 받아서 전처리 과정을 거치고 음성의 feature를 뽑아주는 부분이다. 이 부분과 HMM의 훈련, evaluation은 CSLU Toolkit<sup>1</sup>를 사용하였다. CSLU Tool Kit은 소스가 공개된 freeware이다.



[그림 4] 시스템의 개요

입력 음성에 대해서 Direct current offset을 제거하고 특징 벡터로서 8차원의 PLP 계수와 이것의 1st order derivative를 음성의 특징 벡터로 사용한다. [9]

##### 3.1.2 연속 HMM evaluation

입력 음성이 특징 벡터로 변환 되면 이 특징 벡터를 이용해서 Triphone model HMM의 evaluation을 진행한다. HMM Trie 사전의 state들을 따라가면서 HMM model들을 evaluation하는 과정을 일반적으로 탐색 과정이라고 한다. 본 논문에서 사용된 HMM은 Triphone Model을 기본으로 한다. 그러나 모든 Triphone들이 충분히 훈련될 수 있을 정도로 음성 DB가 크지 않

때문에 Generalization이나 Clustering이 필요하다. 본 논문에서 사용된 Triphone HMM Model은 모두 746개로서 음운 규칙에 의해서 훈련 과정 이전에 클러스터링 되어 훈련된다. 신호 처리기와 마찬가지로 이 부분도 CSLU Toolkit을 이용해서 HMM을 훈련하고 evaluation하게 된다.

##### 3.1.3 Tree-Trellis Search

HMM의 확률값과 HMM Trie 사전을 이용해서 형태소 candidate를 추출하고 형태소 그래프를 만들어가는 과정이다. 이 알고리즘과 탐색 과정에서 사용되는 기법에 대한 자세한 설명은 다음 장에서 설명된다.

#### 3.2 형태소 그래프 생성

HMM에 기반한 연속 음성 인식기는 기본적으로 Viterbi decoding을 이용해서 word hypothesis를 출력하게 된다. Viterbi decoding을 그대로 사용할 경우, 1개의 최고 확률값을 가지는 sentence hypothesis를 구해 줄 수 있으나 자연어 처리 기법이 사용되지 않았기 때문에 인식률이 매우 낮으며 결과로서 의미가 없다. 따라서 일반적으로 인식률을 높이고 인식 시간을 줄이기 위해서 Multi-Pass의 탐색 알고리즘이 많이 사용된다[5][6][8]. 본 논문에서 사용된 탐색 알고리즘은 기본적으로 2번의 탐색 과정을 가지는 Tree-Trellis 탐색 알고리즘[11]의 변형된 형태를 사용한다. 처음 탐색 과정은 Viterbi decoding의 전향 탐색과 유사하게 Morph lattice를 출력하는 과정이며 두 번째 과정을 A\* 탐색을 응용한 탐색 과정이다. 두 번째의 탐색 과정이 끝나면 결과로서 형태소 그래프가 얻어지게 된다. 각각에 대한 알고리즘을 살펴 보자.

##### 3.2.1 전방향 탐색 과정

첫번째 탐색 과정은 입력으로서 HMM의 확률값을 가지고 Morph pair lattice를 출력한다. 그 과정에서 형태소 후보들을 추출하며 형태소 후보들의 접속 검사를 수행한다. 그리고 Beam pruning을 거치고 남은 morph pair에 대해서 각 형태소 쌍의 경계와 누적 확률값, 형태소 쌍 자체만의 확률값을 기록한다. 이 기록은 두 번째 탐색 과정에서 사용된다. 알고리즘은 [그림 5]와 같다. 알고리즘에서 사용된 확률과 형태소의 경계, 점수를 구하는 방법은 [6]에서 소개된 식의 변형이다.

누적 확률 Q는 다음과 같이 구해진다.

$$Q_v(t, s) = \max_{\sigma} \{q(x_t, s | \sigma) \cdot Q_v(t-1, \sigma)\}$$

이 때 Q는 t로 끝나는 마지막 state가 s인 형

<sup>1</sup> The CSLU Toolkit, Center for Spoken Language Understanding Oregon Graduate Institute of Science & Technology

태소 후보  $v$  까지 오는 partial path 의 최고 확률 값 이다. 또한  $q$  는 HMM 의 transition probability 와 emission probability 의 곱이며,  $\sigma$  는 최고 확률 값을 가지는 Partial Path 의 state sequence 를 나타내는 것이다.

시간  $t$  축에 대해서 **time synchronous** 진행

**Acoustic level : HMM Trie 사전의 state 들에 대해**

- 누적 확률  $Q$  를 구한다
- 현재 형태소에 대한 경계를 기록한다
- **Beam** 을 적용한다

**Morph pair level : 현재 Trie 사전의 state 가 한 형태소의 끝인 경우**

- 모든  $(t;v,w)$  에 대해 다음을 기록한다  
형태소의 경계  
형태소의 점수

[그림 5] First Stage 알고리즘

형태소의 경계는 Morph Pair approximation 을 생각하면 간단히 구해진다. 지금 고려되고 있는 형태소 후보의 앞에 존재하는 형태소 후보와의 여러 경계 중에서 현재의 형태소 후보가 최고의 누적 확률 값을 가지게 하는 경계를 선택한다. 이것을 수식으로 표현하면 아래와 같다.

$$\tau(t;v,w) = \tau_v(t-1, \sigma_v^{\max}(t,s))$$

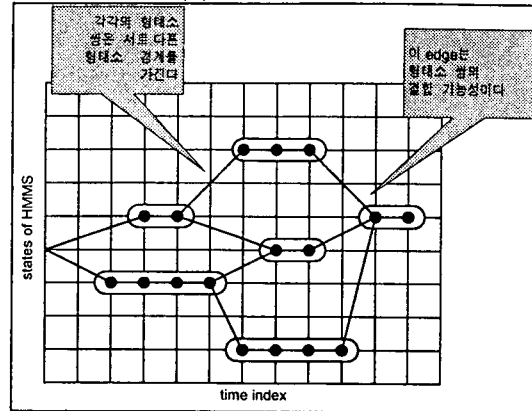
즉, time  $t$  에 끝나며 이전 형태소 후보가  $v$  인 형태소 후보  $w$  는  $v$  가 최고의 확률 을 가지는 시점의 형태소 경계를 그대로 이어 받는다는 뜻이다.

이러한 과정의 첫번째 탐색 과정을 거치면 형태소 쌍이 존재하는 trellis 가 출력된다. 이것을 그림으로 도식하면 [그림 6]와 같다.

### 3.2.2 후방향 탐색

이 탐색 과정은 입력을 첫번째 탐색 과정의 출력인 형태소 쌍의 lattice 로 받고, 형태소 그래프를 출력한다. 입력으로 받은 lattice 는 모든 형태소 후보들이 쌍을 이루고 있어서  $A^*$  탐색을 적용하는 것이 가능하다. 첫번째 탐색 과정과는 달리 두 번째 탐색 과정은 Time asynchronous 이며 right to left 로 진행이 된다. 각각의 형태소 쌍들이 가지고 있는 누적 확률 값을  $A^*$  탐색의

heuristic value( $h$ )로 사용하고 형태소 확률 을  $A^*$



[그림 6] Morph Pair Lattice

탐색의  $g$  값으로 사용한다( $f = g+h$ ).  $A^*$  탐색의 장점은 optimal 탐색을 가능하게 하며 bigram 보다 상위의 언어 모델을 적용하는 것을 가능하게 한다[8].

두 번째 탐색 과정은 rescoring 을 주목적으로 하고, 해당하는 형태소 쌍으로만 탐색 공간이 제한되기 때문에 첫번째 과정에 비하여 수행 시간이 매우 짧게 된다. 일반적인 N-best list 탐색 기법에서는 second phase 가 N 개로 결과를 제한하는 구실을 한다. 여기서도 마찬가지로 Beam 을 적용해서 최종 출력인 형태소 그래프의 크기가 커지는 것을 막는다. 형태소 그래프의 크기가 커질 경우 결과를 분석하는 작업이 어려워지며 상위 자연어 처리의 입력으로 쓰기에 어려워지는 단점이 있다.

## 4. 실험 및 결론

### 4.1 기본 단어 인식 실험

본 논문에서 사용된 신호 처리기와 HMM 음성 인식기의 성능을 평가하기 위해서 소규모 단어의 단어 인식 실험을 수행하였다. 사용된 음성 DB 는 ETRI 에서 제작된 PBW 445DB 를 이용해서 실험하였다. 이 실험에서는 Phone Model HMM 을 사용해서 실험하였으며 아무런 후처리를 가지 않은 상태에서 74%의 단어 인식률을 보였다.

### 4.2 형태소 그래프 생성 실험 (진행중)

#### 4.1 실험에 사용된 음성 DB

본 논문의 실험을 위해서 사용된 음성 DB 는 한국과학기술원 통신연구실에서 제작된 3000 단어급 무역 상담 DB 이다. 이 DB 는 10~20 개의 형태소로 이루어진 연속 음성이며 조용한 환경에

서 녹음된 대화형의 문장들로 구성된다. HMM의 훈련을 위해서 남성 54명의 5300문장을 사용하였으며 실험용으로 남성 10명의 1000문장을 사용했다. 모든 음성은 16Khz로 샘플링되었다.

#### 4.2 샘플 실험 결과

본 논문의 실험을 위해서 작성된 음소열 형태소 사전은 1890여 개의 형태소 entry를 가지고 있다. 음운 접속 정보 테이블은 모두 35항으로 이루어진 표준 발음법에 의거한다. 이때 적절한 Beam Pruning에 의해 완성된 형태소 그래프에 형태소 분석 결과가 맞는 path가 존재하는 확률은 소규모의 샘플 실험 결과 75%~80%로 관측되었다. 여기에 morpheme bigram 언어모델을 적용했을때는 훨씬 더 높아질 것으로 기대된다. 현재 1000문장 전체에 대한 실험을 진행중이다

#### 4.3 결론

본 논문에서는 한국어 연속 음성 인식의 결과로서 형태소 그래프를 정의했고 형태소 그래프를 생성하는 알고리즘에 대해서 기술하였다. 형태소 그래프는 N-best list에 비하여 compact하게 표현되며 상위의 자연어 처리 루틴에 다른 추가 작업 없이 바로 입력으로 사용될 수 있는 장점이 있다. 본 논문에서는 한국어 연속 음성 인식 중에 형태소 접속 검사를 수행하여 범용의 한국어 음성 인식 후처리를 설계하였으며 태깅 확률을 탐색 과정에 추가할 경우 품사 태깅이 음성 인식 중에 이루어질 수 있다.

### 5. 참고 문헌

[1] Xuedong Huang et al, "Microsoft Windows highly Intelligent Speech Recognizer: WHISPER", IEEE ICASSP 1995  
 [2] <http://www.software.ibm.com/is/voicetype>  
 [3] Toshiyuki Takezawa et al, "Conversational Speech Recognition Using Subtree-based CFG Rules and Class Based Bigram Models", 1997 ICSP, 355p~360p  
 [4] 차정원, 이원일, 이근배, 이종혁, "형태소 패턴 사전을 이용한 일반화된 미등록어처리", 인공지능연구회 1997년도 춘계학술발표 논문집, 37p~41p  
 [5] Tohru Shimizu et al, "Spontaneous Dialogue Speech Recognition Using Cross-Word Context Constrained Word Graphs", IEEE ICASSP 96, 145p~148p  
 [6] Stefan Ortmanns et al, "A Word Graph Algorithm for Large Vocabulary Continuous

Speech Recognition", Computer Speech and Language(1997) 11, 43p~72p

[7] 최준기, 이근배, 이종혁, "Tree-Trellis Search와 형태소 그래프를 이용한 한국어 연속 음성 인식기", HCI 97 학술대회 발표 논문집, 303p~307p  
 [8] Mosur K.Ravinshankar, "Efficient Algorithms for Speech Recognition", CMU-CS-96-143, Ph.D thesis, School of Computer Science Computer Science Division, CMU, 1996,  
 [9] The CSLU Toolkit, Center for Spoken Language Understanding, Oregon Graduate Institute of Science & Technology  
 [10] 국어 어문 규정집, 대한교과서주식회사  
 [11] Frank K. Soong and Eng-Fong Hwang, "A Tree-Trellis Based Fast Search for Finding the N Best Sentence Hypothesis in Continuous Speech Recognition", IEEE ICASSP 1991, 705p~708p