

본 챕터에서는 지금까지의 챕터에서 설명한 각종 기법을 취합하여 대어휘 연속 음성 인식 (large vocabulary continuous speech recognition; LVCSR) 엔진을 구성하는 방법에 대해 설명한다.

## 0.1 FST의 합성과 확률모델

지금까지 도입했던 음성인식의 요소를 나타내는 FST를 합성하고, 합성한 FST의 최단 경로 문제의 풀이로서 음성인식결과를 얻는 방법을 고찰한다. 음성인식의 통계모델을 표현하는 FST는 일반적으로는 아래의 4 종류가 있다.

- $G$ : 단어열 Acceptor (6.5)
- $L$ : 문맥에 의존하지 않는 음소열로부터 단어열 변환 (4.2.2)
- $C$ : 문맥에 의존하는 음소열로부터 문맥에 의존하지 않는 음소열로 변환 (5.3)
- $H$ : HMM state 시퀀스로부터 문맥에 의존하는 음소열로 변환 (5.3)

여기에 덧붙여서, 이하의 식에서 보여지고 있는 입력 (관측 벡터열)과 HMM 상태변수의 관계를 나타내는 FST  $E$ 를 가상으로 도입함으로써, 인식할 때의 처리를 모두 FST 형식으로 기술할 수 있게 된다.

$$\begin{aligned} Q[E] &= \{0, \dots, T\}, & I[E] &= \{(0, \bar{1})\}, & F[E] &= \{(T, \bar{1})\}, \\ E[E] &= \{(t-1, t, \sigma, \sigma, -\log p(\mathbf{x}_t | s_t = \sigma)) : t \in 1, \dots, T, \sigma \in \Sigma[H]\} \end{aligned} \quad (1)$$

여기에서  $T$ 는 관측 벡터열의 길이를 의미한다.

### 0.1.1 디코딩 네트워크의 구성과 탐색오류

### 0.1.2 disambiguation 심볼

## 0.2 대어휘 연속 음성인식의 탐색문제

## 0.3 대규모 FST 합성 기술

### 0.3.1 온 더 플라이 합성

### 0.3.2 디스크 기반 인식 시스템

## 0.4 N-Best 리스트 및 lattice 생성

### 0.4.1 lattice 생성

### 0.4.2 lattice로부터 N-Best 리스트 생성

## 인용 및 참고문헌