

퍼지양자화 은닉 마르코프 모델에서 코드워드 종속 거리 정규화와 Instar 형태의 퍼지 기여도에 기반한 출력확률의 평활화

Codeword-Dependent Distance Normalization and Smoothing of Output Probabilities Based on the Instar-formed Fuzzy Contribution in the FVQ-DHMM

최 환 진*, 김 연 준*, 오 영 환*

(Hwan Jin Choi*, Yeon Jun Kim*, Yung Hwan Oh*)

요 약

본 논문에서는 FVQ-DHMM(fuzzy vector quantization-discrete hidden Markov model)에서 강인한 출력확률의 추정을 위해서 코드워드 종속 거리 정규화와 출력확률에 대한 instar 형태의 퍼지 평활화 방법을 제안한다. FVQ-DHMM은 DHMM의 변형된 모델로, 상태별 출력확률이 입력패턴에 대한 각 코드워드와의 가중치와 출력확률의 곱에 대한 합의 형태로 추정된다. FVQ-DHMM의 성능이 가중치 요소와 상태별 출력분포에 영향을 받으므로, 가중치 요소와 상태별 출력분포를 강인하게 추정하는 방법이 필요하게 된다.

실험결과, 제안된 코드워드 종속 거리 정규화(CDDN: codeword dependent distance normalization)를 적용한 방법이 기존의 FVQ-DHMM에 비해 24%의 오인율을 감소가 있었으며, 상태별 출력분포에 대해서 평활화를 적용한 경우 79%의 오인율을 감소 시킬 수 있었다. 이러한 결과는 제안된 CDDN과 퍼지 평활화의 사용이 향상된 인식율을 얻는데 주요하며, 결과적으로 제안된 방법이 FVQ-HMM을 위한 강인한 출력확률의 추정을 위한 대안으로 유용함을 보여준다고 할 수 있다.

ABSTRACT

In this paper, a codeword-dependent distance normalization(CDDN) and an instar-formed fuzzy smoothing of output distribution are proposed for robust estimation of output probabilities in the FVQ(fuzzy vector quantization)-DHMM(discrete hidden Markov model). The FVQ-DHMM is a variant of DHMM in which the state output probability is estimated by the sum of the product of the output probability and its weighting factor for each codeword on an input vector. As the performance of the FVQ-DHMM is influenced by weighting factor and output distribution from a state, it is required to get a method to get robust estimation of weighting factors and output distribution for each state.

From experimental results, the proposed CDDN method has reduced 24% of error rate over the conventional FVQ-DHMM, and also reduced 79% of error rate when the smoothing of output distribution is also applied to the computation of an output probability. These results indicate that the use of CDDN and the fuzzy smoothing of output distribution to the FVQ-DHMM lead to improved recognition, and therefore it may be used as an alternative to the robust estimation of output probabilities for HMMs.

I. 서 론

음성은 인간의 기본적인 의사소통을 위한 수단으로, 그 사용의 편리함으로 인해서 어느 다른 방법들에 비해

서 널리 사용되고 있다. 근간에와서 음성처리와 관련된 기술의 발전, 음성인식을 위한 인지모델, 디지털 처리기술의 발전등에 힘입어 음성 인식에 대한 관심이 점차 증대되고 있는 실정이다.

음성인식은 음향학, 음성학, 언어학 등의 다단계의 지식원을 이용하여 화자에 의해서 발생된 음성으로부터 사람이 이해할 수 있는 표현 방법으로 변환하는 일련의 과정이다. 그러나, 발생상의 변이, 정보의 중복성, 그리고

*한국과학기술원 전산학과/인공지능연구센터
접수일자: 1996년 11월 26일

많은 계산량의 요구등으로 인해서 음성인식은 여전히 어려운 문제로 남아있다[1].

음성인식을 위해서 여러 방법들이 사용되고 있으며, 이 중 가운데에서 은닉 마르코프 모델(HMM: hidden Markov model)에 의한 방법이 우수한 성능을 보이고 있다. 은닉 마르코프 모델은 모델의 출력확률 추정에 따라 이산형(discrete)과 연속형(continuous) HMM으로 나뉜다. 연속형의 경우 학습자료에 따라서 추정 파라미터의 변동이 크며, 많은 계산시간을 요한다는 단점이 있으나 좋은 성능을 나타내고 있다. 반면, 이산형의 경우 적은 계산시간과 적은 자료에 의한 추정에서 성능이 크게 저하되지 않는 장점이 있다. 그러나, 입력패턴에 대한 양자화에 의한 인식을 저하가 일어난다. 은닉 마르코프 모델에서의 이러한 다양한 문제들을 해결하기 위해서, 연속 분포 함수를 이용한 상태의 표현[2][3], 불충분한 학습자료에 의한 부적절한 출력확률의 보정을 위한 출력확률의 평활화[4][5][6], 그리고 인식 모델간의 변별력 학습[7][8]등 다양한 연구들이 수행되고 있는 실정이다.

본 논문에서는 HMM모델의 하나인 FVQ-DHMM(fuzzy vector quantization-discrete hidden Markov model)에서 각 상태에서 강인한 출력확률의 분포를 추정하기 위한 방법으로 입력 패턴과 코드워드 간의 신뢰성 있는 거리 측정과 이와 아울러 각 상태별 출력확률을 평활화하는 방법으로, 코드워드 의존 거리 정규화(CDDN: codeword-dependent distance normalization) 방법과 클래스기반 퍼지 평활화(CFS: class-based fuzzy smoothing)방법의 하나인 instar 형태의 CFS 방법을 제안하고자 한다.

FVQ-DHMM은 DHMM을 변형한 모델로, 각 상태에서의 출력확률이 주어진 입력 패턴과 가까운 K개의 코드워드들에 대한 각각의 출력확률과 코드워드와 입력 패턴 간의 거리를 가중치로 하는 선형 결합의 형태로 계산된다. 몇몇 실험결과들에 의하면[9], FVQ-DHMM이 기존의 DHMM에 비해서 우수한 성능을 나타내고 있으며, 그러한 성능향상은 다수의 출력확률과 그들의 가중치에 의한 출력확률의 계산으로 인해, 기존의 입력패턴에 대한 출력확률을 입력패턴과 가장 유사한 하나의 코드워드의 출력확률로 대치하여 사용하던 방법에 비해서 좋은 성능을 보여주고 있다. 이러한 입력패턴에 대한 다수의 코드워드 출력확률의 고려는 입력패턴에 의한 양자화 오류를 줄여주는 효과를 갖는다.

일반적으로, 이산형 HMM에서 입력패턴에 대한 정확한 출력확률을 추정하기는 어렵다. 그 이유는 입력패턴이 코드북에 의해서 양자화되며, 양자화에 의해서 대응되는 코드워드의 출력확률이 양자화 오류와는 상관없이 입력패턴에 대한 출력확률로 대치되어 사용되기 때문이다. 이러한 문제를 해결하기 위해서 FVQ-DHMM은 입력패턴에 대해서 K개의 근접한 코드워드를 계산하고, 그들의 출력값과 각 코드워드와 입력패턴과의 거리에 대한 함수값을 각 코드워드의 출력확률에 대한 가중치로 사용

하며, 최종적인 출력확률을 계산하게 된다. 따라서, 이러한 추정방법은 기존의 DHMM에 의한 방법과는 다르다.

FVQ-DHMM이 위에서 언급한 몇 가지 장점들을 가지고 있음에도 불구하고, 인식을 향상이라는 측면에서 몇 가지 고려할 사항들이 있다. FVQ의 성능은 크게 직렬한 가중치 값들을 추정하는 것과 강인한 출력분포의 추정이라는 두 가지 요소에 의존한다. 첫 번째 요소로써, FVQ-DHMM에서 가중치 요소의 중요성은 매우 크다. 그 이유는 가중치가 각 대응되는 코드워드 출력확률의 중요도로서 다르게 반영하기 때문이다. 거리의 분포가 각 코드워드마다 다르며, 그러한 분포특성이 거리계산과정에 반영되어야 한다는 가정 하에, 코드워드 의존 거리 정규화(CDDN: codeword-dependent distance normalization)를 제안한다. 이러한 요소와 더불어, 상태별로 강인한 출력확률을 추정하는 것이 필요하다. 기존 DHMM의 경우, 충분하지 못한 학습자료가 종종 unseen 데이터에 대한 zero 확률 값으로 인해서 인식 시에 성능을 저하시키는 경우가 있다. 이러한 문제를 해결하기 위해서, 다양한 출력확률에 대한 평활화 방법들이 제안되어 사용되고 있으며, 몇몇 실험결과 이러한 평활화로 인해서 인식을 향상시키는 것으로 알려져 있다. 본 논문에서는 평활화를 위한 또 다른 방법으로 클래스 의존 퍼지 평활화(CDFS: class-dependent fuzzy smoothing)를 제안한다[10]. CDFS는 특징파라미터 클래스들간의 거리에 기반하여 평활화 행렬을 유도하는 것으로, 이를 위해서 퍼지복합함수를 사용한다. 평활화는 상태별로 이루어지며, 실험결과 제안된 평활화방법이 다른 기존의 방법들에 비해서 향상된 인식을 나타내는 것을 알 수 있었다.

본 논문의 구성은 다음과 같다. 2장에서는 DHMM과 FVQ-DHMM의 정의에 대해서 설명하고, 본 논문의 연구동기에 대해서 기술한다. 3장에서는 본 논문에서 제안된 CDDN과 출력확률 평활화를 위한 CFS에 대해서 설명한다. 4장에서는 실험환경과 실험결과에 대해서 기술하고, 마지막 5장에서는 결론과 향후 연구방향에 대해서 기술한다.

II. FVQ-DHMM의 개요

HMM에 기반한 음성인식에 관한 연구가 1970년에 시작된 이래, HMM은 음성 인식 방법들 가운데에서 가장 우수한 성능을 나타내는 것으로 알려져 있다. HMM의 기본적인 가정은 음성신호는 랜덤 프로세스로 특징 지을 수 있으며, 통계적인 프로세스의 파라미터들을 정확한 방법에 의해서 결정할 수 있다는데 있다. 그리고, 몇몇 실험결과에서 알려진 바와 같이, HMM은 음성인식을 위한 자연스러우며 높은 신뢰성을 갖는 방법을 제공하며, 아울러 인식하고자 하는 영역의 문장구조(syntax)나 의미구조(semantic)들과 통합이 손쉬운 장점을 가지고 있다. 이

절에서는 제안된 방법들의 기본 시스템으로 사용될 FVQ-DHMM에 대해서 기술하고, 아울러 연구배경과 동기에 대해서 설명하고자 한다.

2.1 이산형 DHMM의 정의

HMM은 통계적인 방법의 하나로, 입력된 음성에 대한 음성 파라미터들의 열을 전이를 갖는 상태들의 열로써 표현한 모델이다. HMM의 기본적인 구성요소는 다음과 같다.

음닉마르코프모델 λ 는 5가지 구성요소 (S, V, π , A, B)로 구성되어 있다. S는 N개의 상태들인 s_1, s_2, \dots, s_N 을 나타낸다. V는 M개의 출력 심벌들인 $\bar{v}_1, \bar{v}_2, \dots, \bar{v}_M$ 을 나타낸다. π 는 초기출력확률을 나타내는 벡터인 ($\pi_1, \pi_2, \dots, \pi_N$)을 나타내며, $\pi_i = P(s_i(1))$ 을 의미한다. A는 상태전이 확률을 나타내는 행렬로, $A = [a_{ij}]$, $1 \leq i, j \leq N$ 을 나타내며, $a_{ij} = P(s_j(t+1)|s_i(t))$ 을 의미한다. B는 관측심벌에 대한 출력확률로, $B = [b_{ij}]$, $1 \leq i \leq N, 1 \leq j \leq M$ 을 나타내는 행렬을 나타내며, $b_{ij} = P(\bar{v}_j(t)|s_i(t))$ 을 의미한다. 관측열은 $O = (o_1, o_2, \dots, o_T)$ 을 나타내며, $o_t \in V$ 이다.

인식모델 λ 가 주어졌을 때, 관측열 O의 출력확률은 전향-후향방법(forward-backward)이나 Viterbi알고리즘을 이용해서 구해진다. 전향-후향방법은 EM(estimation maximization)에 기반한 방법으로, 모든 가능한 상태열들에 대해서 동시에 최대 확률값을 갖도록 모델의 파라미터를 갱신한다. 이에 반해서, Viterbi방법은 동적 프로그래밍 방법에 기반하여 최적 상태열에 대해서 출력확률이 최대화되도록 학습한다. 3개의 상태를 갖는 음소 모델과 음소 모델의 접속을 통해서 얻어진 단어모델의 예는 그림 1과 2와 같다.

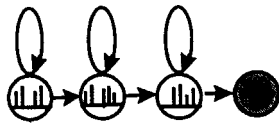


그림 1. 상태간 jump가 없는 세계의 상태를 갖는 음소모델

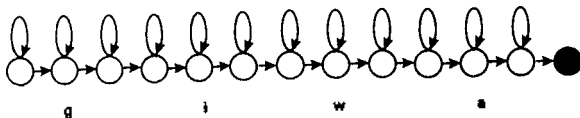


그림 2. 음소모델들을 접속한 단어모델

2.2 FVQ에 기반한 DHMM

FVQ-DHMM[8]은 DHMM의 변형된 모델로, 한 상태에서의 출력확률은 입력패턴에 대해서 K개의 근접 코드워드들에 대한 출력확률과 각 코드워드에 대한 가중치들

의 선형 결합형태로 추정된다. FVQ-DHMM의 가중치 값은 퍼지 목적 함수에 의해서 유도되는데, 이 함수의 인수로써는 입력패턴과 해당 코드워드간의 거리를 사용한다. FVQ-DHMM의 정의는 아래와 같다.

FVQ는 mixture Gaussian VQ(vector quantization)의 단순화된 형태로 볼 수 있다. $d(\bar{x}_i, \bar{v}_j)$ 는 입력패턴 \bar{x}_i 와 코드워드벡터 \bar{v}_j 간의 Euclidean 거리를 나타낸다. FVQ는 입력패턴 \bar{x}_i 에 대해서 아래의 수식에 의해 출력 벡터 $\bar{o}_i = (m_{i1}, m_{i2}, \dots, m_{iM})$ 를 얻는다.

$$m_{ij} = \left[\sum_{k=1}^M [d(\bar{x}_i, \bar{v}_j)/d(\bar{x}_i, \bar{v}_k)]^{1/(F-1)} \right]^{-1}$$

where, $d(\bar{x}, \bar{v}) = \sum_{p=1}^P (x_p - v_p)^2$ (1)

위의 수식(1)에서, F는 퍼지 정도(degree of fuzzy)를 나타내는 상수이며, 벡터 \bar{o}_i 는 아래의 퍼지 목적 함수를 최소화하는 m_{ij} , $1 \leq j \leq M$ 으로 결정된다.

$$\sum_{i=1}^T \sum_{j=1}^M m_{ij}^f \cdot d(\bar{x}_i, \bar{v}_j)$$
 (2)

\bar{o}_i 를 구성하는 각 요소들은 양의 실수값을 가지며, 그 합은 1이다.

관측값이 퍼지값일 때, 관측열은 확률 벡터들의 열이다. 퍼지 관측열을 O로 정의하면, O를 구성하는 \bar{o}_i 는 $\bar{o}_i = (m_{i1}, m_{i2}, \dots, m_{iM})$ 로 나타낸다. 이러한 퍼지 관측열을 출력확률의 계산식에 반영하기 위해서 기존의 출력식을 아래의 식과 같이 변형한다.

$$\omega_t(i) = \sum_{j=1}^M m_{ij} \cdot b_{ij}$$
 (3)

위의 변형된 출력 확률식을 이용해서 Viterbi알고리즘에서 시점 t에서의 전향 확률값에 대한 계산식은 아래와 같다.

$$\tilde{\alpha}_t(i) = \max_{1 \leq j \leq N} [\tilde{\alpha}_{t-1}(j) \cdot a_{ij}] \omega_t(i), 2 \leq t \leq T, 1 \leq j \leq N$$
 (4)

여기서, $\tilde{\alpha}_1(i) = \pi_i \omega_1(i)$, $\forall i \in N$ 이다. 최종적인 출력결과는 $\max_{1 \leq i \leq N} \tilde{\alpha}_T(i)$ 이 된다.

2.3 FVQ-DHMM의 문제점

앞서 언급한 바와 같이, FVQ-DHMM의 성능은 입력패턴에 대한 코드워드 벡터와의 거리에 기반한 가중치와 상태별 출력확률이라는 두 개의 요소에 의존하게 된다. 이러한 요소들이 FVQ-DHMM의 성능에 미치는 영향과 그 해결방법을 고려해 보기로 한다.

먼저 FVQ-DHMM에서 출력확률의 계산과정을 살펴보기로 하자. FVQ-DHMM에서 한 상태에서의 최종 출력값은 입력 패턴에 대한 K개의 인접 코드워드들에 대한 출력확률과 그들의 기여도의 곱에 대한 합의 형태로 표

원된다. 이러한 경우, 상태별 출력확률은 이미 모델의 학습과정에 의해서 고정되는 요소인 상태별 코드워드의 출력확률에 영향을 받기보다는 출력확률을 반영할 기여도를 나타내는 요소에 보다 의존적이 된다. 따라서, 궁극적으로 FVQ-DHMM의 성능향상을 위해서 학습과정에서 보다 강인한 출력확률의 추정과 최종 출력확률의 추정이 사용될 각 코드워드의 반영 기여도를 보다 정확하게 계산하는 것이 필요하게 됨을 알 수 있다. 이를 위해서 본 논문에서는 코드워드의 정확한 반영 기여도의 추정을 위해서 코드워드 의존 거리 정규화(CDDN:codeword-dependent distance normalization)와 상태별로 강인한 출력확률을 얻기 위한 방법으로 퍼지함수에 기반한 평활화 방법을 제안하고자 한다.

FVQ-DHMM에서 K개의 인접한 코드워드들의 출력확률을 반영을 위해서 사용되는 기여도는 거리(distance)에 기반한다. 입력패턴과 코드워드 벡터간의 거리를 위한 척도로는 유클리디언(Euclidean)거리를 사용한다. 이러한 거리척도는 상대적인 거리를 나타낼 뿐, 계산된 상대적 거리가 해당 코드워드의 거리분포를 고려할 때 어느 정도가 되는지를 고려하지 못하고 있다. 즉, 코드워드별 거리에 대한 기준으로 볼 때, 상대거리가 그 거리값이 달라질 수 있으며 보다 정확한 기여도 계산을 위해서는 코드워드별 특성을 반영하도록 계산된 상대거리값을 변환하는 것이 필요하다고 할 수 있다. 이를 위해서, 본 연구에서는 코드북의 생성 시, 각 코드워드에 대한 모든 최적 입력 패턴들에 대한 평균과 분산값을 갖는 정규화 분포(normal distribution)로 코드워드별 거리 분포를 모델링하고 입력 패턴과 해당 코드워드간의 상대적인 거리를 코드워드별 분포특성을 반영한 거리로 변환하는 CDDN을 제안한다.

앞서 설명하였듯이, FVQ-DHMM 성능에 영향을 미치는 요소로는 입력패턴에 대한 코드워드의 기여도 이외에 각 코드워드의 상태 내에서의 출력확률이 있다. 상태별 출력확률은 학습 데이터의 양에 따라서 성능차이가 나타난다. 특히, 이산형 HMM의 경우, 충분한 학습자료로 학습되지 못할 경우 시험 시 unseen 데이터에 대해서 zero 확률을 나타내므로 성능저하의 원인이 되기도 한다. 따라서, 이러한 문제를 해결하기 위해서 상태별 출력확률을 평활화하는 다양한 방법들이 제안되어 사용되고 있으며, 좋은 성능을 보이고 있다. 본 논문에서는 그러한 평활화 방법들 가운데에서 우수한 성능을 나타내고 있는 퍼지 기반 평활화(fuzzy based smoothing)방법을 일반화하고, 그 중에서 instar 형태에 기반한 퍼지 평활화 방법을 제안하고자 한다.

퍼지 기반 평활화는 parallel 방법의 일종으로, 하나의 입력 패턴에 대한 모든 코드워드의 기여도를 가지고 평활화 행렬을 작성하고, 작성된 평활화 행렬과 원래의 상태별 출력분포와의 곱의 형태로 평활화된 출력분포를 유도하게 된다. 이때, 평활화를 위해서 사용될 행렬을 유도하는 방식에는 instar 형태와 outstar 형태의 두 가지가 가능

하다. 본 연구에서는 보다 향상된 결과를 보이는 instar 방식에 의한 평활화 방법을 제안하고자 한다. 보다 자세한 내용은 3.2절에서 설명하고자 한다.

III. 코드워드 의존 거리 정규화와 퍼지기여도에 기반한 출력확률 평활화

이 절에서는 2.3절에서 설명한 FVQ-DHMM과 관련된 문제들을 해결하기 위한 방법들을 기술하고자 한다. 3.1 절에서는 코드워드별 거리분포 특성을 반영한 코드워드 종속 거리 정규화(CDDN)방법에 대해서 기술하며, 다음으로 3.2절에서는 강인한 출력확률을 추정하기 위한 방법으로 클래스기반 퍼지 평활화(CFS:class-based fuzzy smoothing)방법에 대해서 설명하고자 한다.

3.1 코드워드 의존 거리 정규화(CDDN)

FVQ-DHMM에 내제된 문제점들 가운데, 첫 번째로 각 코드워드에 대한 거리 분포 특성을 FVQ-DHMM의 출력확률 계산을 위한 각 코드워드의 가중치 계산에 반영하기 위한 방법으로 본 논문에서 제안된 CDDN(codeword-dependent distance normalization)에 대해서 설명하고자 한다. 기존의 FVQ-DHMM에서, 입력패턴과 코드워드 벡터간의 거리가 유클리디언 거리 척도에 의해서 계산되고, 계산된 상대거리가 입력패턴의 해당 코드워드에 대한 가중치 계산에 사용된다. 코드워드별 거리 분포특성을 고려하지 않고, 단순히 상대거리에 기반한 값을 사용하는 경우에, 상대거리의 의미가 부적절하게 가중치 값으로 사용될 수 있다. 즉, 넓은 거리 분포특성을 갖는 코드워드에 비해 좁은 거리 분포특성을 갖는 코드워드분포에서 동일한 상대거리가 보다 높은 가중치 값을 갖는 것이 바람직하다. CDDN의 유도과정은 크게 3가지 과정을 거쳐서 이루어진다.

[단계 1] 코드북 생성 시, 각 코드워드에 대한 학습 자료의 평균 거리 분포를 위한 파라미터를 계산한다. 각 코드워드에 대한 거리분포를 평균(μ)과 분산(δ)을 갖는 정규분포(normal distribution)라고 가정한다. 각 코드워드의 거리 분포에 대한 평균과 분산에 대한 계산식은 아래와 같다.

$$\mu_i = \frac{1}{P_i} \sum_p^P d(\vec{c}(i), \vec{x}_p(i)), \quad 1 \leq i \leq M \quad (5)$$

$$\sigma_i^2 = \frac{1}{(P_i - 1)} \sum_p^P (d(\vec{c}(i), \vec{x}_p(i)) - \mu_i)^2 \quad 1 \leq i \leq M \quad (6)$$

[단계 2] 하나의 입력패턴에 대한 코드워드와의 유클리디언 거리를 계산한다. 계산된 거리를 단계 1에서 계산된 각 코드워드별 거리에 대한 평균값과 분산값을 이용해서 코드워드 종속 정규화된 확률값으로 변환한다.

$$P(\bar{x}_i, \bar{c}(i)) = \exp \left[-\frac{(d(\bar{x}_i, \bar{c}(i)) - \mu_i)^2}{2\sigma_i^2} \right] \quad (7)$$

여기서, i 는 i 번째 코드워드를 나타내는 색인이며, \bar{x}_i 는 시점 t 에서의 입력패턴이나.

[단계 3] 단계 2에서 계산된 확률값을 음수값을 갖는 로그 함수를 적용해서 거리값으로 변환한다. 이때 계산된 거리값이 코드워드 종속 정규화된 거리값을 나타낸다. 확률에 대한 거리값으로 음수 로그 함수를 사용한 이유는, 단계 1에서 코드워드에 대한 거리분포의 모델과의 호환성을 유지하기 위해서 사용하였다.

$$D_i(\bar{x}_i, \bar{c}(i)) = -\log_c P(\bar{x}_i, \bar{c}(i)) \quad (8)$$

[단계 4] 단계 3에서 계산된 거리값을 FVQ-DHMM을 위한 입력패턴에 대한 코드워드의 가중치를 계산하는 위의 식(1)에 대입된다. 이후의 처리과정은 FVQ-DHMM에서의 출력확률 추정 과정과 동일하다.

CDDN에 사용될 파라미터인 각 코드워드의 평균과 분산의 값은 코드북 생성과정에서 생성되므로, 인식시에 별도의 계산이 불필요하다. 각 코드워드의 거리에 대한 확률모델은 정규분포로 가정하였으며, 이러한 가정에 기반하여 확률과 거리의 변환을 위해서 그러한 분포특성을 반영하는 음수 로그함수로의 변환식을 사용하였다. 즉, 확률과 거리의 변환식도 Gaussian함수의 특성을 갖는 모델링 방법으로 변환됨을 의미한다. 제안된 CDDN의 전체적인 처리과정은 아래의 그림 3과 같다.

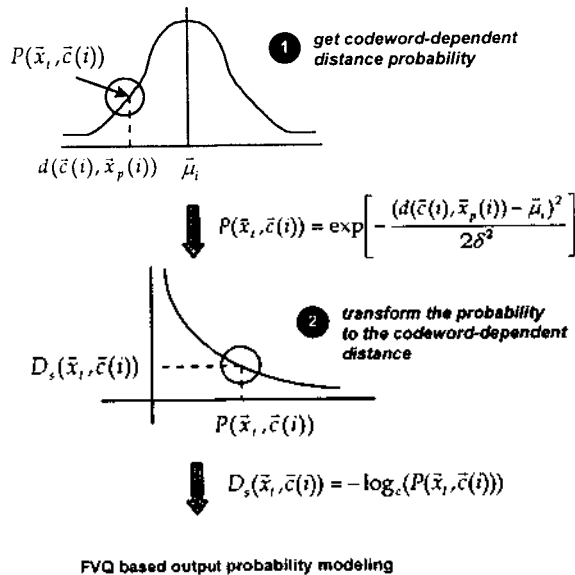


그림 3. 코드워드 종속 거리 정규화(CDDN)의 수행과정

3.2 클래스기반 퍼지 평활화(CFS)

앞서 설명하였듯이, 이산형 DHMM의 경우 부족한 학습 데이터로 인해서 zero 확률을 갖는 출력확률이 존재함으로써 전체적인 인식율을 저하시키는 문제가 있다. 이러한 문제를 해결하기 위해서 다양한 평활화 방법들 -flooring, co-occurrence, fuzzy smoothing-이 제안되어 사용되고 있으며, 이 가운데에서 퍼지에 기반한 평활화가 다른 방법들에 비해서 향상된 인식율을 나타내는 것으로 알려져 있다. 본 절에서는 퍼지를 이용한 평활화 방법을 일반화하고, 이러한 방법들 가운데서 instar에 기반한 방법을 중심으로 기술하고자 한다.

3.2.1 클래스 기반 퍼지평활화의 정의

퍼지에 기반한 평활화 방법으로는 Koo에 의해서 제안된 퍼지 평활화 방법[5]이 있으며, 이러한 방법은 학습자료를 대상으로 최적 코드워드별로 자료들을 분류하고, 각 코드워드들에 대한 입력패턴들의 평균 거리값을 계산하고, 이 값을 퍼지 행렬을 유도하기 위해서 사용한다. 이 방법은 여러 실험에서 다른 평활화 방법에 비해서 좋은 성능을 나타내는 것으로 알려져 있다[9][10]. 본 연구에서는 이러한 퍼지 평활화 방법을 보다 일반화(generalization)시키기 위해서, 클래스 기반 퍼지 평활화(CFS: class-based fuzzy smoothing)라고 명명하고, CFS를 정의하고자 한다.

여기서, 클래스(class)란 “하나의 코드워드에 최적으로 사상되는 입력패턴들의 집합”을 의미하며, 클래스 거리(class distance)란 “클래스 내에 있는 모든 입력패턴과 코드워드간의 거리에 대한 평균값”을 의미한다. CFS는 그 형태에 따라서 instar와 outstar 형태로 나뉘어지며, 그 형태는 아래의 그림 4와 같다.

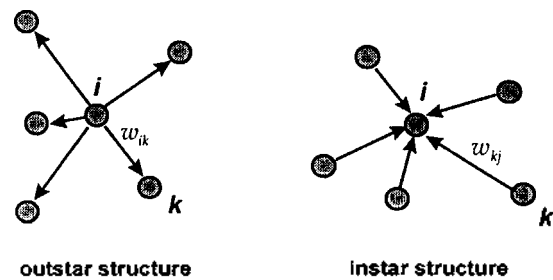


그림 4. CFS의 유형(instar와 outstar-form)

위의 그림 4에서 filled-circle은 클래스를 나타내며, 화살표는 클래스간의 근접관계를 나타낸다. 화살표의 소스(source)부분이 최적 클래스를 나타내고, target 부분이 최적클래스가 주어졌을 때의 다른 클래스의 기여도를 나타내게 된다. Outstar 형태의 경우, 최적 클래스 i 가 주어졌을 때, 클래스 i 에 대한 다른 클래스들의 퍼지 기여도를 나타낸다. 반면, Instar의 경우는 다른 클래스들이 최적일 때, i 번째 클래스의 기여도를 나타낸다. 두 방법간의 차이

는 출력확률의 평활화 시, 최적의 클래스가 주어졌을 때 다른 클래스들의 기여도의 관점에서 출력확률에 대한 평활화를 할 것인가, 아니면 다른 클래스들의 관점에서 한 클래스에 대한 기여도를 중심으로 평활화를 할 것인가에 대한 것이다. 앞서 언급한 Koo[5]가 제안한 퍼지 평활화 방법은 Oustar 형태이고, 본 연구에서는 Instar 방법을 제안하고자 한다.

3.2.2 CFS의 유형: instar-form과 outstar-form

먼저, CFS방법 가운데에서 oustar 형태에 해당되는 Koo에 의해서 제안된 평활화 방법을 설명하기로 한다. 이 방법은 FVQ(fuzzy vector quantization)에서 사용되는 퍼지 목적함수를 사용하여 평활화 행렬을 계산하고, 평활화 행렬과 상태의 출력빈도를 곱해서 최종적인 평활화 출력확률을 계산하는 방법이다. FVQ에 기반한 상태 모델링의 경우, 상태에 대응되는 특징 파라미터에 대해서 모든 코드워드와의 거리를 계산한 후, 각 코드워드와의 거리를 퍼지목적 함수의 입력으로 사용하여 입력패턴에 대한 코드워드들의 퍼지 기여도를 계산하게 된다. 퍼지 평활화는 이러한 방법을 응용한 것으로, 입력패턴에 대한 최적 코드워드별로 입력패턴을 묶고, 이러한 패턴들의 집합과 코드북의 모든 코드워드간의 평균거리를 구한 후, 이러한 패턴집합과 코드워드간의 평균 거리값을 퍼지목적함수의 입력으로 사용한다. 평활화 행렬 T는 다음의 수식을 이용해서 추정된다.

$$t_{ij} = \left[\sum_{k=1}^M [d(C_i, v_j)/d(C_i, v_k)]^{1/(F-1)} \right]^{-1} \quad (9)$$

where, $i, j = 1, 2, \dots, M$

이렇게 유도된 퍼지행렬에 원래의 상태별 출력확률을 곱해서 최종적인 평활화된 코드워드의 출력확률을 구하게 된다.

$$\hat{b}_i^j = \sum_{k=1}^M (b_k^i \cdot t_{kj}) \quad (10)$$

이 방법은 하나의 코드워드를 중심으로 수집된 입력자료들로 구성된 하나의 클래스가 주어졌을 때, 다른 코드워드들과 이 클래스간의 거리값을 구하고 이 값에 퍼지목적 함수를 적용해서 주어진 클래스에 대한 코드워드들의 평활화값을 유도하는 방식으로, 하나의 클래스에 대한 코드워드들의 근접도가 출력확률의 평활화를 위한 요건으로 이용된 경우이다.

이러한 경우와는 달리, instar형태의 CFS에서는 하나의 코드워드에 대한 평활화 값을 다음과 같이 정의한다. 한 코드워드 i 의 출력확률은 식 (11)과 같이 조건부 확률의 합으로 기술할 수 있다.

$$\hat{b}_i = \sum_{k=1}^M P(i|k) \cdot b_k \quad (11)$$

$P(i|k)$ 는 k 번째 코드워드가 존재할 때 코드워드 i 의 출력확률로 k 번째 코드워드와 가장 가까운 입력패턴의 빈도비율로써 평활화를 위해서 사용된다. 여기서, $P(i|k)$ 가 instar형태를 갖게 되며, 이 값은 k 번째 코드워드에 해당되는 클래스에 대한 i 번째 코드워드의 기여도라는 의미로 사용된다. 앞에서 설명한 outstar방법은 하나의 코드워드에 대응되는 클래스에 대해 다른 코드워드와 기여도와 상태에서의 코드워드의 빈도의 곱으로 한 코드워드에 대한 출력확률을 평활화 하는 반면, instar형태에서는 i 번째 코드워드의 출력확률의 평활화 과정은 다른 클래스들에 대한 i 번째 코드워드들의 기여도와 해당 클래스의 코드워드의 출력확률의 곱에 대한 합의 형태로 유도된다. 이 방법은 코드워드의 출력확률 평활화에 영향을 주는 것으로, 다른 코드워드와 관점에서 해당 코드워드와의 관계를 고려하는 것이 해당 코드워드의 관점에서 다른 코드워드와의 관계를 고려하는 것보다 평활화에 중요하다는 것을 나타낸다. Outstar형태와 instar형태에 의한 평활화를 통한 인식실험은 4.2절에서 기술하기로 한다.

IV. 실험 및 결과

4.1 실험환경

본 연구에서 제안한 방법의 성능 평가를 위해서 다른 남자 17명의 화자가 2회씩 발성한 115개의 단어¹⁾(총 3910개)를 대상으로 실험하였다. 학습을 위해서 12명이 2회씩 발성한 자료를 사용하였으며, 실험을 위해서 나머지 5명이 2회씩 발성한 자료를 사용하였다.

음성인식을 위해서 사용된 단위는 문맥독립음소단위(context-independent phoneme like unit)이다. 115단어로부터 얻어진 49개의 음소모델을 사용하였다. 각각의 음소모델은 3개의 상태를 갖는 상태간 점프를 허용되지 않은 left-to-right모델을 사용하였다. 실험에 사용한 특징 파라미터는 14차 멜켑스트럼 계수, 14차 델타 멜켑스트럼 계수와 (energy + delta-energy)를 함께 사용하였다. 멜켑스트럼 계수는 선형 예측 계수를 구한 뒤 Oppenheim이 제안한 양선형 변환(bilinear transformation)[11]을 통해 구하였으며, 주파수의 warping을 위한 변수값은 0.64를 사용하였다. 델타 파라미터는 차분 멜켑스트럼(differenced mel-cepstrum)과 차분 에너지(differenced energy)[12]를 사용하였으며, 차분창(difference window)의 크기는 5로 정하였다.

입력 파라미터의 벡터 양자화를 위해서 일반적으로 널리 사용하는 LBG[13]방법을 사용하였다. 멜켑스트럼과 델타 멜켑스트럼의 경우 각각 256의 코드워드를 사용하였으며, 에너지와 델타 에너지의 경우는 64의 코드워드

1) 이 데이터는 ETRI에서 제공된 것입니다.

를 갖는 단일 코드북을 사용하였다. 코드북의 종료 조건은 $1.0e-6$ 으로 정하였으며, HMM의 각 상태에서 관측되지 못한 심벌의 출력 확률값으로 $1.0e-5$ 를 사용하였다. 특성 파라미터 추출을 위한 분석 조건은 표 1과 같다.

표 1. 실험 환경

sampling rate	16kHz	preemphasis	$1.0-0.972^{-1}$
analysis	20 msec (320 points)	shift interval	10 msec (160 points)
feature parameters	mel-cepstrum(14) delta-mel-cepstrum(14) energy + delta-energy(2)	codebook size	mel-cepstrum(256) delta-mel-cepstrum(256) energy + delta-energy(6)

인식 시스템은 크게 학습단계와 실험단계로 나뉘어지며, 학습단계는 다시 코드워드 작성부와 모델 학습부의 두 부분으로 구성되어 있다. 코드워드 작성부의 경우 학습 데이터를 대상으로 특징 파라미터를 추출하고 LBG 방법에 따라 코드북을 작성하는 부분이다. 모델 학습부에서는 분맥독립 유사 음소 모델을 접속하여 단어모델을 구성하고 단어별 학습자료를 이용해서 유사음소모델을 학습하게 된다. 유사음소모델의 학습을 위해서 segmental K-means방법[14]을 사용하였다. 학습과정을 살펴보면, 단어별 음성 자료로부터 특징을 추출한 뒤, 이를 양자화하여 코드워드의 열로 변환한다. 변환된 파라미터별 코드워드 열을 사용하여 유사음소모델 파라미터의 학습이 이루어지며, 종료조건이 만족될 때까지 학습과정을 반복하게 된다. 일단 모델의 학습이 이루어지면, 실험 데이터에 대한 성능 평가를 위해서 학습과 동일한 과정으로 입력 패턴을 코드 열로 변환하고 각 단어모델별 입력 패턴의 출력 확률을 계산 후, 이들 중 가장 큰 것을 해당 단어로 선정하여 출력한다. 인식 시스템의 전체적인 구성도는 아래의 그림 5와 같다.

4.2 실험결과

4.2.1 출력확률의 클래스 기반 퍼지 평활화에 따른 인식 제안된 방법의 성능평가를 위해서 기존의 평활화 방법들과 비교평가를 수행하였다. 기본적으로 널리 사용되는 flooring방법을 기본 시스템(base system)으로 하여 다른 방법들과의 인식을 측면에서 성능비교를 하였다. 제안된 instar 형태의 평활화 방법의 특성을 보이기 위해서, 기존의 FVQ-DHMM을 대상으로 실험하였다.

선형적 동시출현 평활화의 경우, 평활화행렬을 계산하기 위해서 각 상태에서 얻어진 출력확률을 모두 고려하고 있으나, 제안된 방법에서는 상태 의존적인 코드워드간의 동시 존재 가능성을 고려하고 있다. 퍼지 평활화와 제안된 방법의 경우 퍼지인수(fuzziness factor)의 값은 예비실험을 통해서 좋은 성능을 보인 1.5로 사용하였다. 실험 결과는 표 2와 같다.

표 2. 제안된 방법과 기존 방법들간의 성능비교 (%)

	Flooring	co-occurrence	oustar-CFS	instar-CFS
인식율	94.3%	95.8%	97.1%	98.6%

실험결과, 기준이 되는 flooring방법($1.0-e6$)값을 사용한 방법에 비해서 제안된 방법이 75.4% 우수하였으며, 퍼지 방법의 경우 제안된 instar-형태의 CFS가 기존의 oustar-형태의 CFS에 비해서 55.1%의 성능향상을 보였다. 동시출현방법에 비해서 제안된 방법은 66.7%의 인식율 향상을 보임으로써 상태 의존적인 평활화가 동시출현방법과 같은 전체 음소에 대한 평활화 방법보다 유효함을 알 수 있었으며, CFS방법에서는 제안된 instar형태의 CFS방법이 기존의 oustar형태의 CFS에 비해서 평활화하는 방법으로 효과적임을 알 수 있었다.

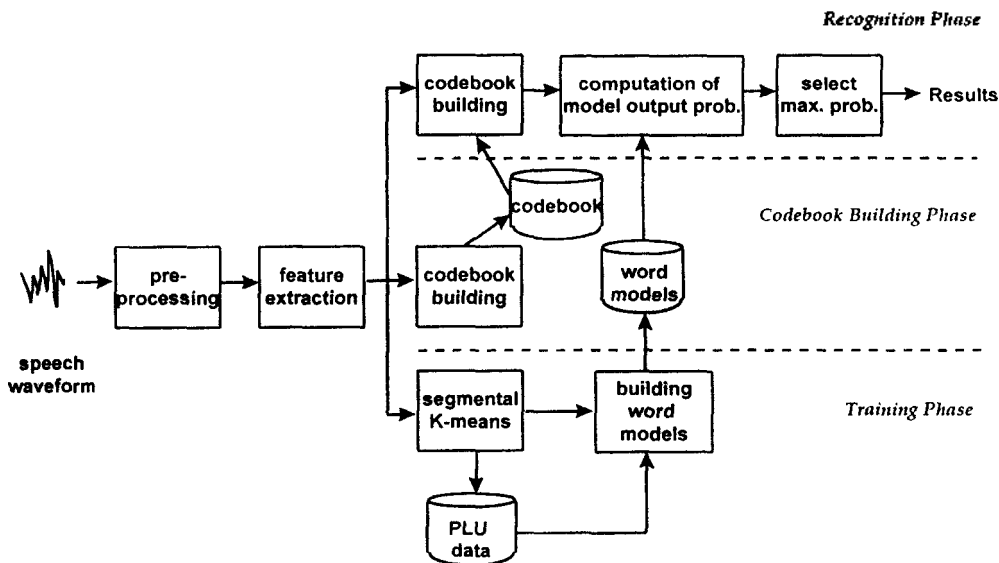


그림 5. 인식시스템의 개요도

4.2.2 FVQ-DHMM, FVQ-CDDN, 그리고 FVQ-CDDN-CFS의 성능비교

제안된 CDDN의 성능을 측정하기 위해서, 기존의 FVQ-DHMM과의 성능비교 실험을 수행하였다. 실험결과와 아래의 그림 6과 같다.

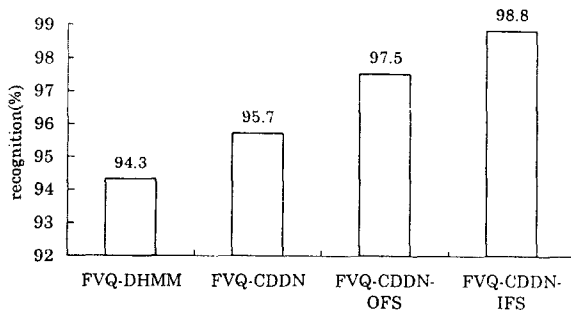


그림 6. 제안된 방법의 비교 실험 결과

기존의 FVQ방법에 CDDN을 적용한 경우(FVQ-CDDN)에 95.7%의 인식율을 얻었으며, FVQ에 oustar형태의 퍼지평활화를 적용한 FVQ-CDDN-OFS의 경우 97.5%의 인식율을 나타내었다. 제안된 instar형태의 퍼지 평활화(FVQ-CDDN-IFS)를 적용한 경우 98.8%의 인식율을 보이므로써, FVQ-DHMM에 비해서 4.5%의 인식율 향상을 보였으며, 클래스 기반 퍼지 평활화 방법에 있어서는 instar 방법이 oustar 방법에 비해서 1.3%의 인식율 향상을 나타내었다. 이러한 실험을 통해서, FVQ-DHMM의 성능향상을 위해서 거리척도의 경우 코드워드의 거리 분포 특성을 고려한 CDDN이 유효함을 알 수 있었으며, 아울러, 출력확률의 평활화를 위한 방법으로 제안된 instar 형태의 퍼지 평활화 방법이 FVQ-DHMM의 성능향상에 기여함을 확인할 수 있었다.

V. 결 론

본 논문에서는 FVQ-DHMM에서 강인한 출력확률의 추정을 위해서 코드워드 종속 거리 정규화와 출력확률에 대한 instar 형태의 퍼지 평활화 방법을 제안하였다. FVQ-DHMM에서 한 상태에서의 최종 출력값은 입력패턴에 대한 K개의 인접 코드워드들에 대한 출력확률과 그들의 기여도의 곱에 대한 합의 형태로 표현된다. 이러한 경우, 상태별 출력확률은 이미 모델의 학습과정에 의해서 고정되는 요소인 상태별 코드워드의 출력확률과 출력확률을 반영할 기여도에 의존하게 된다. 따라서, FVQ-DHMM의 성능향상을 위해서는 학습과정에서 보다 강인한 출력확률을 추정하는 것과 최종 출력확률의 추정에서 사용될 각 코드워드의 반영 기여도를 보다 정확하게 계산하는 과정이 필요하게 된다. 이를 위해서 본 논문에서는 코드워드의 반영 기여도의 정확한 추정을 위해서 코드워드

의존 거리 정규화(CDDN)방법과 상태별로 강인한 출력확률을 얻기 위한 방법으로 instar 형태의 클래스 기반 퍼지 평활화(CFS) 방법을 제안하였다.

실험결과, 제안된 CDDN을 적용한 방법이 기존의 FVQ-DHMM에 비해 24%의 오인식률 감소가 있었으며, 상태별 출력분포에 대해서 평활화를 적용한 경우 79%의 오인식율을 감소 시킴을 알 수 있었다. 이러한 실험들을 통해서, 코드워드 의존 거리 정규화와 instar형태의 출력확률 퍼지 평활화방법이 인식율 향상에 기여함을 알 수 있었으며, 결과적으로 제안된 방법들이 FVQ-DHMM을 위한 강인한 출력확률의 추정을 위한 대안으로 사용 가능함을 알 수 있었다. 향후 연구로는 대어휘와 연속음성을 대상으로 제안된 방법의 유효성을 평가해 보고자 한다.

참 고 문 헌

1. L. R. Rabiner, B. H. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall International Editions, 1993.
2. K. F. Lee, *Automatic Speech Recognition: The Development of the SPHINX System*, Kluwer Publishers, 1989.
3. D. Huang, M. A. Jack, "Semi-Continuous Hidden Markov Models for Speech Signals," *Computer Speech and Language*, Vol. 3, pp. 239-251, 1989.
4. Schwartz, et al., "Robust smoothing methods for discrete hidden Markov models," *Proc. of ICASSP*, pp. 548-551, 1989.
5. J. M. Koo and C. K. Un, "Fuzzy smoothing of HMM parameters in speech recognition," *Electronic Letters*, Vol. 26, pp. 743-744, 1990.
6. J. DAI, "Robust Estimation of HMM Parameters Using Fuzzy Vector Quantization and Parzen's Window," *Pattern Recognition*, Vol. 28, No. 1, pp. 53-57, 1995.
7. H. Applebaum, B. A. Hanson, "Enhancing the discrimination of speaker independent HMMs with corrective training," *Proc. of ICASSP*, pp. S6. 13-16, 1989.
8. R. Bahl, et al., "A new algorithm for the estimation of hidden Markov model parameters," *Proc. of ICASSP*, pp. S11. 2-4, 1988.
9. H. P. Tseng, M. J. Sabin, et al, "Fuzzy Vector Quantization applied to hidden Markov modeling," *Proc. IEEE ICASSP 87*, Dallas, Texas, pp. 15.5.1-15.5.4, 1987.
10. H. J. Choi, Y. H. Oh, "Speech Recognition Using an Enhanced FVQ Based on a Codeword Dependent Distribution Normalization and Codeword Weighting by Fuzzy Objective Function," *Proc. of International Conference on Spoken Language Processing*, Vol. 1, ThPIPI.17, 1996.
11. A. V. Oppenheim, D. H. Johnson, "Discrete Representation of Signals," *The Proceedings of the IEEE*, Vol. 60, No. 6, pp. 681-691, June 1972.
12. K. Shikano, "Evaluation of LPC Spectral Matching Measures for Phonetic Unit Recognition," *Technical Reports*,

Computer Science Department, Carnegie Mellon University, 1986.

- 13. Y. Linde, A. Buzo, R. M. Gray, "An Algorithm for Vector Quantization Design," IEEE Trans. on Communication, Vol. 28, No. 1, pp. 84-95, Jan. 1980.
- 14. L. R. Rabiner, J. G. Wilpon, et al., "A Segmental K-means Training Procedure for Connected Word Recognition Based on Whole Word Reference Patterns," AT&T Technical Journal, Vol. 65, No. 3, pp. 21-31, 1986.

▲최 환 진(Hwan Jin Choi)



1990년 2월: 고려대학교 전산학과(학사)

1992년 2월: 한국과학기술원 전산학과(석사)

1992년 3월~현재: 한국과학기술원 전산학과 박사과정 재학중

※관심분야: 음성인식, 대화관리, 은닉 마르코프 모델, 퍼지이론, 신경 회로망

▲김 연 준(Yeon Jun Kim)



1991년 2월: 한국과학기술원 과학기술대학 전산학과(학사)

1993년 8월: 한국과학기술원 전산학과(석사)

1994년 3월~현재: 한국과학기술원 전산학과 박사과정 재학중

※관심분야: 음성합성, 운율 모델링, 은닉 마르코프 모델, 신경 회로망

▲오 영 환(Yung Hwan Oh)



1972년 2월: 서울대학교 공과대학 전자공학과

1974년 2월: 서울대학교 교육대학원 공업교육학과(석사)

1980년 3월: Tokyo Institute of Technology 정보공학전공(박사)

1981년 4월~1985년 6월: 충북대학교 공과대학 전산학과 조교수

1983년 12월~1984년 11월: University of California(Davis) 연구교수

1985년 7월~현재: 한국과학기술원 전산학과 교수로 재직중

※관심분야: 음성인식, 음성합성, 음성코딩, 화자인식, 대화관리, 신경 회로망, 전문가 시스템