선별적인 관측열 추출을 통한 DHMM 음성인식의 성능 개선

김우창[U] 조선호 고수정 이정현
인하대학교 전자계산공학과
{ kimwc, gangta, sujung }@nlsun.inha.ac.kr, jhlee@inha.ac.kr

Speech Recognition Improvement Using Extraction Selective Observation in DHMM

Woo-Chang Kim[U] Sun-Ho Cho Su-Jeong Ko Jung-Hyun Lee
Dept. of Computer Science and Engineering, Inha University

요약

음성인식 시스템에 사용하는 알고리즘 중에 하나인 DHMM은 코드북을 이용하여 음성의 프레임들에 대한 특징을 관측열로 추출하여 음성의 패턴에 대한 훈련과 인식을 수행하게 된다. 그러나 음성은 유성음과 무성음의 특징 차이가 많이 나게 되므로 하나의 코드북을 이용하게 되면 코드북 오차에 의하여 성질이 전혀 다른 코드북 인덱스를 DHMM의 관측열로 사용하게 된다. 본 논문에서는 음성의 유성음과 무성음에 대한 선별적인 작업을 통해 서로 다른 코드북을 만들어 관측열을 추출하고 선행 관측과 현 관측과의 거리 비교 연산을 통하여 관측의 시간축을 정규화한 관측열을 음성인식에 사용하였다. 본 논문에서 제시하는 인식 방법을 사용하여 실험한 결과, 기존의 인식 방법보다 5.33% 향상된 결과를 얻었다.

1. 서론

음성 정보는 유성음과 무성음의 패턴들의 조합으로 나타낼 수 있다. 이러한 음성 정보를 인식하는 방법은 음성 데이터를 일정한 프레임으로 나누고 각 프레임에 대한 분석을 통하여 특징을 추출한 후 그 특징들의 패턴들을 다양한 방법을 통하여 인식을 할 수 있다. 그 중 DHMM 인식 알고리즘은 이산적인 관측 심볼들의 패턴을 확률적으로 계산한 모델들과 인식패턴과의 비교 확률값을 계산한 결과 중 가장 좋은 값을 선택하는 인식 방법이다.

코드북을 이용한 DHMM 인식 알고리즘은 입력된 음성의 특징들을 추출한 후 그것들을 이용해 코드북을 작성한다. 그리고 입력음성의 각 프레임에 대하여 가장 유사한 코드워드의 대표값을 인덱스로 하여 관측열을 만들어 HMM의 훈련과 인식에 사용한다. 하지만 코드북 오차에 의하여 관측열 내에서 유성음과 무성음의 패턴이 바뀌게 되어 인식에 오류가 생길 수 있다.

본 논문에서는 선형예측방법(linear predictive coding, 이하 LPC)를 이용하여 각 프레임에 대한 특징을 추출하고 유성음과 무성음을 분리하는 과정을 거쳐서 서로 다른 코드북을 구성하여 위와 같은 오차를 줄이며 선행 관측과 현 관측과의 거리 비교 연산을 통하여 선별적인 관측열을 추출하는 방법을 제안하고 이를 DHMM 인식기에 적용함으로써 인식률을 향상시킨다.

2. 기존 인식 시스템 고찰

2.1 LPC를 이용한 음성 특징 추출

LPC는 음성 파형의 파라미터를 상당히 정확하게 추정할 수 있으며 계산 속도도 다른 파라미터들에 비해 상대적으로 빠르므로 음성 분석 방법으로 많이 이용되고 있다[1][2].

LPC는 신호의 현재값을 그 신호의 과거값들의 선형 결합(Linear combination)으로 추정하는 방법이다. 여기서 예측값들과 실제값들 사이의 오차를 최소로 하는 계수들을 LPC계수라 하며 이들을 음성의 특징 벡터로 이용한다[5].

LPC계수를 구하는 방법은 다음과 같다. 입력 신호 s(n)은 p개의 과거값들의 결합으로 구성되며 식(1)과 같다.

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + Gu(n) \qquad (1)$$

여기서 G는 여기 신호의 이득을 나타내며 a_k는 필터계수를 $u(n)$은 정규화된 여기 신호를 나타낸다. 예측값 $\hat{s}(n)$은 식(2)이다.

$$\hat{s}(n) = \sum_{k=1}^{p} a_k s(n-k) \qquad (2)$$

식 (1)과 식(2)를 이용하여 식(3)의 예측 오차를 구할 수 있고 이를 최소로 하는 값에서 LPC계수를 얻을 수 있다.

$$e(n) = s(n) - \hat{s}(n)$$
$$= s(n) - \sum_{k=1}^{p} = Gu(n) \qquad (3)$$

LPC계수는 자기 상관 방법, 공분산법, 격자 필터등의 방법을 이용하여 구한다[1][5].

2.3 벡터 양자화(Vector Quantization)

벡터 양자화는 입력값의 차원이 너무 크거나 그 값의 범위가 매우 큰 경우, 대표 패턴이 저장된 코드북으로부터 이에 대응되는 양자화값으로 차원 수를 줄이거나 범위를 줄이는 방법이다[4][6].

벡터 $x=(x_1, x_2, \ldots, x_d)^T \in R^d$를 d차원의 벡터라고 하면, 벡터 양자화란 벡터 x를 유한개의 원소를 가진 집합 $Z=\{z_i, 1 \le i \le L\}$에서의 하나의 원소로 대치시키는 것이다. 이를 수식으로 나타내면 식(4)와 같다.

$$z = q(x) \qquad (4)$$

식(4)에서 q를 양자화기, Z를 코드북, L을 코드북의 크기, $\{z_i\}$를 코드워드라고 한다. 코드북을 설계하기 위해서는 원래의 벡터 x가 나타나는 공간을 분할하여 L개의 부공간 $\{C_i, 1 \le i \le N\}$으로 나누고, 벡터가 부공간 C_i에 속할 때 코드워드 z_i로 양자화한다.

대표적인 군집화 방법으로는 K-means알고리즘과 LBG 알고리즘이 있다[7].

2.3 HMM(Hidden Markov Model) 인식

HMM의 기본개념은 음성을 마코프 모델의 확률적인 파라미터로 모델링하여 기준 마코프 모델을 만들고 인식과정에서는 입력음성과 가장 유사한 기준 마코프 모델을 추정해 냄으로써 음성을 인식하는 것이다. HMM은 음성의 스펙트럼 변화, 시간 신축 변화 등 실제 음성의 변화를 통계적으로 모델화할 수 있다.

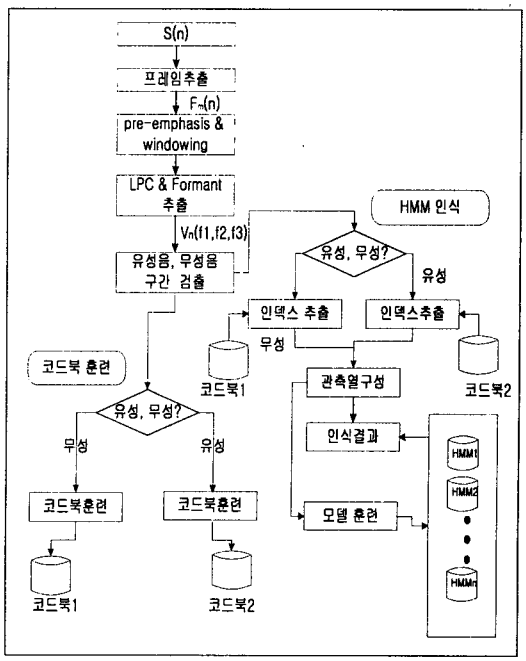
음성 인식의 쓰인 기존의 방식은 음성 데이터를 각 프레임으로 나누고 각 프레임마다 특징을 추출한 후 특징 벡터를 미리 구성된 코드북을 이용하여 관측열을 구성한다. 이를 이용하여 HMM 모델을 훈련하고 인식을 수행한다.

3. 선별적 관측열 추출 시스템

[그림 1]에서와 같이 본 논문에서 제시하는 시스템은 선별적인 관측열을 추출하기 위한 코드북 훈련시스템과 선별적인 관측열을 추출하는 인식시스템으로 나누어진다.

3.1 전처리 과정

음성 신호의 경우, 10ms에서 100ms 간격의 짧은 시간 동안 안정된 주기성을 가지며, 0.5초 이상의 간격에서는 주기성이 변하는 특성을 갖는다. 이러한 음성 신호의 특성을 quasi-stationary라 한다. 입력된 음성 신호는 안정된 구간에서 특징을 추출해야 하기 때문에 이를 20ms 간격으로 프레임화한다[7].



[그림 1] 전체 시스템

음성신호의 주파수 스펙트럼은 일정하지 않고, 주파수 값이 높을수록 그 성분이 작아져 주파수가 2배가 되면 약 6(db)의 기울기로 그 파워의 진폭 특성이 작아진다. 이는 자음 부분의 주파수 정보를 분석하기에 부적절하다. 그러므로 음성신호 분석전에 6(db)기울기를 갖는 고역강조 필터를 통과시켜 음성신호의 스펙트럼이 저역부터 고역까지 같은 S/N비를 갖게 하는 과정이 선강조이며, 전달함수는 다음과 같다.

$$H(Z) = 1 - aZ^{-1} \quad (a: 0.9 \ or \ 0.875) \qquad (5)$$

음성 분석시 각 프레임의 깁스 현상을 제거하기 위해 창함수를 사용하였으며 본 논문에서는 해밍창을 이용하였다.

$$W(i) = 0.54 - 0.46 * \cos(2\pi i * i/N) \qquad (6)$$
$$0 \le i \le N$$

앞장에서 설명한 방법으로 16차의 LPC를 구한 후 이를 이용하여 포만트 주파수를 구한다. 포만트 주파수를 구하는 방법은 역필터 $A(z)$의 근을 계산하는 root solving 방법을 사용하였으며 제1포만트, 제2포만트, 제3포만트 값을 벡터로 하여 각 프레임의 정보를 구성한다.

포만트 정보의 추적을 통하여 유성음과 무성음 구간을

검출하여 각 프레임 정보를 구성한다.

3.2 코드북 훈련 시스템

본 시스템에서 사용하는 코드북 훈련은 LBG알고리즘을 사용하였다.

DHMM을 구성하는 관측심볼의 범위, 즉, 코드북의 전체 인덱스 범위는 1 ~ 128이며, 무성음 구간의 인덱스 범위는 1 ~ 64, 유성음 구간의 인덱스 범위는 65 ~ 128로 구성된다.

훈련 데이터는 15명의 이름을 10명의 남성화자가 각각 10번씩 발성한 것을 사용하여 무성음 구간에 해당하는 코드북1과 유성음 구간에 해당하는 코드북2를 생성한다.

3.3 DHMM 인식 시스템

[알고리즘 1]은 HMM 훈련과 인식에 적용되는 관측열 추출 알고리즘이다. 전처리 과정에서 획득한 프레임 정보를 이용하여 프레임이 무성음 구간이면 코드북1에서 인덱스를 추출하고 유성음 구간이면 코드북2에서 인덱스를 추출한다. 추출한 인덱스의 코드워드와 이전 관측 인덱스가 동일한 구간이면 둘 사이의 거리를 구하여 임계값을 넘지 못하면 관측열에서 제거된다. 관측열의 중복을 제거시켜 음성의 시간 변화를 정규화시키는 기능을 하며 인식 계산의 수를 줄이는 기능을 한다.

[알고리즘 1] 선별적인 관측열 추출 알고리즘

```
i = 0; j = 0;
while(frame != END){
        if(Frame[i] == 자음){
                OBS_TEMP = Get_CodeBook(CODEBOOK1);
        }else if(Frame[i] == 모음){
                OBS_TEMP = Get_CodeBook(CODEBOOK2);
        }
        if(Frame[i] == 모음 && Frame[i-1] == 모음){
                distance = Cal_Dist(OBS[j-1], OBS_TEMP);
                if( distance > THRES_HOLD){
                        OBS[j++] = OBS_TEMP;
                }
        }else OBS[j++] = OBS_TEMP;
        i++;
}
```

코드북 훈련에 사용한 데이터를 가지고 [알고리즘 1]을 적용하여 각 데이터의 관측열을 추출한 후 HTK를 이용하여 DHMM 훈련을 하여 인식에 사용할 모델을 구성하고 인식 단어에 대하여 [알고리즘1]을 적용하여 관측열을 추출한 후 인식을 수행한다.

4. 실험결과

본 논문의 인식 실험에서는 앞장에서 준비한 코드북과 HMM 모델로 HTK를 사용하여 인식 실험을 했으며 인식 데이터는 10명의 화자가 5번씩 각각의 인식 이름을 발성한 후 각 이름에서 10개의 데이터를 선택한다.

각 프레임에서 관측열을 추출하는 기존의 인식 방법과 본 논문에서 제시하는 방법으로 인식을 수행하여 인식 단어의 평균 관측열수와 인식률을 [표 1]과 같이 얻었다.

[표 1]을 가지고 평균 관측열 수의 표준편차를 구하면 기존 시스템의 값은 11.346이고 본 논문에서 제시한 시스템의 값은 4.362로 인식 단어에 대한 관측열 수의 변화 폭이 줄어들었으며 인식률도 5.33% 향상되었다.

[표 1] 관측열의 평균수와 인식률비교

이름	기존방법의 평균관측열수	[알고리즘1] 적용 평균관측열수	기존방법의 인식률	[알고리즘1] 적용 인식률
고수정	108	82	80%	90%
김우창	125	85	90%	90%
김중철	113	80	90%	100%
김진수	108	83	80%	100%
김태욱	84	75	70%	70%
박영규	97	78	80%	80%
서영완	99	74	60%	60%
이정현	112	83	70%	80%
정경용	105	79	70%	70%
정미옥	85	72	60%	60%
조선호	106	77	80%	100%
조영택	89	72	70%	70%
최석용	97	81	80%	80%
한승진	112	83	90%	100%
허준회	108	84	90%	90%
전체	103.2	79.2	77.33% (116/150)	82.66% (124/150)

5. 결론

본 논문에서는 인식 단어의 관측열을 특징에 따라 분류된 코드북을 이용하여 추출하고 유성부의 관측열을 정규화함으로써 기존의 인식 방법보다 인식률을 5.33% 향상시켰다. 하지만 포만트를 이용한 유무성부 검출에서의 오차율을 감소시키기 위해 반자동으로 처리했다. 향후 포만트를 이용한 유무성음 검출을 자동으로 처리하여 인식시스템에 적용하는 것이 필요하다.

6.참고문헌

[1] J. Makhoul, "Liear Prediction : A tutorial Review," Proc. IEEE, Vol.63, pp. 561-580, 1975.

[2] L. R. Rabiner, "On the use of Autocorrelation Analisys for Pitch Detection," IEEE Trans., Vol. ASSP-25, No.1, February 1977.

[3] L. R. Rabiner, R. W. Schafer, *Digital Processing of Speech Signals*, pp. 447-455, Prentice Hall, 1978.

[4] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," IEEE Trans. Comm., Vol, COM-28, pp. 84-95, 1980.

[5] L. R. Rabiner, Biing-Hwang Juang, *Fundamentals of Speech Recognition*, Prentice Hall, pp. 97-140, 1993.

[6] R .M. Gray, "Vector quantization," IEEE ASSP Megazine, pp. 4-29, April, 1984.

[7] 오영환, *음성언어정보처리*, 홍릉과학 출판사, pp. 43-88, 1998.