초기화하지 않은 K-means iteration을 이용한 고립단어 인식

김 진영, 성 굉모
서울대학교 전자공학과

Isolated Words Recognition using K-means iteration without Initialization

Jin Young Kim, Keong Mo Sung
Dept. of Eletronics, Seoul National University

Abstract

K-means iteration method is generally used for creating the templates in speaker-independent isolated-word recognition system. In this paper the initialization method of initial centers is proposed. The concepts are sorting and trace segmentation. All the tokens are sorted and segmented by trace segmentation so that initial centers are decided.

The performance of this method is evaluated by isolated-word recognition of Korean digits. The highest recognition rate is 97.6%.

I. 서론

고립단어 인식에 있어 패턴정합의 방법은 높은 인식률 때문에 많이 쓰이고 있다. 패턴정합에 있어서 중요한 문제는 각 단어에 대해서 적당한 기준 패턴을 선택하는 것이다. 그 방법으로서 집단화 기법을 이용하고 있는데 지금까지 여러가지 기법들이 소개되어 왔다. 그러나 대부분의 방법들이 그 과정이 복잡하거나 임계값(threshold) 또는 초기 중심값을 입력해야 하는 단점이 있다[1,2,3]. 집단화의 기법으로서 K-means iteration은 인식 패턴을 만들기 위해 널리 쓰이고 있으나, 초기 중심값을 입력해야 하며, 초기입력에 따라 그 집단화 성능이 좌우되는 local optimum의 성질을 가지고 있다. 따라서 완전 자동화된 알고리듬을 개발하기 위해서는 초기 중심값을 자동적으로 선택할 수있는 방법이 요구된다. 본 논문에서는 초기 중심을 선택하는 알고리듬으로 sorting과 궤적분할(trace segmentation) [4]을 이용한 방법을 제시하였다. 이 알고리듬을 이용한 방법을 Rabiner가 제안한 UWA(unsupervised clustering withoutavetage) 방법 [3]을 인식실험을 통해 비교하였다.

II. 본론

1. K-means iteration

K-means iteration은 집단으로의 분리, 집단의 중심 계산, 수렴 테스트 등 세가지 부분으로 이루어진 반복과정이다. 만약 M개의 집단을 찾고자 한다면 초기 중심으로 M개의 중심을 지정해야 한다. 일단 초기중심이 주어지면 그 방법은 아래와 같다.

1) 초기화 $X_i=x_i$, $1 \le i \le M$ i : cluster 번호
2) nearest neighbor 법칙에 의해 패턴들을 M개의 집단에 귀속
   즉, x_j W_i iff $d(x_j,X_i) \le d(x_j,X_i)$, $1 \le k \le M$
   W_i : i번째 cluster (거리계산은 DTW를 이용[5])
3) 집단의 중심을 계산 -- minmax center 사용
4) 새로운 중심과 과거의 중심이 같거나, 반복수가 주어진 값을 넘으면 끝, 그렇지 않으면 2) 3)을 반복

위와 같은 K-means iteration은 그 방법이 간단하나 초기값 지정에 번거로움이 있고 그 초기값에 따라 집단들의 중심이 달라지는 성질이 있다.

2. K-means iteration의 초기 중심 선택

초기중심을 선택하기 위해서 여러 가지 방법을 생각할 수있다. 본 논문에서는 초기중심을 선택하기위한 방법으로 sorting과 궤적분할 [4]을 이용한 방법을 제시하였다

(1) sorting에 의한 선택

sorting은 기준 패턴에 대해 일정법칙에 따라 줄을 세우는 방법이다. 따라서 sorting에는 여러가지가 있을 수 있으나 본 논문에서는 세 가지의 sorting 방법을 제시하였다.

입력패턴과 sorting 패턴이 아래와 같을 때

입력 token : $x_1,x_2,,,x_N$
sorting된 token : $x'_1,x'_2,,,x'_N$

첫째, chain map에 의한 방법:

$d(x'_1,x'_{i-1}) < d(x'_1,x'_i), \quad i = 2,3,\ldots,N$

둘째, nearest neighbor에 의한 방법:

$d(x'_i,x'_{i+1}) \le d(x'_i,x_k), \quad x_k \notin \{x'_1,,,x'_i\}$

셋째, 누적거리에 의한 방법:

$\min_k(\sum_{j=1}^{i} d(x'_j,x_k)), \quad x_k \notin \{x'_1,,,x'_{i-1}\}$

이다.

위의 세가지 방법을 이용하여 sorting을 한후 분절(segment)를 이용하여 초기집단을 만들게 된다.

(2) 궤적분할에의한 초기집단 선택

궤적분할은 음성인식에서 데이타수를 줄이기 위해서 쓰여져 왔는데 이것은 스펙트럼의 시간열이 있을 때 전후 스펙트럼간의 변화는 정상상태에서 작다는 성질을 이용한 것이다[4]. 즉 x_1 ,... ,x_N을 패턴열 이라 하자. 이때 패턴열의 변화량의 총합을 C라 하면

$$C = \sum_{i=1}^{N} d(x_i - x_{i+1})$$

그러면 이 패턴열을 패턴열의 거리의 합이 C/M가 되는 간격으로 나누면 M개의 정상 상태를 구할 수 있고 이것이 초기 집단이 되어진다.

(3) 초기화 하지 않는 K-meana iteration의 알고리듬 (K-means iteration without initialization, KWI)

KWI알고리듬 블록도는 그림 1과 같다. 여기서 초기 cluster는 위의 세가지 방법을 이용하여 초기 집단을 구한 후 가장 왜곡(distortion)이 작은 것으로 선택하였다. 여기서 평균화의 과정은 Rabiner가 제안한 방법[3]과 다른 방법을 제시하였다. 즉 한 집단내에서 모든 패턴을 기준패턴에 정규화(warping)하여 정규화 함수에 의하여 기준 프레임으로 정규화 시킨다. 그 다음 정규화된 모든 패턴을 시간 축에 따라 합한 후 집단내의 패턴 갯수로 나누는 것이다. 정규화 함수는 i(k)가 기준패턴, j(k)가 비교패턴의 정규화 함수라 할 때 다음과 같다.

i(k)=1,2,...,I ,단 I는 기준패턴의 길이

j(k)=1,2,...,J ,단 j는 비교패턴의 길이

k:공통 시간축 k=1,2,...K

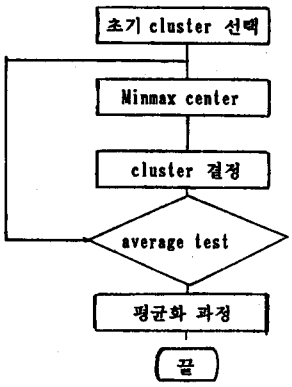

그림 1 KWI 알고리듬

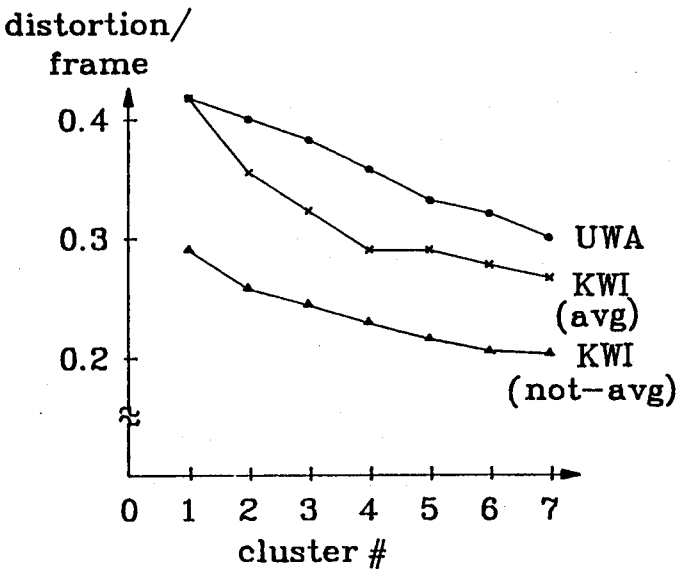

그림 2. KWI와 UWA의 왜곡률 (왜곡률/프레임)

III. 실험및 검토

인식실험은 50명의 화자가 두번씩 발음한 한국어 숫자음(0-9) 1000개의 데이타를 사용하여 수행하였다. 각 음은 20ms의 10차 parcor계수를 통하여 분석하였다. 표 1은 학습과정에 있어서 500 단어에 대한 sorting 방법의 성능을 보여주고 있다. 이 표에 의하면 세가지 방법이 거의 비슷하나 chainmap의 방법이 대체로 우수한 것으로 보여진다. 그러나 각 방법의 성능은 단어나 집단 갯수에 따라 변화하여 어느 하나가 가장 좋다고 결론지을 수는 없었다. 따라서 II장에서 설명을 했듯이 세가지 방법중 가장 왜곡이 작은 것을 초기 중심으로 사용하였다.

그림 2는 Rabiner가 제안한 UWA 방법 [3]과 본 논문에서 제안한 KWI 방법의 프레임당 왜곡을 보여주고 있다. 그림에서 볼 수 있듯이 KWI의 마지막 단계인 평균과정을 생략하여도 왜곡의 입장에서 UWA보다 더 나은 집단화를 수행하고 있음을 알 수 있다. 또한 평균화 과정을 고려하였을 때가 생략한 경우보다 왜곡이 1/3 정도 줄어듬을 알 수 있다.

표 1. 세가지 초기화 방법의 비교

집단갯수	chinmap	nearest-neighbor	누적거리
1	0.417	0.417	0.417
2	0.354	0.364	0.347
3	0.322	0.331	0.330
4	0.288	0.312	0.315
5	0.289	0.292	0.292
6	0.276	0.279	0.279
7	0.266	0.265	0.274

한편 제안한 sorting과 재적분할을 사용한 초기선택이 K-means iteration의 반복 횟수도 2-4회 정도로 작음을 실험과정을 통해 알 수 있었다.

그림 3은 UWA와 KWI의 집단화 방법을 사용했을 때의 인식률을 보여주고 있다. 집단의 갯수는 화자가 50명 정도임으로 5개로 제한하였다. 그림 3에서 볼 수 있드시 KWI의 방법이 평균과정을 이용함에 상관없이 UWA보다 높음을 볼 수 있었다. 또한 평균화의 과정을 사용하여야 화자독립의 인식에 있어 높은 인식률을 얻을 수 있음을 알 수 있었고 제안한 평균화의 과정의 타당성을 보일 수 있었다. 최저 오인식률은 UWA의 경우 11.4%, KWI의 경우 평균을 하지 않았을 때 5.4%, 평균을 했을 때 2.4%이었다.

IV. 결론

본 논문은 집단화 알고리듬인 K-means iteration의 초기화를 선택하는 방법을 제시하였고 인식을 통해 그 타당성을 검토하였다. 초기화는 sorting과 재적분할의 방법을 이용하였다 이러한 초기화를 통했을 때 K-means 수의 반복과정이 작음을 알 수 있었으며 인식률의 면에 있었어도 UWA 방법보다 좋음을 알 수 있었다. 또한 제안한 평균화의 기법이 인식률을 향상시키고 있음을 알 수 있었다. 이 방법을 통한 최고 인식률은 집단의 갯수가 4개일 때 97.6% 이었다.

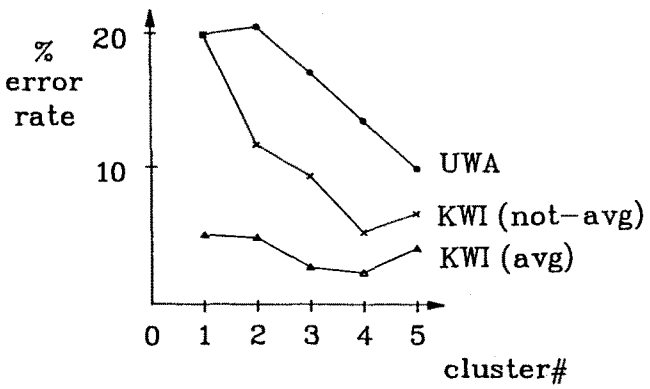

그림 3. KWI와 UWA의 인식률

참고문헌

[1] L.R. Rabiner and A.E. Rosenberg, "Speaker independent recognition of isolated word using clustering techniques IEEE Trans. on ASSP, Vol. ASSP-27, pp. 336-347, Aug. 1979.

[2] S.E. Levinson and L.R. Rabiner, "Interactive clustering techniques for selecting speaker independent reference templates for isolated word recognition," IEEE Trans. on ASSP, Vol. ASSP-27, pp.134-141,1979

[3] L.R. Rabiner and J.G. Wilpon,"Considerations on applying clustering techniques to speaker indenpent word recognition," J.A.S.A., Vol.66, pp.663-673, Sept. 1979.

[4] M.H. Kuhn and H.H. Tomaschewski, "Improvement in Isolated word recognition," IEEE Trans. on ASSP, Vol. ASSP-31, pp.157-167, Feb. 1983.

[5] H.Sakoe and S. Chiba, "Dynamic algorithm optimization for spoken word recognition," IEEE Trans. on ASSP, Vol. ASSP-26, pp.43-49, Feb. 1978.