

Modified ISODATA 방법을 이용한
 불특정화자 단독어 인식
 Speaker-Independent Isolated Word
 Recognition Using A
 Modified ISODATA Method

*황 우 근(Hwang, W. G.)
 안 태 옥(Ann, T. O.)
 이 형 준(Lee, H. J.)
 김 순 협(Kim, S. H.)

요 약

본 논문은 불특정화자의 한국어 단독어인식에 관한 연구로서, 새로운 집단화 방법인 Modified-ISODATA 집단화 방법을 제안한다.

제안된 방법은 종래의 ISODATA 알고리즘에서 외부 고립점 처리 및 분리과정을 단순화 하여, 정확하고도 자동화된 집단의 중심점을 찾는 것을 목적으로 한다.

본 알고리즘을 적용한 결과, 10명의 남성 화자와 4명의 여성 화자가 발음한 11개의 숫자음에 대하여, 최근에 발표된 Modified K-means 방법보다 좋은 인식율을 나타내어, 보다 정확한 집단의 중심점을 찾아 내었음을 입증해 보였다.

ABSTRACT

As a study on Speaker-Independent Isolated Word Recognition, a Modified ISODATA clustering method is proposed.

*광운대학교, 전자계산기공학과

This method simplifies the outlier processing and the splitting procedure in conventional ISODATA algorithm, and eliminates the lumping procedure. Through this method, we could find cluster centers precisely and automatically.

When this method applied to 11 digits by 10 males and 4 females, its recognition rates of 84.42% for K=4 were better than those of the latest Modified K-means, 82.5%. Judging from these results, we proved this method the best method in finding cluster centers precisely.

I. 서 론

인간이 통신을 하는 가장 우선적인 방법은 음성이다. 또한 인간만이 초보 단계를 벗어난 정보를 부호화하고, 전달할 수 있는 음성기관을 발달시켜 왔다.

여기에 디지털 컴퓨터의 개발로 인해, 이 컴퓨터를 이용한 자연스러운 기계와의 통신을 할 수 있도록 하는 연구가 1950년 이후 계속되어 왔다⁽¹⁾. 그때 이후로, 상당히 많은 어휘(100~1,000)까지도 특정 화자 시스템에서는 좋은 성과를 보여 왔으나⁽²⁾, 불특정 화자 시스템인 경우에는 어휘가 상당히 한정되어, 최근 몇몇의 연구자들이 많은 어휘에 대한 불특정 화자 인식시스템을 연구하고 있다. 또한 최근 연구에서는 단어당 다중 표준패턴을 잡아 줌으로써 인식율을 높여 왔다^(3, 4, 5, 6)

집단화 방법에서 Iterative Self-Organizing Data Analysis Techniques A (ISODATA), Chairmap, Shared Nearest Neighbors (SNN) 등은 수작업이 필요하고, Unsupervised Without Averaging (UWA) 은 반자동적인 방법이다. 그러나, 이들 방법은 정확성이 결여되거나 시간소비가 많고 외부고립점 (Outlier) 를 다루지 않았다^(3, 4, 5, 7).

또한 최근에는 제층적 집단화 방법이 소개되었으나, 실제적인 거리계산을 사용하지 않는 순위제 방식으로 신뢰할 수 없고, 집단화를 자동화시킨 최선의 Modified K-means 방법도 외부고립점을 다루지

않아 정확한 집단화를 하지 못하였다^(6, 8).

따라서, 본 논문은 종래의 ISODATA 알고리즘의 외부고립점 처리 및 분리과정에서 파라메타를 줄여, 처리를 간단화시켜 정확하고도 자동화된 집단의 중심점을 찾는 데 그 목적이 있다. 또한, 집단의 중심점을 잡는 방법에 대한 차이를 구별하기 위해 여러 가지 방법 (Minimax, Minsum, Pseudo-Average) 을 사용하였다. 이 방법에 의해 구해진 집단의 중심점은 평균화 과정을 통하여 실제적인 표준패턴을 생성하도록 하였다.

II. 음성 인식을 위한 집단화 방법들

1. 기존의 집단화 방법들⁽⁷⁾

(1) Modified K-means 알고리즘 (MKM)⁽⁸⁾

원하는 집단수가 J일때 K번째 반복인 i번째 집단을 $(W_i^k)^{(i)}$ 으로 표시한다. 여기서 $i=1, 2, \dots, J$ 또한 $K=0, 1, \dots, K_{max}$. (여기서 K_{max} 는 최대 반복 허용 횟수) J값은 2에서 J_{max} (최대 집단갯수)까지 계산한다.

MKM 집단화 과정은 다음과 같다.

- 1) 초기화 $J=1, K=1, i=1$, 거리 행렬 $D=\delta(X_1, X_m), 1 \leq m \leq N$ 계산
- 2) 전체 집합 Ω 의 집단의 중심점 $C(\Omega)$ 를 계산한다. 집단의 중심점은 Minimax 방법과 Pseudo-average, 두 방법으로 각각 구한다.
- 3) $(W_i^k)^{(i)} = \Omega, C[(W_i^k)^{(i)}] = C(\Omega)$ 로 놓음.

- 4) 각 패턴, Ω 의 X_i 을 X_i 에서 집단의 중심점 $C[(W_j^{(k)})]$ 까지 최소 거리를 갖는 집단을 선택 함으로써 집단을 $C(\Omega_j^{(k)})$, $i=1, 2, \dots, J$ 중의 하나에 속하게 분류시킨다.
- 5) 각각의 결과적으로 생성된 집단 $(W_j^{(k)})$, $i=1, 2, \dots, J$ 에 대해 하나의 집단의 중심점 $C[(W_j^{(k)})]$ 을 계산한다. (역시 Minimax 또는 Pseudo-average center 방식으로)
- 6) 만약 이전 반복과 현재의 집단의 중심이 변하지 않았으면 수렴한 것으로 한다. 만약 집단이 변했으면 반복 횟수를 증가시키고 최대 반복 허용 횟수와 비교한다. 만약 반복 횟수가 최대 반복 허용횟수를 넘지 않으면 K-means 반복을 계속 한다. (4~5단계)
- 7) 모든 집단에 대해 평균 집단내의 거리 D_i 를 계산한다.

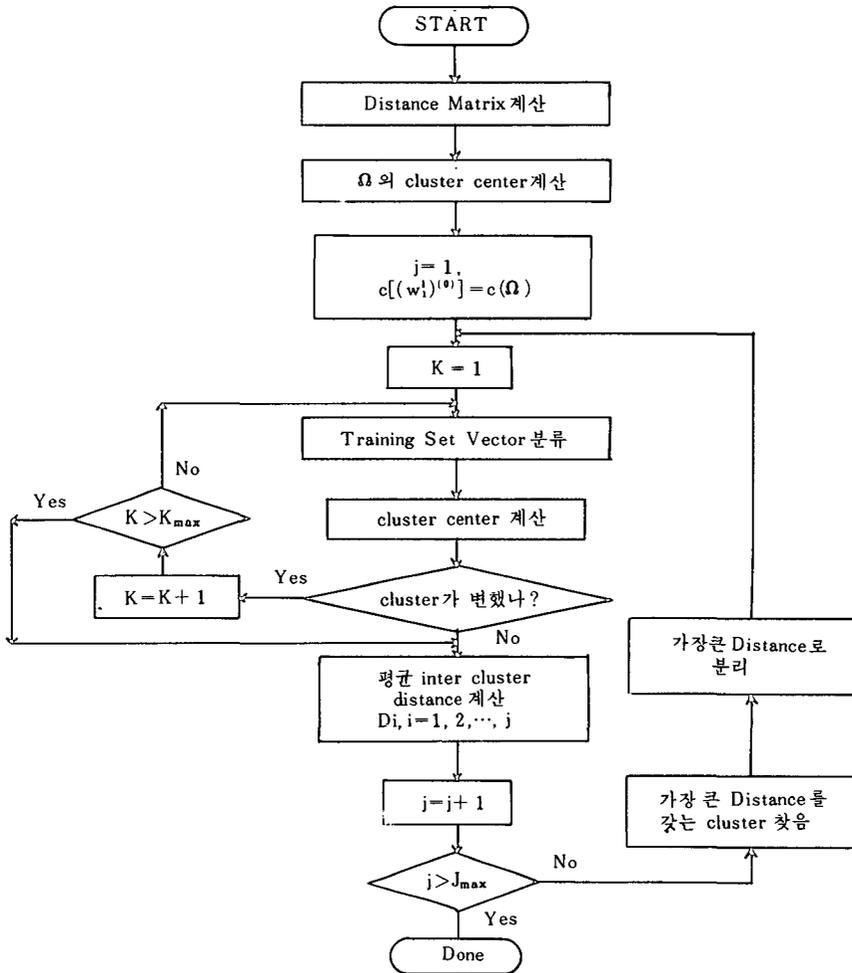


그림 1 Modified K-means 알고리즘의 흐름도.

$$D_i = \frac{1}{|W_j^k|} \sum_{X_l \in W_j^k} \delta(X_l, C[(W_j^k)]) \quad (1)$$

여기서 $|W_j^k|$ 는 마지막 K-means 반복에서 집단 $(W_j^k)^{(k)}$ 에 포함되는 W_j^k 의 패턴 수이다.

- 8) 만약 수렴한 경우는 그 집단 $(W_j^k)^{(k)}$ 을 최적의 J개 집단으로 저장시킨다. 그리고 J를 하나증가시키고 그것의 최대 집단의 크기 J_{max} 와 비교해 본다. 만약 그 비교에서 J가 J_{max} 보다 크면 모든 과정을 끝내고 모든 집단에 대해 마지막 집단의 중심점을 만든다.
 - 9) 만약 J가 J_{max} 보다 적거나 같을 경우는 최대집단내 거리를 갖는 집단 (W_{j-1}^k) 을 찾아서 이 집단은 2개의 집단으로 분리된다. 그리고 반복 횟수는 $M = 1$ 로 다시 정해진다.
 - 10) K-means 반복은 이전에서 처럼 (4~5 단계)정해진다.
- 이것을 흐름도로 표시하면 그림 1과 같다.

(2) ISODATA 알고리즘⁽⁹⁾

ISODATA 알고리즘은 집단의 중심점이 반복적으로 정해진 샘플 평균치라는 면에서 K-means 방법과 유사하다. 그러나 ISODATA는 첨가적으로 'heuristic' 절차를 나타낸다. 이 알고리즘은 실행하기 전에 초기 집단의 중심을 정해야 할 필요가 있다. 그러나 이 초기집단의 중심이, 원하는 집단의 중심과 같은 수일 필요는 없다.

ISODATA는 다음과 같은 원칙적인 단계로 구성 되어 있다.

- 1) 다음의 처리 파라메타를 지정한다.
 - K: 원하는 집단의 중심 수
 - Θ_n : 집단의 영역에서 샘플수와 비교될 파라메타
 - Θ_s : 표준편차 파라메타
 - Θ_c : 묶음 파라메타
 - L: 묶을 수 있는 집단중심의 최대쌍 수

I : 허용되는 반복 회수

- 2) $X \in S_i$, if $\|X - Z_i\| < \|X - Z_j\|$, $i=1, 2, \dots, N_c$; $i \neq j$

관계식을 이용하여 현 집단의 중심에 N 샘플들을 분배한다. (샘플집합에 있어 모든 X에 대해) 여기서 S는 집단의 중심 Z_i 에 할당된 샘플의 부분집합을 나타낸다.

- 3) Θ_n 보다 적은 원소수를 갖는 부분 집합을 버린다. 즉 모든 j에 대해 $N_j < \Theta_n$ 이면 S_j 를 버리고 N_c 를 1 감소한다.
- 4) 각 집단의 중심점 Z_j 가 해당 집합 S_j 의 샘플 평균과 같도록 놓음으로써 각 집단의 중심점 Z_j 를 변경시킨다. 즉,

$$Z_j = \frac{1}{N_j} \sum_{X \in S_j} X, \quad j=1, 2, \dots, N_c \quad (2)$$

여기서 N_j 는 S_j 에 있는 샘플이다.

- 5) $D_j = \frac{1}{N_j} \sum_{X \in S_j} \|X - Z_j\|$, $j=1, 2, \dots, N_c$ (3)

의 관계식을 이용하여 해당 집단의 중심점에서 집단 영역 S_j 에 있는 샘플들의 평균거리 D_j 를 계산한다.

- 6) $D = \frac{1}{N} \sum_{j=1}^{N_c} N_j \cdot D_j$ 의 관계식을 이용해서 그들의 각각 집단의 중심점에서 샘플들의 전체적인 평균거리를 계산한다.
- 7) a) 만약 이것이 마지막 반복이면 $\Theta_c = 0$ 로 놓고 11단계로 간다.
b) 만약 $N_c \leq K/2$ 이면 8 단계로 간다.
c) 이것이 짝수 반복이면 혹은 $N_c \geq 2K$ 이면 11 단계로 간다.
d) 그렇지 않으면 계속
- 8) 각 샘플 부분 집합에 대해

$$\sigma_{ij} = \sqrt{\frac{1}{N_j} \sum_{k \in S_j} (X_{ik} - Z_{ij})^2}$$

$$i = 1, 2, \dots, n; j = 1, 2, \dots, N_c \quad (4)$$

관계식을 이용해서 표준편차 벡터 $\sigma_j = (\sigma_{1j}, \sigma_{2j}, \dots, \sigma_{nj})'$ 를 구한다. 여기서 n은 샘플의 차원수이고, X_{ik} 는 S_j 에 k번째 샘플의 i번째 요소, Z_{ij} 는 Z_j 의 i번째 요소 N_j 는 S_j 에 있는 샘플 수이다. 각 σ_j 요소는 주된 좌표축을 따라 S_j 에 샘플의 표준편차를 나타낸다.

- 9) 각 $\sigma_j, j=1, 2, \dots, N_c$ 의 최대 요소를 발견하고 그것을 σ_{jmax} 로 한다.
- 10) 모든 $\sigma_{jmax}, j=1, 2, \dots, N_c$ 에 대해서 만약 $\sigma_{jmax} > \Theta_s$ 이고 a) $D_j > D$ 이고 $N_j > 2 (\Theta_n + 1)$ 혹은 b) $N_c \leq K/2$ 일때 Z_j 를 Z_{j+} 와 Z_{j-} 의 새로운 2 집단의 중심점으로 나누고 Z_j 를 없애고 N_c 를 1증가시킨다. 집단의 중심점 Z_{j+} 는 σ_j 의 최대 요소에 해당하는 Z_j 의 요소에 주어진 변동량 Γ_j 를 더하여 만들어지고, Z_{j-} 는 Z_j 의 같은 요소에 Γ_j 를 빼서 형성된다. Γ_j 를 정하는 한 방법은 Γ_j 를 σ_{jmax} 의 몇개 부분과 같도록 놓는 것이다. 즉, $\Gamma_j = k \cdot \sigma_{jmax}$ 여기서 $0 < k \leq 1$ 이다. Γ_j 를 고르는데 있어서 주된 요구사항은 임의의 샘플에서 두개의 새로운 집단의 중심점까지의 거리에 감지할 수 있는 차를 충분히 제공해야 한다는 것이다. (그러나 전체적인 집단 영역 배정을 변화시킬 정도로 크면 안된다.) 만약 이 단계에서 분리가 이루어지면 2단계로 가고 그렇지 않으면 계속한다.
- 11) 모든 집단의 중심점들 간의 쌍간의 거리 D_{ij} 를 계산한다.

$$D_{ij} = \|Z_i - Z_j\|, \quad i = 1, 2, \dots, N_c - 1; j = i + 1, \dots, N_c \quad (5)$$

- 12) 거리 D_{ij} 와 파라메타 Θ_c 와 비교해서 Θ_c 보다 작은 L개 가장 작은 거리를 올림차순으로 배열한다.

$$[D_{i_1j_1}, D_{i_2j_2}, \dots, D_{i_Lj_L}] \quad (6)$$

여기서 $D_{i_1j_1} < D_{i_2j_2} < \dots < D_{i_Lj_L}$ 이고 L은 함께 묶어질 수 있는 집단의 중심점들의 쌍의 최대 수이다.

- 13) 각거리 $D_{i_1j_1}$ 을 갖고, 관계된 한쌍의 집단의 중심점들 Z_{i_1} 와 Z_{j_1} 이 존재한다. 이를 거리의 가장 작은 값을 가지고 시작해서 다음 규칙에 따라 한쌍의 묶음동작을 수행한다.
 $l = 1, 2, \dots, L$ 에 대해, 만약 Z_{i_l} 혹은 Z_{j_l} 이 이 반복에서 묶는데 사용되지 않았으면, 다음 관계식

$$Z^* = \frac{1}{N_{i_l} + N_{j_l}} [N_{i_l}(Z_{i_l}) + N_{j_l}(Z_{j_l})] \quad (7)$$

의 관계식을 사용해서 이들 두 집단의 중심점들을 합병한다. Z_{i_l} 과 Z_{j_l} 을 지우고 N_c 를 1 줄인다. 단지 한쌍의 묶음 동작이 허용되고 묶여진 집단의 중심점은 그 영역에 있는 샘플들의 수에 의해 각각의 오래된 집단의 중심점을 가중하여 얻어진다.

실험적인 결과는 더 복잡한 묶음 동작이 불만족스런 결과를 만든다. 위의 과정은 묶여진 집단의 중심점을 결합된 부분 집합들의 실제적인 평균점으로 대표할 수 있게 한다. 집단의 중심점은 단지 한번 묶을 수 있기 때문에 이 단계는 항상 L개 묶여진 집단을 만들수 없을 것이다.

- 14) 만약 이것이 마지막 반복이면, 이 알고리 즘은 끝낸다. 그렇지 않으면 처리 파라메타의 어떤 것이 사용자의 지식에 의해서 변화되기를 원하면 1단계로 가고, 만약 파라메타들이 다음 반

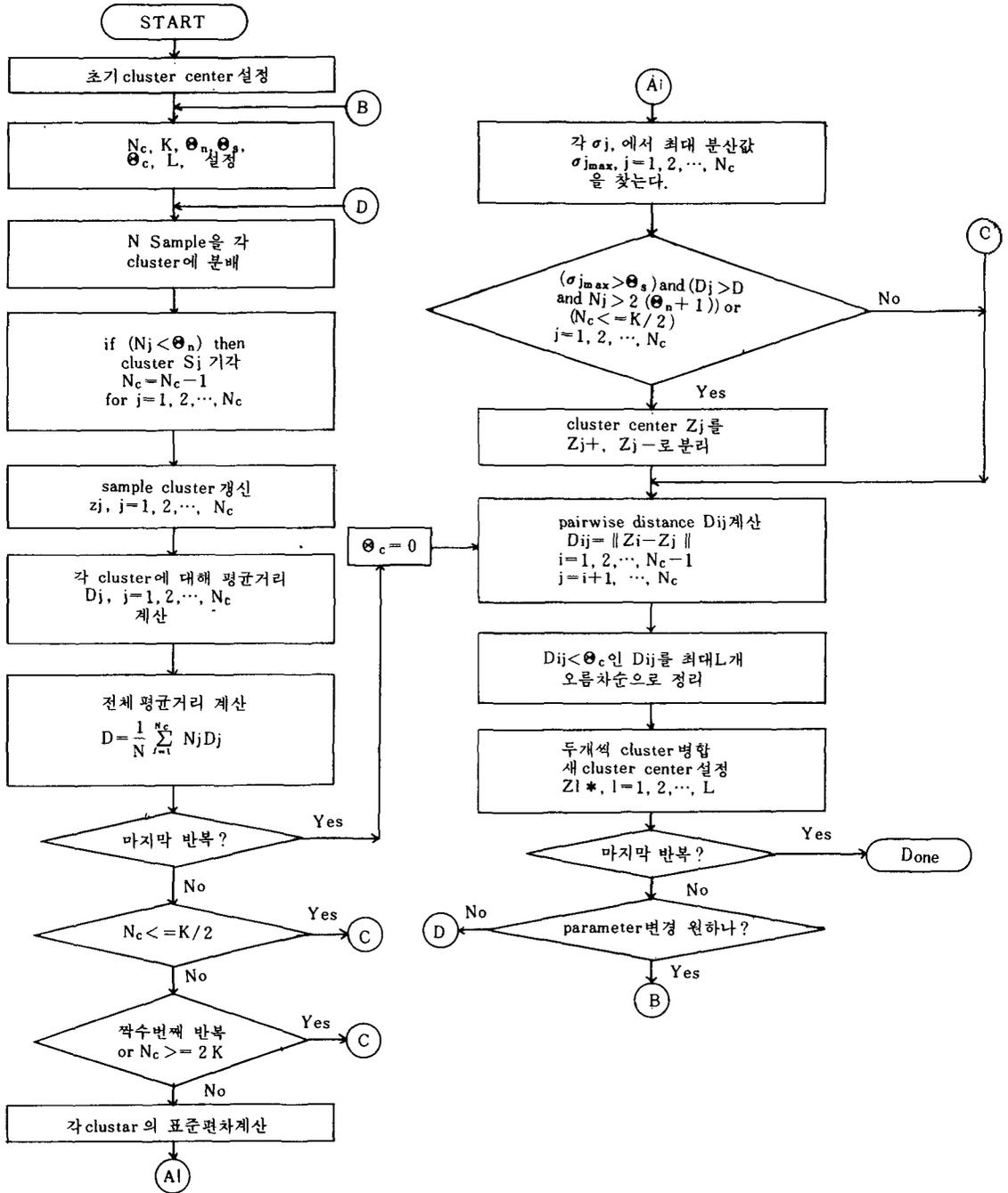


그림 2 ISODATA 알고리즘의 흐름도.

복에서 같은 것으로 남아 있으면 2 단계로 간다. 반복은 절차가 1 단계 혹은 2 단계로 복귀할 때마다 계산한다.

이것을 흐름도로 표시하면 그림 2 과 같다.

2. 기존의 집단의 중심점 계산 방법들

(1) Minimax 방법⁽⁷⁾

만약 n 개 패턴을 가진 집단 A 를 $A = \{X_1, X_2, \dots, X_n\}$ 으로 표시하면

$$\bar{X} = X_1 * \max_{1 \leq l \leq N} d(X_l *, X_l) \leq \min_{1 \leq m \leq N} \max_{1 \leq l \leq N} d(X_m, X_l) \quad (8)$$

여기서 $d(a, b)$ 는 a 와 b 간의 거리를 말한다. 즉, Minimax 중심점은 A 에 있는 다른 모든 패턴에 대한 최대거리가 그 집단에 있는 모든 패턴에 대해 최소가 되는 집단에의 패턴이다.

(2) Minsum 방법⁽¹⁰⁾

집단에 모든 다른 패턴들에 대한 거리의 합이 최소가 되는 패턴을 집단의 중심점으로 정한다.

$A = \{X_1, X_2, \dots, X_n\}$ 일때 Minsum 집단의 중심점은

$$\bar{X} = X_1 * \ni \min_{1 \leq l \leq N} \sum_{i=1}^N d(X_l *, X_i) \quad (9)$$

이 방법과 Minimax 와의 차이는, Minimax 에서의 비교하는 과정이 Minsum 방법에서는 덧셈으로 대체된다.

(3) Pseudo-average 방법⁽⁸⁾

각 집단 W 에 대해 평균거리 D 를 구하고 모든 점들간의 표준편차를 구한다.

$$D = \frac{1}{J(J-1)} \sum_{i=1}^J \sum_{\substack{m=1 \\ m \neq i}}^J \delta(X_i, X_m) \quad (10)$$

$$sd = \sqrt{\frac{1}{J(J-1)} \sum_{i=1}^J \sum_{\substack{m=1 \\ m \neq i}}^J \delta^2(X_i, X_m) - (d)^2} \quad (11)$$

l 번째 패턴에 대해, 패턴 X_l 에 대한 거리가 경험적으로 선택된 문턱치 $T = D + 0.5 * sd$ 보다 작은 패턴들의 수를 세어서 각 토큰에 대한 카운트 CI 을 계산한다. 가장 큰 갯수를 가진 패턴 X_l 의 가상적인 집단의 중심점으로 선택된다. 만약 두 패턴의 갯수가 똑같은 경우에는 그 집단내에서 모든 토큰에 대한 평균거리가 가장 적은 패턴이 집단의 중심점이 된다.

3. 집단의 중심점들을 구하기 위해 사용된 평균화 기법⁽⁷⁾

이 방법을 설명하기 위해서 전체 토큰집합 Ω 에서 두개 토큰 x, y 를 고려해 보면, 토큰 x 는 N_x 프레임의 LPC 계수를 갖고 있고 토큰 y 는 N_y 프레임의 LPC 계수를 갖고 있다고 생각하자. 각 프레임은 $P+1$ 차의 자동상관계수를 갖는다. 만약 x (혹은 y) 의 i 번째 프레임을 $x(i)$ (혹은 $y(i)$) 로 쓰면 x, y 는 각각 다음과 같은 벡터를 갖는다.

$$x = (x(1), x(2), \dots, x(i), \dots, x(N_x)) \quad (12-1)$$

$$y = (y(1), y(2), \dots, y(i), \dots, y(N_y)) \quad (12-2)$$

여기서 $x(i) = (x_0(i), x_1(i), \dots, x_p(i)), y(i)$ 의 경우도 동일함.

토큰 x, y 를 평균하기 위해 x 의 프레임들과 y 의 프레임을 일치시켜야 한다. 그래서 DTW 절차가 정합하기 위해 사용된다.

$$x(i) \rightarrow y(k) = y(w(i)) \quad (13)$$

즉, $k = w(i), i = 1, 2, \dots, N_x$

토큰 x, y 를 평균할 때 다음과 같은 Z 이 얻어진다.

$$Z = (z(1), z(2), \dots, z(Nx)) \quad (14)$$

$$\text{여기서 } Z(i) = 1/2 * [x(i) + y(w(i))]$$

그리고 Z(i)의 k 번째 요소 Zk(i)는

$$Zk(i) = 1/2 * [xk(i) + yk(w(i))] \quad (15)$$

로 얻어진다.

전체집합 Ω에서 Q개 토큰을 평균할 때 각 토큰을 집단의 예상집단에 연속적으로 위평한 다음 1/Q로 나누어 평균화 한다. 여기서 얻어진 집단의 중심점이 실제적인 표준치 값이 된다.

III. Modified ISODATA 집단화 방법을 이용한 템플레이트생성

1. Modified ISODATA 알고리즘의 도입

종래의 ISODATA 알고리즘의 경험적인 파라메타 값을 근거로 한 실험에서, 그 경험적인 값의 애매성 때문에 파라메타 값 변경에 따른 계산시간의 과다 때문에, 단순화시킨 ISODATA 알고리즘을 사용하였다.

즉, 종래의 원하는 집단의 중심점 수를 정하는 파라메타인 K 값을 정할 경우에도 꼭 원하는 개수의 집단 중심점 수가 생기지 않을 수 있는 경우를 제거하여 원하는 개수의 집단의 중심점 수를 위주로 하는 알고리즘으로 변경하였으며, 또한 이 변경된 ISODATA 알고리즘에서는 분리시 파라메타로 사용되는 표준편차 파라메타를 제거시키고, 묶음파라메타에 수반되는 묶을 수 있는 집단의 중심점의 최대 쌍수 파라메타를 자연히 제거시킬 수 있도록하였다. 따라서 사용된 파라메타는 원하는 집단의 중심점수 K, 집단영역에서 샘플수와 비교될 파라메타 Θ_n , 허

용되는 반복횟수 MM 등으로 축소시켜 수행시간 및 계산량을 상당히 줄일 수 있도록 하였다.

2. Modified ISODATA 알고리즘

이 알고리즘은 흐름도는 그림 4에 나타내고 있다. Modified ISODATA 알고리즘의 처리는 다음과 같다.

- 1) 한 음성에 대한 모든 훈련데이터의 거리 행렬을 작성한다.
- 2) 허용하는 반복횟수 $MM = 1$, $C[(W_i)^{(0)}] = C(\Omega)$ 전체반복횟수 $IT = 0$ 로 초기화시킨다.
- 3) 전체 반복횟수를 하나 증가시킨다.
- 4) 모든 집단의 패턴수가 샘플수와 비교될 파라메타 Θ_n 보다 작거나 같으면 그리고 분리 과정에서 분리현상이 일어났다는 표시인 $JFLAG = 1$ 이 아닌 경우에는 외부고립점 처리를 행한다. 이 경우, 전체 집단수를 하나 감소시킨다.

$$N_c = N_c - 1$$

- 5) 모든 데이터들에 대해서 기존의 집단의 중심점에 대해 가장 가까운 집단의 중심점에 속 하도록 재할당 한다.

$$x_p \in W_k \text{ iff } \text{Min } \delta(x_p, C[(W_k)^{(m)}]) \quad (16)$$

$$p = 1, \dots, N, i = 1, \dots, N_c$$

분리과정이 발생했다는 표시인 $JFLAG$ 를 0으로 고정시켜 분리과정의 영향을 받지 않음을 표시한다.

- 6) 모든 집단에 대해 새로운 집단의 중심점을 계산한다.
여기에서는, Minimax와 Minsum 방법으로 모든 집단의 중심점을 구한다.
- 7) 만약, 현재 집단의 수 N_c 가 원하는 집단의 수 K 보다 작으면 분리과정을 수행시킨다.
 i 번째 집단이 분리될때 이것은 식(17)과 같이 두

부분으로 나누어 진다.

$$w_i = w_i + U w_i - \quad (17)$$

이러한 분리의 예를 그림 3 에 제시하고 있다. 다시 말하면 $\delta(x+, x-)$ 가 최대가 되는 두 패턴 $x+$ 와 $x-$ 를 찾고 $x+$ 나 $x-$ 에 대한 거리의 차가 비교적 작은 쪽으로 w_i 의 각 토큰은 w_i+ 나 w_i- 에 배정되며 이때 $x+$ 와 $x-$ 는 새로 결정되어야 하는 집단의 중심점으로는 적합치 못하기때문에 보다 좋은 중심점을 결정해 주기 위하여

$$r+ = \delta(x+, x_p^{(k)}) \quad (18-1)$$

$$r- = \delta(x-, x_p^{(k)}) \quad (18-2)$$

와 같이 정의하고

$$e+ = \delta(x+, x_p^{(k)}) + \delta(x_p+, x_p^{(k)}) - r+ \quad (19-1)$$

$$e- = \delta(x-, x_p^{(k)}) + \delta(x_p-, x_p^{(k)}) - r- \quad (19-2)$$

위 두식이 최소가 될 수 있는 두 패턴 x_p+ 와 x_p- 를 집단의 중심점으로 결정해 준다.

또한 분리과정을 수행한 표시를 JFLAG=1로

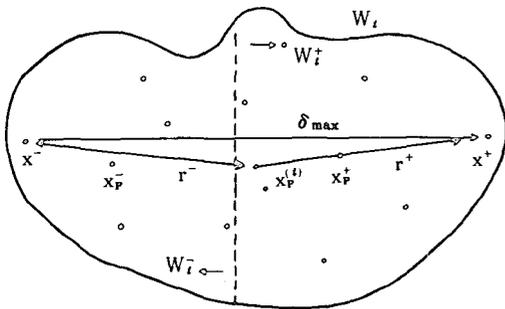


그림 3 분리도.

정한다.

이것은 분리과정 즉시 외부고립점을 제거하는 것보다 안정적이기 때문이다.

- 8) 집단의 중심점이 이전의 반복과정과 비교해서 변하였나를 검사한다. 변했으면 4) 단계로 간다.
- 9) 집단의 중심점이 변하지 않는 경우에는, 이 수렴 과정표시가 반복 최대 허용횟수 M_{max} 보다 큰 값이 될때까지 4) 단계로 되돌아 가는 반복을 수행시킨다.
- 이것은 한번의 수렴으로 안정된 집단의 중심점을 얻을 수 없다는 것에서 기인한 것이다.
- 10) 집단의 중심점이 반복 최대 허용횟수를 넘어 안정된 수렴을 나타낸 경우에는 현재집단의 중심점과 집단에 속한 패턴을 저장시킨다. 원하는 집단의 중심점수를 증가시켜 3) 단계로 가서 이 과정을 반복 수행시킨다.
- 11) 집단의 수가 최대 허용 집단수를 넘으면 동작을 끝낸다.

이 알고리즘의 흐름도는 다음과 같다.

IV. 인식 실험 결과 및 고찰

본 연구에서는 인식 실험에 단독 숫자음 11자 (/영/, /공/, /일/, /이/, /삼/, /사/, /오/, /육/, /칠/, /팔/, /구/)를 성인 남성 10인과 여성 4인이 3번씩 발음한 것을 2번씩의 발음을 표준패턴을 형성하는 데이터로 사용하였고 나머지 1개를 인식을 위한 입력 데이터로 사용하였다.

1. 인식시스템 구성 및 실험

(1) 인식시스템 구성

실험에 사용된 모든 숫자음은 릴테이프로 방음장치가 양호한 스튜디오에서 녹음하였다.

이 릴테이프를 AKAI GK-636 스테레오 테이프

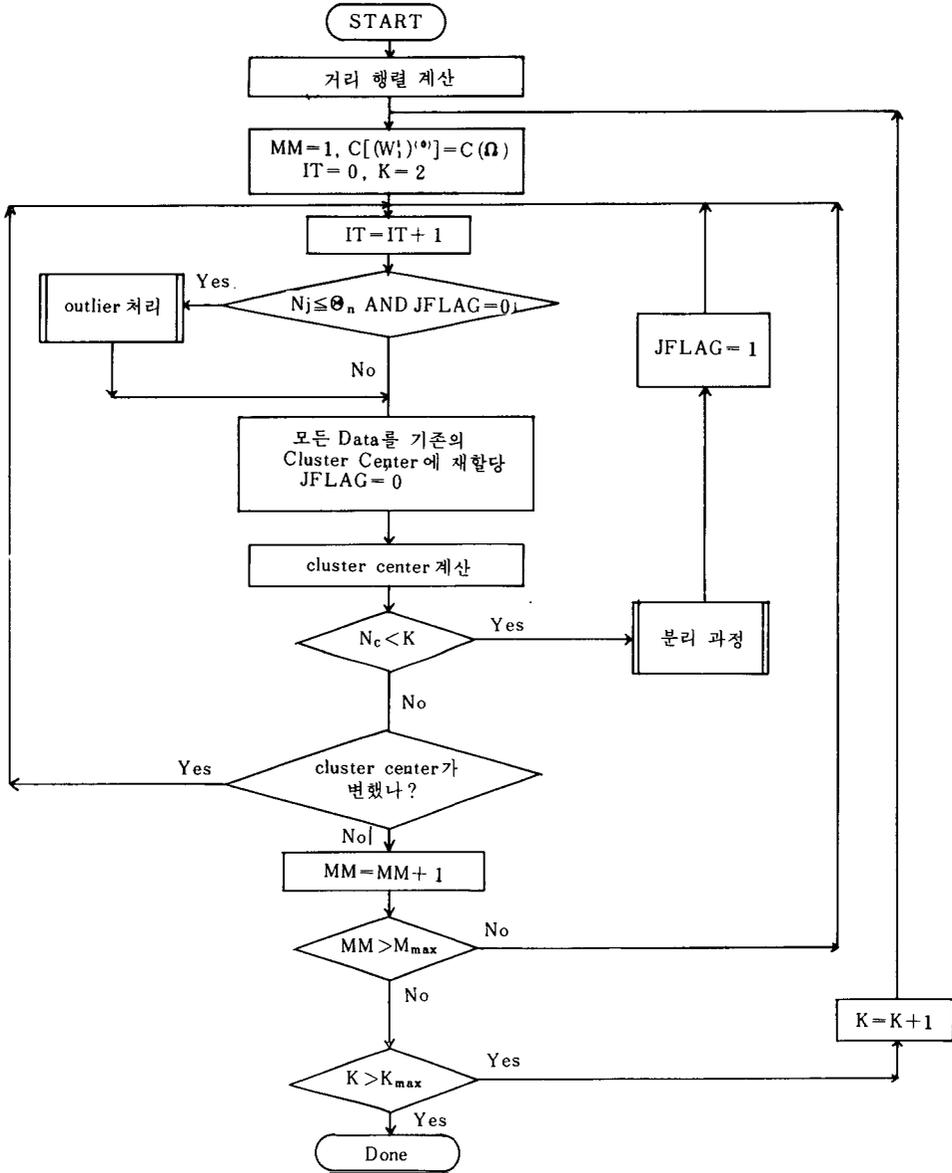


그림 4 Modified-ISODATA 알고리즘의 흐름도.

레코더를 이용하여 최대출력 전압을 1[V]가 되도록 조절하여 차단주파수가 220Hz-4.5KHz인 대역 통과 필터를 통과시킨 다음 MINC에 연결된 Pre-amplifier를 통하여 Preamp되고 12bit A/D 변환기

에 의해 12레벨로 양자화한다. A/D 변환기에서 10 KHz의 샘플링 주파수로 음성신호를 VAX-11/750 미니 컴퓨터에 저장시켜 12차 자동 상관 계수와 12차 LPC 계수를 표준패턴과 시험패턴의 특징 벡터로

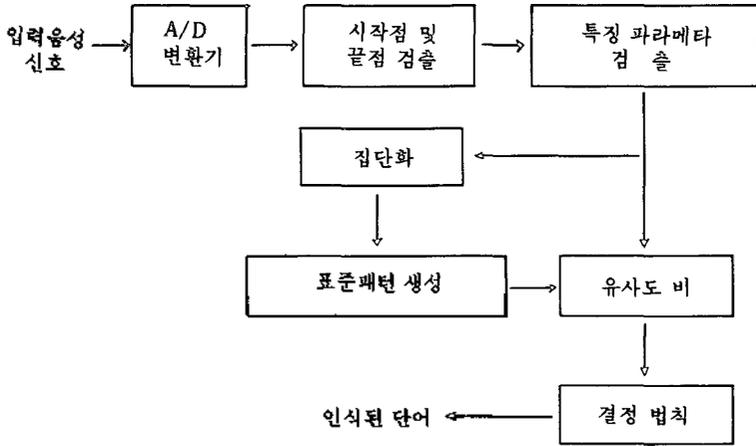


그림 5 인식시스템 구성.

사용하여 집단화과정을 거친 표준패턴과 미지의 입력시험패턴 간의 인식실험을 수행하였다. 이때 사용한 거리함수는 LPC 대수 확률 거리 측정법이다⁽¹¹⁾.

$$d = \log \left[\frac{a_R \cdot V_T \cdot a_R^t}{a_T \cdot V_T \cdot a_T^t} \right] \quad (20)$$

여기서 a_R 과 a_T 는 표준패턴과 시험패턴의 선형예측 계수벡터들이고, V_T 는 시험패턴의 자동 상관 계수 행렬이다.

(2)인식 실험

가) Modified ISODATA 알고리즘에서 다음 경우 각각에 대해 구해진 실제적인 집단의 중심점들을 템플레이트로 잡고 $K=2 \sim 4$ 에 대해 인식실험을 행하였다.

- a) 집단의 중심점 결정 방법을 Minimax로 사용한 경우
- b) 집단의 중심점 결정 방법을 Minsum으로 사

용한 경우

나) 비교실험을 위해서 Modified K-means 알고리즘에 대해 $K=2 \sim 4$ 인 경우에 다음의 두가지 집단의 중심점을 잡는 방법을 사용하여 인식실험을 행하였다.

- a) 집단의 중심점 결정 방법을 Minimax로 사용한 경우
- b) 집단의 중심점 결정 방법을 Pseudo-average로 사용한 경우

2. 인식 결과 및 고찰

(1)수행 시간

수행 반복 횟수에 대해서는 Modified K-means에 비해 Modified ISODATA 방법이 Minimax의 경우에는 많아지나 Modified K-means의 Pseudo-average 방법보다 Modified ISODATA의 Minsum 방법이 수행횟수가 상당히 적어진다.

(2) 인식율

표 1 인식결과(단, 인식율의 단위는(%)이다.)

	Modified K-means						Modified ISODATA					
	MINIMAX			Pseudo-Average			MINIMAX			MINSUM		
K	2	3	4	2	3	4	2	3	5	2	3	4
영	7	11	10	10	8	13	7	11	13	9	11	9
공	9	7	8	3	5	2	9	7	8	8	6	8
알	11	10	11	8	8	7	11	10	10	4	8	11
이	12	14	14	12	13	13	12	13	14	12	12	11
삼	9	8	9	8	7	8	9	8	7	4	7	6
사	10	11	12	10	9	11	10	12	14	10	10	11
오	11	13	13	10	11	13	11	13	13	10	12	13
육	13	12	13	12	12	11	13	13	13	7	12	11
찰	10	12	14	6	14	13	10	11	14	10	13	13
팡	11	11	12	5	12	12	11	11	13	12	11	11
구	10	11	11	12	8	8	10	11	11	10	11	11
계	113	120	127	96	107	111	113	126	130	96	113	115
인식율	73.4	77.9	82.5	62.3	69.5	72.1	73.4	77.9	84.4	62.3	73.4	74.7

(여기서, 최대집단개수는 4, 최대반복횟수는 MKM방법에서는 12(Minum인 경우), $n=1$)

이 인식의 결과표로부터 다음과 같은 사실을 알 수 있었다.

- 1) Modified ISODATA 알고리즘이 Modified K-means 알고리즘보다 최소한 같거나 나은 인식율을 나타내고 있다.
- 2) Modified K-means 알고리즘에서 Minimax 방법이 Pseudo-Average 방법보다 인식율이 훨씬 좋다는 것을 나타내준다. 이것은 외국의 연구 결과와는 다른 현상이며, 이 이유는 Pseudo-Average 방법의 문턱치 T값이 경험적으로 얻어진 값이므로 일반적인 경우에 동일한 값으로 적용하기 어렵다는 것을 보여준 것이다.
- 3) Modified ISODATA 알고리즘에서는 Minimax 방법이 Minsum 방법보다 나음을 알 수 있다.
- 4) 이 실험에서는 K=2, 3, 4에 대해 실험하여 K가 증가함에 따라 인식율이 높아지고 있음을 나

타내 준다.

- 5) 단어별로 볼때, '공', '삼'의 인식율이 다른 단어에 비해 떨어지는 것을 알 수 있다. 이것은 '공'이 '오'나 '구'로 오인되는 경우가 있었고, '삼'이 '사'로 오인되는 경우가 발생되었기 때문이다. 이것은 시작점과 끝점의 잘못된 검출의 영향뿐만 아니라 집단의 중심점을 평균화하는 데서 생기는 오류임을 알 수 있다.
- 6) K=5 이후에는 Modified K-means 알고리즘에서는 집단내의 패턴개수가 1개인 경우가 생겨 집단화의 의미를 부여할 수 없어, K=4까지로 한정시켜 실험하였다.

V. 결 론

본 논문에서는 단순화시킨 ISODATA 알고리즘을 이용한 한국어 숫자음의 불특정 화자에 대한 표준 패턴 생성을 제안하였다.

본 알고리즘의 특징을 외부고립점을 제거하여 정확한 집단을 형성하고 그에 따른 집단의 중심점을 정확히 정할 수 있고, 또한 복잡한 ISODATA 알고리즘을 단순화시킴으로서, 많은 파라메타 값을 변형시킴으로서 생기는 많은 반복횟수와 계산시간, 기억용량 등을 최소화 시킬 수 있다.

그 결과, 최선의 Wilpon이 제안한 Modified K-means 방법보다 좋은 인식률을 나타내었다. 이것은 외부고립점을 제거하여 생기는 정확한 집단의 중심점을 찾았음을 입증하는 것이다. 또한 평균화 과정을 이용하여 실제 패턴을 표준 패턴으로 사용하는 것보다 인식율이 좋다는 사실을 알 수 있었다. 이것은 데이터 구조가 영성한 경우에는 평균화 과정에 의해 집단의 중심점이 정확히 중심되는 부분으로 이동해간다는 것을 보여준 것이다.

앞으로, 음성의 정확한 시작점과 끝점을 검출할 수 있는 표준적인 방법제시와 공통적으로 사용하여

인식결과를 비교할 음성의 데이터 베이스가 필요하며, 이것을 바탕으로 많은 화자(100명 이상)를 대상으로 실험하여 인식 결과의 신뢰성을 확인할 수 있을 것이다. 또한 거리계산을 하는 과정에서 전체 경로제약에 따른 영향까지 보아야 할 것이다.

참 고 문 헌

1. D.R. Reddy, "Speech Recognition by Machine; A Review," Proc. IEEE, Vol.64, No.4, pp.501-531, Apr. 1976.
2. 김순협, "한국어 음성의 분석과 자동인식에 관한연구", 박사학위논문, 연세대학 대학원, 1982. 12.
3. L.R. Rabiner, "On Creating Reference Templates for Speaker Independent Recognition of Isolated Words," IEEE, Trans. on A.S.S.P. Vol. ASSP-16, No.1, pp.34-42, Feb. 1978.
4. L.R. Rabiner and J.G. Wilpon, "Considerations in Applying Clustering Techniques to Speaker-Independent word Recognition, J.A.S.A. Vol.66, No.3, pp.663-673, Sept. 1979.
5. L.R. Rabiner, S.E. Levinson, A.E. Rosenberg and J.G. Wilpon, "Speaker-Independent Recognition of Isolated Words using Clustering Techniques," IEEE. Trans. on A.S.S.P. Vol. ASSP-27, No. 4, pp.336-349, Aug. 1979.
6. T. Kaueko and N.R. Dixon, "A Hierarchical Decision Approach to large-Vocabulary Discrete Utterance Recognition," IEEE, Trans. on A.S.S.P. Vol.ASSP-31, No.5, pp.1001-1066, Oct. 1983.
7. S.E. Levison, L.R. Rabiner, A.E. Rosenberg and J.G. Wilpon, "Interactive Clustering Techniques for Selecting Speaker-Independent Reference Techniques for Selecting Speaker-Independent Reference Techniques for Isolated Word Recognition," IEEE, Trans. on A.S.S.P. Vol.ASSP-27, No.2, pp.134-141, Apr. 1979.
8. J.G. Wilpon and L.R. Rabiner, "A Modified K-Means Clustering Algorithm for use in Isolated Word Recognition," IEEE, Trans. on A.S.S.P. Vol.ASSP-233, No.3, pp.587-594, June, 1985.
9. J.T. Ton and R.C. Gonzalez, Pattern Recognition Principles, Massachusetts, Addison-Wesley Publishing Company. pp.75-109, 1974.
10. 장익현, "독립화자코덱 단어 인식에 있어서의 CLUSTERING 방법에 관한 연구", 석사학위 논문, KAIST, December, 1985.
11. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE, Trans. on A.S.S.P. Vol. ASSP-23, pp.67-72, Feb. 1975.