論 文

論文 69-9-2-1

한글 字體의 特徵抽出의 한 方法

(A Method of Korean Character Pattern Extraction)

姜 麟 求*
(Kang, In Ku)

李 幸 世**
(Lee, Hang Sei)

요 약

한글 자체의 특성을 제시하고 이에 따른 식별방식을 우선 자모에 대해서 전개하였다· 자모의 특징을 중복 없이 나타내는 방식으로 결합 행렬이 적당함을 발견하고, 각 자모에 대한 결합 행렬(行列)을 작성한 것이다]

ABSTRACT

The characteristics of Korean characters are shown and a method of recognition is accordingly developed for the elements.

It is found that a linking matrix is adequate to identify the Characteristics of each element uniquely.

The linking matrics for elements are included in this report.

I. 서 론

전자계산기가 우리나라에도 본격적으로 도입될 단계에 이르렀으므로 멀지 않아 입력 기구로서 한글을 직접 읽는 기계가 개발되리라 믿는다 그런데 한글의 자체는 그 구성이 다른 문자와 다를 뿐 아니라 복잡하다. 그래서 우선 한글 자체의 특성을 규명하고 자모에 대해서 몇 가지의 방법을 시도해 보았다. 그 결과 결합 행렬이 적당하므로 이를 우선 보고하고저 한다.

글씨를 기계로 식별하고저 하는 연구는 전자계산기의 발달과 더불어 활발해지었으며, 벌써 숫자나 영어를 판독하는 기계가 실용되고 있으며 일본 문자에 대한 식별연구도 최근 활발하며 (6, 7) 중국문자에 대한 연구도 이루어지고 있다 (2).

식별의 과정은 보통 두개의 단계로 구분 연구된다. 즉 첫째는 식별하고져 하는 문자의 특징을 가려내는 것이며 둘째로는 이특징과 실제 문자를 비교하여 식별하는 통계학적 방법론이다. 이 글은 첫째 문제에 관한 것으로 특히 자모의 특징 추출(抽出)에 중점을 두었다.

문자의 특징을 나타내는 방법으로 가장 기본적인 것은 문자전체를 하나의 격자(格子)로 겹쳐 봐서 흑백(黑白)에 따라 1과 0으로 나타내는 행렬로 표시한다〔2, 3, 5〕. 이 방법은 상당한 수의 기억이 필요할 뿐만 아니라 문자의 대소에 따라 변하지 않도록 정상화(正常化)의 작업이 필요하다.

주사(Scan)의 방향을 여러개로 정해서 각 방향에 따른 주사(走査)중 문자와 교차하는 회수를 가지고 특징을 구하는 방법도 있다〔1〕. 또한 선의 이동방향의 각도를 나타내는 회진지수(Rotation Index)로 특징을 표시하는 법도 있다〔7〕

일본 가다가나의 식별방식으로는 절점사이를 연결하는 가지(Branch)의 방향으로 특징을 표시하였다〔6〕.

Ⅱ. 한글 글자의 특성

한글 글자는 두개 또는 그 이상의 자모가 합

*海軍士官學校 敎授部
Dept. of Electrical Engineering, ROK Naval Academy

接受 : 1969年 4月20日

성하여 하나의 글자를 이룬다. 그래서 글씨 하나 하나를 하나의 단위로 봐서 식별한다면 흔히 쓰이는 글자만도 1000자가 넘으므로 식별이 매우 비능률적이다. 반면에 자모는 24자이므로 글자를 일단 분해한 다음에 각각의 자모에 대해서 식별을 한 후에 다시 이를 조합하여 한 글자를 인식하는 편이 기억의 용량이나 식별의 절차가 간편하다.

한글 글자의 구성은 표1과 같은 종류로 분류할 수 있다. 이를 요약하여 그림1에서 다음과 같은 일반적인 법칙을 마련할 수 있다.

1. A의 위치에서 반드시 자음 자모가 와야한다(단자음, 중자음)

2. B의 위치에 자모가 있으면 반드시 모음 자모이다.

3. C의 위치에 두개의 자모가 상하로 존재하면 (C_1C_2) C_1은 모음 자모이고 C_2는 자음 자모이다. 두개의 자모가 인접 좌우로 있으면 쌍 받침이다.

4. C의 위치에 자모가 하나이면 그자모의 성격은 확정적이 아니다.

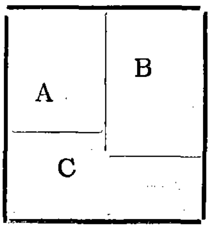
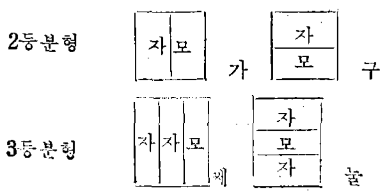
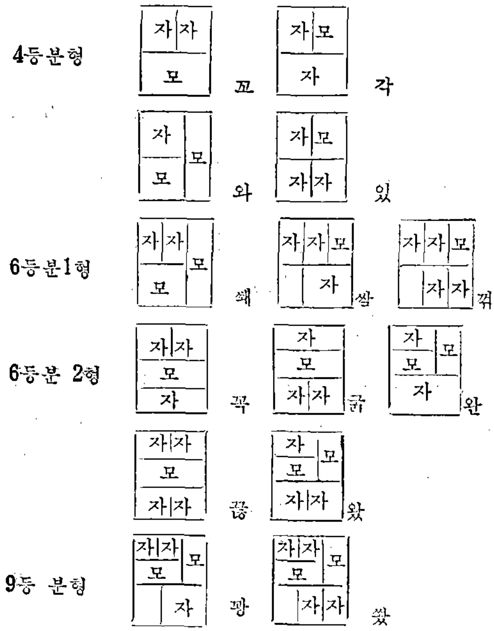


그림 1. 자모의 분포

모음 자모의 복합체중 ㅐ, ㅒ, ㅔ, ㅖ는 단일 자모로 간주하는 것이 글자분해를 쉽게하므로 단일 자모로 간주하여 28개의 자모가 있는 것으로 한다.

표1. 글자의 구성분류





글자를 자모로 풀이하는 방법은 흑점이 응집한 부분을 따로 떼어 놓은 후에 각각을 ABC로 나눈 후 일반 원칙에 어긋남이 없는가 보고 그 다음 해당되는 구성 표1과 대조한 후에 각각의 자모로 분리하여 각 자모를 식별한다.

Ⅲ. 자목의 식별방법

흑백에 따르는 행렬에 의한 방법은 경제적이 아니므로 고려하지 안했으며 주사(走査) 방향에 따른 교차회수도 우리 자모의 단순성 때문에 과히 좋지 못하다. 회전지수에 의한 방법 역시 부적당하다. 그래서 Topological한 방법을 사용하기로 했다.

자모의 절점이란 자획(字劃)의 시작이나 끝 교점, 및 굴절점이며, 가지란 절점 사이를 연결한 선인 자획의 일부 또는 전부를 말한다. 그림 2는 이를 설명한 것이다.

처음 시도한 방법은 가지의 수와 절점의 수에 의한 분류를 시도해 본 결과 자음 자모에서는 비교적 그 분리가 잘된다 모음의 경우에는 많은 중복이 나타남을 발견했다. 모음은 따라서 그자모가 B의 위치에 있는 것인가 C의 위치에 있는 것인가를 구분하고 다시 흑점의 분포를 구분하

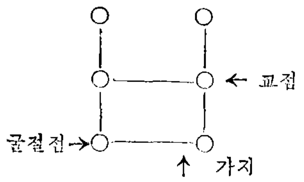


그림 2. 가지와 절점

여 분리가 가능하다. 자음에서도 분리 안되는 것만 이를 따로 적용하여 분리하면 완전한 분리가 가능하다. 표2는 자음을 위의 분리방법으로 그 특징을 추출한 것으로 가지의 수를 열(列)로 하고 절점의 수를 행(行)으로 했을 때 해당되는 자모를 기입했다. 여기서 ㄱ과 ㄴ이 중복되고 ㄷ과 ㅅ이 중복 되는데 이것을 다시 분리하는 방법으로 여러가지 생각할 수 있으나 중심으로 아래 위로 축을 그리고 그 축 좌우 어느 쪽에 더 흑점이 많은가로 구분하면 대개 분리된다 즉 그림4와 같이 ㄷ은 왼편에 흑점이 더 많고 ㅅ은 오른편에 흑점이 더 많다. 그러나 이 방법으로도 ㄹ과 ㅈ의 분리는 불가능하므로 ㅈ으로 자모를 통일해야 한다.

모음의 경우는 표3과 같이 가지수와 자모의

표 3. 자음의 분류

가지 / 절점	1	2	3	4	5	6	7	8
1	○							
2		○						
3		ㄱ ㄴ						
4			ㄷ ㅅ	ㅁ				
5			ㅎ	ㅋ ㅈ				
6				ㅌ	ㄹ ㅈ	ㅂ		
7					ㅊ	(ㅊ)		
8								ㅍ

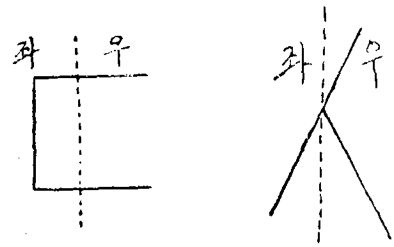


그림 4. 좌우 흑점차

위치에 따라 분류된다. 절점수는 전혀 분리에 도움이 안됨을 발견했다. 여기서 중복된 자모는 B의 위치면 상하 중간축의 좌우 어느쪽에 흑점이 더 많은가 C의 위치면, 좌우 중간축의 상하 어느쪽에 흑점이 더 많은가로 구분이 가능하다. 다만 H는 이때 좌우가 같으므로 3가논리의 현상이 일어나며 Threshold의 규제가 비교적 엄밀해야만 ㅑ, ㅕ와 구분된다.

표 3. 모음의 분류

가지 / 위치	1	2	3	4	5	6	7	8
B	ㅣ		ㅏ ㅓ	ㅔ	ㅑ ㅕ H	ㅖ		ㅒ
C	ㅡ		ㅗ ㅜ		ㅛ ㅠ			

그러나 C의 위치에서는 자음 자모와 혼동될 가능성이 크다.

Ⅳ. 결합 행렬법

이와 같이 한 개의 통일된 방법으로 자모를 전부 일대일로 변환할 수 없는 난점이 있다. 그래서 절점과 절점사이의 연결관계와 그 방향에 의한 특성 추출을 기도한 것이 바로 결합 행렬법이다. 이 결합 행렬은 절점의 번호를 각 절점에 부여하고 그 번호로 행과 열을 이루는 행렬에서 각 요소는 행의 절점과 열의 절점 사이에 연결이 없으면 0, 있으면 그 방향에 따라서 그림 5에서 보이는 바와 같은 방향지수로 표시한다.

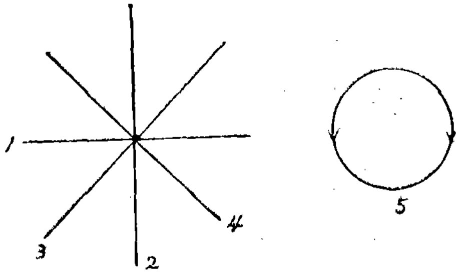


그림 5. 방향지수

이 결합행렬은 아래조건이 만족되는 i행 j열은 전부 0이 되는 삼각행렬이다.

$$i \leqq j$$

절점의 번호를 부여하는 방법이 조직적이어야 한다. 그래서 다음과 같은 규칙에 의해서 절점을 추적하여 번호를 부여한다.

(1) 1번은 최상 최좌(最上 最左)의 절점이다.

(2) 절점은 오른편으로 연결획을 따라 추적함이 최우선이며 다음은 아래로 연결획을 따라 추적하며 그 추적의 순서로 번호를 부여한다.

(3) 절점이 획의 끝일 때에는 다음 절점은 획의 끝의 절점 바로 전 절점과 연결 된것을 전혀 연결안되어 있으면 좌단에서 다시 시작한다.

그림 6은 몇개의 자모에 대해서 위의 규칙을 적용하여 그 번호를 부여한 것이다.

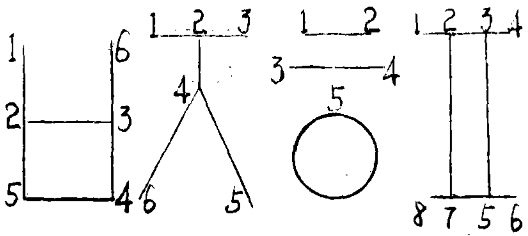


그림 6. 절점 번호 부여의 예

이와 같이 하여 28개의 자모에 대한 결합 행렬은 아래와 같다.

$$ㄱ=\begin{vmatrix} 0 & & \\ & & 0 \\ 1 & 0 & \\ 0 & 2V3 & 0 \end{vmatrix} \qquad ㄴ=\begin{vmatrix} 0 & & \\ & & 0 \\ 2 & 0 & \\ 0 & 1 & 0 \end{vmatrix}$$

$$ㄷ=\begin{vmatrix} 0 & & & \\ & & 0 & \\ 1 & 0 & & \\ 2 & 0 & 0 & \\ 0 & 0 & 1 & 0 \end{vmatrix} \qquad ㅅ=\begin{vmatrix} 0 & & & \\ & & 0 & \\ 3 & 0 & & \\ 0 & 4 & 0 & \\ 3 & 3 & 0 & 0 \end{vmatrix}$$

$$ㅁ=\begin{vmatrix} 0 & & & \\ 1 & 0 & & \\ 0 & 2 & 0 & \\ 2 & 0 & 1 & 0 \end{vmatrix}, \qquad ㅋ=\begin{vmatrix} 0 & & & & \\ 1 & 0 & & & \\ 0 & 2V3 & 0 & & \\ 0 & 0 & 2V3 & 0 & \\ 0 & 0 & 1 & 0 & 0 \end{vmatrix}$$

$$ㅈ=\begin{vmatrix} 0 & & & & \\ 1 & 0 & & & \\ 0 & 3 & 0 & & \\ 0 & 0 & 4 & 0 & \\ 0 & 0 & 0 & 3 & 0 \end{vmatrix}, \qquad ㅇ=\begin{Bmatrix} 0 & 0 \\ 5 & 0 \end{Bmatrix}$$

$$ㄹ=\begin{vmatrix} 0 & & & & & & \\ 1 & 0 & & & & & \\ 0 & 2 & 0 & & & & \\ & 0 & 1 & 0 & & & \\ & & 0 & 2 & 0 & & \\ & & & & & & \\ & & & & 0 & 1 & 0 \end{vmatrix} \qquad ㅂ=\begin{vmatrix} 0 & & & & & \\ 2 & 0 & & & & \\ 0 & 1 & 0 & & & \\ 0 & 0 & 2 & 0 & & \\ 0 & 2 & 0 & 1 & 0 & \\ 0 & 0 & 2 & 0 & 0 & 0 \end{vmatrix}$$

$$ㅈ=\begin{vmatrix} 0 & & & & & \\ 1 & 0 & & & & \\ 0 & 1 & 0 & & & \\ 0 & 3 & 0 & 0 & & \\ 0 & 0 & 4 & 0 & 0 & \\ 0 & 0 & 0 & 3 & 0 & 0 \end{vmatrix}, \qquad ㅌ=\begin{vmatrix} 0 & & & & & \\ 1 & 0 & & & & \\ 0 & 0 & 0 & & & \\ 0 & 0 & 1 & 0 & & \\ 0 & 0 & 2 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 1 & 0 \end{vmatrix}$$

$$ㅎ=\begin{vmatrix} 0 & & & & & \\ 1 & 0 & & & & \\ 0 & 0 & 0 & & & \\ 0 & 0 & 1 & 0 & & \\ 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 5 & 0 \end{vmatrix} \qquad ㅎ=\begin{vmatrix} 0 & & & & & \\ 2 & 0 & & & & \\ 0 & 2 & 0 & & & \\ 0 & 1 & 0 & 0 & & \\ 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 5 & 0 \end{vmatrix}$$

$$ㅊ=\begin{vmatrix} 0 & & & & & & \\ 1 & 0 & & & & & \\ 0 & 0 & 0 & & & & \\ 0 & 0 & 1 & 0 & & & \\ 0 & 0 & 0 & 3 & 0 & & \\ 0 & 0 & 0 & 0 & 4 & 0 & \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 \end{vmatrix} \qquad ㅊ=\begin{vmatrix} 0 & & & & & & \\ 1 & 0 & & & & & \\ 0 & 1 & 0 & & & & \\ 0 & 2 & 0 & 0 & & & \\ 0 & 2V3 & 0 & 0 & 0 & & \\ 0 & 0 & 0 & 0 & 4 & 0 & \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 \end{vmatrix}$$

$$ㅍ=\begin{vmatrix} 0 & & & & & & & \\ 1 & 0 & & & & & & \\ 0 & 1 & 0 & & & & & \\ 0 & 0 & 1 & 0 & & & & \\ 0 & 0 & 2 & 0 & 0 & & & \\ 0 & 0 & 0 & 1 & 0 & 0 & & \\ 0 & 2 & 0 & 0 & 1 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{vmatrix} \qquad ㅏ=\begin{vmatrix} 0 & & & \\ 2 & 0 & & \\ 0 & 1 & 0 & \\ 2 & 2 & 0 & 0 \end{vmatrix}$$

$$ㅓ=\begin{vmatrix} 0 & & & \\ & & & \\ 2 & 0 & & \\ 0 & 2 & 0 & \\ 0 & 1 & 0 & 0 \end{vmatrix}$$

$$ㅗ=\begin{vmatrix} 0 & & & \\ 2 & 0 & 0 & \\ 0 & 1 & 0 & \\ 0 & 1 & 0 & 0 \end{vmatrix}, \qquad ㅜ=\begin{vmatrix} 0 & & & \\ 1 & 0 & & \\ 0 & 1 & 0 & \\ 0 & 2 & 0 & 0 \end{vmatrix}$$

$$\text{ㅑ} = \begin{vmatrix} 0 \\ 2 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{vmatrix}, \qquad \text{ㅕ} = \begin{vmatrix} 0 \\ 2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{vmatrix}$$

$$\overset{①}{\text{ㅛ}} = \begin{vmatrix} 0 \\ 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{vmatrix}, \qquad \overset{①}{\text{ㅠ}} = \begin{vmatrix} 0 \\ 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & {}^{2} & {}^{3} & 0 & 0 & 0 & 0 \end{vmatrix}$$

$$\text{ㅡ} = \begin{vmatrix} 0 & 0 \\ 1 & 0 \end{vmatrix} \qquad \text{ㅣ} = \begin{vmatrix} 0 & 0 \\ 2 & 0 \end{vmatrix}$$

$$\text{ㅐ} = \begin{vmatrix} 0 \\ 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \end{vmatrix} \qquad \text{ㅔ} = \begin{vmatrix} 0 \\ 2 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \end{vmatrix}$$

$$\text{ㅖ} = \begin{vmatrix} 0 \\ 2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \end{vmatrix}, \qquad \text{ㅒ} = \begin{vmatrix} 0 \\ 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

V. 결 론

유일성 (Uniqueness) 을 가지는 자모의 특징 추출은 결합행렬에 의해서 가능함을 보았다. 이것이 기억용량이나 계산의 회수면으로 봐서 최적의 특징 추출방식인가는 앞으로 연구해야 할 문제이다. 또한 글자에서 자모를 분리하는 방법도 더욱 연구해야하며 실제로 글자를 식별하는 기계와 여기에서 사용될 통계학적 수법에도 많은 연구가 계속되어야 하리라고 믿는다.

참 고 문 헌

[1] R.W Weeks; Rotating Raster Character Recognition System, AIEE C&E No.56, P 353-358 Sept. 1961

[2] R. Casey & G. Nagy; Recognition of Printed Chinese Characters, IEEE Trans. on EC Vol. 15, P 91-101, Feb. 1966

[3] S.H. Unger; Pattern detection and Recognition, Proc IRE Vol.47 P 1737-1752, Oct. 1959

[4] L.F. Turner; A System for the Automatic Recognition of Moving Pattern, IEEE Trans. Vol.IT-12, P 195-205, April, 1966

[5] M.Minsky; Steps Tow and Artificial Intelligence, Proc. IRE, Vol.49, P8-29, Jan. 1961

[6] 富田 et al; Recognition of Handwritten Katakana Characters, 電氣通信學會誌 50卷 P 650-663, Apr. 1967

[7] 杉上 et al; Method for the Recognition of Japanese Hiragana Characters, IEEE Trans. IT-14 P 226-233 Mar, 1968

[8] S.Seshu & M.B.Reed; "Linear Graphs Electrical Networks", Addison-Wesley.