

論文95-32B-11-19

# FCM 알고리즘을 이용한 이진 결정 트리의 구성에 관한 연구

## (A Study on the Design of Binary Decision Tree using FCM algorithm)

鄭淳元\*, 朴重祖\*, 金景敏\*, 朴貴泰\*

(Soon-Won Jung, Joong-Jo Park, Kyoung-Min Kim, and Gwi-Tae Park)

### 요 약

본 논문에서는 새로운 이진 결정 트리의 설계 방법에 대하여 제안을 하고 이를 타이어 접지면 패턴 분류에 적용해 보았다. 이 방법에서 이진 결정 트리는 FCM 알고리즘을 이용하여 구성된다. 이진 결정 트리를 구성하는데 있어 초기 군집화에는 모든 특징량이 사용된다. 각 노드에서는 제안되는 유사 지표를 이용하여 하나의 특징량 혹은 특징량 부분 집합이 선택된다. 이렇게 구성된 결정 트리는 미지의 입력 패턴에 대한 분류기로 사용될 수 있다. 제안되는 방법을 타이어 접지면 패턴 분류에 적용시켜 보았으며 특징량 추출을 포함한 설계 과정을 설명하였다. 실험 결과는 본 방법이 유용함을 보여주고 있다.

### Abstract

We propose a design scheme of a binary decision tree and apply it to the tire tread pattern recognition problem. In this scheme, a binary decision tree is constructed by using fuzzy C-means(FCM) algorithm. All the available features are used while clustering. At each node, the best feature or feature subset among these available features is selected based on proposed similarity measure. The decision tree can be used for the classification of unknown patterns. The proposed design scheme is applied to the tire tread pattern recognition problem. The design procedure including feature extraction is described. Experimental results are given to show the usefulness of this scheme.

### I. 서 론

결정 트리(decision tree)는 여러 패턴 인식 분야에서 광범위하게 쓰이는 기법이다.<sup>[1]</sup> 단일 단계 분류기(single stage classifier)에 비해 트리 분류기가 갖는 중요한 장점은 복잡하고 전역적(global)인 결정을 일련의 단순하고 국부적(local)인 결정으로 나누어

주고 빠른 시간 안에 의사 결정을 가능하게 한다는 점이다.

결정 트리 분류기를 설계하기 위해서는 먼저 결정 트리를 구성해야 한다. 결정 트리를 구성하는데 있어 고려해야 할 사항은 크게 두 가지로, 하나는 각 노드에 대한 하부(descendant) 노드의 갯수이고 다른 하나는 각 노드에서의 특징량 선정 문제이다. 본 논문에서는 하부 노드의 갯수가 2인 이진 트리를 구성하였다. 각 노드간의 중복(overlapping)의 문제와 균형(balance)의 문제가 있지만, 이진 트리를 이용하면 기존의 어떠한 단일 단계 분류기도 트리 분류기로 변환할 수

\* 正會員, 高麗大學校 電氣工學科

(Dept. of Electrical Eng. Korea Univ)

接受日字: 1995年2月24日, 수정완료일: 1995年11月7日

있으며, 구성상 구조가 간단하다는 장점을 가지므로 이진 트리를 많이 사용한다.<sup>[12]</sup> 한편 특징량의 선정 문제는 패턴 인식에 있어서 매우 중요한 주제 중의 하나이다. 분류에 꼭 필요한 특징량만을 선정하여 분류기에 사용하면 분류 정밀도(classification accuracy)가 높아지고 특징량 추출 및 분류에 소요되는 계산 시간이 줄어들게 된다. 그러나  $q$ 개의 특징량 중에서 최적의(optimal) 특징량을 추출하는 경우 성능 시험을 위한 특징량의 부분 집합의 갯수는  $(2^q-1)$ 개가 되며 특징량의 갯수가 조금만 커져도 실제 적용하기에는 어려움이 많다. 따라서 최적(optimal)이 아닌, 구할 수 있는 최선의(best) 특징량 선택 기법이 많이 사용되며 분류기가 단일 단계인 경우에 비해 트리 구조인 경우 특징량 선택이 더 용이해진다.<sup>[12,13,14]</sup>

이진 트리 분류기에 대한 많은 연구가 있었으며<sup>[12,3,4,5]</sup>, 특히 80년대 이후 퍼지 개념이 도입된 FCM 군집화 알고리즘을 이용한 이진 트리 분류기에 대한 연구들이 있었다.<sup>[16,7,8,9]</sup> 본 논문에서는 트리 구조의 설계 방법을 FCM 군집화 알고리즘을 이용한 것으로 한정시키도록 하였다. 먼저 [6,7]에서 제시된 방법은 기지의(labelled) 데이터에 적용하기에는 적당치 않다고 볼 수 있으며 특히 [6]에서는 트리의 각 노드에서의 특징량 선정 문제를 고려하지 않고 모든 특징량을 이진 트리 구성에 사용하였다. 또한 임계치(threshold value)의 설정에 따라 최종 노드의 갯수가 입력 패턴 클래스의 갯수와 불일치하는 결과를 보일 수도 있다. [7]에서는 이진 트리의 구성과 특징량 선정이 동시에 이루어졌다. 각 노드에서 이진 군집화를 위하여 처음에 선정된 모든 특징량들이 사용되며 일단 군집화가 된 후 분류 척도(separation index)를 이용하여 단 한개의 특징량이 선정된다. 그러나 여기서 사용한 분류 척도는 단순히 두 군집에 속한 패턴을 이루는 특징량들 간의 유클리디안 거리를 기초로 구해지므로 이러한 방법으로 구한 한개의 특징량이 데이터 전체적인 관점에서 볼 때 전체 특징량을 대표한다고 보기 힘들고 단 하나의 특징량만을 사용함으로써 인해 인식률이 저하되어 실제 적용상의 어려움이 있다. 또한 이 방법은 노이즈(outlier pattern)에 너무 민감하다는 단점이 있다. [8]에서는 위의 단점들을 해결하기 위해 퍼지 분할 행렬간의 유사도를 기반으로 한 특징량 선정 방법을 사용하여 각 노드에서 하나의 특징량을 선정하였다. 이렇게 구해진 특징량은 데이터 전체적

인 관점에서 볼 때 전체 특징량을 대표한다고 보는 것이 타당하며 노이즈에 그다지 민감하지 않다는 장점을 가진다. 그러나 [7]의 경우와 마찬가지로 단 하나의 특징량만을 사용함으로써 인한 문제점을 가지고 있다. 한편 [9]에서는 특징량 선정에 있어 최적의 특징량 부분집합(feature subset)을 구하기 위해 특징량 부분집합의 모든 경우에 대하여 탐색(exhaustive search)하는 방법을 사용하였다. 그러나 이 방법은 앞서 언급한 대로 특징량의 갯수가 커질 경우 실용적인 방법이 되지 못한다.

위의 단점들을 해결하기 위하여 먼저 [7]의 경우와 같이 퍼지 분할 행렬간의 유사도를 기반으로 하여 각 노드에서 하나 혹은 적당한 몇 개의 특징량을 선정하는 방법을 제안한다. 이렇게 함으로써 각 노드에서 전체 특징량을 사용하지 않고 최소의 특징량으로 최대의 효과를 내는 것이 가능해진다.

본 논문은 크게 두 부분으로 나누어져 있다. 하나는 위에서 간단히 언급한 방법으로 이진 결정 트리의 설계를 하고 특징량을 선정하는 부분이고, 다른 하나는 제안되는 방법을 타이어 접지면 패턴 인식에 적용하여 그 유용성을 밝히는 부분이다.

## II. FCM 알고리즘 및 이진 결정 트리의 설계

### 1. FCM 군집화 알고리즘

FCM 알고리즘은 주어진 데이터 집합,  $X = (\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n)$  ( $\vec{x}_k \in R^q$ ,  $q$  is the number of features)에 대하여 특정 내적 노름자(inner product norm metric)를 거리 척도로 사용하여, 정의된 어떤 목적 함수가 근사적 최소값에 도달되도록 퍼지 분할(fuzzy partition)  $U$ 와 군집의 중심값  $V = (\vec{v}_1, \vec{v}_2, \dots, \vec{v}_c)$ 를 반복계산법에 의해 구하는 최적화 퍼지 군집화 알고리즘이다.<sup>[10]</sup> 일반적으로 FCM에서 사용되는 목적함수는 다음과 같다.

$$J_m(U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m (d_{ik})^2 \quad (1)$$

$$d_{ik}^2 = \|\vec{x}_k - \vec{v}_i\|^2$$

여기서  $m$ 은 지수 가중치이며  $c$ 는 군집의 갯수이다.  $k$ -means 알고리즘과 같은 크리스프(crisp) 클러스터링 알고리즘이 각 데이터를 엄격하게 하나의 클러스터에 할당되도록 분류하는 것에 비하여, FCM 알고리즘

은 데이터가 각 클러스터에 얼마만큼 속하는 가를 수치적으로 '0'에서 '1'사이 - [0, 1] -의 값으로 표현하여 실제 데이터의 구조를 나타내는데 더 많은 유연성을 갖게된다.

2. 이진 결정 트리 설계 및 특징량 선정

이진 결정 트리를 설계하는 첫 단계는 주어진 기지 (labelled)의 데이터들을 FCM 군집화 알고리즘을 사용하여 2개의 군집으로 나누는 것이다. 이 단계에서는 처음에 주어진 특징량들을 모두 사용하게 된다. 이 결과로부터 퍼지 분할 행렬을 구할 수 있으며, 이 행렬은 분할된 데이터의 전체적인 구조를 나타낸다고 볼 수 있다.<sup>13)</sup> 분할된 각 노드에서 특징량의 우선 순위 (priority)를 구하기 위하여 각각의 특징량을 이용하여 데이터들을 2개의 군집으로 나누는 작업을 하게 된다. 이로부터 각 특징량에 대한 퍼지 분할 행렬과 유사 지표(similarity index)를 얻을 수 있다.

m개의 데이터에 대하여 전체 특징량 및 k번째 특징량을 이용하여 구한 퍼지 분할 행렬을 각각  $U_i, U_i^k$ 다 하자. 또한 이들 간의 유사 지표(similarity index),  $S_k$ 를 다음과 같이 정의 하자.

$$S_k = \frac{m - d_k}{m} \quad (2)$$

$$\text{where, } d_k = \sum_i \sum_j |u_{ij} - u_{ij}^k|$$

위의 정의에 의해  $S_k$ 가 큰 순서로 특징량을 나열할 수 있으며  $S_k$ 가 큰 특징량일수록 전체 특징량의 종합적인 특성을 많이 가지고 있다고 할 수 있다. 물론 이 방법에 의해 각 노드에서 하나의 특징량만을 선정한다면 가장 좋은(best) 특징량  $f_i$ 은 다음과 같이 선택될 수 있다.

$$l = \text{Arg} \{ \max_{1 \leq k \leq q} \{ S_k \} \} \quad (3)$$

여기서  $\text{Arg} \{ \cdot \}$ 는  $\{ \cdot \}$ 의 조건을 만족하는 변수(argument) k를 의미한다. 각 특징량에 대한 유사 지표가 구해진 후 이를 기반으로 하여 각 노드에서 필요한 특징량을 구할 수 있다. 단 한개의 특징량으로 전체의 특징량을 대신할 수 있다면 좋겠지만 분류상의 예러가 크다면 예러를 줄일 수 있도록 몇 개의 특징량을 선정할 수 있다. 본 연구에서는 먼저 각 노드에서 선택할 수 있는 특징량의 최대 개수를 정해 놓고 유사

지표가 큰 순서대로 특징량 부분 집합(subset)에 추가하여 특징량의 개수에 따른 분류 성능을 평가한다. 특징량 부분 집합 선정 과정은 다음과 같다.

위에서 특징량 공간이 q-차원이라 하였으므로 전체 특징량 집합을 다음과 같이 정의하자.

$$F = \{f_1, f_2, \dots, f_q\} \quad (4)$$

각 노드에서의 최대 사용할 수 있는 특징량의 개수를 r이라 하고,  $S_k$ 가 큰 순서대로 j개의 특징량을 가지는 특징량 부분 집합을  $F_j$ 라 하자.

$$F_i = \{f_{i1}, f_{i2}, \dots, f_{ir}\} \quad (5)$$

단,  $S_{jk} > S_{j'l}$  (where,  $k < l$ )

엄밀히 말해 특징량의 개수가 많고 진다고 분류 성능이 좋아진다는 보장은 없으며<sup>14)</sup> 식(3)과 같이 선정한 단 하나의 특징량이 최고의 분류 성능을 낼 수도 있다. 또한 특징량을 추가함에 따라 어느 시점에서 분류 성능이 갑자기 좋아지는 경우도 있게된다. 따라서  $F_i$ 를 이용하여 군집화를 하였을 때 발생하는 분류 에러를  $err(F_i)$ 라 하면 각 노드에서의 특징량 부분집합  $F_i$ 는 다음과 같이 선정한다.

$$t = \text{Arg} [ \min_{1 \leq i \leq r} err(F_i) ] \quad (6)$$

분류 에러,  $err(F_i)$ 가 같은 것이 여러개 존재할 때는 최소의 특징량을 가지는  $F_i$ 를 선정한다. 즉, 어느 노드에서 특징량 한개만으로도 전체 특징량을 사용한 것 같은 성능을 낸다면 그 노드에서는 2개 이상의 특징량에 대하여 성능 평가를 할 필요없이 단 한개의 특징량만으로 특징량 부분 집합 선정 과정은 끝나게 된다.

이와 같은 특징량 선정 과정은 전체 트리가 완성될 때까지 반복된다. 즉 각 최종(terminal) 노드에 모든 패턴 클래스가 독립적으로 나타나게 되면 이와 같은 과정은 끝나게 된다.

Ⅲ. 이진 결정 트리를 이용한 타이어 접지면 패턴의 분류

1. 타이어 접지면 패턴의 특징량 추출

화상 처리의 관점에서 볼 때, 타이어 접지면 패턴은 부패턴(sub-pattern)이 정렬되어 있으므로 결정적 텍스처(texture)의 한 종류로 볼 수 있다. 패턴의 특징

량을 추출하기 위하여 계조치 동시 발생 행렬(gray-level cooccurrence matrix)라 불리는 2차 통계량을 사용하였다. 이 행렬은 화상내에 존재하는 각 화소들간의 공간적인 관계(spatial relation)를 나타내며 이 행렬로부터 영상의 통계적인 특징들을 추출할 수 있다. [11,12] 특히 원 영상이 이치화되었을 경우 이 행렬을 구하는 과정은 간단하며, 특징량은 쉽게 얻어질 수 있다.

1) 화상의 전처리

타이어 접지 패턴 분류를 위한 데이터 취득 부분은 타이어 접지면의 영상 취득 과정 및 특징량 추출을 위한 전처리 과정으로 구성된다. 전처리 과정은 서브 샘플링과 외곽선 검출로 이루어지며, 전자는 영상 취득후 처리해야 할 데이터량을 줄이기 위하여 3×3 마스크로 마스킹을 취한후 4화소당 한 화소를 취함으로써 행해진다. 한편, 화상의 이치화 과정은 윤곽선 검출(edge detection)을 동시에 행할 수 있는 알고리즘을 사용하였으며 그 알고리즘은 식(7)와 같다.

$$g(i, j) = \begin{cases} 1, & \text{if } (g(i-1, j) + g(i+1, j) - 2g(i, j)) > T_1 \text{ or} \\ & (g(i, j-1) + g(i, j+1) - 2g(i, j)) > T_2 \text{ or} \\ & (g(i-1, j-1) + g(i+1, j+1) - 2g(i, j)) > T_3 \text{ or} \\ & (g(i+1, j-1) + g(i-1, j+1) - 2g(i, j)) > T_1; \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

여기서,  $g(i, j)$  는 처리 대상 화소이며,  $T_1, T_2, T_3$ ,  $T_1$ 는 각각 가로, 세로, 대각선 방향에 민감한 임계치(threshold value)를 나타낸다. 본 실험에서 이 임계치는 모두 15로 같은 값을 선정하였다.

2) 특징량 추출을 위한 계조치 동시 발생 행렬

윤곽선의 방향에 대한 정보는 이치화 영상에 대한 계조치 동시 발생 행렬을 통해 표현될 수 있으며 이러한 행렬은 2×2 행렬이 된다. 특히 행렬 요소중 (2, 2) 요소는 윤곽선의 양에 대한 정보를 가지고 있으므로 이를 특징량으로 사용할 수 있다. 즉 관심있는 여러 각도 성분에 대해 계조치 동시 발생 행렬을 구하고 이로부터 (2, 2) 요소를 추출하여 특징량으로 사용하였다. 예로서 그림 1(a)에 대하여 0°, 45°, 90°, 135°의 각도 성분에 대응하는 계조치 동시 발생 행렬을 구하면 그림 1(b)(c)(d)(e)가 된다. 이로부터 { 8, 2, 8, 2 }라는 특징량 벡터를 추출할 수 있다.

본 실험에서 타이어 접지면 패턴의 특징을 추출하기 위하여 사용된 각도 성분은 0°, 14°, 27°, 45°, 90°, 104°, 117°, 166°의 8종류이다. 그림 2에 사용된 각도 성분을

나타내었다.

0	0	0	0	0	0	0
1	1	1	0	1	1	1
1	0	0	0	1	0	0
1	0	0	0	1	0	0
1	0	0	0	1	0	0
1	1	1	0	1	1	1
0	0	0	0	0	0	0

(a)

	0	1
0	21	5
1	8	8

(b) For 0°

	0	1
0	9	11
1	14	2

(c) For 45°

	0	1
0	14	10
1	10	8

(d) For 90°

	0	1
0	9	14
1	11	2

(e) For 135°

그림 1. (a) 샘플 이치영상, (b)(c)(d)(e) 영상 (a)에 대한 계조치 동시발생 행렬

Fig. 1. (a) A sample binary image, (b)(c)(d)(e) Cooccurrence matrices for (a).

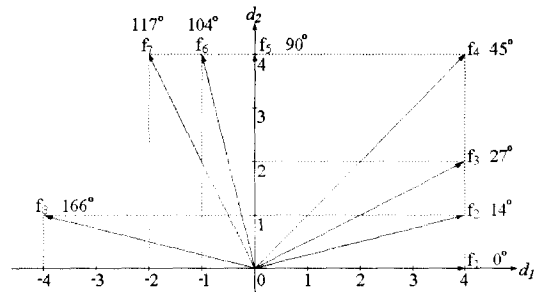


그림 2. 특징량에 대응하는 윤곽선의 각도  
Fig. 2. Directions of edge for the features.

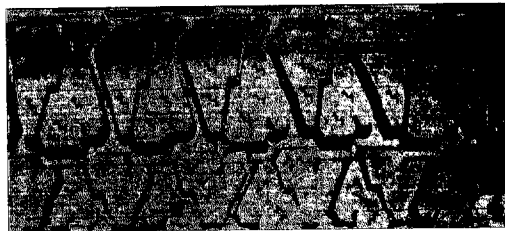
2. 실험 조건

실험에 사용된 타이어 접지면 패턴의 부류는 모두 6 종류이며 각 종류당 10개의 샘플을 가지고 있다. 각 패턴 클래스를  $\{P_1, P_2, \dots, P_6\}$ 로 나타내고  $p_k$ 는  $P_k$ 에 속하는 패턴이라 하자. 그림 3(a)~(f), 그림 4(a)~

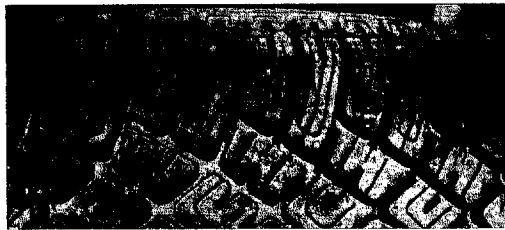
(f)는 각각 샘플 패턴의 원 영상과 전처리 후의 영상을 나타낸다. 표 1에 각 패턴 클래스에 대한 샘플 패턴의 특징량을 나타내었다. FCM 알고리즘에서 지수 가중치  $m$ 은 2로 선정하였으며, 각 노드에서 사용될 수 있는 특징량의 최대 갯수는 전체 특징량 갯수의 절반인 4로 하였다.



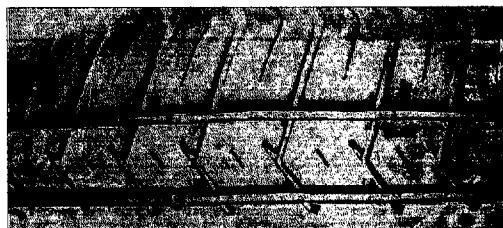
(a) 타이어 접지면 패턴  $p_1$   
(a) Tire tread pattern  $p_1$



(b) 타이어 접지면 패턴  $p_2$   
(b) Tire tread pattern  $p_2$



(c) 타이어 접지면 패턴  $p_3$   
(c) Tire tread pattern  $p_3$



(d) 타이어 접지면 패턴  $p_4$   
(d) Tire tread pattern  $p_4$

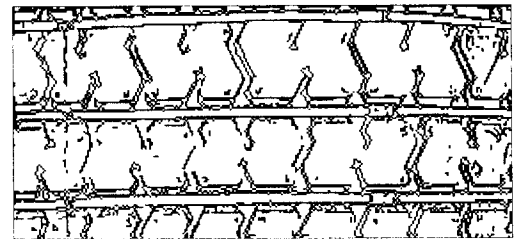


(e) 타이어 접지면 패턴  $p_5$   
(e) Tire tread pattern  $p_5$



(f) 타이어 접지면 패턴  $p_6$   
(f) Tire tread pattern  $p_6$

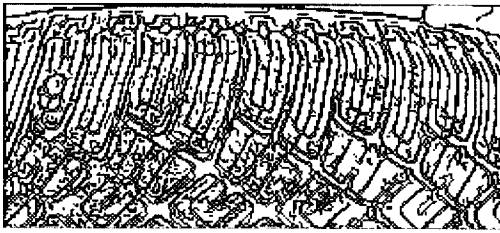
그림 3. 타이어 접지면 패턴의 원영상  
Fig. 3. Original images of tire tread patterns.



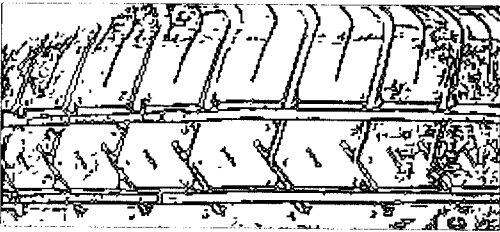
(a) 전처리 후의 패턴  $p_1$   
(a)  $p_1$  pattern after preprocessing



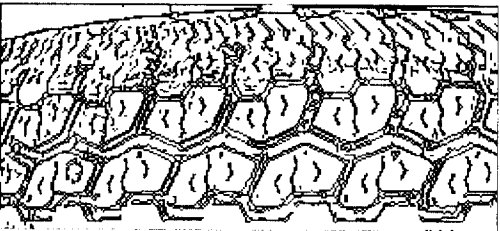
(b) 전처리 후의 패턴  $p_2$   
(b)  $p_2$  pattern after preprocessing



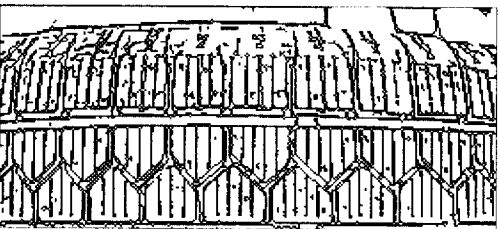
(c) 전처리 후의 패턴  $p_3$   
(c)  $p_3$  pattern after preprocessing



(d) 전처리 후의 패턴  $p_4$   
(d)  $p_4$  pattern after preprocessing



(e) 전처리 후의 패턴  $p_5$   
(e)  $p_5$  pattern after preprocessing



(f) 전처리 후의 패턴  $p_6$   
(f)  $p_6$  pattern after preprocessing

그림 4. 타이어 접지면 패턴의 전처리된 영상  
Fig. 4. Preprocessed images of the tire tread patterns.

표 1. 샘플 타이어 접지면 패턴의 특징량 벡터

Table 1. Feature vectors for the sample tire tread patterns.

Features	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$
Patterns								
$p_1$	1714	931	1750	1743	111	926	2410	2327
$p_2$	822	1163	2195	2134	345	1551	2342	2433
$p_3$	525	1801	4271	3885	983	2237	3927	4616
$p_4$	1427	836	1599	1697	205	882	2055	2137
$p_5$	803	1285	2378	2252	288	1282	2344	2262
$p_6$	1216	895	1907	2046	2680	1218	2267	2415

3. 이진 결정 트리의 구성 및 유사도 측정

먼저 II장 2절에서 기술한 방법으로 타이어 접지면 패턴 데이터에 대한 이진 결정 트리를 구성하였다. 표 2는 각 노드에서의 특징량 각각에 대한 유사도 지표를 나타낸다. 루트 노드에서는  $f_8$ 에 대한 유사도가 가장 크고  $f_8$ 을 이용한 군집화 결과 분류 에러가 최소가 되었으므로 식(3)에 의해 ( $f_8$ )을 루트 노드에서의 특징량 부분 집합으로 선택하게 된다. 루트 노드와 그 하부 노드인  $C_{11}, C_{12}, C_{13}$ 은 모두 특징량 하나만으로 분류가 잘 이루어졌으나  $C_{12}$ 에서는 특징량의 최대 선택 갯수인 네개를 모두 선택해야 됨을 볼 수 있다. 식(3)에 의한 하나의 특징량  $f_2$ 에 의해서는 3개의 분류 에러가 발생되므로 특징량의 갯수를 '1'에서부터 늘려가면서 분류 에러를 검사하게 된다. 한 노드에서 최대 사용할 수 있는 특징량의 갯수를 '4'로 정했으므로  $F_1$ 에서  $F_4$ 까지 분류 에러를 검사하면 각각 '3'개, '5'개, '5'개, '1'개가 되며 식(6)에 의해  $F_3(= \{f_2, f_3, f_4, f_6\})$ 가 분류 에러 '1'개로 특징량 부집합으로 선정되게 된다. II장 2절에서 기술한대로 특징량을 추가함에 따라 오히려 분류 에러가 증가하다가 특징량  $f_6$ 에 의해 갑자기 감소하는 현상을 볼 수 있다. 직관적으로도 알 수 있듯이 전자의 경우는 최대 유사 지표 값이 그 최대값인 1에 가까워 한 개의 특징량으로도 전체 특징량을 대표한다고 볼 수 있으나 후자의 경우는 특징량 대부분이 전자와 달리 작은 유사 지표값을 가짐을 볼 수 있다. 즉 전체 특징량을 사용한 것과 유사한 분류 성능을 내기 위해서는 한개 이상의 특징량이 필요함을 알 수 있다.

이 같은 과정은 전체 트리가 완성될 때까지 반복된다. 즉 각 최종(terminal) 노드에 모든 패턴 클래스

$(P_1, P_2, \dots, P_i)$ 가 독립적으로 나타나게 되면 이와 같은 과정은 끝나게 된다. 그림 5는 구성된 이진 결정 트리 와 각 노드에서 선택된 특징량 부분 집합을 나타낸다.

표 2. 각 노드에서의 유사 지표와 선택된 특징량 부분 집합

Table 2. Similarity index and selected feature subset at each node.

node	$S_i$ for each feature								Selected feature subset
	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$	
Root node	0.25	0.78	0.86	0.86	0.46	0.81	0.83	0.87	$\{f_1\}$
$C_{11}$	0.11	0.28	0.44	0.29	0.84	0.21	0.12	0.65	$\{f_1\}$
$C_{21}$	0.78	0.73	0.73	0.72	0.71	0.78	0.34	0.31	$\{f_1\}$
$C_{31}$	0.93	0.75	0.64	0.49	0.68	0.64	0.98	0.61	$\{f_1\}$
$C_{42}$	0.40	0.67	0.66	0.58	0.16	0.50	0.45	0.38	$\{f_2, f_3, f_4, f_6\}$

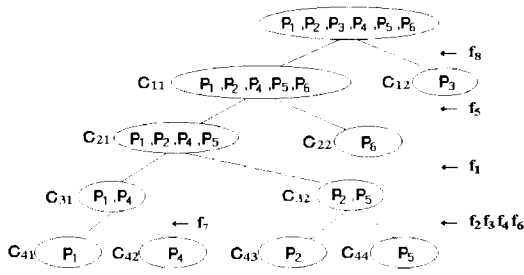


그림 5. 구성된 이진 트리와 선택된 특징량들  
Fig. 5. Constructed binary tree and selected features.

4. 군집화 결과

표 3, 4에 군집화 결과를 보였다. 표 3은 [8]에서 사용한 방법으로 실험한 결과로서 각 노드에서 최대의 유사도를 가지는 특징량 1개만으로 군집화한 경우이고, 표 4는 본 논문에서 제안한 방법으로 군집화한 경우이다. 예상할 수 있듯이  $C_{11}$ 에서의 분류시 특징량 대부분이 작은 유사 지표값을 가짐으로서 한 개의 특징량만으로는 분류 결과가 만족스럽지 못함을 볼 수 있다. 위에서 제안한 방법을 통해 특징량 4개로 분류한 결과 표 4와 같이 오분류가 많이 줄었으며  $C_{11}$ 를 제외한 모든 노드에서 특징량 한개만으로 만족스러운 결과를 얻을 수 있었다. 참고로 [7]에서의 방법으로 트리를 구성한 경우 그림 5와 유사한 트리 구조를 얻을 수 있으나, [8]에서처럼 단 한개의 특징량을 사용함으로써 분류 에러가 '3'이 될뿐 아니라 이렇게 구한 트리

구조 분류기의 성능은 노이즈 패턴에 매우 민감하여 분류 성능의 저하를 초래하게 된다. 한편 두 군집간의 편극 정도(polarization degree)를 기초로 하여 특징량을 선출하고 트리를 구성한 [9]의 방법으로 트리를 구성할 경우 분류 에러는 본 논문에서와 같이 '1'개가 얻어지고 전체 특징량 개수도 감소했으나, 서론에서 기술한 대로 이 방법은 전체적인 탐색(exhaustive search)를 함으로서 전체 특징량의 개수가 커질 경우 실제 적용하기 어렵다는 단점이 있다. 또한 트리를 구성한 후에도 이의 재구성(reconstruction)과정이 필요하고, 패턴 인식시에는 퍼지 패턴 매칭(fuzzy pattern matching)과정을 거쳐야 하므로 인해 전체 알고리즘이 복잡해지는 문제가 있다.

표 3. 각 노드에서 가장 좋은 단 한개의 특징량만으로 군집화시킨 결과

Table 3. Result of binary clustering using only best one feature at each node.

True Class	Cluster					
	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$	$P_6$
$P_1$	10					
$P_2$		7				
$P_3$			10			
$P_4$				10		
$P_5$					10	
$P_6$						10

표 4. 각 노드에서 특징량 부분 집합으로 군집화시킨 결과

Table 4. Result of binary clustering using feature subset at each node.

True Class	Cluster					
	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$	$P_6$
$P_1$	10					
$P_2$		9				
$P_3$			10			
$P_4$				10		
$P_5$					10	
$P_6$						10

IV. 결 론

본 논문에서는 FCM 알고리즘을 이용하여 이진 결정 트리를 구성하고 각 노드에서 필요한 특징량 부분

정 트리를 구성하고 각 노드에서 필요한 특징량 부분 집합을 추출하는 방법에 대해 살펴 보았으며 이를 타이어 접지면 패턴 분류에 적용해 보았다. 이진 결정 트리를 구성하는데 FCM 알고리즘을 이용함으로써 각 노드에서 전체 특징량 및 개별 특징량을 이용한 경우의 퍼지 분할 행렬을 구할 수 있었다. 제안되는 특징량 선정 방법의 가장 큰 장점은 이렇게 구한 퍼지 분할 행렬간의 유사도를 이용함으로써 패턴 전체의 구조적인 특징을 고려하여 특징량 선정을 행할 수 있고 또한 그 과정이 간단히 이루어진다는 것이다. 앞으로의 연구 과제는 오분류 뿐만 아니라 군집 유효도(cluster validity) 개념을 포함하여 특징량의 선정 결과를 개선시키는 일과, 이 실험을 토대로 하여 문자 인식과 같이 부류 수가 많은 패턴에 제안되는 방법을 적용하여 보는 것이라 하겠다.

#### 참 고 문 헌

- [1] S.Rasoul Safavian and David Landgrebe, "A Survey of Decision Tree Classifier Methodology", IEEE Trans. Sys., Man, &Cybern., vol. 21, pp. 660-674, 1991.
- [2] Jack K.Mui and King-Sun Fu, "Automated Classification of Nucleated Blood Cells Using a Binary Tree Classifier", IEEE Trans. PAMI, vol. 5, pp. 429-443, 1980.
- [3] Y.K.Lin and K.S.Fu, "Automatic Classification of Cervical cells using a Binary Tree Classifier", Pattern Recognition vol. 16, pp. 69-80, 1983.
- [4] Harold J.Payne and William S.Meisel, "An Algorithm for Constructing Optimal Binary Decision Trees", IEEE Trans. Computers, vol. 26, pp 905-916, 1977.
- [5] G.H.Landeweerd, T.Timmers, E.S. Gelsema, "Binary Tree Versus Single Level Tree Classification of White Blood Cells", Pattern Recognition, vol. 16, pp. 571-577, 1983.
- [6] D.Dimitrescu, "Hierarchical Pattern Classification", Fuzzy Sets and Systems vol. 28, pp. 145-162, 1988.
- [7] B.B.Devi, "Binary tree design using fuzzy isodata", Pattern Recognition Letter, vol.4 pp. 13-18, 1986.
- [8] 정 순원, 박귀태, "FCM 알고리즘을 이용한 이진 결정 트리의 구성 및 타이어 접지면 패턴 인식에의 적용", 1994년도 한국 퍼지 시스템 학회 추계학술대회, pp. 146-151, 1994
- [9] S.W.Jung, S.W.Bae, G.T.Park, "A Design Scheme for a hierarchical fuzzy pattern matching classifier and its application to the tire tread pattern recognition", Fuzzy Sets and systems vol. 65, pp 311-322, 1994.
- [10] J.C.Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithm", New York, Plenum Press, 1981.
- [11] R.M.Haralick, K.Shanmugam and L. Dinstein, "Textural Features for Image Classification", IEEE Trans. Sys., Man, &Cybern. SMC-3 (6) pp. 610-621, 1973.
- [12] 정 순원, "이치화 영상에 대한 계조치 동시 발생 행렬을 이용한 타이어 접지 패턴의 분류", 석사 학위 논문, 고려 대학교 전기공학과 1992
- [13] R.O.Duda and P.E.Hart, *Pattern Classification and Scene Analysis*, New York: Wiley-Interscience, 1973.
- [14] K.Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic press, 1990.



## 저 자 소 개



鄭淳元(正會員)

1966년 8월 29일생. 1990년 2월  
고려대학교 전기공학과 공학사.  
1992년 8월 고려대학교 대학원 전  
기공학과 공학석사. 1992년 9월~  
고려대학교 대학원 전기공학과 박  
사과정 재학중 주관심분야는 영

상 처리, 패턴 인식, 퍼지 이론 등임.

朴重租(正會員) 제31권 B편 제6호 참조

金景敏(正會員) 제31권 B편 제6호 참조

현재 고려대학교 대학원 전기공학과  
박사과정 재학중

朴貴泰(正會員) 제31권 B편 제6호 참조.

현재 고려대학교 전기공학과 교수