

論文96-33B-1-10

## 유전 알고리즘을 이용한 퍼지 패턴 매칭 분류기의 설계와 응용

## (A design of fuzzy pattern matching classifier using genetic algorithms and its applications)

鄭淳元\*, 朴貴泰\*\*

(Soon-Won Jung and Gwi-Tae Park)

## 요약

본 논문에서는 새로운 퍼지 패턴 매칭 분류기(fuzzy pattern matching classifier) 설계 방법을 제안하였다. 기존의 퍼지 패턴 매칭 분류기를 설계함에 있어 분류기의 성능에 결정적인 영향을 미치는 소속 함수(membership functions)의 모양과 개수에 대한 정확한 정보를 알 수 없었다. 따라서 소속 함수를 구하기 위하여 시행 착오(trial-error)법 혹은 경험에 의존하는(heuristic) 방법이 사용되어 왔다. 그러나 이러한 방법은 다양한 종류의 패턴에 대하여 적용하기에는 한계가 있다. 본 논문에서는 유전 알고리즘을 사용하여 분류 에러를 최소로 줄이는 소속 함수의 적절한 모양과 개수를 찾기 위한 새로운 방법을 제안한다. 유전 알고리즘(genetic algorithms)은 진화라는 생물학적 모델을 기초로한 통계적 알고리즘의 한 부류이다. 이는 여러 함수 최적화 문제에 적용되어 최적 혹은 최적 근처의 해를 찾아 준다. 본 논문에서 유전 알고리즘은 분류 에러에 반비례하는 적합도 함수(fitness function)를 기본으로 소속 함수의 모양과 개수를 결정하는데 쓰인다. 유전 알고리즘에 있어서의 스트링은 소속 함수를 결정하게 되며 인식 결과는 다음 세대의 재생(reproduction) 연산에 영향을 미치게 된다. 제안되는 방법을 타이어 접지면 패턴과 필기체 영문자 인식에 적용하여 보았다. 실험 결과는 본 방법이 유용함을 보여준다.

## Abstract

A new design scheme for the fuzzy pattern matching classifier (FPMC) is proposed. In conventional design of FPMC, there are no exact information about the membership function of which shape and number critically affect the performance of classifier. So far, a trial and error or heuristic method is used to find membership functions for the input patterns. But each of them have limits in its application to the various types of pattern recognition problem. In this paper, a new method to find the appropriate shape and number of membership functions for the input patterns which minimize classification error is proposed using genetic algorithms(GAs). Genetic algorithms belong to a class of stochastic algorithms based on biological models of evolution. They have been applied to many function optimization problems and shown to find optimal or near optimal solutions. In this paper, GAs are used to find the appropriate shape and number of membership functions based on fitness function which is inversely proportional to classification error. The strings in GAs determine the membership functions and recognition results using these membership functions affect reproduction of next generation in GAs. The proposed design scheme is applied to the several patterns such as tire tread patterns and handwritten alphabetic characters. Experimental results show the usefulness of the proposed scheme.

\* 正會員, 高麗大學校 電氣工學科, 서울대 ERC-ACI

研究委員

(Dept. of Electrical Eng. Korea Univ.)

\*\* 正會員, 高麗大學校 電氣工學科

(Dept. of Elec. Eng., Korea Univ.)

接受日字: 1995年8月10日, 수정완료일: 1995年12月20日

## I. 서론

언어의 애매함이나 개념 정의가 불분명한(ill-defined) 현상을 해석적으로 다루기 위한 퍼지 논리는 1965년에 Zadeh 교수에 의해 처음으로 제창되었다. 특히 패턴 인식 분야에 있어서 퍼지 개념은 패턴의 모양이 애매 모호한 경우나, 패턴이 어느 특정 군집에 속하는 정도가 애매 모호한 경우에 이를 효과적으로 모델링할 수 있는 특성이 있다. 퍼지 개념을 패턴 인식 분야에 적용한 대표적 예가 퍼지 패턴 매칭 기법과 FCM(fuzzy c-means) 비교사(unsupervised) 군집화 알고리즘이라 할 수 있다. 본 연구에서는 이들 중 퍼지 패턴 매칭 기법에 대해 다루고자 한다. 기존의 확률 개념을 도입한 방법은 개념 자체가 통계치에 바탕을 두고 있으므로, 인식 절차의 복잡성, 베이시안(Bayesian) 통계에서의 계산의 복잡성을 수반하며 특히 인식에 결정적인 영향을 미치는 특징량(feature)의 실제적인 통계 특성(pdf)을 구하기 힘들다<sup>[1]</sup>. 이러한 문제를 해결하기 위하여 표준 패턴과 입력 패턴의 유사도를 비교하여 패턴 인식을 행하는 퍼지 패턴 매칭(FPM) 기법에 대한 연구가 있어 왔다<sup>[2][3][4]</sup>. 이러한 기법을 이용한 퍼지 패턴 매칭 분류기에 사용되는 소속 함수의 형태 및 개수는 이의 성능을 좌우하는 가장 중요한 요소이지만 현재까지 대부분의 퍼지 패턴 매칭 분류기의 설계에서는 소속 함수를 구하기 위하여 시행 착오(trial-error)법 혹은 경험에 의존하는(heuristic) 방법이 사용되어 왔다. 즉 어떤 종류의 패턴을 분류하기 위하여 퍼지 패턴 매칭 분류기를 설계함에 있어 좋은 성능을 낼 수 있는 소속 함수를 구하기 힘들다는 것이다.

본 논문에서는 이러한 문제점을 해결하기 위하여 최적화 기법의 하나로써 최근 많은 관심을 모으고 있는 유전 알고리즘을 이용하여 패턴 분류에 적합한 소속 함수의 모양, 개수 등을 결정하고자 한다. 또한 제안되는 방법을 타이어 접지면 패턴, 필기체 영문자 인식 등 실제 문제에 적용하여 그 유용성을 보였다.

## II. 유전 알고리즘

유전 알고리즘은 1970년대 미국의 John Holland 교수에 의해 정립된 이론으로 자연의 유전학(natural genetics)과 자연 선택(natural selection)의 원리에

근거한 최적해 탐색 방법이다<sup>[5]</sup>. 기존의 최적해 탐색이 국부 탐색(local search)을 하는데 반해 유전 알고리즘은 여러 해를 동시에 탐색하는 전역 탐색(global search)을 함으로서 전역적인 최적해(global optimal solution)를 찾을 확률이 기존의 최적화 탐색에 비해 큰 것이 특징이다. 특히 파라메타 영역의 연속성과 미분치의 존재와 같은 제약 조건이 없으며 다른 탐색 방법에서 나타나는 보조적인 정보가 필요 없다는 장점을 가진다. 이러한 유전 알고리즘에 쓰이는 연산자는 난수 발생, 문자열 복제, 부분 문자열 교환 등 단순한 형태이지만 그 단순성에도 불구하고 전역적인 최적해를 찾는 데 있어 뛰어난 탐색 기능을 발휘하는데 이러한 유전 알고리즘의 수렴 해석에 관한 연구는 스키마(Schema) 이론에 기초하여 이루어진다. 유전 알고리즘의 성능은 실제 파라미터의 부호화 기법(coding technique), 복제(reproduction), 교배(crossover), 돌연변이(mutation) 등의 유전 연산자(genetic operator)와 평가 함수(evaluation function)의 설정 등에 크게 의존한다<sup>[6]</sup>.

일반적인 이진 부호화 기법(binary coding technique)에 의해 생물과 같은 재생산, 교배, 돌연변이를 거쳐 다음 세대의 자손(offspring)을 만들어 내는 과정은 다음과 같다.

- i) 부호화 및 초기화(coding and initialization)  
생물의 유전 정보를 담고 있는 염색체(chromosome)와 같이 유전 알고리즘에서는 염색체에 해당하는 파라미터를 부호화 한다. 일반적인 부호화 방법은 파라미터를 유한 길이의 이진 스트링(binary string)으로 부호화하며 랜덤하게 N개의 문자열을 생성하여 초기 해집단(initial population)을 구한다.
- ii) 적합도 평가(fitness evaluation)  
각 문자열을 디코딩(decoding)하여 목적 함수에 대한 적합도(fitness)를 계산한다.
- iii) 복제  
자연 선택의 개념을 기반으로 높은 적합도를 가진 문자열에 대하여 다음 세대로 복제될 확률을 높게 한다.
- iv) 교배  
두 문자열을 임의로 선정하여 문자열 안에 있는 유전자 정보를 서로 교환하여 새로운 정보를 갖는 문자열을 만든다.

v) 돌연변이

문자열 안에 있는 유전자의 일부를 임의로 바꾸어 새로운 정보를 갖는 문자열을 만든다.

위와 같은 과정을 반복하여 유전 알고리즘은 최적의 해를 탐색해 나간다. 유전 알고리즘에 관한 자세한 내용은 참고문헌 [5]에 잘 기술되어 있다.

III. 퍼지 패턴 매칭 분류기(FPMC)

퍼지 패턴 매칭의 기본 개념은 두개의 퍼지 집합간의 유사도(similarity)를 비교하는 것이다. 퍼지 패턴 매칭을 수행하기 위하여 표준 패턴(prototype pattern)과 미지의 입력 패턴(input pattern)을 퍼지 소속 함수(fuzzy membership functions)에 대하여 퍼지화 한다.  $X$ 를 퍼지 입력 패턴이라 하고  $A_i$ 를  $i$ 번째 클래스  $\omega_i$ 의 표준 퍼지 패턴이라 하자. 이러한 두 퍼지 패턴  $X$ 와  $A_i$ 의 유사도를 비교하여 퍼지 패턴 매칭을 수행하게 된다. 특징량의 개수를  $q$ 라고 하면 패턴 매칭의 결과는 다음과 같이 벡터로서 나타내어진다<sup>[4]</sup>.

$$M(X|A) = [M(X_1|A_1), M(X_2|A_2), \dots, M(X_q|A_q)]^T \quad (1)$$

여기서  $M(X_j|A_j)$ 는 퍼지 입력 패턴  $X$ 의  $j$ 번째 특징량에 대한 퍼지 집합  $X_j$ 가 표준 퍼지 패턴  $A_j$ 의  $j$ 번째 특징량에 대한 퍼지 집합  $A_j$ 와 어느 정도 일치하는가를 나타낸다.  $M(X|A_j)$ 는 일반적으로 스칼라 형식으로 변환되며 다음과 같은 표현이 가능하다<sup>[2]</sup>.

$$M(X|A_j) = \sum_{i=1}^c M(X_j|A_{ij})/q \quad (2)$$

식 (2)에 나타낸 퍼지 패턴 매칭의 결과는 클래스  $\omega_i$ 에 대한  $X$ 의 소속 정도(grade of belongingness)를 나타내게 된다. 즉  $M(X|A_j)$ 가 크면 클수록  $i$ 번째 클래스에 더 많이 소속된다고 볼 수 있는 것이다. 전체 클래스 개수를  $c$ 라 하면, 최종적으로 클래스  $\omega_i$ 에 대한  $X$ 의 소속 정도를 기반으로 한 분류는 다음과 같이 수행된다.

$$\max_{1 \leq i \leq c} M(X|A_i) = M(X|A_{i_0}) \rightarrow X \in \omega_{i_0} \quad (3)$$

즉, 미지의 입력 패턴을 퍼지 패턴 매칭값이 가장 큰 클래스에 할당하는 것이다.

본 논문에서  $M(X_j|A_j)$ 의 계산은 퍼지 집합 이론으

로부터 정의된 다음과 같은 형태의 퍼지 유사 지표 (fuzzy equality index)를 이용하여 구해진다<sup>[7]</sup>

$$[X_j \equiv A_j](x_0) = \begin{cases} b-a+1 & , \text{if } a > b \\ 1 & , \text{if } a = b \\ a-b+1 & , \text{if } a < b \end{cases} \quad (4)$$

단,  $a = X_j(x_0)$ ,  $b = A_j(x_0)$ .

유한 집합(finite set)의 경우 매칭 값  $M(X_j|A_j)$ 는 다음과 같이 계산될 수 있다.

$$M(X_j|A_j) = \min_{x_0} [X_j \equiv A_j](x_0) \quad (5)$$

간단한 예로서 특징량이 하나일 경우 그림 1과 같이 소속 함수가 주어져 있다고 하자.

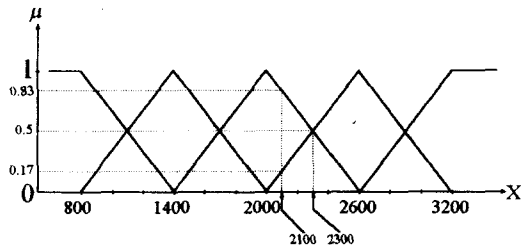


그림 1. 소속 함수의 예  
Fig. 1. Example of Membership functions.

첫번째 표준 패턴의 값이 2100이고 미지의 입력 패턴의 값이 2300이면 표준 퍼지 패턴  $A_{11}$ 과 퍼지 입력 패턴  $X_1$ 을 다음과 같이 구할 수 있다.

$$A_{11} = [0, 0, 0.83, 0.17, 0]$$

$$X_1 = [0, 0, 0.5, 0.5, 0]$$

식 (4)에 의해 등가 정도를 계산하고 식 (5)를 이용하면 다음과 같은 매칭 값을 얻을 수 있다.

$$M(X|A_1) = M(X_1|A_{11}) = \min(1, 1, 0.67, 0.67, 1) = 0.67$$

IV. 유전 알고리즘을 이용한 FPMC의 설계

1. 개요

본 논문에서는 FPMC를 설계하는데 있어서 패턴 분류에 적합한 소속 함수의 형태 및 개수 등을 구하기 위하여 최적화 기법의 하나인 유전 알고리즘을 이용하였다. FPMC와 유전 알고리즘간의 효과적인 결합을 위하여 각 문자열은 그림 2와 같은 의미를 가진다고 가정한다<sup>[8]</sup>

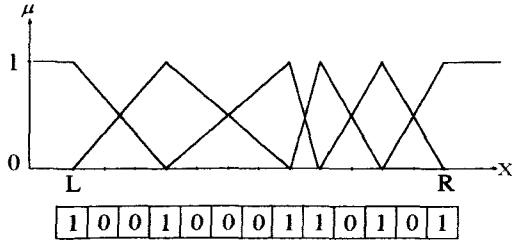


그림 2. 소속 함수의 형태  
Fig. 2. The form of membership function.

그림 2에서 보는 바와 같이 각 소속 함수는 삼각형의 모양을 가지며, 각 소속 함수의 폭은 이웃하는 소속 함수의 중심까지의 거리로 정의된다. 소속 함수들의 배열은 '0'과 '1'로 이루어진 스트링의 형식에 의해 결정되며 이진 숫자열의 '1'에 해당하는 부분은 각 소속 함수의 중심을 의미하게 된다.

2. 유전 알고리즘을 이용한 소속 함수의 설계

각 특징량에 대한 소속 함수를 설계하기 위하여 먼저 결정해야 할 일은 소속 함수에 대한 전체 집합(universe of discourse)을 결정하는 것이다. 주어진 데이터 집합,  $X = \{x_1, x_2, \dots, x_n\}$  ( $x_k \in R^q$ ,  $q$  is the number of features)에 대하여 먼저 각 특징량의 값들을 모두 포함하는 구간의 하한, 상한인  $l$ 과  $r$ 을 정의하도록 하자.

$$l(j) = \min_i x_{ij}$$

$$r(j) = \max_i x_{ij} \tag{6}$$

$l$ 과  $r$ 을 기초로  $j$ 번째 특징량에 대한 전체 집합의 하한, 상한인  $L$ 과  $R$ 을 다음과 같이 정의한다.

$$L(j) = l(j) - \alpha_j \quad (\alpha \geq 0)$$

$$R(j) = r(j) + \beta_j \quad (\beta \geq 0) \tag{7}$$

$\alpha, \beta$ 는 소속 함수의 모양에 여유를 주기 위한 여유분이다. 본 논문에서  $\alpha_j = \beta_j$ 로 하였으며  $\alpha_j$ 는 다음과 같이 구하였다.

$$\alpha_j = \frac{r(j) - l(j)}{2} \tag{8}$$

그림 3에  $L$ 과  $R$ 을 전체 집합의 하한, 상한으로 하여 구한 소속 함수의 예를 보여준다.

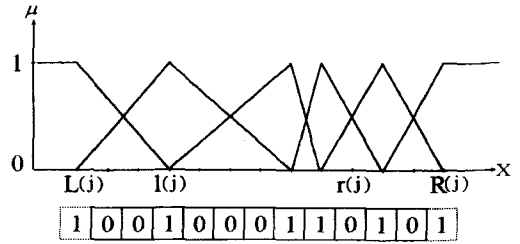


그림 3.  $r(j), l(j), R(j), L(j)$ 를 기반으로 한 소속 함수의 형태  
Fig. 3. The form of membership functions based on  $r(j), l(j), R(j), L(j)$ .

다음에는 유전 알고리즘에서 사용하는 이진 스트링(binary string)을 이용하여 각 소속 함수의 개수와 모양을 결정하는 방법에 대하여 알아보도록 하자. 한 개체(individual)중 한개의 특징량에 대한 이진 스트링이  $m$ -bit로 이루어져 있다고 하자. 그림 3에서 보는 바와 같이 소속 함수의 양끝은 항상 같은 모양을 하고 있으므로  $m$ -bit의 이진 스트링 좌우에 '1'을 첨가한 스트링을 가지고 소속 함수를 구성하게 된다. 즉, 그림 3과 같은 형태의 소속 함수를 만들기 위한 이진 스트링의 비트 수  $m$ 은 13이 아니라 11이 된다. 이렇게 구해진 이진 스트링의 1bit 간은 실제 소속 함수를 구성하는데 있어  $(R(j) - L(j)) / (m + 1)$ 의 간격이 된다. 전체  $(m + 2)$ -bit중 좌로부터  $k$ 번째 비트가 '1'일 경우 이 위치는 각 소속 함수의 중심이 되며 그 값은 다음과 같이 구해지게 된다.

$$L(j) + k * (R(j) - L(j)) / (m + 1) \tag{9}$$

이러한 식으로 각 특징량에 대한 이진 스트링을 소속 함수로 변환시킬 수 있고, 스트링 내의 '1'의 위치는 소속 함수의 모양과 개수를 결정하게 된다. 이렇게 구해진 소속 함수를 기반으로 퍼지 패턴 매칭을 수행할 수 있다. 퍼지 패턴 매칭에서 입력 패턴에 대해 각 클래스의 표준이 되는 표준 패턴  $V_i$ 는 다음과 같이 각 클래스에 속한 패턴들의 평균(numerical mean)으로 구할 수 있다.

$$V_i = \frac{1}{N_i} \sum_{x \in \omega_i} x, \quad i = 1, 2, \dots, c \tag{10}$$

여기서  $N_i$ 는 클래스  $\omega_i$ 에 속하는 패턴의 개수이다. 유전 알고리즘 수행중 적합도(fitness)를 구하기 위한 적합도 함수(fitness function)는 다음과 같이 정

의하였다.

$$fitness = \frac{1}{1 + classification\_error} \quad (11)$$

이 값은 예상할 수 있듯이 적합도는 분류 에러 (classification error)가 '0'일 때 최대값 '1'을 가지며 에러가 커지면 커질수록 작은 값을 갖게 된다. 전체 알고리즘은 다음과 같다.

- i) 스트링 집단(population)을 초기화시킨다.
- ii) 위에서 기술한 방법으로 스트링을 이용하여 각 특징량에 대한 소속 함수를 구성한다.
- iii) 구해진 소속 함수를 이용하여 퍼지 패턴 매칭을 수행한다.
- iv) 퍼지 패턴 매칭의 결과로부터 구해지는 분류 에러로부터 적합도를 계산한다.
- v) 원하는 적합도에 도달한 개체(individual)가 존재하면 알고리즘 수행을 끝낸다.
- vi) 각 개체에 대한 적합도를 기반으로 하여 II장에서 기술한 유전 알고리즘의 세가지 연산인 복제, 교배, 돌연변이를 수행한다.
- vii) 미리 정한 최대 세대수(maximum generation)에 도달하였으면 알고리즘 수행을 끝내고 전체 세대중 가장 좋은 적합도를 가지는 스트링을 최종 결과로 취한다. 그렇지 않으면 ii)로 간다.

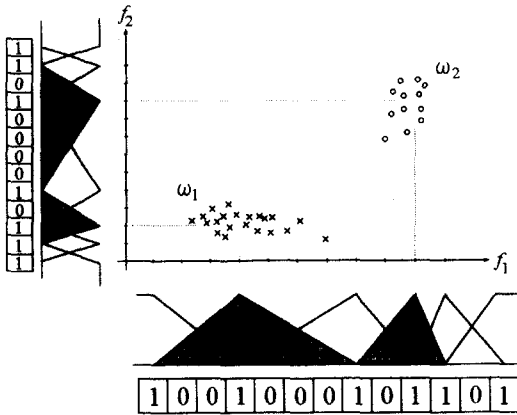
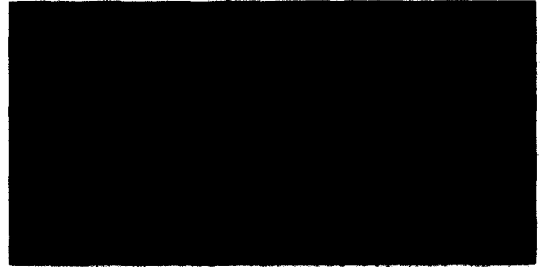


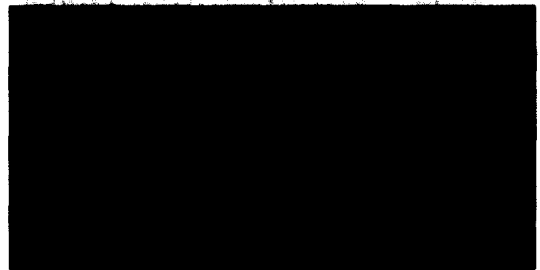
그림 4. GA를 이용한 소속 함수 설계의 예  
Fig. 4. Example of membership function design.

3. 유전 알고리즘을 이용한 소속 함수 설계의 예  
그림 4에 유전 알고리즘을 이용하여 구한 소속 함수의 예를 보였다. 직관적으로 알 수 있듯이 유전 알고리

즘을 통해 두개의 특징량에 대한 소속 함수가 그림 4처럼 구해졌다면  $\omega_1$ 과  $\omega_2$ 를 에러 없이 분류할 수 있는 분류기가 구성될 수 있다. 일반적인 베이시안 (Bayesian) 분류기에서 특히 인식에 결정적인 영향을 미치는 특징량(feature)의 실제적인 통계 특성들—mean, variance, pdf—을 샘플 데이터로부터 구할 필요 없이 소속 함수 내에 이러한 통계 특성이 포함되게 된다.



(a)



(b)

그림 5. 타이어 접지면 패턴의 원영상

(a) 타이어 접지면 패턴  $p_1$

(b) 타이어 접지면 패턴  $p_6$

Fig. 5. Original images of tire tread patterns.

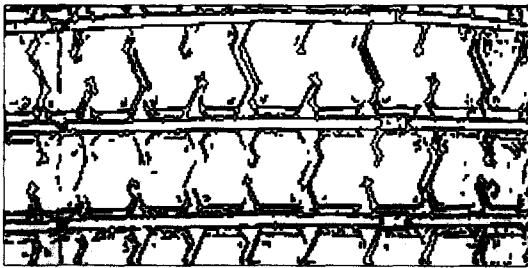
(a) Tire tread pattern  $p_1$

(b) Tire tread pattern  $p_6$

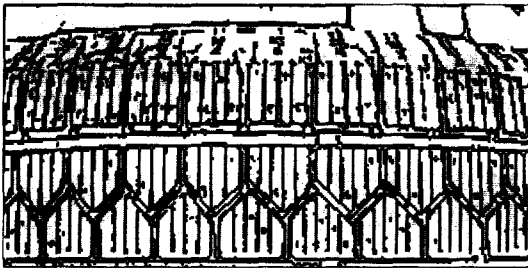
### V. 실험 결과

5장에서는 4장에서 제안한 방법으로 FPMC를 구성하여 두 종류의 패턴에 대하여 실험을 행하고 그 타당성을 보이코자 한다. 두 종류의 패턴은 각각 타이어 접지면 패턴과 필기체 영문자 패턴이다. 유전 알고리즘은 그 특성상 일반적으로 세대수(generation number)가 증가할수록 적합도(fitness)가 증가하는 양상을 보인다. 그러나 유전 알고리즘의 특성상 알고리즘을 실행하는 도중에 가장 좋은 적합도를 가지는 개체가 나타

날 수 있다. 따라서 본 실험에서는 실행할 세대수와 개체 수를 각 경우에 대해 일정하게 정해 놓고 유전 알고리즘의 실행중 가장 높은 적합도를 가지는 개체, 즉 이진 문자열을 선정하였다.



(a)



(b)

그림 6. 타이어 접지면 패턴의 전처리된 영상

(a) 전처리 후의 패턴  $p_1$

(b) 전처리 후의 패턴  $p_6$

Fig. 6. Preprocessed images of the tire tread patterns.

(a)  $p_1$  pattern after preprocessing

(b)  $p_6$  pattern after preprocessing

1. 제안되는 알고리즘의 타이어 접지면 패턴 분류에의 적용

1) 실험에 사용된 타이어 접지면 패턴과 특징량 추출

타이어 접지면 패턴 분류를 위한 데이터 취득 부분은 타이어 접지면의 영상 취득 과정 및 특징량 추출을 위한 전외곽선 검출로 이루어지며, 전자는 영상 취득후 처리해야 할 데이터량을 줄이기 위하여  $3 \times 3$  마스크로 마스킹을 취한 후 4화소당 한 화소를 취함으로써 행해진다. 한편, 화상의 이치화 과정은 윤곽선 검출(edge detection)을 동시에 행할 수 있는 알고리즘을 사용하였다. 그림 5(a), (b), 그림 6(a), (b)는 각각 샘플 패턴중 두종류의 원영상과 전처리 후의 영상을 나타낸다.

위와 같이 전처리된 영상으로부터 알 수 있듯이 전처리된 패턴의 각도 성분은 패턴 분류에 유용한 특징량으로

이용될 수 있음을 알 수 있다. 윤곽선의 방향에 대한 정보는 이치화 영상에 대한 계조치 동시 발생 행렬(gray-level cooccurrence matrix)이라는 2차 통계량을 통해 표현될 수 있으며 이러한 행렬은 이치화 영상에 대해  $2 \times 2$  행렬이 된다<sup>[2]</sup>. 이 행렬은 화상 내에 존재하는 각 화소들간의 공간적인 관계(spatial relation)를 나타내며 이 행렬로부터 영상의 통계적인 특징들을 추출할 수 있다. 특히 행렬 요소중 (2, 2) 요소는 윤곽선의 양에 대한 정보를 가지고 있으므로 이를 특징량으로 사용할 수 있다. 본 실험에서는 타이어에 주로 존재하는 8방향의 각도 성분에 대해 계조치 동시 발생 행렬을 구하고 이로부터 (2, 2) 요소를 추출하여 특징량으로 사용하였다. 표 1에 각 패턴 클래스에 속한 샘플 패턴의 특징량을 나타내었다.

표 1. 샘플 타이어 접지면 패턴의 특징량 벡터

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

Patterns	Features							
	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$
$p_1$	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

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. 유전 알고리즘을 수행하는데 있어서 세대수(number of generation)는 100, 집단수(number of population)는 20으로 하였으며, 교배 확률(crossover probability) 및 돌연 변이 확률(mutation probability)인  $P_c, P_m$ 은 각각 1.0, 0.1로 하였다.

3) 실험 결과

그림 7에 각 세대에 따른 최대 적합도(maximum fitness)와 평균 적합도(average fitness)를 도시하였다. 최대 적합도는 각 세대에서 가장 좋은 적합도를 가지는 개체의 적합도를 의미하며, 평균 적합도는 각 개체에 대한 적합도의 전체 평균이다. 그림에서 알 수 있듯이 5번째 세대에서 이미 분류 에러를 '0'으로 만

드는 개체가 나타나고 있으며 세대수가 증가할수록 평균 적합도는 증가함을 알 수 있다. 최대 세대수는 100이었으나 세대수 10까지만 도사하였다.

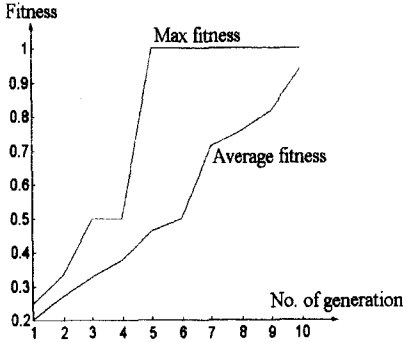


그림 7. 유전 알고리즘 수행 중의 최대 적합도와 평균 적합도  
Fig. 7. Maximum fitness and average fitness in genetic algorithm.

가장 높은 적합도를 가지는 개체의 첫번째 특징량에 대한 이진 문자열 형태와 그에 해당하는 소속 함수의 중심 값을 그림으로 표시하면 그림 8과 같다. 그림 4에서 보여지는 바와 마찬가지로 소속 함수의 개수는 분류하고자 하는 군집의 개수와 일치할 필요가 없는데 왜냐하면 여기서 각각의 소속 함수는 각 군집에 대한 소속 정도를 나타내는 것이 아니라 언어 규칙 (linguistic rule)에 대한 표현이기 때문이다. 참고로 위 패턴을 단순한 최소 거리 분류기(minimum distance classifier)에 의해 분류를 하였을 경우 2개의 분류 에러를 나타내었다.

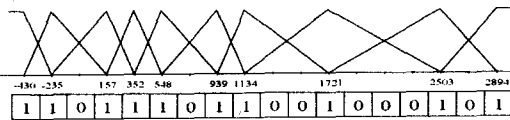


그림 8. 유전 알고리즘을 이용하여 구성된 소속 함수  
Fig. 8. Membership functions using genetic algorithm.

2. 제안되는 알고리즘의 영문자 분류에의 적용

1) 실험에 사용된 영문자와 특징량 추출

필기체 문자 분류에 사용된 문자의 종류는 영문자중 대문자 A~Z, 26종이며 각 문자당 다섯개씩 총 130개의 문자를 스캐너로 취득하였다. 나열된 여러 문자들을 x, y축

상에 투영하여 개별 문자들로 분리해 이치 영상으로 만들었다. 그림 9에 필기체 영문자의 예를 보인다.

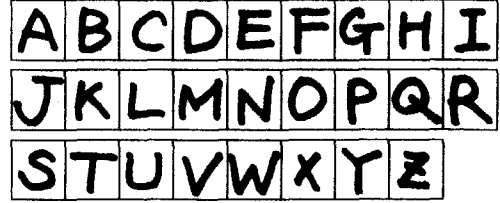


그림 9. 필기체 영문자  
Fig. 9. Handwritten alphabetic characters.

문자의 특성을 나타내는 특징량은 문자의 모멘트(moment), 방향 성분, 화소값 변화 등 여러 가지 방법으로 추출할 수 있으나 본 논문에서는 인식을 위한 특징량을 다음과 같이 간단하게 추출하였다. 문자를 둘러싸는 최소의 사각형에 대하여 각 면을 4등분하는 3점으로부터 문자까지의 거리를 구한다. 상, 하, 좌, 우 4면에 대하여 각각 3개씩의 특징량을 추출하면 그림 10과 같이 각각의 문자에 대해 12개의 특징량이 얻어진다. 그림 10을 보면 이 과정을 쉽게 알 수 있다. 표 2에 각 영문자에 대한 특징량의 예를 보인다.

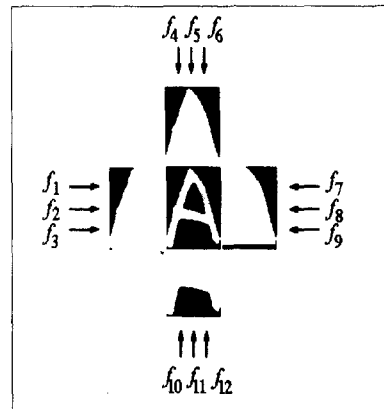


그림 10. 필기체 영문자의 특징량 추출  
Fig. 10. Feature extraction of handwritten alphabetic character.

2) 실험 조건

실험에 사용된 영문자 패턴은 모두 26종류이며 각 종류당 5개의 샘플을 가지고 있어 총 130개의 패턴을 사용하였다. 유전 알고리즘을 수행하는데 있어서 세대수(number of generation)는 300으로 하였으며, 집단수, 교배 확률, 그리고 돌연 변이 확률은 각각 앞서

타이어 접지면 패턴 분류 때와 마찬가지로 각각 20, 1.0, 0.1로 하였다.

표 2. 샘플 영문자 패턴의 특징량 벡터  
Table 2. Feature vectors for the sample alphabetic character patterns.

Patterns	Features											
	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$	$f_9$	$f_{10}$	$f_{11}$	$f_{12}$
A	7	3	0	12	1	11	7	2	3	4	7	5
B	0	2	4	1	2	6	6	1	2	9	0	2
C	3	0	3	4	0	1	1	11	5	3	0	2
D	1	2	3	0	0	2	3	0	8	6	1	6
E	1	0	0	1	0	9	8	9	8	0	0	0
F	2	0	0	3	0	0	8	12	16	2	7	19
⋮												
Z	3	6	4	1	0	0	2	8	7	8	0	2

3) 실험 결과

그림 11에 각 세대에 따른 최대 적합도(maximum fitness)와 평균 적합도(average fitness)를 도시하였다. 그림에서 알 수 있듯이 142번째 세대에서 분류 에러가 '3'인 개체가 나타나고 있으며 181번째 세대 이후부터는 최대 적합도의 변화는 거의 없고 평균 적합도만이 점차 증가하는 것을 보여준다.

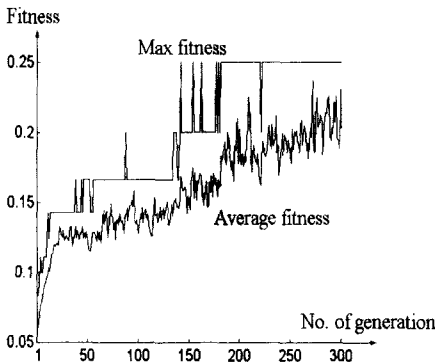


그림 11. 유전 알고리즘 수행 중의 최대 적합도와 평균 적합도

Fig. 11. Maximum fitness and average fitness in genetic algorithm.

가장 높은 적합도를 가지는 개체의 첫번째 특징량에 대한 이진 문자열 형태와 그에 해당하는 소속 함수의 중심 값을 그림으로 표시하면 그림 12와 같다. 참고로 위 패턴을 단순한 최소 거리 분류기(minimum distance classifier)에 의해 분류를 하였을 경우 8개의 분류 에러를 나타내었다.

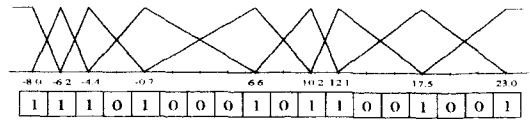


그림 12. 유전 알고리즘을 이용하여 구성된 소속 함수의 예

Fig. 12. Example of membership functions using genetic algorithm.

VI. 결 론

본 논문에서는 최적화 기법의 하나로써 최근 많은 관심을 모으고 있는 유전 알고리즘을 이용하여 퍼지 패턴 매칭 분류기를 구성하여 보았다. 제안된 방법은 퍼지 패턴 매칭 분류기의 성능을 좌우하는 가장 중요한 요소인 소속 함수의 형태 및 개수를 유전 알고리즘을 통해 얻음으로서 각 특징량에 따라 서로 다른 형태의 소속 함수를 얻을 수 있었고 최소 거리 분류기에 비해 좋은 분류 결과를 얻을 수 있었다.

물론 유전 알고리즘은 제한된 세대수와 개체로 인하여 어떤 경우에는 더 좋은 결과를 낼 수 있는 소속 함수를 찾지 못하여 응답이 좋지 않은 경우도 있었다. 그리고 각 이진 문자열이 어떤 실수 값을 나타내는 것이 아니라 소속 함수의 구조를 나타내기 때문에 교배나 돌연변이의 시행시 소속 함수의 구조 자체가 파괴되어 적합도가 향상되지 못하는 경우도 있었다.

향후 연구 과제로는 소속 함수의 모양과 개수뿐만 아니라 전체 집합의 상한, 하한 및 각 패턴의 중심값 결정에도 유전 알고리즘을 사용하여 퍼지 패턴 매칭 분류기의 성능을 향상시키는 것이라고 할 수 있다.

참 고 문 헌

[1] R. O. Duda, and P.E. Hart, *Pattern Classification and Scene Analysis*, New York: Wiley-Interscience, 1973.  
[2] S. W. Jung, S.W.Bae, and 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.



- [3] J. I. Fujimoto, T. Nakatani and M. Yoneyama "Speaker independent word recognition using fuzzy pattern matching". *Fuzzy Sets and Systems*, vol. 32, pp. 181-191, 1989.
- [4] W. Pedrycz, Fuzzy sets in pattern recognition : Methodology and methods, *Pattern Recognition*, Vol. 23, pp 121-146, 1990.
- [5] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Reading, MA : Addison-Wesley, 1989.
- [6] M. Srinivas, and L. M. Patnaik, "Adaptive probabilities of crossover and mutation in genetic algorithms", *IEEE Trans. SMC* Vol. 24, No. 4, pp. 656-667, 1994.
- [7] W. Pedrycz, "Direct and Inverse Problem in Comparison of Fuzzy Data", *Fuzzy Sets and Systems*, 34, pp. 223-235, 1990.
- [8] H. Nomura, I. Hayashi, and N. Wakami, "A learning method fusing fuzzy reasoning and genetic algorithm", in *Proc. of the IMACS/SICE Int. Symp.*, pp. 155-160, 1992.

---

 저 자 소 개
 

---

鄭 淳 元(正會員) 第 32卷 B編 第 11號 參照  
 현재 고려대학교 대학원 전기공학  
 과 박사과정 재학중

朴 貴 泰(正會員) 第 31卷 B編 第 6號 參照  
 현재 고려대학교 전기공학과 교수,  
 서울대 ERC-ACI 연구 위원