

논문 2005-42CI-4-6

# KNN 규칙과 새로운 특징 가중치 알고리즘을 결합한 패턴 인식 시스템

(Pattern Recognition System Combining KNN rules and New Feature  
Weighting algorithm)

이 회 성\*, 김 은 태\*\*, 김 동 연\*\*\*

(Hee-Sung Lee, Euntai Kim, and Dongyeon Kim)

## 요 약

본 논문에서는 유전자 알고리즘을 이용한 새로운 적응적 특징 가중치 방식과 클래스별로 적용된 KNN(Nearest-Neighbor) 규칙을 이용한 새로운 패턴 인식 시스템을 제안한다. 패턴 인식 시스템의 성능을 향상시키기 위하여, 새로운 연산자를 갖는 유전자 알고리즘으로 가중치의 중간값을 결정함으로써 과잉 맞춤(overfitting)을 피하면서, 데이터의 분포에 따라 적절한 특징의 가중치를 찾는 새로운 특징 가중치 알고리즘을 제안한다. 또한, 제안하는 방법은 각각의 클래스를 가장 잘 표현하는 특징 공간들을 개별적으로 찾는다. KNN분류기는 클래스별로 찾은 특징 공간들을 이용하여 클래스에 따라 특징 공간을 변화시켜 미지 패턴의 클래스를 예측한다. 제안된 알고리즘은 Concordia대학의 handwritten numeral database에 적용시켜 그 성능을 확인하였다.

## Abstract

This paper proposes a new pattern recognition system combining the new adaptive feature weighting based on the genetic algorithm and the modified KNN(K Nearest-Neighbor) rules. The new feature weighting proposed herein avoids the overfitting and finds the proper feature weighting value by determining the middle value of weights using GA. New GA operators are introduced to obtain the high performance of the system. Moreover, a class dependent feature weighting strategy is employed. Whilst the classical methods use the same feature space for all classes, the proposed method uses a different feature space for each class. The KNN rule is modified to estimate the class of test pattern using adaptive feature space. Experiments were performed with the unconstrained handwritten numeral database of Concordia University in Canada to show the performance of the proposed method.

**Keywords :** pattern recognition, KNN rules, GA, feature weighting

## I. 서 론

패턴 인식 시스템은 응용분야가 방대하고 활용도가 크기 때문에, 여러 응용분야의 인식이론 및 기술들이 많은 공학자들에 의해 개발되고 있다. 일반적으로 패턴은 특징 공간에서의 벡터로 표현되기 때문에, 특징들의 측정과 측정된 특징의 선택은 패턴 인식 알고리즘의 결

\* 학생회원, \*\* 평생회원, 연세대학교 전기전자공학부  
(School of Electrical and Electronic Eng., Yonsei Univ.)

\*\*\* 정회원, 국립 한양대학교 전자공학과  
(Dept. of Electronic Eng., Hankyong National Univ.)

접수일자: 2005년4월4일, 수정완료일: 2005년7월1일

과에 중요한 영향을 미친다. 따라서 패턴을 정확하게 인식하기 위해서는, 적합한 특징의 선택이 매우 중요하다. 하지만 어느 특징들이 클래스들 간의 가장 좋은 차별성을 제공하는지 알 수 없다. 또한 패턴을 표현 가능하게 하는 하위 특징 공간의 수도 무수히 많다. 그 결과 특징의 측정과 측정된 특징의 추출은 패턴 인식 시스템의 성능에 중요한 역할을 한다<sup>[1]</sup>.

본 논문에서는 유전자 알고리즘<sup>[2]</sup>을 이용한 적응적 특징 가중치(Adaptive-3FW)방식과 클래스별로 적용된 KNN 규칙을 결합한 새로운 패턴 인식 시스템을 제안한다. 우선, 학습 데이터의 특징 벡터에 적응적 3FW 알고리즘을 각 클래스별로 적용하여, 각각의 클래스를 표현하는 특징 공간들을 구한다. 다음으로 미지패턴의 클래스를 결정하기 위하여, 미지패턴의 특징 벡터를 구한다. 클래스에 따라 특징 공간이 클래스에 따라 변하고, 각 클래스별 특징 공간에서 최소 거리를 갖는 학습 데이터의 클래스를 미지패턴의 클래스로 선택하게 된다.

본 논문의 구성은 다음과 같다. II장에서는 유전자 알고리즘과 KNN 규칙 및 FW 알고리즘을 설명하고, III장에서는 새로운 FW 알고리즘을 이용한 패턴 인식 시스템을 제안한다. IV장에서는 제안된 시스템의 효용성을 보이기 위한 실험과 그의 고찰을 한 뒤 마지막으로, V장에서는 결론과 추후 과제에 대한 설명을 한다.

## II. 특징 가중치 방법의 이론적 배경

### 1. KNN 규칙

전체 표준패턴 중에서 패턴  $\mathbf{X}$ 로부터 가장 가까운 거리에 있는  $K$ 개의 패턴을  $\mathbf{X}$ 의 KNN이라 하며, 패턴  $\mathbf{X}$ 의 KNN이 가장 많이 속해있는 클래스를  $\mathbf{X}$ 의 클래스로 결정한다. 이와 같은 기본원리에 따른 최소거리 분류 규칙을 KNN분류 규칙이라고 하며, 이를 위하여 사전에 클래스별 기준이 되는 패턴을 선정하여야 한다<sup>[3]</sup>. 클래스  $S_i$ 의 구성원인 특징벡터  $\mathbf{X}$ 의 가장 가까운 이웃(1NN)이 클래스  $S_i$ 의 구성원일 확률, 다시 말해서 1NN 규칙에 의해  $\mathbf{X}$ 가 옳게 분류될 확률은  $P(S_i | \mathbf{X})$ 이기 때문에 클래스  $S_i$ 의 멤버가 옳게 분류될 확률인  $P(S | S_i)_{NN}$ 은 다음과 같이 표현된다.

$$P(S | S_i)_{NN} = \int_{R^n} P(S_i | \mathbf{X}) p(\mathbf{X} | S_i) d\mathbf{X} \quad (1)$$

식(1)에 Bayes formula를 적용하면 식(2)와 같다.

$$P(S | S_i)_{NN} = P(S_i) \int_{R^n} \frac{p(\mathbf{X} | S_i)^2}{p(\mathbf{X})} d\mathbf{X}, \quad (2)$$

$$p(\mathbf{X}) = \sum_i P(S_i) p(\mathbf{X} | S_i)$$

따라서  $S_i$ 의 멤버가 잘못 분류될 확률은 다음과 같다.

$$P(e | S_i)_{NN} = 1 - P(S | S_i)_{NN} \\ = 1 - P(S_i) \int_{R^n} \frac{p(\mathbf{X} | S_i)^2}{p(\mathbf{X})} d\mathbf{X} \quad (3)$$

$c$ 개의 클래스를 고려하면 전체 클래스에 대한 에러 확률은 식(4)이다.

$$P(e)_{NN} = \sum_{i=1}^c P(S_i) P(e | S_i) \\ = \int_{R^n} [1 - \sum_{i=1}^c P(S_i | \mathbf{X})^2] p(\mathbf{X}) d\mathbf{X} \quad (4)$$

만약,  $\mathbf{X}$ 의 실제 클래스가  $S_m$ 이라고 한다면,  $P(S_i | \mathbf{X})$ 는 식(5)와 같이 표현할 수 있다.

$$P(S_i | \mathbf{X}) = \begin{cases} \frac{P(e | \mathbf{X})_{Bayes}}{c-1} & i \neq m \\ 1 - P(e | \mathbf{X})_{Bayes} & i = m \end{cases} \quad (5)$$

여기서  $P(e | \mathbf{X})_{Bayes}$ 는 식 (6)과 같다.

$$P(e | \mathbf{X})_{Bayes} = \sum_{i \neq m} P(S_i | \mathbf{X}) = 1 - P(S_m | \mathbf{X}) \quad (6)$$

따라서 다음과 같은 부등식을 유도할 수 있다.

$$\sum_{i=1}^c P(S_i | \mathbf{X})^2 \geq (1 - P(e | \mathbf{X})_{Bayes})^2 + \frac{P(e | \mathbf{X})_{Bayes}^2}{c-1} \quad (7)$$

이로부터 Nearest Neighbor Error Rate의 상한은 Bayesian Error Rate가 된다는 식(8)로 유도할 수 있다.

$$P(e)_{NN} \leq 2P(e)_{Bayes} - \frac{c}{c-1} P(e)_{Bayes}^2 \quad (8)$$

실제의 경우, Bayesian Error Rate의 신뢰할만한 추정 하한이 큰 값을 갖는다면, 좋은 Decision making 방법을 찾는 것은 의미가 없다. 그러므로 이런 경우에는 성능을 개선하기 위하여 더 좋은 특징들을 찾아야 한다.

2. 유전자 알고리즘 (Genetic Algorithm)

유전자 알고리즘은 진화 알고리즘의 한 부류이며, 진화 알고리즘은 연산모델을 사용한, 컴퓨터를 기반으로 한 문제 풀이 체계의 설계나 실행에 이미 알려진 진화 메커니즘 중 일부를 중요 요소로 사용하는 것들을 통칭하는 용어이다<sup>[2]</sup>. 유전자 알고리즘은 구조체들의 개체군 (population)을 유지하는데, 이들은 선택의 규칙과 탐색 연산자들 혹은 유전자 연산자들이라 불리는 다른 연산자들의 의해 진화하게 된다.

3. Feature Weighting

Siedlecki와 Sklansky에 의해 처음 소개된 유전자 알고리즘을 이용한 특징의 선택 알고리즘에서, GA는 그림 1과 같이 각각의 비트가 특징과 연관되어있는 최적의 이진 벡터를 찾기 위해 사용되었다<sup>[4]</sup>. 만약,  $i$  번째 비트가 1이라면  $i$  번째 특징은 분류기의 입력에 포함된다. 반대로 0이라면 입력에 포함되지 않는다. GA의 염색체인 이진 벡터에 의해 만들어지는 특징들의 부분집합은 분류기에 의한 정확도에 의해 평가된다. GA는 이 정보를 이용하여 최적의 특징들의 조합을 찾게 된다.

FS알고리즘의 단점은 사용되는 특징들이 모두 같은 중요함(priority)을 갖는다는 것이다. 간단한 해결책은 각각의 특징에 가중치를 부여하는 것이다. Kelly와 Davis<sup>[5]</sup>는 각각의 특징들을 선형 스케일링할 수 있도록 염색체의 각각의 특징에 대응하는 비트들을 실수로 확장하였다.

그림 2에서  $w_i$ 는  $i$  번째 특징과 대응되는 염색체의 비트에 들어가는 수이다. 각각의 특징은 정규화 된 다

음, 대응되는 가중치에 의해 스케일링된다. 이러한 스케일링은 특징들에 다른 중요성을 부여함으로써 보다 정확하게 분류기가 패턴을 분류할 수 있도록 도와준다.

특징 가중치 알고리즘은 특징들을 스케일링 하며 패턴의 클래스의 예측을 변화시킬 수 있기 때문에 보다 정확하게 패턴을 분류할 수 있는 새로운 특징의 집합을 구할 수 있다. 하지만 과잉 맞춤(overfitting) 문제로 일반적인 패턴 인식 문제의 경우 각 염색체의 비트가 실수를 갖는 일반적인 FW 알고리즘보다, 가중치가 0, 0.5, 1의 3개만을 고려하는 3FW 알고리즘이 좋은 성능을 갖는다고 알려져 있다<sup>[6]</sup>.

III. Pattern Recognition system 설계

분류기의 학습 모드는 그림 3과 같은 블록 다이어그램으로 구성되어 있다. 블록 다이어그램의 각 단계를 이번 장에서 서술한다.

1. Feature Measurement

본 논문에서는 제안하는 알고리즘을 숫자 인식에 적용한다. 숫자들은 2차원 공간에서 1차원 선들의 집합이다. 그러므로 선들의 지역적 검출은 숫자의 적절한 특징 중의 하나이다. 본 논문에서는 빠른 계산과 숫자들을 구성하는 선들의 검출을 모두 만족시키는 일차 미분 에지 디텍터를 사용한다. Kirsch edge detector는 8-neighbor들을 모두 고려하기 때문에 가장 정확하게 4 방향의 에지들을 찾아준다고 알려져 있다. Kirsch는 비선형 에지 향상 알고리즘을 다음과 같이 정의 하였다<sup>[7]</sup>.

$$G(i, j) = \max \{1, \max [ 5S_k - 3T_k ] \} \tag{9}$$

여기서,  $S_k, T_k$ 는 다음과 같다.

$$S_k = A_k + A_{k+1} + A_{k+2}, \tag{10}$$

$$T_k = A_{k+3} + A_{k+4} + A_{k+5} + A_{k+6} + A_{k+7}. \tag{11}$$

식(9)에서  $G(i, j)$ 는  $(i, j)$ 픽셀의 그레디언트 값이고,  $A$ 의 서브스크립트(subscript)들은 각각 8의 나머지 값들이다. 그리고  $A_k (k=0, 1, \dots, 7)$ 는 그림 4와 같이 정의되는 픽셀  $(i, j)$ 의 8-Neighbor들이다.

우선, 입력 패턴을  $16 \times 16$ 으로 크기를 정규화 시킨 후, 수평(H), 수직(V), 우-대각(R), 좌-대각(L)의 방향 특징 벡터를 아래와 같은 식으로 계산한다.

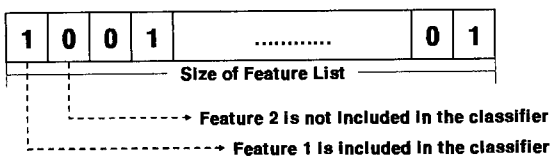


그림 1. FS알고리즘의 염색체의 구성  
Fig. 1. GA chromosome of Feature Selection algorithm.

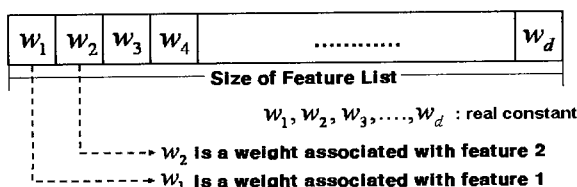


그림 2. FW알고리즘의 염색체의 구성  
Fig. 2. GA chromosome of Feature Weighting algorithm.

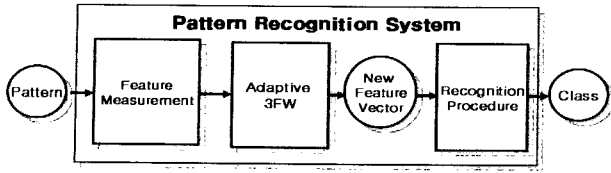


그림 3. 학습모드의 블록 다이어그램  
Fig. 3. Block diagram of Proposed recognition system.

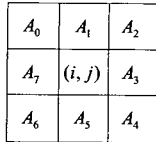


그림 4. 픽셀  $(i, j)$ 의 8-Neighbor  $A_k$  ( $k = 0, 1, \dots, 7$ )  
Fig. 4. 8-Neighbors  $A_k$  ( $k = 0, 1, \dots, 7$ ) of pixel  $(i, j)$ .

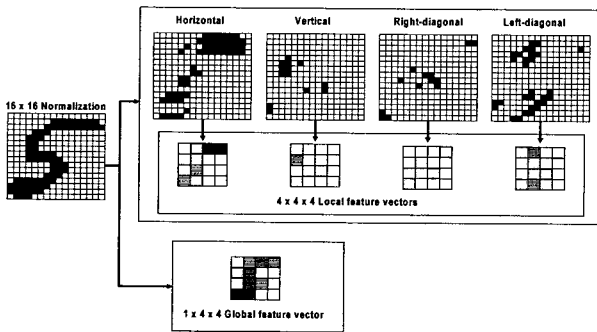


그림 5. 특징 추출의 전체 과정  
Fig. 5. Procedure of Feature Measurement.

$$\begin{aligned}
 G(i, j)_H &= \max[5S_0 - 3T_0, 5S_4 - 3T_4], \\
 G(i, j)_V &= \max[5S_2 - 3T_2, 5S_6 - 3T_6], \\
 G(i, j)_R &= \max[5S_1 - 3T_1, 5S_5 - 3T_5], \\
 G(i, j)_L &= \max[5S_3 - 3T_3, 5S_7 - 3T_7].
 \end{aligned} \tag{12}$$

각각의  $4 \times 4$ 의 픽셀을 하나의 픽셀로 축적시키는 방법으로 식 (12)로 계산된  $16 \times 16$  특징 벡터들을  $4 \times 4$ 의 특징 벡터들로 압축시킨다. 앞에서 계산된 특징 벡터는 입력 패턴의 지역적인 특징만을 고려하였다. 이것을 보완하기 위하여, 정규화 된 입력 영상을  $4 \times 4$ 로 압축시키는 광역 특징 벡터를 고려한다. 그 결과, 최종 특징 벡터는  $4 \times 4 \times 4$ 의 지역 특징 벡터와  $1 \times 4 \times 4$ 의 광역 특징 벡터로 구성되어 있다. 그림 5는 특징을 추출하는 전체 과정을 보여준다<sup>[8]</sup>.

## 2. Adaptive 3FW

제안한 알고리즘은 기존의 3FW의 방법과는 달리 가중치의 중간 값을 유전자 알고리즘으로 결정한다. 또한,

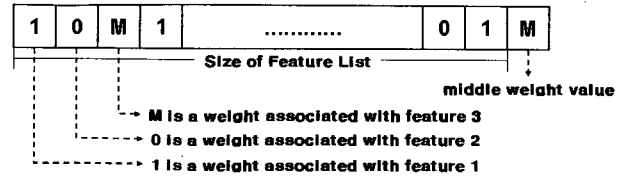


그림 6. 염색체의 구성  
Fig. 6. GA chromosome of our system.

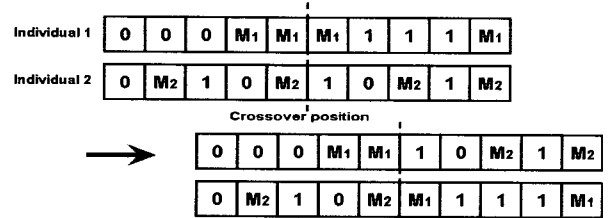


그림 7. 교차 연산  
Fig. 7. The result of classical one-point crossover.

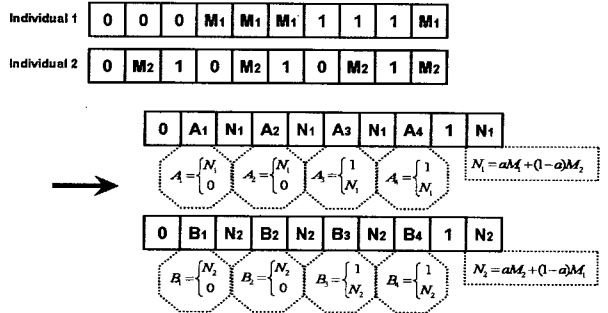


그림 8. 수정된 교차연산  
Fig. 8. Modified Crossover Operator.

새로운 교차 연산자와 돌연변이 연산자를 제안한다.

### 가. Encoding

염색체의 구성은 그림 6과 같다. 여기서  $M$ 은 0과 1 사이의 실수로써 가중치의 중간 값이다. 3FW알고리즘이 0, 0.5, 1의 가중치를 갖는 반면 제안하는 적응적 3FW알고리즘은 0,  $M$ , 1의 가중치를 갖고 있어, 과잉 맞춤을 피하면서, 데이터의 분포에 따라 적절한 특징의 가중치를 찾는 장점을 갖는다.

### 나. Genetic Operator

앞에서 정의된 염색체에 기존의 교차 연산을 적용하면 그림 7과 같다.

일반적인 교차 연산을 적용하면, 자식세대의 염색체가 제안하는 염색체의 구성을 벗어남을 볼 수 있다. 또한, 염색체의 7번째 비트에서 1에서 0으로 0에서 1로의 변화를 볼 수 있다. 이것은 자식세대와 부모세대의 염색체들의 형질이 너무 크게 변한다는 문제점을 갖는다.

더욱 심각한 문제점은 일반적인 교차 연산으로는  $M$ 이 변하지 않는다는 것이다. 이런 문제점을 해결하기 위하여 그림 8과 같은 수정된 교차 연산을 제안한다.

여기서  $N_1, N_2$  는 다음과 같다.

$$\begin{aligned} N_1 &= aM_1 + (1 - a)M_2, \\ N_2 &= aM_2 + (1 - a)M_1, \\ \text{where } a &= \text{random value } [0, 1] \end{aligned} \quad (13)$$

나머지  $A_1, A_2, A_3, A_4$  는 각각 다음과 같이 정의된다.

$$\begin{aligned} A_1 &= \begin{cases} N_1 & M_2 \geq 0.5 \\ 0 & M_2 < 0.5 \end{cases}, \quad A_2 = \begin{cases} N_1 & M_1 \geq 0.5 \\ 0 & M_1 < 0.5 \end{cases} \\ A_3 &= \begin{cases} 1 & M_1 \geq 0.5 \\ N_1 & M_1 < 0.5 \end{cases}, \quad A_4 = \begin{cases} 1 & M_2 \geq 0.5 \\ N_2 & M_2 < 0.5 \end{cases} \end{aligned} \quad (14)$$

$B_1, B_2, B_3, B_4$  도 같은 방식으로 다음과 같이 정의된다.

$$\begin{aligned} B_1 &= \begin{cases} N_2 & M_2 \geq 0.5 \\ 0 & M_2 < 0.5 \end{cases}, \quad B_2 = \begin{cases} N_2 & M_1 \geq 0.5 \\ 0 & M_1 < 0.5 \end{cases} \\ B_3 &= \begin{cases} 1 & M_1 \geq 0.5 \\ N_2 & M_1 < 0.5 \end{cases}, \quad B_4 = \begin{cases} 1 & M_2 \geq 0.5 \\ N_2 & M_2 < 0.5 \end{cases} \end{aligned} \quad (15)$$

교차 연산과 마찬가지로 일반적인 돌연변이 연산을 제안된 알고리즘에 직접적으로 적용할 수 없기 때문에 아래와 같이 정의된 수정된 돌연변이 연산을 제안한다.

```

procedure Modified mutation operator
begin
generate random value r1 from the range [0, 1];
if r1 < mutation_probability
then generate random value  $M_2$  from the range [0, 1];
    last bit( $M_1$ ) =  $M_2$ ; //1st
    while t ≤ (size of chromosome)
        begin
            t=t+1;
            generate random value r2 from the range [0, 1];
            if r2 < mutation_probability
                then mutate bit within {0,  $M_1$ , 1}; //2nd
        end
    end

```

다. Fitness Function

염색체(특징의 가중치 벡터)를 평가하기 위하여, 식 (16)과 같은 평가 함수를 사용하였다.

$$\text{Minimize : } error(w) = C_{pred}N_1 + C_{mask}N_2 \quad (16)$$

여기서  $C_{pred}, C_{mask}$  는 평가 함수의 계수,  $N_1$  은 튜닝 데

표 1. 유전자 알고리즘 파라미터와, 평가 함수의 계수

Table 1. Fitness Coefficient and GA Parameter.

Fitness Coefficient		GA parameters	
$C_{pred}$	20.0	Crossover rate	0.6
$C_{mask}$	1.0	Mutation rate	0.05
		Population size	40
		Generation	200

이터를 잘못 분류한 수,  $N_2$  는 사용되는 특징의 수이다.

$N_1$  을 계산하기 위하여 1NN 분류기를 사용한다. 우선 유전자 알고리즘의 염색체인 특징의 가중치와 학습 데이터를 곱하여 수정된 학습 데이터를 얻는다. 그리고 수정된 학습 데이터와 튜닝 데이터와의 거리를 계산하여 가장 최소 거리를 갖는 학습 데이터의 클래스를 튜닝 데이터의 클래스로 결정한다. 이렇게 계산된 튜닝 데이터의 클래스를 예측한 결과와 실제 튜닝 데이터의 클래스가 다른 개수가  $N_1$  이 된다. 실험에 사용된 유전자 알고리즘의 평가 함수의 계수인  $C_{pred}, C_{mask}$  와 유전자 알고리즘의 파라미터들은 각각 표 1과 같다.

$C_{pred}$  가  $C_{mask}$  에 비해 상대적으로 큰 이유는 분류기의 성능에서 인식률이 가장 중요한 요소이기 때문이다.

3. Classifier 설계

본 논문에서는 모든 클래스의 학습 데이터를 가장 잘 표현하는 특징 공간을 찾는 것과는 달리 각각의 클래스를 가장 잘 표현하는 특징 공간들을 개별적으로 찾는다. 즉, 클래스에 따라 특징 공간이 변하는 KNN 분류 규칙<sup>[5]</sup>을 사용한다. 우선, 적응적 3FW 알고리즘을 학습 데이터의 각 클래스별로 적용하여, 각 클래스를 나타내는 가중치 벡터들을 구한다. 미지 패턴이 들어오면 패턴으로부터  $5 \times 4 \times 4$ 의 특징 벡터를 구한다. 미지 패턴의 특징 벡터는 클래스에 따라 적응적으로 변하고, 각 클래스별 특징 공간에서 최소 거리를 갖는 학습 데이터의 클래스를 미지 패턴의 클래스로 선택한다.

그림 9는 클래스가 2개일 때 제안된 분류기(1NN)의 동작을 보여준다. 특징 a, b는 클래스 1을 특징 a', b'는 클래스 2를 나타내는 특징공간이다.  $d1$  과  $d2$  가 각각의 특징 공간에서 미지 패턴과 학습 패턴의 최소 거리라고 할 때, 미지 패턴의 클래스는 식 (17)과 같이 결정된다.

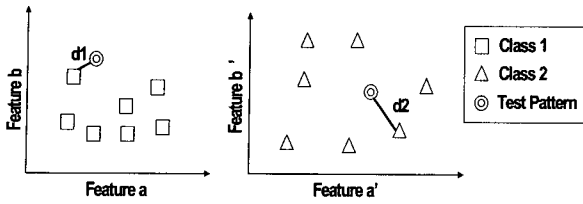


그림 9. 클래스가 2개일 때의 제안된 분류기의 동작  
Fig. 9. The proposed recognition system.

$$Class\ of\ test\ pattern = \begin{cases} class1 & d1 < d2 \\ class2 & d1 > d2 \end{cases} \quad (17)$$

IV. 실험

본 논문에서 제안된 분류 시스템을 검증하기 위하여 Concordia 대학의 numeral database를 사용하였다. 미국 전역에서 수집된 6000개의 손으로 쓴 숫자들로 구성되어 있는 database는 0.153mm의 64 × 224격자로 디지털화 되어있으며, 약 166PPI의 해상도를 갖는다. 분류기의 학습을 위해 사용된 3500개의 학습 데이터, 유전자 알고리즘의 평가를 위해 사용된 500개의 튜닝 데이터 그리고 알고리즘의 평가를 위한 나머지 2000개의 테스트 데이터로 데이터베이스를 나누었다. 그림 12는 숫자의 샘플을 보여준다. 각 숫자들은 모두 다른 필체를 가지고 있으며, 크기와 획의 굵기도 다름을 볼 수 있다.

기존의 방법인 3FW 알고리즘은 1NN 알고리즘에 비해 인식률이 높다. 하지만 제안된 방법들은 3FW 알고리즘보다도 정확도가 높음을 표 2에서 확인할 수 있다.

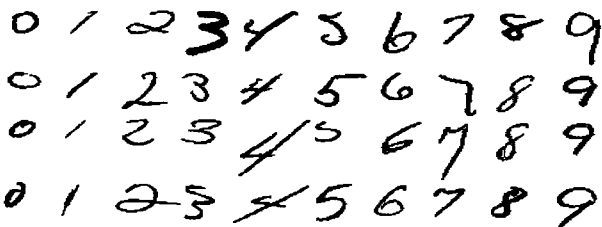


그림 12. 샘플 데이터  
Fig. 12. Sample data.

표 2. 제안된 방법의 인식 결과(%)  
Table 2. Results on The Proposed Method(%).

	1NN	3FW + 1NN	Adaptive 3FW + 1NN	Proposed method
Train	91.80	91.80	92.00	94.00
Test	89.70	90.30	91.25	93.90
feature	80	56	58	57

표 3. 제안한 알고리즘의 클래스별 정확도(%)  
Table 3. Confusion Matrix For The Proposed Method(%).

class	3FW + 1NN	Proposed Method
0	90.00	95.00
1	91.00	94.00
2	90.50	94.00
3	91.00	95.00
4	90.00	93.00
5	90.00	95.50
6	91.00	93.50
7	90.50	92.50
8	90.00	93.50
9	89.00	93.00
total	90.30	93.90

기존의 3FW알고리즘과 본 논문에서 제안한 알고리즘의 클래스별 정확도를 표 3과 같다. 제안된 알고리즘은 기존의 3FW알고리즘에 비해 거의 모든 클래스에서 인식이 잘되는 것을 확인할 수 있다.

V. 결론

본 논문에서는 유전자 알고리즘을 이용한 적응적 특징 가중치 방식과 클래스별로 적용된 KNN규칙을 결합한 새로운 패턴 인식 시스템을 제안하였다. 제안된 적응적 특징 가중치 알고리즘은 기존의 3FW 방법과는 달리 가중치의 중간값을 유전자 알고리즘으로 결정함으로써 과잉 맞춤을 피하면서, 데이터의 분포에 따라 적절한 특징의 가중치를 찾을 수 있는 장점을 갖는다. 이러한 이유로 적응적 특징 가중치 알고리즘은 3FW 알고리즘에 비해 좋은 성능을 보여 주었다. 또한, 지금까지의 유전자 알고리즘을 기반으로 한 FW 알고리즘들은 클래스에 따라 적응적으로 특징 공간을 찾을 수 있는 점을 간과하였기 때문에 학습 데이터의 모든 클래스를 가장 잘 표현할 수 있는 특징 공간을 찾는 데 주력하였다. 이것과는 다르게 본 논문에서는 모든 클래스가 아닌 각각의 클래스들을 가장 잘 표현하는 특징 공간들을 개별적으로 찾았다. 그리고 개별적으로 계산된 클래스별 특징 공간들을 이용해 KNN분류기는 클래스에 따라 특징공간을 변화시켜 미지 패턴의 클래스를 예측하였다.

향후 연구과제로 제안하는 알고리즘을 일반 문자나 한글 인식에 적용한다. 일반 문자나 한글을 인식하기 위해서는 숫자와는 다른 특징이 필요하기 때문에, 일반

문자나 한글의 특징을 추출에 대한 연구가 필요할 것이다.

### 참 고 문 헌

- [1] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *EEE Trans. on Pattern Anal. and Machine Intell.* vol. 19, pp. 153-158, Feb. 1997.
- [2] L. Davis, *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, 1991.
- [3] E. Gose, R. Johnsonbaugh, and S. Jost, *Pattern Recognition and Image Analysis*. Prentice Hall PTR, 1996.
- [4] W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large-scale feature selection," *Pattern Recognition Letter*, vol. 10, pp. 335-347, 1989.
- [5] J. D. Kelly and L. Davis, "Hybridizing the genetic algorithm and the k nearest neighbors classification algorithm," in *Proc. 4th Int. Conf. Genetic Algorithms, Appl.*, 1991, pp. 377-383
- [6] Kohavi, R. Langtry, and Y. Yun, "The utility of feature weighting in nearest neighbor algorithms," *European Conference on Machine Learning, ECML'97*. 1997.
- [7] W. K. Pratt, *Digital Image Processing*. New York: Wiley, 1978.
- [8] S. Lee, "Off-Line Recognition of Totally Unconstrained Handwritten Numerals Using Multilayer Cluster Neural Network," *IEEE Trans. on Pattern Anal. and Machine Intell.*, Vol. 18, pp. 648-652, June 1996.

저 자 소 개



이 희 성(학생회원)  
 2003년 연세대학교 기계전자  
 공학부 졸업(공학사)  
 2005년 연세대학교 전기전자공학부  
 석사과정 졸업(공학석사)  
 2005년~현재 연세대학교 전기  
 전자공학부 박사 과정

<주관심분야 : Computational intelligence, 로봇 비  
 전, 패턴 인식>



김 은 태(평생회원)  
 1992년 연세대학교 전자공학과  
 졸업(공학사, 전체수석)  
 1994년 연세대학교 전자공학과  
 석사과정 졸업(공학석사)  
 1999년 연세대학교 전자공학과  
 박사과정 졸업(공학박사)

1999년 3월~2002년 2월 국립한경대학교  
 제어계측공학과 조교수

2002년 3월~현재 연세대학교 전기전자공학부  
 조교수

2003년 University of Alberta, visiting researcher

1998년~현재 IEEE TFS, IEEE SMC, IEEE

CAS, FSS등에서 심의위원 활동 중.

2003년 대한전자공학회 해동상 수상.

<주관심분야 : Computational intelligence, 지능형  
 로봇>



김 동 연(정회원)  
 대한전자공학회 논문지  
 제35S권 5호 참조