

유전 알고리즘을 이용한 이진 트리 분류기의 설계와 냉연 흠 분류에의 적용

Design of a Binary Decision Tree using Genetic Algorithm for Recognition of the Defect Patterns of Cold Mill Strip

김 경 민, 이 병 진, 류 경, 박 귀 태
(Kyoung-Min Kim, Byung-Jin Lee, Kyoung Lyou, and Gwi-Tae Park)

Abstract : This paper suggests a method to recognize the various defect patterns of a cold mill strip using a binary decision tree automatically constructed by a Genetic Algorithm (GA). In classifying complex patterns with high similarity like the defect patterns of a cold mill strip, the selection of an optimal feature set and an appropriate recognizer is important to achieve high recognition rate. In this paper a GA is used to select a subset of the suitable features at each node in the binary decision tree. The feature subset with maximum fitness is chosen and the patterns are classified into two classes using a linear decision function. This process is repeated at each node until all the patterns are classified into individual classes. In this way, the classifier using the binary decision tree is constructed automatically. After constructing the binary decision tree, the final recognizer is accomplished by having neural network learning sets of standard patterns at each node. In this paper, the classifier using the binary decision tree is applied to the recognition of defect patterns of a cold mill strip, and the experimental results are given to demonstrate the usefulness of the proposed scheme.

Keywords : genetic algorithm, decision tree, feature selection, defect pattern, cold mill strip

I. 서론

냉연 강판의 표면 흠의 유무는 강판의 품질과 직결되므로 생산 라인에서 표면 흠을 신속히 검출하고 발생 원인을 파악하여 결함을 제거해 주는 과정은 생산 공정에서 중요한 부분을 차지한다. 이에 효과적인 냉연 흠 검출 방법이 지속적으로 연구되어 왔으며 현재 사용되고 있는 레이저 반사 신호를 이용한 1차원 검출 방식을 대체하기 위해, 최근 영상처리를 통한 냉연 표면 흠 인식 방법이 연구되고 있다[13].

영상 처리를 통한 기존의 흠 인식 방법은 CCD 카메라로부터 취득된 냉연 강판 영상으로부터 전처리 과정을 거쳐 흠 패턴을 검출하고 실험적으로 결정된 특징 량들을 흠 패턴으로부터 추출하여 이것을 신경회로망에 입력하여 한 단계에 인식하였다. 그러나 기존의 흠 인식 방법은 냉연 표면 흠의 특성을 고려할 때 다음과 같은 두 가지 문제점이 있다. 첫째, 냉연 강판의 표면 흠은 형태가 복잡하고 정형도가 낮아서 흠 인식이 특징량 선정에 민감하고, 또한 분리도가 우수한 개개의 특징량이라도 함께 사용하였을 때 서로 인식에 혼돈을 주는 요인으로 작용할 수 있으므로 실험적으로 특징들을 선정할 경우 결정된 특징들의 최적성을 보장할 수 없다. 둘째, 냉연 흠 패턴은 서로 다른 흠이지만 크기와 모양이 유사하여 같은 패턴으로 오인식 될 수 있는 몇 개의 유사 패턴 군이 존재 하기 때문에 패턴을

한 단계에 분류하면 분류 어려움이 높아진다는 단점이 있다.

이러한 문제점들을 해결하기 위하여 패턴을 한 단계에 분리하지 않고 계층적으로 분리를 진행해 나가는 이진 트리 구조의 인식 방법을 제안한다. 이 방법은 먼저 패턴들을 큰 두 개의 유사 패턴 집합으로 분류하고 분류된 각각을 다시 두 패턴 집합으로 나누는 과정을 반복하여 최종적으로 개개의 패턴으로 분류한다[8]. 이진 트리구조 분류기를 사용하면 각 분류 노드에서 패턴들을 유사한 패턴 집합 별로 분류하는데에 우수한 특징 부집합만을 선택적으로 사용함으로써 혼돈을 주는 특징을 배제하므로 유사한 패턴 그룹이 많은 경우 분류율이 높아진다. 이를 위해 각 단계에서 최적의 특징량을 선별해야 하는데 본 논문에서는 유전알고리즘을 이용한 선형 결정 함수와 특징선정 방법을 사용하여 이진 트리 구조 분류기의 각 노드에서 사용될 우수한 특징 부집합의 선정을 수행하였다.

본 논문은 먼저, 이진 결정 트리에 대해 설명하고 유전 알고리즘[4],[5],[6]을 이용한 선형 결정 함수와 유전 알고리즘을 이용한 특징 부집합 선정 방법에 대해 기술한 후 [2], 이들을 사용하여 이진 결정 트리를 구성하는 방법을 기술하고 이진 결정 트리 분류기를 이용하여 냉연 강판의 표면 흠 패턴을 분류한 실험 결과를 보인다[16].

II. 유전 알고리즘을 이용한 이진 결정 트리의 구성

1. 이진 결정 트리

입력 패턴을 구성하는 특징 전체나 또는 그 일부를 한 번만 사용하여 소속 부류를 결정하는 분류기는 일단계 분류기라고 한다[9],[11]. 그러나 이러한 일단계 분류기의 경우에 사용되는 특징들이 모든 부류에 대하여 최적의 변별

접수 일자 : 1999. 2. 1., 수정완료 : 1999. 9. 2.

김경민 : 여수대학교 전기공학과

류 경 : 동양공업전문대학 전기전자통신공학부 전자분야

이병진 : 고려대학교 전기공학과

박귀태 : 고려대학교 전기전자전파공학부, 서울대학교 ERC-ACI 연구위원

력을 가지는 것은 아니다 또한 일단계 분류기는 패턴을 분류하기 위하여 모든 클래스와 비교하는 과정이 필요하므로 인식기의 효율성이 저하되는 단점이 있다. 이러한 일단계 분류기의 문제점을 해결하기 위하여 두 종류 이상의 결정을 여러번 반복하여 입력 패턴의 부류를 결정하는 분류기를 다단계 분류기라 한다. 이러한 분류기를 '결정 트리 분류기'[10]라 하는데 결정 트리 분류기의 장점은 다음과 같다.

- i) 복잡한 전체 결정 영역을 보다 간단한 국부적인 결정 영역의 합집합으로 근사시킬 수 있다.
- ii) 입력 패턴을 모든 부류와 비교하는 일단계 분류기와 달리, 일부의 부류와만 비교를 하기 때문에 불필요한 비교 연산을 줄일 수 있다.
- iii) 트리의 내부 노드마다 변별력을 최대로 하는 서로 다른 최적 특징 집합을 선정할 수 있기 때문에 인식 능력은 향상시킬 수 있다.

결정 트리 분류기 중 한 노드에 대한 하부 노드의 개수가 '2'인 것을 '이진 결정 트리 분류기'라 한다. 이진 결정 트리 분류기는 모든 후보 특징량 중 각 노드에서 두 부류로 나누는데 가장 적합한 특징 부집합을 선택적으로 사용하여 분리해 나가며 이 과정을 각 패턴 부류가 각 노드에 독립적으로 나타날 때까지 반복한다.

2. 유전알고리즘을 이용한 특징선정

2.1 특징 선정 문제 (feature selection problem)

이진 결정 트리 분류기의 각 노드에서 분류에 꼭 필요한 특징만을 선정하여 사용하면 분류 정밀도가 높아져 분류기의 성능에 대한 신뢰도를 높일 수 있기 때문에 전체 특징들 중에서 각각의 특징 혹은 특징들의 조합의 유효성을 평가하고 그 중 유효한 한 특징 부집합을 선택하는 과정이 필요한데 이를 특징 부집합 선정 문제라 한다.

전체 특징 n 개 중에서 m 개의 원소를 가지는 최적의 특징 부집합을 선정하기 위해서는 m 개의 변수에 대한 모든 조합에 대해 부류간 분리도(class separability)를 평가해야 한다 그러나 n 개의 특징중 m 개를 선정하는 조합의 수는 nCm 으로서 표현되므로 n, m 이 그다지 크지 않을 경우에도 매우 큰 값이 된다. 이러한 소모적인 탐색을 피하는 방법에는 top-down, bottom-up, Branch and bound 탐색 방법 등 여러 가지 방법이 있는데 특징 선정 문제는 결국 최적화 문제라고 할 수 있으므로 본 연구에서는 전역적 최적화 기법인 유전 알고리즘을 이용하여 특징 선정을 수행하였다.

2.2 유전 알고리즘을 이용한 특징 선정

유전 알고리즘은 여러 해를 동시에 탐색하는 전역 탐색을 행함으로써 전역적인 최적해를 찾을 확률이 기존의 최적화 탐색에 비해 크므로 최적의 특징 부집합을 효과적으로 구할 수 있다[3],[7].

그림 1은 유전 알고리즘을 사용하여 최적의 특징 부집합을 구하는 과정을 블록도로 나타낸 것이다. 유전 알고리즘에서 각 개체는 특징 부집합을 나타내는데, 특징 부집합 성능 평가를 통하여 각 개체의 적합도가 계산되고 적합도 함수에 의해 개체의 생존 확률이 결정되며 이 확률에 따라

다음 세대인 진화된 개체, 즉 특징 부집합들이 생성된다. 이러한 진화 과정을 반복하여 최적의 특징 부집합을 얻는다. 여기서 특징 부집합들의 성능을 평가는 최적해를 얻는데 있어서 가장 중요한 과정인데 본 연구에서는 다음 절에서 설명할 유전 알고리즘을 이용한 선형 결정 함수의 결정 방법으로 각 특징 부집합의 특징 공간에서 결정 경계를 구한 후, 이것을 기준으로 패턴을 분류했을 때의 분류 에러로서 특징 부집합의 성능을 평가하였다.

전체 특징의 수가 n 일 때, 유전 알고리즘을 이용한 특징 부집합의 탐색을 위해서 길이 n -bit의 이진 문자열, $a = \{ a_1, \dots, a_n \}$ 을 생각한다. 이진 문자열 a 내에서 a_i 가 '1'이면 i -번째 특징은 전체 특징중 특징 부집합에 속하게 됨을 의미하고 '0'이면 반대로 제외됨을 의미한다. 이러한 가정하에서 유전 알고리즘을 이용한 특징 선정 알고리즘은 다음과 같다.

- i) 문자열 집단을 초기화시킨다.
- ii) 각 문자열을 특징 부집합으로 변환한다.
- iii) 구하여진 특징 부집합을 이용하여 샘플 패턴에 대한 분류 에러 혹은 분류 성능을 구하고 이를 적합도 함수에 반영시켜 적합도를 계산한다
- iv) 원하는 적합도에 도달한 개체가 존재하면 알고리즘 수행을 끝낸다.
- v) 각 개체에 대한 적합도를 기반으로 하여 유전알고리즘의 세 가지 연산인 복제, 교배, 돌연변이를 수행한다.
- vi) 미리 정한 최대 세대수에 도달하였으면 알고리즘 수행을 끝내고 전체 세대 중 가장 좋은 적합도를 가지는 문자열을 최종 결과로 취한다. 그렇지 않으면 ii)로 간다.

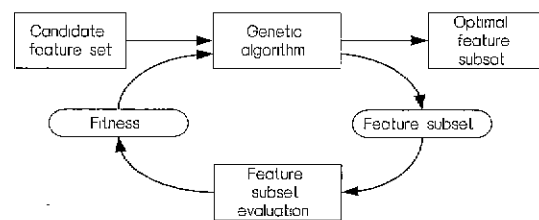


그림 1. 특징 선정 과정의 블록도.

Fig 1 Processing block diagram of feature election.

3. 선형 결정 함수를 이용한 특징 부집합 성능 평가

3.1 유전 알고리즘을 이용한 선형 결정 함수의 결정

유전 알고리즘을 이용하여 패턴 분류 에러를 최소화하는 선형 결정 함수를 결정하는데 있어서 다음과 같은 방법을 사용한다.

주어진 데이터 집합이 $X = \{ x_1, x_2 \dots x_N \} (x_k \in R^n, n$ 은 특징의 개수)라 하면 (1)과 같이 j -번째 특징의 값들을 모두 포함하는 구간의 하한, 상한인 $l(j)$ 와 $r(j)$ 를 정의할 수 있다.

$$\begin{aligned}
 l(j) &= \min_i x_{ij} \\
 r(j) &= \max_i x_{ij}
 \end{aligned} \tag{1}$$

이차원의 경우 j 는 $j=1, 2$ 의 값을 가지며 $l(j)$ 와 $r(j)$ 를 기초로 전체 데이터를 포함하는 최소의 사각형을 생각할 수 있다. 이제 이 사각형 범위 내에서 임의로 두 점을 선택하면 이 점들을 동시에 지나는 한 직선을 구할 수 있으며 이 직선의 기울기와 절편으로부터 2차원 공간의 선형 결정함수인 (2)의 w_1, w_2, w_3 를 구할 수 있다.

$$d(\mathbf{x}) = w_1x_1 + w_2x_2 + w_3 = 0 \quad (2)$$

이를 n 차원 공간으로 확장하면 임의로 n 개의 점을 선택하여 이들을 지나는 초평면을 구할 수 있으며 이로부터 n 차원 공간의 선형 결정함수인 (3)의 w_1, w_2, \dots, w_{n-1} 을 구할 수 있다.

$$d(\mathbf{x}) = w_1x_1 + w_2x_2 + \dots + w_nx_n + w_{n+1} = \mathbf{w}_0' \mathbf{x} + w_{n+1} \quad (3)$$

이 개념을 유전 알고리즘의 이진 문자열에 대응시키면 이진 문자열의 각 세그먼트는 n 차원 공간상의 각 축의 한 위치를 가리키며 n 개의 세그먼트는 한 점을 가리키게 된다. 즉, n 차원의 경우 n 개의 점을 선택해야 하므로, 필요한 세그먼트의 개수는 n^2 이 되며, 각 세그먼트의 길이를 m -bit라 하면 이진 문자열의 총길이는 n^2m -bit이다.

m -bit인 각 세그먼트중 j -번째 세그먼트를 디코딩하여 십진수로 변환한 값을 s_j 라 하면 (4)에 의해 s_j 가 가질 수 있는 최소값 '0'을 $l(j)$ 에 대응시키고 최대값 $(2^m - 1)$ 을 $r(j)$ 에 대응시켜 s_j 를 $[l(j), r(j)]$ 사이의 값 d_j 로 변환시키면 n^2m -bit로 이루어진 한 개체의 이진 문자열은 주어진 데이터를 모두 포함하는 최소 크기의 초입방체(hypercube) 내의 n 개의 점을 나타내게 된다 여기서 $j' = (\text{rem}(j-1, n+1) + 1)$ 이며, $\text{rem}(a, b)$ 는 a 를 b 로 나눈 나머지를 의미한다

$$d_j = \frac{r(j') - l(j')}{2^m - 1} \cdot s_j + l(j') \quad (4)$$

유전 알고리즘을 이용하여 결정 함수의 가중치를 구하는 전체 알고리즘은 다음과 같다

- i) 문자열 집단을 초기화시킨다.
- ii) 각 문자열을 임의의 좌표 점으로 변환한다.
- iii) 위에서 기술한 방법으로 좌표 점들이 이루는 초평면에 대응되는 가중치를 구한다.
- iv) 초평면에 대응되는 가중치를 구하여 결정된 결정 함수로부터 샘플 패턴에 대한 분류 에러를 구하고 적합도를 계산한다.
- v) 원하는 적합도에 도달한 개체가 존재하면 알고리즘 수행을 끝낸다
- vi) 각 개체에 대한 적합도를 기반으로 하여 유전알고리즘의 세 가지 연산인 복제, 교배, 돌연변이를 수행한다
- vii) 미리 정한 최대 세대수에 도달하였으면 알고리즘 수행을 끝내고 전체 세대 중 가장 좋은 적합도를 가지는 문자열을 최종 결과로 취한다. 그렇지 않으면 ii)로 간다.

3.2 특징 부집합의 성능 평가

위에서 기술한 선형 결정 함수는 임의의 특징 부집합이 형성하는 특징 공간에서 표준패턴들을 두 부류로 분류했을 때의 에러를 최소화하는 선형 결정 함수이다. 그러나 실제 입력 패턴은 두 종류의 패턴만 존재하는 것이 아니므로 분류 에러를 결정 경계에 의해 여러 패턴 클래스를 두 부류로 나누었을 때 서로 다른 영역에 존재하는 동일 클래스인 패턴의 개수로 정의한다. 유전 알고리즘은 주어진 특징 공간에서 정의된 분류 에러를 최소화하는 결정 함수를 찾아 내게 되는데 특징 조합에 따라 최소화된 에러는 달라지게 된다. 그러므로 최소화된 분류 에러가 적게 나타나는 특징 조합 개체는 높은 적합도를, 그리고 분류에러가 큰 특징 조합 개체는 낮은 적합도를 가지도록 적합도 함수를 구성한다.

4 이진 결정 트리의 구성과 인식기의 완성

이진 결정 트리의 구성은 표준 패턴에 대해 트리의 각 노드에서 앞서 기술한 방법을 이용하여 두 패턴 그룹으로 나누는데 분류 에러를 최소화하는 특징 부집합을 선정하고, 이것으로 패턴을 두 그룹으로 분류하는 과정을 모든 노드에 대해 실행하여 각 최종 노드에 모든 패턴 부류가 독립적으로 나타날 때 까지 반복함으로써 이루어진다.

전체적으로는 두 부류가 아닌 다 부류에 대해 이진 결정 트리를 구성하는 것이므로 한 노드에서 나누어진 두 군집들은 어느 한쪽으로 치우침 없이 비슷한 수의 부류 혹은 패턴을 가지고 있는 고른 분포를 유지하는 것이 좋다[12]. 이를 정량화시켜 균형 계수를 부류들의 평균수와 생성된 새로운 군집에서의 부류들의 수의 편차로서 (5)와 같이 정의한다. 나누어진 두 군집내의 패턴의 수가 비슷하여, 균형 계수가 작을수록 부류 군집들 간의 최적의 균형 관계를 유지할 수 있고 전체적으로 패턴 인식을 위한 매칭 횟수가 줄어들게 된다[14],[15]. 이 균형 계수의 최소값은 '0', 최대값은 이진 트리의 경우 $\sqrt{2}$ 가 된다.

$$balance = \sqrt{\frac{\sum_{j=1}^h (N_j - \frac{N}{h})^2}{(\frac{N}{h})^2}} \quad (5)$$

(5) h 는 노드의 수, N 은 입력 패턴의 수, N_j 는 j 번째 노드에 속하는 패턴의 수이다. 본 논문에서는 이진 트리를 구성하므로 h 는 '2'가 된다

균형계수를 포함한 적합도 함수는 (6)과 같다.

$$fitness = \frac{1}{1 + w_e \cdot error + w_b \cdot balance} \quad (6)$$

(6)에서 error는 분류 에러, balance는 군집간의 균형 계수를 의미한다. 또한 w_e, w_b 는 각각의 파라미터에 가중을 주기 위한 가중치이다. 만일 한 노드에서 분류 에러와 균형 계수가 모두 '0'이면 적합도는 최대값 '1'을 가지게 된다. 한편 가중치 w_e, w_b 를 조절함에 의해 트리 구성의 결과가 달라질 수 있다. 예로서 w_b 에 큰 값을 준다면 분류 에러를 좀 희생하더라도 군집간의 균형이 좋은 트리

구조를 얻을 수 있는 확률이 크게 된다.

이진 결정 트리가 구성되면 각 노드에서 정해진 최적의 특징 부집합으로 역전과 신경망을 사용하여 입력된 표준 패턴을 두 그룹으로 분리하도록 학습시키는 과정을 모든 노드에 대해 수행함으로써 최종적인 이진 트리 구조의 인식을 완성한다.

III. 이진 트리 분류기를 이용한 냉연 표면 흠 분류

1. 냉연 표면 흠 패턴에 대한 특징 추출

냉연 표면 흠의 종류는 형태에 따라 dull, oil-drop, slip, dent, scale, dirt, scratch 로서 7가지 정도로 나눌 수 있다. 이들에 대해 전처리를 수행하여 아래와 같은 6가지 후보 특징량을 추출하였다.

1.1 기하학적 특징

본 연구에서는 패턴의 크기와 방향에 무관한 기하학적 특징인 패턴의 면적비와 응집도, 그리고 패턴의 면적을 이용한 특징 후보로 선정하였다.

- ① def_area : 대상체의 면적 (흠 패턴의 모든 화소의 개수)
- ② area_rat : def_area와 box_area의 비 (area_ratio = def_area/box_area) 여기서 box_area는 흠을 내부에 포함할 수 있는 최소 사각형의 면적이다.
- ③ compactness : 대상체의 응집도((4π*area) / perimeter²) 여기서 perimeter는 대상체의 윤곽선의 길이이다.

1.2 모멘트를 이용한 특징

확률론에서 정의된 모멘트개념은 영상 처리시 대상체의 형태에 대한 특징량을 구하는 데 매우 실용적이고 유용한 방법으로 사용되고 있다[1]. 모멘트 특징량에서 표면 흠 검사에 유용한 정보는 흠의 장단축비, 장축의 길이 및 퍼짐정도이다. 모멘트로부터 구해지는 특징량은 영상내의 흠의 형태에 대한 중심 모멘트를 사용함으로써 (7)~(11)과 같이 구할 수 있다.

- ① 흠의 장축과 단축의 길이정보. 흠의 방향정보 제공

$$\mu_u = \sum_x \sum_y (x - \bar{x})^u (y - \bar{y})^v f(x, y) \quad (7)$$

$$a = 2\sqrt{2} \sqrt{(\mu_{20} + \mu_{02} + \sqrt{((\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2)})} : \text{장축} \quad (8)$$

$$b = 2\sqrt{2} \sqrt{(\mu_{20} + \mu_{02} - \sqrt{((\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2)})} : \text{단축} \quad (9)$$

여기서 $f(x, y)$ 는 영상의 계조치이며 μ_u 는 대상체의 중심 모멘트(central moment)를 나타낸다.

- ② 흠에 대한 장축과 단축의 길이의 비

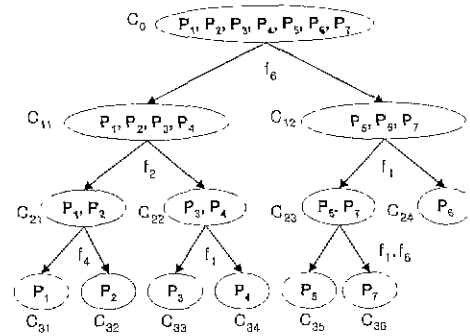
$$\text{axis_ratio} = \frac{b}{a} \quad (10)$$

- ③ Spreadness : 흠의 퍼짐 정도에 대한 특징량

$$\text{spread} = \frac{\mu_{02} + \mu_{20}}{\mu_{00}^2} \quad (11)$$

2. 이진 트리 구조의 인식기 구성

흠 패턴들의 표준 패턴에서 추출한 특징 벡터들이다. 유전 알고리즘을 이용하여 이진트리를 구성하는데 본 연구에서는 (6)의 가중치 w_o 와 w_b 을 모두 1로 설정하였다. 표준 패턴에 대해서 유전 알고리즘을 이용한 선형 결정 함수와 특징 선정 방법을 사용하여 구성한 이진 트리를 그림 2에 나타내었고 표 1에 각 노드에서의 분류에러, 균형계수, 특징 부집합, 그리고 적합도 값을 나타내었다. 표 1에서 Error는 표준 패턴들을 선형 분리를 사용하여 두 부류로 나누었을 때 각 흠 패턴의 군집에서 이탈된 패턴의 개수를 나타낸다. 표 2에서는 각 노드에서 결정된 선형 결정 함수의 가중치값을 보였다



기호	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇
흠패턴	dull	oil	slip	dent	scale	dirt	scratch

기호	f ₁	f ₂	f ₃	f ₄	f ₅	f ₆
특징량	def_area	area_ratio	compactness	axis_ratio	spread	long_axis

그림 2. 표준패턴을 사용하여 구해진 이진 결정 트리.

Fig. 2. Binary decision tree constructed with standard patterns.

그림 2에서 P_i는 패턴의 종류를 나타내고 f_k는 특징량을, 그리고 C_{mm}는 각 노드에서의 패턴 클래스를 나타낸 것이다.

위에서 구성된 이진 트리 각각의 노드에서 선정된 특징 부집합을 사용하여 학습한 신경망으로 각 노드의 분류기를 완성하였다. 신경망의 입력층은 선정된 특징의 개수에 따라 노드의 개수를 정하고, 은닉층의 노드를 10개로 하였으며 출력층은 2개로서 이진 결정을 나타 내었다. target은 bipolar([1 -1] or [-1 1])를 사용하여 표현하였다.

표 3에서 위와 같은 방법으로 구성된 이진 트리구조의 분류기를 사용하여 냉연 강판 흠 패턴을 인식할 결과를 보였다.

표 3에 나타난 7종류의 흠 인식 결과를 보면 Dent와 Slip의 인식율이 매우 낮다. 이는 표 1에서 제시된 바와 같이 이들 흠에 대한 이진 트리를 구성할 때의 선형 분류에러는 노드 C₀, C₁₁, C₂₂에서 모두 0이었는데 그것은 Dent와 Slip의 표준 패턴은 선정된 특징으로서 선형 분리되었다는 것을 의미한다. 이 결과로 알 수 있는 것은 특징

표 1. 각 노드에서의 패턴들과 적합도.

Table 1. The patterns and fitness at each node.

Node	Pattern	Error수	Feature	Fitness
C_0	$P_1 P_2 P_3 P_4 P_5 P_6 P_7$	0	f_6	0.7180
C_{11}	$P_1 P_2 P_3 P_4$	0	f_2	0.5677
C_{12}	$P_5 P_6 P_7$	2	f_4	0.2550
C_{21}	$P_1 P_2$	0	f_4	0.5858
C_{22}	$P_3 P_4$	0	f_1	0.7795
C_{23}	$P_5 P_7$	3	$f_1 f_6$	0.1874

표 2. 각 노드에서 구한 결정 함수.

Table 2. Decision function at each node.

$d(x)$	w_1	w_2	w_4	w_3	w_5	w_6	w_7
$d_0(x)$	0.000	0.000	0.000	0.000	0.000	-1.765	1.000
$d_{11}(x)$	0.000	-2.326	0.000	0.000	0.000	0.000	1.000
$d_{12}(x)$	0.000	0.000	0.000	-1.853	0.000	0.000	1.000
$d_{21}(x)$	-1.713	0.000	0.000	0.000	0.000	0.000	1.000
$d_{22}(x)$	0.000	0.000	0.000	-2.304	0.000	0.000	1.000
$d_{23}(x)$	-4.062	0.000	0.000	0.000	0.000	1.5902	1.000

$$d(x) = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + w_6x_6 + w_7 = 0$$

표 3. 각 흠에 대한 인식율.

Table 3. Recognition rate of each defect pattern.

실험패턴	인식률	
	인식 개수 / 패턴 수	인식률(%)
Dent	0 / 23	0 %
Dull	14 / 28	50 %
Oil_drop	18 / 20	90 %
Slip	6 / 20	30 %
Dirt	21 / 22	95 %
Scale	27 / 34	79 %
Scratch	24 / 28	86 %
전 체	110 / 175	63 %

선정시 표준 패턴에 적합한 최소의 특징만을 선정하기 때문에 이진 트리 구성의 입력으로 사용하는 표준 패턴의 개수가 많지 않으면 노이즈에 민감해지는 특성을 나타낸다는 것이다. 즉, 학습 데이터의 개수에 따라 노이즈에 민감한 정도가 달라진다고 할 수 있다

IV. 결론

본 연구에서는 냉연 흠 패턴 분류에 이진 트리 구조를

이용해 보았다. 이진 트리 구조의 각 노드에서 특징을 선정하는데에 최적화 기법중의 하나인 유전알고리즘을 이용하였고 최적 조건을 판별하는데 역시 유전알고리즘으로 결정한 선형 결정 함수를 사용하였다. 이 방법은 주어진 패턴 군집에 대하여 특징 선정과 이진 결정 트리 구성을 자동적으로 수행한다는 장점이 있다 또 하나의 장점은 유전 알고리즘 내의 적합도 함수를 적절히 설정함에 따라 분류 에러뿐 아니라 분할된 군집의 균형 및 사용된 특징의 개수까지도 고려한 결과를 얻을 수 있다는 것이다.

본 연구에서 각 노드에 대한 신경 회로망을 학습하여 분류기를 설계하는 대신, 이미 구해져 있는 선형결정함수를 직접 사용하여 인식기를 구성할 때 특징이 하나이면 선형 분리가 되어서 의미가 없고 특징이 많아지면 각 노드에서 10% 이상 인식률이 떨어지는 결과를 얻었습니다.

이진 트리 구조 분류기와 한 단계 인식기[17]를 사용할 때, 특징의 수가 많으면 일반적으로 인식률이 높아지며 88% 정도 되었다. 단점으로는 모든 특징에 대해서 처리하므로 시간이 많이 소요되며 선택한 특징이 변별력이 있는 특징인지 알 수 없다는 것이다.

이진 트리 구조 분류기의 성능을 좌우하는 가장 중요한 단계는 각 노드에 대한 특징 선정의 문제라고 할 수 있다. 본 연구에서 사용한 특징 선정방법은 특징의 차수가 n 인 경우 한 개체의 세그먼트 개수가 n^2 이 되어 특징의 개수가 늘어나면 유전 알고리즘의 계산 시간이 많이 늘어나는 단점이 있다. 그리고 특징 공간상에 패턴의 특징 벡터의 분포가 산만하여 선형 분리가 가능하지 않은 경우 선형 결정함수를 이용하여 선정한 특징들에 대해 신뢰할 수 없게 된다.

앞으로 이러한 문제점을 효과적으로 해결하기 위한 방안의 연구와 더불어 이진 결정 트리 구성시 비선형 결정 함수를 이용하는 방법에 관한 연구로 이진 트리 구조의 패턴 인식기의 성능을 향상시키는 것이 향후 연구 과제로 남아 있다.

참고문헌

- [1] M. D Levine, "Feature extraction : a survey", *Proc of the IEEE*, pp 1391-1407, 1969
- [2] A. K. Jain and R. Dubes, "Feature definition in pattern recognition with small sample size". *Pattern Recog.*, vol 10, pp. 85-97, 1978
- [3] L. Yao, "Nonparametric learning of decision regions via the genetic algorithm," *IEEE Trans. Sys., Man, Cybern.*, vol. 26, pp. 313-321, Apr. 1996
- [4] L. Chambers, *Practical Handbook of Genetic Algorithms*, CRC Press, 1995
- [5] L. Yao and W. A. Sethares, "Nonlinear parameter estimation via the genetic algorithm," *IEEE Trans. Signal Proc.* vol. 42, pp. 927-935, Apr. 1994.
- [6] J. J. Grefenstette, "Optimization of control parameters for genetic algorithms," *IEEE Trans. Sys.*

Man, Cybern., vol. 16, pp. 122-128, Jan./Feb. 1986.

[7] J. K. Muir and K. S. Fu, "Automated classification of nucleated blood cells using a binary tree classifier," *IEEE Trans. Patt. Anal. Mach. Intl.*, vol. 5, pp. 429-443, Sep. 1980.

[8] S. R. Safavian and D. Landgrebe. "A survey of decision tree classifier methodology," *IEEE Trans. Sys., Man, Cybern.*, vol. 21, pp. 660-674, May/June. 1991.

[9] H. J. Payne and W. S. Meisel, "An algorithm for constructing optimal binary decision trees," *IEEE Trans. Computers*, vol. 26, pp. 905-916, Sep. 1977.

[10] G. H. Landeweerd, T. Timmers, and E. S. Gelsema, "Binary tree versus single level tree classification of white blood cells," *Pattern Recognition*, vol. 16, pp. 571-577, 1983.

[11] P. H. Swain and H. Hauska, "The decision tree classifier design and potential," *IEEE Trans. Geosci. Elec.*, vol. 15, pp. 142-147, Jul. 1977.

[12] C. Y. Suen and W. R. Wang. "ISOETRP : An interactive clustering algorithm with new object," *Pattern Recognition*, vol. 7, pp. 211-219, 1984.

[13] G. P. Babu and M. N. Murty, "Clustering with evolution strategies," *Pattern Recognition*, vol. 27, pp. 321-329, 1994.

[14] S.Z. Selim and K. Alsultan, "A simulated annealing algorithm for the clustering problem," *Pattern Recognition*, vol. 24, pp. 1003-1008, 1991.

[15] 정순원, 박귀태, "유전 알고리즘을 이용한 이진 결정 트리의 설계와 응용." 전자공학회 논문지, 제33권, B 편, 6호, pp. 1122-1130, 1996.

[16] B. J. Lee, K. Lyoo, G. T. Park, K. M. Kim, "Design of a binary decision tree for recognition of the defect patterns of cold mill strip using genetic algorithm", *Proc. of the Third Asian Fuzzy Systems Symposium*, pp. 208-212, Korea, June 1998.

[17] 김경민, 박귀태, 박중조, 외, "냉연 표면흠 검사 알고리즘의 개발," 제어·자동화·시스템공학회, 제3권, 제2호, pp. 179-186, 1997, 4.



김 경 민

1988년 고려대 전기공학과 졸업. 동대학원 석사(1991), 동대학원 박사(1996). 1997-현재 국립 여수대학교 전임강사. 관심분야는 컴퓨터 비전, 퍼지 및 신경회로망, 스테레오 비전 등임.



이 병 진

1996년 고려대 전기공학과 졸업. 동대학원 석사(1998). 1998년~현재 전기전자전파공학부 박사과정. 관심분야는 신호처리 및 컴퓨터비전, 유전자 및 신경회로망 응용.



류 경

1988, 1990, 1999 고려대 전기공학과 학사 석사 박사. 1999년~현재 동양공업전문대학 전기전자통신공학부 조교수. 관심분야는 컴퓨터비전, 영상처리, 유전자 및 신경회로망, 시스템 설계.



박 귀 태

1975년, 1977년, 1981년 고려대학교 전기공학과 학사, 석사, 박사학위. 1975년-1978년 한국원자력연구소 연구원. 1978년-1981년 광운대학교 전기공학과 조교수 1981년~현재 고려대학교 전기·전자·전파공학부 교수. 주요 관심분야는 컴퓨터비전, 퍼지 및 신경망이론 및 지능제어 등임.