

## 하드 디스크 결함 분포의 패턴 인식에 관한 연구

이재두, 문운철, 이승철  
중앙대학교 전자전기공학부

### A Study on Pattern Recognition of Hard Disk Defect Distribution

Jae-Du Lee, Un-Chul Moon, Seung-Chul Lee  
Dept. of Electrical and Electronics Engineering, Chung-Ang Univ.

**Abstract** - 본 논문에서는 다층 퍼셉트론(Multi-Layer Perceptron)을 이용한 하드 디스크 결함 분포의 패턴 인식 기법을 제시한다. 결함 분포로부터 5 가지의 특징들을 추출하고, 이를 이용하여 퍼셉트론의 입력을 구성하였으며, 미리 분류된 표준 패턴 클래스를 이용하여 퍼셉트론의 출력을 구성하였다. 테스트 결과, 제시된 신경망은 하드 디스크의 패턴 분류에 만족할 만한 성능을 나타내었다.

여기서,  $d_i$ 는  $i$  번째 결함,  $r_i$  와  $\theta_i$  는 각각  $i$  번째 결함의 거리 및 각도이다. 이 때, 각도는 HDD의 3시 방향을 기준으로 설정한 후, 반시계 방향으로 각도가 증가하는 것으로 설정하였다.

#### 1. 서 론

패턴 인식 (Pattern Recognition)은, 미지의 입력 패턴을 여러 클래스 중 적절한 하나의 클래스로 대응시키는 정보처리 과정으로 정의될 수 있다[1-3]. 인공신경망 (Artificial Neural Network)의 일종인 다층 퍼셉트론 (Multi Layer Perceptron)은 오차역전파 (Error Back Propagation) 학습 알고리즘의 개발과 더불어, 일반적인 함수의 학습에 효율적으로 적용될 수 있음이 알려졌다. 이와 같은 학습 능력을 기반으로, 퍼셉트론은 여러 분야의 교사 분류 문제에 성공적으로 적용되었다 [4-7].

본 연구에서는 컴퓨터의 기억장치인 HDD (Hard Disk Drive) 결함의 패턴 분류에 다층 퍼셉트론을 적용한 결과를 제시한다. HDD는 제조 공정의 마지막 단계에서 HDD의 읽기/쓰기/찾기 등의 검사를 수행하며 정상적인 읽기/쓰기/찾기가 수행되지 않는 섹터는 결함으로 처리되고, 결함의 개수가 주어진 임계치를 초과하게 되면, 그 HDD는 불량품으로 처리된다. 불량 HDD에 대한 진단과 처리는 수리공정에서 경험이 축적된 수리사가 수작업으로 진단하여 불량 종류를 판정하고, 이에 근거하여 해당 HDD의 제작업 및 부품 교체가 이루어지게 된다. 따라서, HDD 결함 분포의 분류는 불량 종류를 나누는 중요한 정보를 제공하며, 정확한 수리판정 여부는 생산 비용에 중요한 영향을 미치게 된다.

결함 패턴으로부터 패턴 분류의 근거가 되는 5가지의 특징을 선정 한 후, 퍼셉트론 신경망의 교사 학습(Supervised Learning)을 위하여, 5개의 특징 값들을 입력으로, 그리고 수리사에 의해서 미리 분류된 클래스 종류를 출력 값으로 갖는 학습 데이터를 구성하여 다층 퍼셉트론 신경망의 오차 역전파 학습에 사용하였다. 제시된 알고리즘은 실제 제조 공정에서 발생한 불량 HDD의 결함 분포를 대상으로 평가되었고, 만족할 만한 패턴 분류 성능을 확인하였다.

#### 2. 본 론

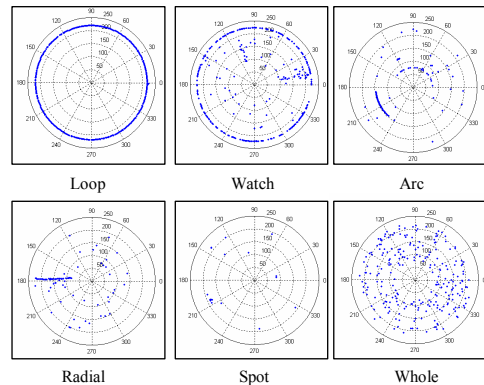
##### 2.1 결함 분포의 표준 패턴 클래스

디스크는 표면상에 동심원들로 구성된 실린더(Cylinder)와 각 실린더 내의 물리적인 최소 저장 단위인 섹터(Sector) 들로 구성되어 있다. 고속으로 회전하는 디스크와 디스크의 중심에 수직인 방향으로 왕복운동을 하는 HAS (Head Suspension Assembly)의 제어에 의해 고밀도로 집적된 디스크 섹터에 접근하여 데이터의 입출력이 수행된다. 이러한 회전체로서의 특성 때문에 HDD 결함 분포는 전반적으로 디스크 중심에 대칭적인 형태를 띠고 있다.

##### 2.2 전처리 및 5 가지 패턴 특징

먼저, 결함 분포의 전처리(Pre-processing)를 통하여 결함 분포 데이터 과일을 작성하였다. 디스크를 극좌표(Polar Coordinate)상으로 나타낸 후, 반지름은 150 구간으로, 각도는 1°씩 360 구간으로 각각 이산화시켜서 구역화하였다. 각 구역에 기준 이상의 결함들이 존재하는 경우, 그 구역에는 결함이 존재하는 것으로 처리하였다. 이와 같은 전처리를 거치면, 일반적으로 500개 이하의 결함들이 다음과 같은 형태로 나타내어진다.

$$d_i = (r_i, \theta_i) \tag{1}$$



〈그림 1〉 전처리된 표준 패턴 클래스

그림 1은 전처리된 표준 패턴의 전형적인 데이터이다. 그 후, 다음과 같은 다섯 가지의 특징이 선정되었다.

##### 2.2.1 필터링후 결함의 개수

본 연구에서는 산발적인 결함들을 제거하기 위하여 극좌표를 직교좌표(Cartesian Coordinates)로 변화하였다. 그 후, 종축과 횡축이 각각 5칸의 크기를 갖는 5\*5 정사각형 형태의 이동창(Moving Window)을 이용하여, 결함 개수가 3개 이하인 이동창의 결함은 제거하는 필터링(filtering) 방식을 사용하였다. 첫 번째 특징은, 필터링된 패턴 데이터의 결함 개수  $n$ 으로 선정하였다.

##### 2.2.2 필터링 전후의 결함 개수 차이

두 번째 특징으로는, 필터링 전후의 결함 개수 차이를 선정하였다.

##### 2.2.3 r 성분의 표준편차

이는 반지름 성분의 관점에서 결함들이 얼마나 분포되어 있는지를 의미하며, 다음과 같이 일반적인 표준편차로 계산된다.

$$\sigma_r = \sqrt{\sum_{i=1}^n (r_i - m_r)^2 / n} \tag{2}$$

여기서,  $n$ 은 필터링 후 결함의 총 개수,  $m_r$ 은  $r_i$ 의 평균값이다.

##### 2.2.4 $\theta$ 점유도

이는  $\theta$  성분으로 계산한 결함의 점유도이다. 취득된 데이터의 각도는 0° ~ 360° 사이의 정수 단위로 이산화 되었기 때문에 전체 360°에서 결함을 포함한 각도를 다음과 같이 계산하였다.

$$\theta \text{ 점유도} = (\text{결함을 포함한 각도} / 360) * 100 \tag{3}$$

##### 2.2.5 $\theta$ 성분의 분포도

이는 각도의 관점에서 결함들이 얼마나 퍼져있어 있는 지를 나타내는 정도이다. 루프형이나 분침형의 경우에 전 각도에 걸쳐서 결함들이 분포

되어 있으며, 찍힘형이나 방사형의 경우에는 좁은 각도에 집중되어 있다는 사실에 근거하여 매우 유용한 특징이 될 수 있다.

하지만, 극좌표계의 특성상 0°와 359°는 매우 가까운 각도임에도 불구하고 데이터 상에서는 큰 차이를 나타내는 것으로 처리되게 되므로, 이러한 문제를 해결하기 위해서, 결합의 각도를  $m_\theta$ 를 기준으로 새롭게 정의하여  $\sigma_\theta$ 를 구하였다. 즉, 물리적으로  $m_\theta$ 로부터 가장 멀리 떨어진 각도는 대칭방향인 ( $m_\theta \pm 180^\circ$ )이므로, 모든 결합의  $\theta$ 가 평균  $m_\theta$ 를 기준으로  $180^\circ \sim +180^\circ$ 로 사상(Mapping)되도록 다음과 같이 새롭게 정의하였다.

$$\theta' = \begin{cases} \theta - m_\theta, & \text{if } -180^\circ \leq (\theta - m_\theta) \leq 180^\circ \\ \theta - m_\theta - 360^\circ, & \text{if } 180^\circ < (\theta - m_\theta) \\ \theta - m_\theta + 360^\circ, & \text{if } (\theta - m_\theta) < -180^\circ \end{cases} \quad (4)$$

(4)와 같이 재정의된 각도를 이용하여 표준편차를 구하면,

$$\sigma_\theta = \sqrt{\sum_{i=1}^n (\theta'_i)^2 / n} \quad (5)$$

과 같이 표현되며, 이는 “평균에서 분포된 정도”라는 표준편차의 물리적인 의미에 부합하게 된다.

또한, 극좌표계의 특성을 고려한 평균을 구하기 위해서, 입력 데이터를 직접 이용하여 구한 평균  $m_{\theta 1}$ , 그리고 9시 방향을 극좌표계의 새로운 기준 축으로 변환하여 구한 평균  $m_{\theta 2}$ 를 각각 계산하였다.

그 후, 두 개의 기준 축에 의해서 각각 구해진 평균  $m_{\theta 1}$ 과  $m_{\theta 2}$ 를 기반으로 (5)를 이용하여 각각  $\sigma_{\theta 1}$ 과  $\sigma_{\theta 2}$ 를 구하였다. 최종적으로는, 두  $\sigma_{\theta 1}$ 과  $\sigma_{\theta 2}$  중 작은 값을 각도 성분의 분포도로 정의하였다.

$$\theta \text{ 성분의 분포도} = \min(\sigma_{\theta 1}, \sigma_{\theta 2}) \quad (6)$$

### 3. 퍼셉트론을 이용한 패턴 분류

본 연구에서는, 앞장에서 기술된 5가지의 특징들을 기반으로 각 패턴을 구분하는 다층 퍼셉트론을 구성하였다. 이를 위하여, 5가지 특징 값들을 입력 받아, 두 개의 은닉층을 거쳐서, 6개의 출력 노드를 갖는 다층 퍼셉트론을 구성하였다. 즉, 5가지의 특징 값 ( $x_1, x_2, x_3, x_4, x_5$ )를 입력 받아, 15개와 8개의 은닉층 노드를 통과한 후, 6개의 노드 값 ( $p_1, p_2, p_3, p_4, p_5, p_6$ )을 출력하도록 정의하였다.

이와 같은 구조에서, 패턴의 클래스에 따른 출력 값을 (7)과 같이 정의하여 학습에 사용하였다.

$$\begin{cases} \text{Loop} : \text{Pattern1} : [100000] \\ \text{Watch} : \text{Pattern2} : [010000] \\ \text{Whole} : \text{Pattern3} : [001000] \\ \text{Radial} : \text{Pattern4} : [000100] \\ \text{Spot} : \text{Pattern5} : [000010] \\ \text{Arc} : \text{Pattern6} : [000001] \end{cases} \quad (7)$$

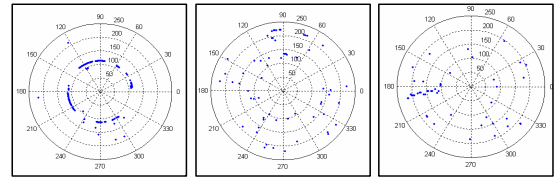
실제 학습을 위하여, 각 패턴별로 수리사가 미리 분류한 100개씩의 샘플을 수집하여 총 600개의 학습 데이터를 선정하였다. 퍼셉트론의 학습을 위해서는 Matlab에서 제공하는 toolbox의 LM (Levenberg Marquardt) 오차 역전파 함수를 이용하였다.

새롭게 입력된 분류 대상 패턴에 대해서, 먼저 5개의 특징값 ( $x_1, x_2, x_3, x_4, x_5$ )이 계산되고, 이는 학습된 퍼셉트론 신경망의 5개의 입력노드에 각각 입력되어, 6개의 출력 값이 계산된다. 이 때, 최종적인 패턴은 6개의 노드중 최대값을 갖는 노드로 결정되게 된다. 즉, 출력값이 최대로 나타나는 패턴을 최종적인 클래스로 결정하게 된다.

### 4. 실험 결과

학습된 퍼셉트론은 실제 제조공정에서 발생한 불량 HDD의 결합 분류에 적용되었다. 그림 2는 세 개의 예제 패턴을 나타내며, 표 1은 이러한 세 개의 패턴이 제시된 알고리즘에 의하여 분류되는 과정을 나타낸 것이다. 구현된 결과를 확인하기 위하여, 학습에 사용되지 않은 250개의 결합 분포에 대해서 실험한 결과, 수리사의 판별과 238개가 일치하여

95.2%의 적중률을 나타내었다. 이 때, 일치하지 않은 12개 패턴은 대부분 실제 수리사의 판단으로도 적절하게 나누어지지 않는 불분명한 패턴으로 분석되어, 패턴 분류의 결과는 신뢰할 만한 것으로 입증되었다.



<그림 2> 세 개의 예제 패턴

<표 1> 예제 패턴 분류 과정

		Pattern 1	Pattern 2	Pattern 3
특징	$x_1$	124	14	11
	$x_2$	15	49	42
	$x_3$	3.4109	42.7240	31.4174
	$x_4$	84.3144	10.5473	4.9074
	$x_5$	32.22	3.889	1.667
신경망출력	$p_1$	0.0001	0.0000	0.0002
	$p_2$	0.9440	0.0010	0.0032
	$p_3$	0.0928	0.3962	0.2997
	$p_4$	0.0040	0.2180	0.9441
	$p_5$	0.0092	0.0329	0.5130
	$p_6$	0.9171	0.0185	0.0021
결과		watch	whole	radial

### 5. 결론

본 연구에서는 다층 퍼셉트론 (Multi-Layer Perceptron)을 하드 디스크 드라이브(Hard Disk Drive, HDD) 결합 분포의 패턴 분류에 적용한 연구 결과를 제시하였다. 실험 결과, 만족스러운 분류 성능을 확인하였으며, 본 연구 결과는 수리업무에 기초한 수리관정 작업의 자동화에 유용하게 응용될 수 있을 것으로 기대된다.

#### 감사의 글

본 연구는 산업자원부의 지원에 의하여 기초전력연구원 (R-2005-B-103) 주관으로 수행된 과제임.

#### [참고 문헌]

- [1] S. Theodoridis and K. Koutroubas, *Pattern Recognition*, Academic Press, 1999.
- [2] R. Xu and D. Wunsch II, “Survey of Clustering Algorithm”, *IEEE Transactions on Neural Networks*, Vol. 16, No. 3, pp.645-678, May 2005.
- [3] George Nagy, “State of the Art in Pattern Recognition”, *Proceedings of the IEEE*, Vol. 56, No. 5, May 1968.
- [4] S. K. Pal and S. Mitra, “Multilayer Perceptron, Fuzzy Sets, and Classification”, *IEEE Transactions on Neural Networks*, Vol. 3 No. 5 pp. 683-697, Sep. 1990.
- [5] O. J. Murphy, “Nearest Neighbor Pattern Classification Perceptron”, *Proceedings of IEEE*, Vol. 78, No. 10, pp. 1595-1598, Sep. 1990.
- [6] A. M. Sharf and T. T. Lie, “ANN Based Pattern Classification of Synchronous Generator Stability and Loss of Excitation”, *IEEE Transactions on Energy Conversion*, Vol. 9, No. 4, Dec. 1994.
- [7] T. Auld, A. W. Moore and S. F. Gull, “Bayesian Neural Networks for Internet Traffic Classification”, *IEEE Transactions on Neural Networks*, Vol. 18, No. 1, pp.223-239, Jan. 2007.