

HDD (Hard Disk Drive) 결함 분포의 패턴 분류에 관한 연구

권 현 태, 문 운 철, 이 승 철
 중앙대학교 전자전기공학부

A Study on a Pattern Classification of HDD (Hard Disk Drive) Defect Distribution

Hyun-Tae Kwon, Un-Chul Moon, Seung-Chul Lee
 Dept. of Electronics & Electrical Eng. Chung-Ang University

Abstract - This paper proposes a pattern classification algorithm for the defect distribution of Hard Disk Drive (HDD). In the HDD productions, the defect pattern of defective HDD set is important information to diagnosis of defective HDD set. In this paper, 5 characteristics are determined for the classification to six standard defect pattern classes. A fuzzy inference system is proposed, the inputs of which are 5 characteristic values and the outputs are the possibilities that the input pattern is classified to the standard patterns. Classification result is the pattern with maximum possibility. The proposed algorithm is implemented with a PC system for defective HDD sets and shows its effectiveness.

1. 서 론

HDD(Hard Disk Drive)는 데이터의 기록/재생, 보관, 접근 속도 등을 고려하였을 때, 아직까지 가장 유력한 기록장치이다[1],[2]. 각 공정 단계에서 HDD의 읽기/쓰기/찾기 등의 검사를 수행하여 HDD의 이상 유무를 확인한다[3]. 수리공정에서 HDD 결함 분포 분류는 불량률의 종류를 나누는 중요한 정보를 제공한다. 패턴분류는 임의로 주어진 미지의 패턴을 여러 클래스 중의 한 클래스로 대응시키는 정보처리 과정이다. 인식대상의 특성에 따라 다양한 패턴인식의 기법이 적용된다[4]-[7]. 수리공정에서 관리하는 결함 분포의 표준 패턴 클래스(Standard Pattern Class)는 루프(Loop)형, 분침(Watch)형, 아크(Arc)형, 방사(Radial)형, 점침(Spot)형, 전체(Whole)형의 여섯 가지이다.

수작업에 기초한 수리판정 작업을 자동화하게 되면, 부정확한 수리판정에 기인한 추가 공정비용 발생 및 부품 추가소요, 수리사의 숙련도 편차에 따른 공정능력 저하 등을 예방할 수 있다. 결함 분포의 패턴 분석을 바탕으로, 패턴 분류의 근거가 되는 5가지의 특징을 기반으로 한 퍼지추론(Fuzzy Inference)을 이용하였다[8]-[11]. 제시된 알고리즘은 실제 제조 공정에서 발생한 HDD를 대상으로 적용할 수 있도록 PC 시스템으로 구현되었고, 실 적용 결과, 만족할 만한 패턴 분류 성능을 확인하였다.

2. 본 론

2.1 HDD결함분포의 표준패턴클래스

디스크는 표면상에 동심원들로 구성된 실린더(Cylinder)와 각 실린더내의 물리적인 최소 저장 단위인 섹터(Sector) 들로 구성되어 있다. 고속으로 회전하는 디스크와 디스크의 중심에 수직인 방향으로 왕복운동을 하는 HSA(Head Suspension Assembly)의 제어에 의해 고밀도로 집적된 디스크 섹터에 접근하여 데이터의 입출력이 수행된다[1],[2].

그림 1은 전형적인 여섯 가지 표준 패턴 클래스를 나타낸다.

본 연구에서는, HDD결함 분포의 전처리(Pre processing)를 통하여 결함 분포 데이터 파일을 작성하였다. 일반적으로 500개 이하의 결함들이 다음과 같은 형태로 나타내어진다.

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

d_i 는 i 번째 결함, r_i 와 θ_i 는 각각 결함의 거리 및 각도이다. 이때, HDD의 3시 방향을 기준으로 하여 반시계 방향으로 각도가 증가하는 것으로 설정하였다.

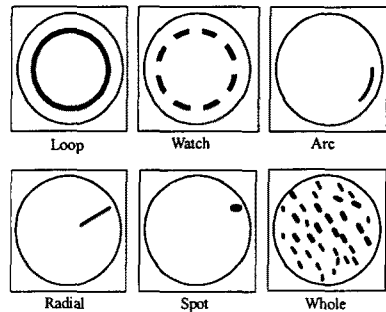


그림1. HDD 결함분포의 표준 패턴 클래스

2.2 패턴 분류를 위한 5 가지 특성

A. 필터링(Filtering)후결함의개수

극좌표를 직교좌표(Cartesian Coordinates)로 변화하여 횡축은 각도, 종축은 반지름으로 변형한 후, 반지름 축은 1, 각도 축은 각각 1°씩으로 이산화시켜서 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. 필터링 된 패턴 데이터의 결함 개수를 특징으로 선정 하였다.

B. 필터링 전후의 결함개수차이

필터링 전후의 결함 개수 차이이다.

C. r성분의 표준편차

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

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

n 은 필터링 후 결함의 총 개수, m_r 은 결함 데이터 r 성분의 평균값이다.

D. θ 점유도

전체 360°에서 결함을 포함한 각도를 다음과 같이 계산

하였다.

$$\text{점유도} = (\text{결합을 포함한 각도}/360) * 100 \quad (3)$$

E. θ 성분의분포도

각도의 관점에서 결합들이 얼마나 퍼져있어 있는 지를 나타내는 정도이다. θ 의 표준 편차를 직접 사용할 경우에는 극좌표계의 특성상 0° 과 359° 는 매우 가까운 각도임에도 불구하고 데이터 상에서는 큰 차이를 나타내는 것으로 처리되게 된다. 이러한 문제를 해결하기 위해서, 결합들의 각도를 m_θ 를 기준으로 새롭게 정의하여 σ_θ 를 구하였다. 즉, 물리적으로 m_θ 로부터 가장 멀리 떨어진 각도는 대칭방향인 ($m \pm 180^\circ$)이므로, 모든 결합의 θ 가 평균 m_θ 를 기준으로 $-180^\circ \sim +180^\circ$ 로 사상(Mapping)되도록 다음과 같이 재 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)과 같이 재정의 된 각도 표준편차를 구하면

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

이는 "평균에서 분포된 정도"라는 표준편차의 물리적인 의미에 부합하게 된다.

입력 데이터를 직접 이용하여 구한 평균 m_{θ_1} , 그리고 9시 방향을 극좌표계의 새로운 기준 축으로 변환하여 구한 평균 m_{θ_2} 를 각각 계산하였다.

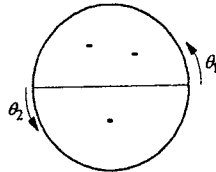


그림 2. 두 개의 각도 기준 축 θ_1 과 θ_2 .

두 개의 평균 m_{θ_1} , m_{θ_2} 를 바탕으로 (5)를 이용하여

σ_{θ_1} 와 σ_{θ_2} 를 각각 구하였다.

최종적으로는, 두 σ_{θ_1} 과 σ_{θ_2} 중 작은 값을 각도 성분의 분포도로 정의하였다.

$$\text{성분의 분포도} = \min(\sigma_{\theta_1}, \sigma_{\theta_2}) \quad (6)$$

2.3퍼지추론을이용한패턴분류

하나의 특징이 각 패턴을 독립적으로 결정할 수 없고 각 특징들의 상호 연관을 통하여 패턴이 구분되어야 하는 상황으로 분석되었다. 이러한 경우, 퍼지추론은 매우 유용한 도구가 될 수 있다[8]-[11]. 이는 각 입력 변수를 결정적으로 고려하는 대신, 각 집합에 속하는 가능성(Possibility)을 종합적으로 고려하여 추론할 수 있는 장점이 있다[12],[13]. 5가지 특징들을 입력받아, 각 패턴에 속할 가능성이 출력되는 퍼지규칙(Fuzzy Rule)을 다음과 같이 정의하였다.

Rule 1) If x_1 is A_1 , x_2 is B_1 , x_3 is C_1 , x_4 is D_1 and x_5 is E_1 Then 루프형. (7)

Rule 2) If x_1 is A_2 , x_2 is B_2 , x_3 is C_2 , x_4 is D_2 and x_5 is E_2 Then 분침형. (8)

Rule 3) If x_1 is A_3 , x_2 is B_3 , x_3 is C_3 , x_4 is D_3 and x_5 is E_3 Then 아크형. (9)

Rule 4) If x_1 is A_4 , x_2 is B_4 , x_3 is C_4 , x_4 is D_4 and x_5 is E_4 Then 방사형. (10)

Rule 5) If x_1 is A_5 , x_2 is B_5 , x_3 is C_5 , x_4 is D_5 and x_5 is E_5 Then 적힘형. (11)

Rule 6) If x_1 is A_6 , x_2 is B_6 , x_3 is C_6 , x_4 is D_6 and x_5 is E_6 Then 전체형. (12)

여기서, x_1 : 필터링 후 결합의 개수,
 x_2 : 필터링 전후의 결합 개수 차이,
 x_3 : r 성분 표준편차,
 x_4 : 점유도,
 x_5 : 성분의 분포도.

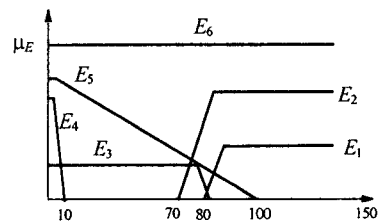
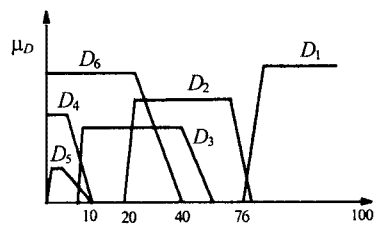
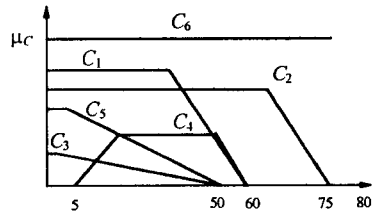
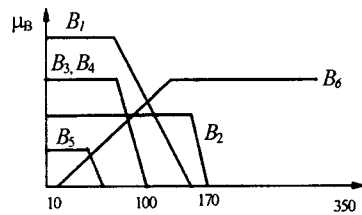
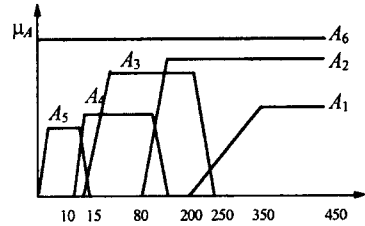


그림 3. 퍼지 집합 A, B, C, D, E.

퍼지 집합 A, B, C, D, E는 취합된 패턴 데이터의 반복적인 수정을 통하여 수작업으로 선정하였다. 그림 2는 각 입력에 대한 퍼지 집합 A, B, C, D, E를 나타낸다. 퍼지 집합 $A_1, A_2, \dots, E_5, E_6$ 의 소속정도(Membership Degree) 최대값은 모두 1이지만, 편의상

각 집합들의 구분을 위하여 적절하게 스케일링 하여 나타내었다. 새로운 패턴에 대해서 $(x_1, x_2, x_3, x_4, x_5)$ 이 계산되고, 이는 퍼지 추론 시스템의 여섯 개의 규칙에 각각 입력된다. 각 규칙에서는 다음과 같은 적합도(Truth Value)를 계산하게 된다[9]-[11].

$$\omega_i = \min[\mu_{Ai}(x_1), \mu_{Bi}(x_2), \mu_{Ci}(x_3), \mu_{Di}(x_4), \mu_{Ei}(x_5)]$$

for i -th rule

(13)

따라서, 각 규칙의 적합도가 최대로 나타나는 패턴을 최종적인 출력으로 도출하게 된다.

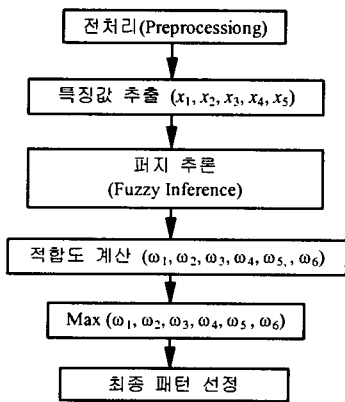


그림 4. 제시된 패턴 분류시스템의 전체 과정

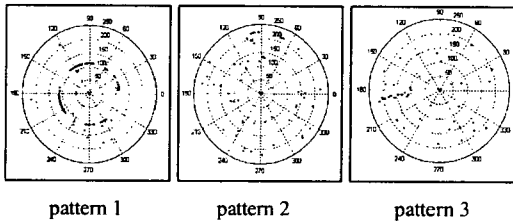


그림 5. 세 개의 테스트 패턴

표 1. 패턴 분류 과정

		Pattern 1	Pattern 2	Pattern 3
특 징	x_1	124	14	11
	x_2	15	49	42
	x_3	3.4109	42.7240	31.4174
	x_4	32.22	3.889	1.667
	x_5	84.3144	10.5473	4.9074
적 합 도	ω_1	0	0	0
	ω_2	1	0	0
	ω_3	0.3428	0.1545	0
	ω_4	0	0	1
	ω_5	0	0.10	0.404
	ω_6	0.05	0.390	0.320
결 과		watch	whole	radial

그림 5 는 제시된 알고리즘의 적용을 설명하기 위한 세 개의 테스트 패턴이다. 표 1 에서 첫 번째 패턴의 특징

값 $(x_1, x_2, x_3, x_4, x_5)$ 는 각각 (124, 15, 3.4109, 32.22, 84.3144)로 계산되었으며, 이 값이 퍼지 추론에 입력된 결과, 각 규칙의 적합도가 (0, 1.0, 0.3428, 0, 0, 0.05)로 나타나 최종적으로 분침형 으로 판정된 결과를 나타낸다

구현된 결과를 500개의 Sample HDD에 대해서 실험한 결과, 5년 이상 수련도를 갖는 수리사의 판별 결과와 483개가 일치하여 96.6%의 적응률을 나타내었다.

3. 결 론

본 연구에서는 하드디스크 드라이브(Hard Disk Drive, HDD) 결합 분포의 패턴 분류에 관한 연구 결과를 제시하였다. 먼저, 각 패턴을 구분할 수 있는 5개의 특징들을 추출하였다. 또한 패턴 분류의 특성에 부합하도록 퍼지 추론 시스템을 구성하여 각 특성들의 상호 연관을 고려하여 각 패턴에 속할 가능성을 추론할 수 있도록 하였다. 본 연구 결과는 수작업에 기초한 수리판정 작업의 자동화를 가능하게 하여, 차후 실제 생산 현장에 유용하게 응용될 것으로 기대된다. 본 연구에서 제시한 각 성분의 집중도는 유사한 극좌표계의 패턴 분류 문제에도 유용하게 사용될 수 있을 것으로 기대되며, 각 특징들의 퍼지 추론에 의한 패턴 분류 또한 유사한 분야의 패턴 분류에 유용하게 적용될 수 있다고 판단된다.

[참 고 문 헌]

- [1] 정정주, 이승희, 추상훈, "디스크 드라이브의 기술 동향과 제어기술", *제어자동화시스템 공학회지*, 1998년 5월호, pp. 15-21.
- [2] 이호성, "하드 디스크 드라이브 서보 제어기의 설계", *제어자동화시스템 공학회지*, 1998년 5월호, pp. 22-28.
- [3] 문운철, 이승철, 남창우, "Hard Disk Drive 검사 시스템의 고장 진단을 위한 퍼지-이진 논리 결합형 전문가 시스템에 관한 연구" *대한전기학회 논문지*, 53-D권 6호, pp. 457-464, 2004년 6월.
- [4] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, 1999.
- [5] M. Nadler and E. P. Smith, *Pattern Recognition Engineering*, John Wiley and Sons, 1993.
- [6] J. Schurmann, *Pattern Classification*, John Wiley and Sons, 1996.
- [7] 김상운, *패턴인식 입문*, 홍릉과학출판사, 1995.
- [8] L. A. Zadeh, "Fuzzy Sets", *Inform. Contr.*, Vol. 8, pp. 338-353, 1965.
- [9] H. J. Zimmermann, *Fuzzy set theory and its applications*, Kluwer Academic Publishers, 1991.
- [10] W. Pedrcz, *Fuzzy logic and fuzzy systems*, JOHN WILEY & SONS, 1989.
- [11] B.Kosco, *Neural network and fuzzy systems*, Englewood Cliffs, Prentice-Hall, 1992.
- [12] A. Baraldi and P. Blonda, "A survey of fuzzy clustering algorithms for pattern recognition-Part I & II", *IEEE Trans. On Systems, Man and Cybernetics-Part B: Cybernetics*, Vol. 29, No. 6, pp. 778-801 Dec. 1999.
- [13] K. S. Ray and T. K. Dinda, "Pattern classification using fuzzy relational calculus", *IEEE Trans. On Systems, Man and Cybernetics-Part B: Cybernetics*, Vol. 33, No. 1, pp. 1-16, Feb. 2003