

# 패턴인식법에 의한 압축기의 이상진단에 관한 연구

김태구 · 김광일

인제대학교 산업안전보건학과

## 1. 서 론

엔진이나 콤프레셔 등과 같은 기기 진동의 동특성은 불규칙적으로 변동하는 성분을 갖는 랜덤 프로세스로 그것을 수학적으로 명확히 기술하는 것은 어렵다<sup>1)</sup>. 하지만 통계학적인 입장에서 시계열 데이터를 보면 시계열 데이터가 확률변수로서 각각의 모집단에 속한다. 따라서 이 점에 주목하여 시계열의 확률적인 특징을 추출하는 것으로, 각각의 시계열 데이터를 확률공간으로 구별하는 것이 가능하다면 시계열 데이터에 의해 표현되어진 상태의 식별가능하다는 이론이 성립된다<sup>2)</sup>. 이럴 때에 특히 신호데이터 간의 차이를 효율 높게 검출하는 것이 중요하다. 이를 위하여 정보이론중의 하나인 쿨백식별정보량(Kullback Discrimination Information, 이하KDI)<sup>3),4)</sup>을 이용을 검토한 것이 제안

되어져 왔다. 쿨백식별정보량은 2개의 확률밀도함수를 비교하기 위한 척도로서, 신호데이터의 분류를 모집단의 선별(選別)문제로 하여 확률적 입장에서 확률공간의 차이를 계산하는 것이다<sup>5)~9)</sup>. 본 논문에서는 콤프레셔를 실험대상으로, 계측한 진동데이터를 기초로 KDI이상진단시스템을 바탕으로 이상검지를 시험했다.

본 논문에서는 실험장치에 있어서, AR모델<sup>10)</sup>에 의해 동정하고 시계열 모델을 식별정보량인 KDI를 사용하여 식별하기 위한 최적의 측정조건을 시험했다. 또한 실제로 이상검지시스템의 실용성을 고려한 경우, 검지의 다기능화가 요구되어 진다. 즉, 1개의 이상검지장치로 대상이 된 기기의 정상과 이상상태만의 진단이 아닌, 이상의 종류 및 정도의 식별까지 요구되어 진다. 이럴 경우에는 패턴인식법이론의 유효하다는 것을 주목하게되는 데<sup>11)</sup> 본 논문에서는 확률공간에 있어서 모델간의 거리의 크기를 최근접분류법(The Nearest Neighbor Classification Rule, NN rule)에 의해, 최근접 모델의 라벨을 부여하는 방법으로 정상과 이상을 검지하는 동시에 이상의 종류까지의 식별의 가능성을 고찰했다.

## 2. 이 론

### 2.1. Kullback Discrimination Information 및 패턴인식

쿨백식별정보량(Kullback Discrimination Information, 이하 KDI)은 실제의 확률분포가 쿨  $f_0(x)$ 일 때, 그 분포를  $f_1(x)$ 로 가정했을 때의 부적합(dissimilarity)정도를 나타내는 척도로서, 다음식과 같이 표현된다.

$$I(f_0, f_m) = \int f_0(x) \log[f_0(x)/f_m(x)] dx \quad (1)$$

여기서, 확률분포  $f_0(x), f_m(x)$ 를 정규분포로하고  $x_0, x_m$ 를 그것을 가정으로 발생하는 시계열이라 하면,  $x_0, x_m$  간의 거리를 나타내면 식(2)와 같다.

$$2d(x_0, x_m) = \log \frac{\hat{\Sigma}_m}{\hat{\Sigma}_0} + tr \hat{\Sigma}_0 (\hat{\Sigma}_m^{-1} - \hat{\Sigma}_0^{-1}) + tr \left[ \hat{\Sigma}_0^{-1} (\hat{\mu}_0 - \hat{\mu}_m) (\hat{\mu}_0 - \hat{\mu}_m)^T \right] - n \quad (2)$$

여기서,  $x_0$  과  $x_m$  을  $n$ -duration 시계열,  $\hat{\mu}_j = (\hat{\mu}_j(1), \dots, \hat{\mu}_j(n))$ ,  $\hat{\Sigma}_j (j=0 \text{ or } m, m=1, \dots, M)$ 는 각각의 시계열  $x_j$ 의 추정평균과 추정공분산 행렬이다. 식(2)로 부터 시계열간의 거리를 구 할 경우에는 동식으로부터 구한다는 가정으로부터, 시계열  $x_0$  과  $x_m$ 의 확률분포는 정규분포여야 한다. 또, 관측된 시계열 데이터로부터 계산되어진  $n \times n$  공분산 행렬의 계산이 포함되어 있다. 그러므로, 관측데이터에 AR Model를 fitting시킴으로 잔차시계열(residual time series)의 정규성을 만족하고 계산의 간략화를 기대 할 수 있다. AR 모델을 fitting시킨 거리의 계산식을 식(3)에 표시하면<sup>11)</sup>,

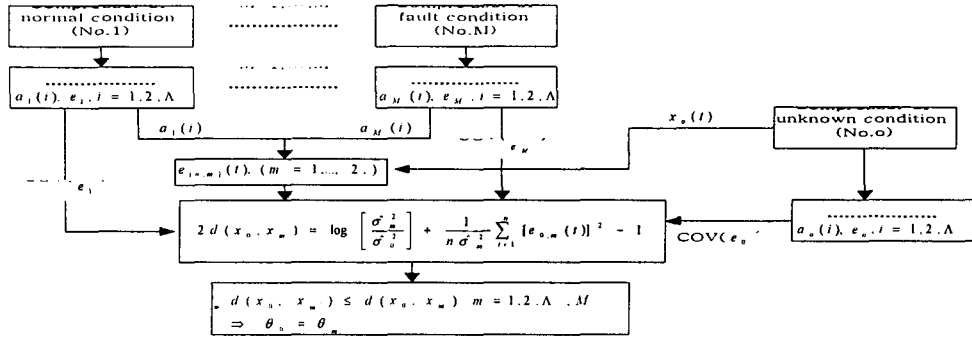
$$2d(x_0, x_m) = \log \left[ \frac{\hat{\sigma}_m^2}{\hat{\sigma}_0^2} \right] + \frac{1}{n \hat{\sigma}_m^2} \sum_{t=1}^n [e_{0,m}(t)]^2 - 1 \quad (3)$$

여기서, 기호  $\hat{\sigma}_0^2, \hat{\sigma}_m^2$  각각의 시계열  $x_0, x_m$ 의 추정분산,  $e_{0,m}(t)$ 는 표본시계열  $x_0$ 에 Fitting된 모델로부터 테스트용 시계열  $x_m$ 을 Filtering함으로 발생하는 잔차시계열이다.

패턴인식의 분야에서 사용되고 있는 최근접분류법(the Nearest Neighbor Classification Rule, 이하 NN Rule)은 다음과 같다.  $x_m, m=1, \dots, M$ 을 표본시계열(labelled time series),  $\theta_m$ 을 그 시계열에 대응하는 라벨(label; 정상, 이상등)  $x_0, \theta_0$ 를 새로 분류되는 테스트용 시계열과 그것의 라벨이라고 한다. 또,  $x_0$  과  $x_m$ 의 거리를  $d(x_0, x_m)$ 라고 하면,

$$d(x_0, x_m) \leq d(x_0, x_n) \quad m=1, 2, \dots, M \\ \Rightarrow \theta_0 = \theta_m \quad (4)$$

즉, 새로 분류되는 테스트용 시계열은 그것의 최근접의 라벨이 주어진다. 패턴인식과 KDI를 이용한 이상검지의 흐름을 그림1에 나타냈다.



### 3. 실험

본 연구의 실험에서 실험장치도(Fig. 2)를 이용해서 콤프레셔의 정상상태(No.1-Normal)와 보통 자주 발생하는 7종류(①No.2-Suction filter choking, ②No.3-Valve spring failure, ③No.4-Wear of piston ring 5wt-%, ④No.5-Wear of piston ring 10wt-%, ⑤No.6-Suction valve failure, ⑥No.7-Suction filter choking and Wear of piston ring 5wt-%, ⑦No.7-Suction filter choking and Wear of piston ring 10wt-%)이상상태의 콤프레셔의 진동가속도를 계측조건(Table. 1)에 따라 측정하였다. 여기서 얻어진 데이터를 각각의 표본시계열로 하고, 즉 패턴인식에서의 표본시계열을 정하는 문제에 있어 랜덤추출의 방법을 택하여, 각각의 시계열 데이터의 AR 파라미터를 추정하였다. 다음으로, 각각 8종류의 상태에 대응하는 테스트 시계열을 각 10회 측정하였다. 그리고 각 상태에서의 표본시계열의 AR 파라미터를 사용하여, 테스트시계열 데이터를 동정하고 그로부터 얻은 잔차시계열을 계산했다. 마지막으로 구한 잔차시계열 간의 KDI를 계산해서, NN rule에 의해 각 테스트용 시계열의 상태를 분류했다.

Table 1. Experimental conditions

1. Number of data	8000, 4000, 2000, 1600
2. Sampling periods	10 $\mu$ sec., 20 $\mu$ sec., 50 $\mu$ sec.
3. Piston speed	2800 RPM
4. Nozzle pressure	No load 80% (All condition)

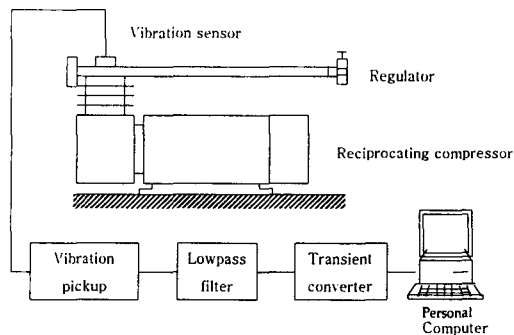


Fig.2 Schematic diagram of experimental apparatus

#### 4. 실험결과

계측조건으로 데이터 수 8000개, 샘플링 주기 10  $\mu$ sec., 20  $\mu$ sec., 50  $\mu$ sec.시의 각각 10회씩 측정하여 총80회의 판정을 행하여 판정한 결과는 Table 2.에 나타낸다. 이 표를 보면, 데이터수가 같은 8000개의 경우, 어떤 상태든지 샘플링주기가 10  $\mu$ sec.일 때 오분류가 가장 많았고, 50  $\mu$ sec.일 때 정확히 판정하는 경우가 많음을 알았다.

그리고, 데이터의 수를 다르게 하여 즉, 샘플링 주기와 1개의 시계열에 포함되어 있는 데이터의 수를 8000, 4000, 2000, 1600개로 다르게 해서 얻은 결과를 Table. 3에 나타냈다. Table. 3으로부터 다음과 같은 특징을 볼 수 있었다. 즉, a)샘플링주기가 길수록, b)샘플링 데이터의 수가 많을수록 대상을 파악하기 쉽다는 것이다. 따라서, b)는 일반적으로 당연한 귀결이지만, a)의 이유에 대해서는 다음에서 고찰하였다.

Table 2. Experimental results (Samples=8000)

	KDI between measured signal and each labeled class								
	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	Correcting decision rate[%]
10 $\mu$ sec.	40%	60%	100%	0%	70%	80%	20%	60%	53.8
20 $\mu$ sec.	90%	100%	100%	100%	100%	100%	100%	50%	92.5
50 $\mu$ sec.	90%	100%	100%	100%	100%	80%	100%	100%	96.5
Average of vibration[gal]	3300	3000	3000	2400	1700	1000	2600	1700	

Table 3. Summary of experimental results

Samples	Samples period[ $\mu$ sec.]			Samples	Samples period[ $\mu$ sec.]		
	10	20	30		10	20	30
1600	-	-	92.5%	4000	50.0%	80.0%	98.8%
2000	40.0%	76.3%	95.0%	8000	53.8%	92.5%	96.3%

#### 5. 고찰

##### 5.1. 최적 계측조건의 결정

실험결과(Table. 3)으로부터 다음과 같은 특징을 알 수 있다.

- 1)시계열을 포함하는 데이터의 수가 같을 때에는 샘플링 주기가 길수록 정보율이 높다.
- 2)샘플링 주기가 같을 때에는 샘플링 데이터의 수가 많은 수록 정보율이 증가하는 경향이 있다.

즉, 본 실험에 사용한 계측기 및 PC등에 있어, 콤프레셔 각각의 상태를 분류한 결과로부터, 10  $\mu$ sec. ~ 50  $\mu$ sec. 에서는 샘플링 주기가 길수록 정보율이 높아지고, 데이터의 수가 많을수록 판정의 정보율이 높아지는 것을 고려하면, 데이터수 8000개, 샘플링 주기 50  $\mu$ sec.의 계측조건은 본 실험에 있어서 최적의 조건으로 사료되어진다.

## 5.2. 상태판정의 효율향상

본 논문에서 사용한 방법에서는, 새로 관측한 데이터를 여러종류의 표본시계열의 원 세트의 계산으로 분류 가능한 특징이 있지만, 실제의 시스템, 플랜트에서는 상정하지 못한 이상의 상태도 발생되어진다. 이 방법으로는 시스템의 정상상태의 KDI의 분포를 구해놓고 정상상태의 분포 외에 새롭게 관측된 시계열과 정상상태와 KDI를 계산하여 역치보다 큰 KDI치가 발생하면 이상으로 분류하여 판단하는 것이다. 여기서, 본 실험에서는, 정상상태의 표본시계열과 각각의 상태에서 10 세트의 테스트 시계열 데이터 간의 KDI를 계산하여, 그 분포를 그림 3에 나타냈다. 예를 들어 KDI치가 0.1일 때를 역치로 설정하면(그림의 점선), 정상이 아닌 경우 신속히 검지가 가능하다. 이 방법도 하나의 이점으로 계산시간을 절약하고 상태판정이 조기에 가능한 방법이라 하겠다.

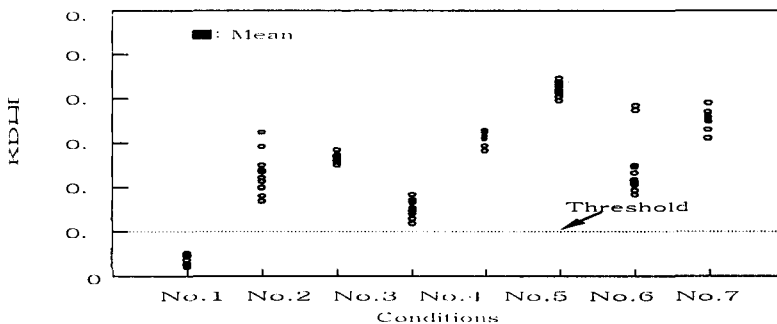


Fig. 3 KDI between normal condition and each fault condition (Sampling period=10  $\mu$ sec., Samples=8000)

## 6. 결론

본 논문은 쿨백식별정보량을 이용하여 이상검지시스템을 구축, 콤프레셔의 이상을 진단한 것이다. 본 연구로 얻은 결과를 종합하면 다음과 같이 요약할 수 있다.

- 1) 계측한 진동데이터와 표본시계열로써 모델화한 참조모델과의 KDI를 계산하여 비교함으로써 이상의 종류를 식별할 수 있었다. 구체적으로 콤프레셔의 이상진단실험에 있어 7 종류의 이상발생시 각각의 상태를 검지하고 분류하는 것이 가능했다.
- 2) 불규칙적으로 변동하는 신호계측에 있어 최적계측조건을 설정하는 프로세스를 나타내었다. 이 프로세스를 콤프레셔의 실험으로 검증하여 데이터 수 8000개, 샘플링주기 50  $\mu$ sec.의 계측조건이 최적인 결과를 얻었다.
- 3) 상정하지 못한 이상을 검지하고 판정하는 방법을 제안했다. 정상상태만의 KDI의 분포를 구해놓고 역치를 설정하여 정상상태와 다른 상태일 때 조기에 상태를 검지하고 경보(Alarm)할 수 있는 시스템을 제안하였다.

## 참고문헌

- 1) 柴田壁: 不規則現象, 日本機械學會誌, Vo.73, No.621, pp.1391~1400 1970
- 2) 山崎弘郎: 異常の檢出と豫知(センサと設備診斷技術), 工業調査會, pp.190~194 1988
- 3) S.Kullback: Information Theory and Statistics, pp.3~11, John Wiley and Sons Inc. USA 1959
- 4) S.kullcack, R.A.Leibler: On Information and Sufficiency, The Annals of Mathematical Statistics, Vol.22, pp.79~86 1951
- 5) W.Gersch: Kullback Leibler Number Cluster/Classification Analysis of Stationary Time Series, Proceedings of the 4th International Joint Conference on Pattern Recognition, pp.270~275, Kyoto, JAPAN 1978
- 6) W.Gersch, S.Braun: Nearest Neighbor-Time Series Analysis Classification of Faults in Rotating Machinery, Transactions of the ASME Journal of Vibration, Acoustics, Stress and Reliability in Design, Vol.105, pp.178~184 1983
- 7) S.Braun: Mechanical Signature Analysis-Theory and Applications-, pp. 288~318, Academic Press Inc. London 1986
- 8) Christian Doncarli and Eric Le Carperntier: An Optimal Approach for Random Signals Classification, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.13, No.11, pp.1192~1196 1991
- 9) Y.C.Jeon and C.Jams Li: Non-linear ARX Model-Based Kullback Index for Fault Detection of a Screw Compressor, Mechanical Systems and Signal Processing, Vol.9, No.4, pp.341~358 1995
- 10) 熊丸耕介, Torsten SODERSTROM: カルバック識別情報量規範を用いた故障檢出とモデル評價, 計測自動制御學會論文集, Vol.22, No.10, pp.1058~1065, 昭和61
- 11) 김태구, 김광일: 기어의 이상검지 및 진단에 관한 연구-Wavelet Transform 해석과 KDI의 비교-, 한국산업안전학회지, Vol.15, No.2, pp.1~7 2000