A Study on Application of Discrete Wavelet Decomposition for Fault Diagnosis on a Ship Oil Purifier

Paper Songho Lee, Taehyun Lee1, Jeongyeong Kim, Jongjik Lee, Kyungha Ryu, Yongjin Kim*, and Jong-Won Park*
Presentation by Tobramycin

NCUT

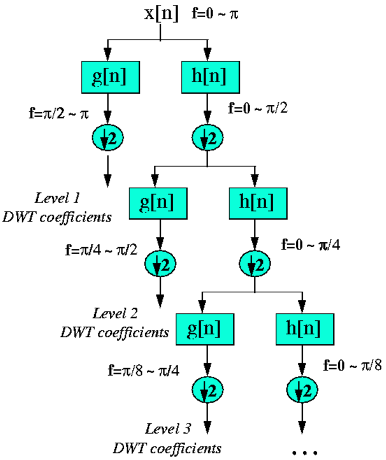
2022.7

Contents

1. Background
2. Framework
3. Establishment of Fault Diagnosis Framework using Fault Simulation

Background

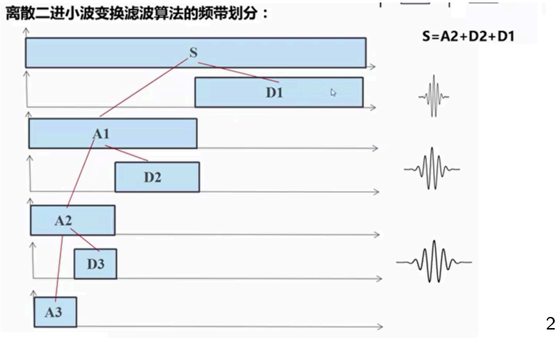
Discrete Wavelet Transform



[1]

[1] https://blog.csdn.net/qq_42874547/article/details/106942497

Discrete Wavelet Transform

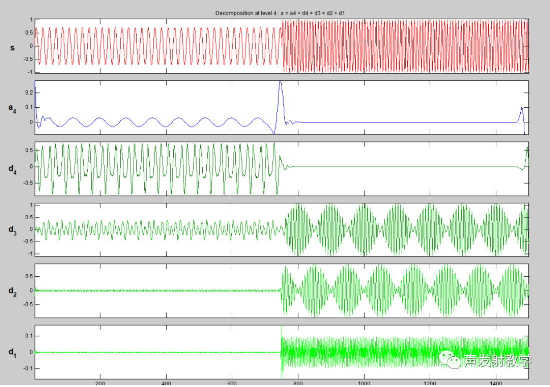


2

DWT 用小波函数（wavelet fuction）和尺度函数（scale function）来分别分析高频信号和低频信号，也即高通滤波器和低通滤波器。
和傅里叶变换的区别：保留了频率的时间位置信息

[2] https://blog.csdn.net/qq_42874547/article/details/106942497

Discrete Wavelet Transform

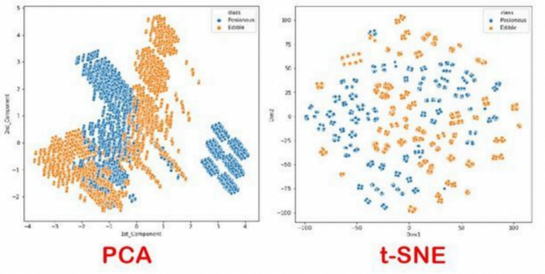


[3]

[3] https://blog.csdn.net/weixin$_3 9640262/article/details/111580713$

t-distributed Stochastic Neighbor Embedding (t-SNE)

基本思想: t-SNE 中主要是将“距离的远近关系”转化为一个概率分布，每一个概率分布就对应着一个“样本间距离远近”的关系。



PCA 的拥挤现象 [4]

[4] https://zhuanlan.zhihu.com/p/426068503

t-distributed Stochastic Neighbor Embedding (t-SNE)

- 1. 定义高维和低维概率分布:

$$P(j \mid i)=\frac{S\left(x_i, x_j\right)}{\sum_{k \neq i} S\left(x_i, x_k\right)}, j \neq i, i=1,2, \ldots, n \tag{1}$$

$$Q(j \mid i)=\frac{S'\left(z_i, z_j\right)}{\sum_{k \neq i} S'\left(z_i, z_k\right)}, j \neq i, i=1,2, \ldots, n \tag{2}$$

  - x 是高维数据点，z 是低维数据点。
  - $S\left(x_i, x_j\right)$ 是数据 x_i 与数据 x_j 的相似度，距离越近越相似。

t-distributed Stochastic Neighbor Embedding (t-SNE)

- 2. 计算 KL 散度 (Kullback-Leibler Divergence):

$$D_{KL}(p\|q) = \mathbb{E}_{X\sim p}\left[\log\frac{p(X)}{q(X)}\right] = \mathbb{E}_{X\sim p}[\log p(X) - \log q(X)] \tag{3}$$

$$\begin{aligned}
\sum_{i=1}^{n} p\left(x_i\right)\log\left(\frac{p\left(x_i\right)}{q\left(x_i\right)}\right) &= -\sum_{i=1}^{n} p\left(x_i\right)\log\left(\frac{q\left(x_i\right)}{p\left(x_i\right)}\right) &(4)\\
&\geq -\sum_{i=1}^{n}\log\left(\sum_{i=1}^{n} p\left(x_i\right)\cdot\frac{q\left(x_i\right)}{p\left(x_i\right)}\right) &(5)\\
&= 0 &(6)
\end{aligned}$$

  - 一般用于度量两个概率分布函数之间的“距离”
  - $D_{KL}(p\|q) \geq 0$
  - 公式4 是 KLD 的离散形式
  - 公式5 来自于琴生不等式 (Jensen Inequality): 若 $f(x)$ 是下凸函数，则 $\frac{\sum_{i=1}^{n} f(x_i)}{n} \geq f\left(\frac{\sum_{i=1}^{n} x_i}{n}\right)$

t-distributed Stochastic Neighbor Embedding (t-SNE)

- 3. 利用 KL 散度定义 t-SNE 的损失函数

$$L\left(z_1, z_2, \ldots, z_n\right)=\sum_{i=1}^n D_{KL}(P(j \mid i) \| Q(j \mid i))$$

$$=\textstyle\sum_{i=1}^n \sum_{j \neq i} P(j \mid i) \log \left(\frac{P(j \mid i)}{Q(j \mid i)}\right) \quad (7)$$

我们要目标是找到一组低维数据 $\left(z_1^*, z_2^*, \ldots, z_n^*\right)$ 使得上面这个损失函数达到最小，即 t-SNE 的最优解为

$$\left(z_1^*, z_2^*, \ldots, z_n^*\right)=\underset{z_1, z_2, \ldots, z_n}{\arg \min } L\left(z_1, z_2, \ldots, z_n\right) \tag{8}$$

梯度下降：

$$Z^{(t)}=Z^{(t-1)}+\eta \frac{\partial L}{\partial Z^{(t-1)}}+\alpha(t)\left(Z^{(t)}-Z^{(t-1)}\right) \tag{9}$$

Framework

Framework

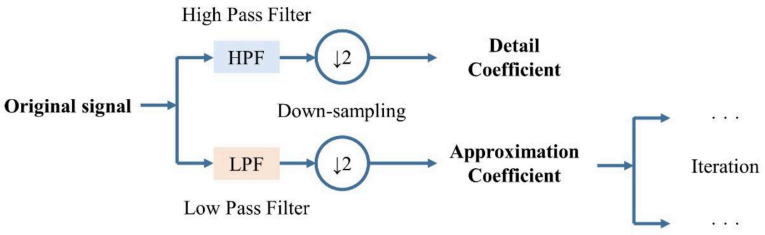


Figure 1. Schematic diagram describing the mechanism of DWT

Framework

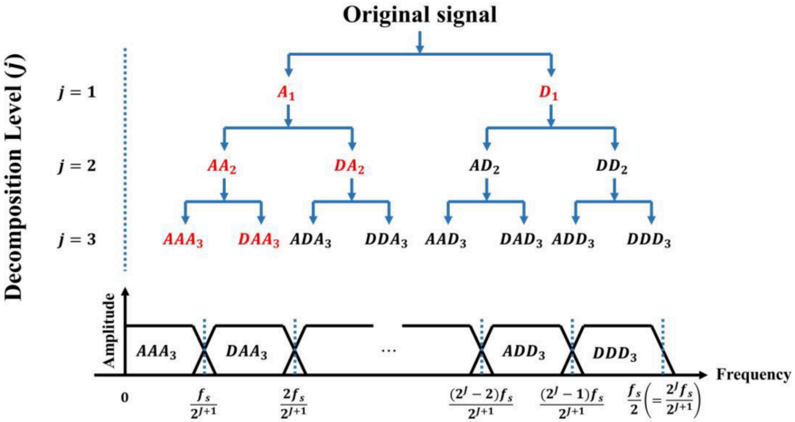


Figure 2. Schematic diagram describing the wavelet decomposition tree and its frequency band

t-SNE

$$p_{j|i} = \frac{\exp\left(-\left\|x_i - x_j\right\|^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\left\|x_i - x_k\right\|^2 / 2\sigma_i^2\right)} \tag{10}$$

$$q_{j|i} = \frac{\exp\left(-\left\|y_i - y_j\right\|^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\left\|y_i - y_k\right\|^2 / 2\sigma_i^2\right)} \tag{11}$$

σ_i is the standard deviation of the Gaussian distribution centered on x_i.

Support Vector Machine

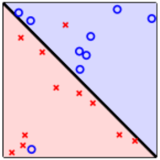
超平面:

$$\omega \cdot x_i + b \geq 1, y_i = 1 \tag{12}$$
$$\omega \cdot x_i + b \leq -1, y_i = -1 \tag{13}$$

损失函数:

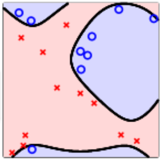
$$\text{minimize } \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}\xi_i \tag{14}$$

$$\text{subject to} y_i \left(\langle \Phi\left(x_i\right), \omega\rangle + b\right) \geq 1, i = 1, 2, \ldots, \mathrm{N} \tag{15}$$



$\mathbf{\Phi}_1$

- part of reasons: $\mathbf{\Phi}$
- other part: **separable**



$\mathbf{\Phi}_4$

Metrics

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{16}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{17}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{18}$$

$$F_1 \text{ score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{19}$$

Measure

Table 1. Wavelet spectrum based on discrete wavelet transform

Measure	Wavelet Spectrum
Mean	$w_j = \log_2\left[\frac{1}{n_j}\sum_{k=1}^{n_j} d_{j,k}^2\right]$
Median	$w_j = \text{median}\left(d_{j,k}^2\right)$
Variance	$w_j = \log_2\left[\frac{1}{n_j - 1}\sum_{k=1}^{n_j}\left(d_{j,k}^2 - \bar{d}_{j,k}^2\right)^2\right]$
Interquartile range (IQR)	$w_j = Q_3\left(d_{j,k}^2\right) - Q_1\left(d_{j,k}^2\right)$

Establishment of Fault Diagnosis Framework using Fault Simulation

Dataset

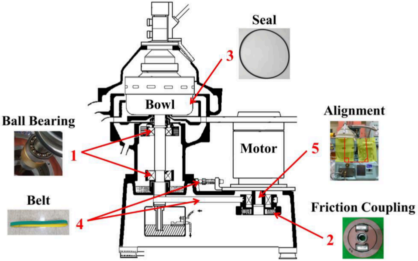


Figure 3. Schematic showing parts with high failure criticality of the oil purifier

Table 2. 5 types of failure components, modes, mechanisms, and effects

Mode Number	Component	Failure Mode	Failure Mechanism	Test Condition
0	-	Normal state	-	-
1	Bearing	Boundary lubrication	Poor lubrication	Reducing lubricant 100% compared to full capacity
2	Friction coupling	Poor power transmission	Wear of friction elements	Reducing the element mass 100% compared to the original state
3	Seal	Tension degradation	Thermal aging	Thermal aging at 120°C for 900 hours
4	Flat belt	Fatigue	Tension degradation	Shifting the motor towards the bowl by 13.62mm
5	Shaft (= Spindle)	Fatigue Crack	Misalignment	Tilting the motor shaft by 0.9° counterclockwise

Dataset & Configuration

"The total data set consisted of 22,800 data, 13,200 data for training and 9,600 data for testing. Then, in the train set, it was divided again by 8:2, so that 10,560 data were used for training and 2,640 data were used for validation. "

"The learning parameters of the OAO SVM (one-against-one SVM) were fixed as $=0.1$ and $C=10$."

Results

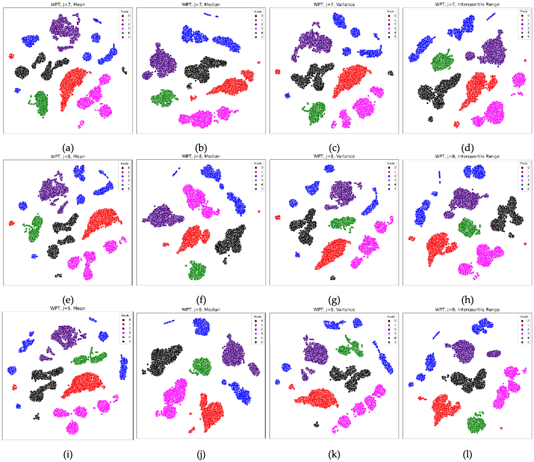


Figure 10. The t-SNE result of wavelet spectrum measures with respect to the decomposition level *j* based on the WPT. Wavelet spectrum measure varies in the row direction. (a),(e),(i) mean; (b),(f),(j)

Results

Table 3. The *Accuracy* and F_1 *score* for performance evaluation of the trained OAO SVM model

Decomposition Method		Wavelet Spectrum Measure	Bowl Axial dir.		Bowl Radial dir.		Motor Axial dir.		Motor Radial dir.	
			Accuracy	F_1 *score*	*Accuracy*	F_1 *score*	*Accuracy*	F_1 *score*	*Accuracy*	F_1 *score*
DWT	$j = 7$	Mean	0.9992	0.9992	0.9991	0.9991	0.9961	0.9961	0.9975	0.9975
		Median	0.9979	0.9979	0.9997	0.9997	0.9929	0.9929	0.9950	0.9950
		Variance	0.9993	0.9993	0.9994	0.9994	0.9998	0.9998	0.9990	0.9990
		IQR	0.9973	0.9973	0.9996	0.9996	0.9999	0.9999	0.9954	0.9954
	$j = 8$	Mean	0.9597	0.9573	0.9994	0.9994	0.9135	0.8967	0.9970	0.9970
		Median	0.9983	0.9983	0.9999	0.9999	0.9878	0.9877	0.9947	0.9947
		Variance	0.9979	0.9979	0.9994	0.9994	0.9083	0.8888	0.9984	0.9984
		IQR	0.9976	0.9976	0.9995	0.9995	0.9988	0.9987	0.9953	0.9953
	$j = 9$	Mean	0.8833	0.8742	0.9809	0.9812	0.8815	0.8391	0.9968	0.9968
		Median	0.9975	0.9975	0.9997	0.9997	0.9896	0.9895	0.9947	0.9947
		Variance	0.8820	0.8832	0.9961	0.9962	0.8961	0.8683	0.9980	0.9980
		IQR	0.9989	0.9989	0.9994	0.9994	0.9977	0.9977	0.9954	0.9954
WPT	$j = 7$	Mean	0.9658	0.9666	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		Median	0.9246	0.9296	0.9095	0.9039	0.9602	0.9599	0.9375	0.9387
		Variance	0.8855	0.8865	1.0000	1.0000	0.9996	0.9996	0.9994	0.9994
		IQR	0.9298	0.9342	0.8988	0.8891	0.9902	0.9903	0.9209	0.9167
	$j = 8$	Mean	0.8610	0.8592	0.8748	0.8748	0.8761	0.8772	0.8764	0.8767
		Median	0.8390	0.8511	0.8609	0.8424	0.9305	0.9341	0.8532	0.8451
		Variance	0.8590	0.8574	0.8744	0.8746	0.9531	0.9547	0.8747	0.8773
		IQR	0.7788	0.7927	0.8488	0.8186	0.9425	0.9459	0.7740	0.7647
	$j = 9$	Mean	0.7495	0.7141	0.7833	0.7565	0.6547	0.6345	0.6890	0.6870
		Median	0.8186	0.8371	0.7131	0.6765	0.8073	0.8173	0.6992	0.7079
		Variance	0.7698	0.7363	0.8132	0.8006	0.7595	0.7553	0.7542	0.7500

Thank you!