AI for Predictive Maintenance

Nirav Bhatt
MTech Industrial AI
IIT Madras, Chennai, India

From Traditional tools to AI

- Data-rich world
- Can we use the data for Predictive Maintenance?
- Predictive Maintenance:
  - Data Acquisition and Processing
  - Diagnostics: Anomaly or Fault detection and localization
  - Prognostics and forecasting: Monitoring the health and predicting the remaining useful life
  - Decision Support and Human/Software Machine interface
    - Decision-making through control (corrective action) and optimization

Data-driven Predictive Maintenance

- Diagnostics: Data-driven Fault detection and localization using key statistics and formulation
  - Unsupervised Approaches to detect, and localize faults with unlabelled data
  - Supervised Approaches: classification to faulty conditions with labeled data
- Prognostics and Forecasting
  - Remaining useful life is an important indicator of health
  - Regression-based formulation (Predict RUL) from time series of futures
  - Classification-based formulation (Predict failure within given classes)
  - Survival Analysis based approaches
    - Traditional Approaches Kaplan-Meier Model, Cox hazard model,
    - Regression models for survival response
    - DL based formulations

Data-driven Fault Diagnosis

Unsupervised ML:

T^2 and Q Statistic for application of PCA in fault detection

Traditional Approach:
Univariate Statistical Monitoring

Traditional Approach:
Univariate Statistical Monitoring

Traditional Approach:
Univariate Statistical Monitoring

T^2 Statistic

Training dataset $X \in \mathcal{R}^{n\times m}$ (m variables , n observations)

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}$$

Sample covariance matrix S is given by

$$S = \frac{1}{n-1} X^T X$$

Eigenvalue decomposition of the sample covariance matrix S is

$$S = V \Lambda V^T$$

T^2 Statistic

Quantifies the deviation of observed data from the reference point in a multivariate space

- Also referred to as Hotelling's T^2 statistic
- Follows a Hotelling's T^2 distribution, which is a generalization of the univariate Student's t-distribution to multiple dimensions
- Scaled squared 2-norm of an observation vector from its mean
- Takes into account both mean and covariances of the variables

T^2 Statistic

Assuming S is invertible

$$z = \Lambda^{1/2} V^T \mathrm{x}$$

Where $\mathrm{x} \in \mathcal{R}^m$ is the observation vector

Hotelling's T^2 statistic is given by

$$\boldsymbol{T^2 = z^T z}$$

- Scaling on x is in the direction of the eigenvectors which is inversely proportional to the standard deviation along the eigenvectors

Thresholds for T^2 Statistic

T^2 statistic follows a χ^2 distribution with m degrees of freedom

$$T^2_\alpha = \chi^2_\alpha(m)$$

Assumptions

- Observations are randomly sampled from a multivariate normal distribution
- Sample mean vector and covariance matrix for normal operations are equal to actual mean vector and covariance matrix

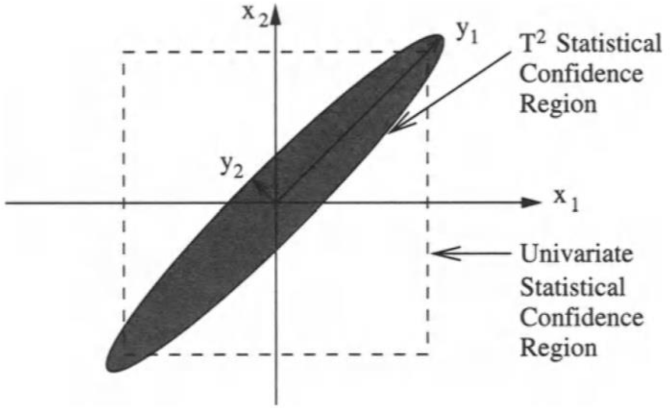
As degree of correlation between variables increases, elliptical confidence region becomes elongated and the amount of conservatism increases

Thresholds for T^2 Statistic

T^2 statistic follows a χ^2 distribution with m degrees of freedom

$$T_\alpha^2 = \chi_\alpha^2(m)$$

Assumptions



For a level of significance α, and m = 2, $T^2 \leq T_\alpha^2$ is an elliptical confidence region

As degree of correlation between variables increases, elliptical confidence region becomes elongated and the amount of conservatism increases

Thresholds for T^2 Statistic

Unknown covariance matrix

- Covariance matrix can be estimated from the sample covariance matrix.
- Faults can be detected using the threshold

$$\boldsymbol{T_\alpha^2 = \frac{m(n-1)(n+1)}{n(n-m)} F_\alpha(m, n-m)}$$

Where F_α (m, n-m) is upper 100α% critical point of the F-distribution with m and n-m degrees of freedom

Outliers detection

$$\boldsymbol{T_\alpha^2 = \frac{(n-1)^2 (m/(n-m-1)) F_\alpha(m, n-m-1)}{n(1+(m/(n-m-1))}}$$

Data requirements for T^2 Statistic

- Quality and quantity of the training dataset influences the effectiveness of T^2 statistic
- For a given α, the relative error is calculated by

$$\boldsymbol{\epsilon} = \frac{\frac{m(n-1)(n+1)}{n(n-m)} F_{\alpha}(m, n-m) - \chi^2_{\alpha}(m)}{\chi^2_{\alpha}(m)}$$

- If $\boldsymbol{\epsilon}$ is high, more data should be collected
- Number of observations is approximately 10 times the dimensionality of the observation space
- In case some diagonal elements of Λ are small, T^2 values will be erratic.
- In that condition, dimensionality reduction is recommended.

PCA: Revisit

PCA: Revisit

Application of PCA in fault detection

- T^2 statistic can be calculated from the PCA representation

$$\boldsymbol{T^2} = \boldsymbol{X^T V (\Sigma^T \Sigma)^{-1} V^T x}$$

Where V $\epsilon\ \mathcal{R}^{m\times m}$ unitary matrix, $\boldsymbol{\Sigma}\ \epsilon\ \mathcal{R}^{n\times m}$contains non-negative real singular values of decreasing magnitude along its main diagonal

Application of PCA in fault detection

- T^2 statistic can be calculated from the PCA representation

$$\boldsymbol{T^2} = \boldsymbol{X^T V (\Sigma^T \Sigma)^{-1} V^T x}$$

Where V $\epsilon\ \mathcal{R}^{m\times m}$ unitary matrix, $\boldsymbol{\Sigma}\ \epsilon\ \mathcal{R}^{n\times m}$contains non-negative real singular values of decreasing magnitude along its main diagonal

- Smaller singular values are prone to errors because these values contain small signal-to-noise ratio
- Therefore, loading vectors associated with larger singular values should be retained

Application of PCA in fault detection

- T^2 for the lower-dimensional space by including the loading vector (P) with only a largest singular values

$$\boldsymbol{T^2 = X^T P \Sigma_a^{-2} P^T x}$$

Where Σ_a contains first a rows and columns of Σ

- The above equation measures the variations in the scores space only
- If the actual mean and covariance are known, then

$$\boldsymbol{T^2 = \chi_\alpha^2(a)}$$

- T^2 statistic threshold when actual covariance matrix is estimated from sample covariance

$$\boldsymbol{T_\alpha^2 = \frac{a(n-1)(n+1)}{n(n-a)} F_\alpha(a, n-a)}$$

Application of PCA in fault detection

Application of PCA in fault detection

- For detecting outliers

$$\boldsymbol{T_{\alpha}^2} = \frac{(\boldsymbol{n}-\boldsymbol{1})^{\boldsymbol{2}}(\boldsymbol{a}/(\boldsymbol{n}-\boldsymbol{a}-\boldsymbol{1}))\boldsymbol{F_{\alpha}}(\boldsymbol{a},\boldsymbol{n}-\boldsymbol{a}-\boldsymbol{1})}{\boldsymbol{n}(\boldsymbol{1}+(\boldsymbol{a}/(\boldsymbol{n}-\boldsymbol{a}-\boldsymbol{1}))\boldsymbol{F_{\alpha}}(\boldsymbol{a},\boldsymbol{n}-\boldsymbol{a}-\boldsymbol{1})}$$

- T^2 statistic is sensitive to inaccuracies in the PCA space corresponding to the smaller singular values.
- This is because, it directly measures the scores corresponding to smaller singular values.

Squared prediction error and Q statistics

Q Statistic

Measures the sum of squared residuals or the sum of squared standardized residuals

- Used for monitoring the portion of the observation space corresponding to the m-a smallest singular values

$$\boldsymbol{Q} = \boldsymbol{r}^{\boldsymbol{T}}\boldsymbol{r},$$
$$\boldsymbol{r} = \left(\boldsymbol{I} - \boldsymbol{P}\boldsymbol{P}^{\boldsymbol{T}}\right)\boldsymbol{x}$$

Where r is the residual vector, a projection of the observation x into the residual space, P is the loading matrix

- Does not directly measure the variations along each loading vector
- Measures the total sum of variations in the residual space
- Does not suffer from an over-sensitivity to inaccuracies in the smaller singular values

Threshold for Q Statistic

- Also known as squared prediction error (SPE)
- Squared 2-norm measuring the deviation of the observations to the lower-dimensional PCA representation.
- Distribution for the Q statistic

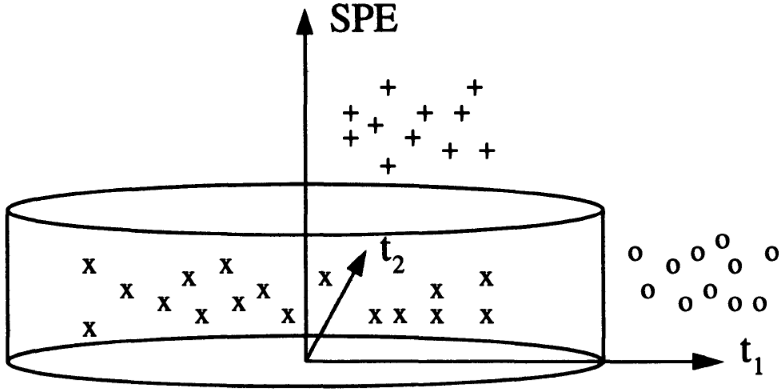
$$\boldsymbol{Q_\alpha = \theta_1 \left[\frac{h_0 C_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{1/h_0}}$$

Where $\theta_i = \sum_{j=a+1}^{n} \sigma_j^{2i}$, $h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$, C_α is the normal deviate corresponding to the (1-α) percentile

- The above equation can be used for computing the threshold for a given α.

T^2 and Q Statistic

- When two statistics are utilized along with their respective thresholds, cylindrical in-control region is produced



‘o’ indicates T^2 statistic violation, ‘+’ indicates Q statistic violation

Application of PCA in fault identification

Objective

Determine which observation variables are most relevant for diagnosing the fault

Contribution Plot

- Used to identify the variables that contribute most to a specific fault
- Applied in response to a T^2 violation
- Typically, it is based on quantifying the contribution of each process variable to the individual scores
- For each variable summing the contributions only of those scores responsible for the out-of-control status

Contribution Plot - Steps

Check the normalized scores $(t_i/\sigma_i)^2$

Determine r ≤ a score responsible for out of control status

Calculate contribution to out of control

$$cont_{i,j} = \frac{t_i}{\sigma_i^2} p_{i,j}(x_i - \mu_i)$$

When $cont_{i,j}$ is negative, set it equal to zero

Calculate total contribution of the j^{th} process variable, x_j,

$$CONT_j = \sum_{i=1}^{r} cont_{i,j}$$

Plot $cont_{i,j}$ for all m variables, x_j

Where r = Residual vector

$p_{i,j} = (i,j)^{th}$ element of loading matrix

- Variables are prioritized by the total contribution values
- Only sensitive to smaller singular values

PCA approach to fault identification

- Based on quantifying the total variation of each of the variables in the residual space
- Assumption: m-a smallest singular values are all equal
- Variance of each variable inside the residual space is given by

$$\boldsymbol{\hat{s}_j^2 = \sum_{i=a+1}^{p} p_{i.j}\sigma_i^2}$$

- Given q new observations, out-of-control variable is indicated by

$$\boldsymbol{s_j^2/\hat{s}_j^2 > F_\alpha(q-a-1, n-a-1)}$$

$\boldsymbol{s_j^2}$ and $\boldsymbol{\hat{s}_j^2}$ are variance estimates for new and training set observations, respectively

$\boldsymbol{F_\alpha(q-a-1, n-a-1)}$ is (1-α) percentile limit using F distribution

PCA approach to fault identification

$$\boxed{s_j^2/\hat{s}_j^2 > F_\alpha(q-a-1, n-a-1)}$$

- The above equation is testing the null hypothesis $s_j = \hat{s}_j$
- One side alternative hypothesis = $s_j > \hat{s}_j$
- Null hypothesis is rejected if the equation holds
- In two sided hypothesis testing, the two-sided alternative hypothesis $s_j \neq \hat{s}_j$ is concluded if

$$\boxed{s_j^2/\hat{s}_j^2 > F_{\alpha/2}(q-a-1, n-a-1)} \quad \text{or} \quad \boxed{s_j^2/\hat{s}_j^2 > F_{\alpha/2}(n-a-1, q-a-1)}$$

PCA approach to fault identification

For mean

- In two sided hypothesis testing, the two-sided alternative hypothesis is concluded if

$$\frac{\mu_j - \hat{\mu}_j}{\hat{s}_j\sqrt{\frac{1}{q-a}+\frac{1}{n-a}}} > t_{\alpha/2}(q+n-2a-2)$$

or

$$\frac{\mu_j - \hat{\mu}_j}{\hat{s}_j\sqrt{\frac{1}{q-a}+\frac{1}{n-a}}} < -t_{\alpha/2}(q+n-2a-2)$$

μ_j and $\hat{\mu}_j$ are means of x_i for new and training set observations respectively

$t_{\alpha/2}(q+n-2a-2)$ is (1-α/2) percentile limit using t distribution

- The fault identification approach discussed require a group of q≫ 1 observations

PCA approach to fault identification

- Fault identification measure based on an observation vector at a single time instant is the normalized error

$$\boldsymbol{RES_j = r_j / \hat{s}_j}$$

Where $\boldsymbol{r_j}$ is the jth variable of the residual vector

- RES can be used to prioritize the variables where the variable with the highest normalized error is given priority

PCA for fault diagnosis

- One model based on the data from all fault are stacked into matrix X
- Maximum likelihood classification for an observation x is fault class *i* with the maximum score discriminant is given by

$$\boldsymbol{g_i(x) = -\frac{1}{2}(x - \bar{x}_i)^T P(P^T S_i P)^{-1} P^T (x - \bar{x}_i) + \ln p_i - \frac{1}{2} ln[det(P^T S_i P)]}$$

$$\boldsymbol{\bar{x}_i = \frac{1}{n_i} \sum_{x_j \in \chi_i} x_j}$$

$\boldsymbol{n_i}$ is the number of data points in fault class *i*, $\boldsymbol{\chi_i}$ is the set of vectors $\boldsymbol{x_j}$ which belong to fault class *i* , $S_i \in \mathcal{R}^{m \times m}$ is the sample covariance matrix for fault class *i*

PCA for fault diagnosis

- If P is selected to include all of the dimensions of the data and the overall likelihood from all fault classes is same,

Then the previous equation reduces to discriminant function for multivariate statistics

$$\boldsymbol{g_i(x) = -\,(x - \bar{x}_i)^T (S_i)^{-1} (x - \bar{x}_i) - \ln[det(S_i)]}$$

- Multivariate statistics serves as a benchmark for the other statistics
- The score discriminant, residual discriminant , and combined discriminant are used with multiple PCA models

PCA for fault diagnosis

- An observation x is classified as a fault class *i* with the maximum score discriminant

$$g_i(x) = -\frac{1}{2}(x)^T P_i \sum_{a,i}^{-2} P_i^T x - \frac{1}{2}\ln\left[det\left(\sum_{a,i}\right)\right] + \ln(p_i)$$

P_i is the loading matrix for fault class *i*,, $\sum_{a,i}$ is the diagonal matrix, $\sum_{a,i}^{2}$ is the covariance matrix, $\mathbf{p_i}$ is the overall likelihood of fault class *i*

Note: Assumption – observation vector x has been auto scaled according to mean and SD of the training set for fault class *i*

Weakness: Useful information for other classes is not utilized when each model is derived

PCA for fault diagnosis

- The previous equation reduces to

$$\boldsymbol{T_i^2} = (\boldsymbol{x})^T \boldsymbol{P_i} \sum_{\boldsymbol{a,i}}^{-2} \boldsymbol{P_i^T x}$$

Assumptions: Overall likelihood for all fault classes and sample covariance matrix for all class is the same

- **Residual discriminant** represents the observation if important variations in the discriminating between the faults are assumed contained in the residual space

$$\boldsymbol{Q_i/(Q_\alpha)_i}$$

- **Combined discriminant** represents the observation if the important variations are contained both within the score and residual space

$$\boldsymbol{c_i}\left[\boldsymbol{T_I^2}/\left(\boldsymbol{T_\alpha^2}\right)_{\boldsymbol{i}}\right] + (\boldsymbol{1} - \boldsymbol{c_i})[\boldsymbol{Q_i}/(\boldsymbol{Q_\alpha})_{\boldsymbol{i}}]$$

Where c_i is a weighting factor between 0 and 1 for fault class *i*

PCA for fault diagnosis

- When a fault is diagnosed as fault *i*, it is not likely to represent a new fault when

$$\left[T_I^2/\left(T_\alpha^2\right)_i\right] \ll 1$$

$$[Q_i/(Q_\alpha)_i] \ll 1$$

- **If either of these conditions is not satisfied, it is likely that the observation represents a new fault**
- It is recommended to assess the likelihood of successful diagnosis before application of pattern classification

Similarity index = $f = \frac{1}{m}\sum_{j=1}^{m} \tilde{\sigma}_j$ is used to quantify similarity between the covariance structures (Range: 0-1)

Where $\tilde{\sigma}_j$ is the j^{th} singular value of $V_1^T V_2$, V_1 and V_2 contain all m loading vectors for classes 1 and 2 respectively

Reduction order and PLS prediction

- In fault diagnosis, the reduction order is determined by the value that minimizes the information criterion

$$\boldsymbol{f_m(a) + \frac{a}{\tilde{n}}}$$

Where $\tilde{\boldsymbol{n}}$ is the average number of observations per class, $\boldsymbol{f_m(a)}$ is the misclassification rate for the training set

- Misclassification rate $= \frac{No. of\ incorrectly\ assigned\ classes}{Total\ \ no. of\ classifications\ made}$

$$\textbf{Estimated score vector } (\hat{\boldsymbol{t}}_{\boldsymbol{j}}) = \boldsymbol{E_{j-1} w_j}$$

$$\textbf{Matrix residual } (\boldsymbol{E_j}) = \boldsymbol{E_{j-1} - \hat{t}_j q_j^T}$$

$\mathbf{q_j}$ is the loading vector

Reduction order and PLS prediction

- Prediction of the predicted block $Y_{train2,a}$ of the training set using PLS1 with a PLS components

$$\boldsymbol{Y_{train2,a} = F_j = \sum_{j=1}^{a} b_j \hat{t}_j q_j^T}$$

or

$$\boldsymbol{Y_{train2,a} = XB2_a}$$

b$_{\mathbf{j}}$ is the regression coefficient, **B** is the diagonal regression matrix

- Prediction of the predicted block $Y_{train1,a}$ of the training set using PLS1 with a PLS components

$$\boldsymbol{Y_{train1,a} = [Y_{train1,a}\ Y_{train2,a} \cdots \cdots \cdots Y_{trainp,a}]}$$

$$\boldsymbol{y_{traini,a} = f_{i,j} = \sum_{j=1}^{a} b_{i,j} \hat{t}_i q_{i,j}}$$

Fault detection, identification and diagnosis using PLS

- All fault detection, identification and diagnosis techniques for PCA can be applied for PLS

Overestimation: When the element of Y_{train} for an in-class member >1 or element Y_{train} for non-class member is >0

Underestimation: When the element of Y_{train} for an in-class member <1 or element Y_{train} for non-class member is < 0

Taking underestimation and overestimation of Y into account into a second cycle of PLS algorithm (NIPALS) is done if some of the elements are underestimated while others are overestimated

THANK
YOU!