2018 UKSim-AMSS 20th International Conference on Modelling & Simulation

Application of Principle Component Analysis in Resolving Influential Factor Subject to Industrial Motor Failure

Nurfatihah Syalwiah Binti Rosli
nurfatihah_g03577@utp.edu.my

[a] Rosdiazli Bin Ibrahim, [b] Idris Ismail
[a]rosdiazli@utp.edu.my, [b]idrisim@utp.edu.my

Department of Electrical & Electronics Engineering
Universiti Teknologi PETRONAS
32610 Seri Iskandar, Perak, Malaysia

***Abstract* - Predictive maintenance is very important towards industrial economy by improving equipment efficiency, reliability and reducing downtime. In recent years, abundant of data of rotating equipment is readily available from various sources. However, these data are not being utilized and analyzed for improving maintenance performance. This requires advanced techniques to analyze a variety of data in order to transform into relevant information. Most problems with a lot of parameters involved were not being specific to analyze the contribution of motor failure. Therefore, this research proposed an efficient data analysis using Principle Component Analysis (PCA) in determining the most influential factor to the failure of the industrial motor. The result will show the parameters that influence the motor failure. This finding can be used as a guideline for predictive maintenance in order to mitigate the risk of the plant shutdown.**

Keywords - PCA; influential factor; predictive maintenance; industrial motor.

I. INTRODUCTION

Recent emergence of the Internet of Things (IoT) concept has changed the trend of development in industry technology. Nowadays, industry 4.0 concept is applied for improving industrial communication and automation in order to create smart products and services [1]. In the context of industry 4.0, new mechanism is needed to accomplish the assigned task effectively. Therefore, all information from devices and equipment must communicate to configure and manage network resources. For Oil and Gas processes, information on every equipment within the plant is critically considered in order to ensure high reliability and performance through smart maintenance. The continuous attention to equipment failure from operator and engineer are very crucial for identification and reduction of downtime of plants.

Rotating equipment being essential element of industrial process, are a common sight in applications such as industrial fans, pumps, blowers, power and machine tools, turbines, compressors, movers, mills etc. The advantages of these rotating equipment are low acquisition and maintenance costs, high reliability, and easily adaptable to several load conditions. However, these equipment are subject to mechanical and electrical failures. For instance, failure of motors can cause complete system downtime if they are located in critical place within the plant. Example of electrical faults that may lead to failure are winding insulation problem and broken rotor bars [2]. On the other hand, mechanical faults involve, misalignment of shaft, air gap eccentricity and bearing faults [3]. These faults occur commonly due to current or voltage unbalance, harsh operating condition, prolonged activity times, and so on [4]. These failures may lead to significant economic losses. For example, capital equipment in semiconductor industry suffered 7% of scheduled maintenance that cost USD 100,000 for each downtime [5]. Bevilacqua [6] reported that the cost of maintaining equipment is between 15 to 70 % of total production cost depending on the type of industry. In view of aforementioned, adequate prognostic methods for rotating equipment maintenance are required for any detection of faults and to prevent interruptions in production.

Figure 1 shows the one of the example sources of failure statistics of AC Motors in industry. Bearings fault resulted in up to 41% of the total cases. This is followed closely by stator faults with 37% [7].

In recent years, in order to detect equipment failure before occurrence, the industry utilizes Predictive Maintenance (PdM) approach. This is to reduce the overall cost of maintenance. The main aim of PdM is to forecast when the equipment will experience failure and to prevent its occurrence by performing maintenance ahead. Ideally, possible failures are predicted and prevented without incurring so much maintenance cost. Thus, PdM emerges as a solution to the production inefficiency.

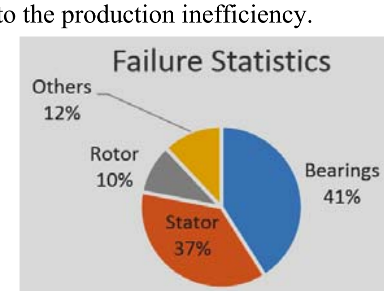

Figure 1: Failure Statistics of AC Motor [7]

0-7695-6405-4/18/$31.00 ©2018 IEEE
DOI 10.1109/UKSim.2018.00016

Despite the advantages of PdM, it is however faced with the limitation of sufficient technology to analyze huge amount of data from various sources in order to extract information without knowing the condition of the process. Data Analytics (DA), being relevant to the analysis of big data, has been incorporated to predictive maintenance of rotating equipment. This is to give early warning of failure and to reduce the impact and cost of abnormal situations [8]. The DA is a process of examining data to uncover the abnormalities or problem associated with the data, in order to draw conclusions based on meaningful ways of data exploration. In the next generation of industry 4.0, "big data" has been highlighted by many companies that have millions of values in their databases. These big data required an advanced computer solutions and effectively mine data in order to solve millions data point and complex problem [9].

The 4th generation maintenance is faced with some challenges for example, large amount of data available to be used for automated data analysis. This becomes impediment to providing robust maintenance system. As reported by Bevilqua and Braglia [10], predictive maintenance is the best practical maintenance strategy used in industry to predict equipment failure and Remaining Useful Life (RUL). However, low quality level of analysis and prognostics affect the technical features of PdM. These limitations of PdM gave impact toward maintenance performance. Thus, to reduce economic losses, industry needs to improve its unplanned downtime. This can be achieved by developing failure classification or using data to predict failure and RUL for each equipment [10].

The function of data analytics is to extract information from multiple sources to be further used for the purpose of prediction and/or decision making. To be compared with traditional data analytics, the massive and multiple sources characteristic of big data give challenges to achieving the goal of data mining and knowledge discovery [11]. Therefore data analytics using PCA is proposed in solving the problem of reducing high dimensionality of the data in order to get more precise and accurate prediction.

The rest of this paper is organized as follows: Section II discusses about PCA, while Section III presents the methodology of applying data analytic in order to determine the most influential factor for motor failure. The last two sections present the results and conclusion respectively.

II. PRINCIPLE COMPONENT ANALYSIS

The development of PdM includes the process of feature extraction and prognostics of prediction model. A significant research has been reported in literature to select feature extraction method and develop prognostics model [12], [13]. The purpose of Feature Extraction (FE) is to reduce dimensionality of data. When large datasets and multi-dimensional operating space have been used in model development, it is often faced a problem known as curse of dimensionality [14]–[16]. FE extracts a subset of new or reduced features from the original feature or large set of data. It keeps data to be informative, non-redundant and sufficiently accurate [17]. Several survey papers have been reported to compare the effectiveness of the feature extraction techniques in dimensionality reduction [12]. These surveys classified feature extraction techniques into two: linear and non-linear methods. Linear method such as Principle Component Analysis, Independent Component Analysis, Linear Discriminant Analysis [18], [19] and Fisher Discriminant Analysis [20], [21] have been widely used for dimensional reduction. PCA is being chosen as its low noise sensitivity and lack of redundancy of data given by its components [22].

The following steps shows the application of PCA algorithm in MATLAB:

Step 1: Centre and Standardize: The method begins calculating z-scores of input. It removes the influence location and scale from variables in raw data. The calculation of z-scores using MATLAB is written as Equation (1) below:

$$B = \text{zscore}\ (A) \qquad (1)$$

Step 2: Covariance: The second step will produce covariance matrix of centered and standardized data calculated in Step 1. Covariance is getting from the Eigen value of previous matrix as Equation (2):

$$C = \text{eig}(\text{cov}(B)) \qquad (2)$$

III. RESEARCH METHODOLOGY

To determine the influential parameters that contribute to industrial motor failure, the PCA techniques was applied with the following methods:

1. Data collection: Data is collected from 6th Dec 2015 until 19th July 2017 with one-minute interval. The total data is about 807840. There are14 parameters involved in investigating the most influential factor of industrial motor failure which are: Air To Plant (ATP), Air Booster Discharge (ABD), Current, Power, RTD1-5 Temperature, Hot Air Temperature (HAT), Cold Air Temperature (CAT), Voltage, Pressure Discharge (PD) and Temperature Discharge (TD).
2. Data filtering: All the outliers and missing data is filtered in order to remove the insignificant value that can affect the correlation between parameters [23-24]
3. Data normalization: All data is being processing by normalize the mean and unit variance. The data is scaled between [-1 1] in order to have similar impact of vary data with different range. Therefore, the input data have same level of data distribution [25].

4. Input all data into PCA function: All data is tested using PCA in order to find the principle components and explained variance. In this research, the raw data is ⟦X ∈R⟧^(m ×n) where m is number of samples or observations and n is the number of parameters [26].
5. Determine the most influential parameters subject to industrial motor: Output from PCA such as value of components, explained variance and coefficient can be used to determine the correlation of parameters. While eigenvectors value will explain the variability of total variance. The results is presented based on actual data given by industry [27].

IV. RESULTS AND DISCUSSIONS

A. Data Collection and Data Filtering

The initial stage is data gathering for model development and it was taken one minute interval data from December 2015 until July 2017. These data have been divided into two categories, which are healthy and unhealthy data. The unhealthy data has been further categorized into 3 types of common faults as shown in Table I.

TABLE I. PERCENTAGE OF FAULT DATA

Type of Fault	**Description of Fault**	**Percentage of Fault (%)**
Shutdown	Readings read zero value when the plant is shutdown	0.97
Outliers	The data lies in abnormal distance from other values	0.20
Out of range	The readings value is outside the normal operating range of the process parameter	4.50

TABLE II. PERCENTAGE OF FAULT DATA

PC#	Eigen value	Explained variance (%)	Cumulative
1	6.5495	46.7826	46.7826
2	3.4568	24.6918	71.4744
3	1.1974	8.5533	80.0277
4	0.9946	7.1048	87.1326
5	0.9113	6.5093	93.6420
6	0.4047	2.8911	96.5331
7	0.1504	1.0747	97.6079
8	0.1210	0.8645	98.4724
9	0.1104	0.7886	99.2611
10	0.0582	0.4159	99.6771
11	0.0230	0.1645	99.8416
12	0.0210	0.1507	99.9923
13	0.0008	0.0059	99.9983
14	0.0002	0.0016	100

The unhealthy data was figured out and filtered out by putting upper and lower range of each parameter. In order to have an optimized model, good historical data is needed from the well-functioned instrument. Therefore all fault data is removed in order to get more robust prediction model. The total data available is about 807840 used for data analysis. There are 14 parameters involved in investigating which parameter contribute to the shutdown or motor failure.

From the Table I above shown that most fault happen is out of range. This fault data can be further analyzed in order to know the most influence parameter that lead to the failure of the system. However, the fault data for each parameter has been discovered as shown in Figure 2. The parameter of HAT gives the highest percentage of fault data compared to the other parameters. HAT value is the input of motor as its parameter as indicator of the capacity to discharge cold air for the system. The failure of this data showed the problem in motor process.

TABLE III. PERCENTAGE OF FAULT DATA

PC#	Group of Parameters	PC value
1	RTD1	0.37
	RTD2	0.38
	RTD3	0.38
	RTD4	0.35
	RTD5	0.36
	HAT	0.33
	CAT	0.38
2	ATP	0.46
	ABD	0.50
	CURRENT	0.48
	POWER	0.50
3	CAT	-0.33
	TD	0.87
4	VOLTAGE	0.80
	TD	0.48

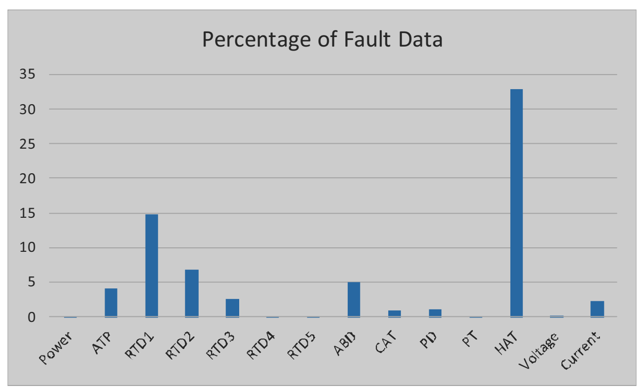

Figure 2: Percentage of Fault Data

B. Data Analysis using PCA

Data from actual plant in one industrial company in Malaysia are analyzed using PCA. Based on Table II, number of principle components produced are 14 because of number of parameters involved in the analysis. Refer to Table II, the total of these eigenvalues is 14. The percentage of explained variance is calculated by dividing the eigenvalues with total variance. For example, first Eigen value, 6.5495 divided by 14 equals to 0.4678 or about 46.78% of variation. The cumulative percentage is the successive proportions of explained variance. Therefore, to determine how many principle components should be considered is based on the drop of eigenvalue, or explained variance. For instance, subtracting the second eigenvalue 3.4568 from the first Eigen value, 6.5495 gives 3.0927. While the difference between the second and third value is 2.2594. The subsequent differences are even smaller. The first four principle components explain 87% of the variation. This is an acceptably large percentage to be considered for the correlation analysis.

The correlation is obtained from the principle component interpretation. Due to standardization, all principle components will have men 0 and standard deviation is the square root of Eigen value. Based on the principle components interpretation, the numbers are large in magnitude either positive or negative direction are being chosen as the most variables correlated with each component. Based on Table III, the level of correlation value is determined based on the larger value from each component. For first principle component analysis PC#1, seven of the original parameters are strongly correlated. When one of these parameters increase, the other parameters also increase. Whereas, second principle component analysis PC#2, increasing with one of the values will increase the other value. The PC#3 components, decreasing the Hot Air Temperature will increase the Temperature Discharge. Lastly, PC#4 increases with increasing Voltage and Temperature Discharge.

V. CONCLUSION

Predictive Maintenance can be improved its accuracy and function by adding analytics algorithm to predict the failure based on the most influential parameters to the equipment failure. Factors that contribute to the industrial motor failure are successfully determined using PCA techniques. The resolved of this influential parameters is beneficial for plant management and can used for the purpose of predictive maintenance. In future, the implementation of the technique can be applied in actual plant to validate its accuracy.

ACKNOWLEDGMENT

The authors wish to thank PETRONAS Utility Kertih (UK), Petronas Chemical Sdn Bhd (PETLIN) for research collaboration and data utilization and members and staffs of Universiti Teknologi PETRONAS for the continuous support of assistantship and facilities.

REFERENCES

[1] J. Wan et al., “Software-Defined Industrial Internet of Things in the Context of Industry 4.0,” IEEE Sens. J., vol. 16, no. 20, pp. 7373–7380, 2016.

[2] R. S. Kraleti, M. Zawodniok, and S. Jagannathan, “Model based diagnostics and prognostics of three-phase induction motor for vapor compressor applications,” PHM 2012 - 2012 IEEE Int. Conf.on Progn. Heal. Manag. Enhancing Safety, Effic. Availability, Eff. Syst. Through PHM Technol. Appl. Conf. Progr., 2012.

[3] S. Karmakar, S. Chattopadhyay, M. Mitra, and S. Sengupta, “Induction Motor Fault Diagnosis,” no. i, 2016.

[4] G. H. Bazan, P. R. Scalassara, W. Endo, A. Goedtel, W. F. Godoy, and R. H. C. Palácios, “Stator fault analysis of three-phase induction motors using information measures and artificial neural networks,” Electr. Power Syst. Res., vol. 143, pp. 347–356, 2017.

[5] S. Munirathinam and B. Ramadoss, “Big data predictive analtyics for proactive semiconductor equipment maintenance,” Proc. - 2014 IEEE Int. Conf. Big Data, IEEE Big Data 2014, pp. 893–902, 2015.

[6] M. Bevilacqua and M. Braglia, “The analytic hierarchy process applied to maintenance strategy selection,” Reliab. Eng. Syst. Saf., vol. 70, no. 1, pp. 71–83, 2000.

[7] E. H. Mayoral et al., “World â€TM s largest Science , Technology & Medicine Open Access book publisher Fourier Analysis for Harmonic Signals in Electrical Fourier Analysis for Harmonic Signals in Electrical Power Systems Power Systems.”

[8] Z. Rawi, “SPE 128559 Machinery Predictive Analytics,” no. March, pp. 23–25, 2010.

[9] A. Brodsky, G. Shao, M. Krishnamoorthy, A. Narayanan, D. Menasce, and R. Ak, “Analysis and optimization in smart manufacturing based on a reusable knowledge base for process performance models,” Proc. - 2015 IEEE Int. Conf. Big Data, IEEE Big Data 2015, pp. 1418–1427, 2015.

[10] M. Kans and D. Galar, “The Impact of Maintenance 4.0 and Big Data Analytics within Strategic Asset Management,” … Perform. Meas. Manag. 28 …, no. November 2016, pp. 96–103, 2017.

[11] C.-W. Tsai, C.-F. Lai, H.-C. Chao, and A. V. Vasilakos, “Big data analytics: a survey,” J. Big Data, vol. 2, no. 1, p. 21, 2015.

[12] Z. Ge, Z. Song, and F. Gao, “Review of Recent Research on Data-Based Process Monitoring,” Ind. Eng. Chem. Res., vol. 52, no. 10, pp. 3543–3562, 2013.

[13] L. Liao and F. Köttig, “Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction,” IEEE Trans. Reliab., vol. 63, no. 1, pp. 191–207, 2014.

[14] E. Zorarpacı and S. A. Özel, “A hybrid approach of differential evolution and artificial bee colony for feature selection,” Expert Syst. Appl., vol. 62, pp. 91–103, 2016.

[15] R. S. Blasone, J. A. Vrugt, H. Madsen, D. Rosbjerg, B. A. Robinson, and G. A. Zyvoloski, “Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling,” Adv. Water Resour., vol. 31, no. 4, pp. 630–648, 2008.

[16] M. S. Kan, A. C. C. Tan, and J. Mathew, “A review on prognostic techniques for non-stationary and non-linear rotating systems,” Mech. Syst. Signal Process., vol. 62, pp. 1–20, 2015.

[17] F. Boldt, T. Rauber, and F. Varejão, “Feature Extraction and Selection for Automatic Fault Diagnosis of Rotating Machinery,” Lbd.Dcc.Ufmg.Br, vol. 1, pp. 1–8, 2013.

[18] W. Klecka, “Discriminant Analysis,” Adv. Neural Inf. Process. Syst., vol. 17, no. 60, pp. 1569–1576, 1980.

[19] V. T. Tran and B. S. Yang, “An intelligent condition-based maintenance platform for rotating machinery,” Expert Syst. Appl., vol. 39, no. 3, pp. 2977–2988, 2012.

[20] Z. Li, R. Outbib, S. Giurgea, and D. Hissel, “Diagnosis for PEMFC Systems: A Data-Driven Approach With the Capabilities of Online Adaptation and Novel Fault Detection,” IEEE Trans. Ind. Electron., vol. 62, no. 8, pp. 5164–5174, 2015.

[21] A. P. E, E. D. Dan, and O. I. Sidney, “A Review of the Limitations of Some Discriminant Analysis Procedures in Multi-Group Classification .,” vol. 5, no. 9, pp. 199–204, 2015.

[22] S. Karamizadeh, S. M. Abdullah, A. A. Manaf, M. Zamani, and A. Hooman, “An Overview of Principal Component Analysis,” J. Signal Inf. Process., vol. 4, no. 3, pp. 173–175, 2013.

[23] G. Shao, S. J. Shin, and S. Jain, “Data analytics using simulation for smart manufacturing,” Proc. - Winter Simul. Conf., vol. 2015–Janua, no. Smlc 2012, pp. 2192–2203, 2015.

[24] X. Dai and Z. Gao, “From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis,” IEEE Trans. Ind. Informatics, vol. 9, no. 4, pp. 2226–2238, 2013.
[25] A. K. Ronay and R. Bhinge, “Data analytics and uncertainty quantification for energy prediction in manufacturing,” Proc. - 2015 IEEE Int. Conf. Big Data, IEEE Big Data 2015, pp. 2782–2784, 2015.
[26] A. Cubillo, S. Perinpanayagam, and M. Esperon-Miguez, “A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery,” Adv. Mech. Eng., vol. 8, no. 8, p. 168781401666466, 2016.
[27] “P-RMPDC - Potential Cost Compression rev1.” .