Integrated Frame Work for Data Driven Process Monitoring and Diagnosis System using Machine Learning and Cloud Computing

Yash Kansagara[1], Professor V.P.Patel[2]

[1]*Student, Dept. of Instrumentation and control Engineering, L.D. College of Engineering, Ahmedabad, India*
[2]*Professor, Dept. of Instrumentation and control Engineering, L.D. College of Engineering, Ahmedabad, India*

---***---

Abstract— *Today's manufacturing processes are much complex and high efficient. Because most of process are work on automation and artificial intelligence system and for that user and developers needs to required monitoring or diagnosis of that system. In given Paper we use two system for monitoring of process and diagnosis of that processes. 1. Machine learning 2. Cloud computing. Process monitoring is an assessment of the process of intervention. In this, by using ML programing Developers are monitoring process program.*

Also diagnosis meaning is find the fault and error from that process Machine learning is an application of AI which provides systems the ability to automatically learn and improve from experience without being explicitly. Programmed. Machine learning focuses on the development of computer programs that when system getting unstable during that time change in parameter by use of program. User & developers can also monitoring process by the use of cloud computing technique. Cloud Computing has emerged as new paradigm of Automation technology. In this dissertation user and developers are also monitoring process and diagnosis the process using through cloud computing.

Keywords— *SVM, PCA, RFE Etc*.

1. INTRODUCTION

A combined measure of the original Support Vector Machine (SVM) and Principal Component Analysis (PCA) is provided to carry out fault classification, and compare its result. With SVM-RFE (Recursive Feature Elimination) method. RFE method is used for feature extraction, and PCA is utilized to project the original data onto a lower dimensional space.

PCA *T*2, SPE statistics, and original SVM are proposed to detect the faults. Some common and basic faults of the Tennessee Eastman Process (TEP) are analysed in terms of the practical system and reflections of the dataset. PCA and RFE can effectively detect and diagnose these common faults. In Recursive feature Elimination algorithm, all variables are in decreasingly ordered according to their contributions.

The classification accuracy rate is improved by choosing a reasonable number of features. Based on the literature review, it's observed that in the paper there output and the work is mainly focus on some of techniques. Many more techniques are available for get system stable as well as process fault detection and monitoring.

Here we are purpose an efficient solution of system monitoring and diagnosis by using of classification techniques with deferment Machine learning algorithms. And for that we look for Some algorithms like – Support Vector Machine, Artificial Neural Network, K nears Neighbourhood, Decision Tree, Principal component analysis, Recursive Feature Elimination. We select SVM, PCA, and RFE of the algorithms for further work.

2. PROPOSED METHOD

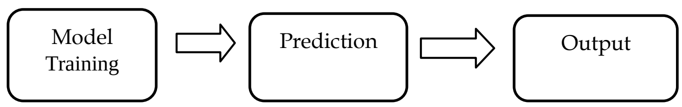
2.1 Method –

Support Vector Machine

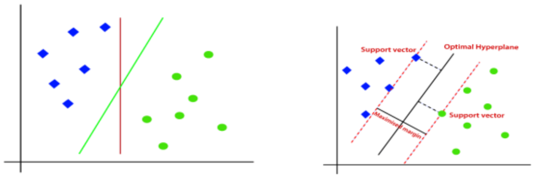
Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

Type of SVM --- (1) Linear

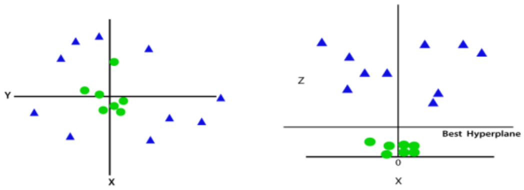
(2) Non linear



Block diagram



Linear SVM



Non Linear

(2) Recursive Feature Elimination

RFE arithmetic is to calculate the sorting coefficient at first according to the weight vector *w*, which is got from the SVM model training process, then take out the feature variable which has minimal sorting coefficient in each iteration, and get the descending order of feature variables. The classical SVM-RFE is based on the linear kernel function, while, in nonlinear case, the RBF is used as the kernel.

In every iteration, the feature variable with minimal sorting coefficient will removed. The new sorting coefficient will obtained by using Support vector Machine to train the rest of the feature variables by executing this all process iteratively, a feature ranking list is got. According to this ranking list, the relative degree for each feature with the category will be known. The quality of these subsets can be assessed by the classification accuracy. So, the optimal feature subset can be obtained. To TE process, such a feature ranking list is got for each type of fault.

Here A, B, C, D, E are Normal Attributes.
And T is Target attributes.

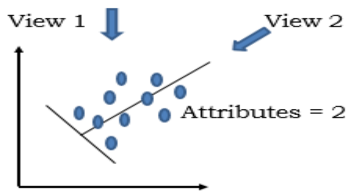
A	B	C	D	E	T

Recursive Feature Elimination
Now select (A & T)→ + → M1 (model)
(B & T)→ + → M2 (model)
(C & T)→ + → M3 (model)
.
(A, B, C, D, E & T)→ + → M24 (model)
After finding M1, M2 up to M24 find model from then which have god accuracy.

(3) Principal Component Analysis

PCA was proposed by Pearson in 1901. PCA aims at DE correlating the industrial process data. PCA is highly correlated, multivariate, and reducing the dimension of the data. In order to reduce dimension and keep more information of the original image at a same time, a low dimension data model needs to be established and the variance of the data reaches the maximum value through the projection. After dimensional reduction, the noise and redundant information in the original higher dimensional space can be eliminated.

Because of this the error can be reduce and the recognition accuracy will be improve. To realize the purpose of improving the calculation speed, PCA statistics are used as metrics to determine whether the sample are faulty or not.



2.2 Process System : Tennessee Eastman

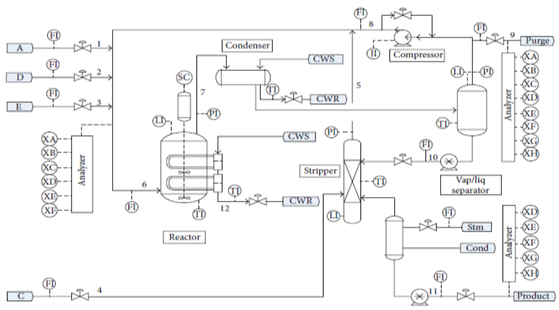


Fig. TE Process

The TE process was created by the Eastman Chemical Company. This Process provide a realistic industrial process for evaluating process control and also monitoring methods. It was firstly proposed by Downs and Vogel on the AICHE (American Institute of Chemical Engineers). It has been extensively used in the aspects of control and optimization, process control, fault diagnosis, statistical process monitoring, data driven, and so on.

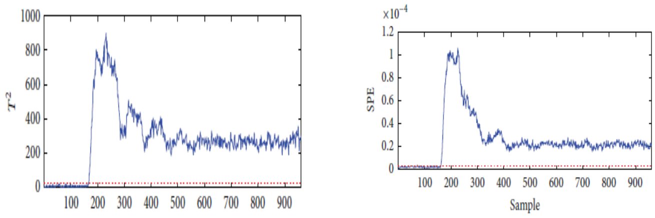
The Tennessee Eastman plant mainly consists of five units. There are 8 components. A, C, D, and E are gaseous materials. B is an inert substance. F is a liquid reaction byproduct. G and H are liquid reaction products. The reaction of H has a lower energy than that of G. G has a higher sensitivity to temperature.

Fault No.	Fault	Fault Type
1	A/C Feed ratio	Step
2	B composition	Step
3	D feed Temperature	Step
4	Reactor Cooling Water Inlet Tem.	Step
5	Condenser Cooling Water Inlet Tem.	Step
6	A Feed Loss	Step
7	C Header Pressure Loss	Step
8	A,B,C Feed Composition	Random Variation
9	D feed Temperature	Random Variation
10	c feed Temperature	Random Variation
11	Reactor Cooling Water Inlet Temp.	Random Variation
12	Condenser Cooling Water Inlet Temp.	Random Variation
13	Reaction kinetics	Slow Drift
14	Reactor Cooling Water valve	Sticking
15	Condenser Cooling Water valve	Sticking
16	Unknown	N/A
17	Unknown	N/A
18	Unknown	N/A
19	Unknown	N/A
20	Unknown	N/A

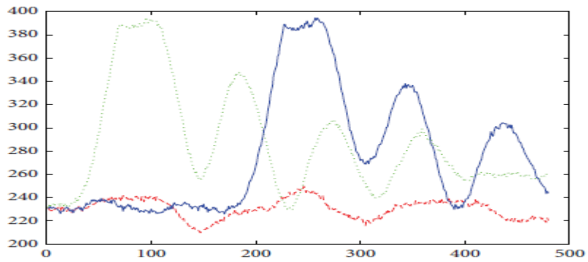
Due to the presence of a catalyst, the gaseous reactants will become liquid product when it enters into the reactor. This catalyst is permanent and soluble in liquid. The reactor has an internal condenser. It is used to remove the heat generated by the reaction. Along with the component which is not completely reacted, the product leaves the reactor in the form of steam. The product comes to the gas-liquid separator through a condenser.

3. Experiment and Result

Fault 1:-



PCA Result

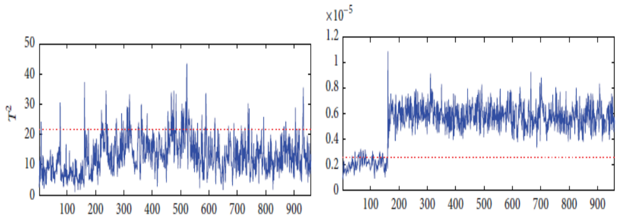


PE : Plot For result

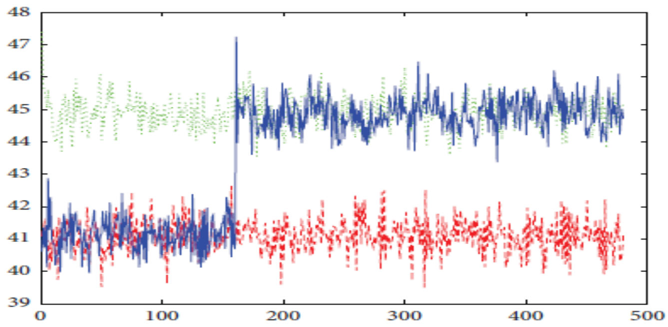
-- Normal Data ... Train Data _ Test Data

- Fault 1 is the ratio of A/C feeds takes a step change. But, under normal circumstances, variable 19 does not have such a change, which results in the occurrence of the fault. The training samples also have a step change, so SVM model can learn this change through the training samples which results in that the fault samples can be detected.
- In PE plot we can show there are three type of waveform are there. For normal database, for Train databased and f Or Test Data so based o that we can easily detect.

Fault 4:-



PCA Result



PE: Plot For result

- - Normal Data ... Train Data _ Test Data

- Fault 4 is the temperature step changes of the reactor cooling water inlet. When the fault happens, that temperature increases quickly, but the standard Deviation of the other variables is similar to that under Normal conditions. This enhances the difficulty of fault detection.
- Value of variable was maintained between 44 and 46, while the normal dataset does not have a step change, and its normal value is between 40 and 42. For the training samples of fault 4, the value ranges between 44 and 46 and is different from the normal values.
- The PCA-based statistics are utilized for fault4 detection; After 160 sets of normal samples, the value of statistics appears a step variety.

4. CONCLUSIONS

- Process monitoring, specifically fault detection and diagnosis, is one of the major fields in process systems engineering that benefits the advances in data-driven modeling and dimensionality reduction techniques with the increased availability of process data.
- In this Presentation, we present theoretical advances in the feature selection algorithm based on Support Vector Machines, describe a data driven framework for fault detection and diagnosis in continuous processes, and finally apply it to the Tennessee Eastman benchmark process.
- Depending on the method of PCA-SVM and PCA-RFE, these Common fault in the TEP are effectively diagnosed. The original SVM method is also used to detect the fault. The effect is good. According to RFE the most relevant variables to each fault has been found.

5. REFERENCES

1. Pham, D.T. and Chan, A.B., (1998), "Control Chart Pattern Recognition Using a New Type of Self-Organising Neural Network", Proc. Instn. Mech. Engrs. Vol. 212, No.1, pp. 115 - 127.
2. Pham, D.T. and Oztemel, E., (1994), "Control Chart Pattern Recognition Using Learning Vector Quantization Networks", International Journal of Production Research, Vol. 32, No.3, pp. 203-212.
3. Alcock, R.J. and Manolopoulos, Y., (1999), "Time-SeriesSimilarity Queries Employing a Feature-Based Approach", In 7th Hellenic Conference on Informatics, Greece.
4. Liao, T.W., (2005), "Clustering of time series data - a survey", Pattern Recognition, 38(11), 1857-1874.
5. Kalpakis, K., Gada, D. and Puttagunta, V., (2001), "Distance measures for effective clustering of ARIMA time-series" Proceedings of 2001 IEEE International Conference on Data Mining, San Jose, CA, pp. 273-280.
6. Warren Liao T., Chi-Fen Ting and Pei-Chann Chang (2006), "An adaptive genetic clustering method for exploratory mining of feature vector and time series data", International Journal of Production Research, Vol. 44, No. 14, 2731-2748
7. Hayashi, A., Mizuhara, Y., & Suematsu, N. (2005), "Embedding Time Series Data for Classification", Machine Learning and Data Mining in Pattern Recognition, pp. 356-365.
8. Laizhong Cui, Shu Yang and et al. "A survey on application of machine learning for Internet of Things". International Journal of Machine Learning and Cybernetics, August 2018, Vol. 9, Issue 8, pp 1399–1417.
9. Luigi At zori, Antonio Iera, Giacomo Morabito. "The Internet of Things: A survey". Computer Networks, 2010. Vol. 54, Issue 15, pp 2787-2805, ISSN 1389-1286.
10. Chen, Feng and et al. " Data Mining for t he Internet of Things: Literature Review and Challenges." International Journal of Distributed Sensor Networks, August 2015.
11. Mohammad Saeid Mahdavinejad and et al. "Machine learning for internet of things dat a analysis: a survey". Digital Communications and Networks, 2018.Vol. 4, Issue 3, pp 161-175,ISSN 2352-8648.
12. Hyo-Sik Ham, Hwan-Hee Kim, Myung-Sup Kim, and Mi-Jung Choi, " Linear SVM-Based Android Malware Detect ion for Reliable IoT Services," Journal of Applied Mathematics, 2014. Art icle ID 594501
13. Khanna, Divyansh, et al. "Comparative study of classification techniques (SVM, logistic regression and neural networks) to predict the prevalence of heart disease." International Journal of Machine Learning and Computing. 2015. Vol. 5, Issue 5, pp 414.
14. J. Bughin, M. Chui, and J. Manyika. Clouds, big data, and smart assets: Ten tech-enabled business trends to watch, Aug. 2010.