Available online at www.elixirpublishers.com (Elixir International Journal)

Control Engineering

Elixir Control Engg. 61 (2013) 17079-17082

ISSN: 2229-712X

Fault diagnosis in benchmark process control system using quantitative process history based statistical techniques

Tarun Chopra

Department of Electrical Engineering, Govt. Engineering College Bikaner, India-334004.

ARTICLE INFO

Article history:

Received: 11 June 2013;

Received in revised form: 24 July 2013;

Accepted: 6 August 2013;

Keywords

Fault Diagnosis,
Benchmark study,
Control Valves.

ABSTRACT

Fault diagnosis research deals with real world problems in terms of plant efficiency, maintainability and reliability. Fault in process and manufacturing industries are crucial in order to improve production efficiency, quality of the product and the cost of production. This work is focused on the development of a fault diagnosis method for application to an actuator benchmark specifically designed for industrial fault diagnosis. The performance of the proposed approach is demonstrated on the DAMADICS benchmark problem.

Introduction

With the increased complexity of industrial systems and the high level of process quality, reliability and safety requirements the automation of diagnostics is warranted in order to make it possible to determine the location, reason and time of the fault precisely; whereas earlier in the case of simple technical systems, human inspection was enough[1].

Real processes are usually dynamic, non-linear and stochastic. In such cases, prompt fault diagnosis requires accurate models of processes and for this interrelations between process parameters should be apparent.

Statistical measures like Correlation and principal factor analysis, Clustering and Discriminant analysis are examples of unsupervised learning. These measures aid in looking for patterns, groupings or other ways to characterize the data that may lead to understanding of the way the data interrelates in Complex Systems [2].

In this paper ,the above mentioned Statistical techniques have been chosen as tool for analyzing the information related to fault diagnosis in DAMADICS Benchmark Process Control System .This Benchmark is concerned with applications of diagnosis methods on chosen actuators in 5-stage evaporisation plant used in the Lublin sugar factory, Poland [3].

State of Art

Correlation and principal factor analysis is used for exploratory analysis where the nature and relationship between variables is to be understood. Correlation is a measure of the association between two variables. The purpose of Factor Analysis is to identify a set of underlying factors that explain the relationships between correlated variables[4]. K-Means clustering, one of the older predictive modeling methods, was developed during 1975 - 1977 by Hartigan and Wong[5]. The basic idea of K-Means clustering is that clusters of items with the same target category are identified, and predictions for new data items are made by assuming they are of the same type as the nearest cluster center.

Discriminant Analysis was originally developed in 1936 by R.A. Fisher [6].

This analysis finds a linear transformation ("discriminant function") of the predictors, which yields a new set of transformed values that provides a more accurate discrimination than either predictor alone.

Methodology

During the dynamic sugar production process, there is a possibility that there will occur faults in the actuator valve block, with different types of strength i.e. abrupt {small, medium, big} and incipient. The early diagnosis (detection and isolation) of those faults minimize damages in the industrial line.

The dataset used for this case study have been generated in this study by employing the MATLAB-SIMULINK model of the actuator as shown in Fig 1.

In accordance with the scope of the defined objective for this work, only the data related to fault categories F7 (medium evaporation or critical flow), F12 (electro-pneumatic transducer fault), F13 (stem displacement sensor fault), F15 (positioner spring fault) have been considered.

Results

The results after applying the above mentioned three techniques have been summarized as follows:-

Correlation and Principal Factor Analysis

The summary of variables has been presented in Table 1 and results have been tabulated in Tables 2-5.

Project Parameters

Target variable: Type of Fault
Number of predictor variables: 6

Input Data

Number of variables (data columns): 7
Data subsetting: Use all data rows
Number of data rows: 80
Total weight for all rows: 80
Rows with missing target or weight values: 0
Rows with missing predictor values: 0

Tele:
E-mail addresses: tarun_ecb@rediffmail.com

Fig 1: MATLAB- Simulink Model

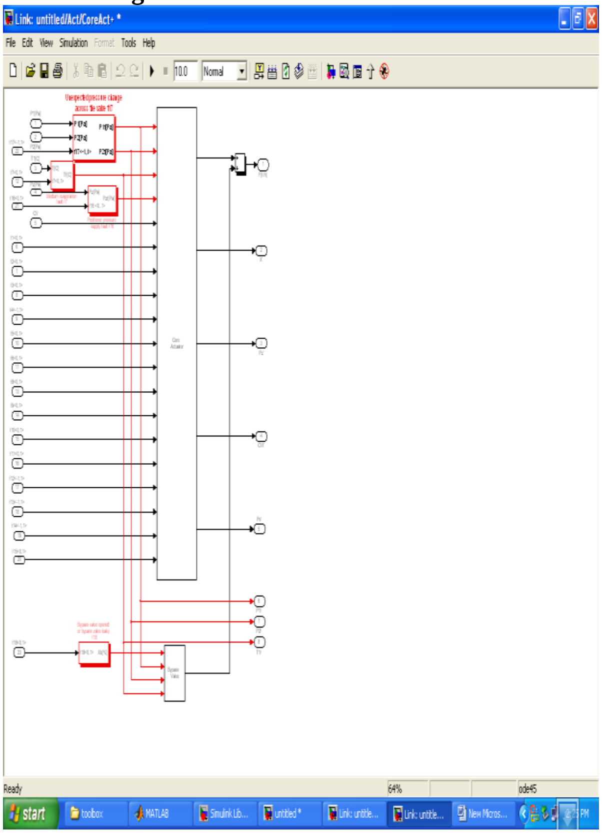


Table 1: Summary of Variables

S.No.	Var	Class	Type	Missing rows	Category
1	CV	Predictor	Continuous	0	20
2	P1	Predictor	Continuous	0	20
3	P2	Predictor	Continuous	0	20
4	T	Predictor	Continuous	0	77
5	X	Predictor	Continuous	0	63
6	F	Predictor	Continuous	0	32
7	Type of fault	Target	Categorical	0	4

Table 2: Correlation Matrix

	CV	P1	P2	X	F	T
CV	1.0000	-0.3719	-0.0419	0.4973	-0.5184	0.0004
P1	-0.3719	1.0000	-0.0518	-0.2102	0.2624	0.0006
P2	-0.0419	-0.0518	1.0000	-0.0001	-0.0096	-0.0006
X	0.4973	-0.2102	-0.0001	1.0000	-0.8584	-0.6589
F	-0.5184	0.2624	-0.0096	-0.8584	1.0000	0.4314
T	0.0004	0.0006	-.0006	-0.6589	0.4314	1.0000

Table 3: Factor Importance

Factor extraction method: Principal Factor Analysis

Factor	Eigen value	Variance %	Cumulative %	Scree Plot
1	2.37766	83.156	83.156	********************
2	0.46142	16.137	99.293	***
3	0.02020	0.707	100.000	
4	-0.00057	. .		
5	-0.16852	. .		
6	-0.33434	. .		

Stopped at cumulative explained variance = 80%

Minimum allowed eigenvalue = 0.50000

Number of factors retained = 1

Table 4: Communalities

	Initial	Final	Common Var. %	Unique Var. %
CV	0.4784	0.2506	25.065	74.935
P1	0.1529	0.0753	7.535	92.465
P2	0.0083	0.0000	0.000	100.000
X	0.8655	1.0000	100.000	0.000
F	0.7715	0.7887	78.871	21.129
T	0.6127	0.2411	24.114	75.886

Communalities converged after 13 iterations

Table 5: Rotated Factor Loading Matrix

Rotation method: Varimax

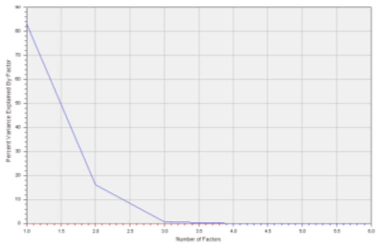
	Fac1
CV	0.5006
P1	-0.2745
P2	0.0008
X	1.0108 *
F	-0.8881 *
T	-0.4911

Eigen value: - 2.378

Variance : - 83.156

Cum. Var. : - 83.156

Fig 2: Scree Plot



K-Means Clustering

The results after applying K-Means Clustering have been tabulated in Tables 6-9.

Project Parameters

Target variable: Type of Fault

Number of predictor variables: 6

Type of analysis: Classification

Category weights (priors): Data file distribution

Misclassification costs: Equal (unitary)

Validation method: Cross validation

Number of cross-validation folds: 10

K-Means Clustering Parameters

The model was created with 3 clusters per target category.

Predictions were made by using the closest cluster.

Misclassification Tables

Table 6: Training Data

Actual			Misclassified		
Category	Count	Wt.	Count	Wt.	%
F12	20	20	1	1	5.00
F13	20	20	0	0	0.00
F15	20	20	1	1	5.00
F7	20	20	0	0	0.000
Total	80	80	2	2	2.500

Table 7: Validation Data

Actual			Misclassified		
Category	Count	Wt.	Count	Wt.	%
F12	20	20	2	2	10.00
F13	20	20	0	0	0.00
F15	20	20	3	3	15.00
F7	20	20	0	0	0.000
Total	80	80	5	5	6.250

Model Size Summary Report

Model evaluation was performed using 4-fold cross-validation.

Clusters	Misclassification %	Probability error %	
2	8.75000	16.04143	
3	3.75000	15.82462	<-- Lowest misclassification
4	6.25000	15.73564	
5	6.25000	15.71177	
6	6.25000	15.66267	
7	5.00000	15.64611	
8	5.00000	15.62520	
9	5.00000	15.58750	
10	5.00000	15.57692	
11	5.00000	15.57289	
12	5.00000	15.56973	
13	5.00000	15.56492	
14	5.00000	15.54949	
15	5.00000	15.54591	
16	5.00000	15.54120	
17	5.00000	15.54120	
18	5.00000	15.54120	
19	5.00000	15.54120	
20	5.00000	15.54120	
21	5.00000	15.54120	
22	5.00000	15.54120	
23	5.00000	15.54120	
24	5.00000	15.54120	
25	5.00000	15.54120	
26	5.00000	15.54120	
27	5.00000	15.54120	
28	5.00000	15.54120	
29	5.00000	15.54120	
30	5.00000	15.54120	
31	5.00000	15.54120	
32	5.00000	15.54120	
33	5.00000	15.54120	
34	5.00000	15.54120	
35	5.00000	15.54120	
36	5.00000	15.54120	
37	5.00000	15.54120	
38	5.00000	15.54120	
39	5.00000	15.54120	
40	5.00000	15.54120	
41	5.00000	15.54120	
42	5.00000	15.54120	
43	5.00000	15.54120	
44	5.00000	15.54120	
45	5.00000	15.54120	
46	5.00000	15.54120	
47	5.00000	15.54120	
48	5.00000	15.54120	
49	5.00000	15.54120	
50	5.00000	15.54120	
51	5.00000	15.54120	
52	5.00000	15.54120	
53	5.00000	15.54120	

The model has been built using 3 clusters.

Confusion Matrix

Table 8: Training Data

Actual Category	Predicted Category			
	F12	F13	F15	F7
F12	19	0	1	0
F13	0	20	0	0
F15	1	0	19	0
F7	0	0	0	20

Table 9: Validation Data

Actual Category	Predicted Category			
	F12	F13	F15	F7
F12	18	0	2	0
F13	0	20	0	0
F15	3	0	17	0
F7	0	0	0	20

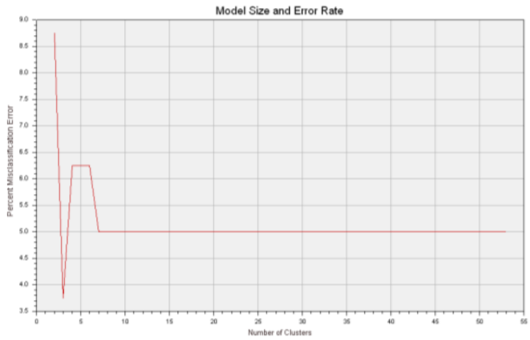


Fig 3:- Model Size and Error Rate

Linear Discriminant Analysis (LDA)

The results obtained after application of LDA have been presented in Tables 10-15.

Project Parameters

Target variable: Type of Fault
Number of predictor variables: 6
Type of analysis: Classification
Category weights (priors): Data file distribution
Misclassification costs: Equal (unitary)
Validation method: Cross validation
Number of cross-validation folds: 10

LDA Parameters

Table 10: Category weights (prior probabilities)

Category	Probability
F7	0.25
F12	0.25
F13	0.25
F15	0.25

Table 11: Standardized Coefficients of Linear Discriminant Functions

Predictor	DF1	DF2	DF3
CV	1.883818	-20.814485	11.179469
P1	8.480973	7.945867	1.952741
P2	-33.424995	-15.639614	2.985998
X	-0.262934	-1.102951	-14.299424
F	-0.612769	-9.254251	-5.827017
T	-912.376894	3.552316	-4.458468
% Var.	99.995	0.004	0.001

Misclassification Tables

Table 12: Training Data

Actual			**Misclassified**		
Category	Count	Wt.	Count	Wt.	%
F12	20	20	0	0	0.00
F13	20	20	0	0	0.00
F15	20	20	1	1	5.00
F7	20	20	0	0	0.00
Total	80	80	1	1	1.25

Table 13: Validation Data

Actual			**Misclassified**		
Category	Count	Wt.	Count	Wt.	%
F12	20	20	0	0	0.00
F13	20	20	0	0	0.00
F15	20	20	1	1	5.00
F7	20	20	0	0	0.00
Total	80	80	1	1	1.25

Confusion Matrix

Table 14: Training Data

Actual Category	**Predicted Category**			
	F12	F13	F15	F7
F12	20	0	0	0
F13	0	20	0	0
F15	1	0	19	0
F7	0	0	0	20

Table 15: Validation Data

Actual Category	**Predicted Category**			
	F12	F13	F15	F7
F12	20	0	0	0
F13	0	20	0	0
F15	1	0	19	0
F7	0	0	0	20

Discussions

The strength of above approach lies in the accuracy manifested in handling the classification (discrimination) task of faults inside overlapping areas with fine precision.

Future research needs to focus on further improvement of fault diagnosis results on DAMADICS benchmark in other categories of overlapping faults by using perception based decision making.

References

[1] Isermann, R., Fault Diagnosis Systems. An Introduction from Fault Detection to Fault Tolerance. Springer, Berlin, 2005.

[2] Phillip H. Sherrod, “DTREG Predictive Modeling Software”, 2003

[3] Michal Bartys et al, “Introduction to the DAMADICS actuator FDI benchmark study” Control Engineering Practice 14, 577–596, 2006.

[4] Gorsuch, Richard L. Factor Analysis. W. B. Saunders Co. 1974

[5] Hartigan, J.A. and Wong, M.A. A K-Means Clustering Algorithm. Applied Statistics 28. 1979.

[6] Fisher, R.A ,The use of multiple measures in taxonomic problems, Ann. Eugenics, 7:179—188, 1936.