Preprints of the 10th IFAC International Symposium on Dynamics and Control of Process Systems
The International Federation of Automatic Control
December 18-20, 2013. Mumbai, India

Alarm allocation for event-based process alarm systems

Pradeep Dalapatu * , **Salim Ahmed** *[,1] , **Faisal Khan** *

* *Faculty of Engineering & Applied Science*
Memorial University, St. John's, NL, Canada A1B 3X5

Abstract: The ability to monitor large numbers of variables and the flexibility to assign alarms to each variable led to a substantial increase in the numbers of alarms in industrial plants. This, in turn, increased the numbers of false and redundant alarms. In plant operations, the numbers of annunciated alarms regularly exceed the acceptable rates that operators can handle. To reduce the number of assigned alarms, a risk-based alarm system has been proposed in the literature (Ahmed et al. (2011); Chang et al. (2011)) where alarms are assigned to groups of variables instead of individual variables. This articles explores the options for grouping variables for alarm allocation. Several grouping methods are discussed and an event-based grouping procedure is detailed. Selection of the key variables for a group is performed using the information that the variables can have to distinguish between an abnormal and a normal condition. The concept of mutual information is used to quantify the information. Variables with high information gain are grouped together for each respective abnormal event. To identify the redundant variables within the groups to further reduce the number of variables to be monitored, the maximum cross-correlation between pairs of key variables are used. A case study using the example of a continuous stirred tank reactor is used to demonstrate the methodology.

Keywords: Alarm; Variable allocation; Alarm Flooding; Information theory; Mutual information.

1. INTRODUCTION

Continuous developments in technologies such as the supervisory control and data acquisition (SCADA) and the distributed control system (DCS), along with the low cost of sensors have increased the ability to monitor and store large number of variables during plant operations. In a complex process plant the number of observed variables may be in thousands. Some of the process variables are used to detect faults, which can be defined as 'deviations of a monitored variable or a calculated parameter from its normal range'. A fault is a symptom of a system failure that can be identified as 'changes in process parameter with disturbances from external processes, equipment failures or control system failures'. One or more failures can lead a process to an abnormal event which is hazardous in safety, economical, social and environmental aspects (Venkatasubramanian et al. (2003)).

Designing an alarm system to detect failures and implementing corrective actions are critical tasks in chemical, process, and oil and gas industries. Alarms are used to notify the operator about any faulty condition from the respective failure before its propagation to a more hazardous event. Due to the ability to monitor large numbers of variables with low cost, plant designers tend to assign alarms to as many variables as they can. It simply costs less to add an alarm than to discuss whether it is needed or not. However many alarms that are allocated to different variables may indicate minor faults that have no significant effect on any major abnormal event. With a small disturbance, many minor 'false' alarms can be triggered. On the other hand, a major failure, which can be propagated to a severe event can trigger many redundant minor alarms along with a primary alarm. Due to the false and the redundant alarms, alarm rates can be increased to a level above the operators' physical ability to handle alarms. This phenomena is called 'alarm flooding' that can reduce the motivation of the operators to check on alarms. Also it can reduce the ability of the operators to effectively detect the root causes of any major failure (Izadi et al. (2009)). According to EEMUA (2007), an operator should not handle more than six alarm per hour. In reality, according to different reviews, numbers of alarms exceed this value by large numbers, during both normal and abnormal conditions. Chang et al. (2011) has discussed about the recent situation in industries and the standards which they have for alarm management.

Minimizing the number of alarms without compromising the ability to identify the significant failures is a crucial factor in alarm system design. Many alarm management methods have been proposed in the literature. Alarm management life cycle is proposed by ISA (2009) to manage alarms at the design stage and alarm processing techniques like grouping of alarms, alarm suppression and shelving are discussed in EEMUA (2007).

To reduce the number of alarms a risk-based alarm design procedure has been proposed in the literature (Ahmed et al. (2011)). In this approach, alarms are allocated

[1] author to whom all correspondence should be addressed. E-mail: sahmed@mun.ca, Tel: 1-709-864-7652, Fax:1-709-864-8759

to groups of variables instead of individual variables. However, the above reference does not outline how to define such groups to assign alarms. In this article the methodologies for grouping of variables are explored. An event-based grouping option is detailed as a suitable option for grouping of variables to allocate alarms. The remainder of the article is organized as follows. Section 2 discusses the different grouping options followed by the details of an event-based grouping procedure in section 3. A case study is presented in section 4 followed by concluding remarks.

2. GROUPING VARIABLES TO ASSIGN ALARMS

In this paper we focus on the concept of grouping variables and assign alarms to groups of variables. Allocation of an alarm to a group of variables will result in the annunciation of one alarm even when one or more variables within the group has deviated. Grouping can be performed considering various factors. Variables can be grouped according to their types, or the equipment that they are associated with, or according to their correlation. In addition, alarms can be event-based where a group of variables associated with an abnormal event can be allocated an alarm. This concept has been used to develop a risk-based alarms system (Ahmed et al. (2011)) where measurements of each variable in a group are used to calculate its associated risk; an overall risk associated with all of the variables in a group is then evaluated and an alarm is annunciated if the overall risk exceeds its acceptable value. The details of the risk-based alarm system design methodology is part of an ongoing research project. In this article, we focus on the options to grouping of variables.

Identification of the key variables related to an abnormal event is a challenging task. The most important variables can be identified by various techniques such as the variable selection methods, or based on expert knowledge. Chen and Wang (2000) proposed a method based on the principle component analysis (PCA) and the resulting contribution plot to detect the important variables to classify fault conditions. Orantes et al. (2008) used the concept of entropy from the information theory to estimate the most informative variable related to a failure for the purpose of selecting sensor locations. However grouping of variables for alarm allocation has not been addressed in the literature. In this article the mutual information concept (Pérez et al. (2006)) for key variables selection by using information theory for Gaussian random variables is used for grouping variables to assign alarms.

Based on their information contents, variables can be grouped together to represent different abnormal events. However, there can be highly correlated variables within a group. In order to identify the redundant variable within a group, correlation analysis is needed to be done. Various methods have been proposed in the literature to cluster process variables or alarms according to their correlation. Yang et al. (2012) and Noda et al. (2011) proposed methods to analyze correlated alarms by using binary alarm data. Geng et al. (2005) proposed a method to cluster variables by fuzzy clustering method. Independent grouping analysis is proposed by Alhoniemi et al. (2007) considering mutually dependent variables using a cost function. Yu and Liu (2003) discussed about information redundancy between variables using the concept of mutual information. In this article, the maximum cross-correlation (Swift et al. (2001)) among the variables is used to identify the redundant variables. The following section discusses the different grouping methods.

2.1 Grouping methods

Grouping based on variable types: Different types of measurements such as temperatures, pressures and levels are available from industrial plants. By grouping variables according to their types and assigning alarms to groups may significantly reduce the number of alarms in a plant. For example, if there are a number of thermo-couples along the length of a distillation column, instead of assigning alarms to each of the measurements, one alarm can be allocated to the set of temperature measurements. Annunciation of the alarm would indicate an abnormality related to the temperature in the column. Thus, in a particular system which has a high number of monitored variables of the same type, the operator can efficiently identify a faulty situation without causing alarm flooding. However the operator will need more information to identify the root causes of the failures.

Grouping by plant unit or equipment: In a complex process plant monitoring system, variables can be grouped unit- or equipment-wise. For example, measurements from the stripping section of a distillation column can be grouped together to assign an alarm whose annunciation would direct the operator to focus on that section and take actions. Thus the operator can effectively identify the failure location and further analyze the situation to find the root cause without having many alarms from the same unit or system. But due to correlation of the variable, one unit failure can be affected by other upstream variables and this can mislead operators.

Grouping based on correlations: Strong correlations exist among plant variables due to their interactions and also due to plant connectivity. For example, the composition of the feed to a reactor may affect the conversion in the reactor leading to a changed product composition, product flow rate and/or the temperature in the reactor. If alarms are assigned to each of the variables, a change in the feed may cause a number of alarms to annunciate. Thus one failure may lead to many alarms. If variables are grouped according to their correlations, number of redundant alarms can be significantly reduced. But the information from the alarm will be unclear. Also prioritizing of alarms can be ineffective to the operator.

Grouping by abnormal events: Variables related to an abnormal event may be grouped together to assign an alarm. For example, for a simple tank process, the flow rates of the inlet and the outlet streams along with the level of liquid in the tank may be related to an overflow condition of the tank. However, instead of assigning alarms to each of the variables, an overflow alarm can defined based on the above measurements. Thus number of alarms can be reduced. In addition, the annunciation of the alarm would inform the operator about a defined event.

The above mentioned group-based alarm assignment would require that a single indicator be defined from the measurements of the variables to annunciate an alarm. Such exercise for the first three options would require further study; for the event-based grouping the risk of the event associated with the variables in the group can be used for alarm annunciation. However, challenges remain on how to estimate the risk associated with a set of variables. Once the risk can be estimated, alarm annunciation, its prioritization and the diagnosis to find the root cause can be done with less difficulties. Considering this aspect, the abnormal event-based grouping seems to be more promising. The following section details the event-based grouping procedure.

3. EVENT-BASED ALARM ASSIGNMENT

In order to select variables to form a group, the first step is to identify the failure that can occur in the unit or equipment or system. This can be done by various risk assessment methods e.g. the HAZOP and the FMEA. Then the abnormal events which may result from the failures are identified. Once the failures and the abnormal events are identified, process data are required to group variables. If the plant is at the design stage, simulation can be carried out to generate data for the abnormal events. For an operational plant, historical data can be used along with simulations to meet data requirements. Using data for both normal operations and for abnormal events, the information theory can be used to select the key variables associated with an event.

3.1 Selection of the key variables

The information theory, proposed by Shannon (1948) which is routinely used in communication systems, measures the information of a random variable in a quantitative manner. According to the theory, uncertainty associated with a random variable X can be measured by its entropy $H(X)$ using Eq.1.

$$H(X) = -\sum_{x} P(x)\, log_2(P(x)) \tag{1}$$

Here, X is assumed to be a discrete random variable. $P(x)$ is the probability distribution of $X = x$ occurrence. Entropy is measured with unit 'bits', therefore log_2 is considered in calculations. For two random variables X and Y, the joint entropy $H(X, Y)$ can be defined as

$$H(X, Y) = \sum_{x}\sum_{y} P(x, y)\, log_2(P(x, y)) \tag{2}$$

$P(x, y)$ is the discrete joint probability distribution of $X = x$ and $Y = y$. The mutual information which one random variable contains about the other random variable can be derived as outlined in (Cover and Thomas (1991)).

$$I(X; Y) = \sum_{x,y} P(x, y)\, log_2 \frac{P(x, y)}{P(x)P(y)} \tag{3}$$

$I(X; Y)$ is the mutual information between random variable X and Y. Eq.3 can be simplified to give

$$I(X; Y) = H(X) - H(X/Y) \tag{4}$$

$H(X/Y)$ is the entropy of the random variable X given Y. The mutual information between two random variables can be used in abnormal event-based alarm design to select the key variables that consist most information regarding an event. In order to do that, variable Y need to be defined as a random variable that indicate the failures that can propagate to specific abnormal event. For an example, if there is only one failure that can be propagated to an abnormal condition then Y can be defined by two random numbers, $Y = 0$ (normal) and $Y = 1$ (failure) or else if there are $k - 1$ failures that can be propagated to the specific abnormal event then Y can have k random numbers (k being the number of failure plus the normal condition).

If the variable does not have the ability to distinguish an abnormal event from the normal condition, then the amount of uncertainty do not change. If a variable can distinguish between conditions, then the amount of uncertainty will be reduced. Reduction of the uncertainty or entropy is the information gain that a variable consists.

Entropy of a continuous random variable having a normal distribution has been defined in (Cover and Thomas (1991)) as

$$H(X) = -\int_{x} p(x)\, log_2(p(x)) = \frac{1}{2}\, log_2(2\pi e \sigma^2) \tag{5}$$

$p(x)$ is the probability distribution of continues random variable X and σ is the standard deviation of X.

Pérez et al. (2006) proved that if the variable X has a normal distribution and if the class C is a multi-nomial random variable having 1 to k finite outcome with a probability distribution of $P(C = c)$, and $p(c, x)$ is the joint probability distribution of $C = c$ and $X = x$, then the information that the variable can have for all the classes is given by

$$I(X; C) = \sum_{c=1}^{k} \int_{x} p(c, x)\, log_2 \frac{p(c, x)}{P(c)p(x)} \tag{6}$$

$I(X; C)$ is the mutual information between X and C. During normal conditions, variations of data occur only due to measurement noise which is typically small white noise. But if the variable contains high information, then for each failure condition the variation of data will be significant. Therefore entropy between failures and the variable will decrease. Hence from Eq. 5, the following equation can be derived to calculate the information gain, Pérez et al. (2006),

$$I(X; C) = \frac{1}{2}\left[log(\sigma^2) - \sum_{c=1}^{k} P(c) log_2(\sigma_c^2) \right] \tag{7}$$

where σ is the standard deviation of the random variable X and σ_c is the standard deviation of the random variable X given $C = c$. $P(C)$ is assumed to have uniform distribution implying that the information about the normal and failure conditions are unknown and their probability of occurrence are the same. It is also assumed the data acquired from simulation for each variable in different conditions are normally distributed and the classes are considered as multinomial random variables.

Finally variables having high I values can be selected as the most suitable variables to monitor the respective event. Thus the abnormal event alarm can be assigned to the group of variables.

3.2 Identification of the redundant variables

Different variables selected within a group may contain the same information and thus can be considered to be redundant. The redundant variables can be identified using the correlation analysis. The purpose is to identify redundant variables within a group and thus to exclude all but one from a redundant set for monitoring.

To perform the correlation analysis, data are standardized to have zero mean and the cross-correlation between pairs of variables are estimated. There can be time lags between variables. Hence to calculate the maximum correlation, time lag is varied and the correlations are calculated to get the maximum positive or negative value. Maximum time lag can be decided using process knowledge (Swift et al. (2001)). Pearson correlation coefficient is used to calculate the similarity. At the maximum positive coefficient ϕ_{max} time lag is lag_{max} and at maximum negative coefficient ϕ_{min} time lag is lag_{min}. Then the maximum absolute correlation is calculated as follows (Yang et al. (2010)),

ϕ_{max} is taken at lag_{max} if $\phi_{max} \geq -\phi_{min}$
ϕ_{min} is taken at lag_{min} if $\phi_{max} < -\phi_{min}$

Correlation matrix can be used to develop a correlation color map and variables that are highly correlated with each other are grouped together. The high correlated color cluster indicate the variables that have high correlations. For the purpose of better visual representation, grouping can be done by calculating the similarity distance between each pair of variable. After getting the distance between variables in the data, variables close to each other can be linked and presented in clusters in a hierarchical tree (Yang et al. (2012)). A dendrogram is used to present correlated variables. Fig. 1 present the complete methodology for variable allocation for the event-based alarm system.

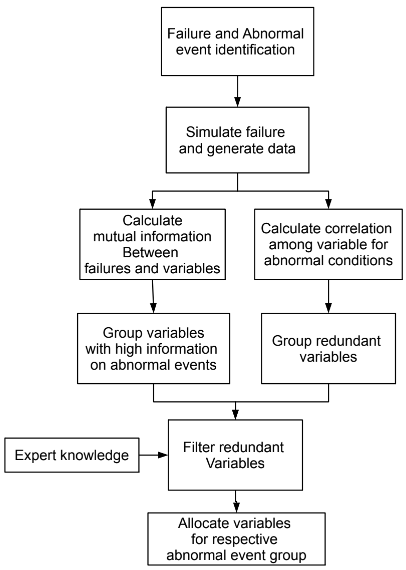

Fig. 1. Methodology for grouping variables to assign alarms

4. CASE STUDY AND RESULTS

4.1 Case study

As a case study, a jacketed continuous stirred tank reactor (CSTR) is considered. An irreversible exothermic reaction $A \rightarrow B$ is assumed to take place in the reactor with a first order kinetics. A temperature controller is used to control the reactor temperature by manipulating the coolant flow rate. The level of the reactor tank is also maintained by manipulating the reactor outlet flow. Heat losses are considered negligible and a perfect mixing condition is assumed. All the parameter for the model is taken from (Luyben (1996)) and the controller PI parameters are taken from (Chang and Yu (1990)). To demonstrate the methodology, Simulink is used to built a plant model. Different failure conditions are simulated with Simulink to generate data. For the CSTR, 11 variables are identified. Table 1 lists the variables.

Table 1. List of variables for the CSTR.

No	Measuring variable
1	Reactor liquid percent level
2	Coolant utility outlet temp
3	Reactant concentration
4	Reactor vessel temperature
5	Reactor output flow rate
6	Coolant utility flow rate
7	Reactant feed temperature
8	Reactant feed flow rate
9	Coolant inlet temperature
10	Level controller output
11	Temperature controller output

Variables 5 and 6 are manipulated to control variables 1 and 4, respectively. Other variables are uncontrolled variables.

Ten failures are considered for this study. Using Simulink, data for all the failures and the normal condition are generated. Table 2 presents the failures.

Table 2. Possible failure conditions for the CSTR process

No	Failure
F1	Reactant feed flow disturbance - High flow
F2	Reactant feed flow disturbance- Low flow
F3	Coolant system failure - High coolant temperature
F4	Coolant system failure - Low coolant temperature
F5	Reactor out flow valve failure- High flow
F6	Reactor out flow valve failure- Low flow
F7	Coolant flow valve failure- High flow
F8	Coolant flow valve failure- Low flow
F9	Reactant feed quality failure - High concentration
F10	Reactant feed quality failure - Low concentration

Some of these failure can propagate to more sever abnormal events. Table 3 presents the abnormal events and the respective failures that can lead to the events.

4.2 Grouping variables

Using the concept of the information theory, variables are grouped according to their information gain. First a pair-

Table 3. Possible abnormal events for the CSTR process

Abnormal Event	Failure
Runaway	F3,F8
Flooding	F1,F6
Low quality products	F9,F4,F7

wise comparison for the normal and a failure is carried out to identify variables that can reduce the uncertainty of the failure under consideration. Fig. 2 presents the information gain of all of the variables in the form of a bar chart for the failure F8: coolant valve failure - low coolant flow.

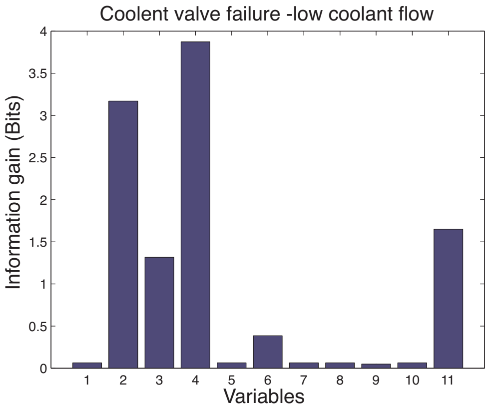

Fig. 2. Information gain of different variable corresponding to the failure F8: coolant valve failure - low coolant flow

As observed from Fig. 2, there are 5 main variables 4, 2, 11, 3 and 6 have the significant information about the failure. Accordingly, this set of five variables are considered as the key variables for the for the failure F8. Following the same procedure, key variables are identified for all the failures.

In order to group the variables according to the abnormal events, information gains are calculated by considering all the failures that can propagate to the corresponding abnormal event. Fig. 3 presents the information gain for all of the listed variables for the abnormal event, runaway reaction.

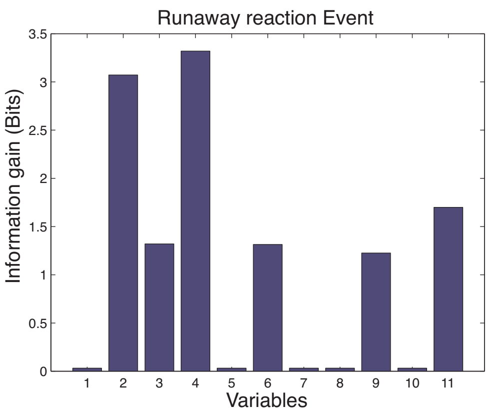

Fig. 3. Information gain of different variable for the event - runaway reaction

Table 4. Selected variable groups for different abnormal events.

Event	Key Variables	Redundant variables	Chosen group
Runaway	(2,3,4,6,9,11)	(2,3,4) and (6,9)	(2,6,11)
Flooding	(1,3,5,8,10)	(3,5,8)	(1,8,10)
Low Quality	(2,3,4,11)	(2,4)	(3,4,11)

As shown in Fig. 3, the variables 4, 2, 11, 6, 3 and 9 can give significant information about the event. Accordingly, this set of six variables are considered as the key variables for the event runaway reaction. Following the same procedure, key variables are identified for all the events.

4.3 Redundant variable selection

Correlation analyses are carried out to identify the redundant variables within a group of key variables which are selected for each events. The maximum cross correlation matrix is generated by varying the time-lag between each pair of variables for each abnormal condition data. It is converted to a correlation color map. For the purpose of visualization a hierarchical cluster tree is developed in the form of a dendrogram as shown in Fig. 4 that show the correlations among variables in a runaway reaction. Finally it is required to choose one variable from each redundant group for the purpose of minimizing the monitored variables. Choosing the most suitable variable may become a challenging task. Process knowledge as well as sensor characteristics may be required to consider for this purpose. In this case the variable with the highest information gain among the redundant variables is chosen. Table 4 presents the list of variables that can be allocated to the groups corresponding to the individual events and the highly correlated variables within different groups. It also shows the final group selection for each of the events.

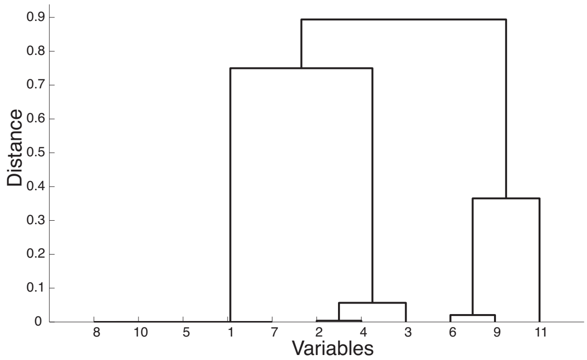

Fig. 4. Correlation among variables for runaway event

4.4 Discussion

From the results, reactor vessel temperature (variable 2), coolant utility flow rate (variable 6) and temperature controller output (variable 11) are the main variables that have most information regarding runaway reaction. It is obvious that the main variable that can be used to detect a runaway is the reactor temperature. Main root cause for the runaway reaction is the failure of the coolant system. Variable 6 and 11 are directly related to the coolant system failures.

Primary variable for flooding condition monitoring is level of the reactor (Variable 1), therefore it should be a key variable. The main root cause is the level controller failure and feed flow valve failure. Variables that are directly related to both failures are identified as the key variables that are reactant feed flow rate (variable 8) and level controller output (variable 10).

Low quality production can be quantified by reactant concentration (variable 3) which is a key variable according to the methodology. Incomplete reaction due to the low temperature is the main reason for low quality production. Present methodology has identified reactor vessel temperature (variable 4) and temperature controller output (variable 11) as other key variables to detect low quality production.

The case study demonstrates that process knowledge justifies the selection of the key variables by the proposed methodology.

4.5 Other grouping types

As mentioned in Section 2, there are other options to group variables. Table 5 presents the selected group of variables according to different grouping methods. As shown in the table, different methods may result in significantly different results.

Table 5. Results on group formation using different methodologies.

Groups	G1	G2	G3	G4	G5
Variable type	1	2,4,7,9	3	5,6,8	10,11
Plant/Unit	4,2,6,9,11	1,4	3,7,8	1,5,10	
Correlation	2,3,4	6,9	7,11	5,8,10	1
Event base	2,6,11	1,8,10	3,4,11		

5. CONCLUSION

Methodologies to allocate variables to groups for the purpose of designing group-based alarm systems are discussed. A procedure for selection of variables to form groups for an event-based alarm system is detailed. The method uses the information theory and the concept of mutual information to select the key variables to allocate to a group. Correlation analyses are then carried out to select the redundant variables within a group. A case study using the example of a CSTR is used to elaborate the proposed methodology. Following the same procedure, variable selection to design an event-based alarm system can be carried out for an entire plant. Once variables are selected to form groups, one alarm will be assigned to each group. Finally, a risk-based approach will be used to estimate the risks associated with the variables in a group and the overall risk associated with a group will be evaluated. The alarm will be annunciated if the overall risk is higher than a pre-chosen threshold. This article outline the grouping methods; risk estimation, overall risk evaluation, choice of thresholds and diagnosis of alarms will be addressed in future works.

ACKNOWLEDGEMENTS

Salim Ahmed acknowledges the financial support from the Research and Development Corporation (RDC) of Newfoundland and Labrador for this work.

REFERENCES

S. Ahmed, H. A. Gabbar, Y. Chang, and F. I. Khan. Risk based alarm design: A systems approach. 2011.

E. Alhoniemi, A. Honkela, K. Lagus, J. Sepp, P. Wagner, and H. Valpola. Compact Modeling of Data Using Independent Variable Group Analysis. pages 1–14, 2007.

C. Chang and C. Yu. On-Line Fault Diagnosis Using the Signed Directed Graph. pages 1290–1299, 1990.

Y. Chang, F. Khan, and S. Ahmed. A risk-based approach to design warning system for processing facilities. *Process Safety and Environmental Protection*, (5), 2011.

F Chen and X Wang. Discovery of Operational Spaces from Process Data for Production of Multiple Grades of Products. *Industrial & Engineering Chemistry Research*, July 2000. ISSN 0888-5885.

T. Cover and J. Thomas. *Elements of Information Theory*. Wiley-Interscience, August 1991.

EEMUA. Alarm Systems, a guide to design, management and procurement. *EEMUA*, (191), 2007.

Z. Geng, Q. Zhu, and X. Gu. A fuzzy clustering-ranking algorithm and its application for alarm operating optimization in chemical processing. March 2005.

ISA. Management of Alarm Systems for the Process Industries. Technical Report June, International Society of Automation, 2009.

I. Izadi, S. L. Shah, D. Shook, and T. Chen. An Introduction to Alarm Analysis and Design. pages 645–650, 2009.

I. Luyben. *Process Modelling, Simulation and Control for Chemical Engineers*. 1996.

M. Noda, F. Higuchi, T. Takai, and H. Nishitani. Event correlation analysis for alarm system rationalization. 2011.

A. Orantes, T. Kempowsky, M. Le Lann, and J. Aguilar-Martin. A new support methodology for the placement of sensors used for fault detection and diagnosis. *Chemical Engineering and Processing*, (3), March 2008.

A. Pérez, P. Larrañaga, and I. Inza. Supervised classification with conditional Gaussian networks: Increasing the structure complexity from naive Bayes. *International Journal of Approximate Reasoning*, September 2006.

C. E. Shannon. A Mathematical Theory of Communication*. 5(I):3–55, 1948.

S. Swift, A. Tucker, N. Martin, and X. Liu. Grouping multivariate time series variables : applications to chemical process and visual field data. 14, 2001.

V. Venkatasubramanian, R. Rengaswamy, and K. Yin. A review of process fault detection and diagnosis Part I : Quantitative model-based methods. 27:293–311, 2003.

F. Yang, S. L. Shah, D. Xiao, and T. Chen. Improved correlation analysis and visualization of industrial alarm data. *ISA transactions*, 51(4):499–506, July 2012.

Fan Yang, L. Shah Sirish, and Deyun Xiao. Signed Directed Graph modeling of industrial processes and their validation by data-based methods. October 2010.

L Yu and H Liu. Feature Selection for High-Dimensional Data : A Fast Correlation-Based Filter Solution. 2003.