Transactions of the Korean Nuclear Society Spring Meeting
Jeju, Korea, May 18-19, 2017

Accident Classification and Clustering Using RSM in NPPs

Moath Kassim , Gyunyoung Heo*
Kyung Hee Univ., Yongin-si, Gyeonggi-do, 17104, Republic of Korea
**Corresponding author:* gheo@khu.ac.kr

1. Introduction

The conventional way for tracking an accident during emergency operation is to manually track important safety related parameters whether they are approaching the unacceptable limits that are provided by vendors/designers in the technical specifications of the plant such as; observing the pressure, temperature, SG water level and so on [1]. However, operators may not be able to recognize the start time of accident and may take much time to recognize the type of accident, therefore, the need of a reliable computational aid system is with high value as to help operators to identify the initiation time of accident and quickly estimate the type of accident as to tackle the task and mitigate the situation with minimum time.

Regarding the objectives of this paper, a computerized algorithm can be helpful for prognostic, classifying, and clustering accidents during nuclear power plant (NPP) operation; detecting the initiation time of accident as fast as before the reactor get tripped, and identifying the accident type would minimize the time needed to assess the accident properly.

The Residual Sign Matrix (RSM) is a matrix obtained from the difference between the normal operation state and transient/accident state, this method can detect the change in plant parameters' pattern and it can also work as a transformation matrix for the purpose of classification and clustering. The RSM was developed with two phases; the 0^{th} RSM is a simple matrix explaining the signals of accident pattern into up (+1), down (-1), and (0) no change with respect to normal operation condition pattern. However, many accidents may share same numbers of positive patterns' trends and negative patterns' trends that make it hard to classify and cluster accidents accurately due to the overlapping of points.

The second phase is the 1^{st} RSM which represent the exact increase and the exact decrease in term of angle. The idea of orientation mode of a signal was inspired by the Edge Detection (ED) method which widely used in image processing. This method's techniques are mathematically based on the first numerical derivatives.

The 1^{st} RSM was used as transformation of all signals of the accident pattern to be able to classify them in one scale and same unit, then the classification was performed using the K-NN (K-Nearest Neighbor). Clustering was performed using the K-mean clustering method considering unknown accident.

Sets of PWR simulated data in normal condition and accident conditions such as Loss Of Coolant Accident (LOCA), Steam Generator Tube Rupture (SGTR), Steam Line Break Inside Containment (SLBIC), Steam Line Break Outside Containment (SLBOC), Load Rejection (LR), and Moderator Dilution (MD) were used to apply these methods using in-house code and Rapidminer software.

2. Methods and Results

2.1 Accidents Prognostic using RSM

Detecting accidents during the operation time of NPP is an important issue to maintain the safety of the plant and to prevent any undesirable consequence on the environment surrounded. Although the Emergency Operation Procedures (EOP) are applied when the reactor trip, an early prognostic is more important as it can save assessment time and cost and prevent any complex consequence. For this purpose, this paper presents an approach for detecting an accident with respect to stored qualified length of normal operation condition data as soon as before the reactor get tripped.

The RSM is a matrix obtained by calculating the residual and the angular residual between normal operation condition data and the expected accident. This RSM was developed in two phases; first is the 0^{th} RSM that represent the trends of accident signals with respect to the normal operation signals, these trends are represented by trinomial values either +1 which explains "increase", -1 which explains "decrease", and 0 which explains "No change" [2]. Equation (1) is showing how to generate these values.

$$0^{th}RSM(t,j) = sig[X_{acc}(t,j) - X_{norm}(\mathrm{t,j})] \quad (1)$$

Where; t refers to time, j refers to the attribute (plant parameter), X_{acc} is the accident data, X_{norm} is the normal operation data

The second phase of RSM is the 1^{st} RSM which is a novel idea developed by the author, converts all signals no matter their scales and their physical units into an angel space, this angle space explains the angular change or we can say the orientation mode of the accident's signals with respect to the normal operation signals. The 1^{st} RSM can be also used as a transformation matrix for the purpose of accidents classification and clustering. This method was inspired by the ED method since it tracks the orientation change of colors intensity in image processing and likewise it can track the orientation change in transient situation (accident) in data processing [3, 4] which analyses images and the intensity change in their colors. The mathematical formula derived for generating the 1st RSM is as following:

$$1st\ RSM_i(t,j) = \tan^{-1}\left(\frac{X_{acc}(t,j) - X_{norm}(t,j)}{t_i - t_{i-1}}\right) \quad (2)$$

Where; i is the time step

It is important to emphasize here, that the 0th RSM and the 1st RSM will be only generated at the start time of accident and for a few seconds before the reactor is tripped.
As to consider the efficiency of accidents detection minimizing the noise and degradation effects, an accuracy band was applied to the normal operation sample data and the residual and the angular residual was calculated with respect to the upper and lower limits of this band.

2.2 Accidents Classification using RSM & K-NN

In case there are some available data sets of accidents happened in the same type of plant under the same operation condition, these stored data sets which can be called as "Training data" along with a qualified normal operation data will be processed by the 1st RSM method to generate their respective matrix, and then will be processed and classified by any available classification method, and then the results will be stored in the plant database for the use against any upcoming accident.

In this paper, we propose the use K-Nearest Neighbor (K-NN) classifier. The K-NN is a classification approach that classify groups of data based on the distance between their points and then predict for any similar upcoming group [5]. This idea would be helpful for the developing of online monitoring aid system based on the 1st RSM.

In figure (1), the process of K-NN accidents classification with respect to some stored accidents is illustrated using Rapidminer software [6].

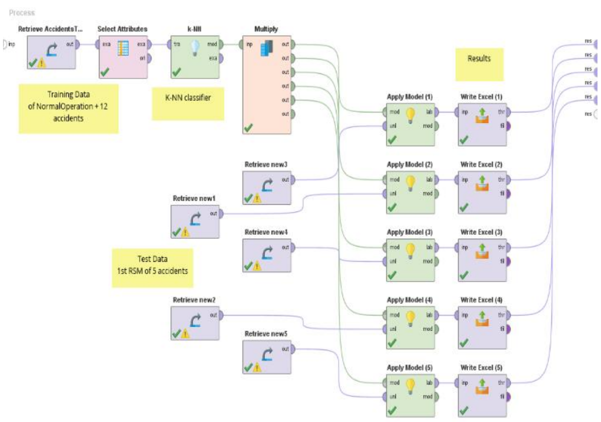

Figure 1. Classification procedures of five accidents using K-NN generated by Rapidminer.

In this procedure, 12 accidents data sets plus a normal accident data set were processed and converted to 1st RSM forms and then stored and, in addition, 5 new accident data sets were also processed by the 1st RSM algorithm and connected along with the training data sets stored, they all underwent the K-NN classification procedure to generate the results shown in section (3). It is important to emphasize here that the new accidents which usually called "Test data" are related to the training data so the K-NN will assign them to one of those 12 groups of accident that are stored previously.

2.3 Accident Clustering using RSM & K-means

In case a new accident occurs and this accident does not belong to any of the stored accidents, the need of clustering method is with high value. However, the clustering method will identify the new group as unknown accident or as there are 12 clusters stored in the database it will assign "cluster 13" for the unknown accident group.

For clustering purpose, we used the K-mean clustering approach that is simple and widely popular. The K-means is typically a class of procedures that are known as clustering algorithms [5]. The k-means algorithm tries to find a user-specified numbers of clusters (k) that are represented by their means. Figure (2) shows the procedure implemented in Rapidminer [6] for clustering an unknown accident among stored accidents RSMs in plant's database, and because K-mean clustering method doesn't need splitting the data into training and test data, all normal operation and accidents data including the new accident were stored dynamically in one file and the only thing we need to do is to specify the number of K groups we have.

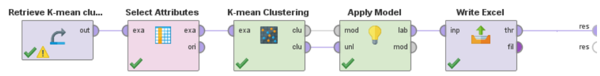

Figure 2. Clustering procedures of an unknown accident using K-mean generated by Rapidminer.

2.4 Data Preparation

A set of PWR simulated data was used to conduct this study. 12 training accidents data sets were generated from 4 malfunctions of the same plant condition in addition to the normal operation data set, these accident malfunctions are; LOCA (cold leg), LOCA (hot leg), Steam Line Break inside containment (SLBIC), Steam Line Break outside containment (SLBOC). For test data sets, malfunctions such as Steam Generator Tube Rupture (SGTR), Load Rejection (LR), and Moderator Dilution (MD) were used. All data sets are in time domain and have 94 parameters (attributes), and the accidents were generated as a single failure accident to avoid complexity at this part of the study.

Training and test data used in this study are illustrated with their respective and given names in table (1).

Table 1. Training and test data sets used for obtaining the 1st RSMs

Cluster	Data set given name	Scenario	Failure rate (%)
0	NormalOperation	Normal condition	0
Training data sets			
1	LOCA-CL1%	LOCA Cold Leg	1
2	LOCA-CL10%		10
3	LOCA-CL30%		30
4	LOCA-HL1%	LOCA Hot Leg	1
5	LOCA-HL10%		10
6	LOCA-HL30%		30
7	SLBIC10%	Steam Line Break Inside Containment	10
8	SLBIC30%		30
9	SLBIC50%		50
10	SLBOC10%	Steam Line Break Outside Containment	10
11	SLBOC30%		30
12	SLBOC50%		50
Classification Test data sets			
	New1	LOCA Cold Leg	20
	New2	LOCA Hot Leg	20
	New3	LOCA Cold Leg	40
	New4	Steam Line Break Inside Containment	20
	New5	Steam Line Break Outside Containment	20
Clustering Test data sets			
13	New6	Load Rejection	50
14	New7	Moderator Dilution	50
16	New8	Steam Generator Tube Rupture	10
17	New9		30
18	New10		50

2.5 Rapidminer Studio

Rpidminer is a visual environment software for where most of the predictive, classification, clustering analysis can be obtained based on built in programs. Its interface is intuitive graphical and very friendly that no extra programming is required to perform this analysis.

In this study and as figure (3) shows, Rapidminer Studio version 7.4 was used as the tool for performing classification and clustering analysis while the RSMs were generated using in-house code.

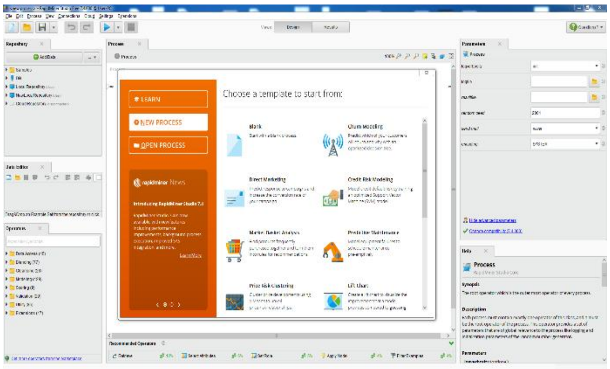

Figure 3. Rapidminer Studio 7.4 interface.

3. Results & Discussion

The 0th RSM can detect the initiation time of accident since the residual between the accident signals and their respective normal operation signals is generated. But the 0th RSM has only three values that can make overlapping problem at the time of classification and clustering, so, the 1st RSM was used instead.

The 1st RSM also can detect the initiation time of accident since it calculates the angle of deviation between accident signals and their respective signals of normal operation pattern. The results of 1st RSM was only considered for 5 time steps, one before the accident (T-1), one is the accident initiation time (T), and three time steps after the accident initiation time (T+1, T+2, T+3) in order to perform the analysis before reactor trip. Table (2) is showing a representative result of five attributes (plant parameters) of the 1st RSM in case of LOCA-CL 10%.

Table 2. The 1st RSM results of five selected parameters of LOCA-CL 10%

Plant Variable	1st RSM				
	T -1	T	T +1	T +2	T +3
P_θ	0	-1	-2	-2	-3
$Tavg_\theta$	0	0	1	1	1
$LVPZ_\theta$	0	-7	-19	-30	-39
$QMWT_\theta$	0	-38	-73	-74	-72
RM_θ	0	1	3	4	6

Where; θ is the dynamic angle P_θ is Reactor Coolant System (RCS) pressure, $Tavg_\theta$ is RCS average temperature, $LVPZ_\theta$ is Pressurizer level, $QMWT_\theta$ is total thermal power, RM_θ is Rad Monitor for reactor building air.

As the K-NN was used as a classification method for this study, the use of 1st RSM data sets was really helpful to generate good results. The K-NN could dynamically estimate the relation between the five new accidents (new1, new2, new3, new4, and new5) with the 12 accidents stored in database. In table (3) we can see that for each step of time the K-NN estimates the type of malfunction that is really near to the actual scenario generated. The actual scenario for the accident named “new1” was LOCA-CL 20%, and the K-NN with the help of 1st RSM estimate this accident at the time it occurred as LOCA-CL 30%, at the second time step it estimated the accident as LOCA-CL 10%, and then at the third and fourth time steps it estimated the accident as LOCA-CL30% again.

Table 3. K-NN classification results of accident LOCA-CL 20% based on RSM

Class	Time of accident	K-NN prediction
New1	T-1	NormalOperation
New1	T	LOCA-CL30%
New1	T+1	LOCA-CL10%
New1	T+2	LOCA-CL30%
New1	T+3	LOCA-CL30%

For the other new accidents, the K-NN classifier with the help of 1st RSM also succeeded to give a perfect estimation of the type of malfunction with good estimation of the failure fraction.

K-means is a major clustering method; it can deal with many cases of data clustering. K-mean here clustered the new accident with respect to the stored accidents and normal operation clusters. Usually K-mean accuracy is decreasing if the clusters are more than two, but here K-means accuracy increases when the number of accident were increased. Table (3) is showing K-mean dynamic prediction results for a new accidents (new6).

Table 4. K-mean Clustering results of accident “new6” based on RSM

1st RSM data	K-mean clustering prediction				
	T-1	T	T+1	T+2	T+3
Normal Operation	C#0	C#0	C#0	C#0	C#0
LOCA-CL1%	C#0	C#1	C#1	C#1	C#1
LOCA-CL10%	C#0	C#2	C#2	C#2	C#2
LOCA-CL30%	C#0	C#3	C#3	C#3	C#3
LOCA-HL1%	C#0	C#3	C#3	C#3	C#3
LOCA-HL10%	C#0	C#4	C#4	C#4	C#4
LOCA-HL30%	C#0	C#4	C#4	C#4	C#4
SLBIC10 %	C#0	C#7	C#7	C#7	C#7
SLBIC30 %	C#0	C#8	C#8	C#8	C#8
SLBIC50 %	C#0	C#9	C#9	C#9	C#9
SLBOC10 %	C#0	C#10	C#10	C#10	C#10
SLBOC30 %	C#0	C#11	C#11	C#11	C#11
SLBOC50 %	C#0	C#12	C#12	C#12	C#12
new6	C#0	C#13	C#13	C#13	C#13

Where C refers to “Cluster”

Conclusions

As the time of emergency in NPP, a fast estimation and decision of which procedure will be followed to assist and mitigate the accident; a computational aid system would be helpful for operators to identify the initiation time of accident and to help them identifying the type of accidents using classification or clustering methods. This computational engine is based on the RSM specifically the 1st RSM that can identify the increasing and decreasing pattern with their orientation angle and can capture the initiation time of accident.

The 1st RSM was able to detect the initiation time of accident by the deviation of angle from the normal operation pattern sample. It was used as a transformation method for classification and clustering, too. Based on RSM, classification and clustering was successfully performed with good results using the K-NN classifier and K-mean clustering methods.

As it can make a contribution in the EOP, operators who would follow this aid system can identify the type of accident and its initiation time and they would also infer the range of failure fraction of an accident.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIP) (Grant Number: NRF-2016M2A8A2953047)

REFERENCES

[1] IAEA. (2006). Development and Review of Plant Specific Emergency Operating Procedures, 103.

[2] Kim, M. C., & Seong, P. H. (2008). A method for identifying instrument faults in nuclear power plants possibly leading to wrong situation assessment. *Reliability Engineering and System Safety*, *93*(2), 316–324.

[3] Kassim M. & Heo G., ESREL (2017). Detection and Classification of NPP Accidents Scenarios based on Residual Sign Matrices (RSMs).

[4] Trucco, E., & Verri, A. (1998). Introductory techniques for 3-D computer vision. Prentice Hall.

[5] Duda, R. O., Hart, P. E., Stork, D. G. (2001). Pattern classification. Wiley.

[6] Rapidminer. (2014). RapidMiner Studio Manual, 110.