Accident Classification and Clustering Using RSM in NPPs

Moath Kassim

Gyunyoung Heo

Mainformatics Lab.

경희대학교
KYUNG HEE UNIVERSITY

OBJECTIVES

The main goal of this study is to approach reliable methods for making a safety aid system that can help operators **to detect and identify the accident as soon as before the reactor trip**. This scope including:

- Identifying the exact time of accident using Residual Sign Matrix (RSM).
- Classifying known and stored plant accidents data using K-Nearest Neighbor (K-NN) classification method.
- Clustering known & unknown plant accident data using K-mean clustering method.

INTRODUCTION

- During Emergency Operating Procedure (EOP), operators detect an accident visually by tracing the important safety related parameters when they approach unacceptable limits. This would include human errors estimating the real time of accident.
- As time and human errors can contribute badly in plant safety during accidents, a safety aid system can be used to help operators and decision makers to identify and estimate the time and the type of accidents as fast as before the reactor trip.
- Residual Sign Matrix (**RSM**) method including 0^{th} and 1^{st} RSMs can help converting plant data into a trends and angle spaces, simulating "increasing & decreasing trends" and "fast or slow increasing & decreasing.
- In addition, when the 1^{st} RSM undergoes a classification or clustering process, it can deliver an estimation message of what kind of accident occurs.
- PWR simulation data sets were generated in steady state condition and used for analysis
- In-house code was developed to generate RSMs and store them, then Rapidminer software was used for the classification and clustering processing

0^{th} RSM vs. 1^{st} RSM

- 0^{th} RSM explain the trends of signals during accident (Kim & Seong) with respect to the normal condition as (+1) for increasing, (-1) for decreasing, and (0) for no change.
- It can be calculated from the residual between the normal and accident signals in time domain (t):

$$0^{th}RSM(i,j) = sig[X_{acc}(i,j) - X_{norm}(i,j)]$$

Where; X_{acc} is a plant variable value in accident time

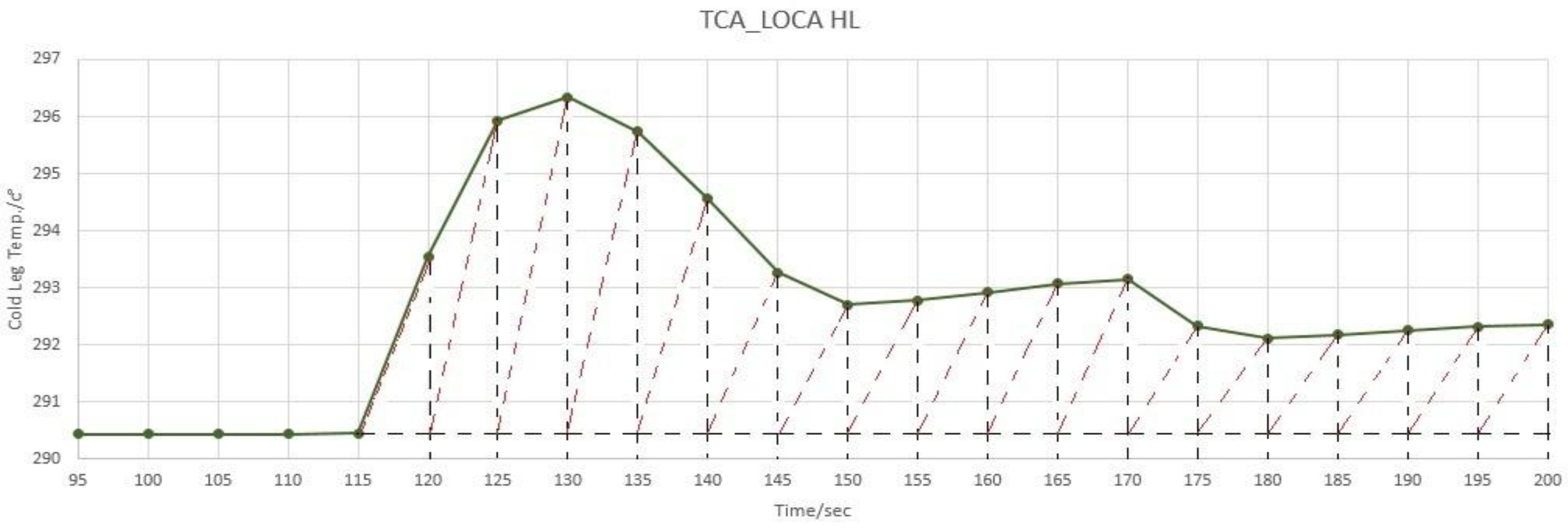
X_{norm} is a plant variable value at normal condition

- 0^{th} RSM was found not helpful in term of classification and clustering methods since it has only 3 values that make huge overlapping and decrease the accuracy of classification and clustering to minimum.
- 1^{st} RSM is novel idea inspired by the edge detection method of image processing.
- It is not accurate to only consider trends of signals when we deal with tens of signals coming at the same time; like in the NPP.

0^{th} RSM vs. 1^{st} RSM

- 1^{st} RSM calculates the angle of the residual between normal and accident signals in time domain (t):

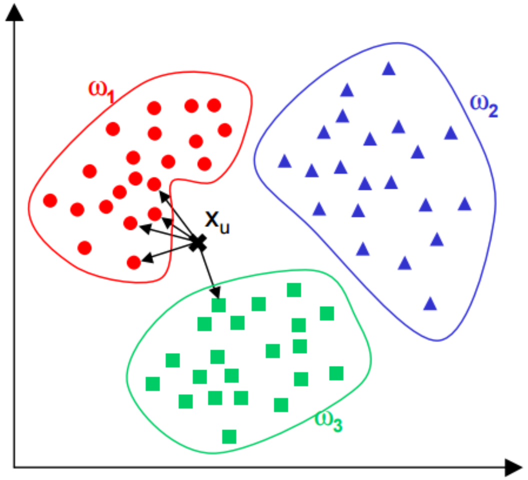
$$1st\ RSM(i,j) = tan^{-1}\left(\frac{X_{acc}(i,j) - X_{norm}(i,j)}{t(i) - t(i-1)}\right)$$



- 0^{th} RSM is imbedded in the calculation of 1^{st} RSM as the 1^{st} RSM generate the angles in term of (+), (-), and (0).
- Both 0^{th} RSM and 1^{st} RSM can detect the initiation of accidents by detecting the difference from upper or lower limits of normal operation accuracy band.

K-Nearest Neighbor

- Classification method that considers only a certain number of classes' nearest points (local neighborhood) to our new observation.
- Training samples are classified to certain classes using one of the similarity measure distances (metric) :



- Euclidian distance : $d(x_u, \omega_j) = \sqrt{\sum_i (x_{u,i} - \omega_{j,i})^2}$
- Manhattan distance: $d(x_u, \omega_j) = \sum_i |x_{u,i} - \omega_{j,i}|$

- The dimensions of each class are determined by:
  - K here refers to the nearest neighbor points to the new observation (k is a positive integer, typically small)
  - The similarity measure (such as Euclidian distance) is used to calculate the distance between the new observation point and the K neighbor points.

Mainformatics Lab

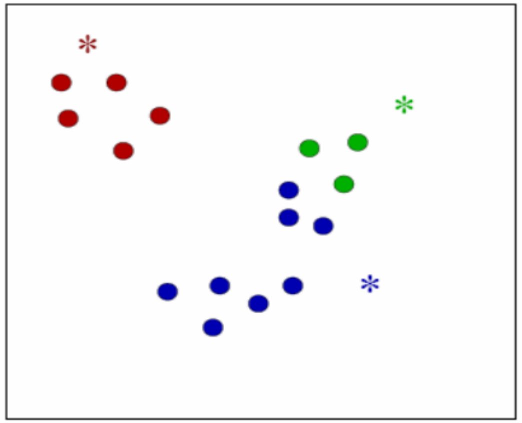
K-mean

- Clustering method that clusters data into group based on calculation of the mean of each group a long with the new observation.
- Calculation of the mean (centroid) is done for many iteration allocating and reallocating the mean.

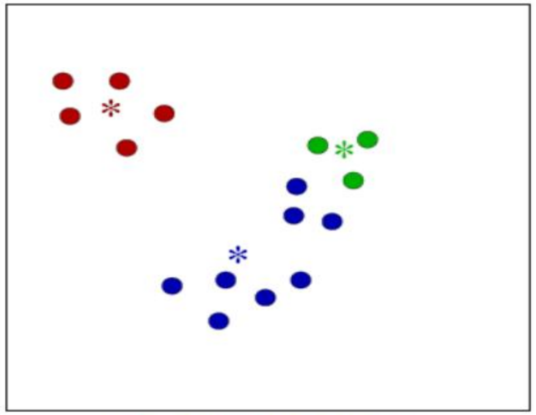
$$\mu_j = \frac{1}{n}\sum_{i=1,\, x_i \in w_j}^{n} x_i$$

where; (*)In the figures refers to the assigned and estimated mean (μ)

- The optimal (nearest) distance to the mean is the measure that allocates the data point in specific cluster.
- Euclidian distance is usually used to calculate the distance between the assigned mean and the surrounding points.
- K here refers to the number of clusters we wish to have.



Assign to nearest representative



Re-estimate means

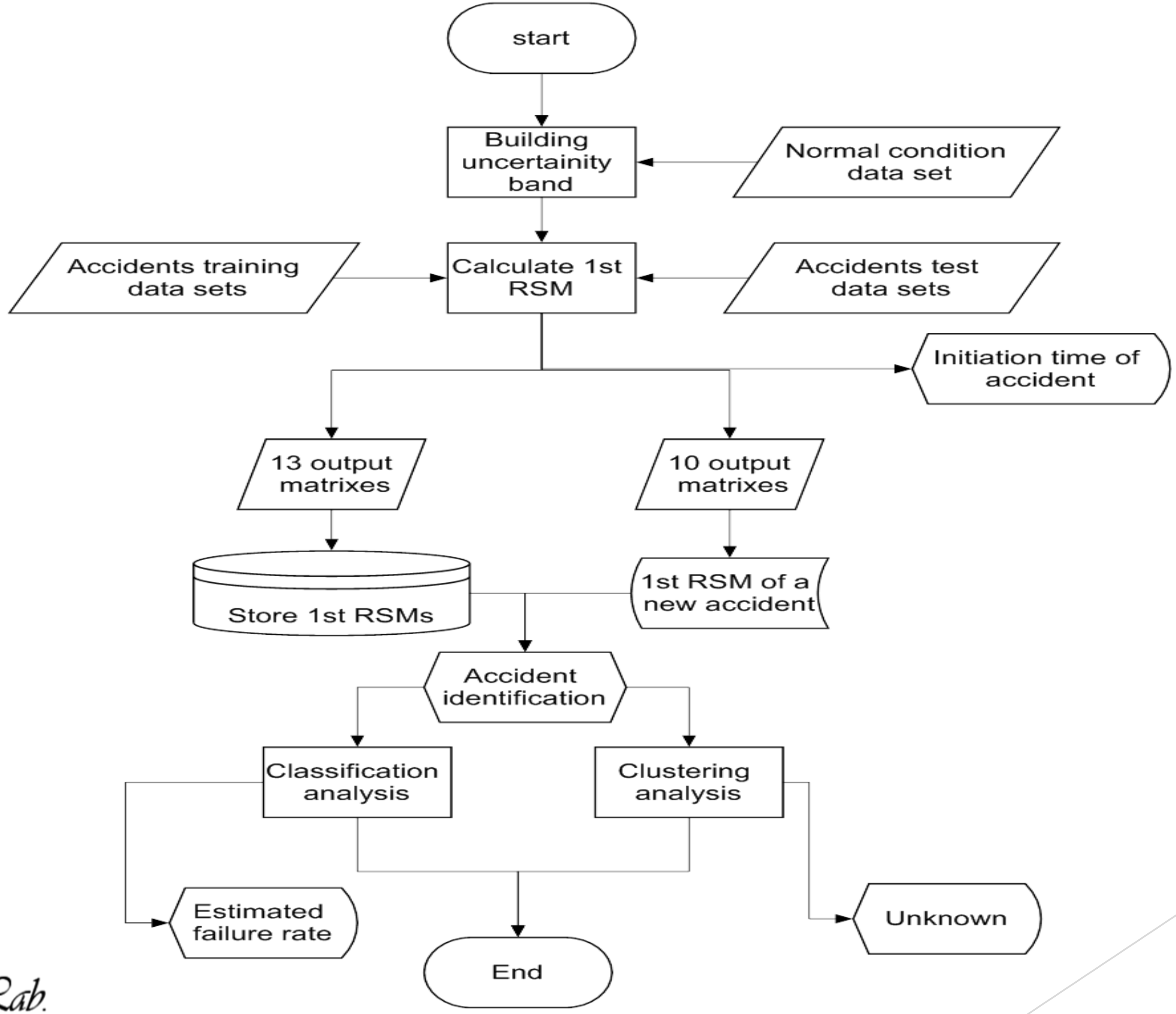
Mainformatics Lab.

Simulation Data

- PWR simulation data sets were collected with steady state plant condition of 100% power MOC.
- 23 data sets were obtained, and divided into two parts, training sets and test sets:
  - Training sets includes 1 set of normal operation data, and 12 accidents' sets.
  - Test data are 10 accidents' sets divided into two parts, 5 accidents sets were used for classification purpose and the other 5 sets were used for clustering.
- Each set was in a matrix form of (100sce,94 plant variables).
- The time interval of analysis before the initiation of accident to reactor trip was 5 time steps (25 sec).

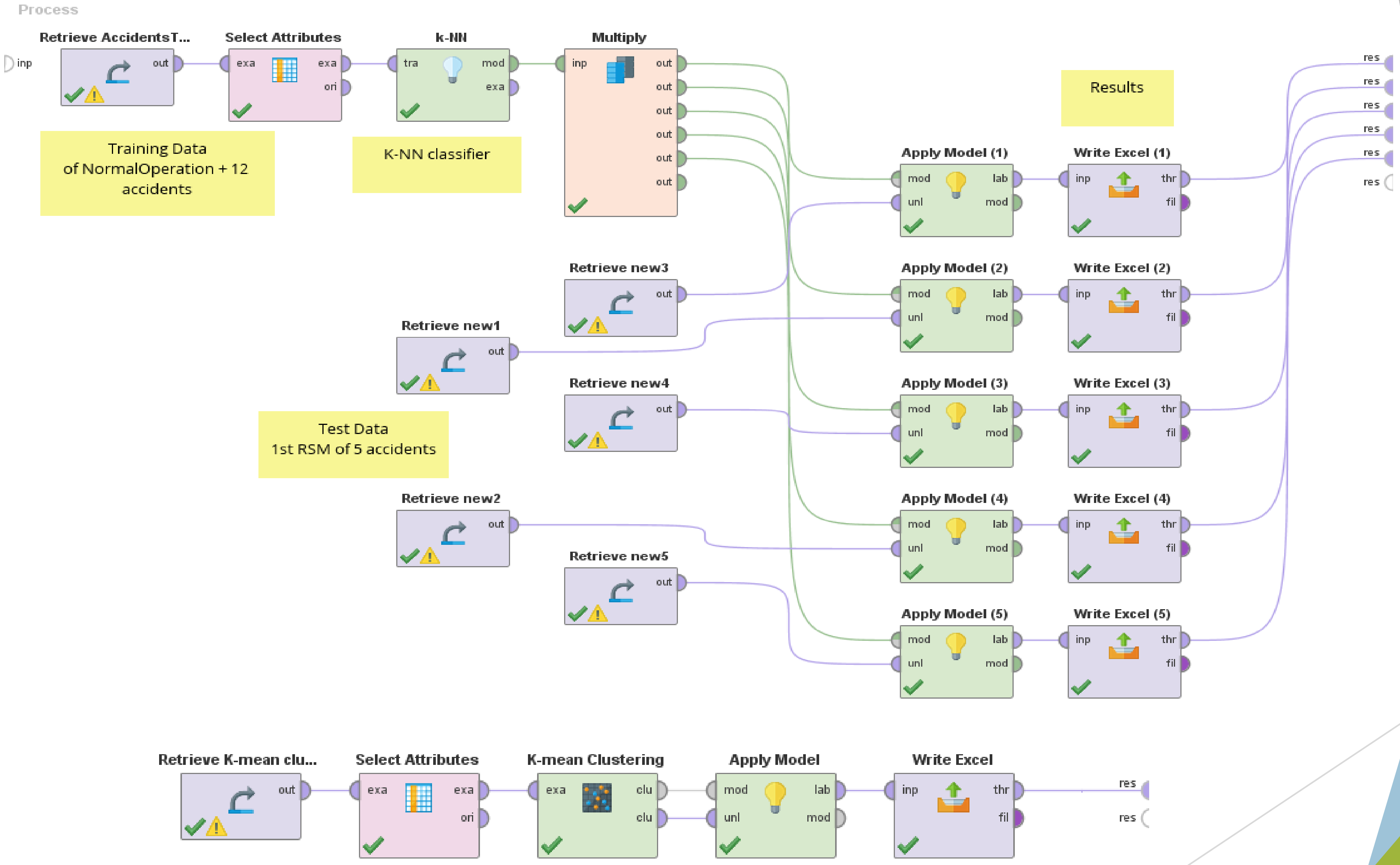
Cluster	Data set given name	Scenario	Failure rate (%)
0	NormalOperation	Normal condition	0
Training data sets			
1	LOCA-CL1%	LOCA Cold Leg	1
2	LOCA-CL10%		10
3	LOCA-CL30%		30
4	LOCA-HL1%	LOCA Hot Leg	1
5	LOCA-HL10%		10
6	LOCA-HL30%		30
7	SLBIC10%	Steam Line Break Inside Containment	10
8	SLBIC30%		30
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
15	New8	Steam Generator Tube Rupture	10
16	New9		30
17	New10		50

1st RSM Procedure



Mainformatics Lab.

Classification & Clustering Processes



Mainformatics Lab.

Classification Results

Accident new1	
Accident Time	Prediction(Class)
T-1	Normal Operation
T	LOCA-CL30%
T+1	LOCA-CL10%
T+2	LOCA-CL30%
T+3	LOCA-CL30%

Accident new2	
Accident Time	Prediction(Class)
T-1	Normal Operation
T	LOCA-HL1%
T+1	LOCA-HL1%
T+2	LOCA-HL1%
T+3	LOCA-HL1%

Accident new3	
Accident Time	Prediction(Class)
T-1	Normal Operation
T	LOCA-CL30%
T+1	LOCA-CL30%
T+2	LOCA-CL30%
T+3	LOCA-CL30%

Accident new4	
Accident Time	Prediction(Class)
T-1	Normal Operation
T	SLBIC30%
T+1	SLBIC30%
T+2	SLBIC30%
T+3	SLBIC30%

Accident new5	
Accident Time	Prediction(Class)
T-1	Normal Operation
T	SLBOC30%
T+1	SLBOC30%
T+2	SLBOC30%
T+3	SLBOC30%

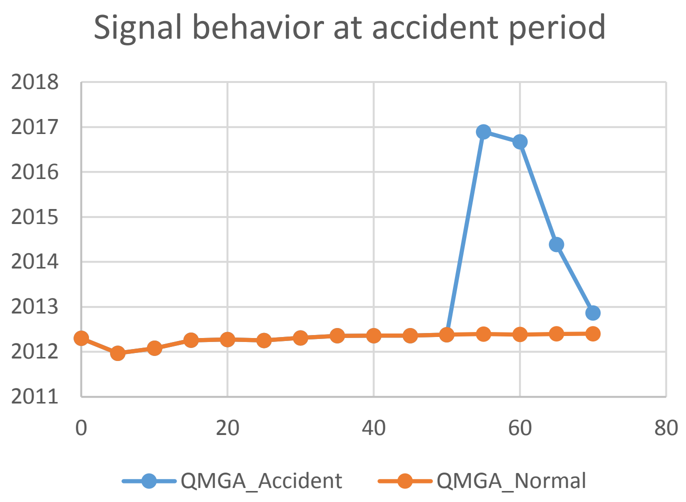
Clustering Results

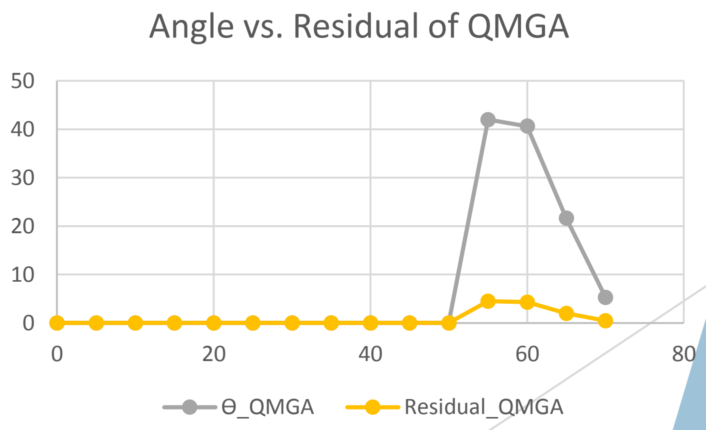
Accident	Accident Time				
	T-1	T	T+1	T+2	T+3
New6	Cluster#0	Cluster#13	Cluster#13	Cluster#13	Cluster#13
New7	Cluster#0	Cluster#13	Cluster#13	Cluster#13	Cluster#13
New8	Cluster#0	Cluster#13	Cluster#13	Cluster#13	Cluster#13
New9	Cluster#0	Cluster#13	Cluster#13	Cluster#13	Cluster#13
New10	Cluster#0	Cluster#13	Cluster#13	Cluster#13	Cluster#13

id	Class	cluster
1	NormalOpera...	cluster_0
2	LOCA-CL1%	cluster_1
3	LOCA-CL10%	cluster_2
4	LOCA-CL30%	cluster_3
5	LOCA-HL1%	cluster_4
6	LOCA-HL10%	cluster_4
7	LOCA-HL30%	cluster_4
8	SLBIC10%	cluster_7
9	SLBIC30%	cluster_8
10	SLBIC50%	cluster_9
11	SLBOC10%	cluster_10
12	SLBOC30%	cluster_11
13	SLBOC50%	cluster_12
14	new10	cluster_13

Pros and Cones;1st RSM

- ❑ Advantages:
- Dimension reduction
- Uniting scales.
- Uniting physical attributes (parameter) in one space; one unit.
- Angles when drawn as a dimension in time domain shows similar behavior of the real accident signal behavior (but not exactly similar).
- It shows the **approximate** change of signals increasing or decreasing in term of angles.
- 1st RSM can be performed dynamically keeps the information of trend directions as they are.





Pros and Cones:1st RSM

- 1st RSM gives better results than analyzing raw data
- Expected delay of calculation is one sec, and computational time cost may exceeds 3 minutes for training the data, and less than 3 sec for identifying the accident. (the more data the more computational time)

❑ Limitation:

- It is a combination method based on normal and abnormal signals; so it can not be generated without them both.
- Analysis of 5 time steps are not enough to considered reliable unless LOCA accident failure rate decreased up to 10 or 20%

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

Accident new4	
Accident Time	prediction(Class)
T-1	Normal Operation
T	SLBIC30%
T+1	SLBIC30%
T+2	SLBIC30%
T+3	SLBIC30%

Accident new4	
Accident Time	prediction(Class)
T-1	Normal Operation
T	LOCA-CL1%
T+1	SLBIC10%
T+2	SLBIC30%
T+3	SLBIC30%

Accident new5	
Accident Time	prediction(Class)
T-1	Normal Operation
T	SLBOC30%
T+1	SLBOC30%
T+2	SLBOC30%
T+3	SLBOC30%

Accident new5	
Accident Time	prediction(Class)
T-1	Normal Operation
T	SLBIC10%
T+1	SLBOC10%
T+2	SLBOC30%
T+3	SLBOC30%

CONCLUSION & FUTURE WORK

- Time of accidents was identified by 0th RSM & 1st RSM.
- Normal condition and accident data were converted into an angle space (1st RSM).
- 1st RSM method would be applied also to start-up/shut down condition with the implementation of angle shift.
- Classification was performed using K-NN and 1st RSM giving better prediction than using K-NN with raw data.
- K-mean was performed successfully to detect a new accident that is not belonging to the stored accidents, but it couldn't cluster the stored accidents correctly.
- **Work to be done in the near future:**
  - Appling reliable simulation data to satisfy the V&V, and enhancing 1st RSM by performing preprocessing techniques to qualified data before main processing.
  - Clustering by K-mean should consider all stored data as one group and the new accident data as a second group.
  - Other classification & clustering methods would be used with 1st RSM to check the validity 1st RSM with machine learning algorithms.

Thank You