Pattern Identification of Dynamic Event Tree Scenarios with Clustering

Junyung KIM[a] and Hyun Gook KANG[a]
[a] Rensselaer Polytechnic Institute, Troy, USA

Abstract: The large number of scenarios generated from a single initiating event due to time-related scenario evolutions has been a challenge of dynamic probabilistic risk assessment (DPRA) for years. Many researchers have reported that the risk insight which enhances safety of the entire system requires not only the full understandings of scenario evolutions but also the principal characteristics of the events. Since the time-related scenario evolution brings a lot of difficulties not only to organize such large amounts of information but also to analyze and interpret its physical meaning, clustering analysis has been considered to be useful to group scenarios with similar characteristics and to identify key features of each group so that an analyst can understand entire scenario behaviors by groups. The performance of clustering analysis is highly related on which distance matrix is used. For a given scenario dataset, this paper will perform clustering analysis with global alignment (GA) kernel distance matrix specialized for time-series data so as to identify and classify scenarios generated in a dynamic event tree (DET) analysis. Global alignment kernel is to assess similarity between time series data by casting the dynamic time warping (DTW) distances and similarities as positive definite kernels. In this study, 2500 scenarios are generated by the MOSAIQUE (Module for SAmpling Input and QUantifying Estimator) coupled with RELAP5 that simulates the thermal hydraulic behavior of the nuclear power plant (NPP), APR-1400. An application is considered with regards to the scenarios generated following a small break LOCA event in the NPP. The proposed classification and identification approach has grouped the 2500 scenarios with 53 clusters and the result can show marginal grace time of safety component which static PRA cannot present.

Keywords: Scenario clustering, Dynamic PRA, Transient analysis, Mitigation coverage

1. INTRODUCTION

Dynamic methodologies for probabilistic risk assessment (PRA) are those that account for possible coupling between triggered or stochastic event through explicit consideration of the time element in the system evolution [1]. There are many dynamic PRA methods and dynamic event tree (DET) is used in many of them. DET is to determine the risk associated with sophisticated plant system and seeks to timing and process relationships in the stochastic system. The branching conditions in a DET are decided by preliminarily specified rules, such as success/failure of safety components on demand, system recovery timing or when state variables reach predefined set points during the simulation.

The one of the major challenges in DET is due to the large number of scenarios generated and many researchers have already pointed out that the number of scenarios in DET to be analysed is much larger than that of the classical fault/event tree approaches [2][3]. Each branch in a random scenario can contain time evolutions of a large number of variables and hence infinite number of scenarios can be generated theoretically. Such large amounts of information are very difficult to organize and interpret in regard to the main trends in scenario evolutions and the main risk contributors for each initiating event [4]. In order to tackle aforementioned challenge, clustering analysis has been considered a possible solution for many years. In 2005, Kopustinkas et al. [5] proposed to group the DET-generated scenarios in classes of "similarity", by combining information from both the event sequences and the patterns of evolution of the process variables and in 2013, Diego Mandelli et al. [1] proposed a solution which is to partition the set of scenarios into groups, called clusters, and analyze each group individually rather than all the scenarios simultaneously. The partition is performed by identifying similarities among scenarios and grouping them according to predefined similarity criteria.

Since clustering analysis has been deemed a possible solution for handling a large number of scenarios, several clustering methods have been introduced with dynamic event trees. In 2009, fuzzy c-means algorithm coupled with possibilitic clustering method is proposed to group the scenarios from a DET for a steam generator tube rupture event [6] and in 2013, another clustering algorithm, the mean-shift (MS) method has been proposed in DETs [1]. However, both approaches have limitations: Fuzzy c-means algorithm is susceptible to noisy data and unable to identify clusters that are non-spherical and mean-shift method is sensitive to both the choice of clustering parameter (bandwidth) and to noise in the data. In addition, both algorithms are based on Euclidean distance, which is not adequate for clustering time series data. For most time series data mining algorithms, the quality of the output depends almost exclusively on the distance measure used [7] and many in the research community have determined that DTW is a superior choice as a time series distance measure, and it has been found to regularly outperform the Euclidean distance [8].

The dynamic time warping distance, however, has limitation to be considered as a distance and similarity measurement [9] as well in that (1) DTW distance does not satisfy the triangular inequality and (2) similarity function with the dynamic time warping distance is not positive definite, which is against the condition to be a similarity function that Jeff M. Phillips et al [11] mentioned.

This paper will present a scenario clustering algorithm with partition around medoids (PAM) and global alignment kernel distance which does not have the issues that DTW distance has. With the clustering algorithm, identification of the scenarios which have a similar behavior and association of scenarios within one cluster will be done first. After clusters are defined, similarity among the scenarios in each cluster will be identified and key patterns of clusters are discussed. With this analysis, clusters can help understand how the changes in sequence or timing of variables action (such as sequence of safety system running and recovery timing of firstly failed safety system) impact the overall system dynamics.

2. PAM clustering algorithm with Global Alignment Kernel distance

2.1 K-Medoids

Clustering is the process of grouping a set of objects into clusters so that objects within a cluster are similar to each other but are dissimilar to objects in other clusters [12]. The k-medoids algorithm is a medoid-shift clustering algorithm which is related to k-means algorithm [19]. Both the k-means and k-medoids algorithms are partitional and attempt to minimize the distance between points labeled to be in a cluster and a point designated as the center of that cluster. However, K-means clustering is known “sensitive” to the outliers even though its computation time is quite efficient. For this reason, K-medoids clustering are sometimes used, where representative objects called medoids are considered instead of centroids. Because it is based on the most centrally located object in a cluster, it is less sensitive to outliers in comparison with the K-means clustering [13].

2.2 Partitioning around Medoids (PAM) algorithm

The most common realization of K-medoids clustering is the partitioning around medoids (PAM) algorithm [14] and is as follows:

1) Initialize: randomly choose K of the n data points as the medoids
2) Associate each data point to the closest medoid. ("closest" here is defined using any valid distance metric)
3) For each medoid m
4) For each non-medoid data point
   i) Swap m and o
   ii) Compute the total cost of the configuration
5) Select the configuration with the lowest cost.
6) Repeat steps 2 to 5 until there is no change in the medoid.

2.3 Similarity of time series data

Cuturi [9] proposed an algorithm to assess similarity between time series using kernels, which is called global alignment (GA) kernel. The algorithm starts by formalizing an alignment between two time series x and y as π, and defined the set of all possible alignments as A(n,m), which is constrained by the lengths of x and y. It is shown that the DTW distance between x and y which is expressed in the simplest form in equation (1) can be understood as the cost associated with the minimum alignment [10].

$$\mathrm{DTW}(\mathrm{x},\mathrm{y}) = \min_{\pi\in A(n,m)} D_{x,y}(\pi) \tag{1}$$

The cost

$$D_{x,y}(\pi) = \sum_{i=1}^{|\pi|} \varphi(x_{\pi 1(i)} y_{\pi 2(i)}) \tag{2}$$

is defined by a local divergence φ, that measures the discrepancy between any two points x_i and y_j observed in x and y where |π| is the length of π.

2.4 Global Alignment Kernel distance

A Global Alignment (GA) kernel is defined as in equation 3(a) and 3(b), where κ is a local similarity function. Cuturi et al. [15] argue that the similarity described by K_{GA} incorporates the whole spectrum of costs and provides a richer statistic than the minimum of that set, which is the sole quantity of DTW distance. In contrast to DTW, this kernel considers the cost of all possible alignments by computing their exponentiated soft-minimum, so it is argued that it quantifies similarities in a more coherent way [10].

$$K_{GA}(x,y) \stackrel{\text{def}}{=} \sum_{\pi\in A(n,m)} e^{-D_{x,y}(\pi)} \tag{3a}$$

$$K_{GA}(x,y) = \sum_{\pi\in A(n,m)} \prod_{i=1}^{|\pi|} \kappa(x_{\pi 1(i)} y_{\pi 2(i)}) \tag{3b}$$

In order to reduce the GA kernel's complexity, Cuturi (2011) [9] proposed using the triangular local kernel for integers shown in equation (4), where T represents the kernel's order. By combining it with the kernel κ in equation (5), the Triangular Global Alignment (TGA) kernel in equation (6) is obtained. Such a kernel is parameterized by the triangular constraint T and the Gaussian's kernel width σ [10].

$$\omega(i,j) = \left(1 - \frac{|i-j|}{T}\right)_{+} \tag{4}$$

$$\kappa(x,y) = e^{-\Phi_\sigma(x,y)} \tag{5a}$$

$$\Phi_\sigma(x,y) = \frac{1}{2\sigma^2}\|x-y\|^2 + \log\left(2 - e^{\frac{-\|x-y\|^2}{2\sigma^2}}\right) \tag{5b}$$

$$\mathrm{TGAK}(\mathrm{x},\mathrm{y},\sigma,\mathrm{T}) = \tau^{-1}\left(\omega \otimes \frac{1}{2}\kappa\right)(i,x;j,y) = \frac{\omega(i,j)\kappa(x,y)}{2-\omega(i,j)\kappa(x,y)} \tag{6}$$

The similarity returned by the TGAK can be normalized with equation (7) so that its values lie in the range [0, 1]. Hence, a distance measure for time-series can be obtained by subtracting the normalized value from 1. The resulting distance is symmetric and satisfies the triangle inequality. The interested reader is referred to Cuturi (2011) [9] for an extensive literature review.

$$\text{similarity} = \exp\left(\log\big(\mathrm{TGAK}(x, y, \sigma, T)\big) - \frac{\log(\mathrm{TGAK}(x,x,\sigma,T)) + \log(\mathrm{TGAK}(y,y,\sigma,T))}{2}\right) \quad (7)$$

3. Identification and Classification of key features

3.1 The Algorithm

The objects to be classified are the DET scenarios which have been generated by system analysis code. The basic steps for their classification are illustrated in Figure 1. The Module for SAmpling Input and QUantifying Estimator (MOSAIQUE) [16] is used in this study as the DET generator tool while system dynamics was modelled using RELAP5 [17]. Variables of interest such as the number of safety injection (SI) pump available and recovery timing of failed SI pump are selected and relevant randomness are given in each variable. After running the entire thermal hydraulic simulation codes generated by MOSAIQUE, system parameters such as core liquid volume and water level in pressurizer are selected for clustering purpose. Since none of the clustering algorithm automatically decides the optimal number of clusters, an analyst should utilize additional information: end state of each scenario (core failure or core safe). Every scenario is either in core failure state or core safe state after the simulation running time. In general, it is considered as core failure if peak cladding temperature (PCT) is greater than 1477K. The novel algorithm using PCT of each scenario proposed in Table 1 can help decide the optimal number of clusters. The summation term for calculating ε_j in Table 1 starts from i =2 since i denotes cluster number and the minimum possible cluster number is two.

Figure 1: the scenario classification approach

	Procedure	Examples
1 Variable Selection	• Select variables in focus	• The number of SI pump successful at the beginning • Recovery time of failed SI pump
2 Assigning Randomness	• Decide randomness (distribution) of success/fail indicator for selected variables • Recovery timing is randomly assigned	• SI pump success / fail — Uniform(-0.1,0.9)* • Recovery timing of firstly failed component — Between 1,800 sec and 10,000 sec after failure notification
3 Running Scenarios	• Select the number of scenarios • Generate scenarios with selected variables — RELAP 5 embedded in MOSAIQUE • Run scenarios (TH codes)	• 2,500 scenarios with different variables run
4 *Select system parameters for clustering*	• Selecting system parameters for clustering algorithm — Selected system parameters will be embedded in clustering algorithm.	• Water level of S/G, water level of Vessel, S/G pressure of simulation data are selected for clustering purpose.
5 Classification (Clustering)	• Cluster scenarios with clustering algorithm • Evaluate clustering result • Choose the number of clusters	• TGAK algorithm is utilized. • The number of clusters from 2 to 60 are tested. • The number of clusters are decided after evaluation.
6 Post Processing	• Investigate relations between clusters • Analyze major contributor of a specific cluster	• In SBLOCA, SI must be recovered within 3,500 sec if non of safety valves are worked.

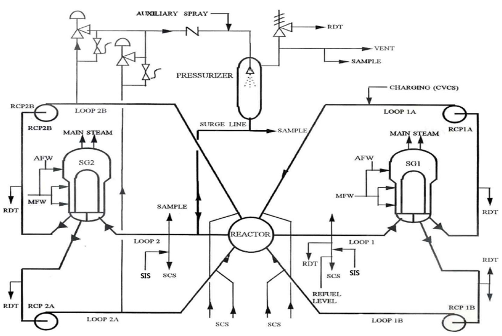
3.2 Case Study: SBLOCA in a nuclear power plant

For the case study, actual nuclear power plant system, APR-1400 is used for the illustration of the proposed clustering methodology. APR-1400 is a standard evolutionary advanced light water reactor and Figure 2 shows the system configuration of an APR-1400 [18]. In this case study, one top event has been considered: peak cladding temperature reaches the limit of 1477K. Figure 3 shows the temporal behavior of the PCT temperature for all the 2500 scenarios. Several assumptions are made as below:

Table 1: the algorithm of deciding the optimal number of clusters

Input	**N**	total number of scenarios
	i	The number of clusters
	j	the largest number of cluster in a cluster group (e.g., entire scenarios are grouped with 33 clusters, j is 33)
	k	the number of cluster to be tested
	$n_{1477\uparrow,\,\mathrm{i}}$	the number of scenarios where PCT reached in 1477K in i-th cluster
	$n_{1477\downarrow,\,\mathrm{i}}$	the number of scenarios where PCT did not reached in 1477K in i-th cluster
Output	ϵ_{j}	the error rate of the case where the largest number of clusters is j
1	For j in 2:k	
2	$\varepsilon_j = \dfrac{\sum_{i=2}^{j} \min(n_{1477\uparrow,i}, n_{1477\downarrow,i})}{N}$	
3	End	
4	Select smallest j for the number clusters which has ϵ_{j}=0.	

Figure 2: the system configuration of a PWR plant



1) At time zero, small break loss of coolant accident (SBLOCA) (break size is 0.002027 m^2) happens at cold leg.
2) Mission time for this system analysis has been set to 10,000 sec. In the original dataset, however, scenarios that reach PCT of 1477K are stopped even though they did not reach to 10,000 sec. For those scenarios that did not reach 10,000 sec, it was decided to extend in time these scenarios up to 10,000 sec with the last value simulated.
3) In this study, the status of seven safety components are of interest: safety injection (SI) pump (4 units), auxiliary feed water pump (2 units), shutdown cooling pump (SCP) (2 units), safety injection tank (SIT) (4 units), atmosphere dump valve (ADV) (4 units), main steam safety valve (MSSV) (6 units), and pilot operated safety relief valve (POSRV) (8 units).[*] Each component is in one of the three states: [1)] Work well when it is requested, [2)] fail first and then recovered after a certain time period or [3)] fail to be recovered until after the mission time. There is a uniform distribution of recovery timing between 1,800 sec and 10,000 sec after SBLOCA happens. In total, 60 variables are selected and the variables of interest are listed in Table 2. The recovery time refers to the amount of time required to fix the component after the SBLOCA event begins.
4) The availability and the recovery timing of components are assumed to be independent.
5) Data from each scenario are monitored at every 60 second
6) For the clustering purpose, 6 system parameters are used: [1)]steam generator A dome pressure, [2)]steam generator B dome pressure, [3)]steam generator A inventory level, [4)]steam generator B inventory level, [5)]core liquid volume and [6)]pressurizer collapsed water level.

[*] In this research, pilot involved operation of POSRV is not considered for the simplicity of simulation. POSRV is only open at a predetermined condition: Pressure in pressuriser is greater tha17.23686×10^6 Pa.

Table 2: variables of interest

Number	Variable	Distribution
1	AFW Pump 1 success/fail indicator*	Uniform(-0.1, 0.9)
2	AFW Pump 2 success/fail indicator	Uniform(-0.1, 0.9)
3	SI Pump 1 success/fail indicator	Uniform(-0.1, 0.9)
4	SI Pump 2 success/fail indicator	Uniform(-0.1, 0.9)
5	SI Pump 3 success/fail indicator	Uniform(-0.1, 0.9)
6	SI Pump 4 success/fail indicator	Uniform(-0.1, 0.9)
7	SC pump 1 success/fail indicator	Uniform(-0.1, 0.9)
8	SC pump 2 success/fail indicator	Uniform(-0.1, 0.9)
9	SIT 1 success/fail indicator	Uniform(-0.1, 0.9)
10	SIT 2 success/fail indicator	Uniform(-0.1, 0.9)
11	SIT 3 success/fail indicator	Uniform(-0.1, 0.9)
12	SIT 4 success/fail indicator	Uniform(-0.1, 0.9)
13	ADV 1 success/fail indicator	Uniform(-0.1, 0.9)
14	ADV 2 success/fail indicator	Uniform(-0.1, 0.9)
15	ADV 3 success/fail indicator	Uniform(-0.1, 0.9)
16	ADV 4 success/fail indicator	Uniform(-0.1, 0.9)
17	MSSV 1 success/fail indicator	Uniform(-0.1, 0.9)
18	MSSV 2 success/fail indicator	Uniform(-0.1, 0.9)
19	MSSV 3 success/fail indicator	Uniform(-0.1, 0.9)
20	MSSV 4 success/fail indicator	Uniform(-0.1, 0.9)
21	MSSV 5 success/fail indicator	Uniform(-0.1, 0.9)
22	MSSV 6 success/fail indicator	Uniform(-0.1, 0.9)
23	POSRV 1 success/fail indicator	Uniform(-0.1, 0.9)
24	POSRV 2 success/fail indicator	Uniform(-0.1, 0.9)
25	POSRV 3 success/fail indicator	Uniform(-0.1, 0.9)
26	POSRV 4 success/fail indicator	Uniform(-0.1, 0.9)
27	POSRV 5 success/fail indicator	Uniform(-0.1, 0.9)
28	POSRV 6 success/fail indicator	Uniform(-0.1, 0.9)
29	POSRV 7 success/fail indicator	Uniform(-0.1, 0.9)
30	POSRV 8 success/fail indicator	Uniform(-0.1, 0.9)
31	AFW Pump 1 recovery time**	Uniform(1,1800, 20,000)
32	AFW Pump 2 recovery time	Uniform(1,1800, 20,000)
33	SI Pump 1 recovery time	Uniform(1,1800, 20,000)
34	SI Pump 2 recovery time	Uniform(1,1800, 20,000)
35	SI Pump 3 recovery time	Uniform(1,1800, 20,000)
36	SI Pump 4 recovery time	Uniform(1,1800, 20,000)
37	SC Pump 1 recovery time	Uniform(1,1800, 20,000)
38	SC Pump 2 recovery time	Uniform(1,1800, 20,000)
39	SIT 1 recovery time	Uniform(1,1800, 20,000)
40	SIT 2 recovery time	Uniform(1,1800, 20,000)
41	SIT 3 recovery time	Uniform(1,1800, 20,000)
42	SIT 4 recovery time	Uniform(1,1800, 20,000)
43	ADV 1 recovery time	Uniform(1,1800, 20,000)
44	ADV 2 recovery time	Uniform(1,1800, 20,000)
45	ADV 3 recovery time	Uniform(1,1800, 20,000)
46	ADV 4 recovery time	Uniform(1,1800, 20,000)
47	MSSV 1 recovery time	Uniform(1,1800, 20,000)
48	MSSV 2 recovery time	Uniform(1,1800, 20,000)
49	MSSV 3 recovery time	Uniform(1,1800, 20,000)
50	MSSV 4 recovery time	Uniform(1,1800, 20,000)
51	MSSV 5 recovery time	Uniform(1,1800, 20,000)
52	MSSV 6 recovery time	Uniform(1,1800, 20,000)
53	POSRV 1 recovery time	Uniform(1,1800, 20,000)
54	POSRV 2 recovery time	Uniform(1,1800, 20,000)
55	POSRV 3 recovery time	Uniform(1,1800, 20,000)
56	POSRV 4 recovery time	Uniform(1,1800, 20,000)
57	POSRV 5 recovery time	Uniform(1,1800, 20,000)
58	POSRV 6 recovery time	Uniform(1,1800, 20,000)
59	POSRV 7 recovery time	Uniform(1,1800, 20,000)
60	POSRV 8 recovery time	Uniform(1,1800, 20,000)

* If indicator is greater than 0, component does not work.

** A component is recovered after the recovery time, which is follow uniform distribution. If recovery timing is bigger than 10,000 sec, it implies that a component has not recovered until the end of mission time.

After running scenarios with 60 variables shown in Table 2, clustering analysis is followed using the 6 system parameters stated above. For calculating similarity presented in equation (7), σ is set as

$$\sigma = med(||x - y||) \cdot \sqrt{med(|x|)} \tag{8}$$

which is a strategy that Cuturi [9] proposed and T is given without constraint. After clustering analysis, it turns out that 53 are optimal number for clusters: 53 is the first total number of clusters with $\varepsilon = 0$. Figure 4 shows the ε with different number of clusters.

Figure 3: graphical representation of PCT of the scenarios

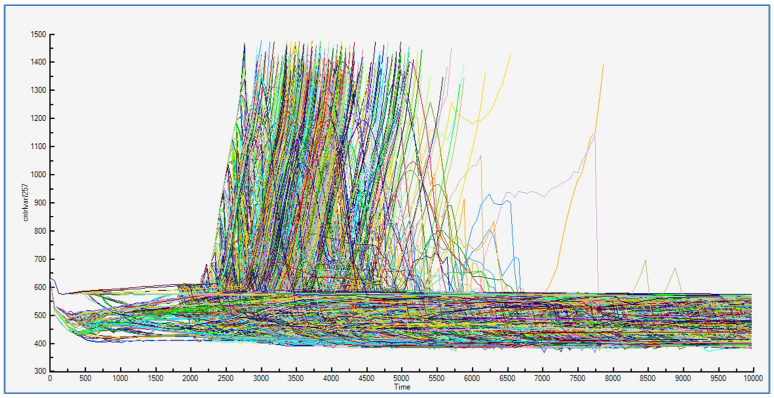
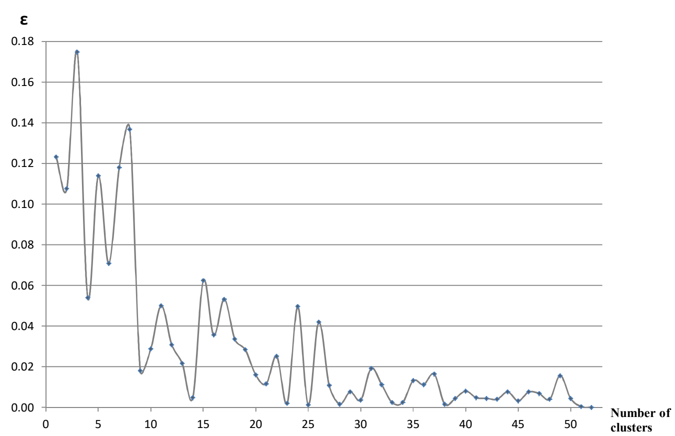


Figure 4: clusters numbers with ε



At this point, it is possible to analyse the characteristics of the clusters rather than the full dataset individually. It is assumed that there shall be principal characteristics to decide core failure or success. In this study, the timing of first unit of component j available is mainly considered to find key features of each cluster. In order to find key features of each cluster, novel index, called influence factor (IF), is developed as below:

$$IF_{component\ j\ of\ cluster\ i} = 1 - \frac{0.5 \times \exp\left(-\frac{1+Var_{component\ j,\ all}}{1+Var_{component\ j,\ cluster\ i}}\right)}{\exp(-1)} \tag{9}$$

where $Var_{component\ j,\ all}$ is variance of the timing of first unit of component j available in all scenarios and $Var_{component\ j,\ cluster\ i}$ is variance of the timing of first unit of component j available in cluster i. In equation (9), it is assumed that component j played major role in cluster i, or in other words, component j's behavior is a key characteristic of scenarios in cluster i in case that IF is close to 1. In contrary, if component j does not play major role in cluster i, then IF will be equal or less than 0.5. Table 3 and Table 4 show the result and Table 4 indicates the following:

- SI pump, ADV, MSSV, and auxiliary feed water pump are major components which affect clustering scenarios. In contrary, SCP, SIT and POSRV have minor impact on clustering SBLOCA event scenarios.
- It is noticeable that mean timings of the first unit of SI unit available when core fail statuses are "safe" (safe group) are less than 3500 seconds after SBLOCA happens. The mean timings of the first unit of SI unit available when core fail statuses are "fail" (failure group) stand in contrast. In most clusters in failure group, the numbers are bigger than 5800 seconds except a few outliers. Therefore, it is safe to say that fist unit of SI pump should be recovered no more than 3500 seconds on average.
- Especially, the timing of the first unit of SI pumps available gets important if the timing of the first unit of auxiliary feed water pumps available get delayed. Comparing cluster 19 in safe group and cluster 25 in failure group in Table 3 and Table 4, it can be concluded that as long as the first SI pump is available within 1600 sec on average, the operators can have grace time within around 7000 seconds for making the first unit of aux feed water pumps available.
- Cluster 10, 40 and 50 in core failure group indicate that core may get failed if ADV and MSSV are not available (stuck closed) within around 5000 seconds on average even if first unit SI pump out of four units is available right after SBLOCA.

4. CONCLUSTION

Many researchers have tried to address the problem of identifying and grouping scenarios resulting from dynamic safety and reliability assessments for which the number of scenarios that are simulated is much larger than that of the classical fault/event tree approaches.

In this paper, a methodology based on clustering algorithm with GA kernel based distance matrix which has been proven to be more accurate for clustering time series data is proposed. An analyst may need information of the plant system in order to find the vulnerable point of the safety system and the marginal grace time for repairing safety-related component. For this information to be of practical use, the scenarios should be grouped together on the basis not only of the occurred events and their end states as done in the classical fault/event tree approaches but also of the physical evolution of the process variables, which may depend on the order and timing of occurrence of the events [6].

The case study presented in the paper considers scenarios generated in a dynamic event tree analysis of a SBLOCA event. A total of 2500 scenarios are generated, resulting from the combination of the possible realization of safety component: success, failure and repaired. The PAM clustering algorithm with GA kernel based distance is used for clustering 2500 scenarios. New algorithm deciding the optimal number of clusters is proposed and 53 clusters are generated after clustering process. Novel index for measuring influence level of each safety component in a cluster is developed.

It is shown that grouping scenarios into clusters may be a resort to identify and evaluate the key characteristics of the entire dataset. The methodology presented in the paper may give an aid to an analyst to understand the characteristics of each cluster better and apply the result for finding recovery priority of the safety components and its marginal grace time for recovery.

Table 3: IF of components in each cluster*

K-number	Aux FP	SI	SCP	SIT	ADV	MSSV	POSRV	Core Fail
3	0.646371	**1**	0.526712	0.437835	**1**	**0.868517**	0.470322	Safe
4	0.690876	**0.996791**	0.42988	0.557764	**1**	**0.734935**	0.442253	Safe
6	0.544335	**0.966202**	0.638127	0.517412	0.60855	0.697685	0.547238	Safe
7	**0.798049**	**0.865012**	0.612585	0.474229	**0.791628**	0.477204	0.427014	Safe
11	0.582621	**0.773381**	0.516235	0.315024	**0.760609**	**0.993502**	0.52605	Safe
12	**0.917561**	**0.82512**	0.450314	0.658776	**0.996188**	0.396277	0.544812	Safe
16	0.575146	**0.912888**	0.496333	0.499161	**0.998592**	0.519354	0.687422	Safe
17	**0.758911**	**0.999507**	0.528873	0.610588	**1**	**0.703852**	0.69568	Safe
19	**0.870212**	**1**	0.480732	0.36383	**0.918533**	**1**	0.058527	Safe
21	**0.767465**	**1**	0.496869	0.573261	**1**	0.434527	0.415733	Safe
22	0.597834	**0.946081**	0.58687	0.474971	**0.889121**	**0.890969**	0.517287	Safe
23	0.447878	**0.990907**	**0.740057**	0.506381	0.400544	**1**	**0.754987**	Safe
24	0.648202	**0.968632**	0.525485	0.510644	**0.881964**	0.693814	0.55749	Safe
26	0.537891	**1**	0.494798	0.529123	**1**	0.372105	0.492823	Safe
27	0.512786	**0.809907**	0.480875	0.49776	**0.980668**	0.28832	0.469414	Safe
28	**0.934935**	0.317929	0.461364	0.426198	**0.980822**	0.630963	0.37269	Safe
30	0.482637	**0.784302**	0.532283	0.496718	**0.81008**	**0.725733**	0.621754	Safe
33	0.460475	**0.979087**	0.43368	0.505023	**0.952191**	**0.971108**	0.224587	Safe
34	0.504218	**0.812427**	0.412073	0.527199	0.574999	0.438996	0.320715	Safe
35	0.53227	**0.989846**	0.391875	0.637681	**1**	0.662777	0.4478	Safe
36	**0.981847**	0.592328	0.506616	0.515991	**0.936138**	0.613595	0.301908	Safe
37	0.57618	**0.964217**	0.6004	0.53346	**0.789074**	0.681297	0.388058	Safe
38	0.57502	**0.829476**	0.51247	0.569054	**0.965838**	**0.744086**	0.59757	Safe
39	0.508675	**0.8135**	0.502652	0.569146	0.628613	**0.943816**	0.608564	Safe
41	**0.864454**	**0.997797**	0.514999	0.385986	**0.700505**	**1**	0.516858	Safe
42	0.416191	**0.999996**	0.552259	0.443779	**0.783862**	**1**	0.581906	Safe
43	0.647463	**0.89509**	0.481124	0.452354	0.632964	**1**	0.434609	Safe
45	0.575723	**0.997888**	0.552283	0.448211	**0.885871**	**0.88315**	0.464105	Safe
47	0.67906	**0.861604**	0.504069	0.421309	0.595192	**0.880479**	0.680196	Safe
48	0.620535	**0.803047**	0.50859	0.490701	0.620499	**0.991876**	**0.751307**	Safe
51	**0.962983**	0.498448	0.44178	0.51731	**0.732342**	**0.706917**	0.460661	Safe
53	0.415607	**0.757046**	0.548888	0.59855	**0.819974**	0.478233	0.563024	Safe
1	0.472116	**0.776312**	0.473665	0.514885	**0.863567**	**1**	0.339338	Fail
2	0.571929	**0.905582**	0.561643	0.407179	0.506223	**0.999999**	**0.760517**	Fail
5	0.620756	**0.90146**	0.522011	0.490006	0.597854	0.302455	0.526003	Fail
8	0.645316	**0.896034**	0.530497	0.578242	**0.826057**	**1**	0.686242	Fail
9	0.452906	**0.750839**	0.461541	0.464576	**0.811952**	**0.766377**	0.491597	Fail
10	0.47329	**1**	0.450095	0.494041	**0.881776**	**0.937406**	0.388072	Fail
13	**0.999201**	**0.987295**	0.393129	0.427494	**1**	**0.947878**	**1**	Fail
14	**0.968805**	**0.930005**	0.51199	**0.729519**	**1**	**0.945901**	0.362565	Fail
15	0.454398	**0.80123**	0.456538	0.423823	0.617745	**0.844857**	0.609217	Fail
18	0.539568	**0.877763**	0.491745	0.515444	**1**	0.615957	0.408052	Fail
20	0.541085	**0.707088**	0.511315	0.590092	0.610183	**1**	0.532445	Fail
25	**0.724446**	**0.892652**	**0.807135**	0.608277	**0.743954**	**1**	0.188348	Fail
29	0.570253	**0.897347**	0.555421	0.310313	**0.703507**	**0.881806**	0.513902	Fail
31	0.524109	**0.806369**	0.499955	0.485703	0.654309	**0.721975**	0.49799	Fail
32	0.583503	**0.808567**	0.37559	0.454574	**0.993863**	**0.999917**	0.453612	Fail
40	0.578989	**1**	0.524926	0.462031	**0.938659**	0.676882	**0.730515**	Fail
44	0.327444	**0.799642**	0.19414	0.633544	**1**	**1**	0.455613	Fail
46	**0.739623**	**0.887598**	0.477618	0.510282	**0.999993**	**0.940717**	0.425062	Fail
49	**0.812288**	**0.857094**	0.42649	0.601534	**1**	**0.99999**	0.473768	Fail
50	0.437904	**1**	0.483438	0.576736	**0.802902**	**0.863724**	0.583212	Fail
52	0.317011	**0.939673**	0.371521	0.407749	**0.92657**	**0.989084**	0.376866	Fail

* The highlighted one implies that IFs are greater than 0.7.

Table 4: mean timing of the first unit of component j available in cluster i.

K-number	Aux FP	SI	SCP	SIT	ADV	MSSV	POSRV	Core Fail
3	7,885.5	**0.0**	5,662.7	3,907.2	**88.0**	**834.4**	1,680.8	Safe
4	7,192.2	**2,704.7**	5,005.3	2,667.3	**0.0**	**2,360.1**	1,910.0	Safe
6	6,236.7	**1,706.8**	6,537.2	3,343.3	4,784.4	2,856.0	1,738.7	Safe
7	**2,871.9**	**2,184.1**	7,212.4	3,675.7	**1,008.6**	2,379.8	1,432.7	Safe
11	7,340.1	**807.0**	6,217.9	4,942.4	**5,334.6**	**593.1**	1,601.4	Safe
12	**4,536.1**	**1,863.1**	4,867.7	3,151.1	**414.7**	2,685.5	1,476.0	Safe
16	3,839.6	**1,227.8**	5,080.6	3,746.2	**476.4**	1,971.0	1,429.4	Safe
17	**7,936.1**	**417.2**	5,078.8	3,696.1	**0.0**	**1,891.9**	1,187.0	Safe
19	**5,906.8**	**0.0**	6,973.8	4,650.0	**6,815.6**	**239.6**	3,037.6	Safe
21	**7,128.1**	**52.8**	6,315.1	3,786.2	**0.0**	2,962.1	1,696.6	Safe
22	6,401.6	**1,609.3**	5,177.5	3,696.8	**4,114.4**	**1,077.6**	1,625.5	Safe
23	6,899.3	**1,263.8**	**5,677.7**	4,327.4	6,758.1	**0.0**	**1,442.9**	Safe
24	3,881.9	**1,377.2**	4,949.5	2,718.4	**4,523.1**	2,252.9	1,752.3	Safe
26	6,667.5	**2,957.0**	5,036.1	3,730.2	**0.0**	2,249.6	1,721.4	Safe
27	3,670.6	**1,132.8**	5,472.4	3,235.1	**435.8**	2,828.4	1,412.4	Safe
28	**911.7**	2,830.8	4,957.0	3,337.6	**303.6**	1,823.2	1,544.5	Safe
30	6,423.8	**1,753.9**	5,663.0	3,457.2	**3,999.0**	**2,024.7**	1,294.6	Safe
33	4,661.1	**697.5**	5,101.9	3,801.4	**3,255.2**	**956.4**	2,045.3	Safe
34	7,183.6	**3,452.9**	6,149.6	3,600.6	1,370.2	2,281.3	1,932.1	Safe
35	7,189.4	**2,575.3**	3,873.9	3,910.0	**0.0**	2,162.9	1,552.3	Safe
36	**2,346.4**	2,590.2	4,253.8	3,051.0	**717.6**	2,358.9	2,192.2	Safe
37	4,232.1	**845.8**	4,760.6	2,570.5	**902.6**	1,560.9	1,568.4	Safe
38	6,371.7	**2,301.2**	5,783.0	3,658.7	**580.3**	**1,372.6**	1,997.3	Safe
39	5,628.9	**2,048.4**	5,291.7	3,164.2	6,214.6	**702.3**	1,277.5	Safe
41	**7,439.2**	**1,600.1**	5,705.4	4,049.2	**4,536.0**	**0.0**	1,234.5	Safe
42	6,626.5	**485.9**	6,718.2	2,947.0	**5,250.5**	**0.0**	1,366.5	Safe
43	6,432.7	**1,157.5**	6,308.2	3,886.7	5,675.1	**54.1**	2,168.3	Safe
45	4,473.4	**1,000.2**	5,880.3	3,478.5	**5,449.2**	**1,571.1**	1,815.2	Safe
47	5,549.5	**1,370.6**	5,180.7	3,421.5	4,506.1	**1,162.0**	1,097.6	Safe
48	5,994.1	**1,248.9**	5,205.8	3,371.9	5,705.3	**336.7**	**1,331.5**	Safe
51	**1,500.6**	2,731.0	5,473.8	3,825.7	**2,461.4**	**1,031.8**	1,562.5	Safe
53	5,244.1	**1,521.2**	5,061.7	2,975.3	**5,632.7**	1,433.4	1,194.5	Safe
1	6,197.8	**7,039.4**	5,304.3	3,122.6	**6,929.4**	**81.7**	1,563.1	Fail
2	6,079.4	**6,862.2**	5,286.2	4,266.9	3,851.0	**207.0**	**908.5**	Fail
5	6,314.1	**6,548.0**	6,024.1	3,354.3	1,601.8	2,647.3	1,613.5	Fail
8	7,212.0	**6,580.9**	5,062.9	3,484.5	**4,531.3**	**0.0**	1,112.4	Fail
9	5,577.5	**5,128.8**	5,393.3	3,507.5	**5,896.2**	**5,163.1**	1,694.4	Fail
10	4,913.9	**0.0**	5,665.9	3,144.1	**5,789.7**	**5,243.1**	2,360.8	Fail
13	**8,902.5**	**6,696.9**	5,280.3	3,911.3	**0.0**	**2,329.3**	**0.0**	Fail
14	**8,576.2**	**7,391.4**	7,006.5	**3,443.2**	**0.0**	**1,217.8**	1,401.0	Fail
15	6,284.2	**7,016.2**	4,313.6	4,404.5	4,281.6	**3,684.7**	1,141.5	Fail
18	6,641.5	**6,764.9**	5,315.0	3,168.7	**0.0**	3,726.4	1,632.8	Fail
20	6,578.4	**6,089.0**	5,339.7	3,664.4	4,584.3	**72.3**	1,815.3	Fail
25	**2,460.6**	**6,982.5**	**5,036.4**	3,303.8	**6,556.3**	**0.0**	2,311.5	Fail
29	6,847.9	**6,268.0**	6,224.3	4,902.5	**4,485.4**	**2,153.4**	1,404.6	Fail
31	5,980.8	**6,261.0**	5,655.6	3,913.3	5,196.6	**1,700.8**	1,713.4	Fail
32	6,849.7	**7,587.2**	4,360.4	3,159.9	**3,051.7**	**402.5**	1,965.5	Fail
40	6,466.3	**0.0**	5,274.1	4,335.5	**6,334.2**	5,160.1	**1,554.1**	Fail
44	7,467.5	**5,885.1**	4,286.8	3,239.5	**0.0**	**0.0**	5,513.4	Fail
46	**6,991.0**	**7,022.8**	5,300.0	3,302.8	**283.8**	**926.5**	2,306.8	Fail
49	**7,088.0**	**7,105.1**	5,691.5	3,190.5	**0.0**	**261.1**	2,214.6	Fail
50	5,669.2	**0.0**	5,774.6	3,061.5	**6,636.5**	**5,470.4**	1,423.0	Fail
52	5,377.0	**7,588.1**	4,247.6	5,474.4	**4,780.7**	**703.7**	1,647.4	Fail

References

[1] Mandelli, Diego, et al. "Scenario clustering and dynamic probabilistic risk assessment." Reliability Engineering & System Safety 115 (2013): 146-160.

[2] Labeau, P.E., Smidts, C., Swaminathan, S., 2000. Dynamic reliability: towards an integrated platform for probabilistic risk assessment. Reliability Engineering and Systems Safety 68, 219–254.

[3] Diego Mandelli, Alper Yilmaz, Tunc Aldemir et al, 2013. Scenario clustering and dynamic probabilistic risk assessment. RESS 115, 146-160.

[4] Zio, Enrico. "Reliability engineering: Old problems and new challenges." Reliability Engineering & System Safety 94.2 (2009): 125-141.

[5] Kopustinkas, V., Augutis, J., Rimkevicius, S., 2005. Dynamic reliability and risk assessment of the accident localization system of the Ignalina NPP RBMK-1500 reactor. Reliability Engineering and System Safety 87, 77–87.

[6] Mercurio, Davide, et al. "Identification and classification of dynamic event tree scenarios via possibilistic clustering: application to a steam generator tube rupture event." Accident Analysis & Prevention 41.6 (2009): 1180-1191.

[7] Shieh, J., & Keogh, E. iSAX: Indexing and Mining Terabyte Sized Time Series. ACM SIGKDD, pp. 623 – 631, 2008.

[8] Begum, Nurjahan, et al. "Accelerating dynamic time warping clustering with a novel admissible pruning strategy." Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.

[9] Cuturi, Marco. "Fast global alignment kernels." Proceedings of the 28th international conference on machine learning (ICML-11). 2011.

[10] Sardá-Espinosa, Alexis. "Comparing Time-Series Clustering Algorithms in R Using the dtwclust Package." (2017).

[11] Phillips, Jeff M., and Suresh Venkatasubramanian. "A gentle introduction to the kernel distance." arXiv preprint arXiv:1103.1625 (2011).

[12] Han, J., M. Kamber, and K. H. Tung. "Spatial Clustering Methods in Data Mining: A Survey. Harvey J. Miller and Jiawei Han (eds.), Geographic Data Mining and Knowledge Discovery." (2001).

[13] Park, Hae-Sang, and Chi-Hyuck Jun. "A simple and fast algorithm for K-medoids clustering." Expert systems with applications 36.2 (2009): 3336-3341.

[14] Al Abid, Faisal Bin. "A Novel Approach for PAM Clustering Method." International Journal of Computer Applications 86.17 (2014).

[15] Cuturi, Marco, et al. "A kernel for time series based on global alignments." Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on. Vol. 2. IEEE, 2007.

[16] Lim, Ho-Gon, Sang-Hoon Han, and Jae Jun Jeong. "MOSAIQUE–A network based software for probabilistic uncertainty analysis of computerized simulation models." Nuclear Engineering and Design 241.5 (2011): 1776-1784.

[17] RELAP5-3D Code Development Team, RELAP5-3D Code Manual; 2005

[18] Lee, Sang-Seob, Sung-Hwan Kim, and Kune-Yull Suh. "The design features of the advanced power reactor 1400." Nuclear Engineering and Technology 41.8 (2009): 995-1004.

[19] E. M. Mirkes, “K-means and K-medoids (Applet)”, University of Leicester, 2011.