Proceedings of the Korean Nuclear Society Autumn Meeting
Seoul, Korea, October 1995

Applications of a Methodology for the Analysis of Learning Trends in Nuclear Power Plants

Hang Youn Cho, Sung Nam Choi, Won Yong Yun
Korea Institute of Nuclear Safety

Abstract

A methodology is applied to identify the learning trend related to the safety and availability of U.S. commercial nuclear power plants. The application is intended to aid in reducing likelihood of human errors. To assure that the methodology can be easily adapted to various types of classification schemes of operation data, a data bank classified by the Transient Analysis Classification and Evaluation(TRACE) scheme is selected for the methodology.

The significance criteria for human-initiated events affecting the systems and for events caused by human deficiencies were used.

Clustering analysis was used to identify the learning trend in multi-dimensional histograms. A computer code is developed based on the K-Means algorithm and applied to find the learning period in which error rates are monotonously decreasing with plant age.

1. Introduction

The objective of this study is to apply a methodology for the identification of the learning trend related to the safety and availability of U.S. commercial nuclear power plants with emphasis on human factors contributions. The application is to assist in reducing likelihood of operator errors and to help reduce the frequency of significant human errors. Quantitative human error analysis is concerned with identifying the weaknesses and strengths of human response in performing operation and maintenance tasks and with determining relevant human reliability characteristic such as human error rates.

In the first stage using developed criteria by the Nuclear Safety Analysis Center(NSAC) and other work(WASH-1400), the significance criteria are developed and the following factors are considered: loss of safety function, abnormal or unexpected transient, loss of indication, overriding engineered safety features, operation degradation. A classification scheme was used to extract numerically coded data from the Licensee Event Reports(LERs). The Transient Analysis Classification and Evaluation(TRACE) scheme, developed by Sabri et al.(1), was designed to provide appropriate means of encoding and quantification of human errors.

The classified data are collated to identify possible types of problems and to find means for statistical analysis using tabulating, plotting, and histograms. In addition, the human error rates are estimated and tabulated. The learning period of the human in nuclear plants is found by checking the slopes of intensity error rate curves. After reviewing the classified data and the tabulated results, methodologies for statistical analysis are reviewed and a methodology such as an analysis of clustering is developed for trend analysis.

The K-Means algorithm was selected as the clustering methodology. Data clustering can be used to provide reduced data for the analysis of multiway contingency tables by grouping of the plants of similar age groups. A computer program is developed to accommodate a clustering algorithm developed by McQueen as cited in (2).

After the TMI accident, the need for analyzing the operating experience became more evident. The operation experience used in the present work for nuclear power plants has been extracted from the Licensee Event Reports(LERs) and monthly outage reports. Several attempts (1, 3-6) have been made to identify significant problem areas using LERs and most of them have based the significance on the basis of a high number of reported events. However, the number of events occurring is not a sufficient index in itself because it lacks a base for comparison. Husseiny et. al (3) suggested the analysis of sparse contingency tables to identify the categories of action that are highly associated with the particular components that are generically involved in the reactor systems during maintenance and testing (M&T) using the LERs. In a contingency table of two dimensions, the number of cells was high while the average number of observations per cell was small and many cells had zero entries, because of the large number of combinations for components and systems. To overcome this sparsity, it is necessary to cluster the data according to similarity of the data.

Efforts to collect and update meaningful data are of prime importance to extract statistical trends from the compiled data of occurrences in nuclear power plants. The LERs have been reviewed to assess the impact of human factors on nuclear power plant's performance and reliability (7-8). Operator errors reported for the 18-year period (1961-1978) were analyzed. It was concluded that all human errors need to be analyzed including M&T for a complete and meaningful evaluation of the impact of human errors on the plant's performance. It also stressed that the frequency count of errors alone cannot provide any meaningful conclusions.

Operator data from LERs for all U.S. commercial light water reactor for the period between 1972 and 1977 was used to compute error rates and then employed to examine effects due to age, power level, and reactor type on the operator hazard function. A Bayes methodology was proposed to apply to available scarce data to update, predict, and smooth failure rates (9).

Husseiny et al. (3) suggested that fitting a model essentially smooths in the information provided by the sparse data set across the entire contigency table yielding better estimates of the probabilities. Several authors (3,8-14) indicated that the total available human error data were somewhat scarce for human reliability analysis. Therefore, for the appropriate analysis it is better to use a clustering technique for the data reduction so that each cell in the cross-classified categorical data has more frequencies and the zero entries in the table also can be reduced.

2. Description of Cluster Technique for Trend Analysis

A. Introduction to Cluster Analysis

The objective of cluster analysis is to separate a set of objects into constituent groups so that the members of any one group differ from one another as little as possible, according to a chosen criterion. Clustering, called unsupervised pattern recognition and classification, is an important tool in exploratory data analysis. Hartigan (15) indicated that one meaning of clustering is the computation of multivariate histograms, which may be useful and revealing even if there are no "real" clusters.

The reasons for using cluster analysis for human error data for commercial nuclear power plants are as follows: to idetify significant problems on human performance for the facilities by checking out of ordered age groups, to provide reduced data for the analysis of multiway contingency tables by clustering of similar plant age groups, to investigate existence of distinct clusters of nuclear power plants. A clustering algorithm was recommended by Bayne et al. (16). Monte Carlo methods were used to estimate the probability of misclassification of 13 clustering algorithms for six types of parameterization of two bivariate normal populations. The K-Means partitioning method was found to be the best with the smallest probability of misclassification.

B. K-Means Algorithm for Clustering

K-Means algorithm minimizes a performance index which is defined as the sum of the squared distances from all points in the cluster domain to the cluster center (17). The input data is represented by arranging measurements in the form of a matrix: X=(Xik) where Xik denotes the value of the kth variable (k=1, ..., ℓ) of the ith object (i=1, ..., m). A partition of n groups ($0 < n \leq m$) of the set M={1, ..., m} of m objects is a set of index sets p1, p2, ...,

$P_n \subset M$ where each of the n groups is assigned to one and only one of the j groups. The minimum variance criterion is described by

$$e=\sum_{j=1}^{n} \sum_{i\in p_j} \sum_{k=1}^{\ell} (X_{ik} - \overline{X}_{jk})^2$$

$$\text{where } \overline{X}_{jk} = \frac{1}{|P_j|} \sum_{i\in p_j} X_{ik}.$$

Thus, the sum of the sum of the squared Euclidean distances of the cluster members from their centroids is minimized by using an algorithm to assign cases to clusters. The K-Means algorithm is to search for a partition with small e by moving cases from one cluster to another. The search ends when no such movement reduces e. Optimal solution is not guaranteed, but a near optimum is almost always found with the K-Means algorithm. The advantage of the K-Means algorithm is that, with a predefined data matrix $x(m, \ell)$ whose rows are processed in sequence, K-Means can be reorganized so that those rows can be stored sequentially on an external medium(tape or disk) and also processed sequentially at each pass. As a result there is no practical limitation on m and ℓ because of storage requirements.

3. Identification of the Learning Trend in Human Related Events in Nuclear Power Plants

The impact of human error on the safety of LWR power plants requires that specific types of human error problems be identified to mitigate recurrent problems. For the purpose of this study it is performed to identify transient periods in the learning trend for human related events using developed K-Means clustering program.

Preliminary trending analysis needs to be performed for the plant age, reactor mode, and initiation action opposed to other various variables such as: initiator, significance (Significant, Potentially Significant, and Insignificant) on human deficiency and event consequence, significant criteria, status of system involved, systems, and components. The coarse screening includes analyses of tables and polygons, and indicates needs for grouping and matching of plant experience into reasonably homogeneous groups to support meaningful analysis. As an effort clustering analysis using K-Means algorithm is used to identify cut points between two phases: earlier failure and stationary status. Table 3.1 indicates that the first three years after commercial operation in nuclear power plants can be considered as an earlier failure phase (generally high error rate) and stationary status (low error rate) follows for over three-year experience. The number of clusters is selected on the basis of sum of squared distances empirically. For BWRs, the data for the first three years are lack of information (e.g., no power plants operating on age groups $0 < x < 1$ and $2 \leq x < 3$ for 1978 and age group $1 \leq x < 2$ for 1977). However, the human error rates on BWRs is likely to behave those of PWRs based on the results from clustering analysis in Table 3.1.

4. Conclusions and Recommendations

Clustering technique is applied to identify the learning trend related to safety and availability of U.S. commercial nuclear power plants. The application is intended to aid in reducing likelihood of human errors. To assure that the application can be easily adapted to various types of classification schemes of operation data, a data bank classified by TRACE is selected for reduction of the data using clustering for the homogeneous grouping of the data.

The significance criteria for human-initiated events affecting the systems and for events caused by human deficiencies were applied. For the consequences of human initiated events, an event was considered significant (potentially significant) if that event led (could have led) to one or more of the following consequences: loss of safety function, abnormal or unexpected transient, loss of indication, or operation degradation. For the nature of human defieciency, a significant event is considered as one which has led to any of the above consequences or which was caused by symptomatic problems or involved overriding engineered safety features. These criteria can provide preliminary flagging for significance of events as they were being reviewed.

Error rates (error/plant-year operation) for different initiator (operator, maintenance, and technical groups) for PWRs and BWRs were calculated and tabulated according to plant age.

The first three years of reactor operation for both reactor types were identified as the learning period with higher rates than those which had approximately low constant rates in the range of 0.3 to 0.5 error/plant-year after that period. Rates of significant, potentially significant events showed the same trend for event consequences and human deficiences. It is concluded that clustering analysis was useful to verify the learning trend in multidimensional histograms and the first three years of the commercial operation of reactors are identified as the learning period.

It is also recommended to apply this clustering technique for Korean environment to identify the learning trend.

5. Literature Cited

1. Sabri, Z. A. and Others. September 1980. LER Review and Analysis with TRACE Summaries for Human-Related Events. SAI-106-81-AM. Science Applications, Inc., Ames, Iowa.
2. Tou, J. T. and R. C. Gonzalez. 1975. Pattern Recognition Principles. John Wiley, New York.
3. Husseiny, A. A., Z. A. Sabri, and J. J. Turnage. February 1980. The Identification of the Impacts of Maintenance and Testing upon the Safety of LWR Power Plants (Final Report Part 1) SAI-100-81-AM. Science Applications, Inc., Ames, Iowa.
4. Nuclear Safety Information Center Staff. 1973-1980. Recent Occurrences at Nuclear Reactors and Their Causes. Nuclear Safety, Nov. - Doc. 1973 through Jan. -Feb. 1980. (Various Issues)
5. Sabri, Z. A. 1979. Characterization of Operator Errors from Actual Plant Operation Experience. Trans. Am. Nuc. Soc. 33:562-564.
6. Oh, C. B. and M. C. Cullingford. 1979. An Overview of the Probabilistic Analysis Staff Programs on Human Factors. Trans. Am. Nuc. Soc. 33:556-557.
7. Cho, H. Y. 1979. Human Error Data Retrieval for the U.S. Commercial Nuclear Power Plant. Unpublished M. S. Thesis, Library, Iowa State University, Ames, Iowa.
8. Sabri, Z. A. 1981. Evaluation of Gross Operator Error Rates Based on Past Experience in Commerical Nuclear Power Plants (Final Report). NUREG/CR-2143. Engineering Research Institute, Iowa State University, Ames, Iowa.
9. Sabri, Z. A. and Others. 1980. Bayesian Inference and Updating of Reliability Data. CONF-800403. Proceedings of the ANS/ENS Topical Meeting on Reactor Safety, Knoxville, Tennessee 1:478-484.
10. Cho, H. Y. 1981. Development and Applications of a Methodology for the Analysis of Significant Human Related Event Trends in Nuclear Power Plants. Unpublished Ph.D. Thesis, Library, Iowa State University, Ames, Iowa.
11. U. S. Nuclear Regulatory Commission. 1975. Reactor Safety Study-- An Assessment of Accident Risks in U. S. Commercial Nuclear Power Plants. WASH-1400. U. S. Government Printing Office, Washington, D. C.
12. Cho, H. Y., A. A. Husseiny, and Z. A. Sabri. November 1979. Depicting Operator Performance by Weibull Distribution. Trans. Am. Nuc. Soc. 33:576-579.
13. Azarm, A. A., A. A. Husseiny, and Z. A. Sabri. 1980. Estimation of Operator Hazard Function Based on Commercial Nuclear Power Plant Experience. CONF-800403.

Proceedings of the ANS/ENS Topical Meeting on Reactor Safety, Knoxville, Tennessee 2:1072-1078.

14. Wong, S. M. and Others. April 1980. Statistical Analysis of Reportable Events due to Maintenance and Testing Activities in Nuclear Power Facilities. CONF-800403. Proceedings of the ANS/ENS Topical Meeting on Reactor Safety, Knoxville, Tennessee 1:446-454.
15. Hartigan, J. A. 1975. Clustering Algorithms. John Wiley, New York.
16. Bayne, C. K., J. J. Beauchamp, C. L. Begovich and V. E. Kane. 1980. Monte Carlo Comparisons of Selected Clustering Procedures. Pattern Recognition 12:51-62.
17. Spath, H. 1980. Cluster Analysis Algorithms for Data Reduction and Classification of Objects. John Wiley, New York.

Table 3.1 Summary of Clustering on Human Error Rates in PWRs/BWRs

Description	Rx. Type	Variables[a]	Mean Error Rate (Error/Plant-Year)			
			1st	2nd	3rd	d[b]
Operator (Op)	PWR	Op	6.96	2.70	0.97	1.215
and		Main	1.58	1.27	0.98	
Maintenance (Main)	BWR	Op	2.75	0.75		6.297
		Main	2.55	1.45		
Reactor Mode	PWR	1	6.33	2.59	1.58	
		2	0.63	0.73	0.34	
		3	0.63	0.07	0.05	1.510
		4	0.95	0.15	0.07	
		5	0.63	0.94	0.25	
		6	0.00	0.34	0.13	
		7[c]	0.32	0.20	0.10	
	BWR	1	4.80	1.44		
		2	0.40	0.27		
		3	0.20	0.05		5.307
		4	0.40	0.16		
		5	0.20	0.35		
		6	2.00	0.17		
		7	0.00	0.10		
Initiation Action	PWR	Perception (Pe)	0.11	0.17		
		Cognition (C)	0.47	0.12		
		Decision (D)	0.53	0.07		6.050
		Response (R)	3.16	1.33		
		Behavior (B)	2.02	0.70		
	BWR	Pe	0.80	0.22		
		C	0.60	0.39		4.619
		D	0.20	0.03		
		R	2.80	0.94		
		B	2.80	0.76		
	PWR	Sig (S)	1.35	0.33		
Event		Potential (P)	1.18	0.53		7.756
Consequence		Insig. (I)	3.98	1.67		
	BWR	S	0.85	0.42		
		P	0.95	0.49		3.420
		I	4.70	1.24		
	PWR	S	2.22	1.11	0.36	
Human Deficiency		P	2.85	1.10	0.78	3.462
		I	4.43	2.04	0.77	
	BWR	S	0.85	0.52		
		P	1.80	0.74		3.980
		I	3.85	0.89		

[a] Refers to classification scheme in Chapter 3 (10).
[b] Sum of sum of squared distance.
[c] Unknown reactor mode