Transactions of the Korean Nuclear Society Spring Meeting
Jeju, Korea, May 29-30, 2014

A Comparison of AOP Classification Based on Difficulty, Importance, and Frequency by Cluster Analysis and Standardized Mean

Sun Yeong Choi* and Wondea Jung

Korea Atomic Energy Research Institute, Integrated Safety Assessment Div., Daedeok-daero 989-111, Yuseong-Gu, Daejeon, Republic of Korea, 305-353
sychoi@kaeri.re.kr

1. Introduction

The number of abnormal operation procedures (AOPs) has been increased as operators establish AOPs additionally to reflect an abnormal plant situation which has first appeared in a nuclear power plant (NPP). In Korea, there are plants that have more than one-hundred kinds of AOPs. Therefore, operators have started to recognize the importance of classifying the AOPs [1]. They should pay attention to those AOPs required to take emergency measures against an abnormal status that has a more serious effect on plant safety and/or often occurs.

We suggested a measure of prioritizing AOPs for a training purpose based on difficulty (D), importance (I), and frequency (F) [2]. A DIF analysis based on how difficult the task is, how important it is, and how frequently they occur is a well-known method of assessing the performance, prioritizing training needs and planning [3]-[6]. We used an SDIF-mean (Standardized DIF-mean) to prioritize AOPs in the previous paper. For the SDIF-mean, we standardized the three kinds of data respectively. We called them the SD-value, SI-value, and SF-value. We then calculated their mean.

In this paper, we performed a cluster analysis to classify AOPs. A cluster analysis is one of the common techniques for a statistical multivariate analysis, which is used in many scientific fields. Clustering is the classification of similar objects into groups so that each group shares some common characteristics [7]. The results of this research will be utilized not only to understand the AOP characteristics at a job analysis level but also to develop an effective AOP training program.

The purpose of this paper is to perform a cluster analysis for an AOP classification and compare the results through a cluster analysis with that by a standardized mean based on difficulty, importance, and frequency.

2. Methods and Results of Cluster Analysis

2.1 Data collection for D, I, and F and Standardization

For a cluster analysis, we applied the D, I, and F data collected for the SDIF-mean [2]. To collect data for D and I, a survey targeting a twelve MCR operation crew of a reference plant was carried out. We drew up a questionnaire in which each AOP is scored on a semantic differential scale of 1 to 5 for D and I respectively. Semantic differential questions measure the respondents' attitude or thought towards the given objects. The meaning of each score for D and I is as follows:

- 1: very easy / much less important
- 2: somewhat easy / less important
- 3: neither easy nor hard / neither less important nor important
- 4: somewhat hard / somewhat important
- 5. very hard / very important

For the F-value, we applied the existing research results [8]. The F-value collected from the Korea Nuclear Information System (KONIS) has a 1 to 5 score, and each score has the following meanings:

- 1: the event occurred more than ten years
- 2: the event occurred within ten years
- 3: the event occurred within five years
- 4: the event occurred within two years
- 5: the event occurred within one year

We standardized the three kinds of data, respectively to rescale data since the F-value shows different value ranges for the average and standard deviation compared to the D-value and I-value.

2.2 Cluster Analysis

A cluster analysis is a statistical analysis to group a set of objects in such a way that objects in the same group called a cluster are more similar (in some sense or another) to each other than to those in other groups (clusters). In this paper, we applied a cluster analysis to classify AOPs. A statistical computing program, R, was utilized to conduct a cluster analysis using the complete linkage method using the 'Euclidean' distance as a distance matrix. R, statistical software commonly used in scientific fields, is available as a free download from the www.r-project.org [9].

The syntax used for a cluster analysis in this paper was as follows:

```
> dsdif <— dist (sdif, method = "euclidean")
> sdifx = hclust (dsdif, method = "complete")
> plot (sdifx)
```

For the syntax, "sdif" is a data set for standardized D, I, and F and "dsdif" and "sdifx" are arbitrary names for the process. The first line is to compute a distance matrix. We used the 'Euclidean' method for a distance matrix for AOPs. The second line is to apply hierarchical clustering. The 'hclust' function in R uses the 'complete linkage' method for hierarchical clustering by default. This clustering method defines the cluster distance between two clusters to be the maximum distance between their individual components. At each stage of the clustering process, the two nearest clusters are merged into a new cluster. The process is repeated until the whole data set is formed into one single cluster. The third line is for plotting a dendrogram. A dendrogram is a tree diagram frequently used to illustrate the arrangement of the clusters produced by hierarchical clustering.

2.2 Result of Cluster Analysis

Fig. 1 describes a dendrogram presenting the results of the cluster analysis of 101 kinds of AOPs of a reference plant with their SD-value, SI-value, and SF-value. From Fig. 1, three clusters were produced broadly. The first group (A) consists of 36 AOPs, the second group (B) 24 AOPs, and the third group (C) 41 AOPs. Table 1 is a summary table of groups (A), (B), and (C).

AOPs of group (A) show characteristics indicating that their SD-value, SI-value, and SF-value are generally lower than average. AOPs of group (B) have a higher SD-value, SI-value, and SF-value than average. In particular, group (B) has the highest SF-value than those of other groups. AOPs of group (C) show a different pattern. They have the highest SD-value and SI-value but the lowest SF-value compared to the others. Like the SF-value of group (B), the SD-value of group (C) is remarkably high.

In summary, group (A) consists of AOPs less important and less difficult to perform, and their related abnormal events occur rarely. It can be assumed that the need for training regarding abnormal situations related to the AOPs of group (A) is lower than others. Group (B) is composed of AOPs whose occurrence frequency is the highest and difficulty and importance are higher than average. Finally, AOPs in group (C) have the highest difficulty and importance, but the related abnormal events occur very rarely. From an educational point of view, AOPs in group (C) can be more important than those of group (B) since increasing the ability to react in the case of an accident is the purpose of training even though an accident very rarely happens.

3. Comparison of AOP Classification by Cluster Analysis with Standardized Mean

In the previous paper, we classified AOPs into four groups by the SDIF-mean for a training plan. Table 2 shows the results. Four types of training interval, that is, six months, one year, two years, and none, are candidates for the groups in Table 2.

Table2. Classification of AOPs for Training

Group	Criterion	%
A′	$\mu+1\sigma <$ SDIF-mean	16
B′	$\mu <$ SDIF-mean $\leq \mu+1\sigma$	42
C′	$\mu-1\sigma <$SDIF-mean $\leq \mu$	27
D′	SDIF-mean $\leq \mu-1\sigma$	16

The sixteen AOPs of group D′ having the lowest SDIF-mean are all included in group (A) from the cluster analysis in this paper. However, seven AOPs in group C′ are not involved in group (A) of Table 1. Five AOPs are categorized into group (C) while two AOPs are categorized into group (B). As mentioned above, since the AOPs of group (C) are more important than those of group (B), we investigated the five AOPs. Table 3 shows the AOPs of group C′ in Table 1 in order of SDIF-mean.

Table 3. AOPs of Group C′

AOP No.	D	I	F	SDIF-Mesn
3842C	4.12	3.92	1	-0.0111
3842D	4.12	3.89	1	-0.0358
3518B	3.54	3.92	3	-0.1104
3451H	3.65	3.80	3	-0.1160
3522A	3.77	3.92	2	-0.1266
3731B	3.46	3.95	3	-0.1724
3596B	3.74	4.15	1	-0.1828
3463B	3.24	3.85	4	-0.2481
3515A	3.49	4.08	2	-0.2783
3712C	3.48	3.83	3	-0.2820
3431K	3.64	3.90	2	-0.2949
3712A	3.73	4.05	1	-0.3044
3711D	3.62	3.90	2	-0.3086
3463A	3.31	3.46	5	-0.3123
3451G	3.27	3.76	4	-0.3146
3531B	3.49	3.72	3	-0.3820
3741A	3.38	3.81	3	-0.4028
3522B	3.43	3.69	3	-0.4704
3431J	3.30	4.09	2	-0.4725
3721	4.03	3.56	1	-0.4799
3451F	3.16	3.63	4	-0.5661
3741E	3.62	3.90	1	-0.5715
3461	3.43	3.56	3	-0.6003
3691C	3.39	3.34	4	-0.6029
3841C	3.70	3.79	1	-0.6048
3731A	3.41	3.80	2	-0.6435
3841D	3.70	3.72	1	-0.6720

From Table 3, the marked five AOPs are categorized into group (C) by a cluster analysis. This means that they should be considered more carefully than other AOPs of group C′ to set a training program, since their D and I value are relatively high although their SDIF-means are lower than average.

3. Conclusions

In this paper, we categorized AOPs into three groups by a cluster analysis based on D, I, and F. Clustering is the classification of similar objects into groups so that each group shares some common characteristics. In addition, we compared the result by the cluster analysis in this paper with the classification result by the SDIF-mean in the previous paper. From the comparison, we found that a reevaluation can be required to assign a training interval for the AOPs of group C′ in the previous paper those have lower SDIF-mean. The reason for this is that some of the AOPs of group C′ have quite high D and I values while they have the lowest frequencies. From an educational point of view, AOPs in group (C) which have the highest difficulty and importance, but their related abnormal events occur very rarely are important, since increasing the ability to react in the case of an accident occurrence is the purpose of training even though accidents very rarely occur.

Acknowledgement

This work was supported by Nuclear Research & Development Program of the National Research Foundation of Korea (NRF) grant funded by the Korean government, Ministry of Science, Ict & future Planning (MSIP). (Grant Code: 2014M2A8A4025991)

REFERENCES

[1] Korea Hydro & Nuclear Co., "Operation procedure for Shinkori #2 plant," Rev.0. 기행-2001-09, 2001.

[2] S.Y. Choi and W. Hung, A Study on Rating of AOP (Abnormal Operation Procedure) Based on Difficulty, Importance, and Frequency, Transaction of the KNS Autumn Meeting, October 2013.

[3] John A. Yoder "Training program handbook: A systematic approach to training," DOE-HDBK-1078-94, 1994.

[4] Rail Safety & Standards Board, "Good Practice in Training: A Guide to the Analysis, design, delivery and management of training," RS/220, Issue 2, June 2007.

[5] Dennis K. Neitzel, "How to Develop an Effective Training Program." IEEE Industry Applications Magazine, May/ June 2006.

[6] Y.J. Lee, K.S. Jo, N.Y. Lim, D. W. Kim, S. B. Kwon, and E. H. Lee, " Analysis of Importance, Difficulty, and Frequency of Nurses' Job in Outpatient Departments, " Journal of Korean Academy of Fundamentals of Nursing Vol. 16 No. 2, 232-241, May 2009.

[7] R.A. Johnson and D.W. Wichern, "Applied multivariate statistical analysis," Pearson, 2007.

[8] Korea Hydro & Nuclear Co., "Task-to-training matrix for YGN 2,3 plants," 2011-50011863-Dan-0782, 2011.

[9] W. N. Venables, D. M. Smith and the R Core Team, "An Introduction to R, Notes on R: A Programming Environment for Data Analysis and Graphics," Version 3.0.1 (2013-05-16), 2013.

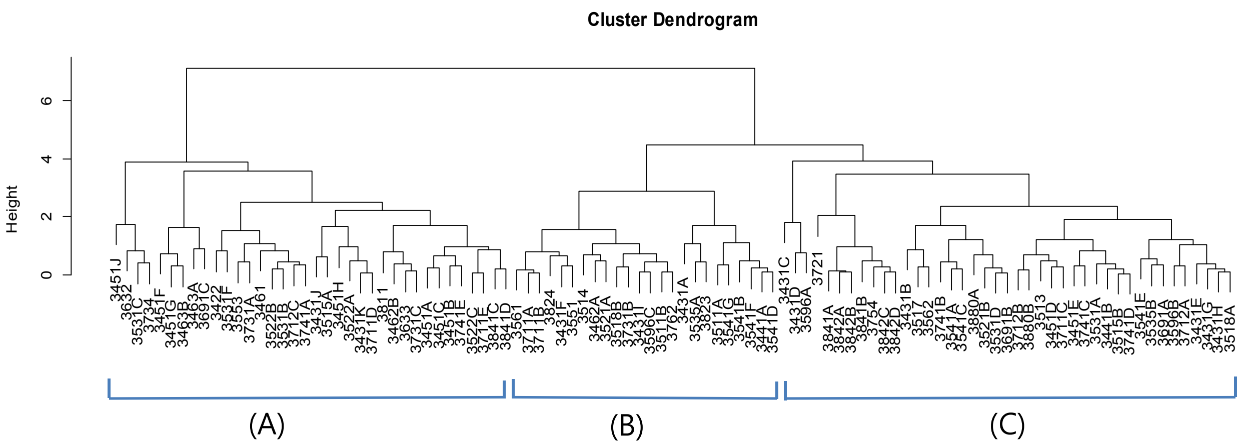

Figure 1. Dendrogram of AOPs of Reference Plant Based on D, I, and F

Table 1. Summary of Mean and Standard Deviation of SD, SI, and SF of Three Groups

Group	Number of AOPs	SD-mean (0.00±0.99)	SI-mean (0.00±0.99)	SF-mean (0.00±1.00)	SDIF-mean (0.00±0.74)
A	36	-1.00±0.63	-1.01±0.72	-0.30±0.92	-0.77±0.50
B	24	0.18±0.70	0.50±0.57	1.20±0.54	0.63±0.53
C	41	0.77±0.56	0.59±0.66	-0.44±0.66	0.31±0.35