

# THE SELF-ORGANIZING MAP ANALYSIS OF ACCIDENTS INVOLVING HAZARDOUS AND NOXIOUS SUBSTANCES

JEWAN RYU & HEEKYUNG PARK

Department of Civil and Environmental Engineering, Korea Institute of Science and Technology, South Korea

## ABSTRACT

Recently, the marine transportation of hazardous and noxious substances (HNS) has increased continuously. This has raised the risk of HNS spill accidents significantly and emphasized the issue of creating an emergency response system that is specialized to HNS. However, various types and properties of HNS have made it difficult to respond to the accidents. To develop the response system for HNS accidents, some qualitative considerations and statistical approaches are required. However, the previous studies have limitations due to the infrequency of accidents and lack of real-case data. The present study was performed with 103 cases of global HNS accidents from online database. In addition, 58 cases were nonlinearly projected using a self-organizing map, after some steps of data handling. Information about the date and year of accidents, location, cause, type of HNS, amount of spill, type of vessels, and the age of vessels was gathered and learned by SOM after the handling process. The learning parameters were optimized by trial-and-error, and the results of clustering showed 6 types of HNS accidents. The weight vectors of neurons in different clusters showed differences ways to classify HNS accidents. The type of vessel is a particular priority in the classification process.

*Keywords: hazardous and noxious substances (HNS), clustering, self-organizing map, feature of accident.*

## 1 INTRODUCTION

The issues of hazardous and noxious substances (HNS) have been highlighted since the definition of HNS has been defined as “Any substance other than oil, which, if introduced into the marine environment, is likely to create hazard to human health, to harm living resources and marine life, to damage amenities or to interfere with other legitimate uses of the sea” by International Maritime Organization 2000 [1]. There are variety of physical and chemical characteristics of HNS with about 6,500 different types of HNS. Internationally, the HNS volume of marine transportation has continuously increased and, in case of South Korea, has been reached about 19% of the total marine transportation [2].

The development of diagnosing system to response to HNS accident is needed to be and has been improved in recent years, but reflecting to response system that of oil spilling accident, the current response system to HNS accident does not show much improvements as expected as a considerable extent [2].

Previous researches were mainly focused on the qualitative analysis to build response system of HNS spill accidents [3]–[5]. Cunha et al. [3] analysed the 119 accidents around the world with the frequency of HNS according to classification standards such as Standard European Behaviour Classification (SEBC) and International Bulk Chemical (IBC) code. Cho et al. [6] performed Event Tree Analysis based on probability of the cause of accidents. Lee and Jung [7] categorized the severity and frequency of HNS accidents and suggested the future worst spill case scenario. Some studies suggested a response resources system based on the risk evaluation that is calculated from frequency of the inherent hazard of spilled HNS [8], [9].

Recent studies have been carried out to analyse the HNS accidents and have calculated the risk of the accidents. However, there is a major problem that there is lack of samples to



process qualitative or statistical analysis with guaranteed reliability of the model. In this study, self-organizing map (SOM), an unsupervised learning technique, was introduced to overcome the limitation and to provide clusters of the models by using the validity of the features of HNS accidents.

## 2 METHODOLOGY

### 2.1 Brief description of self-organizing map

The self-organizing map, as known as Kohonen's map, is an unsupervised machine learning technique that reflects the relations and similarity of the input data [10]. It projects the topology of the input data into 2-dimensional map, called array, of neurons. For initiation step, weight vectors,  $m_i$ , are randomly assigned to  $M$  neurons. In addition, input vectors find their winner neuron that whose distance is the closest from each input vector by eqn (1).

$$c = \arg \min_{1 \leq i \leq M} \|x(t) - m_i(t)\|, \quad (1)$$

where  $m_i$  is a weight vector of a neuron,  $x$  is an input vector,  $t$  is iteration number and  $M$  is number of neurons.

After finding the winner neuron, the weight vectors of winner and neighbours are updated by eqn (2) and iterate until the weight vector is converged. Neighbours are determined by the neighbourhood function,  $h_{ci}(t)$ . The  $\alpha(t)$  is monotonically decreased while the iteration is processed and this leads the weight vectors converge.

$$\begin{aligned} m_i(t+1) &= m_i(t) + h_{ci}(t)[x(t) - m_i(t)], \\ &\quad \forall_i \in \text{neighborhood}. \\ h_{ci}(t) &= \alpha(t) \exp\left(-\frac{\|c - i\|^2}{2\sigma^2(x)}\right). \\ 0 &< \alpha(t) < 1. \end{aligned} \quad (2)$$

### 2.2 Data collection and handling process

The global historic HNS accident data was collected at the online database of Centre of Documentation, Research and Experimentation on Accidental Water Pollution (CEDRE) [11]. 68 cases of accidents related to HNS was gathered with the attributes of; Name of vessel, date, location, cause, spilled HNS, amount of spill, type of vessel, and the birth year of vessel. The names of vessel clearly have no relation to the accident so they were just substituted by numbering index. The amounts of HNS spill were also ignored because of missing values and difficulty of unifying the measures, especially the case of containers. To handle the various spilled HNS, the classification of HNS convention was adopted [1], because its coverage was larger than other classification standards.

After all, continuous numerical values were scaled from 0 to 1 and discrete functional values were converted to binary vectors with appropriate dimension. Some new attributes were calculated from original attributes. The date of accident was split into month and year. Month was converted to four seasons and year was scaled between 0 to 1. Four seasons were adjusted once again as computational language. They were converted to 4-dimensional binary vectors that they can be differently recognized and calculated by self-organizing map. The location had three functional values; offshore, inshore, and port. Therefore, it was converted to 3-dimensional vectors; respectively [0 0 1], [0 1 0], and [1 0 0]. Furthermore, the cause of

Table 1: The structure of input data to self-organizing map.

Attributes	Values	Input vectors
Season	Spring, Summer, Fall, Winter	4-dimensional binary vectors
Year	Numerical values	Scaled in [0, 1]
Location	Offshore, Inshore, Port	3-dimensional binary vectors
Cause	Grounding, Structural damage, Collision, Weather condition, Load and unload, Fire and explosion	6-dimensional binary vectors
Type of HNS	i, ii, iii, iv, v, vi, vii	7-dimensional binary vectors
Type of vessel	Cargo vessel, Bulk carrier, Gas tanker, Tanker, Container ship, Passenger ship	6-dimensional binary vectors
Age of vessel	Numerical values	Scaled in [0, 1]

accidents, type of HNS, and type of vessel were also vectorised. The final input data after handling were 58 samples of 25 dimension (Table 1).

### 2.3 Optimization of learning parameters

The overall parameters of self-organizing map were optimized by trial-and-error. The size of map was  $8 \times 6$  that clearly showed the cluster boundaries. The total iteration number of updating weight vectors in learning process was set as 80 because there were any changes of the model with further iteration.

## 3 RESULT AND DISCUSSION

The two-visualizing matrixes, P- and U-matrices, were used to describe the result of SOM (Fig. 1). P-matrix showed the number of assigned input samples on each neuron. U-matrix showed the unified distance of each neuron, which means the Euclidean distance to weight vectors of neighbouring neurons. The neuron with the largest input sample had 6 samples in its neighbourhood. P-matrix was a measure of how the input data are spread on each neuron. It was considered that the inputs were assembled with some ravines. U- matrix showed the cluster structure more clearly. Imaginary neurons were inserted to emphasize the distance. Three ridges were observed and it was valid to consider at least 3 clusters. The ridges were hollow areas in input data space. Considering the change of weight vectors along with the ridge, it was concluded that the most effective classifying attributes were the types of HNS and cause of accidents. Samples assigned across ridges showed clear contrast in those two attributes. The horizontal ridge divided the map into two; upper and lower area. Upper area showed the accidents that caused by collision or structural damage. Lower area showed the accidents that caused by grounding or weather condition. The vertical ridge differentiated the types of HNS across itself. However, the effect caused by the age of vessel was not observed. Thus, it was difficult to consider the aging of vessel as an influence factor to HNS accidents.

Furthermore, the existence of two dense spot was observed in right-lower area. In addition, observed some of weakly dense spots were implicated the probability of more sub-clusters.

The unsupervised learning provided intuition for clustering, but this required more quantitative determination. Silhouette coefficient, which measured the validity of clustering, was used to compared the number of clusters. The model was starting with the 3 clusters, and the model ended with 6 clusters (Fig. 2). This result was also corresponded with the conventional rule of thumb, which is  $k \approx \sqrt{n/2}$ . Clusters were indexed in counter clockwise.

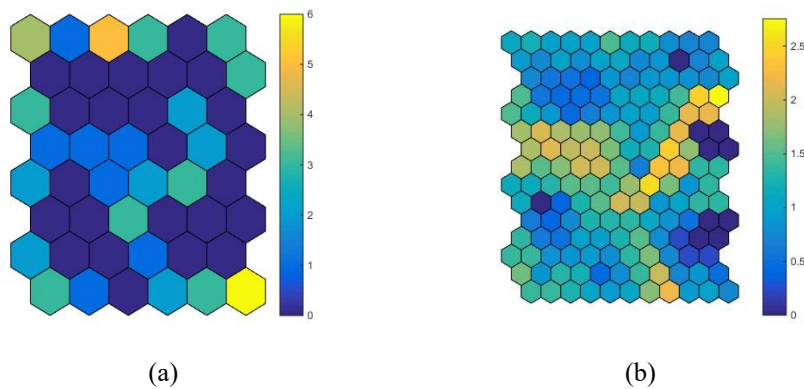


Figure 1: Visualization of the SOM result. (a) P-matrix; (b) U-matrix.

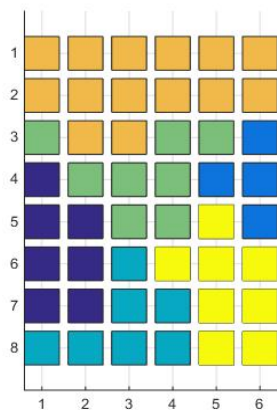


Figure 2: Cluster structure of SOM.

Table 2: Dominant attributes for each cluster.

	Location	Cause	Type of HNS	Type of vessel
1 <sup>st</sup> cluster	Offshore	Structural damage, Collision	ii, iii, vi	Chemical Tanker
2 <sup>nd</sup> cluster	Offshore	Structural damage, Collision	iv, v	N/A
3 <sup>rd</sup> cluster	Inshore	Weather condition, Grounding	iv	N/A
4 <sup>th</sup> cluster	N/A	Grounding	vii	Bulk carrier
5 <sup>th</sup> cluster	Offshore	N/A	vii	Cargo vessel
6 <sup>th</sup> cluster	N/A	N/A	v	Gas tanker

The dominant attributes along with the cluster in Fig. 2 were described in Table 2. The attributes more than half of all samples in one cluster was selected as dominant one. Because of the four attributes were dividing clusters as well, they could be a feature to distinguish accidents cases and should be considered as a preparing scenario. Some remarks from the clustering result were listed:

- The location of accidents had strong correlation with the cause of accident.
- The type of HNS had strong correlation with the type of vessel accident.
- The cause of accident had little correlation with the type of HNS.
- The location of accidents had little correlation with the type of vessel.
- The season, year, and age of vessel had no effect to accident occurrence.

#### 4 CONCLUSION

In recent years, the potential risks of HNS accidents have increased dramatically and consequently in the necessity for the appropriate response system has been emphasized. Thoroughly, there are many studies to diagnose the past accident cases, but they have limited results due to quantitative analysis and short of data. Thus, it was the main contribution to perform quantitative analysis and overcoming the statistical limitation in this study. The self-organizing map suggested 6 clusters to distinguish the accidents. The main influencing factors to make cluster was concluded location, cause, type of HNS, and type of vessel. The study results showed that the time of the accidents and the age of vessels did not distort the cluster structure. Therefore, it was concluded that these factors only implied small impacts on clustering. In conclusion, the present study found the dominant features that is effective in the creation of clusters of HNS accidents by using self-organizing map. Furthermore, the study results can give some insights to institute to develop response system, which can perform its task according to the dominant features from our study results. Because there is no applicable labelling to HNS accidents, the self-organizing map are simply projected the topology and relationship of the accidents. It is expected to predict future accident by labelling the accidents.

#### ACKNOWLEDGEMENTS

This research was supported by a grant [MPSS-CG-2015-01] through the Disaster and Safety Management Institute funded by Ministry of Public Safety and Security of Korean government. And also, this work is financially supported by Korea Ministry of Land, Infrastructure and Transport(MOLIT) as U-City Master and Doctor Course Grant Program.

#### REFERENCES

- [1] International Maritime Organization, *Protocol on Preparedness, Response and Co-operation to Pollution Incidents by Hazardous and Noxious Substances (OPRC-HNS protocol)*, 2000.
- [2] Korea Coast Guard, *Research on advanced HNS response system*, 2011.
- [3] Cunha, I., Moreira, S. & Santos, M.M., Review on hazardous and noxious substances (HNS) involved in marine spill incidents – An online database. *Journal of hazardous materials*, **285**, pp. 509–516, 2015.
- [4] Neuparth, T., et al. Hazardous and Noxious Substances (HNS) in the marine environment: Prioritizing HNS that pose major risk in a European context. *Marine pollution bulletin*, **62.1**, pp. 21–28, 2011.



- [5] Rocha, A., et al. Toxicity of seven priority hazardous and noxious substances (HNSs) to marine organisms: Current status, knowledge gaps and recommendations for future research. *Science of The Total Environment*, **542**, pp. 728–749, 2016.
- [6] Cho, S.J., Kim, D.J. & Choi, K.S., Hazardous and Noxious Substances (HNS) Risk Assessment and Accident Prevention Measures on Domestic Marine Transportation. *Journal of the Korean Society of Marine Environment & Safety*, **19.2**, pp. 145–154, 2013.
- [7] Lee, M. & Jung, J.Y., Risk assessment and national measure plan for oil and HNS spill accidents near Korea. *Marine pollution bulletin*, **73.1**, pp. 339–344, 2013.
- [8] Lee, E.B., Yun, J.H. & Chung, S.T., A study on the development of the response resource model of hazardous and noxious substances based on the risks of marine accidents in Korea. *Journal of Navigation and Port Research*, **36.10**, pp. 857–864, 2012.
- [9] Harold, P.D., et al., Development of a risk-based prioritisation methodology to inform public health emergency planning and preparedness in case of accidental spill at sea of hazardous and noxious substances (HNS). *Environment international*, **72**, pp. 157–163, 2014.
- [10] Kohonen, T., *MATLAB Implementations and Applications of the Self-Organizing Map*, Unigrafia Oy, Helsinki, 2014.
- [11] Database of spill incidents and threats in waters around the world, Centre of Documentation, Research and Experimentation on Accidental Water Pollution, Online. <http://wwz.cedre.fr/en/Our-resources/Spills>.

