MARY KAY O'CONNOR PROCESS SAFETY CENTER
TEXAS A&M ENGINEERING EXPERIMENT STATION

20th Annual International Symposium
October 24-26, 2017 • College Station, Texas

Prediction on Emergency Evacuation Orders Using Naïve Bayes Classification and Deep Learning

Chuntak Phark, Seungho Jung†
Department of Environmental and Safety Engineering, Ajou University, Suwon 443-749, Korea

Abstract
Emergency response to chemical accidents is proceeded in order of prevention, mitigation, preparedness, response and recovery. One of the methods of response is emergency evacuation orders. In order to minimize the loss of life, it is important to issue prompt and precise evacuation orders when chemical accidents such as toxic gas emissions occur near populated areas.
This paper presents a method and results for predicting emergency evacuation orders using naïve bayes classification, one of the statistical analysis methods, and Deep-learning, one of the artificial neural network analysis methods. The study was conducted using 61,563 useful data extracted from 115,569 accidents that occurred between 1996 and 2014 in ATSDR's National Toxic Substance Incidents Program(NTSIP) dataset. Rapidminer 7.5, a big data analysis program, was employed for big-data analysis. Through the analysis, it was predicted whether emergency evacuation orders were issued or not with high accuracy.
This study demonstrates that the technique can be used to identify the factors which affect the actual evacuation orders in the past and eventually provide a systematical decision-making process for rapid and accurate orders in the future accidents. In addition, as a result of the analysis, the accuracy of the method using Deep-learning has been proven higher than that of using Naïve bayes classification.

Key words: Emergency evacuation order, naïve bayes classification, deep learning, big-data analysis, NTSIP dataset

1. Introduction

The chemical industry is the third largest field of manufacturing of South Korea and affects a variety of industries, including automobiles, textiles, electronics, and construction. The scale of this field has been gradually increasing, and accordingly, the number of business places where chemicals are handled has been also increasing. This situation leads to increases in the total amount and kinds of chemical substances used in the chemical industry or in the world every year. The increase in the number of business places where chemical substances led to the increase in the number of chemical accidents occurred per year, the increase in the amount of chemical substances used led to the increase in the potential risks of chemical accidents, and the increase in the kinds of chemical substances led to the increase in the diversity or complexity of accidents. This trend is expected to continue hereafter too.

The ways to reduce the damage due to chemical accidents, which occur gradually more frequently, seriously, diversely, and complexly, comprise prevention, preparedness, responses, and recovery. Among them, prevention and preparation are effective before a chemical accident occurs, and responses and recovery are used after the occurrence of a chemical accident. Emergency evacuation, which is an act of quickly getting out of the area where the accident occurred, corresponds to a response, and the successful issuance of an emergency evacuation order can contribute to significantly reducing the damage due to the accident.

However, the issuance of an emergency evacuation order poses one problem, that is, the impact of the emergency evacuation order on the community and the neighboring citizens. Suppose you were sleeping at dawn and an emergency evacuation order has been issued. If the aftermath of the chemical accident does not get out of the business place where the accident occurred although you woke up and evacuated to a nearby shelter, you will be very angry. On the contrary, in the opposite situation, that is, if you did not evacuate but your house was within the scope of a chemical accident, the outcome should be much more than making you angry. It is clear that both the former and the latter cases have a bad influence on the overall social atmosphere. Therefore, the issuance of an emergency evacuation order undergoes very complicated decision making processes to avoid such a problem. According to Sorensen, et al. (2004), factors that determine the level of protection offered by protective actions are the characteristics of the released chemical, potential meteorological conditions at the site, the characteristics of structures surrounding the facility, the age of the building, air exchange in residential buildings, air exchange in office buildings, wind speed and temperature differentials, air exchange in vehicles, air replacement time, and time available before the public is exposed. The decision making related to an emergency evacuation order sometimes takes quite some time since an emergency evacuation order should be appropriately issued considering all these conditions when an accident has occurred. For this reason, we often encounter cases where the golden time for responses to accidents is missed thereby failing to reduce damage.

The problem of decision making in complex situations is not just a problem of the field of process safety. Attempts to analyze big data based on facts and solve such a problem using models that help decision making have been made frequently in other fields. In the field of medicine, Bekir KARLIK (2011) conducted a study to diagnose hepatitis disease using the Naïve Bayes classifier and neural networks. In the field of electronics, Selina S.Y.NG, et al. conducted a study to predict the residual useful life of lithium-ion batteries using the Naïve Bayes model. The attributes used in the study are the use conditions of the Li-ion battery and the ambient temperature. According to their study, the accuracy of the Naïve Bayes model is high because this model derives more stable results than the support vector machine (SVM), which is another big data analysis technique.

Big data analysis means processing high volume and complex data using PCs to find useful information. Here, the concept of machine learning, similar to humans' information learning, is applied, which refers to a series of processes through which information is gathered, refined, adapted, and generalized. Big data analysis and machine learning have become possible thanks to advances in computer technology. Advances in computer technology have made it possible to store rapidly produced high quality information and efficiently refine large amounts of information with complex interrelationships to extract and use useful data.

In particular, accident cases have been diversely analyzed recently through data mining techniques based on such data. Veltman (2008) studied the relationship between the attributes in HSEES and how they affected human safety in “Incident Data Analysis Using Data Mining Techniques.”. This study showed that the data mining analysis was able to address questions with regard to types of events that occur without having to read detail data attributions. Khan (2010) studied findings from chemical process incidents in the past and building up incident database and analysis in “Active and Knowledge – based Process Safety Incident.”. This paper suggests structuring unstructured data (such as sentences) through text mining. In addition, the database can be used in conjunction with management of change software, allowing users to more actively use the data.

Syukri (2012) studied incident patterns of the HSEES chemical incident database using data and text mining methodologies, and suggested correlations of each attribute. This paper has shown the possibility of analyzing the possible scenarios of incidents and the severity of incidents that may be caused when the required data is provided.

However, none of methodology has been developed in order to help decision making of the emergency evacuation orders though many researches have been conducted in the area of using accident database.

In this study, a model is proposed to predict emergency evacuation orders for future accidents using the accident database and machine learning technology.

2. Materials for machine learning

2.1. Database for study

The Hazardous Substances Emergency Events Surveillance (HSEES) system was operated by the Agency for Toxic Substances and Disease Registry (ATSDR) from January 1991 to September 2009 to describe the public health consequences of chemical releases, and to develop activities aimed at reducing the harm. An acute chemical release is an uncontrolled or illegal spill or release lasting <72 hours of an uncontrolled or illegal spill or release of any hazardous substance meeting specific predefined criteria. Releases of petroleum (e.g., crude oil or gasoline) were excluded from the HSEES system because the Comprehensive Environmental Response, Compensation, and Liability Act (Superfund legislation) excludes them from the Agency for Toxic Substances and Disease Registry authority(ATSDR, HSEES Database, 1996-2009).

Beginning in 2010, ATSDR replaced HSEES with the National Toxic Substance Incidents Program (NTSIP) to expand on the work of HSEES. NTSIP helps states in the US to collect surveillance data and to promote cost-effective, proactive measures such as converting to an inherently safer design, developing geographic mappings of chemically vulnerable areas, and adopting the principles of green chemistry (design of chemical products and processes that reduce or eliminate the generation of hazardous substances). Because the more populous states such as New York and Texas had the most incidents, areas with high population density should be carefully assessed for preparedness and prevention measures. NTSIP develops estimated incident numbers for states that do not collect data to help with state and national planning. NTSIP also collects more detailed data on chemical incidents with mass casualties(*ATSDR, NTSIP Database, 2010-2014*).

HSEES and NTSIP data can be used by public and environmental health and safety practitioners, worker representatives, emergency planners, preparedness coordinators, industries, emergency responders, and others to prepare for and prevent chemical incidents and injuries.

In the HSEES database, there are 102,037 incident cases that occurred between 1996 and 2009 for the states of Alabama, Colorado, Florida, Iowa, Louisiana, Michigan, Minnesota, Missouri, Mississippi, North Carolina, New Hampshire, New Jersey, New York, Oregon, Rhode Island, Texas, Utah, Washington, and Wisconsin. In the NTSIP database, there are 13,532 incident cases for three years between 2010 and 2014 for the states of Louisiana, North Carolina, New York, Oregon, Tennessee, Utah, and Wisconsin.

These databases express an accident with maximum 89 attributes (some changes every year) and the description of each attributes are given in table 1.

Table 1. HSEES/NTSIP databases description (*ATSDR, NTSIP Database, 2014).*

Attribute Class	Description
STAE	State where event occurred
EVNTCNTY	County where event occurred
FIPSCODE	Five digit FIPS county code
EVNTTYPE	Type of event
NOTF_TYP	Who notified the health department? – Primary source
NOTF_2_TYP	Who notified the health department? – Supplementary source
NOTF_THR	Primary source ID in other database
NOTF_2_THR	Supplementary source ID in other database
THRTACTU	Was the release actual or threatened
YEAR	Year when event occurred
SEASON	Season when event occurred
WEEKDAY	Portion of week when event occurred
TIME	Time range that event occurred
AREATYP1	Description one of type of area where event occurred
AREATYP2	Description two of type of area where event occurred
AREA_RES	Residential area within ¼ mile of event
PRIM_FACT	First contributing factor
SEC_FACT	Secondary contributing factor
PRIM_SPECIFY	Primary factor specify
SEC_SPEFICY	Secondary factor specify
FIXTYPE1, 2	Fixed facility type one, two
TRNTYPE1, 2	Transportation type one, two
NAICS	2-3 digit NAICS code for event location
NAICS_DESC	NAICS description assigned to the NAICS 2-3 digit code
LIVEQTR	Number of people living within ¼ mile of event
EVAC_ORD	Evacuation ordered : TARGET VALUE of this study
EVAC_PPL	Total number of people evacuated as a result of the event
SHLT_ORD	In-place sheltering ordered
DCON_SCTOTR	Rang of number of people decontaminated at the scene
DCON_MFTOTR	Rang of number of people decontaminated at a medical facility
TOT_CHEM	Total number of chemicals spilled
SUB_CAT	Substance category
CHEM1~6	Chemical name #1 ~ #6
CHM_QCAT1 ~ 6	Category for the amount of Chemical #1 ~ #6
CHM_UNIT1 ~ 6	Unit of measure for the amount of Chemical #1 ~ #6
RELS1CHEM1 ~ 6	First type of release for Chemical #1 ~ #6
RELS2CHEM1 ~ 6	Second type of release for Chemical #1 ~ #6

TOT_VICT	Total number of victims of the event
TOT_FATAL	Total number of fatality in the event
AGE_CAT1	Number of victim under 18 years old
AGE_CAT2	Number of victim older than 18.
VICT_EMP	Number of employee victims
VICT_RESP	Number of responder victims
VICT_GP	Number of general public victims
VICT_STD	Number of student victims
INJ_TRA	Number of victims with trauma injuries
INJ_RESP	Number of victims with respiratory system irritation
INJ_EYE	Number of victims with eye irritation
INJ_GASTRO	Number of victims with gastrointestinal problems
INJ_HEAT	Number of victims with heat stress injuries
INJ_BURN	Number of victims with burn injuries
INJ_SKIN	Number of victims with skin irritation injuries
INJ_CNS	Number of victims with dizziness or other CNS symptoms
INJ_HACHE	Number of victims with headaches
INJ_HRT	Number of victims with heart problems
INJ_SOB	Number of victims with shortness of breath
SEV_HOSPA	Number of victims where injury severity required treatment at hospital and admittance
SEV_HOSPR	Number of victims where injury severity required treatment at hospital without being admitted or victim was transported to hospital for observation with no treatment
SEV_NHOSP	Number of victims where injury severity required treatment on the scene (first aid); or victim was seen by a private physician within 24 hrs; or injuries were experienced within 24 hrs of the event and reported by an official
VDCON_SN	Number of injured people decontaminated at the scene
VDCON_MF	Number of injured people decontaminated at a medical facility
VDCON_BOTH	Number of injured people decontaminated at both the scene and a medical facility

The HSEES / NTSIP database (1996 - 2014) collected by ATSDR in the United States contains 115,569 incident cases. The attributes classes were modified to facilitate data analysis.

(1) Added NFPA rating and deleted accident chemical substance names(CHEM1)

(2) Unified industrial codes(NAICS) and deleted industrial code descriptions(NAICS_DESC)

(3) Removed recently collected attributes classes

(4) Removed geographical information(EVNTCNTY, FIPSCODE, etc)

(5) Removed the years of occurrence of accidents

(6) Unknown information at the beginning of the accident: the total amount of accident substances(CHM_QCAT) or the total number of victims(TOT_VIC).

After the modification of the attributes classes, those incident cases that fall under the following were excluded from the analysis.

(1) Incident cases with no information on whether an emergency evacuation order was issued or not

(2) Incident cases with unknown NFPA rating

In the database ('modified HSEES / NTSIP database') with attributes class and incident cases modified as described above, 61,563 incident data remain. Machine learning was performed with this database. Descriptions of the 'modified HSEES / NTSIP database' used in the analysis are listed in Table 2.

Table 2. Descriptions of the 'modified HSEES / NTSIP database'

Attribute Class	Description
EVNTTYPE	Type of event
THRTACTU	Was the release actual or threatened
SEASON	Season when event occurred
WEEKDAY	Portion of week when event occurred
TIME	Time range that event occurred
AREATYP1	Description one of type of area where event occurred
AREATYP2	Description two of type of area where event occurred
AREA_RES	Residential area within ¼ mile of event
FIXTYPE1	Fixed facility type one
TRNTYPE1	Transportation type one
FIXTYPE2	Fixed facility type two
TRNTYPE2	Transportation type two
NAICS	2-3 digit NAICS code for event location
LIVEQTR	Number of people living within ¼ mile of event
EVAC_ORD	Evacuation ordered : TARGET VALUE of this study
SUB_CAT	Substance category
RELS1CHEM1	First type of release for Chemical #1
RELS2CHEM1	Second type of release for Chemical #1
H	Newly added. NFPA rating; Health hazard(0~4)
F	Newly added. NFPA rating; Fire hazard(0~4)
R	Newly added. NFPA rating; Instability hazard(0~4)

3. Algorithms for study and Evaluation methods

3.1. Deep learning over view

Deep learning, or machine learning, is one of the big data analysis techniques, which means a technology in which the program analyzes data based on a given algorithm, learns the data through the analysis, and makes judgments or predictions based on the learning. It is largely divided into supervised learning, unsupervised learning, reinforcement learning, and evolutionary learning depending on the purposes of use. Among them, supervised learning is a way to generalize the training set and target value by learning them so that correct answers can be inferred (Stephen Marsland, 2014.), and predicting emergency evacuation orders corresponds to supervised learning. Among supervised learning algorithms, the Naïve Bayes classifier, which is based on statistics, and the multilayer perceptron model, which is based on Neural Artificial Networks, were used to conduct this study.

3.2. Analysis algorithm: Naïve Bayes classifier

Naïve Bayes Classifier is a simple supervised learning method based on statistics (D. Lowd, P. Domingos, 2005). It is based on the Bayes rule, which is a method of finding the posterior probability using the prior probability and the probability of a single event that can be easily obtained by using the database. The Bayes rule has a disadvantage that its application becomes difficult when the number of factors that affect the situation to be predicted increases. Therefore, the posterior probability is obtained assuming that all factors affecting the situation are independent, and this method is called Naïve Bayes classifier. The accuracy of Naïve Bayes classifier is known to be quite high despite that it infers the posterior probability assuming situations that are not independent in fact as being independent from each other.

3.3. Analysis algorithm: Artificial neural networks

Artificial neural networks are models designed by simulating the neural networks of living things. Artificial neural networks derive their results through the interactions among parallel nodes consisting of the input layer that receives data, the summing junction and activation function that perform calculations, and the output layer that outputs results. The signals that came into the input layer are multiplied by the weights of individual neurons and summed up at the summing junction. The resultant value is entered into the activation function, where it is judged based on the threshold to derive output values. Equations (1) and (2) and Figure (1) are the expressions of the foregoing as formulas and a figure, respectively.

$$u_k = \sum_{j=1}^{p} w_{kj} x_j \tag{1}$$

$$y_k = \varphi(u_k - \theta_k) \tag{2}$$

Where, x_j = input signals; w_{kp} = synaptic weights of neuron; u_k = linear combiner output; θk = threshold; $\varphi(\cdot)$ = activation function, and y_k = output signal.

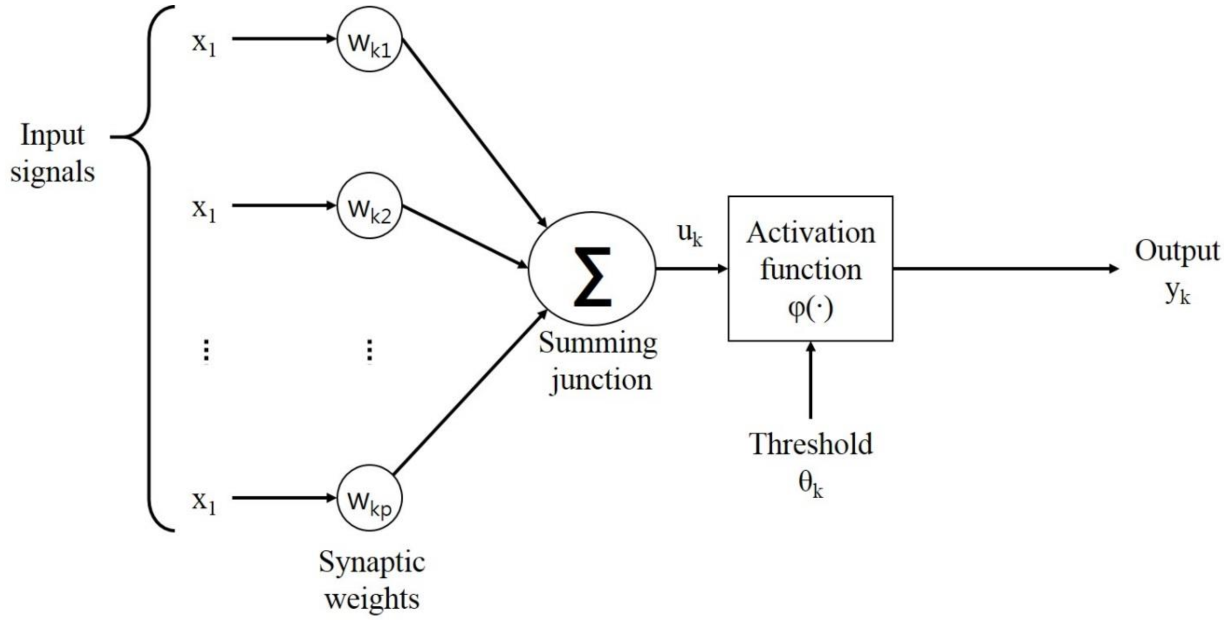

Figure 1. Perceptron model (Simon Haykin. 1994.)

Artificial neural networks enable learning even using incomplete data because the results are derived through interactions among multiple nodes. Therefore, artificial neural networks become to have the characteristics termed fault tolerance and adaptability. This is particularly in contrast to the Naïve Bayes classifier.

The artificial neural networks used in this study are multi-layer perceptron (MLP). The information input to the MLP flows from the input layer, goes through the hidden layer (s), and flows to the output layer. Figure 2 is a schematization of the foregoing.

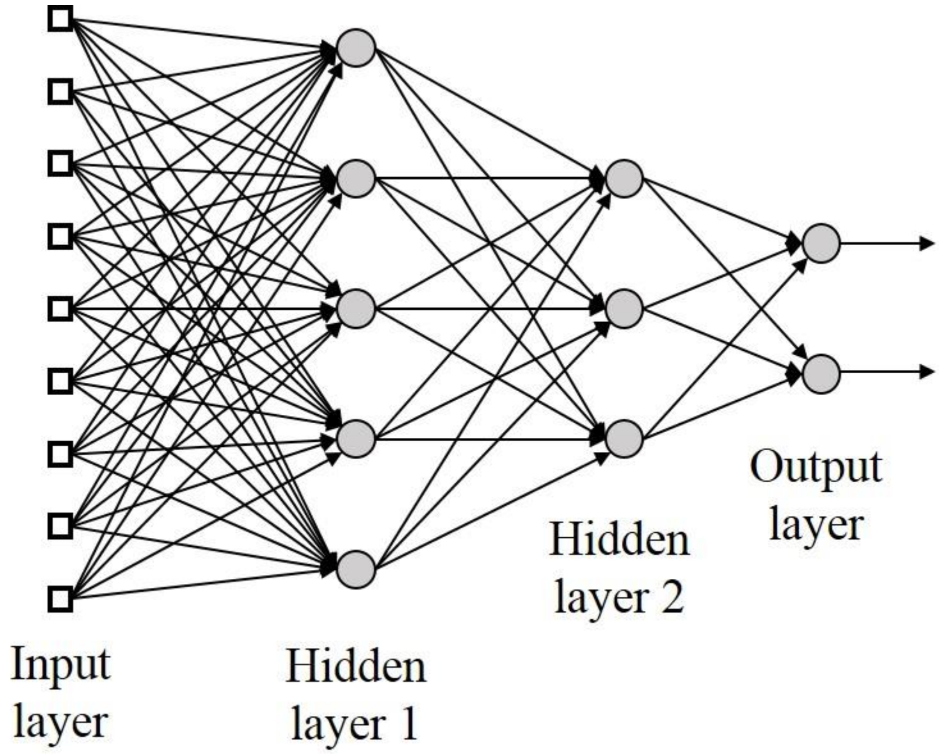

Figure 2. Multi-layer Perceptron artificial neural networks structure

The MLP used in this study is a back propagation algorithm. Back propagation is based on the error-correction learning rule, which is a learning method that reduces the error between the target value and the learning outcome. According to this method, in the back propagation, calculations are performed in the forward direction (input to output) using the initially set synaptic weights and the errors between the calculation results and the target value are calculated. Using these errors, the synaptic weights are modified in the reverse direction (output to input) (Simon Haykin, 1994). While repeating the foregoing, the learning is conducted until the synaptic weights that minimize errors are found by calculating the Mean Square Errors (MSE) of all data. The learning is finished when the point where the MSE is the smallest and the output value at this time becomes the learning result of back propagation.

3.4. Accuracy calculation: Overall percent agreement, Sensitivity, Specificity, ROC curve, AUC

The overall percent agreement, sensitivity and specificity, and AUC were used as evaluation indicators for prediction models. The overall percent agreement is the percentage of correct predictions. The sensitivity is the ability to correctly identify those with true results (true positive rate) while the specificity is the ability to correctly identify those with untrue results (true negative rate).

The receiver-operating characteristic curve (ROC) is a graph that expresses the relationship between the sensitivity and the specificity. The ROC curve is useful for visualizing machine learning models. The AUC is a value that means the area below the ROC. It is used as a measure of discrimination of prediction models. In general, tests can be classified based on AUC values into non-informative (AUC=0.5); less accurate ($0.5<AUC\leq 0.7$); moderately accurate

(0.7<AUC≤0.9); quite accurate (0.9<AUC<1); and perfect (AUC=1).(Hosmer, D. W. and S. Lemeshow, 2000)

4. Data analysis procedure

The analysis was conducted using Rapidminer 7.5, a big data analysis program. This program supports more than 500 operators to support various big data analysis tasks and provides diverse operators such as web mining, text mining, and time series analysis. In the present study, the Naïve Bayes classifier, which is one of the predictive operators, and H2O algorithm, which is one of the back propagation MLP algorithm, were used.
The 'modified HSEES/NTSIP database' was learned and tested using the Naïve Bayes classifier and artificial neural networks to derive the accuracy. Machine learning was conducted after dividing learning data: test data in a ratio of 6: 4. Whereas the Naïve Bayes classifier does not require the optimization of the analysis algorithm because it is based on statistical values, artificial neural networks require the optimization of the algorithm. In this study, the activation function and the size of hidden layers were optimized. In general, there is no model to determine size of hidden layers. Therefore, the rule-of-thumb 'to determine the size of hidden layers as 70~90% of the size of input layers'(Saurabh Karsoliya. Approximating Number of Hidden layer neurons in Multiple Hidden Layer BPNN Architecture. International Journal of Engineering Trends and Technology. Vol 3. Issue 6. pp.714 – 717. 2012.) was followed. After trying the four activation functions (tanh, rectifier linear, maxout, exponential rectifier linear) which are widely used in machine learning research, the activation function (rectifier linear) was finally selected and used since it had the highest AUC. Artificial neural networks were optimized using multi-fold cross-validation. Multi-fold cross-validation is a method to divide a data set into k data sets, conduct learning using k-1 data sets, conduct validation using the remaining one data set, and repeated the foregoing k times. In this study, the learning data was divided into 10 data sets for validation(k=10). Figure 3. Shows the overall study flow.

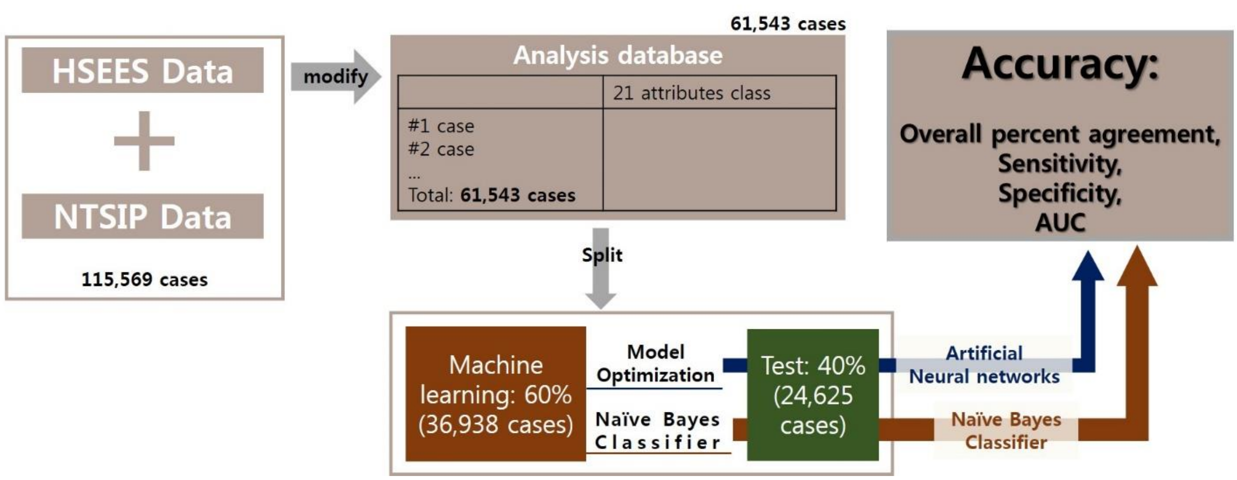

Figure 3. Methodology overview

5. Results and discussions

5.1. Artificial neural networks optimization

Since the database used in the analysis has 21 attribute classes, the size of the input layer becomes 21 and the size of hidden layer was determined in a range of 15 ~ 18. Since the AUC was the highest at 0.892 when the activation function was set as the rectifier (hidden layer size: 17 x 17), it was regarded as the optimum value. Table 3 shows the AUC values by activation function and hidden layers size.

Table 3. Model optimization results

Hidden layer size	15 x 15	16 x 16	17 x 17	18 x 18
Tanh	0.887	0.884	0.89	0.886
Rectifier	0.889	0.89	**0.892**	0.885
Maxout	0.891	0.885	0.885	0.891
Exponential Rectifier	0.882	0.881	0.88[A1]	0.878

5.2. Case study

This chapter demonstrates how to use artificial neural networks methodology to predict emergency evacuation orders for the following situations:
One Monday afternoon in the winter, the accident occurred in the metal product facility. The facility is located in undeveloped area where 3 peoples live within a quarter of a mile. The unit of the accident was the ancillary process equipment and accident material was ammonia. In the aftermath of the accident, the leaked ammonia has spread to the atmosphere.

1) Collect the information and make dataset according to the table 2. Table 4 shows the dataset of this situation.

Table 4. Dataset of case study

Situation	**Data transformation**
Facility	EVNTTYPE = Fixed facility
Spread into the atmosphere	THRTACTU = Actually released into the environment RELS1CHEM1 = Air emission
Winter	SEASON = Winter
Monday	WEEKDAY = Yes
Afternoon	TIME = Daytime
Facility is located in undeveloped area	AREATYP1 = Undeveloped
The ancillary process equipment	FIXTYPE1 = Ancillary process equipment
At metal product facility	NAICS = 311
3 peoples live within a quarter of a mile	AREA_RES = Yes, LIVEQTR = 3
Accident material was ammonia	H = 3 F = 1 R = 0

2) Set 'modified HSEES / NTSIP database' as the learning dataset. Deep learning is used as a learning algorithm and set the parameters to the optimal values found earlier (Activation function: Rectifier, Hidden layer size: 17 x 17, nfold = 10, epoch = 16). Set ‘Dataset of case study’ as the test dataset. An example of model setting using rapidminer 7.5 is shown in figure 4.

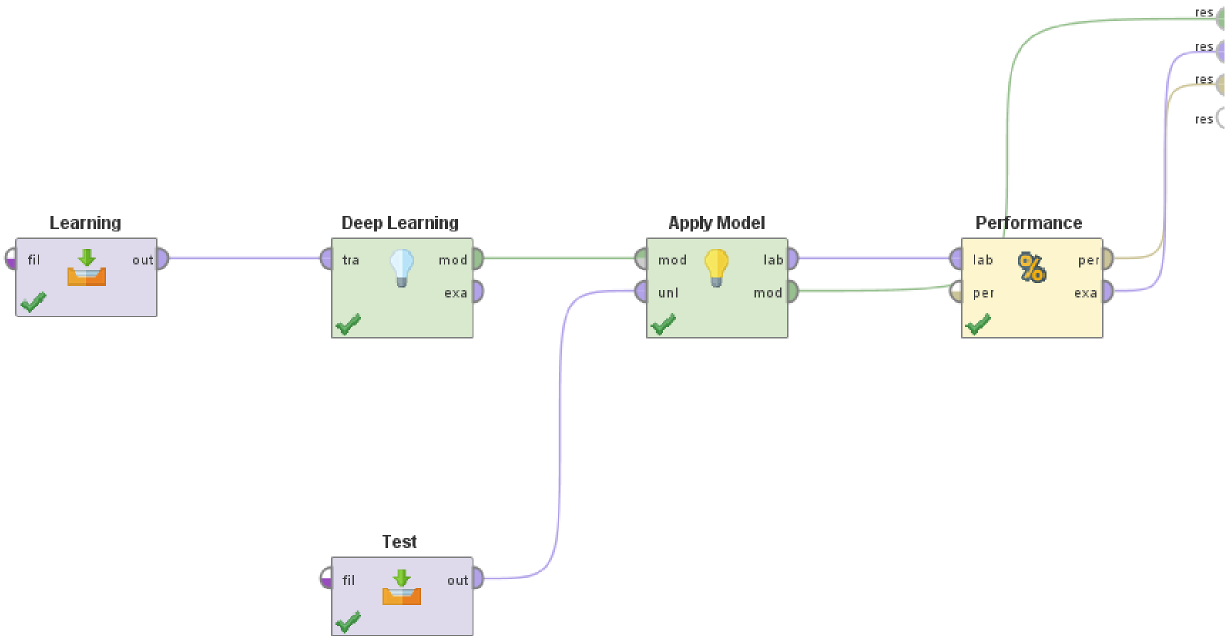

Figure 4. Example of model setting

3) Check the result.

The artificial neural network model predicted that an emergency evacuation order had to be issued, and in fact an emergency evacuation order was issued.

5.3. Results on prediction of emergency evacuation orders

The optimized value of artificial neural networks is the rectifier (hidden layer size: 17 x 17) and the accuracy levels of artificial neural networks and the Naïve Bayes classifier in the prediction of emergency evacuation orders were contained in Figure 5. In addition, the analysis using artificial neural networks took approximately 1 minute while the analysis using the Naïve Bayes classifier took approximately 10 sec.

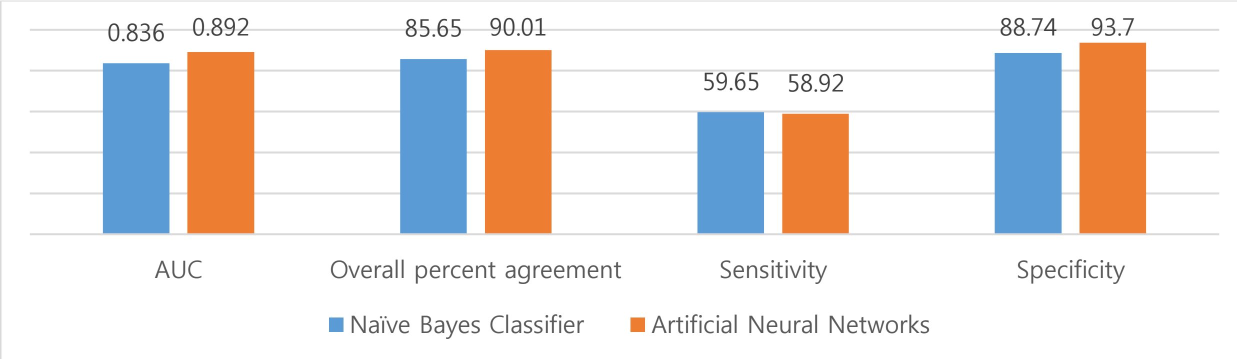

Figure 5. Accuracy of Machine learning using 'modified HSEES/NTSIP database'

The results of prediction of emergency evacuation orders using artificial neural networks were more accurate than those using Naïve Bayes classifier. Since the AUC value obtained using artificial neural networks is close to 0.9, it is believed that emergency evacuation orders can be predicted using the NFPA rating. Since the analysis took a very short time, it can be helpful for quick responses to accidents[A2].

5.4. Limitations and recommendations

Although the issuance of emergency evacuation orders were predictable at high levels of accuracy, the effectiveness of the emergency evacuation orders could not be judged. This means that the results of emergency evacuation order issued wrongfully might also have been learned. Therefore, it is believed that if the effectiveness of the emergency evacuation order is studied; models more helpful for decision making can be derived.

Emergency evacuation orders are an engineering issue but they are also a social issue. Different results of judgment may be produced depending on the safety culture of the workplace or the community where the accident occurred. Studies to derive quantitative or qualitative methods that can express safety culture are judged necessary.

Emergency response methods against chemical accidents include sheltering in addition to emergency evacuation orders. Sheltering orders are more efficient than evacuation orders when high concentrations of toxic gases pass in the form of puff. However, there was no consideration of sheltering in this study. Therefore, studies of sheltering orders are also judged necessary.

6. Conclusions

Emergency evacuation orders were predicted through machine learning using a modified HSEES / NTSIP database and the following major conclusions were derived.

(1) Emergency evacuation orders can be predicted using machine learning even when there are only the information obtained at the initial stage of the accident. The AUC values, which are an index of discriminatory power of the predicted values, obtained using both machine learning methods were considered to be 'quite accurate'.

(2) Quite high accuracy could be obtained even when the accident substances were replaced by NFPA for analysis. This method is considered to be applicable to chemical accidents caused by new substances without any accident history.

Reference

ATSDR, HSEES Database, 1996-2009, Accessed from http://www.atsdr.cdc.gov/hs/hsees/

ATSDR, NTSIP Database, 2010-2014, Accessed from https://www.atsdr.cdc.gov/ntsip/

Bekir KARLIK, Hepatitis Disease Diagnosis Using Backpropagation and the Naive Bayes Classifiers, Journal of Science and Technology, 1, 2011, 49-62.

D. Lowd, P. Domingos. "Naive bayes models for probability estimation", In 22th International Conference on Machine Learning, 2005

Hosmer, D. W. and S. Lemeshow. Applied logistic regression (2nd ed.) New York: Willy. 2000

Khan, S. S, Active and knowledge-based process safety incident retrieval system, 2010.

Selina S.Y.Ng, Yinjiao Xing, Kwok L. Tsui, A naive Bayes model for robust remaining useful life prediction of lithium-ion battery, Applied Energy, 118, 2014, 114-123.

Simon Haykin. Neural Networks. MacMillan. 1994.

Sorensen, J. H., Shumpert, B. L., & Vogt, B. M, Planning for protective action decision making: evacuate or shelter-in-place, Journal of Hazardous Materials, 109(1) 2004, 1-11.

Stephen marsland, Machine learning: An algorithmic perspective, second edition, Chapman & Hall/CRC, 2014.

Syukri, M, Analysis of the HSEES Chemical Incident Database Using Data and Text Mining Methodologies, 2012.

Veltman, L. M, Incident data analysis using data mining techniques, 2008.