Classification of Ammonia in water for Oil and Gas Industry using Case Based Reasoning (CBR)

Muhamad Faruqi Zahari[1], Muhammad Sharfi Najib[1], Kamarul Hawari Ghazali[1], Mohd Razali Daud[1], Fathimah Abdul Halim[1], Abd. Aziz Mohd Azoddein[2]
[1] Faculty of Electrical and Electronic Engineering, University Malaysia Pahang, Pekan, 26600, Pahang
[2] Faculty of Chemical and Natural Resources, University Malaysia Pahang, Gambang, 26300, Pahang
E-mail: sharfi@ump.edu.my[1]

***Abstract*-- Toxic gasses are exists in environment such as benzene, ammonia and others. Ammonia highly dissolves in water which is sources of human and other species. If the ammonia have high concentration, the effect of human health will be dangerous. Then, using proper monitoring and wastewater management the hazard can be prevented. This paper proposed the intelligence classification technique using an Electronic Nose (E-nose) measurement. The sensor array in the E - nose are used for the inputs of the Case Based Reasoning (CBR) for intelligent classification. The experimental result shows that the technique accomplished to classify with high accuracy which is 100% of accuracy.**

***Index Terms*--CBR, e-nose, ammonia, classification, water.**

I. Introduction

Nowadays, there are common toxic gasses exists in an environment such as benzene, ammonia, and others [1]. There are several industries that contribute to this toxic gasses in an environment such as medical , food packaging, agriculture, oil and gas and others [2]–[4]. The industries have own regulations to protect human or earth from hazards such as wastewater regulations, hazard managements and others [5]–[8]. Furthermore, water environmental issues have continued since long years ago, which source of the clean water being contaminated due to poor management and monitoring by various unpermitted activities [9]–[12].

One of the famous chemicals in water is Ammonia which caused lung edema, failure of the nervous system, acidosis and kidney damage [13], [14]. In addition, ammonia have highly dissolve in water, which have colorless fluid and pungent smell [15]–[17]. Thus, monitoring and management of wastewater that consist of toxic substance are compulsory to prevent environmental pollution [10], [17].

There are several methods for monitoring such as thermal image processing, Light Detection and Ranging (LIDAR), Electronic Nose (e-nose) and others [18]. E-nose using a concept sensor array for odor classification device [13]. The widely used sensor for chemical detection was metal-oxide gas sensors which used for varied application [19]. Metal-oxide sensor have advantages of low cost, short time response and high sensitivity [20], [21].

The E - nose was a device that functions as human olfactory system which detect the odor and the human brain will classify the odor based on the knowledge [22]. Thus, e-nose will take over the human nose to detect toxic gasses and using intelligence classification in order to define the odor of toxic gasses [23].

The e-nose detection data will analyze using normalization techniques which enables the comparison between ammonia concentration [24], [25]. In addition, the normalization technique was corrected systematic error which often present and can be removed by normalization method [26], [27]. Furthermore, normalization has several methods such as Range scale, relative scale, baseline subtraction, global method and local method [24], [28]. While, the data analyze using range scale which is [0, 1] as a fixed range for all samples [29], [30].

The simple boxplot displays the several categories of a discrete variable by separating the continuous variable of five statistic which is minimum, first quartile, median value, third quartile and maximum value [31]. The boxplot can differentiate the features of each sensor for each sample by statistical features as mentioned[32], [33].

Regression one of the methods to determine the validity of data point through a set of data point, regression analysis can be used in combination with statistical techniques [34], [35].

There are several classification methods such as an Artificial Neural Network (ANN), k-Nearest Neighbor (KNN), Case Based Reasoning (CBR) and others [36]. CBR was created as a four step process of human or computer reasoning which called CBR cycled [37]. CBR cycled contain of retrieve, reuse, revise and retain which the crucial step of CBR was the retrieval case in order to classify similar cases that stored in the library[36], [38]. CBR used the retrieval method using similarity percentage which based on nearest distance between new cases and stored cases was illustrated in Eq. 1 below [37], [39]–[41]:

$$\text{Similarity}(N_i, S_i) = \frac{\sum w_i \times f(N_i, S_i)}{\sum w_i} \quad (1)$$

where,

w_i is an index weight for matching feature
N_i is the new cases features
S_i is the source cases features
F is the similarity function between N and S cases.

II. Experimental Setup

A. Ammonia Preparation

The Ammonia will be prepared based on Oil and Gas industry environment. Ammonia in water sample is used for experimental process using 28% - 30% of ammonia in ammonia hydroxide which is a laboratory specification. Then, the ammonia in water was diluted into specific concentration for 10ml for each sample cell. The concentration was diluted using dilution equation. Each sample of concentration has five repeated measures data using an e-nose.

The concentration validated using chemical method which is called Total Nitrogen Test (TNT) to confirm the concentration of ammonia in water. Then, the confirmed concentration was inserted to sample cell and being moved to the e-nose detection for further process.

E- Nose were consist of array sensor that will measure the different odor of ammonia in water concentration. Data measured by e-nose for each sample has dimension of 4x5 which in total 20 measured data. Each sensor will measure the concentration in 1000 x 4 of data which in total 4000 raw data at a time. Then, the raw measured data were employed with preprocessing method in orde to reduce the error between the measured data.

Subsequently, the raw measured data were normalized in range [0 to 1] for reduction of error and noise of the measured data. Then, it was used to obtain the profile graph for visualizing the different of concentration. After that, boxplot technique was used for summarizing normalized data into five statistical features, which is minimum, maximum, first quartile, median and third quartile.

Then, the mean of the normalized data was used as features in CBR technique. By implementing a CBR algorithm, the classification of ammonia in water concentration can be accomplished.

B. Data Collection

The data collection will be recorded as in Table 1. The sample of ammonia water will contain different percentages of ammonia.

Table I
Sample data Collection

Sensors ***Data (n)***	***Sensor 1***	***Sensor 2***	***Sensor 3***	***Sensor 4***
$\boldsymbol{D_1}$	D_{11}	D_{12}	D_{13}	D_{14}
$\boldsymbol{D_2}$	D_{21}	D_{22}	D_{23}	D_{24}
$\boldsymbol{D_3}$	D_{31}	D_{32}	D_{33}	D_{34}
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
$\boldsymbol{D_N}$	D_{N1}	D_{N2}	D_{N3}	D_{N4}

III. Result and Discussion

Collected data had been normalized into 0 to 1 interval. Then, normalized data was plotted into graph as in Figure 1 and Figure 2 which can visualized the pattern between low concentration of ammonia in water and high concentration of ammonia in water.

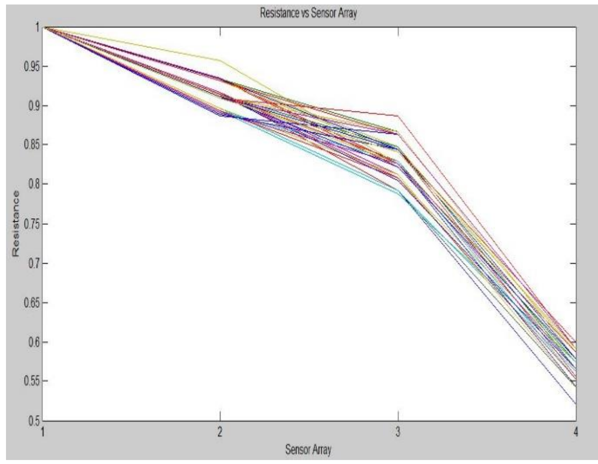

Fig. 1. Low ppm of ammonia pattern

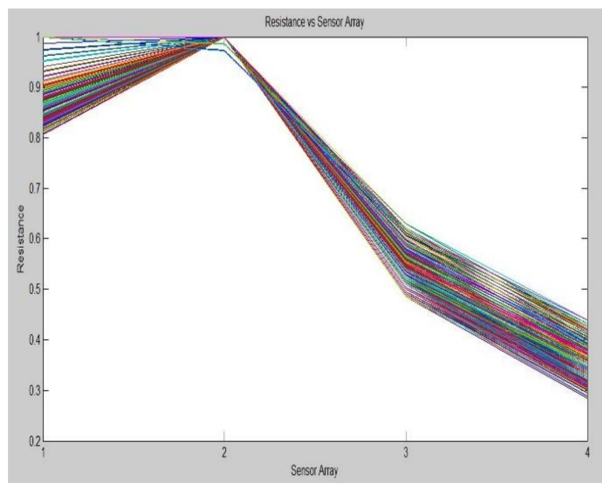

Fig. 2. High ppm of ammonia pattern

The raw data have been normalized to obtain a unique profile of odor concentration of ammonia in water as shown in Figure 1 and Figure 2 for high and low concentration of ammonia. The graph shows that there are differences between two classes based on pattern and values.

Based on pattern recognition, low ppm of ammonia have a different pattern compared to high ppm of ammonia as shown in Figure 1 and Figure 2. The graph features were improved by statistical analysis using boxplot as shown in Figure 3 and Figure 4.

Boxplot consists of minimum, maximum, first quartile, median and third quartile features point. Based on Figure 3 and Figure 4, the blue line of the boxplot indicates the first quartile and third quartile. The median was indicated as a red line in the boxplot graph. While, the minimum and maximum point was indicated as black line which known as a whisker.

Based on the Figure 3 and Figure 4, there are significantly different has been detected between the low ppm and higher ppm, which is the most responsive sensor for low ppm ammonia in water was the sensor number one while the most responsive sensor for high ppm ammonia in water was sensor number 2.

To validate the data, linear regression technique was used to differentiate the high ppm and low ppm as shown in Figure 5. The regression show of 100% of classification data between low and high ppm of ammonia.

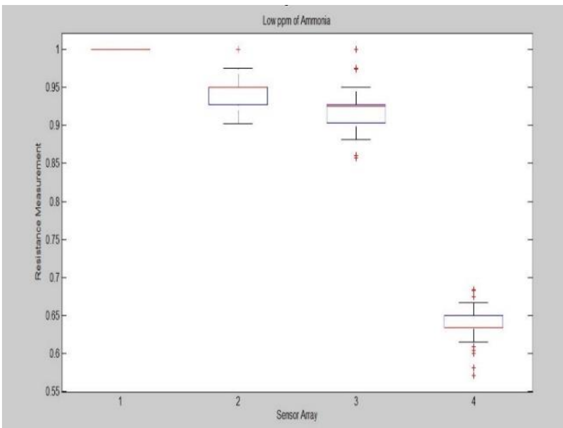

Fig. 3. Low concentration of ammonia boxplot

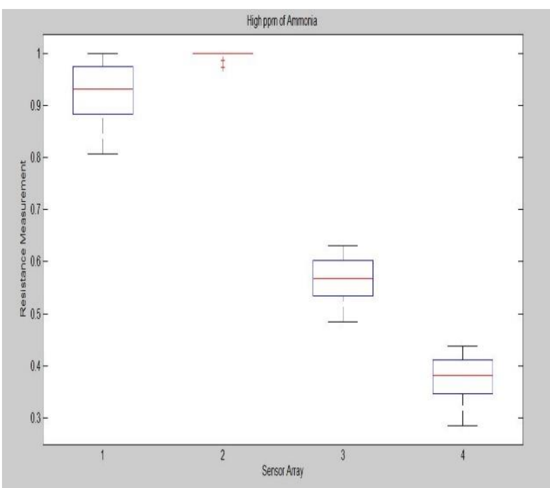

Fig. 4. High concentration of ammonia boxplot.

The regression and the CBR data based on the same features (mean) in order to validate the classification. The results have shown that the CBR has produced 100% of accuracy classification rate between low and high ppm of ammonia by using the Equation 1 as mentioned before.

Table 2 show the statistical analysis of CBR which based on nearest percentage similarity function of CBR classifier. Its show that the average of percentage shows the total accuracy to classify between low ppm and high ppm of ammonia in water which is approaches 100% of classifying.

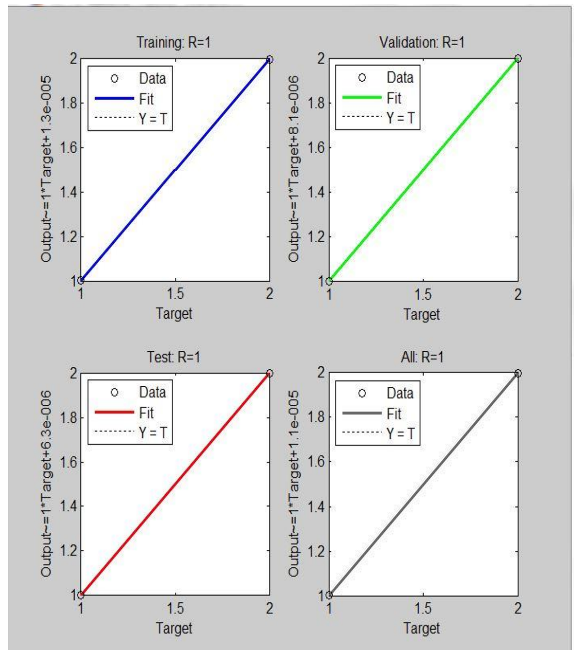

Fig. 5. Regression between Low and High ppm of ammonia

Table II

Summary Analysis on the Case Based Reasoning

Performance Evaluation	**Highest Percentage (K1)**	**Second Highest Percentage (K2)**	**Third Highest Percentage (K3)**	**Average**
Criteria	*Values*	*Values*	*Values*	*Values*
Total Cases	20	20	20	20
High PPM (P)	10	10	10	10
Low PPM (N)	10	10	10	10
True Positive (TP)	10	10	10	10
True Negative (TN)	10	10	10	10
False Positive (FP)	0	0	0	0
False Negative (FN)	0	0	0	0
Sensitivity = TP/(TP+FN)	1	1	1	1
Specificity = TN/(FP+TN)	1	1	1	1
Accuracy = (TP+TN)/ (P+N)	1	1	1	1

IV. Conclusion

This paper presents the reliability of CBR technique in classification of ammonia in water concentration using mean features. The analysis of the profile, boxplot and ANN regression shows the different between two concentration classes. Then, using intelligent classification which is CBR the result was successful approach 100% of classification between two classes.

V. Acknowledgment

The authors would like to thank all staff and student involved, Faculty of Electrical and Electronics Engineering, Faculty of Chemical and Natural Resources, UMP and excellent fund from UMP Research Grant coded RDU130606.

VI. References

[1] V. Ceorgieva, L. Spassov, and V. Georgiev, "An Attempt For NH3 Detection Based on Quartz Resonator With Thin SnO2 Film," Jt. Meet. EFTF- IEEE IFCS, pp. 1086–1088, 1999.

[2] A. Yoshio, "The Formation of Heavy Industry: One of the Processes of Industrialization in the Meiji Period," Dev. Econ., vol. 3, no. 4, pp. 450–470, Dec. 1965.

[3] R. O. A. Rahman, H. A. Ibrahium, and Y.-T. Hung, "Liquid Radioactive Wastes Treatment: A Review," Water, vol. 3. pp. 551–565, 2011.

[4] C. Klotz-Ingram and K. Day-Rubenstein, "The changing agricultural research environment: What does it mean for public-private innovation?," AgBioForum, vol. 2. pp. 24–32, 1999.

[5] U. Cohen, D. Allison, and J. Witte, "Critical Issues in Healthcare Environments," 2009.

[6] H. Roopadevi and R. K. Somashekar, "Assessment of the toxicity of waste water from a textile industry to Cyprinus carpio," J. Environ. Biol., vol. 33, pp. 167–171, 2012.

[7] G. Liu, Z. Yang, B. Chen, Y. Zhang, M. Su, and L. Zhang, "Emergy evaluation of the urban solid waste handling in liaoning province, China," Energies, vol. 6, pp. 5486–5506, 2013.

[8] C. H. Drew, D. A. Grace, S. M. Silbernagel, E. S. Hemmings, A. Smith, W. C. Griffith, T. K. Takaro, and E. M. Faustman, "Nuclear waste transportation: Case studies of identifying stakeholder risk information needs," Environmental Health Perspectives, vol. 111. pp. 263–272, 2003.

[9] C. Wochnowski, S. Metev, and G. Sepold, "UV–laser-assisted modification of the optical properties of polymethylmethacrylate," Appl. Surf. Sci., vol. 154–155, pp. 706–711, 2000.

[10] J. A. Awomeso, A. M. Taiwo, A. M. Gbadebo, and A. O. Arimoro, "Waste disposal and pollution management in urban areas: A workable remedy for the environment in developing countries," Am. J. Environ. Sci., vol. 6, pp. 26–32, 2010.

[11] N. P. Randall and K. L. James, "The effectiveness of integrated farm management, organic farming and agri-environment schemes for conserving biodiversity in temperate Europe - A systematic map," Environmental Evidence, vol. 1. p. 4, 2012.

[12] S. Broadmeadow and T. R. Nisbet, "The effects of riparian forest management on the freshwater environment: a literature review of best management practice," Hydrology and Earth System Sciences, vol. 8. pp. 286–305, 2004.

[13] Q. Ameer and S. B. Adeloju, "Polypyrrole-based electronic noses for environmental and industrial analysis," Sensors Actuators B Chem., vol. 106, no. 2, pp. 541–552, May 2005.

[14] I. Trebs, F. X. Meixner, J. Slanina, R. Otjes, P. Jongejan, and M. O. Andreae, "Real-time measurements of ammonia, acidic trace gases and water-soluble inorganic aerosol species at a rural site in the Amazon Basin," Atmospheric Chemistry and Physics Discussions, vol. 4. pp. 1203–1246, 2004.

[15] A. P. Seitsonen, A. M. Saitta, T. Wassmann, M. Lazzeri, and F. Mauri, "Ammonia," Water, vol. 5, p. 1966, 2010.

[16] Health Canada, "Ammonia in Drinking Water," Canada, 2013.

[17] K. Jahn, "Ecologically Sound Disposal Methods in Ammonia Technology," 2009.

[18] P. Murvay and I. Silea, "A survey on gas leak detection and localization techniques," J. Loss Prev. Process Ind., 2005.

[19] A. Vergara and E. Llobet, "Sensor Selection and Chemo-Sensory Optimization: Toward an Adaptable Chemo-Sensory System," Frontiers in Neuroengineering, vol. 4. 2012.

[20] J. R. Cordeiro, M. I. V. Martinez, R. W. C. Li, A. P. Cardoso, L. C. Nunes, F. J. Krug, T. R. L. C. Paixão, C. S. Nomura, and J. Gruber, "Identification of Four Wood Species by an Electronic Nose and by LIBS," Int. J. Electrochem., vol. 2012, pp. 1–5, 2012.

[21] K. Song, Q. Wang, Q. Liu, H. Zhang, and Y. Cheng, "A Wireless Electronic Nose System Using a Fe2O3 Gas Sensing Array and Least Squares Support Vector Regression," Sensors, vol. 11. pp. 485–505, 2011.

[22] H. V. Ramamoorthy, S. N. Mohamed, and D. S. Devi, "E-Nose and E-Tongue: Applications and Advances in Sensor Technology," Journal of NanoScience and NanoTechnology, vol. 2. pp. 370–376, 2014.

[23] D. Hodgins, "The development of an electronic ' nose ' for industrial and environmental applications," Sensors Actuators B Chem., vol. 26–27, pp. 255–258, 1995.

[24] B. Tudu, B. Kow, N. Bhattacharyya, and R. Bandyopadhyay, "Normalization techniques for gas sensor array as applied to classification for black tea," vol. 2, no. 1, pp. 176–189, 2009.

[25] F. S. Cabo, Z. Trajanoski, and K. Cho, "A Graphical Users Interface to Normalize Microarray Data," in Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), 2003, no. Dsc, pp. 1–16.

[26] L. Leydesdorff, "On the Normalization and Visualization of Author Co-Citation Data : Salton ' s Cosine versus the Jaccard Index," J. Am. Soc. Inf. Sci. Technol., vol. 131.

[27] D. Venet, "MatArray: a Matlab toolbox for microarray data," Bioinformatics, vol. 19, no. 5, pp. 659–660, Mar. 2003.

[28] M. Astrand, "Normalization and Differential Gene Expression Analysis of Microarray Data," Chalmers University of Technology and Goteborg University, 2008.

[29] J. C. Miecznikowski and K. F. Sellers, "Statistical Analysis of Chemical Sensor Data," in Advances in Chemical Sensors, 2002, pp. 327–347.

[30] J. Yosinski and R. Paffenroth, "Nonlinear Estimation for Arrays of Chemical Sensors," 2000.

[31] H. M. Rodriguez and A. J.O, "Simple Boxplot," in Building SPSS Graphs to Understand Data, J. P. Donnelly, Ed. 2013, pp. 200–209.

[32] M. Pfannkuch, "Comparing Box Plot Distributions: A Teacher's Reasoning," Stat. Educ. Res. J., vol. 5, no. 2, pp. 27–45, 2006.

[33] F. Marmolejo-ramos and T. S. Tian, "The shifting boxplot . A boxplot based on essential summary statistics around the mean .," Int. J. Psychol. Res., vol. 3, no. 1, pp. 37–45, 2010.

[34] L. Pace, "Correlation and Regression," in Beginning R: An Introduction to Statistical Programming, no. i, Apress, 2012, pp. 215–242.

[35] K. Y. Wang and T. T. Chau, "An Association between Air Pollution and Daily Outpatient Visits for Respiratory Disease in a Heavy Industry Area," PLoS One, vol. 8, 2013.

[36] M. S. Najib, M. N. Taib, N. A. M. Ali, M. N. M. Arip, and A. M. Jalil, "Classification of Agarwood grades using ANN," Int. Conf. Electr. Control Comput. Eng. 2011, pp. 367–372, 2011.

[37] M. S. Najib, M. U. Ahmad, P. Funk, M. N. Taib, and N. A. M. Ali, "Agarwood Classification : A Case-based Reasoning Approach Based on E-nose," in IEEE 8th International Colloqium on Signal Processing and Its Application, 2012, pp. 120–126.

[38] M. G. Voskoglou, "Fuzzy Sets in Case-Based Reasoning," 2009 Sixth Int. Conf. Fuzzy Syst. Knowl. Discov., vol. 6, pp. 345–373, 2009.

[39] N. Xiong and P. Funk, "Combined feature selection and similarity modelling in case-based reasoning using hierarchical memetic algorithm," Evolutionary Computation (CEC), 2010 IEEE Congress on. pp. 1–6, 2010.

[40] P. Funk and N. Xiong, "Case-Based Reasoning and Knowledge Discovery In Medical Applications With Time Series," Comput. Intell., vol. 22, no. 3/4, pp. 238–253, 2006.

[41] C. Tsatsoulis and A. B. Williams, "Case-Based Reasoning," in Knowledge- Based Sysyems, Volume 3., C. T. Leondes, Ed. Academic Press, 2000, pp. 807– 837.