Sensors & Transducers, Vol. 193, Issue 10, October 2015, pp. 86-92

Sensors & Transducers

http://www.sensorsportal.com

An Empirical Study for Quantification of Carcinogenic Formaldehyde by Integrating a Probabilistic Framework with Spike Latency Patterns in an Electronic Nose

[1] Muhammad HASSAN, [1, 2] Amine BERMAK, [3] Amine Ait Si ALI and [3] Abbes AMIRA

[1] School of Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong
[2] College of Science and Engineering, Hamad bin Khalifa University, Education City, Doha, Qatar
[3] College of Engineering, Qatar University, Doha, Qatar
[1] Tel.: 00852 23588526
E-mail: mhassan@connect.ust.hk

Received: 31 August 2015 /Accepted: 5 October 2015 /Published: 30 October 2015

Abstract: Recently, exposure to formaldehyde has appeared as a major concern since it has been listed as a human carcinogen. Conventional methods for its long-term monitoring are not feasible due to their high operational cost, long analysis time and the requirement of specialized equipment and staff. In this paper, we develop an electronic nose, containing an array of commercially available low cost Figaro gas sensors, to support autonomous and long-term monitoring of formaldehyde. Hardware friendly gas quantification without requiring any manual tuning of parameters is the major challenge with the electronic nose. We handle this challenge by treating it as a classification problem because data acquisition at continuously varying concentrations may incur large expense and a great deal of time. Instead, twenty different concentrations of formaldehyde with 0.25 ppm increment step in the target range between 0.25 to 5 ppm, spanning commonly found formaldehyde levels in indoor and outdoor environments, are input to obtain its signatures in order to quantify/classify its levels within this target range. A computationally efficient bio-inspired spike latency coding scheme, in which spike latencies corresponding to sensitivity patterns of the sensors in the array shift with the change in concentration, is targeted for this purpose. However, stochastic variability in the spike latency patterns, corresponding to repeated exposure to the same formaldehyde concentration level, is observed. We target two Bayesian inference methods, namely multivariate Bayesian and naive Bayes, to express the uncertainty about the spike latency patterns in terms of a probability encoding framework. These methods do not require any manual tuning of parameters, in contrast to other state of the art methods. A best performance of 95.75 % is achieved with the naive Bayes method on the experimentally obtained data set of formaldehyde.

Keywords: Carcinogenic formaldehyde, Sensor array, Spike latency pattern, Bayesian inference.

1. Introduction

Formaldehyde (CH_2O) is one of the most ubiquitous and reactive aldehydes in the environment. It is a colorless and strong-smelling chemical, which is widely used in building and furniture construction materials [1]. Building residents may be exposed to CH_2O gas when it is emitted from materials containing this chemical upon thermal or chemical decomposition. Unvented fuel burning appliances and

tobacco smoke may also cause CH_2O inhalation. One scientific study reveals that overall CH_2O emission from different building sources may exceed 2 parts per million (ppm) [2]. Short-term exposure to CH_2O may cause skin, nasal, throat, and eye irritation. The United States Occupational Safety and Health Administration (OSHA) has set its short-term exposure limit (15-30 minutes) at 2 ppm and permissible exposure limit (up to 8 hours) to 0.75 ppm [2]. Precautionary measures are required for healthy living when the CH_2O concentration reaches one half of the permissible exposure limit.

Formaldehyde received great attention in 1980, when its carcinogenicity was reported in rats and mice after its long-term inhalation [3]. Since that time, its carcinogenicity has also been studied in humans. In 2004, the International Agency for Research on Cancer (IARC) classified CH_2O as a human carcinogen based on sufficient evidence that its long-term exposure may cause nasopharyngeal cancer and leukemia in humans [4]. A recent study reported increasing concentration levels of CH_2O in the indoor environment of urban areas [5]. This alarming situation highlights the importance of formaldehyde monitoring with a low cost and robust solution on a long-term basis for healthy living. Unfortunately, traditional methods of gas chromatography and spectro-fluorimetry [5-7] cannot be used for the long-term monitoring of formaldehyde because specialized equipment and staff are required for the analysis of air samples collected from the area being monitored. Moreover, the cost and analysis time associated with these methods is very high.

Electronic nose systems, containing an array of gas sensors, emerged as a successful platform for the fast identification of gases in the last two decades, and they are targeted at many applications like food quality checking [8], diseases diagnosis [9], bacteria identification [10], environmental monitoring [11], beverages classification [12], paper quality inspection [13] and identification of health endangering indoor gases [14]. In this paper, we develop an electronic nose system, containing an array of six commercially available Figaro gas sensors, to quantify CH_2O concentration.

Motivated by the recent experimental findings in the field of neuroscience which report a logarithmic relationship between odor concentration and the spike latency of mitral cells [15], a logarithmic time domain scheme has been previously presented for gas classification by translating a sensor array response into a spike latency pattern [16]. Hardware friendly rank order based classifiers have then been developed for gas identification by using this technique [16-18]. In these classifiers, the temporal sequence of spikes is utilized to distinguish gases. We adopt this scheme to retrieve concentration information by utilizing the shifts in the spike latencies with the change in concentration. However, there is no straightforward relationship between the shift in the spike latency and the formaldehyde concentration as there is in the rank order based classifiers, where the change in relative times between spikes does not change the classification performance as long as their temporal order is not changed. Fig. 1 demonstrates the basic concept of this approach. The figure shows that the spike latency of each sensor is changed with the change in concentration but the arbitrary temporal sequence remains fixed.

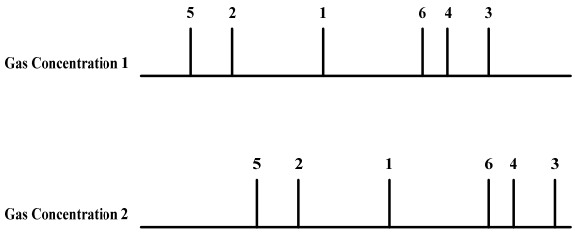

Fig. 1. Demonstration of change in spike latency of each sensor with the change in concentration.

Generally, gas sensors exhibit randomness in their responses due to inherent issues in this technology (such as drift, aging, change in operating conditions etc.); and as a result, stochastic variability is observed in the latency patterns, which makes gas quantification more challenging.

In this paper, we formulate this problem as a classification problem because the use of discrete values of concentration, spanning regular intervals within the target range of interest, is the most feasible option for obtaining an experimental data set in order to save time and cost. Data corresponding to each discrete value of concentration is treated as single class data.

A Bayesian inference approach [19] is targeted to deal with the randomness in the latency patterns because this provides an analytical solution and no manual tuning of parameters is required. This approach has been successfully used in neuroscience to build computational theories for perception and action [20]. There are two major steps in this approach. The first step is to learn the probability encoding model or the tuning curve for the spike latency patterns at each predefined concentration value of CH_2O from the experimental data obtained through the sensor array.

The second step is to use a Bayesian decoding model to estimate the formaldehyde concentration for a newly test latency pattern by using the learned probability encoding model. Multivariate Bayesian or quadratic discriminant analysis (QDA) and a naive Bayes classifier (NBC) are typically used under this framework when there are no shared parameters between the classes. Multivariate Bayesian assumes a relationship between features, while the naive Bayes classifier considers independence among features. The performance of both of these methods, along with those of other state of the art approaches, is evaluated by acquiring CH_2O data at twenty different concentration values, spanning from 0.25 to 5 ppm.

The paper is organized as follows. Section 2 explains the probabilistic inference approach for CH_2O concentration estimation. Next, Section 3 describes the experimental setup for data acquisition and evaluates the performance of the Bayesian inference methods along with other state of the art methods. Finally, the conclusion is drafted in Section 4.

2. Probabilistic Framework

A logarithmic time-domain encoding scheme has been used for gas identification in rank order based classifiers [16-18]. In these classifiers, the spike latency l_i (expressed in arbitrary units) of the i-th sensor corresponding to a target gas is represented as

$$l_i = \frac{\log x_i}{a_i}, \tag{1}$$

where x_i denotes the sensitivity of the i-th sensor and a_i is a sensor dependent parameter which is extracted through linear regression between the average log sensitivity of the sensors across the array as an explanatory variable and the sensitivity of the i-th sensor as an output variable. The resultant spike latency patterns carry information about the gas identity and its concentration. In rank order based classifiers [16-18], the temporal sequence of spikes is used for gas identification. In this paper, we utilize the change in spike latency of each sensor to estimate the concentration level of CH_2O.

The potential challenge with this scheme is that gas sensors usually exhibit randomness in their responses, which results in stochastic variability in the latency patterns. We use a probabilistic inference approach to retrieve concentration information from the random latency patterns.

The main objective of using probabilistic inference is to find the most probable concentration class or level of the newly test latency pattern by learning the distribution of latency patterns corresponding to each concentration level from the available measurements taken with the electronic nose. Let us consider the following notations for this probabilistic inference problem: Suppose we have a set of concentrations $\boldsymbol{c}=\{c_1, c_1, \ldots, c_n\}$, and the experimentally obtained latency pattern $\boldsymbol{l}=\{l_1, l_2, \ldots, l_d\}$, where d denotes the total number of sensors and n represents the total number of concentration classes or levels. For gas quantification algorithms, l_i corresponds to one single feature and $\boldsymbol{l}$ to a d-dimensional feature vector.

Probabilistic inference is a two-step process [20]. In the first step, we learn a model fitting that captures the mapping from $\boldsymbol{l}$ to $\boldsymbol{c}$ from the available sensor array measurements. In the second step, we use Bayesian decoding to estimate the concentration level c_j for the newly observed test latency pattern.

In order to learn the model fitting, we need to know the distribution or probability encoding model of the latency patterns conditioned on the CH_2O odor intensity. With a particular model, parameterized by a vector θ, we can use maximum likelihood (ML) to obtain the optimal estimate $\hat{\theta}$ for which the latency patterns are most likely:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, p(\boldsymbol{l}|\boldsymbol{c}, \theta) \tag{2}$$

In the multivariate Bayesian method [21], we assume that latency patterns follow multivariate Gaussian distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and hence second order statistics, that is, mean and covariance, are sufficient to learn this distribution. We use ML to estimate these parameters from the sensor array measurements. If $\boldsymbol{m}$ is the ML estimate of the true mean ($\boldsymbol{\mu}$) and $\boldsymbol{S}$ is the ML estimate of the true covariance matrix ($\boldsymbol{\Sigma}$), then the conditional density of the latency patterns with a given CH_2O odor intensity class c_j is given by

$$p(\boldsymbol{l}|c_j) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{S}_j|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{l}-\boldsymbol{m}_j)^T \boldsymbol{S}_j^{-1}(\boldsymbol{l}-\boldsymbol{m}_j)\right] \tag{3}$$

For a new latency pattern, Bayesian decoding is used to compute the posterior probability $p(c_j|\boldsymbol{l})$ of each concentration class c_j in the set with a given latency pattern $\boldsymbol{l}$. It can be described as

$$p(c_j|\boldsymbol{l}) = \frac{p(\boldsymbol{l}|c_j)p(c_j)}{p(\boldsymbol{l})}, \tag{4}$$

where $p(c_j)$ is the prior probability of the j-th class. In our case, we consider the same prior for each class. The denominator term $p(\boldsymbol{l})$ is for normalization, and the same for all classes, and hence can be ignored. As a result, the posterior probability ($p(c_j|\boldsymbol{l})$) only depends on the likelihood term ($p(\boldsymbol{l}|\, c_j)$).

The multivariate Bayesian approach considers correlations between latencies, and hence involves substantial computations and requires large storage for estimated parameters, i.e., $n \times (d \times d)$ for the covariance matrix and $(n \times d)$ for the mean vector for n classes. Moreover, the sample covariance matrix is considered as a poor estimate of the true covariance matrix in the case when the available measurements and feature vector size are comparable or their ratio is not extremely large [22]. On the other hand, the naive Bayes classifier assumes that features are independent and require comparatively few parameters, i.e., $(n \times d)$ for the mean vector and $(n \times d)$ for the variance. Due to the independent assumption, the likelihood term can be written as:

$$p(\boldsymbol{l}|c_j) = \prod_{i=1}^{d} p(l_i|c_j) \tag{5}$$

After estimating the sample mean (denoted as m_{ij}) and sample variance (denoted as v_{ij}) through ML for the i-th sensor and j-th class, the $p(l_i|c_j)$ can be written as for univariate Gaussian distribution:

$$p(l_i|c_j) = \frac{1}{\sqrt{2\pi {v_{ij}}^2}} \exp\left[-\frac{(l_i - m_{ij})^2}{2{v_{ij}}^2}\right] \quad (6)$$

Finally, the concentration class with the maximum posterior probability (represented as $\hat{c}$) is selected as an estimated concentration of the newly observed latency pattern:

$$\hat{c} = \underset{j}{\operatorname{argmax}}\, p(c_j|\boldsymbol{l}) \quad (7)$$

3. Experimental Setup and Performance Evaluation

We use six commercially available Figaro gas sensors to build an array for CH_2O concentration estimation. The description of these sensors is listed in Table 1, along with the names of the target compounds for which they are mainly marketed. Different part numbers are used in an attempt to achieve a unique latency pattern with varying sensitivity.

Table 1. Gas sensors used to acquire CH_2O signatures

No.	Figaro Part Number	Target Compounds
1.	TGS 826	Ammonia
2.	TGS 2600	Air contaminants
3.	TGS 2602	Volatile organic compounds
4.	TGS 2610	Liquefied petroleum gas
5.	TGS 2611	Methane
6.	TGS 2620	Solvent vapors

The experimental setup for acquiring the response of CH_2O at different concentrations is shown in Fig. 2. The sensor array is embedded in a glass container with an inlet valve for CH_2O exposure and outlet valve for its outflow. The cylinders of CH_2O and air are connected to mass flow controllers (MFCs), which are used to control the CH_2O concentration by mixing it with air in different proportions. A computer with a data acquisition board is used for MFCs programming in order to achieve the desired concentration of CH_2O and digitize the response of the sensor array.

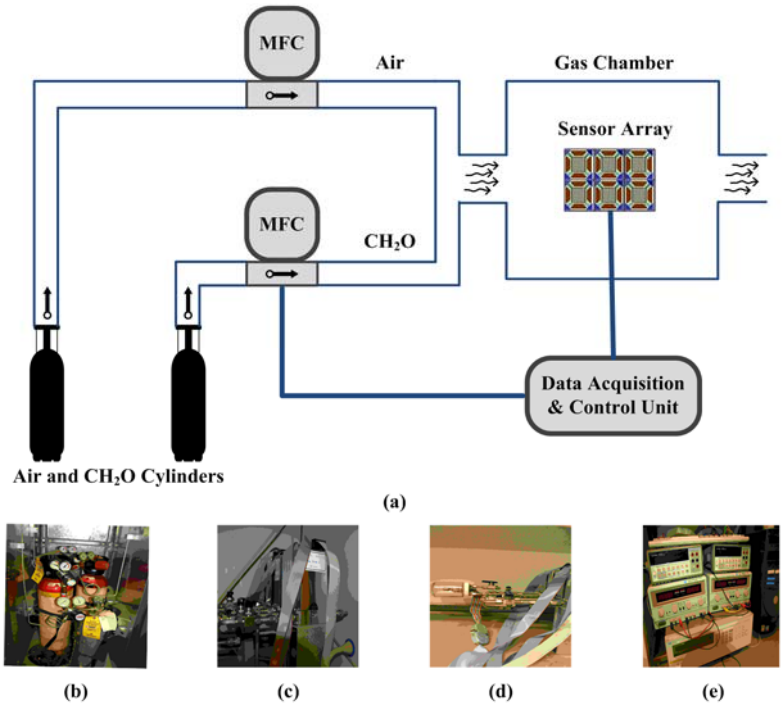

Fig. 2. (a) Block diagram of experimental setup to acquire data of CH_2O at different concentrations; (b) Gas cylinders; (c) MFCs; (d) Sensor array embedded in a glass chamber; (e) Data acquisition system.

We expose the sensor array to twenty different concentration values of CH_2O in the range between 0.25 ppm to 5 ppm, with a 0.25 ppm increment step. At the start of the experiment, the sensor array is firstly exposed to air for 750 seconds to obtain the response without any target gas vapors. This response is referred to as a baseline response. The sensor array is then exposed to CH_2O at a specified concentration for 500 seconds to obtain its response. Then air is again injected for 750 seconds to remove the CH_2O molecules trapped on the sensors in the previous gas injection phase and to recover the baseline response, i.e., the response without CH_2O vapors. All the sensors in the array respond to the target concentrations of CH_2O with different values of sensitivity. A typical response of the sensors in the array to CH_2O at two different concentrations is shown in Fig. 3. In the figure, the sensor array response corresponding to air

exposure is shown up to 750 seconds. Upon injection of CH_2O, with concentration value of 0.25 ppm, at 751 seconds, the resistance of each sensor starts decreasing. At 1250 seconds, air is again injected to allow the sensors to recover their baseline response. At 2000 seconds, CH_2O is again injected with increased concentration value of 0.5 ppm, which results in a further drop in the resistance of the sensors. We repeat this process to acquire 200 response patterns for the whole concentration range.

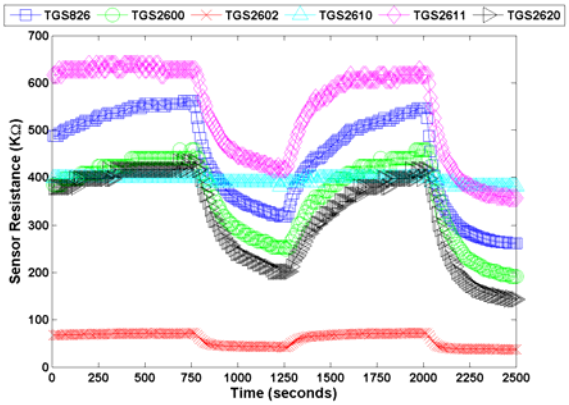

Fig. 3. Sensor array response to air and at two different concentrations of CH_2O.

From the resistance values, the sensitivity of each sensor is computed by dividing the value of the sensor resistance at the end of gas exposure by the sensor resistance at the end of air exposure. The regression coefficient a_i of each sensor is computed through linear regression between the log sensitivity of the *i*-th sensor and the average of the log sensitivity across the sensor array, as shown in Fig. 4.

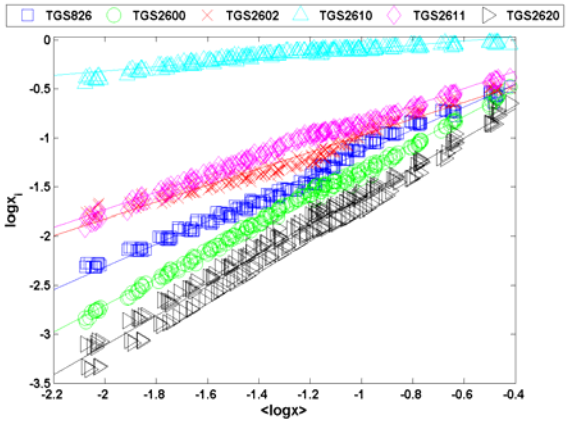

Fig. 4. Extraction of regression coefficients for each sensor in the array through linear regression.

The resultant regression parameters are used to transform the sensitivity pattern of the sensor array into a spike latency code or pattern by using Equation (1). Fig. 5 shows the latency values at each concentration level, where the latency is represented in arbitrary units resulting from Equation (1). There is no fixed value of spike latency at any concentration level. Instead, variability is shown at each concentration level and the mean value of the latency varies with the change in concentration.

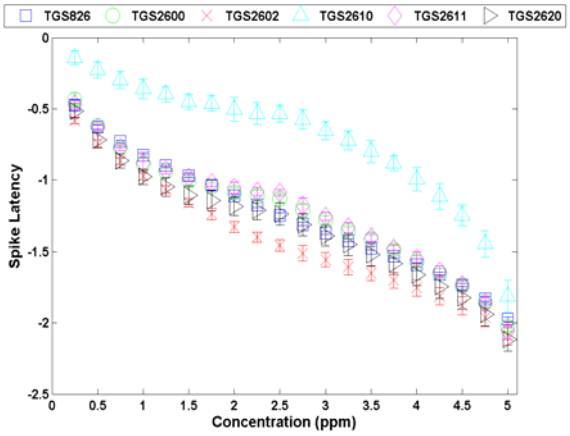

Fig. 5. Spike latency of each sensor at twenty different concentrations of CH_2O.

Bayesian inference methods along with other state of the art methods, including Gaussian mixture models (GMM), multi-layer perceptron (MLP) and support vector machines (SVM) with a linear and radial basis function (RBF) kernel, are used to estimate the CH_2O concentration class with the resultant data set. A 5 × 2 cross validation technique is used to evaluate the performance of these methods by dividing the experimental data into training, validation and testing sets. The performances of all these methods are summarized in Table 2.

Table 2. Performance comparison of algorithms for CH_2O concentration quantification/classification

Classification Method	Classification Performance (%)	Maximum Concentration Error (ppm)
GMM	91.25	0.25
MLP	89.25	0.50
SVM (Lin)	88.5	0.25
SVM (RBF)	92.5	0.25
QDA	92.75	0.25
NBC	95.75	0.25

In terms of maximum concentration error, the performance of most of the classifiers is comparable, but a maximum accuracy of 95.75 % is achieved with naive Bayes classifier (NBC) to correctly classify the true concentration class. The better performance of the naive Bayes classifier among other state of the art methods has been reported in many different domains

[23-27], even when the independence assumption does not hold, which is justified in [26-27] by showing its optimality under dependence conditions. The simpler implementation of the naive Bayes classifier as compared to other methods also facilitates the integration of a simple hardware solution with the sensor array.

A simplified gas quantification system can be developed with the naive Bayes classifier as shown in Fig. 6. Parameters for latency formation and univariate Gaussian distribution are computed from the training data and stored in memory. A newly test sensitivity pattern is first transformed to a latency pattern by using the corresponding regression coefficients and then the stored distribution parameters of each class are used to compute the conditional probability. Finally, the concentration class with maximum probability is assigned to the new test pattern.

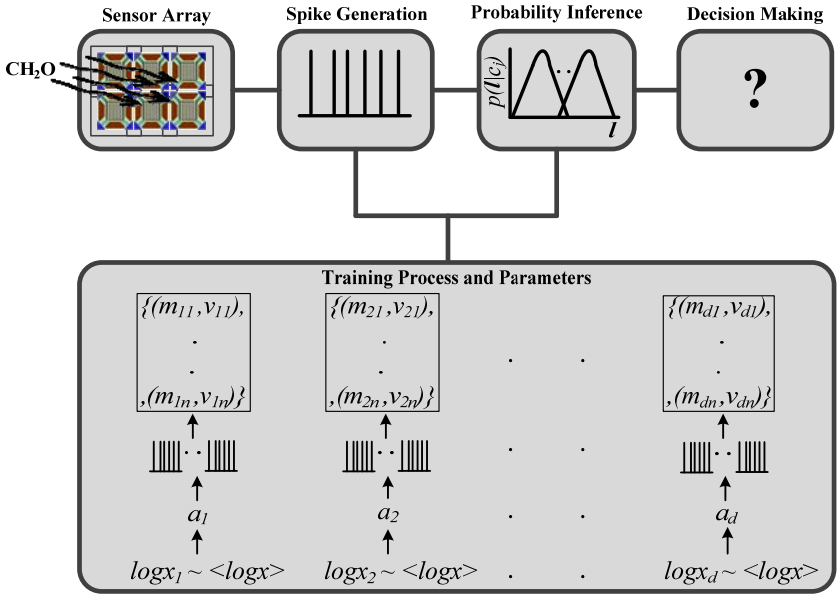

Fig. 6. Block diagram for CH_2O quantification with naive Bayes classifier by utilizing spike latency patterns.

4. Conclusion

In this paper, we have proposed a low cost and compact solution in the form of an electronic nose, containing an array of commercially available low cost gas sensors, to estimate the concentration of health endangering formaldehyde. As compared to this solution, commercial methods for formaldehyde monitoring are more accurate but cannot be adopted for continuous monitoring on a long-term basis due to the time, cost and complex sequence of procedures involved in these methods. Instead of focusing on exact knowledge of concentration, we treat this problem as a classification problem by dividing the target concentration range at evenly spaced intervals. A bio-inspired spike latency coding scheme is used to extract knowledge of concentration. In order to estimate concentration in the presence of stochastic variability in the latency patterns, different classification methods are explored. The best performance is achieved with a naive Bayes classifier. Besides providing an analytical solution and not requiring any manual tuning of parameters, the naive Bayes classifier also facilitates hardware friendly gas quantification/classification which can be realized with a simple hardware platform. A building resident can continuously monitor formaldehyde concentration levels with the electronic nose and can set a threshold to indicate when to take preventive action, such as removal of the formaldehyde source or provision of proper ventilation to reduce its concentration.

Acknowledgements

This paper was made possible by the National Priorities Research Program (NPRP) grant No. 5 - 080 - 2 - 028 from the Qatar National Research Fund (a member of the Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

[1]. M. N. Indang, A. S. Abdulamir, A. A. Bakar, A. B. Salleh, Y. H. Lee, Y. N. Azah, A review: methods of determination of health-endangering formaldehyde in diet, *Research Journal of Pharmacology*, Vol. 3, No. 2, 2009, pp. 31-47.

[2]. T. Salthammer, S. Mentese, R. Marutzky, Formaldehyde in the indoor environment, *Chemical Reviews*, Vol. 110, No. 4, 2010, pp. 2536-2572.

[3]. W. D. Kerns, K. L. Pavkov, D. J. Donofrio, E. J. Gralla, J. A. Swenberg, Carcinogenicity of

formaldehyde in rats and mice after long-term inhalation exposure, *Cancer Research*, Vol. 43, No. 9, 1983, pp. 4382-4392.
[4]. International Agency for Research on Cancer, IARC classifies formaldehyde as carcinogenic to humans, Press release. June15, 2004, URL: http://www.iarc.fr/en/mediacentre/pr/2004/pr153.html [accessed: 2015-09-09].
[5]. P. Guo, K. Yokoyama, F. Piao, K. Sakai, Sick building syndrome by indoor air pollution in Dalian, *International Journal of Environmental Research and Public Health*, Vol. 10, No. 4, 2013, pp. 1489-1504.
[6]. J. G. Dojahn, W. E. Wentworth, S. D. Stearns, Characterization of formaldehyde by gas chromatography using multiple pulsed-discharge photoionization detectors and a flame ionization detector, *Journal of Chromatographic Science*, Vol. 39, No. 2, 2001, pp. 54-58.
[7]. H. L. Pinheiro, M. V. de Andrade, P. A. de Paula Pereira, J. B. de Andrade, Spectrofluorimetric determination of formaldehyde in air after collection onto silica cartridges coated with Fluoral P, *Microchemical Journal*, Vol. 78, No. 1, 2004, pp. 15-20.
[8]. S. Capone, P. Siciliano, F. Quaranta, R. Rella, M. Epifani, L. Vasanelli, Analysis of vapours and foods by means of an electronic nose based on a sol-gel metal oxide sensors array, *Sensors and Actuators B-Chemical*, Vol. 69, No. 3, 2000, pp. 230-235.
[9]. A. P. F. Turner, N. Magan, Electronic noses and disease diagnostics, *Nature Reviews Microbiology*, Vol. 2, No. 2, 2004, pp. 161-166.
[10]. M. Trincavelli, S. Coradeschi, A. Loutfi, B. Söderquist, P. Thunberg, Direct identification of bacteria in blood culture samples using an electronic nose, *IEEE Transactions on Biomedical Engineering*, Vol. 57, No. 12, 2010, pp. 2884-2890.
[11]. R. E. Baby, M. Cabezas, E. N. Walsöe de Reca, Electronic nose: a useful tool for monitoring environmental contamination, *Sensors and Actuators B-Chemical*, Vol. 69, No. 3, 2000, pp. 214-218.
[12]. M. Penza, G. Cassano, Classification of food, beverages and perfumes by WO3 thin-film sensors array and pattern recognition techniques, *Sensors and Actuators B-Chemical*, Vol. 73, No. 1, 2001, pp. 76-87.
[13]. M. Holmberg, F. Winquist, I. Lundström, J. W. Gardner, E. L. Hines, Identification of paper quality using a hybrid electronic nose, *Sensors and Actuators B-Chemical*, Vol. 27, No. 1, 1995, pp. 246-249.
[14]. M. Hassan, A. Bermak, Threshold detection of carcinogenic odor of formaldehyde with wireless electronic nose, in *Proceedings of the IEEE Sensors*, Valencia, November 2014, pp. 1368-1371.
[15]. T. W. Margrie, A. T. Schaefer, Theta oscillation coupled spike latencies yield computational vigour in a mammalian sensory system, *Journal of Physiology*, Vol. 546, No. 2, 2002, pp. 363-374.
[16]. H. T. Chen, K. T. Ng, A. Bermak, M. K. Law, D. Martinez, Spike latency coding in biologically inspired microelectronic nose, *IEEE Transactions on Biomedical Circuits and Systems*, Vol. 5, No. 2, 2011, pp. 160-168.
[17]. M. Hassan, S. Brahim Belhaouari, A. Bermak, Probabilistic rank score coding: A robust rank-order based classifier for electronic nose applications, *IEEE Sensors Journal*, Vol. 15, No. 7, 2015, pp. 3934-3946.
[18]. J. A. Yamani, F. Boussaid, A. Bermak, D. Martinez, Glomerular latency coding in artificial olfaction, *Frontiers in Neuroengineering*, Vol. 4, art. 18, 2012, pp. 1-9.
[19]. E. Alpaydin, Introduction to Machine Learning, *MIT Press*, 2010.
[20]. D. C. Knill, A. Pouget, The Bayesian brain: the role of uncertainty in neural coding and computation, *Trends in Neuroscience*, Vol. 27, No. 12, 2004, pp. 712-719.
[21]. M. Hassan, A. Bermak, A. A. S. Ali, A. Amira, Bayesian inference using spike latency codes for quantification of health endangering formaldehyde, in *Proceedings of the 6th International Conference on Sensor Device Technologies and Applications (SENSORDEVICES' 2015)*, Venice, Italy, 23-28 August 2015, pp. 79-82.
[22]. A. P. Dempster, Covariance Selection, *Biometrics*, Vol. 28, No. 1, 1972, pp. 157-175.
[23]. B. Turhan, A. Bener, Analysis of Naive Bayes' assumptions on software fault data: An empirical study, *Data & Knowledge Engineering*, Vol. 68, No. 2, 2009, pp. 278-290.
[24]. J. M. A. Luz, M. S. Couceiro, D. Portugal, R. P. Rocha, Hélder Araújo, Gonçalo Dias, Comparison of classification methods for golf putting performance analysis, in *Computational Intelligence and Decision Making*, Springer, Netherlands, 2013, pp. 35-45.
[25]. Y. Li, R. Anderson-Sprecher, Facies identification from well logs: A comparison of discriminant analysis and naive Bayes classifier, *Journal of Petroleum Science and Engineering*, Vol. 53, No. 3, 2006, pp. 149-157.
[26]. H. Zhang, Exploring conditions for the optimality of naive Bayes, *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 19, No. 2, 2005, pp. 183-198.
[27]. P. Domingos, M. Pazzani, Beyond independence: conditions for the optimality of the simple Bayesian classifier, in *Proceedings of the International Conference on Machine Learning*, 1996, pp. 105-112.