ISSN: 2456 – 3137
International Journal of Computational Research and Development
Impact Factor 4.775, Special Issue, January - 2017
International Conference on Smart Approaches in Computer Science Research Arena
On 5th January 2017 Organized By
Department of Computer Science, Sri Sarada College for Women (Autonomous), Salem, Tamilnadu

NEURAL NETWORK BASED VOC ANALYSIS FOR LUNG CANCER DETECTION

V. K. Uma* & D. Arul Pon Daniel**
Department of Computer Science, Loyola College of Arts & Science, Mettala, Namakkal, Tamilnadu

Cite This Article: V. K. Uma & D. Arul Pon Daniel, "Neural Network Based VOC Analysis for Lung Cancer Detection", International Journal of Computational Research and Development, Special Issue, January, Page Number 7-10, 2017.

Abstract:

The Volatile Organic Compounds (VOC) is analyzed for industrial and medical purpose. In the medical field it is found to be an important biomarker of lung cancer. This VOC's can be analyzed using array of gas sensors. In an open sampling system, when the sensor array is directly exposed to the environment being analyzed, the identification of chemical substances present a more difficult challenge due to the dispersion of gaseous chemical and the environmental changes. This problem can be overcome by combining the gas sensor array with an algorithm to compensate for the environmental changes. Here artificial neural network is used to train the sensor array to detect VOC's under various conditions. This system can be used to detect various volatile organic compounds under complex environmental conditions.

Key Words: Gas Sensors, VOC, Lung Cancer & Exhaled Breath

Introduction:

The design of inexpensive, reliable, quick responding, highly sensitive and power efficient chemosensory array systems also referred to as electronic nose systems (e-nose) is of important task in the industrial and medical field. The e-nose system is of important one in the medical research since it is found to be capable of detecting Lung cancer. The Lung cancer can be identified using analyzing VOC's present in the exhaled breath of human. The e-nose systems have neither achieved the potential in application tasks that go beyond the identification and differentiation of chemical substances nor the market penetration expected by the pioneers. Part of the reasons responsible for the crash of these applications are the natural mechanisms dominating the dispersion of chemical gaseous analytes in environmental conditions, namely diffusion, turbulence, and advection, as well as the sensitivity of chemical sensors to temperature, air flow and humidity. There are numerous computational algorithms have been proposed to enhance the performance of e-nose systems, which are important for sensor drift reduction as well as minimizing the impact of sensor failures, it still remains un conclusive to what extent the information obtained from these chemosensory systems can be exploited to reliably discriminate gases in realistic sampling scenarios, e.g., wind tunnels. Gas-phase chemical analysis by means of electrical transduction involving the prediction of the identity and quantity of the chemical compound has been traditionally performed in highly tight-controlled sensing test chambers that isolate the chemical analyte from its natural, predominantly complex environmental condition. Because it ensures a strict, tight control over some critical sensing conditions, like environmental factors such as temperature, pressure, and ambient flow, such isolation enables the chemical sensory system to exhibit chemical signatures that are related to the analyte being monitored and the sensing material used. The other methods used in analytical chemistry include Traditional methods of analytical chemistry, includes mass spectrometry and gas chromatography, are other prominent examples of such isolating techniques that are very useful in the identification and quantification of chemical analytes .These types of analysis, however, not only require indirect, rather complicated sampling procedures, also including cases involving the destruction of the tested sample and most importantly this methods cannot reveal the spatial and temporal structure of the chemical stimulus in its natural ambient. A gaseous chemical plume emitted from a fixed location conveys two critical pieces of information of the sensory world enclosed within its own volume: information about the identity of the analyte and information revealing the spatial coordinates of, or distance between, the source point and the observer to some extent.

Lung Cancer Detection System:

Lung Cancer is the leading cause of cancer-related death, not only in males all around the world but also in females. Lung cancer causes 1.4 million deaths per year worldwide. Early stage disease is amenable to curative surgery, but to date no screening process has been able to detect disease at a stage which has altered the overall survival. The measurement of VOCs in exhaled breath will provide a noninvasive technique for assessing lung pathology, and some of which are associated with Lung Cancer. There was a significant difference in the smell print of patients with lung cancer compared with healthy subject. These VOC's can be analyzed using array of gas sensor. The signal from the sensors can be classified to differentiate cancer and healthy patients using neural network. The Lung Cancer detection system is shown in the figure 1.

The analysis of VOC in open atmosphere is a challenging task since it is affected varying atmospheric conditions such as temperature, humidity etc. In order to overcome this problem the sensor array can be combined with an algorithm to compensate the changes. The sensor array can be calibrated and trained for analyzing VOC using the open dataset available at UCI repository.

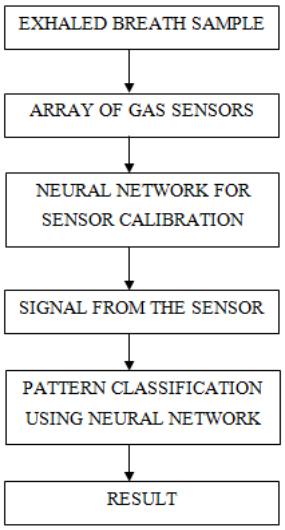

Figure 1: Lung Cancer Detection System

ISSN: 2456 – 3137
International Journal of Computational Research and Development
Impact Factor 4.775, Special Issue, January - 2017
International Conference on Smart Approaches in Computer Science Research Arena
On 5th January 2017 Organized By
Department of Computer Science, Sri Sarada College for Women (Autonomous), Salem, Tamilnadu

Data Set:

The dataset freely available at the UCI repository is used for training the sensor. The dataset is collected utilizing nine portable sensor array modules, each endowed with eight metal oxide gas sensors manufactured by Figaro Inc. The sensor array is positioned at six different line locations normal to the wind direction, creating there by a total number of 54 measurement locations called uniform measurement grid uniformly distributed throughout the entire wind tunnel test-bed facility. In particular the dataset consists of a very extensive selection of multiple mission/scenario-representative chemical analyte species, namely, acetone, acetaldehyde, ammonia, butanol (butyl-alcohol), ethylene, methane, methanol, carbon monoxide, benzene, and toluene, which in addition to their industrial applications as precursors in the manufacture of explosives, narcotics, and polymers, these chemical agents are highly recognizable to pose an immediate danger to life and health in public and military places. The dataset ultimately induces a 10-class gas discrimination problem, in which the goal is to identify and discriminate the 10 distinct, high-priority chemical analyte hazards at relevant concentrations in real-world operating environments regardless of the location of the sensory system platform within the annotated wind tunnel test-bed facility. The entire list of chemical analyte hazards as well as their nominal concentration values at the outlet of the gas source is in parts-per-million by volume (ppmv). The dataset contains18000 times-series measurements recorded from a 72 metal-oxide gas sensor array-based chemical detection platform. Every measurement contains 72 time-series recorded during 260 seconds, each collected at a sample rate of 100 Hz(samples per second).The dataset also contains time, temperature, and relative humidity information. The resulting dataset ultimately includes 75-time series composed of 26000 points. For manipulation purposes, the data is organized into eleven folders, each containing the number of measurements per chemical class identity and nominal concentration indicated. Each folder contains 6 folders, each representing the line location within the test area of the wind tunnel from which the set of time-series were recorded.

Neural Network:

The feed forward neural network is used here for the analysis purpose. The gradient descent learning algorithm is used for training the neural network. The neural network consists of seventy two inputs which are the important features used for analysis of volatile organic compounds. The seventy two inputs are the output from the nine sensor arrays each consisting of the eight sensors. The inputs are normalized before feeding to the neural network. The training and the testing data are randomly chosen and the network is trained using the training data. The network after training is tested against the test data. The structure of the network used is shown in the below figure 2.

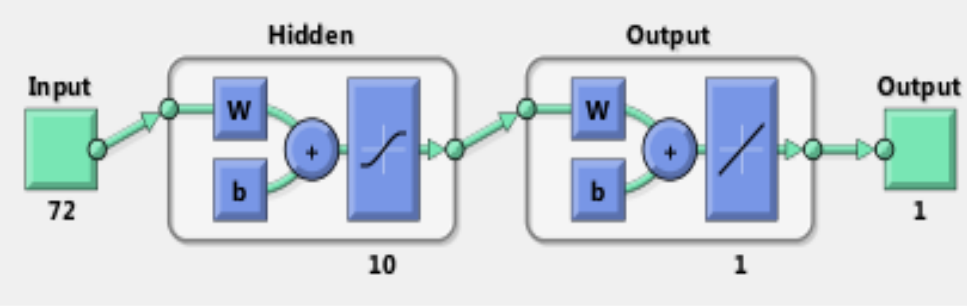

Figure 2: Structure of Neural Network used for Analysis

There is no generalized method to determine the optimum values for number of hidden layers, neurons in each hidden layer, etc., as they are working of expected intelligence.

Apply the input vector to the input units. Let $X_p = (x_{p1}, x_{p2}, \ldots, x_{pN})^t$ is be an input vector.

Calculate the net input values to the hidden layer units:

$$net_{pj}^{h} = \sum_{(i=1)}^{N} w_{ji}^{h} x_{pi} + \theta_{j}^{h}$$

Where net_{pj}^{h} net is the net input to hidden layer, w_{ji}^{h} is the weight on the connection from i^{th} input unit θ_j^h is the bias term and "h" refers to quantities on the hidden layer.

Calculate the outputs from the hidden layer:

$$i_{pj} = f_{j}^{h}\left(net_{pj}^{h}\right)$$

Where i_{pj} is the output from hidden layer and f_{j}^{h} is the activation function. Move to the output layer.

Calculate the net-input values to each units:

$$net_{pk}^{o} = \sum_{(j=1)}^{L} w_{kj}^{o} i_{pj} + \theta_{k}^{o}$$

where net_{pk}^{o} is the net input to the output layer, w_{kj}^{o} is the weight in the connection from j^{th} hidden unit, θ_k^o is the bias term and "o" refers to quantities on the output layer.

Calculate the outputs:

$O_{pk} = f_{k}^{o}\left(net_{pk}^{o}\right)$ where O_{pk} is the output obtained from the output layer

Calculate the error terms for the output units:

ISSN: 2456 – 3137
International Journal of Computational Research and Development
Impact Factor 4.775, Special Issue, January - 2017
International Conference on Smart Approaches in Computer Science Research Arena
On 5th January 2017 Organized By
Department of Computer Science, Sri Sarada College for Women (Autonomous), Salem, Tamilnadu

$\delta_{pk}^{o} = (y_{pk} - O_{pk}) f_j^{o'}\left(net_{pk}^{o}\right)$ where δ_{pk}^{o} is the error at each output unit,

$\delta_{pk}^{o} = y_{pk} - O_{pk}$ where y_{pk} is the desired error and O_{pk} is the actual error

Calculate the error terms for hidden units:

$\delta_{pj}^{h} = f_j^{h\prime}\left(net_{pj}^{h}\right)\sum_k \delta_{pk}^{o} w_{kj}^{o}$ where δ_{pj}^{h} is the error at each hidden unit

Notice that the error terms on the hidden units are calculated before the connection weights to the output-layer units have been updated.

Update weights on the output layer:

$$w_{kj}^{o}(t+1) = w_{kj}^{o}(t) + \eta\delta_{pk}^{o} i_{pj}$$

Update weights on the hidden layer:

$$w_{ji}^{h}(t+1) = w_{ji}^{h}(t) + \eta\delta_{pj}^{h} x_i$$

where η is the learning rate parameter. The order of the weight updates on an individual layer is not important. Be sure to calculate the error term

$$E_p = \frac{1}{2}\sum_{k=1}^{M} \delta_{pk}^{2}$$

since this quantity is the measure of how well the network is learning. When the error is acceptably small for each of the training-vector pairs, training can be discontinued [9]. The neural network training is completed at 896 epoch. The 75% of the data is used for training and the remaining 25% of the data is used for testing.

Results and Discussions:

The sensor array is trained and calibrated by fixing the heater voltage to 5V and for three wind speeds & six different positions. Here the array is trained for single heater voltage, three variable wind speeds and six different locations. The results after training and testing of network are given in terms of accuracy. The following table 1 shows the results of the neural network.

Table 1: Accuracy of the Neural Network for various conditions

Training Position	Wind Speed (m/s)	Acurracy (%)
1	0.10	74.25
	0.21	80.51
	0.34	80.22
2	0.10	76.54
	0.21	80.21
	0.34	80.01
3	0.10	82.34
	0.21	84.09
	0.34	83.76
4	0.10	88.56
	0.21	89.86
	0.34	89.65
5	0.10	82.45
	0.21	80.78
	0.34	81.67
6	0.10	79.45
	0.21	80.11
	0.34	79.67

From the Table 1 it is inferred that the system gives the maximum accuracy of 89.86% for sensor positioned at the location 4 and for the wind speed of 0.21m/s and for the heater voltage of 5V. The sensor array can analyze the VOC's with a accuracy of 89.86% when trained at position normal to the wind direction.

Conclusion:

In this paper neural network has been used to classify the lung cancer as malignant or benign or normal. Based on the obtained results the resilient algorithm produced, up to the mark of classification accuracy 89.69%, during training Feed-forward Back-propagation neural network. It was also observed that, in general, single hidden layer Back-propagation neural neural network provided the better classification accuracy to classify the breath samples of lung cancer. In the near future, we need to standardize the procedures and develop a learning system widely acceptable by breath analyst throughout the world. In this way, we will be able to reduce deaths due to lung cancer, the first leading cause of cancer deaths worldwide.

References:

1. Alexander Vergara, Jordi Fonollosa, Jonas Mahiques, Marco Trincavelli, Nikolai Rulkov, Ramon Huerta, " On the performance of gas sensor arrays in open sampling systems using Inhibitory Support Vector Machines", Sensors and Actuators in 2013.

2. Vanessa H. Tran, Hiang Ping Chan, Michelle Thurston, Paul Jackson, Craig Lewis, Deborah Yates, Graham Bell, and Paul S. Thomas, "Breath Analysis of Lung Cancer Patients Using an Electronic Nose Detection System", IEEE Sensors Journal in 2010.
3. Hiang Ping Chan, Craig Lewis and Paul S. Thomas, "Exhaled breath analysis: Novel approach for early detection of lung cancer", Journal of Lung Cancer in 2009.
4. Marco.S, Gutierrez-Galvez. A, "Signal and data processing for machine olfactionand chemical sensing: a review", IEEE Sensors Journal in 2012.
5. Ziyatdinov. A, Marco. S, Chaudry. A, Persaud. K, Caminal. P, Perera. A, "Drift compensation of gas sensor array data by common principal component analysis", Sensors and Actuators B: Chemical in 2010.
6. Susan C. van't Westeinde and Rob J. van Klaveren , "Screening and Early Detection of Lung Cancer", The Cancer Journal in 2011.
7. Hiang Ping Chan, Vanessa Tran, Craig Lewis and Paul S. Thomas, "Elevated Levels of Oxidative Stress Markers in Exhaled Breath Condensate", Journal of Thoracic Oncology in 2009.
8. Peled N, Hakim M, Tisch U, Bunn Jr PA, Miller YE, Kennedy Timothy C, Mattei, Jane , Mitchell, John, Hisrch, Fred R and Haick Hossam , "Non-invasive breath analysis of pulmonary nodules", Journal of Thorac in 2012.
9. James A. Freeman and David M. Skapura. "Neural Networks Algorithms, Applications and Programming Techniques", pp 115-116, 1991.