Data Analytics for Home Air Quality Monitoring

Petya Mihaylova[2], Agata Manolova[3](✉), and Petia Georgieva[1]

[1] Department of Electronics, Telecommunications and Informatics, University of Aveiro, Avero, Portugal
[2] English Language Faculty of Engineering, Technical University of Sofia, Sofia, Bulgaria
[3] Faculty of Telecommunications, Technical University of Sofia, Sofia, Bulgaria
amanolova@tu-sofia.bg

Abstract. Modern air quality monitoring systems are characterised by high complexity and costs. The expensive embedded units such as sensor arrays, processors, power blocks, displays and communication units make them less appropriate for small indoor spaces.

In this paper we demonstrate that two widely available, in private houses, sensors (for Humidity and Temperature) are promising alternative, to the expensive indoor air quality solutions, provided with intelligent data processing tools. Our findings suggest that neural network based data analytics system can learn to discriminate unusual indoor gases from normal home air components based only on temperature and humidity measurements.

Keywords: Indoor air quality · Data analytics · Neural network · Deep Autoencoder Neural Network

1 Introduction

Nowadays, people spend much time in closed spaces, therefore monitoring of indoor air quality attracted much attention in recent years. Standards, guidelines and requirements, defined by international agencies, are used to evaluate the acceptable quality of air in indoor as well as outdoor environments. Research is carried out to bring dust free, noxious free and smell free environment at home, hospitals, schools, cars, etc. Several air quality monitoring systems have been recently proposed.

A grey model to indoor air quality management in rooms based on real-time sensing of nano and micro particles and volatile organic compounds is proposed in [1]. Pollution sources are analysed and a management model is defined to minimize the time during which the pollutant concentration falls below threshold value. An embedded system model for air quality monitoring is proposed in [2] using low cost gas sensors and Arduino microcontroller. The system is tailored to study the long-term impacts of bad air quality on health particularly with respect to allergic patients. An enhancement of the gas sensitivity and selectivity of a piezoelectric micro-cantilever by using chemically-modified carbon nanotubes as a sorbing layer and a gas sensitive film is studied in [3]. Two measurement modes are compared, the frequency mode that

V. Poulkov (Ed.): FABULOUS 2019, LNICST 283, pp. 79–88, 2019.
https://doi.org/10.1007/978-3-030-23976-3_8

requires a significant amount of nanotubes to coat the cantilever and trap target molecules and the resistance mode that needs a small amount of nanotubes for the film not to be too conductive. An improved sensor based on resistance mode is demonstrated to achieve good sensitivity to gases.

Smart home system with embedded gas sensors arrays to control not only the indoor air quality but also to discriminate room occupancy and human activities in presented in [4]. A portable chemical-based monitoring system built of 32 gas sensors array has been tested in NASA space craft cabin simulator. The compact autonomous system (3.6 L volume, 3.4 kg weight) is composed of polymer-carbon composite elements and is suitable for a long term continuous operation. The system detects the number of individuals present in the room and the number of people exercising. To correct for sensor drift and improve the precision, during periods of lack of activity in the room, the sensors' baselines can be adjusted. Due to air circulation, odours travel from one room to another and thus the sensing range of chemical sensors appear to be wider than video camera-based systems. Interestingly, the system is able to detect human behaviours that caused higher concentration levels of ethanol. Gas sensor arrays are usually used for discrimination of gas mixtures composed of air and single chemical such as hexane, ethanol, acetone, ethyl acetate and toluene. Method for gas mixtures discrimination based on sensor array, temporal response and data driven approach is proposed in [5]. Furthermore in [6], gas recognition by activated thin-film sensors array is studied. Principle Component Analysis (PCA) is applied to cluster target gases. Gas discrimination using nano-electronic nose has been applied in [7]. The integration of nanowire and carbon nanotube sensors, precise control of the sensor temperature, and the use of PCA for data processing resulted in effective discrimination between a wide variety of gases, including explosive ones and nerve agents. The response of these sensors to hydrogen, ethanol, and NO2 were measured at different concentrations and both at room temperature and at 200 °C.

A method for online de-correlation of chemical sensor signals from the effects of environmental humidity and temperature variations is proposed in [8]. The accuracy of electronic nose measurements for continuous monitoring is improved taking into account the simultaneous readings of environmental humidity and temperature. The electronic nose setup, built for this study, consists of eight metal-oxide (MOX) gas sensors, temperature and humidity sensors with a wireless communication link to external computer. The wireless electronic nose was used to monitor the air for two years in one residence and collected data continuously during 537 days with a sampling rate of 1 sample per second. To test the benefits of de-correlating humidity and temperature measurements from the MOX sensors' responses, a scenario with three gas stimuli has been designed – banana, wine and baseline responses. Multiclass inhibitory support vector machine (ISVM) method [9] is used to discriminate between the presence of banana, presence of wine, and baseline activity. To compare the performance of the classifier with and without decorrelation of humidity–temperature, four subsets of data were created by combining raw sensor responses, filtered sensor data, and temperature and humidity. Experimental results show that including the filtered data in the classification model improves significantly the discrimination capability of the model. In summary, it has been shown that simultaneous humidity and temperature recordings are promising to extract relevant chemical signatures.

The reviewed air quality monitoring systems are complex and costly. They integrate expensive components such as sensor networks, processors, power blocks, displays and communication units. Such technology is more appropriate for public building and less suitable for private houses. Moreover, the large number of sensors requires significant computational resources and processing time for data analytics.

In this paper we propose data analytics models for air quality monitoring adequate for small indoor spaces such as private houses. The models are based on machine learning approach (deep neural networks) and are demonstrated with real data provided by the authors of [8] in UCI Machine Learning Repository (Gas sensors for home activity monitoring Data Set). The task is to detect gas changes due to different home activities based on measurements of ten sensors (eight MOX gas sensors, temperature and humidity sensors). The proposed data analytics models differ in terms of number of sensors as inputs and temporal length of the sensor readings. The goal is to build reliable gas discrimination model based only on short time measurements of temperature and humidity sensors.

The rest of the paper is organized as follow. In Sect. 2 deep neural networks are introduced as the proposed data analytics model. In Sect. 3 the sensor array of data is described. The implementation aspects and obtained results are discussed in Sect. 4. Conclusions are drawn in Sect. 5.

2 Deep Neural Networks

Over the last decade, deep learning techniques have become very popular in various application domains such as computer vision, automatic speech recognition, natural language processing, and bioinformatics where they achieved excellent results on various tasks. For example, neural networks with multiple hidden layers (deep neural networks) are very successful in solving classification problems for high dimensional data. Each layer learns to represent the data at a different level of abstraction. The idea of having one algorithm that first maps data into a representative feature space and then solve recognition tasks gained the great success of deep neural networks (DNNs). DNN models have been applied within a wide range of applications including images, videos, speech, text, [10–12], and recently also in neuro-imaging domain [13, 14]. The DNN success is due to their ability to extract representations that are robust to partial translation and deformation of input patterns.

In the present study we explore the advantages of Deep Autoencoder Neural Network (DANN) in the context of sensor array data modelling and compare it with a shallow neural network (NN) model. The following training procedure for DANN was implemented:

(a) Training the first hidden layer of the auto-encoder without providing the labels. An auto-encoder is a neural network which attempts to replicate its input at its output. Thus, the input and the output have is the same size. The auto-encoder is comprised of an encoder followed by a decoder. The encoder maps an input to a hidden representation and the decoder attempts to reverse this mapping to reconstruct the original input.
(b) The next hidden layers (auto-encoders) are trained in a similar way. The main difference is that the features that are generated from the previous auto-encoder are the training data for the next auto-encoder. The size of each subsequent encoder is decreasing, so that it learns compressed input data representation.
(c) Unlike the auto encoders, the final (output) layer is trained in a supervised fashion using training data labels. Softmax function was used as a processing unit in this layer.
(d) Final retraining of the whole DANN is performed in a supervised mode applying error-backpropagation. This step is referred as fine DNN tuning.

3 Data and Experimental Scenario Description

The complete data set [15] consists of recordings of ten sensors (8 MOX gas sensors, temperature and humidity sensors). The sensors were exposed to two specific stimuli (wine or banana smells) and background home smells. The responses to banana and wine stimuli were recorded by placing the stimulus close to the sensors. The duration of each stimulation varied from 7 min to 2 h, with an average duration of 42 min. The dataset contains a set of time series from three different conditions: wine, banana and background activity. There are 35 inductions with wine, 33 with banana and 31 recordings of background activity, corresponding to measurements along 99 days. The dataset is composed of 99 snippets of time series, each being a single induction or background activity. In total, there are 919438 samples. For each induction, the time when the stimulus was presented is set to zero.

The system requirement is to detect on-line early change of indoor air composition and give an alarm. Therefore, the sensor response to a stimulus at the beginning of each experiment (the first few minutes) is of major importance. Sensor records over the first 160 s. of the experiment taken each 5 s were extracted from the original dataset. Figure 1 shows samples from all sensors taken from one experiment with banana stimulus. Samples for negative values of time correspond to sensor responses before the stimulus presentation.

Data structure (Table 1) consists of 99 examples (99 days of experiments with different stimulus) and time series of 32 readings per sensor for a total of ten sensors. Three hypotheses are studied: detection of air quality variation based on (i) all sensors; (ii) MOX sensors, and (iii) Temperature and Humidity sensors.

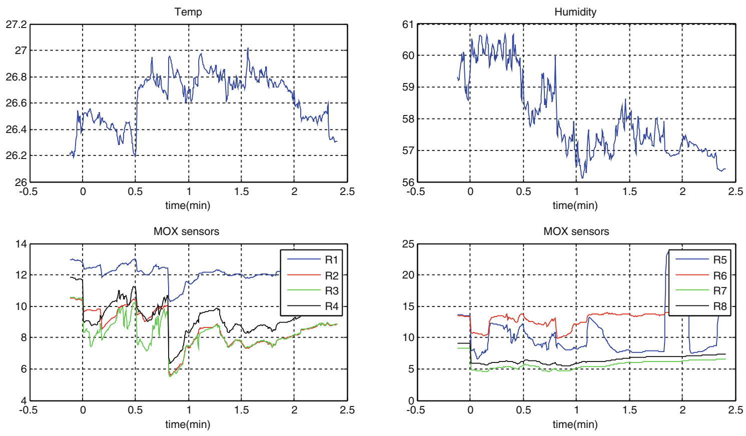

Fig. 1. Data visualisation (MOX, Temperature and Humidity sensor samples)

Table 1. Data structure

Exper.	T (temp.)	H (humidity)	MOX sensor R1	…..	MOX sensor R8
day1	Tt1, Tt2….. Tt32	Ht1, Ht2….. Ht32	R1t1, R1t2….. R1t32	…..	R8t1, R8t2….. R8t32
…….	Tt1, Tt2….. Tt32	Ht1, Ht2….. Ht32	R1t1, R1t2….. R1t32	…..	R8t1, R8t2….. R8t32
day99	Tt1, Tt2….. Tt32	Ht1, Ht2….. Ht32	R1t1, R1t2….. R1t32	…..	R8t1, R8t2….. R8t32

4 Gas Discrimination – Implementation and Experimental Results

Two data analytics models for indoor gas discrimination were built- Deep Autoencoder Neural Network (DANN) and a shallow Neural Network (NN). The models were implemented in RapidMiner - open-source software environment. Training and hyper-parameter optimisation steps for both models are outlined in DANN/NN Training Algorithms. The PCA reduction of the feature space (Table 2) is a helpful step to speed up the analysis in online implementation.

Table 2. Original and PCA (95% accumulated variance) reduced feature space

Feature set	Feature space (# of features)	PCA reduced feature space (#)
Hum & Tem	64	5
MOX Sensors	256	17
H & T & MOX	320	19

DANN/NN Training Algorithms

1. *Load dataset (Retrieved operator)*
2. *Data normalization into the range (0,1] (Normalize operator)*
3. *PCA-based feature space reduction (PCA operator, results in Table 2)*
   Iterate
4. *Model hyper-parameter optimization (grid-based Optimize Operator)*
   Optimization of cost function parameters (learning rate and momentum) in ten steps sampled from a linear scale range of [0.1 0.99] , (Table 3).
5. *Cross Validation (10 folds CV)*
   If shallow NN model: single hidden layer L1 (9 units)
   or
   If DANN: two autoencoder layers, L1 (300 units) and L2 (100 units)
6. *Performance assessment (CV test based on model accuracy)*

Table 3. Model optimal hyper-parameters

Feature sets	Optimisation parameters	NN hyper-parameters
Hum & Tem	learning rate = 0.278, momentum = 0.1	rho = 0.9693, eps = 0.1
MOX Sensors	learning rate = 0.1, momentum = 0.634	rho = 0.2789, eps = 0.42
H & T & MOX	learning rate = 0.723, momentum = 0.189	rho = 0.18, eps = 0.74

4.1 DANN Model

The implemented training sequence for the DANN model is schematically outlined in Fig. 2. The module structure represents the complete data analysis process, starting from data normalization, feature space reduction (PCA), model hyper-parameter optimization, training, validation and finally model performance assessment on test data. The computationally most demanding step is the hyper parameter optimization. This procedure is repeated for the following data sets:

(i) All sensors (MOX, Temperature, Humidity sensors) – data structure of 99 examples with 320 features (10 sensors × 32 readings).
(ii) MOX sensors – data structure of 99 examples with 256 features (8 sensors × 32 readings).
(iii) Temperature and Humidity sensors – data structure of 99 examples with 64 features (2 sensors × 32 readings).

4.2 Shallow NN Model

The DANN hypothesis is compared with a shallow NN model. The NN process workflow is schematically represented in Fig. 3. The procedure is similar to the DANN model, however the Neural Net module undergoes different parameter optimization technique.

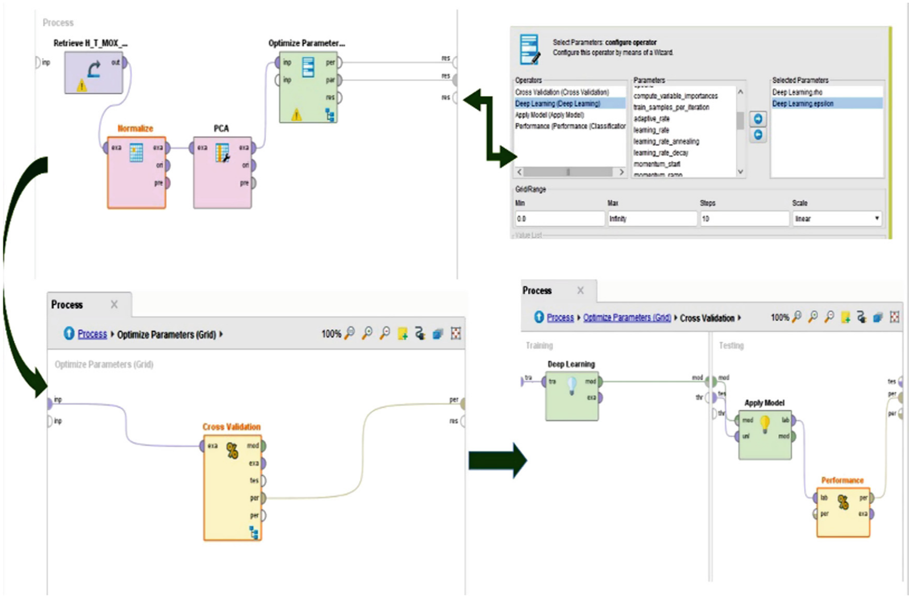

Fig. 2. DANN model fitting (RapidMiner)

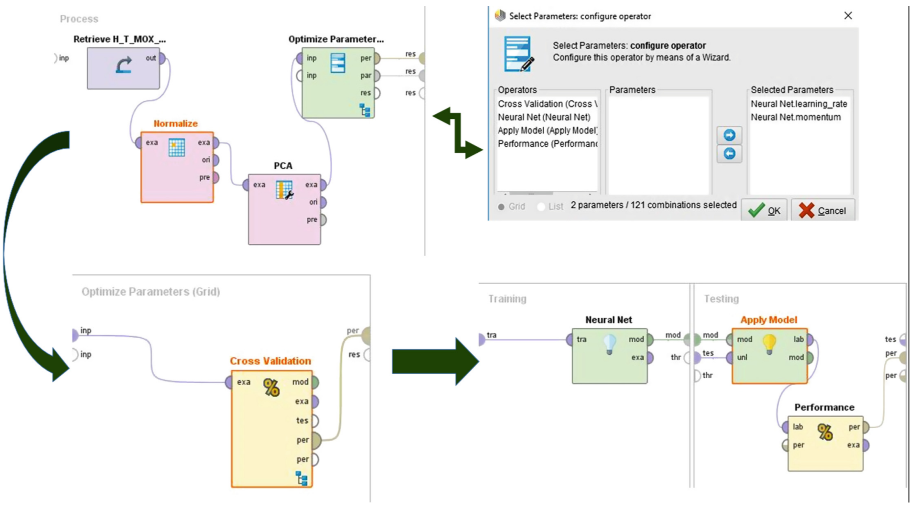

Fig. 3. Shallow NN model fitting (RapidMiner)

4.3 Experimental Results

The motivation behind the present study is to assess if the humidity and temperature sensors are sensitive to variations in the air composition and may account for changes in the air quality.

The obtained experimental results are rather promising (see Table 4). Both DANN and shallow NN gas discriminative model based on records only from two sensors (Hum & Temp) are overall more accurate in detecting unusual (banana and wine) from usual (background) gases than when the models are provided with more sensor data. Class precision and recall performance indicators (summarized in Tables 5 and 6) are also more favourable with respect to (Hum & Temp) sensor scenario. Shallow NN (with a single hidden layer) model outperforms the deep NN (two autoencoders) which is somehow expected due to the low number of features and training data.

Our results show that MOX sensors degrade the accuracy of the system. A possible explanation for this unexpected outcome may be the quality of the MOX sensor or the existence of periods of faulty (unregistered) states.

Table 4. Accuracy (%)

Sensors	NN	DANN
Hum & Tem	69.78	69.67
MOX Sensors	68.78	55.44
H & T & MOX	65.67	55.89

Table 5. Class precision (%)

Sensors	Class	NN	DANN
Hum & Tem	banana	58.06	61.76
	wine	58.33	64.52
	background	93.75	82.35
MOX Sens.	banana	68.98	48.39
	wine	65.71	58.45
	background	71.43	59.36
H & T & MOX	banana	66.67	55.56
	wine	61.11	51.11
	background	69.7	66.67

Table 6. Class recall (%)

Sensors	Class	NN	DANN
2 sensors (Hum & Tem)	banana	54.55	63.64
	wine	58.33	55.56
	background	100	93.33
8 MOX Sensors	banana	60.61	45.45
	wine	63.89	66.67
	background	71.43	53.33
10 sensors (H & T & 8 MOX)	banana	60.61	60.61
	wine	61.11	63.89
	background	76.67	40.00

5 Conclusion

Proper values of hydro-thermal parameters and good air quality are known to have a great influence on human health and comfort. Major efforts in the area of indoor air quality are focused into making our homes smart so that the healthy level of indoor air is automatically controlled.

In this paper we demonstrate that two widely available, in private houses, sensors are feasible to discriminate unusual gases from the usual house air composition. Humidity and Temperature sensors are a promising alternative to the expensive indoor air quality solutions provided with intelligent data analytics tools. Variations of the air composition due to new stimuli are encoded in trivial Hum & Temp readings and can be discriminated by a ML model trained to recognise the background home air composition.

We are aware that "wine" and "banana" are not widely accepted as typical stimuli to produce "unusual gases", however in the experimental scenario they have been selected as the new stimuli and the sensor recordings during their presence are labelled as anomaly.

For now, the proposed data analytics system is confident in binary discrimination between what has been learned as normal and abnormal home air composition. This research can be extended focusing on better recognition of various abnormal air states. Further to that considering more relevant unusual gas cases particularly those that may cause health risks is a research direction with high social impact.

Acknowledgements. This study has been done during the traineeship program of PhD student Petya Mihaylova in University of Aveiro funded by ERASMUS+EU programme for education, training, youth and sport, supported by technical University of Sofia, Bulgaria.

References

1. Mikuckas, A., et al.: A grey model approach to indoor air quality management in rooms based on real-time sensing of particles and volatile organic compounds. Appl. Math. Model. **42**, 290–299 (2017)
2. Jangid, S., Sharma, S.: An embedded system model for air quality monitoring. In: International Conference on Computing for Sustainable Global Development (INDIACom), pp. 303–308 (2016)
3. Clémenta, P., et al.: Gas discrimination using screen-printed piezoelectric cantilevers coated with carbon nanotubes. Sens. Actuators, B **237**, 1056–1065 (2016)
4. Fonollosaa, J., et al.: Human activity monitoring using gas sensor arrays. Sens. Actuators, B **199**, 398–402 (2014)
5. Szczurek, M.M., Flisowska-Wiercik, B.: Method of gas mixtures discrimination based on sensor array, temporal response and data driven approach. Talanta **83**, 916–923 (2011)
6. Penza, M., Gassano, G., Tortorella, F.: Gas recognition by activated WO_3 thin-film sensors array. Sens. Actuators, B **81**, 115–121 (2001)
7. Chen, P.-C., Ishikawa, F.N., Chang, H.-K., Ryu, K., Zhou, Ch.: A nanoelectronic nose: a hybrid nanowire/carbon nanotube sensor array with integrated micro-machined hotplates for sensitive gas discrimination. Nanotechnology **20**(12), 125503 (2009)

8. Huertaa, R., Mosqueiroa, T., Fonollosab, J., Rulkova, N.F., Rodriguez-Lujand, I.: Online decorrelation of humidity and temperature in chemical sensors for continuous monitoring. Chemo-metrics Intell. Lab. Syst. **157**, 169–176 (2016)
9. Huerta, R., Vembu, S., Amigó, J.M., Nowotny, T., Elkan, C.: Inhibition in multiclass classification. Neural Comput. **24**(9), 2473–2507 (2012)
10. Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L.: Large-scale video classification with convolutional neural networks. In: Computer Vision and Pattern Recognition (2014)
11. Krizhevsky, A., Sutskever, I.G., Hinton, E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems (2012)
12. Hinton, G., Osindero, S., The, Y.-W.: A fast learning algorithm for deep belief nets. Neural Comput. **18**, 1527–1554 (2006)
13. Bozhkov, L., Koprinkova-Hristova, P., Georgieva, P.: Learning to decode human emotions with Echo State Networks. Neural Networks **78**, 112–119 (2016)
14. Bozhkov, L., Koprinkova-Hristova, P., Georgieva, P.: Reservoir computing for emotion valence discrimination from EEG signals. Neurocomputing **231**, 28–40 (2017)
15. UCI Machine Learning Repository: Data Sets. https://archive.ics.uci.edu/ml/datasets.html