Electronic Nose Odor Classification with Advanced Decision Tree Structures

Selda GÜNEY[1], Ayten ATASOY[1], Radim BURGET[2]

[1] Dept. of Electrical and Electronics Engineering, Karadeniz Technical University, 61080, Trabzon, Turkey
[2] Dept. of Electrical Engineering and Communication, Brno University of Technology, 61200, Brno, Czech Republic

seldakul@ktu.edu.tr, ayten@ktu.edu.tr, burgetrm@feec.vutbr.cz

Abstract. *Electronic nose (e-nose) is an electronic device which can measure chemical compounds in air and consequently classify different odors. In this paper, an e-nose device consisting of 8 different gas sensors was designed and constructed. Using this device, 104 different experiments involving 11 different odor classes (moth, angelica root, rose, mint, polis, lemon, rotten egg, egg, garlic, grass, and acetone) were performed. The main contribution of this paper is the finding that using the chemical domain knowledge it is possible to train an accurate odor classification system. The domain knowledge about chemical compounds is represented by a decision tree whose nodes are composed of classifiers such as Support Vector Machines and k-Nearest Neighbor. The overall accuracy achieved with the proposed algorithm and the constructed e-nose device was 97.18 %. Training and testing data sets used in this paper are published online.*

Keywords

Electronic nose, odor classification, machine learning, data-mining.

1. Introduction

Electronic nose holds great promises for many fields of our lives. Although there are already many industrial applications, the mammalian olfactory system is much more advanced and accurate. For example, the human nose can distinguish between approximately 10,000 different odors. In this regard the e-nose systems are just in their beginnings.

Electronic nose (e-nose) devices are used in many areas of industry today. They are used, for example, in agriculture, airline transportation, cosmetics, environmental, food and beverage (food freshness, quality, ripeness and shelf-life) manufacturing, medical and clinical, military, pharmaceutical, regulatory, scientific research, and automotive applications. These devices are commonly composed of three main parts: a sensor unit, electronic unit and pattern recognition unit [1].

The sensor unit is responsible for transforming the presence of chemical compounds (molecules) into electrical signals. In the unit there are usually several sensors, where each sensor is sensitive to different compounds and therefore also produces different electrical signals by changing their properties such as conductance, capacity, voltage, tolerance to measuring weight, thermal exchange, radial exchange or measuring current with odor.

Another part of the electronic nose, the electronic unit, is a signal pre-conditioning unit, which is responsible for producing electrical signals in electronic circuits [2]. It consists of converters, filter circuit, differentiation circuit and oscillator circuit.

The third part of the electronic nose, the pattern recognition unit, takes signals obtained from the sensor unit and performs signal analysis in order to recognize and classify odor type. Although the sensors cannot cover all the chemical compounds, appropriately selected sensor types in combination with artificial intelligence can be used to detect a substance (or its smell) and/or differentiate between them.

The quality of a proposed e-nose system is commonly evaluated by six important parameters: sensor response, sensitivity, accuracy, linearity, repeatability, and reliability [3]. Today's e-nose systems still do not achieve the required parameters in many applications and there are many issues which have not been solved yet. The most important are: baseline drift, parametric drift, aging, poisoning, poor selectivity, cross sensitivity, temperature and humidity dependence. One way to overcome these problems is to use better pattern recognition algorithms. An effective pattern recognition algorithm is also a good solution to problems which occur for other reasons.

In this study, a new electronic nose device was designed and constructed for the purpose of classifying into 11 different odor classes. These classes were: moth, angelica root, rose, mint, polis, lemon, rotten egg, egg, garlic, grass, and acetone. For the purpose of gathering data a total 8 sensors were used. Using the sensors, a total of 104 different experiments were performed and stored as training and testing sets, which have been released online[1]. Using several feature extraction algorithms, different features were obtained from the raw signal. Using the

[1] Database download link: http://eee.ktu.edu.tr/personel_pages/selda_guney/database.html

domain knowledge about the chemical composition of the odors, the odors were grouped into class sets and formed into a decision tree structure. For each node of the decision tree structure an optimal classifier, its optimal parameters and optimal feature subset were selected. Many learning algorithms were examined, including the decision tree, Support Vector Machines (SVM), *k*-Nearest Neighbors (*k*-NN), Random Forest, and others. Also, a special learning algorithm was used, which is a combination of *k*-NN and SVM in the DT structure.

The main contribution of this paper is the proposal of an e-nose device and an innovative odor classification algorithm, with selected sensors for the classification task, and the finding that the highest accuracy was achieved with a new herein introduced decision tree algorithm, which is combined with SVM and *k*-NN. This learning algorithm has not been applied for the classification of odors in e-nose until now. The achieved accuracy of the proposed e-nose system with the given sensors was 97.18 % among 11 classes.

Using the proposed decision tree, it is possible to incorporate some domain knowledge in the learning algorithm and to make it much cheaper to upgrade to new sensors in future without expansive re-training on a large data set. This is particularly advantageous in the current situations, when sensor development is quite rapid and each year new and better sensors are put on the market.

The rest of the paper is structured as follows. Section 2 gives an overview of some other related works. Section 3 describes the experimental framework and the way how the data were obtained, preprocessed and how the features were extracted and used with artificial intelligence algorithms. Section 4 discusses the results achieved and Section 5 concludes the paper.

2. Related Work

In recent years, many works related to e-nose systems and odor sensing have been published and also many learning algorithms have been used in order to obtain some knowledge from e-nose systems [4].

In recent years, many works related to e-nose systems and odor sensing have emerged. The limitation of most of them is given by the very limited training sets; imperfect and relatively simple electronic sensors (especially when compared to the mammalian olfactory system) are unable to gather accurate results, there is still a big semantic gap that prevents the sensor-measured values being easily mapped to some natural gasses. Together with all this, the progress in sensor quality is relatively fast. Therefore when any database emerges, it gets old relatively fast.

The most of the recent work deals with detection of single component or relatively simple odors [5]-[14]. All these works were relatively successful and accurate. But the most of the natural odors are mixture of the simple odorant molecules, which forms so-called complex odors [15].

There are also many works dealing with the classification of complex odors [16]-[21]. For example, 7 different types of complex coffee odors are classified by several different algorithms [22]-[24]. Another example of classifying complex odors is the classification of sesame oil, maize oil and their combination [25]. In [26], fruity odors were classified into 3 classes (banana, lemon and litchi). Also, 4 different types of milk were recognized by e-nose in [27], and there are many other similar works [28], [29].

In this study, the data set was retrieved using 11 different complex odors, which is more than in other works.

3. Experiment Description

The following section describes the proposed electronic device and provides details about its components and their technical parameters. It also describes the process of data preprocessing, feature extraction and machine learning in order to automatize the process of odor classification.

3.1 Measurement System

The experimental set-up of the e-nose device consists of (1) dry air tube, (2) odor source, (3) valves, (4) mass flow controller, (5) sensor chamber, (6) DAQ, and (7) computer (see Fig. 1). In this study, eight metal oxide Figaro Gas Sensors were used. Details about the selected sensors are provided in Tab. 1. In addition to the above sensors, humidity and temperature sensors were used to measure the surrounding environment conditions.

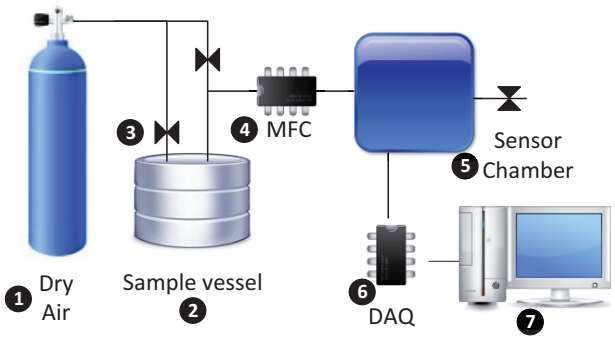

Fig. 1. Scheme of constructed electronic nose device. It consists of (1) dry air tube, (2) odor source, (3) valves, (4) mass flow controller, (5) sensor chamber, (6) DAQ and (7) computer.

All the sensors were embedded in the sensor chamber (see Fig. 2). Teflon tubes were used as an odor delivery system to avoid the effect of the previously used odors.

The data obtained by the sensor are transferred to the computer by using a DAQ 6259National Instrument with a Sampling frequency of 1 Hz. The program developed to control the whole data acquisition process and the algorithms used for data processing are implemented in Matlab [30].

Fig. 2. Sensor chamber and Figaro gas sensors.

For this study, 11 odor classes were used. The classes are: moth, angelica root, rose, mint, polis, lemon, rotten egg, egg, garlic, grass, and acetone. The reason why these odors were used in this study is to show success of the proposed algorithm for different complex odors. For each class, the experiments were repeated 9 times on average. The aim of this work is to work with a limited number of training samples, since it reflects the situation in the industry. We limited the measurements for each odor to nine-times. This number is acceptable from the viewpoint of statistical confidence of results and, at the same time, does not mean much work and significant investments for a user who will need to retrain the nose to other odors. A total of 104 measurements were performed, 63 samples (60 %) of these measurements were used for training, the remaining 41 samples (40 %) were used for testing. The same amount of each material is tested in every experiment.

The experiment stages are the pre-processing stage, odor sampling stage, breathing stage and post-processing stage. The duration of these stages is 130 s, 30 s, 30 s, and 410 s, respectively. The total experiment duration is thus 600 s and the total number of samples obtained is 600. Sensor response reaches the sensor baseline value in the pre-purging stage. In the flow delivering system, the sample odor moves from the source to the sensor chamber. The sample odor is stored in the sensor chamber for 30 s during the odor sampling stage. Thus, the value of sensor response reaches the maximum value at the breathing stage. Finally, the odor stored in the sensor chamber is cleaned in the post purging stage.

No.	*Sensor name*	*Target Gasses*
1	TGS 880	Cooking Vapors
2	TGS 2620	Organic solvents
3	TGS 825	Toxic gasses
4	TGS 2602	Indoor Pollutants
5	TGS 826	Toxic gasses
6	TGS 2104	Automobile Ventilation
7	TGS 830	Chlorofluorocarbons
8	TGS 2610	Combustible gasses

Tab. 1. Sensors used in the experiment.

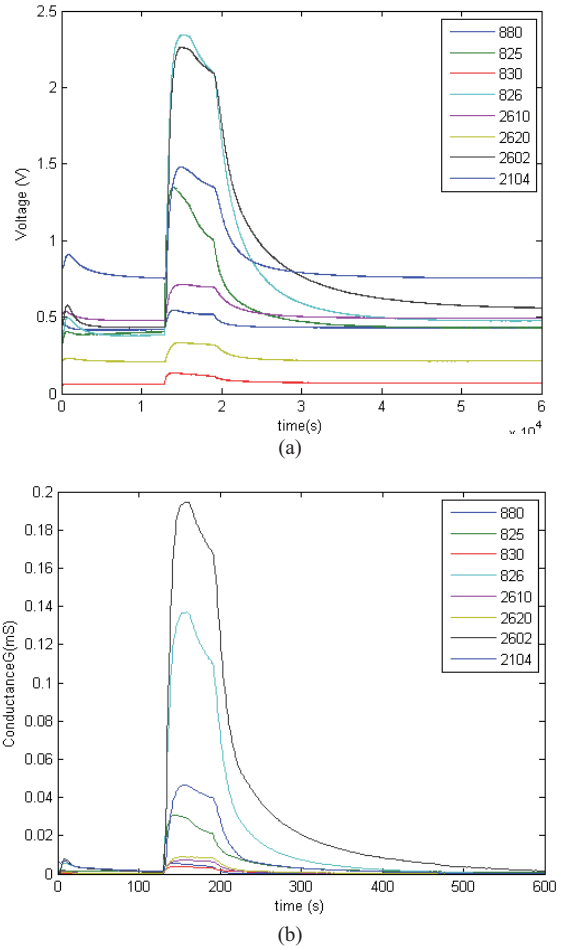

Fig. 3. Signal responses obtained from sensors: a) the raw signal b) the pre-processed signals.

3.2 Data Preprocessing

The baseline manipulation is applied to the raw signal in order to decrease sensor drift in the signal pre-processing stage [2]. Then the zero normalization is applied to the processed data. The raw signal and the pre-processed data can be seen in Fig. 3. In this study, the conductance of sensors are considered and the sensor response is given by the formula

$$G' = \frac{G - G_0}{G_0} \tag{1}$$

where G_0 and G are the beginning conductance value of sensor response and the conductance value of sensor response, respectively. The sampling frequency used in the study was 1 Hz and the measurements took 600 s. Therefore 600 data samples were obtained for each sensor. Since 8 different sensors were used, 8 × 600 = 4800 data samples were obtained for each sample odor.

3.3 Feature Extraction and Selection

As mentioned before, current sensors have many weaknesses and are far from giving absolutely accurate values. In order to improve the accuracy of classification,

not only the steady-state response of the sensor was used. For the analysis, the dynamic response of the sensor was used. This is the period that was used for the measurement and is also acceptable for many industrial applications. From these time periods, many algorithms of feature extraction were used to extract additional knowledge about the signal.

The feature extraction algorithms used in the electronic nose applications can be broadly divided into 3 groups as sub-sampling methods, parameter extraction methods and system identification methods [31]. In this study, the sub-sampling methods and parameter extraction methods were examined.

In the sub-sampling method, sensor responses are sampled for different frequencies throughout the time of odor measurement. Data reduction is done using the sub-sampling method. From the original 600 data samples each sixtieth sample was used and therefore 10 samples were selected from each sensor. For each odor, 80 (= 10 × 8 sensors) features were obtained. The experimental section revealed that better results were achieved with the sub-sampling method. Fig. 4 shows two feature responses to two different odors using sensor no. 1. The values in figures a) to j) stand for different features. For this purpose a radar plot was used. The radar plot shows feature values gathered from sensor no. 1 a) for angelica root, and b) for polish odors. It is clear from the figure that the two odors can in this case be easily distinguished using these features. Other feature extraction methods were also applied to the constructed e-nose system and they were compared with one other in the sense of the resulting classification accuracy.

The second group, the parameter extraction algorithms, extracts information regarding the transition response from each sensor. The extracted features were either maximum value or minimum value. The original signal data were filtered using differentiation filters of the first, second and third orders. From the original and the differentiated signals taken from the previous step, the following features were extracted: arithmetic mean, quadratic mean, mean between the maximum and the minimum value, geometric mean, harmonic mean, median, spectral roll-off, sum, centroid, (two highest) peak indexes and amplitudes, maximum value, minimum value, signal amplitude and linear coefficients, gradients and interception points of a piecewise linear smoothing number of zero-crossings (i.e. 4 × 20 = 80 features for each sensor).

Of course, not all the features were used for the resulting classification system and only a subset of features was used. The feature selection process is described in Section 3.5.

3.4 Data Analysis and Machine Learning

For the data obtained, many algorithms were examined in order to find a classification model with the highest accuracy. Besides the well-known algorithms such as *k*-NN, random forest, decision tree, and SVM, a newly proposed algorithm was also used, which combines other learning algorithms into a decision tree structure.

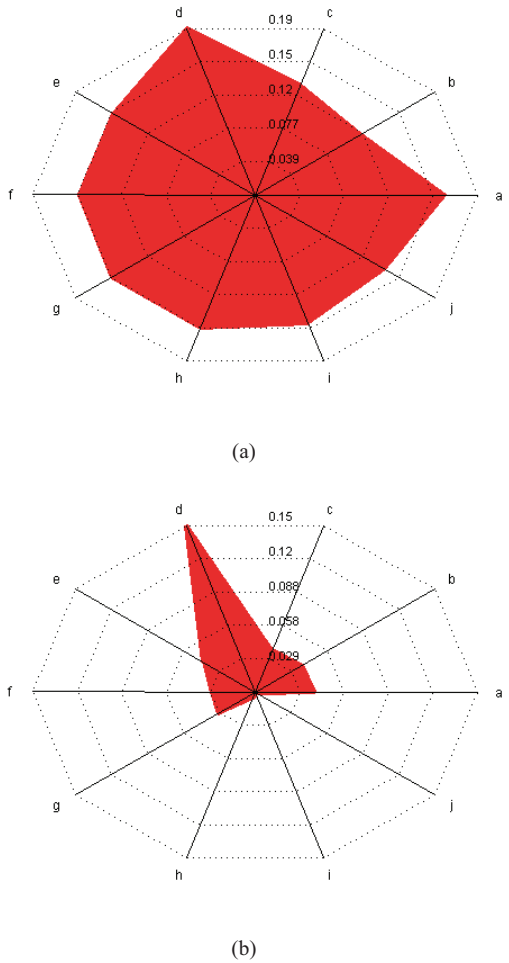

Fig. 4. Radar plots of: a) features extracted from angelica root for sensor no. 1, b) features extracted from polish for sensor no. 1.

The *k*-NN algorithm is one of the commonly used classification algorithms, which determines the class of a new observation according to the distance from the training set.

Random forest [32] is a classifier which consists of many different decision trees and applies a bagging meta-learning concept to these decision trees. The features of the decision trees are a randomly selected subset of features. This algorithm is often successful with data with low signal-to-noise ratio.

SVM method classifies the data with the help of the linear or non-linear kernel function, which tries to separate (commonly two) classes by hyper-planes [33]. For the purpose of this work, poly-nominal classification is needed [33]-[35]. In the OVA-SVM method, the number of classifiers is selected as the number of classes. Every classifier separates one class from another. Accordingly, the method is used to identify the class to which the data belongs. A disadvantage of this approach is the higher training time because of the fact that all the classifiers are

trained with a complete data set. In OVO-SVM, if the class number is n, the classifier number is designated as $n(n-1)/2$. Every classifier separates data into two classes. Whichever class suggestion is the highest, the class of classified data is assumed to belong to this class. In case the number of class increases, the number of classifiers which will be designed greatly increases. In this study, 55 classifiers are needed for the separation of 11 classes.

In industrial applications it is often necessary to build a classifier on a limited training set. At the same time, the data obtained from the sensors contain a relatively low signal-to-noise ratio. The presence of noise in odor sensing together with the limited training set makes the classification more complicated. Especially in medical applications, it becomes a significant problem.

For this reason, a new algorithm was proposed in order to increase the tolerance to nonlinear behavior of the signal on the one hand, and to increase accuracy even on limited data set on the other hand. In this study, the proposed structure combines the k-NN and SVM algorithms into a decision tree structure. The success rate of the proposed classifier increases with the proposed decision tree structure. The reason is supposed because of combining two different approaches – the method robust to noisy environments (decision tree) and method able to accurately divide classes even with limited training samples.

In order to effectively utilize most of the data, the cross-validation (CV) statistical technique was used to find the best parameters of models and data attributes. In particular, the leave-one-out cross validation (LOO-CV) technique was chosen in this study. Using LOO-CV, a classifier is trained using all training samples except one sample in each loop. This approach was used with the training data set to find the optimal parameters of each learning algorithm and to select the optimal subset of attributes from the data. This process is performed on the training data set. To prevent over-fitting the results, the trained model is validated on the validation set.

First, the odors are separated into two groups by the proposed algorithm. After examining the data, the sensor responses of odors which are similar to each other are taken as one group. In this study, three of odor samples are similar to each other. So the other odor samples are similar to each other. These similarities can be seen on sensor responses (see Fig. 5). This tree structure was first introduced in [2]. Originally [2] the tree structure was used for the classification of different odor concentration values. In this paper, the utilization of the structure is expanded to distinguish between different complex odors.

Thus the classifier first separates the data into two groups. Each classifier in every node of the decision tree structure separates the data into two groups until each node contains only one group. 10 classifiers are used for classification among 11 classes (see Fig. 6).

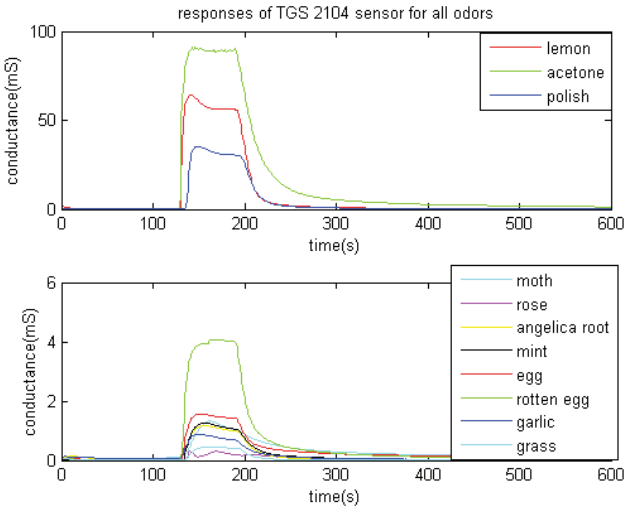

Fig. 5. The response signals of the TGS 2104 sensors to lemon, acetone, polish, moth, rose, angelica root, mint, egg, rotten egg, garlic and grass.

The proposed decision tree structure was inspired by the work published in [2] and can be summarized as follows.

Step 1: The algorithm starts at node 1 (see Fig. 6) with all the 11 possible odor classes. IDs of the classes are assigned as follows: 1 = moth, 2 = angelica root, 3 = rose, 4 = mint, 5 = polis, 6 = lemon, 7 = rotten egg, 8 = egg, 9 = garlic, 10 = grass, and 11 = acetone.

Step 2: In each node of the tree, the odor classes are divided into two sub-groups. These divisions were constructed based on the knowledge of the chemical compositions of substances in each class. For example, node ID = 2, where 3 possible classes form the input (class 1 = moth, 2 = angelica root and 3 = rose) are divided into groups {1, 2} and leaf {3}. It is clear that the leaf (containing only a single class) is not split any more. Similarly, this is repeated for the other nodes until each class has its own leaf (see Fig. 6). In each node, SVM was used as the classifier.

Step 3: For each node of the tree, an optimal parameter and optimal feature set were selected. These parameters were determined using the training set using the LOO-CV technique with accuracy measure, i.e. the number of true positives and negatives divided by the number of all training samples. Only the best result achieved was reported in this study.

Step 4: Several algorithms were used in this study, including SVM with different kernel types: k-NN with different distance measures, Decision tree, Random Forest, and others. Details about model selection, optimization and feature selection are described in section 3.5.

Step 5: Using the optimal classifiers found in steps 3 and 4, the resulting classifier is constructed (see Fig. 6 and Tabs. 4-6).

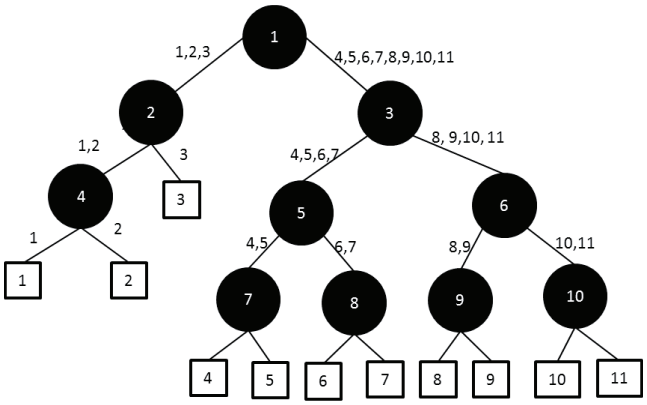

Fig. 6. The proposed learning model, which combines decision tree and Support Vector Machine classifier at its nodes. Leaf IDs stand for a particular odor, node numbers are identifications of nodes and comments of paths stand for possible odors in a given path.

3.5 Model Optimization

Model selection and optimization and feature selection (that was mentioned in the previous section, algorithm steps 3 and 4) were performed as follows. All the training and model evaluation were trained on the training set with the use of LOO-CV. The evaluation criterion was accuracy A:

$$A = \frac{N_{\mathrm{TP}} + N_{\mathrm{TN}}}{N} \tag{2}$$

where N_{TP} stands for the number of true positively classified samples, N_{TN} stands for the true negatively classified samples, and N stands for the total number of samples (in the case of the training set it is 63).

At each tree node described in the previous section, different learning algorithms were examined for the odor group classification. The experiments include the k-NN algorithm, where the k parameter ranged from 1 to 7, using only odd numbers, and numerical vector distance measures including Euclidean, Manhattan, Canberra, Chebyshev, correlation, dice, cosine and overlap measures. Another learning algorithm examined was SVM. Different kernel types were experimented with, including dot, radial (RBF), linear, poly-nominal with degree 2 and 3 and Gaussian kernels. The complexity parameter C was examined in the range $C \in \langle 0.0;\ 100.0 \rangle$ and the ε parameter in the range $\varepsilon \in \langle 0.001;\ 0.01 \rangle$. Other learning algorithms were Decision Tree and Random Forest, where their structure was determined using: gain ratio, information gain, GINI index and accuracy criteria. Experiments were conducted with different confidence thresholds t ranging in the interval $t \in \langle 0.01;\ 0.25 \rangle$ and minimal gain g ranging in $g \in \langle 0.01;\ 0.1 \rangle$. As expected, neither the Decision Tree nor the Random Forest performed well in the case of limited training data sets. Neural networks were examined using different numbers of neurons in the hidden layer, different learning rates and different numbers of training cycles and momentums.

Other learning models were also examined, for example Naïve Bayes, Linear regression, and some meta-learning algorithms such as Bagging, Stacking, Boosting, and others. No significant results were achieved with this study.

The parameters of the learning algorithms were optimized using the genetic algorithm (size of population 15, crossover method: uniform, selection method: tournament with elitism). In the case of determining k in the k-NN algorithm the grid search method was used.

For each algorithm, the sequential forward feature selection (SFFS) algorithm with 2 additional speculative rounds was performed. The algorithm selects a subset of features from the data set that best classifies the data by sequentially selecting and adding features until there is no improvement in classification accuracy. In addition, two speculative rounds were performed before the algorithm was stopped, to reduce the risk of premature SFFS termination. In case the algorithm gave the same accuracy for two subsets, the one with fewer features was preferred because of lesser computational demand. The experiments confirmed that using all the features for learning algorithms gave worse classification results. As a result, the SFSS algorithm reduced redundancy in the feature set, saved a portion of computational requirements and increased the accuracy of a particular classifier.

Results of the optimization part are shown in Tab. 4, where node IDs correspond to the tree depicted in Fig. 6.

4. Results and Discussion

The resulting model, which was trained as described in the previous section, was validated using a test set containing 41 measurements. The test set is independent of the training set.

In general, chemical compounds of odors are described well in the literature these days and therefore it is not necessary to retrieve lengthy training sets for odor classifier training. With the chemical domain knowledge combined with machine learning algorithms (as described in this paper), the trained model can achieve a relatively high accuracy even with limited training sets. The benefit here is clear since data labeling and data acquisition are quite a lengthy and arduous process for man. The feature extraction was done using two types of feature extraction: sub-sampling and the parameter extraction method.

First, data obtained by the sub-sampling method was used. The results retrieved using the algorithm described in section 3.4 and the optimization process using sub-sampling feature extraction method in section 3.5 are shown in Tab. 3. The table shows the best learning algorithm chosen for a particular tree node (see Fig. 6) and selected features. Comparisons of the proposed decision tree with other general purpose learning algorithms (SVM, k-NN, etc.) are shown in Tab. 4. Details about the results

Node ID	1	2	3	4	5	6	7	8	9	10
Feature 1	74	14	22	4	11	1	3	1	2	1
Feature 2	-	-	34	-	36	17	11	-	3	2
Feature 3	-	-	-	-	64	-	-	-	15	14
Classification algorithm	SVM	*k*-NN	SVM	*k*-NN	SVM	SVM	SVM	*k*-NN	*k*-NN	SVM

Tab. 3. Features selected from extracted data by using the sub-sampling method and the classification algorithms used for every node in the proposed decision tree structure.

Classification Algorithm	Parameters	Used Features	Success (%)
Decision tree	default	All	76.19
Decision tree	default	Reduced set	88.90
k-NN	$k = 1$, Euclidean distance	34,64,75	93.65
k-NN	$k = 1$, Chebyshev distance	5, 26, 34, 75	93.65
k-NN	$k = 1$, Cosine similarity measure	4, 34, 64, 74, 75	96.83
k-NN	$k = 1$, Dynamic Time Warping	14, 34, 74, 75	92.06
k-NN	$k = 1$, Manhattan distance	14, 34, 74, 75	93.65
k-NN	$k = 3$, Manhattan Distance	All	96.83
Random Forests	$I = 5$	All	92.06
Proposed Decision tree	-	See Tab. 3	97.18

Tab. 4. The accuracy of classifiers using the features selected from extracted data by using the sub-sampling method.

Classification algorithm	Parameters	Features	Success rate (%)
Random Forest	$N = 10$,	All	80.95
k-means	$k = 1$, Euclidean distance	5, 34, 64	88.89
SVM	RBF kernel, $c = 60$	34, 45, 66, 75	92.06
Proposed Decision Tree	-	Tab. 6	95.24

Tab. 5. Accuracy of classifiers with using selected features. The feature extraction is described by Section 3.3.

Classifier	1	2	3	4	5	6	7	8	9	10
Feature 1	1	12	6	169	9	2	1	1	1	1
Feature 2	19	30	30	-	12	53	3	2	93	12
Feature 3	53	-	53	-	73	71	12	19	-	-
Feature 4	81	-	-	-	162	-	-	70	-	-
Classification algorithm	*k*-NN	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM

Tab. 6. Features selected from extracted data by using the parameter extraction method and the used classification algorithms for every node of the proposed decision tree structure.

of classification algorithms using the testing set are shown in Tab. 4. In Tab. 4 a) the algorithm used, b) selected features and c) resulting accuracy are shown.

The same experiment as described above was performed using the parameter extraction method for feature extraction. Again, the algorithm described in Section 3.4 and its optimization is described in Section 3.5 were performed. The trained decision tree model is shown in Tab. 5. Its comparison to other general learning algorithms is shown in Tab. 6.

5. Conclusion

In this study, an electronic nose system was designed and constructed for the purpose of classifying into 11 different odor classes. These classes are: moth, angelica root, rose, mint, polis, lemon, rotten egg, egg, garlic, grass, and acetone. For this purpose, a total of 8 sensors were selected and used to obtain a training set numbering 104 different experiments. With the use of the chemical domain knowledge about the odors, a decision tree structure was proposed (see Fig. 6). Chemical compounds of odors are relatively well described in the literature these days. Classifiers at each particular node of the tree were trained and optimized using a training set (60 % of the obtained data set) and validated using the rest of the data set (40 % of the obtained data set). The data set used in this study was released online.

Using the proposed decision tree, it is possible to incorporate some domain knowledge in the learning algorithm and to make it much cheaper to upgrade to new sensors in future without expansive re-training on a large data set. This is particularly advantageous in the current situations, when sensor development is quite rapid and each year new and better sensors are put on the market.

The main contribution of this paper is constructing an e-nose device with selected sensors for the complex odor classification task, and the finding that the highest accuracy was achieved with utilizing domain knowledge about odors. Since the development of sensors is quite rapid it can save significant resources and make it much cheaper to upgrade to a newer sensors in future.

This learning algorithm has not been applied to the classification of odors in e-nose until now. The achieved accuracy of the proposed e-nose system with the given sensors was 97.18 % among 11 complex classes.

Acknowledgements

This study was funded by project No: 2009.112.004.3, Research Projects Unit at Karadeniz Technical University and by projects SIX CZ.1.05/2.1.00/03.0072, EU ECOP EE.2.3.20.0094, CZ.1.07/2.2.00/28.0062, and MPO FR-TI4/151, Czech Republic.

References

[1] GUTIERREZ-OSUNA, R., NAGLE, H. T., KERMANI, B., SCHIFFMAN, S. S. Pattern analysis for electronic noses. In PEARCE, T. C., SCHIFFMAN, S. S., NAGLE, H. T., GARDNER, J. W. (Eds.), *Handbook of Machine Olfaction: Electronic Nose Technology*. Weinheim (Germany): Wiley, 2003, p. 133 - 160.

[2] GUNEY, S., ATASOY, A. Multiclass classification of n-butanol concentrations with k-nearest neighbor algorithm and support vector machine in an electronic nose. *Sensors and Actuators B: Chemical*, 2012, vol. 166-167, p. 721 - 725.

[3] GARDNER, J. W., YINON, J. *Electronic Noses and Sensors for the Detection of Explosives*. Dordrecht (Netherlands): Kluwer, 2004.

[4] HINES, E. L., LLOBET, E., GARDNER, J. W. Electronic noses: A review of signal processing techniques. *IEE Proceedings – Circuits, Devices and Systems*, 1999, vol. 146, no. 6, p. 297 - 310.

[5] SRIVASTAVA, A. K., SHUKLA, K. K., SRIVASTAVA, S. K. Exploring neuro-genetic processing of electronic nose data. *Microelectronics Journal*, 1998, vol. 29, no. 11, p. 921 - 931.

[6] KERMANI, B. G., SCHIFFMAN, S. S., NAGLE, H. T. Using neural networks and genetic algorithms to enhance performance in an electronic nose. *IEEE Transactions on Biomedical Engineering*, 1999, vol. 46, no. 4, p. 429 - 439.

[7] CHO, J. H., KURUP, P. U. Decision tree approach for classification and dimensionality reduction of electronic nose. *Sensors and Actuators B: Chemical*, 2011, vol. 160, p. 542 - 548.

[8] KHALAF, W., PACE, C., GAUDIOSO, M. Least square regression method for estimating gas concentration in an electronic nose system. *Sensors*, 2009. vol. 9, no. 3, p. 1678 - 1691.

[9] GAUDIOSO, M., KHALAF, W. On the use of the SVM approach in analyzing an electronic nose. In *Proceedings of the 7th International Conference on Hybrid Intelligent Systems*. Kaiserslautern (Germany), 2007.

[10] ZHANG, S., XIE, C., HU, M., LI, H., BAI, Z., ZENG, D. An entire feature extraction method of metal oxide gas sensors. *Sensors and Actuators B: Chemical*, 2008, vol. 132, p. 81 - 89.

[11] ZHANG, S., XIA, X., XIE, C., CAI, S., LI, H., ZENG, D. A method of feature extraction on recovery curves for fast recognition application with metal oxide gas sensor array. *IEEE Sensors Journal*, 2009, vol. 9, no. 12, p. 1705 - 1710.

[12] MUEZZINOGLU, M. K., VERGARA, A., HUERTA, R. A unified framework for volatile organic compound classification and regression. In *The 2010 International Joint Conference on Neural Networks (IJCNN)*. Barcelona (Spain), 2010.

[13] CHO, J. H., KURUP, P. U. Decision tree approach for classification and dimensionality reduction of electronic nose data. *Sensors and Actuators: B Chemical*, 2011, vol. 160, p. 542 - 548.

[14] GÜNEY, S., ATASOY, A. Classification of n-butanol concentrations with k-NN algorithm and ANN in electronic nose. In *Proceedings of Innovations in Intelligent Systems and Applications*. Istanbul (Turkey), 2011.

[15] DUTTA, R., HINES, E. L., GARDNER, J. W., KASHWAN, K. R., BHUYAN, M. Tea quality prediction using a tin oxide-based electronic nose: an artificial intelligence approach. *Sensors and Actuators B:Chemical*, 2003, vol. 94, p. 228 - 237.

[16] DEBSKA, B., SWIDER, B. G. Decision trees in selection of featured determined food quality. *Analytica Chimica Acta*, 2011, vol. 705, no. 1-2, p. 261 - 271.

[17] WANG, D., WANG, X., LIU, T., LIU, Y. Prediction of total viable counts on chilled pork using an electronic nose combined with support vector machine. *Meat Science*, 2012, vol. 90, no. 2, p. 373 - 377.

[18] KUSKE, M., RUBIO, R., ROMAIN, A. C., NICOLAS, J., MARCO, S. Fuzzy k-NN applied to moulds detection. *Sensors and Actuators B: Chemical*, 2005, vol. 106, p. 52 - 60.

[19] XIE, L., WANG, X. Gas quantitative analysis with support vector machine. In *Proceedings of Chinese Control and Decision Conference*. Jinhua (China), 2009.

[20] TANG, K. T., LIN, Y. S., SHYU, J. M. A local weighted nearest neighbor algorithm and a weighted and constrained least-squared method for mixed odor analysis by electronic nose systems. *Sensors*, 2010, vol. 10, no. 11, p. 10467 - 10483.

[21] BRUDZEWSKI, K., OSOWSKI, S., MARKIEWICZ, T., ULACZYK, J. Classification of gasoline with supplement of bio-products by means of an electronic nose and SVM neural network. *Sensors and Actuators B*, 2006, vol. 113, p. 135 - 141.

[22] PARDO, M., SBERVEGLIERI, G. Classification of electronic nose data with support vector machines. *Sensors and Actuators B*, 2005, vol. 107, p. 730 - 737.

[23] JHA, S. K., YADAVA, R. D. S. Power scaling of chemiresistive sensor array data for odor classification. *Journal of Pattern Recognition Research*, 2011, vol. 6, no. 1, p. 65 - 74.

[24] WANG, X., CHANG, J., WANG, K., YE, M. Least square support vector machines in combination with principal component analysis for electronic nose data classification. In *Proceedings of 2nd International Symposium on Information Science and Engineering*. Shanghai (China), 2009.

[25] HAI, Z., WANG, J. Electronic nose and data analysis for detection of maize oil adulteration in sesame oil. *Sensors and Actuators B: Chemical*, 2006, vol. 119, p. 449 - 455.

[26] TANG, K. T., CHIU, S. W., PAN, C. H., HSIEH, H. Y., LIANG, Y. S., LIU, S. C. Development of a portable electronic nose system for the detection and classification of fruity odors. *Sensors*, 2010, vol. 10, p. 9179 - 9193.

[27] BRUDZEWSKI, K., OSOWSKI, S., MARKIEWICZ, T. Classification of milk by means of an electronic nose and SVM neural network. *Sensors and Actuators B*, 2004, vol. 98, p. 291 - 298.

[28] ZHANG, L., TIAN, F., NIE, H., DANG, L., LI, G., YE, Q., KADRI, C. Classification of multiple indoor air contaminants by an electronic nose and a hybrid support vector machine. *Sensors and Actuators B*, 2012, vol. 174, p. 114 - 125.

[29] PAPADOPOULOU, O. S., PANAGOU, E. Z., MOHAREB, F. R., NYCHAS, G.-J. E. Sensory and microbiological quality assessment of beef fillets using a portable electronic nose in tandem with support vector machine analysis. *Sensors*, 2008, vol. 8, p. 142 - 156.

[30] GUNEY, S., ATASOY, A. An electronic nose system for assessing horse mackerel freshness. In *International Symposium*

on Innovations in Intelligent Systems and Applications (INISTA). Trabzon (Turkey), 2012.

[31] GUTIERREZ-OSUNA, R., NAGLE, H. T. A method for evaluating data-preprocessing techniques for odor classification with an array of gas sensors. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 1999, vol. 29, no. 5, p. 626 - 632.

[32] BREIMAN, L. Random forests. *Machine Learning*, 2001, vol. 45, p. 5 - 32.

[33] MATHUR, A., FOODY, G. M. Multiclass and binary SVM classification: Implications for training and classification users. *IEEE Geoscience and Remote Sensing Letters*, 2008, vol. 5, p. 241 - 245.

[34] HSU, C. W., LIN, C. J. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 2002, vol. 13, no. 2, p. 415 - 425.

About Authors ...

Selda GÜNEY received the BSc degree in Electrical and Electronics Engineering in 2004 and the MSc degree in 2007 from Karadeniz Technical University of Trabzon, Turkey, where she is currently working toward the Ph.D. degree at the Department of Electrical and Electronics Engineering. Her research interests include pole placement, control systems, pattern recognition, and electronic noses.

Ayten ATASOY received the BSc, MSc, and PhD degrees, all in Electronics Engineering, from Karadeniz Technical University, in Trabzon-Turkey, in 1987, 1992, and 1997, respectively. She is currently an assistant professor with the Department of Electrical and Electronics Engineering, Karadeniz Technical University, Trabzon, Turkey. Her research interests include system design and analysis, system identification and estimation, model reduction, control systems, pole placement, biomedical signal processing and analysis, pattern recognition, and electronic noses.

Radim BURGET is currently assistant professor with the Department of Electrical Engineering and Communication, Brno University of Technology. He received the MSc degree in 2006 and Ph.D. in 2010 from Brno University of Technology, Czech Republic. His research interests include data-mining and image mining methods.