Neural Network Ensembles for Online Gas Concentration Estimation Using an Electronic Nose

Chaibou Kadri[1], Fengchun Tian[1], Lei Zhang[1], Guojun Li[2], Lijun Dang[1], and Guorui Li [1]

[1]College of Communication Engineering, Chongqing University
ShaZheng street 174, ShaPingBa district, Chongqing 400044, China

[2]Chongqing Communication Institute
Chongqing, 400035, China

Abstract

Ensemble method is a learning paradigm that has been shown to improve the performance of classical learning methods which are based on single model. However, for an ensemble method to be effective, it is essential that the base models are sufficiently accurate and error-independent (i.e. diverse) in their predictions. Moreover, ensemble integration is one of the most critical steps in ensemble learning. In this paper, a dynamic integration method is proposed and applied in electronic nose for online concentration estimation of some indoor air pollutants namely formaldehyde, benzene, toluene, and carbon monoxide. For comparison purpose, other integration methods were also evaluated. Experimental results show that this method is attractive, and with additional improvement it can be a good alternative for online air quality monitoring using electronic nose systems.

Keywords: *electronic nose, neural network ensembles, dynamic integration method, online monitoring*

1. Introduction

Ensemble learning is an attractive research field in various research communities. In ensemble learning, instead of using one model for prediction, several models are employed. This learning method has been found to improve the performance of single models, provided that the base models are sufficiently accurate and diverse in their predictions [1,2]. It consists of two main phases: ensemble generation and ensemble integration. Homogeneous learning (i.e. the base models are trained using the same learning algorithm) and heterogeneous learning (i.e. the base models are trained using different learning algorithms) are the two main ensemble generation methods. With regard to the second method, due to high computational complexity associated with combining models derived from different algorithms, less
research has been done in this area. As for the ensemble integration, one of these two approaches is generally used: combination approach where the predictions of the base models are combined using some rules to obtain the ensemble prediction, and the selection approach where the prediction of one or more good model(s) is (are) selected based on some criteria to form the final prediction. It is worth mentioning that both combination and selection can be either static or dynamic. In static approach, the trained models do not change, whereas in dynamic approach the prediction strategy is adjusted based on some rules for each test sample (pattern).

Although there has been much research work on ensemble learning, less research has been carried out on its application in E-nose signal processing. In [3,4] Gao. et al., and Gao and Wei used ensemble method to predict simultaneously both the classes and concentrations of several kinds of odors. In the first approach, they used specifically MLP networks as base models, whereas in the second approach each ensemble (therein referred as panel) is a hybridization of four base models namely MLP, MVLR, QMVLR, and SVM. In [5] Shi, et al. used heterogeneous classifiers including density models, KNN, ANN and SVM for odor discrimination. In another study, Hirayama et al.[6], demonstrated that it was possible to detect liquid petrol gas (LPG) calorific power with high recognition rate (up to 99%) using an E-nose and a committee of machines, even with the failure of one random sensor, or when a sensor loses its sensitivity to the target gas. Vergara et al. [7] proposed an ensemble method which uses support vector machines as base classifiers to cope with the problem of drift in chemical gas sensors. Very recently, A. Amini et al. [8] used an ensemble of classifiers on data from a single metal oxide gas sensor (SP3-AQ2, FIS Inc., Japan) operated at six different rectangular heating voltage pulses (temperature modulation), to identify three gas analytes viz. methanol, ethanol and 1-butanol, at range of 100-2000 ppm. Among these prior studies, only two [3,4] focused on concentration estimation (i.e. regression context). Moreover, simple averaging combination method, a static integration approach, was used. Therefore, investigation of dynamic integration approaches is of critical importance. In this paper, a new dynamic integration method called dynamic weighting of base models DWBM is proposed

and compared with two other integration methods: basic ensemble method (BEM) and the dynamic selection (DS).

2. Datasets generation

To generate the datasets used in this paper, several experiments were carried out using a self-made electronic nose. Detailed description of the experimental setup and procedure can be found in our previous publications [9,10]. However, to make the paper self-contained, we reproduce the experimental setup (see Fig. 1).As for the experimental procedure it is worth noting that during all the experiments, the respective ranges of the temperature and humidity were 15-45℃ and 25-80%. Also, a single experiment consists of three phases: exposure to clean air for 120s to stabilize the sensors, exposure to gas analyte for 480s, and another exposure to clean air for 120s to allow the sensors recover. Between any two consecutive experiments, the chamber is cleaned for about 10mins to avoid (minimize) interference from any chemical remnant. Last but not least, a flash memory is used for real-time data storage. It is worth mentioning that the real concentration of benzene was determined by gas chromatography method; while that of formaldehyde was determined using two different methods: acetylacetone spectrophotometric method for concentrations greater than 0.5ppm, and the 3-methyl-2-benzothiazolinone hydrochloride (MBTH) method for concentrations less than 0.5ppm. This is the aim of using organic gas sampler. For the other gases, standard measurement equipments were placed inside the chamber and the displayed concentrations were recorded. The sampling rate during data acquisition is one point every three seconds. This data can be transferred to a personal computer (PC) using NiosII IDE and the Joint Test Action Group (JTAG) cable, for further processing. The number of measurements for formaldehyde, benzene, toluene, and carbon monoxide is 126, 72, 66, and 58, respectively. Therefore, for each gas analyte, an original dataset is obtained, which is referred as raw measurements from the sensor array.

The raw measurements from the sensor array are first filtered (to remove measurement noise) prior to feature extraction. For gas concentration estimation, Szczurek et al. [11] demonstrated that features from the steady-state portion of a gas sensor response are more informative. Taking this into account, we selected one feature from that portion (see Fig. 2). For the auxiliary sensors (temperature, humidity) we selected features at the same time positions with other sensors. The extracted features are normalized to have values in the interval [0, 1]. Having an array of eight sensors, an $8\times m$ (m is the number of measurements or samples) feature data matrix is formed for each dataset. Then we used Kennard and Stone (K-S) algorithm [12] to divide each dataset into three sub-datasets: 40% for training, 40% for validation, and 20% for test.

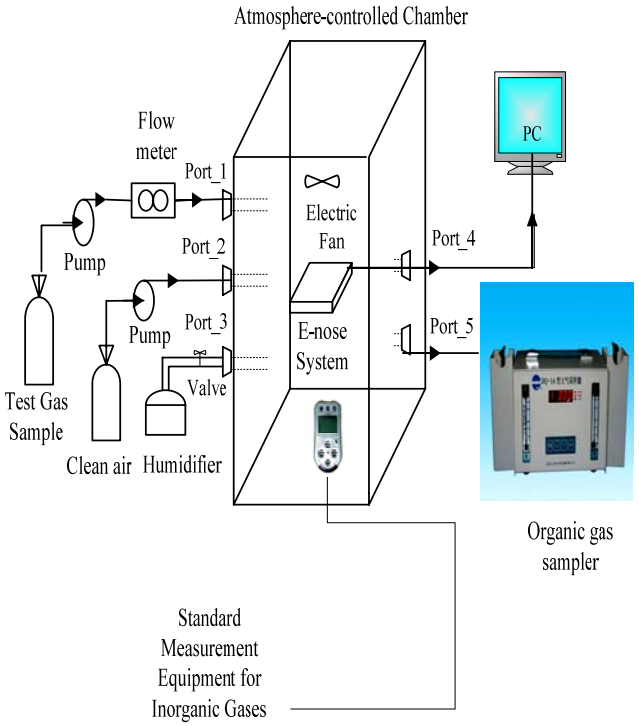

Fig. 1 Experimental setup

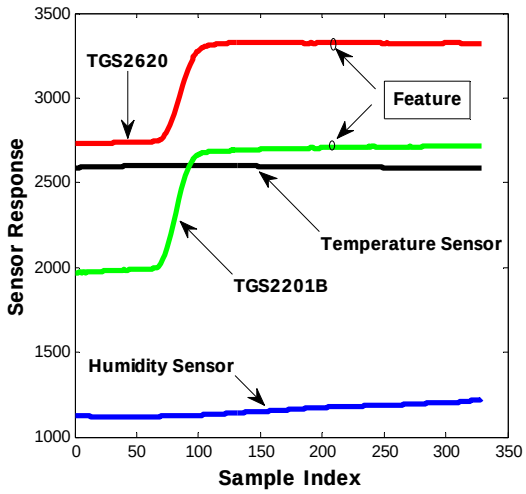

Fig. 2 Responses of four sensors

3. Methodology

3.1 Ensemble generation methods

The main objective of any ensemble generation method is to build sufficiently accurate and diverse base models. Ensemble generation methods can be grouped into two categories: methods based on the modification of the learning dataset, and methods based on the modification of the training algorithm. The first category includes re-sampling methods such as bagging [13], and adaptive boosting [14], among others. The second category includes negative correlation learning [15], and ADDEMUP [16], among others. If the ensemble is based on neural network models, a simple yet often effective method to generate diverse base models is to train each base model with different initial random weights [1], or a different topology [17]. In this paper we combine these two alternatives. More specifically, five base models were trained, each with different initial weights and different topology. Classical Levenberg Marquardt back-propagation algorithm was used during the training process of each base model. For detailed discussion on neural networks in general and back propagation algorithm in particular, we refer the reader to [18]. To ascertain that the base models are relatively accurate, early-stopping option on validation dataset is used. Furthermore, only models with validation errors less than or equal to a preset threshold are kept. As for the topology, it was changed by varying the number of hidden neurons from five to twenty five with an increment of five neurons. Fig. 3 shows the topology of one the base models used, with eight input nodes (the number of sensors), five hidden neurons, and one output neuron.

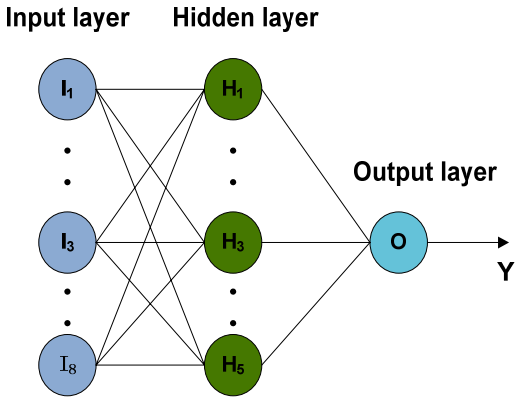

Fig. 3 Topology of one base model

It is worth mentioning that, in this paper the mean absolute relative error (MARE) is used as performance measure, on both the validation and test datasets. More formally, for an m samples datasets, the MARE is defined as follows:

$$\mathrm{MARE} = 100 \times \frac{1}{m} \sum_{i=1}^{m} \left(\frac{y_i - \hat{y}_i}{y_i} \right) \quad (1)$$

where m is the number of samples, y_i is the real concentration, and $\hat{y}_i$ is the predicted concentration.

3.2 Ensemble integration methods

Basic ensemble method (BEM) and generalized ensemble method (GEM) were the first approaches to ensemble integration for regression [1]. In BEM the output of the ensemble is obtained by averaging the outputs of all the base models; whereas in GEM, the output of the ensemble is obtained through a weighted sum of the outputs of the base models. More formally, given a set of m base models, the output of the ensemble on a novel sample x can be expressed as:

$$f_{ens}(x) = \sum_{i=1}^{m} w_i f_i(x) \quad (2)$$

where w_i is the weight assigned to the i^{th} model. There are many ways to find the values of w_i. The simplest way is to set them to the same value, which is known as simple averaging or BEM. Another alternative to find the weights is by minimizing the mean square error of the ensemble on either the training dataset or a set of validation data [1].

Another approach to ensemble integration is through stacking where a meta-model is trained to learn how to combine/select the outputs from the base models [19]. In stacking, level zero models (base models) produce meta-data consisting of the target value and the base models' predictions, obtained using cross-validation on training data. A linear or non-linear regression algorithm is then used to build the meta-model based on the meta-data. Consequently, this method is applicable in situations where large size datasets are available, so we did not consider it.

Recently, dynamic integration techniques have shown to be promising, especially in the context of classification [20]. In dynamic integration methods, local accuracy is used to either select one or many base models from a pool of base models, or to combine some of or all the base models using some weighting rules. More specifically, the prediction of a novel sample is obtained based on the prediction(s) of similar sample(s) in either the training or the validation dataset. The standard method for obtaining similar data in the context of ensemble learning is the well known k-nearest neighbors (k-NN) algorithm with Euclidean distance [21, 22]. In this paper, two distance metrics are considered: the Euclidean distance and the cosine distance.

Given an *mx*-by-*n* data matrix **X**, which is treated as mx (1-by-*n*) row vectors $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{mx}$, and *my*-by-*n* data matrix **Y**, which is treated as *my*(1-by-*n*) row vectors $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{my}$, the Euclidean and cosine distances between the vector x_s and y_t are respectively defined as:

$$Ed_{st} = \sqrt{\left(\mathbf{x}_s - \mathbf{y}_t\right)\left(\mathbf{x}_s - \mathbf{y}_t\right)^T} \tag{3}$$

where Ed_{st} is the Euclidean distance between the vector $\mathbf{x}_s$ and $\mathbf{y}_t$, $\left(\mathbf{x}_s - \mathbf{y}_t\right)^T$ is the transpose of vector $\left(\mathbf{x}_s - \mathbf{y}_t\right)$.

$$Cd_{st} = \left(1 - \frac{\mathbf{x}_s \mathbf{y}_t^T}{\sqrt{\left(\mathbf{x}_s \mathbf{x}_s^T\right)\left(\mathbf{y}_t \mathbf{y}_t^T\right)}}\right) \tag{4}$$

where Cd_{st} is the cosine distance between the vector x_s and y_t.

It is worth mentioning that the performance of k-NN algorithm is affected not only by the distance metric but also by the number of nearest neighbors, k. Therefore, for the DS and DWBM methods (note that the k-NN algorithm is not used on BEM method) the optimal value of k was found empirically using the validation dataset, for each air pollutant. Thus, the DS method selects the best base model based on its performance on k similar sample(s) in the validation dataset; while the new method the DWBM method assigns a weight to each base model based on its performance on k similar sample(s) in the validation dataset.

Drift is one of the most critical problems associated with semiconductor gas sensors which constitute the core part of our E-nose system. In order to investigate the robustness of these methods against drift, we simulated the drift by adding random Gaussian noise to the input attributes of the test dataset of each air pollutant. The noise standard deviation was set to 0.002, a value capable of causing quite sufficient data shift.

3.2.1 Calculation of the weight for each base model

The output of a model is weighted by the inverse of the errors of similar points in the validation dataset. More specifically the weight used for the i^{th} example and k^{th} model is defined as:

$$w_i^{\,k} = \frac{\dfrac{1}{\sum_{j=1}^{N} erv_{ij}^{\,k}}}{\sum_{m=1}^{M}\sum_{j=1}^{N} erv_{ij}^{\,m}} \tag{5}$$

where M is the number of base models, N is the number of nearest neighbors, $erv_{ij}^{\,m}$ is the error of the m^{th} base model on the j^{th} similar sample (nearest neighbor) in the validation dataset.

4. Empirical results and comparison analysis

All computations were carried out using MATLAB R2010a (MathWorks Inc.) software on a desktop computer with the Intel(R) Core(TM) i3 T2450 2.93 GHz CPU, 2 GB RAM and Windows XP professional operating system.

To avoid biased comparison, for every dataset five ensembles were generated using the procedure described in section 3.1. The performance of each method is evaluated on these five ensembles, and the MARE average values were computed. Furthermore, for the DS and DWBM methods, the average values of the optimal number of nearest neighbors (k) were also computed. Table 1 and Table 2 show the MARE average values of three methods when evaluated on original test dataset, using cosine distance and Euclidean distance, respectively. The average values of the optimal number of nearest neighbors (k) are also shown for the DS and DWBM methods. Table 3 and Table 4 show the MARE average values of three methods when evaluated on noisy test dataset, using cosine distance and Euclidean distance, respectively. The average values of the optimal number of nearest neighbors (k) are also shown for the DS and DWBM methods.

Table 1: MARE average values of five ensembles evaluated on original dataset, using cosine distance

Datasets	Methods		
	DS	BEM	DWBM[a]
Formaldehyde	47.87(12)	42.71	43.47(1)
Benzene	12.25(4)	10.60	10.50(1)
Toluene	13.8(17)	10.43	11.40(1)
CO	17.75(4)	19.24	18.57(1)

[a] our method, the numbers in parentheses represent the average value of k

Table 2: MARE average values of five ensembles evaluated on original dataset, using Euclidean distance

Datasets	Methods		
	DS	BEM	DWBM[a]
Formaldehyde	49.38(19)	42.71	44.10(1)
Benzene	11.74(6)	10.60	10.70(1)
Toluene	14.05(10)	10.43	11.62(1)
CO	17.75(3)	19.24	18.15(1)

[a] our method, the numbers in parentheses represent the average value of k

Table 3: MARE average values of five ensembles evaluated on noisy dataset, using cosine distance

Datasets	Methods		
	DS	BEM	DWBM[a]
Formaldehyde	51.44(6)	43.47	48.65(1)
Benzene	14.33(9)	12.14	12.07(1)
Toluene	15.82(5)	11.24	12.52(1)
CO	16.67(4)	18.21	17.66(1)

[a] our method, the numbers in parentheses represent the average value of k

Table 4: MARE average values of five ensembles evaluated on noisy dataset, using Euclidean distance

Datasets	Methods		
	DS	BEM	DWBM[a]
Formaldehyde	53.16(19)	43.47	48.02(1)
Benzene	12.97(7)	12.14	11.67(1)
Toluene	15.82(4)	11.24	12.79(1)
CO	16.67(3)	18.26	16.69(1)

[a] our method, the numbers in parentheses represent the average value of k

From results depicted in Tables 1-to-4 we can see that, regardless of the distance metric or the dataset (test), DS method requires more nearest neighbors than DWBM method which requires only one nearest neighbor. This implies that the new method has less computational complexity than the dynamic selection method. It is worth mentioning that BEM method, which does not require finding nearest neighbors, is the simplest yet effective method. Both the DWBM and BEM methods perform almost comparably over all datasets. Moreover, these two methods outperform the dynamic selection (DS) method on most of the datasets, except on carbon monoxide dataset. A sound reason for this is the ensemble built using this dataset is composed of many redundant base models, which could have adversely affected the performance of the averaging and the dynamic weighting method (new method). Indeed, both the averaging and the dynamic weighting methods are more sensitive inaccurate and/or redundant base models than the dynamic selection method. Furthermore, in [17], it was pointed out that base models (nets) trained on different training sets are more likely to possess high levels of diversity than those created through variations in their initial conditions (weights), or their topology. This finding is in line with the results obtained in this paper. To support our hypothesis, we decided to plot the predictions from all base models in each ensemble for CO validation data, as shown in Fig. 4. It can be seen that, within each ensemble, there is strong similarity among the base models. As for the distance metrics, we can see that there is no metric that is globally better. The effect of these metrics on the performance of either the dynamic selection method or the dynamic weighting method is data-dependent.

As can be seen from Tables 3 & 4, the performance of these methods degraded when evaluated on noisy data. However, the degradation was not very serious (a maximum increment of 5% on only formaldehyde dataset). Indeed, the robustness of these methods can be enhanced by generating ensembles composed of significantly accurate and diverse base models.

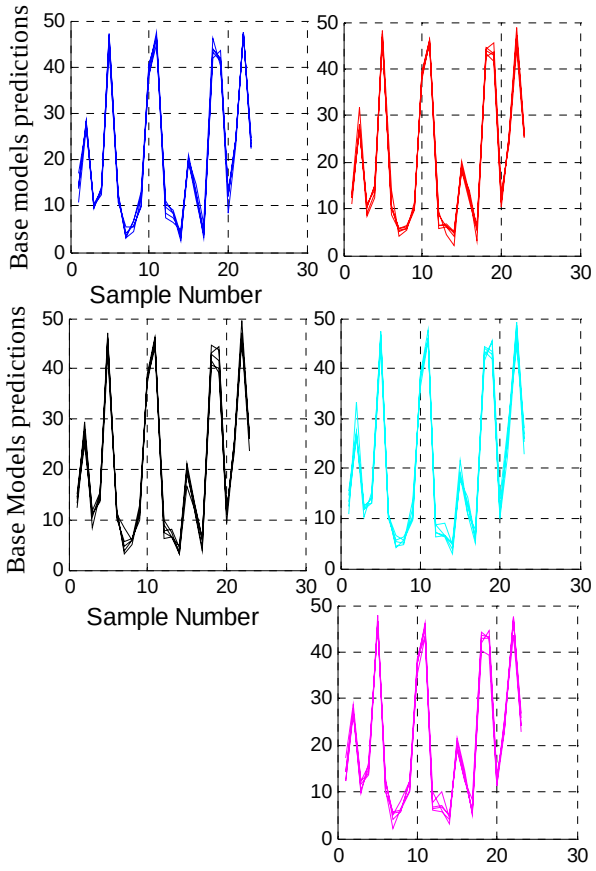

Fig. 4 Predictions from all base models in ensemble for CO validation data

5. Conclusion

Ensemble learning is a promising alternative for online prediction in general, and for online air quality monitoring using electronic nose systems in particular. However, if ensemble learning is to be effective, the generation and integration of bases models should be optimally done. In this paper, a new dynamic integration method is proposed, the performance of which was compared with that of other integration methods. Experimental results show that the new method is not only effective but also appealing. Moreover, owing that the evaluation was done on ensemble composed of relatively accurate and diverse base models, and dynamic weighting methods are sensitive to inaccurate and/or redundant base models, we believe that this method would have performed better if the ensemble generation method was done using more efficient techniques. This will be considered in our future work, along with evaluation on large-scale datasets and comparison with many methods.

Acknowledgments

The authors would like to acknowledge financial supports from the Key Science and Technology Research Program (CSTC2010AB2002, CSTC2009BA2021), the Chongqing University Postgraduates' Science and Innovation Fund (CDJXS12160005), the Fundamental Research Funds for the Central Universities of China (CDJXS10161114), and the Natural Science Foundation Project of CQ CSTC (cstc2012jjB0123). Special thanks to the anonymous reviewers for their valuable comments and suggestions.

References

[1] R. J. Mammone, Artificial Neural Networks for Speech and Vision, London: Chapman & Hall, pp.126-142, 1993.
[2] Hansen, Salamon, "Neural network ensembles", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 12, No. 10, 1990,pp.993-1001.
[3] D. Gao, et al., "Simultaneous estimation of classes and concentrations of odors by an electronic nose using combinative and modular multilayer perceptrons", Sensors and Actuators B, Vol.107, 2005, pp.773-781.
[4] G. Daqi, C. Wei, "Simultaneous estimation of odor classes and concentrations using an electronic nose with function approximation model ensembles" Sensors and Actuator B, Vol.120, 2007 , pp.584-594.
[5] Shi M., Brahim-Belhouari S., Bermak A. and Martinez D, "Committee machine for odor discrimination in gas sensor array", in Proceedings of the 11th International Symposium on olfaction and electronic nose (ISOEN), 2005, pp. 74-76.
[6] Vitor Hirayama, Francisco J. Ramirez-Fernandez, Walter J. Salcedo, "Committee machine for LPG calorific power classification", Sensors and Actuators B, Vol.116, 2006, pp. 62-65.
[7] A.Vergara et al., "Chemical gas sensors drift compensation using classifier ensembles", Sensors and Actuators B, Vol.166-167, 2012, pp. 320-329.
[8] A. Amini, et al., "Improving gas identification accuracy of a temperature-modulated gas sensor using an ensemble of classifiers", Sens. Actuators B: Chem. (2012).
[9] L.Zhang, F.C.Tian, H.Nie, et al., "Classification of multiple indoor air contaminants by an electronic nose and a hybrid support vector machine", Sensors and Actuators B, Vol.174, 2012, pp.114-125.
[10] L. Zhang, F.C. Tian, S. Liu, et al., "Chaos based neural network optimization for Concentration Estimation of Indoor Air Contaminants",Sensors and Actuators A, Vol.189,2013, pp.161-167.

[11] A. Szczurek et al., "The stop-flow mode of operation applied to a single chemiresistor", Sensors and Actuators B, Vol. 148, 2010, pp.522-530.
[12] R.W. Kennard, L.A. Stone, "Computer aided design of experiments", Technometrics, Vol. 11, 1969, pp. 137-148.
[13] L. Breiman, "Bagging predictors", Machine Learning, Vol. 24, 1996, pp.123-140.
[14] Freund, Y., and R. E. Schapire, "Experiments with a new boosting algorithm", in Proceedings of the thirteenth International Conference on Machine Learning,1996,pp. 148-156.
[15] Liu, Y. & Yao, X., "Ensemble Learning via Negative Correlation", Neural Networks, Vol.12, 1999, pp.1399-1404.
[16] Opitz, D., Shavlik, J., "Actively Searching for an Effective Neural-Network Ensemble", Connection Science, Vol. 8, 1996, pp.337-353.
[17] A. Sharkey and N. Sharkey, "Combining diverse neural networks", The Knowledge Engineering Review, Vol.12, No.3, 1997, pp.231-247.
[18] S. Haykin, Neural Networks, A Comprehensive Foundation, 2nd Edition, Prentice-Hall, 1999.
[19] Wolpert, D., "Stacked Generalization", Neural Networks, Vol.5, 1992, pp. 241-259.
[20] L. Didaci, G. Giacinto, F. Roli, G.L. Marcialis, "A study on the performances of dynamic classifier selection based on local accuracy estimation", Pattern Recognition, Vol. 38, No.11, 2005, pp. 2188-2191.
[21] Woods,K., "Combination of multiple classifiers using local accuracy estimates", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.19, No.4,1997, pp. 405-410.
[22] N. Suguna, and K. Thanushkodi, "An Improved k-Nearest Neighbor Classification Using Genetic Algorithm", International Journal of Computer Science Issues, Vol.7, Issue 4, No 2, 2010.