Neural Processing Letters **19:** 25–35, 2004.

Ensemble Learning for Chemical Sensor Arrays

S. BERMEJO★ and J. CABESTANY
Department of Electronic Engineering, UPC, Jordi Girona 1-3, 08034 Barcelona, Spain.
e-mail: sbermejo@eel.upc.es

Abstract. Electrochemical sensors, like ion-selective field transistors (ISFET), are electronic devices that merge solid-state electronic technology with chemical sensors so as to be sensitive to the concentration of a particular ion in a solution. However, as it has been previously reported, their response does not only depend on a single ion but also is affected by several interfering ions found in the solution to be measured. These interfering ions can be considered as noise and consequently, a post-processing stage that increases the SNR is obligatory. Our work shows how ensemble learning methods could be used in an array of chemical sensors in order to deal with this problem. In particular, we introduce a novel neural learning architecture for ISFET arrays, which employ ISFET models as prior knowledge. The proposed ensemble learning systems are RBF-like solutions based on bagging and optimal linear combination. Several experimental results are included, which demonstrate the interest and viability of the proposed solution.

Key words. array processing, bagging, ensemble learning, optimal linear combination, radial basis functions (RBF), sensor arrays

1. Introduction

Traditionally, the aim of sensor designers has been the production of single, highly selective sensors. However, there is an alternative method developed in the last decade based on the design of arrays of less-selective sensors and combined with a post-processing stage that allows you to obtain an overall response that is better than that obtained by the individual sensors. Such a post-processing stage performs a kind of statistical array processing [1] to overcome the inherent limitations of sensors by enhancing their output signals in order to facilitate the extraction of relevant information in further stages. Hence, the incorporation of these intelligent processing techniques into traditional sensor systems gave rise to the so-called smart sensors [2]. Electronic noses [3, 4] are a recent example of this approach in chemical sensing, which aims to obtain qualitative or quantitative information about particular chemical components [5]. This kind of analysis of a chemical solution can be performed with electrochemical sensors [6, 7], like ISFETs [8], which convert chemical information into an electrical signal using a transductor capable of recognizing some chemical components.

Before ISFETs, or other chemical sensors, can be employed in normal conditions, they must be calibrated in order to obtain an output linearly dependent on the ion

★Corresponding author.

concentration. In order to do this, some regression methods are usually employed to recover the original concentration from the output of the ISFET conditioning circuit [9,10]. However this task is not easy because they are sensitive, not only to a single ion but, to several interfering ions making their response a non-linear function of a mixture of ion concentrations found in the solution. Additionally, their output response varies from one device to another given a collection of them, which strongly links the calibration process to a particular device that has a limited lifetime (months). Nevertheless, these two problems could be tackled with array processing techniques widely employed in the Neural Networks community such as ensemble learning.

Ensemble learning [11] is a well-established method for improving the generalization performance within a set of learning machines. The idea is to combine a collection of learning systems (e.g., linear regressors) that have been trained in the same task to obtain a stabilized solution from a set of single predictors, which suffer a strong dependence on the particular training set, initial conditions and parameters of the learning system. All these problems found in a single predictor limit its final performance, which is in principle undesirable [12, 13]. It is important to stress that, in order to gain something from using an ensemble, the single predictors to combine must be different in order to produce a better-combined solution, which happens in sensor arrays since there is always a spatial diversity within their single members. Since ensemble methods typically reduce the variance term of the generalization error [14], an improvement in the array performance is often achieved in comparison with the single predictors of the ensemble. Bagging [15, 16] and linear combination of predictors [17] are two simple forms of ensemble learning for regression problems.

In this work, we explore some simple forms of array processing based on ensemble learning for recovering original ion concentrations from an ISFET array. The rest of the paper is organized as follows: Section 2 introduces how ensemble methods can be employed in ISFET arrays with the help of an empirical ISFET model which is used as a prior knowledge to the learning system. In Section 3, some preliminary experiments are introduced to demonstrate the viability of the proposed solution. Finally, Section 4 presents some conclusions.

2. Ensemble Learning for ISFET Arrays

2.1. PRIOR KNOWLEDGE BASED ON A LOCAL ISFET MODEL

Ion-sensitive field-effect transistors (ISFETs) [18] are among the most interesting chemical sensors that have appeared in the last decades. They are built on MOSFETs [19] and achieve selectivity to a particular ion activity in a chemical solution by replacing the metallic gate in a MOSFET with a structure of membranes. Many different types of ISFETs have been presented varying the materials and structure of the sensitive membrane, e.g., MEMFETs [20], REFETs [21] and CHEMFETs [22]. However all these devices operate with the same basic principles. The classical approach in ISFET modeling [22, 23] consists in developing an equivalent circuit formed by two (relatively) uncoupled parts: the electrochemical part that models the membrane, which is responsible of the electrochemical transduction, and the electrical one based on a MOSFET model. According to this model, the steady-state drain current of an ISFET can be expressed in the linear range of operation using the MOSFET equations as follows:

$$I_{\mathrm{d}} = \beta(\mathrm{V}_r - \mathrm{V}_{T'} - 0.5\mathrm{V}_{\mathrm{ds}})\mathrm{V}_{\mathrm{ds}} \quad \text{for } \mathrm{V}_{\mathrm{ds}} < \mathrm{V}_r - \mathrm{V}_{T'} \tag{1}$$

where V_r is the voltage applied to the reference electrode, V_{ds} is the drain-source voltage, $\mathrm{V}_{T'}$ is the chemically modified threshold voltage and $\beta = \mu C_{\mathrm{ox}} W/L$. The threshold voltage which can be expressed as

$$\mathrm{V}_{T'} = \mathrm{V}_T - \mathrm{EMF} \tag{2}$$

where EMF is the electrochemically induced voltage. When more than one ion is present in the solution, EMF can be locally described by the Nickolsky equation:

$$\mathrm{EMF} = \mathrm{E}_i + \mathrm{E}_m = \mathrm{E}_i + \frac{\mathrm{RT}}{n_i\mathrm{F}}\ln\left(a_i + \sum_j \mathrm{K}_{ij} a_j^{z_i/z_j}\right) \tag{3}$$

where E_i is a constant than depends on several chemical constants, E_m is the voltage across the membrane surface, R is the gas constant, T is the temperature in Kelvin, n_i is the charge of the measured ion, F is the Faraday constant and a_i is the ion activity, K_{ij} is the selectivity coefficient which relates the response to the interfer ing ions a_j, Z_i is the valence of the main ion and Z_j is the valence of the disturbing ion j in the solution. Consequently, (1) can be written as

$$I_{\mathrm{d}} = \beta\left(\mathrm{V}_r + \mathrm{E}_o + \Phi\log_{10}\left(a_i + \sum_j \mathrm{K}_{ij} a_j^{Z_k/Z_j}\right) - \mathrm{V}_T - 0.5\mathrm{V}_{\mathrm{ds}}\right)\mathrm{V}_{\mathrm{ds}} \tag{4}$$

where $\{\beta, \mathrm{E}_o, \Phi, \mathrm{K}_{ij}, \mathrm{V}_T\}$ denote the parameter set of the model to be determined from measurements. As Figure 1 shows, the model for a potassium ISFET fits the training data ($\mathrm{V}_r = 2\mathrm{V}$ and $\mathrm{V}_{\mathrm{ds}} = 0.5\mathrm{V}$) well and departs slightly from test data ($\mathrm{V}_r = 2\mathrm{V}$ and $\mathrm{V}_{\mathrm{ds}} \neq 0.5\mathrm{V}$) which denotes a dependency on the Nikolsky model's parameters of the operating point. As one can observe in Figure 2 the local model is accurate and simple enough to be incorporated into an ensemble learning architecture as a prior knowledge to form a model of how mixing is produced.

2.2. AN ENSEMBLE LEARNING ARCHITECTURE

We suppose that a number of ISFETs are included in an I/V conversion circuit that provides an output voltage proportional to the ISFET membrane potential [26, 27]. Additionally, a steady-state preprocessing and A/D conversion is also performed in order to eliminate the transient part of the ISFET response, as well as digitalizing the array analog signals. According to the above assumptions, the input of our neural processing system can be modeled, with the incorporation of prior knowledge based

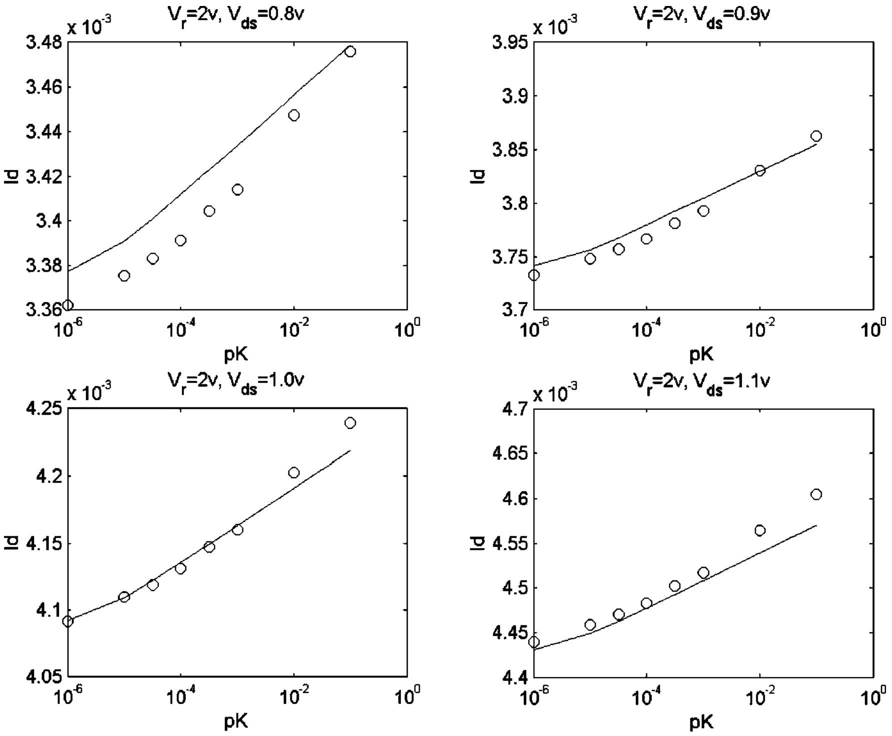

Figure 1. Fitting the Nikosky model in a potassium ISFET [24, 25] in which a sodium interference a_{Na} was present: the model was fitted for $V_r = 2V$ and $V_{ds} = 1V$ while it was tested for $V_r = 2V$ and $V_{ds} = 0.8, 0.9$ and 1.1V. The estimated parameters of the model are $\beta = 0.0013$, $E_o = 0.3160V$, $V_t = -1.4595V$, $K = 1.66e-005$ and Φ=.0211. Note: $pK = -\log_{10}(a_K)$ and $a_{Na} = 0.1$.

on (1), by m discrete-time signals $x_1[n], \ldots, x_m[n]$ that correspond to a non-linear mixture of p source signals $a_1[n], \ldots, a_p[n]$ based on (1), i.e.,

$$\begin{aligned} x_1[n] &= b_{11} + b_{12}\log_{10} s_1[n] \\ &\vdots \\ x_m[n] &= b_{m1} + b_{m2}\log_{10} s_m[n] \end{aligned} \tag{5}$$

with

$$s_u[n] = a_i[n] + \sum_j K^u_{ij} a_j[n]^{Z_i/Z_j}, \quad u = 1, \ldots, m \tag{6}$$

or expressed in a vector form

$$\mathbf{x}[n] = (x_1[n] \ldots x_m[n])^T = \mathbf{F}(a[n]) \tag{7}$$

where $\mathbf{F}$ is a non-linear mixing function from $\Re^p$to $\Re^m$ and T denotes transpose. Two additional remarks on (4)–(7) are worth noting. First, the I/V conversion

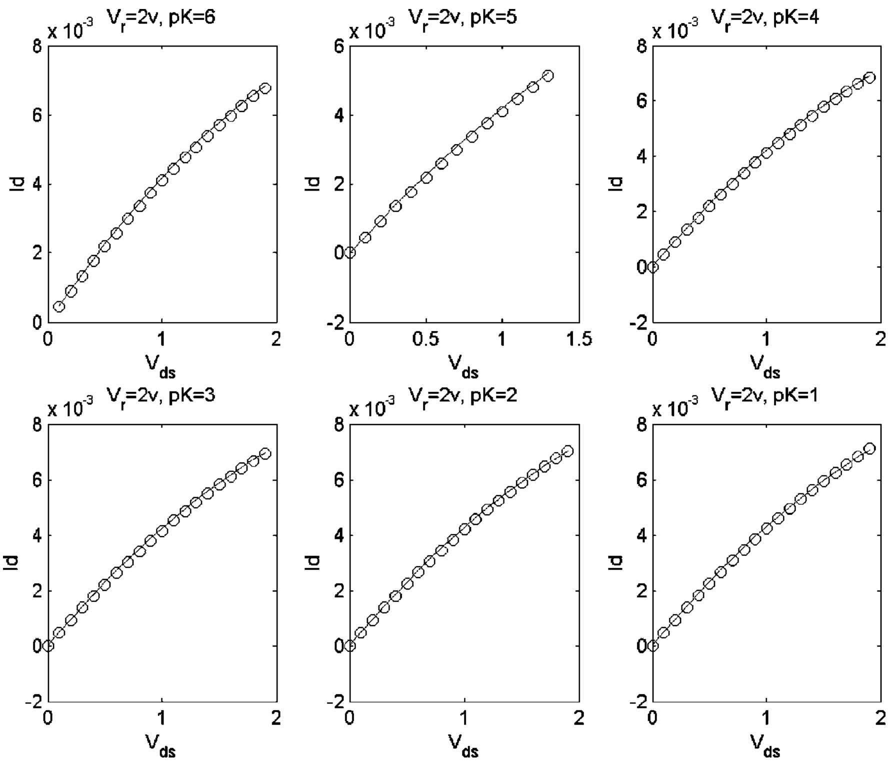

Figure 2. Fitting the I_d-V_{ds} to data with the K coefficient and Φ extracted from a previously computed Nikolski model for $V_r = 2V$ and $V_{ds} = 1V$. The final estimated parameters of the model are $\beta = 0.0013$, $E_o = 0.3263V$, $V_t = -1.4698V$, $K = 2.46e\text{-}005$, $\Phi = 0.021$.

performed by an ISFET conditioning circuit is a linear function of the drain current in the linear range of the ISFET so any linear function of (1) can be arranged in the form of (5). On the other hand, it is implicitly assumed in (4)–(6) that all the ISFETs in the array are sensitive to the same ion activity a_i and consequently the other sources are considered in principle as noise.

According to (7), our goal is to create an inverse of **F** from a set of measurements {**a[n]**, **x[n]**} in order to predict ion activities from new input data measurements. This recovery process can be achieved by a pxm separating function **G** which allows an estimation of the source signals $\{a_i[n], i = 1, \ldots, p\}$ through the following reconstruction algorithm,

$$\mathbf{y}[n] = (y_1[n] \ldots y_p[n])^T = \mathbf{G}(\mathbf{x}[n]) \approx \mathbf{a}[n] \tag{8}$$

where clearly **G** must tend to $\mathbf{F}^{-1}$ to obtain a good estimation of **a** with **y**. As a first step on constructing **G**, a NN predictor for a single element ($m = 1$) can be obtained as

$$y[n] = 10^{\frac{x[n]-w_1}{w_2}} \tag{9}$$

where y is an estimator of a_i based on the inverse mapping of (1) for $m = 1$ and the weights w_u are computed minimizing the MSE from a (training) set of samples $\{(x_u, ai_u), u = 1, \ldots, N\}$. Note that the ultimate performance of the non-linear regressor will be limited by the amount of noise caused by the interfering ion activities in (6). A straightforward extension could be formulated when $m > 1$ with the inclusion of an output layer that 'averages' the results of the hidden nodes into a Radial basis function-like architecture:

$$\text{Hidden layer:} \quad z_u[n] = 10^{\frac{x_u[n]-w_{u1}}{w_{u2}}} \quad u = 1, \ldots, m \tag{10}$$

$$\text{Output layer:} \quad y[n] = \sum_{u=1}^{m} f_u(z_u[n]) \tag{11}$$

with

$$f_u(z) = 1/m \tag{12}$$

or

$$f_u(z) = w'_u \tag{13}$$

The resulting NN in (10), (11) can be noted as bagged NN if it employs (12) and linear NN if it is based on (13) since the set of predictors $\{z_u\}$ are fused through the ensemble methods of bagging [15] and optimal linear combination [17], respectively. This modular NN can be trained in two steps with the following algorithm: (1) train the individual predictors $\{z_u\}$ minimizing the MSE, and (2) train then the output layer that incorporates the individual predictors using bagging and optimal linear combination. Note that the proposed RBF-like NN is directed at predicting the ion activity a_i to which ISFETs are most sensitive.

The overall NN, given by (10)–(13), forms an ensemble of simple non-linear predictors, although it has a similar architecture than a RBF. However, it cannot be considered as a RBF since the single predictors are trained separately and the final linear combination is performed with fusion techniques provided by ensemble learning. More importantly, all the single predictors are trained in the same task and the aim of the linear combination is to improve the response of the best of the single predictors of the ensemble. In fact, the problem could be solved with a single non-linear predictor and consequently the goal of the overall NN is to improve the single sensor response with the help of combining several sensors through ensemble techniques.

3. Experiments

We present some simple experiments with an m-potassium ISFET array in which training data has been artificially created from real measurements. Interferences are reduced only to sodium ions. We have extracted the ISFET model parameters of (1) using the potassium ISFET measurements. Then, in order to simulate the differences between the real devices of the array, we have generated random parameters uniformly distributed at ±10% around the extracted parameters, which is representative of the observed behavior in a population of ISFETs. Potassium and sodium activities were uniformly distributed over $[10^{-6}, 10^{-1}]$ for simulating a weak interference in the presence of potassium activity with $K \sim 10^{-5}$.

In Figures 3 and 4, we show the predicted and real potassium activities computed for a single training and test set of 100 samples for m-bagged NN, *m*-linear NN with $m = 2$ and 6. The test mean squared errors for linear, bagged and single NNs were 2.3e-5, 4.9e-5, 5.35e-5 for $m = 2$, and 5.16e-5, 9.93e-5, 1.14e-4 for $m = 6$ respectively. These results denote that bagged NN slightly improves the behavior of a single predictor but it is inferior to the linear NN. Additionally, it can be also observed that the increase of the elements in the ensemble does improve the overall response with a certain optimal number since 6-NN performances are worse than those achieved by 2-NN.

Several additional experiments were performed in order to analyze the behavior of the three supervised NN approaches and the impact of the number of elements

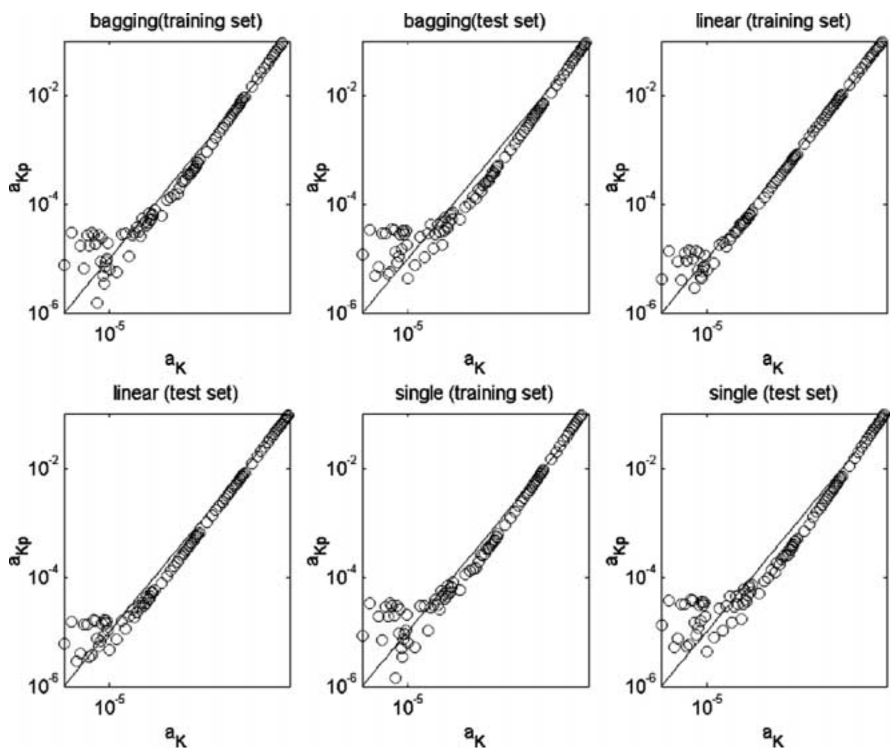

Figure 3. Predicted vs. real potassium activities in the training and test set for the supervised approach based on a single element, 2-bagged NN and 2-linear NN.

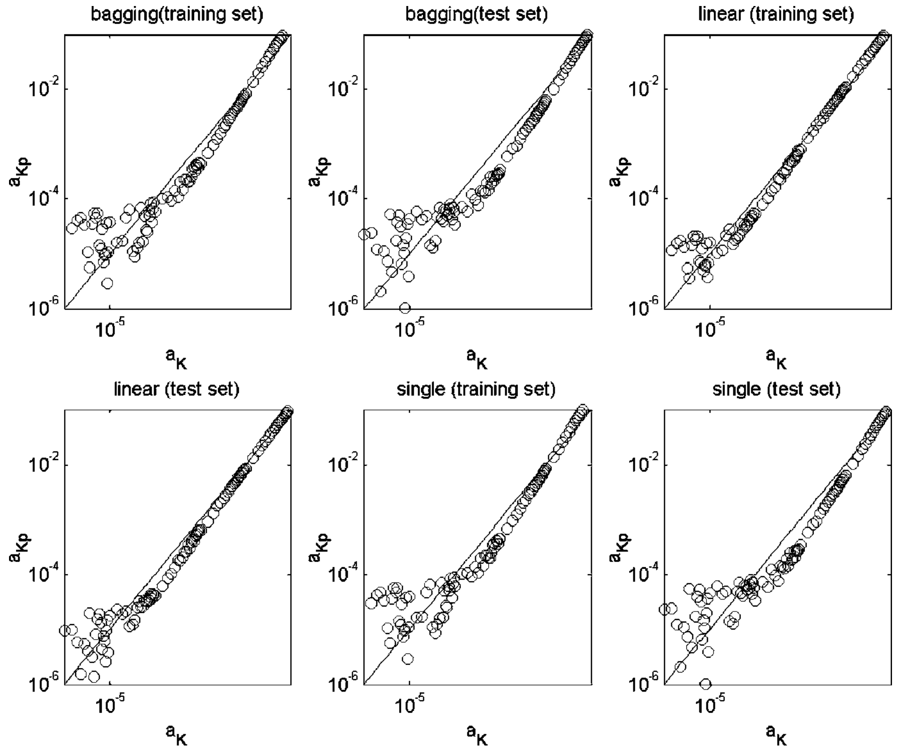

Figure 4. Predicted vs. real potassium activities in the training and test set for the supervised approach based on a single NN, 6-bagged NN and 6-linear NN.

in the ensemble. Figure 5 shows the test MSE for different sizes (N) of the training and test sets over 1000 runs for $m = 2$. As before, the 2-linear NN outperforms the 2-bagged NN which achieves almost the same behavior as a single NN. Another important result observed in Figure 5 is that the increase of the number of training samples does not decrease very much the MSE. This behavior can be related to the use of linear regressors since they achieve better generalization performance with few training samples in comparison with non-linear learning systems. So, in practice, we can calibrate the sensor array with few training samples, which is a very interesting feature in chemical sensing since multiple-point calibration is a laborious and time-consuming process. On the other hand, the existence of an optimal m given a fixed N is clearly demonstrated in Figure 6, in which the optimal numbers of the ensemble are achieved for 8 and 9. However, as one can also observe in Figure 6, the number of elements can be reduced for practical purposes since performance is not considerably improved upon 3 or 4 elements, which is a common choice in real applications of sensor arrays.

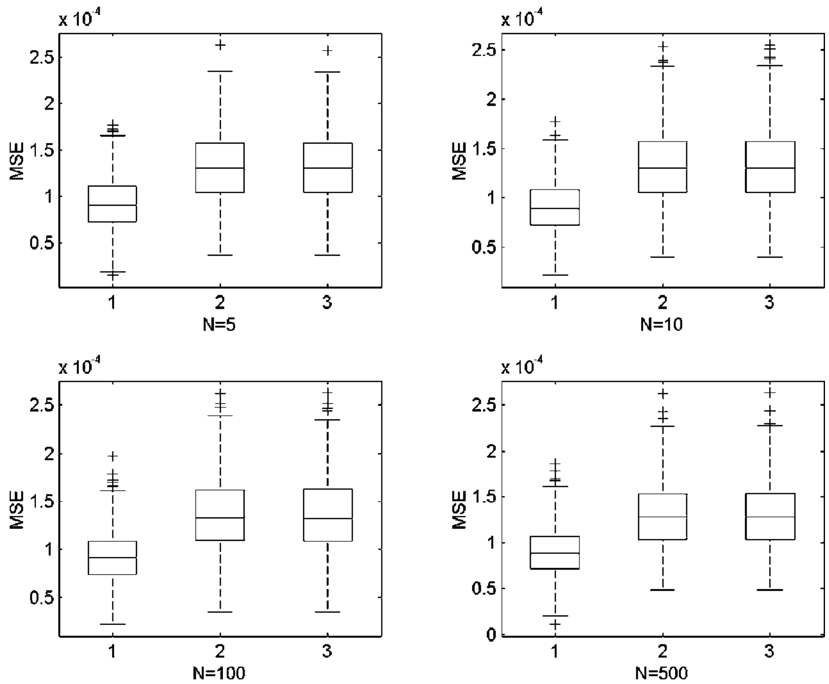

Figure 5. Test MSE computed over 1000 repetitions for several training set-sizes (N). Notes: 1=2-linear, 2=2-bagged NN, 3=single NN.

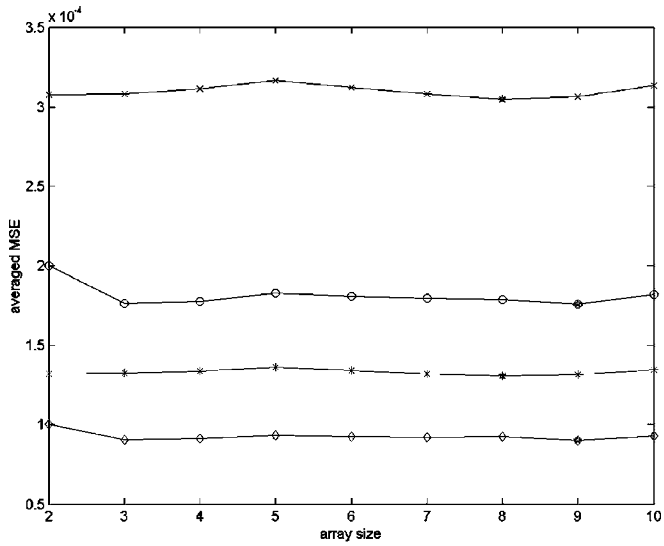

Figure 6. Average training (circle and x-mark) and test (diamond and star) MSE computed over 1000 repetitiions for m-linear NN and m-bagged NN respectively with N=100, m=2,...,10. The best results are double marked.

4. Conclusions

We have presented two novel learning architectures for the problem of recovering ion activities in ISFETs arrays. The development of these neural architectures were based on incorporating prior knowledge extracted from local ISFET models, which bias the learning algorithm to a feasible solution using a small set of measurements. The proposed learning system, which aims to recover the main ion activity of the ISFET array, incorporates linear regression into an ensemble learning method using bagging or optimal linear combination. Experiments with artificial data generated from real measurements in m-ISFET arrays demonstrate the viability of our proposal since a small ensemble formed by 3 or 4 elements calibrated with few training samples (5–10) considerably improves the performance of a single element.

Acknowledgements

This research has been partially supported the SEWING EU project (contract no. IST-2000-28084). The authors want to acknowledge Dr. Leszek Opalski from the Institute of Electronic Systems of the Warsaw University of Technology (WUT) for supplying the ISFETs measurements.

References

1. Chen, T., et al.: Highlights of Statistical Signal and Array Processing, *IEEE Signal Processing Mag*. (Sep. 1998), 21–64.
2. Gardner, J. W.: *Microsensors: Principles and Applications*, John Wiley & Sons, Chichester, 1994.
3. Nagle, H. T., Schiffman, S. S. and Gutierrez-Osuna, R.: The how and why of electronic noses, *IEEE Spectrum*. (1998), 22–31.
4. Gutierrez-Osuna, R. and Nagle, H. T.: A method for evaluating data-processing techniques for odor classification with an array of gas sensors, *IEEE Trans. Sys. Man, Cybern. B*. SMVB-29, No.5, (1999), 626–632.
5. Wiemhöfer, H. D. and Cammann, K.: Ion-sensitive Electrodes, In: W. Göpel J. Hesse and J. N. Zemel (eds.), *Sensors: A Comprehensive Survey*, Vol. 2, Chemical and Biochemical Sensors Part I. Edited by W. Göpel, T. A. Jones, M. Kleitz, J. Lundström, T. Seiyama, VCH, Weinheim, pp. 177–183, 1991.
6. Cattrall, R. W.: *Chemical Sensors*, Oxford University Press, Oxford, 1997.
7. Wilson, D. M., Hoyt, S., Janata, J., Booksh K. and Obando, L.: Chemical Sensors for Portable, Handheld Field Instruments, *IEEE Sensors J*., **1** (4), (December 2001).
8. van den Berg, A.: Ion-sensitive Field Effect Transistors, In: W. Göpel J. Hesse and J. N. Zemel (eds.), *Sensors: A Comprehensive Survey*, Vol. 2, Chemical and Biochemical Sensors Part I. Edited by W. Göpel, T. A. Jones, M. Kleitz, J. Lundström, T. Seiyama, VCH, Weinheim, pp. 474–501, 1991.
9. Dybko, A.: *Chemical Sensors: Technology and Metrology*, Ph.D. Thesis, Department of Analytical Chemistry, Warsaw University of Technology, 2001.
10. Baucke, F. G. K., Naumann, R. and Alexander-Weber, Ch.: Multiple-point calibration with linear regression as a proposed standardization procedure for high-precision pH measurements, *Anal. Chem*. **65** (1993), 3244–3251.

11. Sharkey, A. (ed.).: *Combining Artificial Neural Nets*, Springer-Verlag, Berlin, 1999.
12. Breiman, L.: The Heuristics of Instability in Model Selection, *Ann. Stat.*, **24** (1997), 2350–2381.
13. Breiman, L.: Bias-Variance, Regularization, Instability and Stabilization, In: C. M. Bishop (ed.), *Neural Networks and Machine Learning*, NATO ASI Series, Series F: Computer and Systems Sciences, Vol. 168, Springer-Verlag, Berlin, 1998.
14. Geman, S., Bienenstock, E. and Doursat, R.: Neural Networks and the bias/variance dilemma, *Neural Comp*. **4** (1992).
15. Breiman, L.: Bagging Predictors, *Machine Learning*. **26** No. 2, (1996), 734–751.
16. Natftaly, U., Intrator, N., Horn, D.: Optimal Ensemble Averaging of Neural Networks, *Technical Report*, *Israel Institute of Technology*, (1997).
17. Hashem, S.: Optimal linear combinations of neural networks, *Neural Networks*, **10**(4), (1997), 599–614.
18. Bergveld, P.: Development of an Ion-sensitive solid-state device for neurophysiological measurements, *IEEE Trans. Biomed. Eng*. BME-17 (1970), 70–71.
19. Warner, R. M. Jr. and Grung, B. L.: MOSFET: *Theory and Design*, Oxford University Press, New York, 1999.
20. Bousse, L., de Rooij, N. F. and Bergveld, P.: Operation of chemically sensitive field-effect sensors as a function of the insulator-electrolyte interface, *IEEE Trans. Electron. Devices*. ED- 30 (1983), 1263–1273.
21. Smith, R. L. and Scott, D. C.: An integrated sensor for electrochemical measurements, *IEEE Trans. Biomed. Eng*. BME-33 (1986), 83–90.
22. Sibbald, A.: Chemical-sensitive field-effect transistors, *IEE Proceedings*. **130**(5), Part I (1983), 233–244.
23. Martinoia, S. and Massobrio, G.: A behavioral macromodel of the ISFET in SPICE, *Sens. Act. B*. **62** (2000), 182–189.
24. Brzózka, Z., Holterman, A. J., Honing, W. N., Verkerk, U. H., can den Vlekkert, H. H., Engbersen, J. F. J. and Reinholut, D. N.: Enhanced performance of potassium CHEMFET by optimization of a polysiloxane membrane, *Sens. Act. B*. **18–19** (1994), 38–41.
25. Opalski, L. J.: Personal communication.
26. Palán, B., Santos, F. V., Karam, J. M., Courtois, B. and Husák, M. : New ISFET sensor interface circuit for biomedical applications, *Sens. Act. B*. **57** (1999), 36–42.
27. Chin, Y.-L., Chou, J.-C., Sun, T.-P., Liao, H.-K., Chung, W.-Y. and Hsiung, S.-K.: A novel SnO2/Al discrete gate ISFET pH sensor with CMOS standard process, *Sens. Act. B* **75** (2001), 36–42.