Adaptive Stochastic Classifier for Noisy pH-ISFET Measurements

Tong Boon Tang, Hsin Chen, and Alan F. Murray

School of Engineering and Electronics,
The University of Edinburgh,
Edinburgh EH9 3JL, UK
{tbt, hc, afm}@ee.ed.ac.uk

Abstract. Sensor drift is an inevitable measurement problem and is particularly significant in the long term. The common practice is to have an auto-calibration facility (including standard buffers or accurate integrated actuators) mounted on the monitoring system. However, this approach may not be feasible when the monitoring system is miniaturized to the size of a capsule. In this paper, we develop an adaptive stochastic classifier using analogue neural computation to produce constantly-reliable classification for noisy pH-ISFET measurements. This classifier operates at the signal-level fusion and auto-calibrates its parameters to compensate the sensor drift, with simple learning rules. The ability of the classifier to operate with a drift of 85 % of the pH-ISFET's full dynamic range is demonstrated. This sensor fusion highlights the potential of neural computation in miniaturized multisensor analytical microsystems such as Lab-in-a-Pill (LIAP) for long-term measurements.

1 Introduction

Driven by current Lab-on-a-Chip and System-on-Chip (SoC) technological trends, it is now possible to shrink a complex multisensor microsystem into the size of a capsule [1]. However, it is then inherently more difficult to extract useful information from what is now far noisier sensor data. Our primary interest is in a simple data-fusion algorithm that is robust to noise, hardware-amenable and thus able to underpin an intelligent sensor fusion system (ISFS). Since sensor drift is an unavoidable problem especially in the long term, an ISFS should be **autonomous adaptive** and possibly capable to **classify** noisy sensory data into categories.

Our prototype capsule [1] contains standard PN-junction silicon diode temperature sensor [2] and ion-sensitive field effect transistor (ISFET) pH sensor [3]. The ISFET suffers from four different types of sensor drift. The type of drift, which we confront most, is due to the instability of the reference electrode. One approach to eliminate the drift is to use two different pH-sensitive layers (e.g. Ta_2O_5 and oxynitride/Si_3N_4) in differential mode. However, this design suffers from crosstalk between the transistors when some protons (by-product of enzyme reaction) diffuse to the ion-insensitive transistor and cause a false signal. This, combined with our interest in miniaturization of the multisensor microsystem [1], means that we choose to use a single ISFET for each pH measurement.

O. Kaynak et al. (Eds.): ICANN/ICONIP 2003, LNCS 2714, pp. 638–645, 2003.

In this study, we choose to employ neural computation, in particular Continuous Restricted Boltzmann Machine (CRBM) [4], to calibrate the ISFET's drift. The CRBM has been shown to be a stochastic generative model that can model continuous data with a simple training algorithm. The following sections present our investigation on the potential of CRBM in our application - performing reliable classification of noisy data that suffers from sensor drift.

2 Neural Computation

This section briefs the CRBM model, while detailed description should be referred to [4]. The CRBM has one visible and one hidden layer with inter-layer connection defined by a weight matrix $\{\mathbf{W}\}$. A stochastic neuron j has the following form:

$$s_j = \varphi_j \left(\sum_i w_{ij} s_i + \sigma \cdot N_j(0,1) \right), \tag{1}$$

$$\text{with} \quad \varphi_j(x_j) = \theta_L + (\theta_H - \theta_L) \cdot \frac{1}{1 + \exp(-a_j x_j)} \tag{2}$$

where s_i refers to input from neuron i, and $N_j(0,1)$ represents a unit Gaussian noise with zero mean. The noise component $\sigma \cdot N_j(0,1)$ allows the CRBM to perform probabilistic *analogue* neural computation without quantization and hence avoid unnecessary loss of information which a binary RBM suffers from [5,6,7]. To enhance efficient learning, the noise scaling factor σ in visible layer is set to a constant value close to the input data's standard deviation, while σ in hidden layer is set to 0.4 in order to avoid over-fitting problem [8]. Parameter a_j is the noise-control factor which controls the slope of the sigmoid function, such that a neuron j behaves deterministically (small a_j), or continuous-stochastically (moderate a_j), or binary-stochastically (large a_j). θ_H and θ_L are then simply two constants defining the sigmoid function's asymptotes.

Both $\{a_j\}$ and $\{w_{ij}\}$ can be trained by "Minimizing Contrastive Divergence" (MCD) learning [5]. The simplified MCD learning rule [8] requires only addition and multiplication, and is therefore hardware-amenable.

3 Sensor Model

This section introduces the mathematical models of the temperature and the pH sensors.

3.1 Temperature Sensor

The signal conditioning circuit is illustrated in Fig. 1a. The output voltage V_{out} is linearly proportional to the environmental temperature T. In our application [1], the sensor should operate within a dynamic range of 0 - 70 ^{o}C with a sensitivity of 31.5 mV/^{o}C. Results from fabricated integrated sensors [1] show that the temperature sensor can be represented by:

$$V_{out}(t) = 31.5 \times T + 1030 + \tau_{temp}(t) + a_{temp} N_{temp}(t) \tag{3}$$

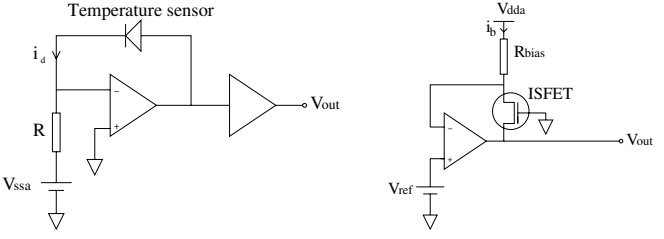
at time t. The terms $\tau_{temp}(t)$, a_{temp} and $N_{temp}(t)$ refer to non-linear sensor drift function, noise magnitude and random Gaussian noise function (due to background noise of the sensing environment). Units for $V_{out}(t)$ and T are mV and ^{o}C respectively.

3.2 pH Sensor

The signal conditioning circuit is depicted in Fig. 1b. This sensor provides a dynamic range of pH 1 - 10. Experiment [1] reveals that the ISFET has a sensitivity of 23.5 mV/pH with 16 μA of excitation current. Thus, a first-order model for the output voltage V_{out} of a particular ISFET (with an unique V_T) at time t will be:

$$V_{out}(t) = 18 \times T - 23.5 \times pH + V_T + \tau_{pH}(t) + a_C N_C(t) \tag{4}$$

The term $\tau_{pH}(t)$ refers to non-linear pH sensor drift function while a_C and $N_C(t)$ refer to the composition noise magnitude and random Gaussian noise function for both temperature and pH sensing variances. Units for $V_{out}(t)$ and T are mV and ^{o}C respectively.



(a) Temperature (b) pH-ISFET

Fig. 1. Signal conditioning circuits for the sensors

4 Experiments and Discussion

In this study, we investigate the CRBM's ability to model the measurement signals from one temperature sensor and ten pH-ISFET sensors. The 11-dimensional measurement data are generated according to (3) and (4), with a_{temp} and a_c being set to 0.013 and 0.053 which correspond to their experimentally-recorded levels (20 mV and 80 mV) [1]. With $\theta_H = 1$ and $\theta_L = -1$, all generated data are also renormalized into ± 1.

Fig. 2a shows the training data consisting of two sets of measurements. The sensors are exposed to a pH 4 liquid in the first set (cluster A), while to a pH 10 liquid in the second set (cluster B). Since both sets of measurement have the same ambient temperature of 37 ^{o}C, the mean values of the temperature sensor (first visible unit in Fig.2a) in both sets

are the same. As the characteristics of the sensors are defined by the fabrication process, the only variation between the ten ISFETs is the threshold voltage. Therefore, the two data clusters have similar distributions (Fig. 2a).

A CRBM model (implemented in Matlab) with five hidden unit, including the bias unit, is trained to model the two clusters of data. The noise-scaling constant σ is set to 0.05 for visible units, in accordance with a_{temp} and a_c, and to 0.4 for the hidden units. To test CRBM's ability to classify noisy sensory data, both weights $\{w_{ij}\}$ and noise-control factors $\{a_j\}$ are updated in the first 30,000 epochs. The results of CRBM functioning as a classifier after 30,000 epochs' training are discussed in Sec.4.1. Sec.4.2 then examines the trained CRBM's ability to trace any subsequent shift in training data after 30,000 epochs, by turning ON the learning of some particular parameters. Finally, Sec.4.3 further complicates the task by removing one cluster in the training data, and demonstrates that the CRBM remains able to trace the shift and classify noisy data.

4.1 CRBM as a Classifier

After the initial 30,000 training epochs, the bias unit H0 in the hidden layer encodes the underlaying shape (i.e. mean) of the training data distribution, as depicted in Fig. 2b. This is due to its state (permanently '+1') which allows it to learn faster (with a larger weight change $\Delta\hat{\omega_{i0}}$) than the other hidden units that have near-zero initial states. On the other hand, the hidden unit H1 has a set of large values ($\sim$ -0.6) of receptive field for visible units V2-11 (Fig. 2c) and a large noise control factor a_h (2.6956) relative to other hidden units' as shown in Fig. 3a. These point to the mechanism whereby a CRBM models two such clusters of data. The large values imply that the activity/state of the hidden unit H1 is very sensitive to the particular elements in the input data vector. Consequently, it is possible to classify the unknown input data into cluster A or B by simply observing the activity of the hidden unit with large a_h, in this case H1. Furthermore, we realize that the activity of the hidden unit can be a measure of confidence for this high-level abstraction of information. As indicated in Fig. 3b, a clear separation (lowest upper boundary - highest lower boundary = 0.65) of H1's response to two different clusters shows that a 100 % accuracy on classification is easily achieved.

4.2 Tracing Sensor Drift

After the initial 30,000 training epochs, sensor drift is introduced to the first pH sensor (second visible unit V2). For simplicity, we set $\tau_{temp}(t)$ = 0 and $\tau_{pH}(t)$ = 5 mV per 5000 epochs. If all weights ω_{ij} remain updated simultaneously as in Sec. 4.1, all ω_{ij} will "compete" to respond to any shift in the mean of the data distribution, due to their stochastic nature. Since the experiment in Sec. 4.1 indicates that the receptive field of the bias unit H0 in the hidden layer describes the mean of the training data distribution, we thus propose that only H0's receptive field be updated with the latest distribution after the initial 30,000 epochs.

After 230,000 epochs, the reconstruction data distribution, as depicted in Fig. 4b, has a notable shift of 0.2 units approximately in the V2 axis from the reconstructed data at epoch 30,000 (Fig. 4a). This shift coincides with the controlled shift of 0.2 unit in the training data. Further evidence is shown in Fig. 4c where the receptive field of the

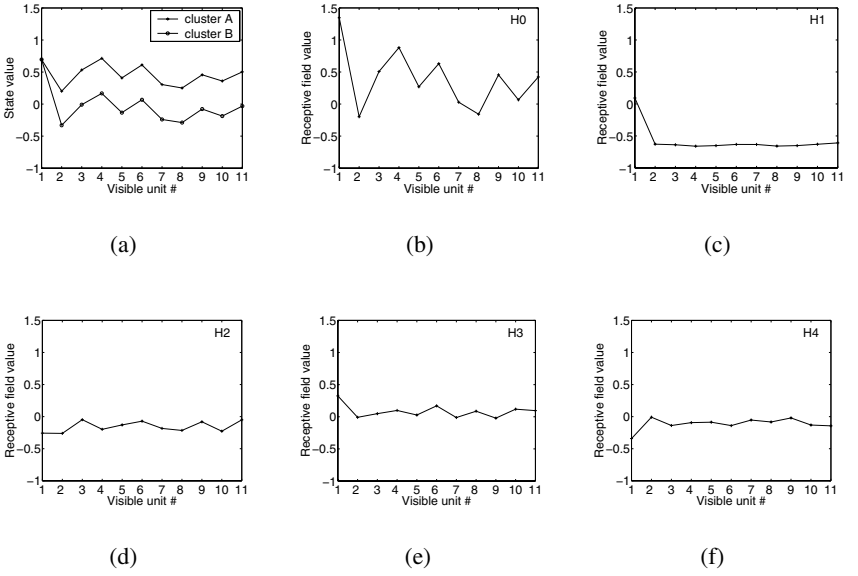


Fig. 2. (a) The mean values for the two clusters training data in the initial 30,000 epochs. (b) - (f) The receptive field values for all the hidden units H0-4 respectively after its initial training period 30,000 epochs

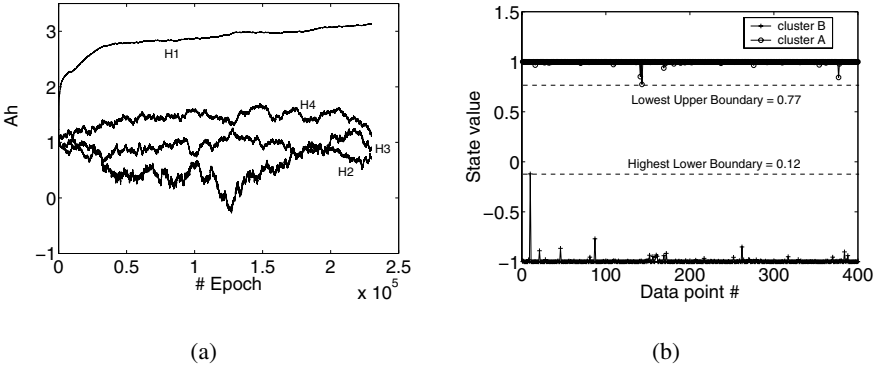
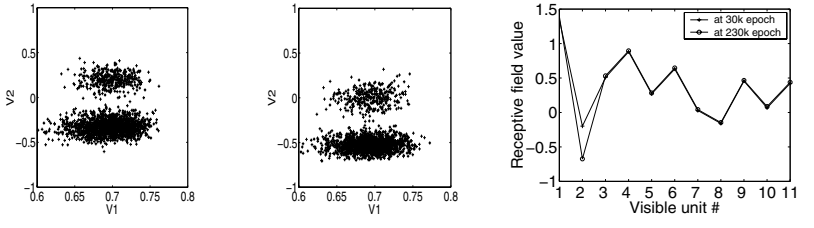


Fig. 3. (a) The noise control factor for hidden units over the entire experimental period. (b) The activities for the hidden unit H1 for a set of 400 training data points for both clusters A and B after its initial training period 30,000 epochs

bias unit in the hidden layer for visible unit V2 has been adjusted by 0.4752 unit to "trace" the change in the input data whilst the other ω_{ij} remain almost unchanged. This result consolidates the argument that a learning CRBM, under a useful and constrained configuration, can adapt to environmental changes such as sensor drift whilst present an

unchanging (autonomously re-calibrated) representation of drifting data to subsequent layers of processing - at least in this relatively simple real example.



(a) (b) (c)

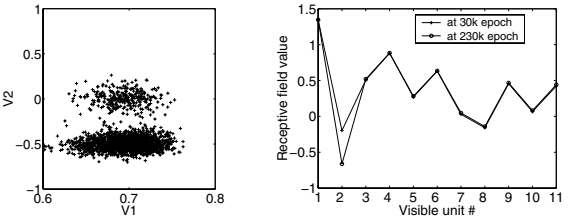
Fig. 4. (a) Reconstruction data distribution for the first two visible units after 30,000 epochs. (b) Reconstruction data distribution for the first two visible units after 230,000 epochs. (c) Receptive fields for the bias unit H0 in the hidden layer at 30,000 and 230,000 epochs

4.3 Single Cluster Updating

When a microsystem is monitoring on its surrounding environmental parameters, often there is only a single class of data available over a long period of time. This may cause a serious problem for a **learning** system such as above because of catastrophic interference. When new data distribution (a single cluster) is presented, the system tends to adopt a new set of parameters to re-generate the new distribution, and consequently lose its ability to classify.

To investigate the CRBM's ability to learn an incomplete training data without losing classifying ability, cluster B in training data is removed after epoch 30,000, and the same drift as in Sec. 4.2 is introduced to the remaining training cluster A. As predicted, despite the effort of inhibiting the learning for all (except bias unit) weights, the CRBM parameters are adjusted to re-generate only one cluster of data. We found that this is due to a strong "compensation" from the noise-control factor a_j. We therefore repeated the experiment with learning in a_j switched OFF after the initial 30,000 epochs.

Fig. 5a shows promisingly that the CRBM is still able to re-generate two clusters at epoch 230,000 despite the current training data has only single drifting cluster of data for updating its parameters. Besides, a notable shift in the V2-axis is learnt. Further evidence is given in Fig. 5b whereby the receptive field of bias unit H0 in hidden layer has a large change (0.4651) for V2 and very small adjustment (less than 0.0199) for the rest. This is encouraging and reinforces the potential of this form neural computation in dealing with noisy and drifting integrated-sensor data. One example of its possible application will be a location indicator for a capsule that monitors the gastrointestinal (GI) tract [9].



(a) (b)

Fig. 5. Experiment on using a single cluster to update a CRBM's parameters. (a) Reconstruction data for the first two visible units after 230,000 epochs. (b) Receptive fields for the bias unit H0 in the hidden layer at 30,000 and 230,000 epochs

5 Conclusion

The stochastic neural algorithm CRBM shows its ability to learn the 11-dimensional probabilistic distribution of the analogue sensor data with merely 5 hidden units. Additionally, we have demonstrated its ability to classify noisy data with reasonable confidence. With proper configuration, this classifier is further capable to counteract at least simple sensor drift to improve its reliability, by "auto-calibrating" key parameters instead of re-learning a new set of receptive fields for every newly-drifted data distribution. Finally, we demonstrated that a CRBM is able to adapt to sensor drift despite the training data is incomplete. Therefore, we anticipate that neural computation such as CRBM can find a niche in the increasingly complicated distributed multisensor systems as a form of intelligent sensor fusion.

Acknowledgment. This work is supported by the Scottish Higher Education Funding Council (RDG 130). The authors would like to thank Dr. Erik A. Johannessen for helpful discussion on chemical sensors and signal conditioning circuits.

References

1. Tang, T.B., et al: Towards a Miniature Wireless Integrated Multisensor Microsystem for Industrial and Biomedical Applications. IEEE Sensors Journal: Special Issue on Integrated Multisensor Systems and Signal Processing **2(6)** (December 2002) 628–635
2. Rashid, M.H.: Microelectronics Circuit Analysis and Design. Boston: PWS Publishing Company (1999)
3. Bergveld, P.: Development, operation, and application of the ion-sensitive field effect transistor as a tool for electrophysiology. IEEE Trans. on Biomedical Eng. **19** (1972) 342–351
4. Chen, H., Murray, A.F.: A Continuous Restricted Boltzmann Machine with a Hardware-Amenable Learning Algorithm. Proc. of the International Conference on Artificial Neural Networks (2002) 358–363
5. Hinton, G.E.: Training Products of Experts by Minimizing Contrastive Divergence. Neural Computation **14(8)** (2002) 1771–1800

6. Smolensky, P.: Information processing in dynamic systems: Foundations of harmony theory. Parallel Distributed Processing: Explorations in the Microstructure of Cognition **1** (1986) 195–281
7. Murray, A.F.: Novelty detection using products of simple experts – a potential architecture for embedded systems. Neural Networks **14** (2001) 1257–1264
8. Chen, H., Murray, A.F.: A Continuous Restricted Boltzmann Machine with an Implementable Training Algorithm. IEE Proc. of Vision, Image, Signal Processing, in press (2003)
9. Lei, W., et al: Integrated micro-instrumentation for dynamic monitoring of the gastro-intestinal tract. Proc. of the 2nd Annual International IEEE EMB Special Topic Conference on Microtechnologies in Medicine and Biology (2002) 219–222