35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Statistical Machine Learning for Removing Drift from Biosensor Signals

Xu Chen, Dean Messing, and Peter van Beek

Abstract — **Electro-chemical signals are often contaminated by drift, making accurate estimation of signal parameters challenging. We propose a method based on statistical machine learning to predict and remove drift from biosensor signals. The proposed method decomposes the observed signal into a sparse linear combination of pure drift and signal basis vectors. First, drift and signal basis sets are constructed using nearby pure drift and parametric signal models, respectively. We then build on the LASSO framework and formulate drift prediction as an optimization problem by projecting the observed signals into the drift and signal basis sets. This minimizes MSE between the predicted and observed signal under the constraints that weights on the drift and the signal bases are non-negative and have sparse representation. Validation tests with synthetic data demonstrate the effectiveness of the proposed framework. Experimental results over a large number of real data demonstrate robustness of the proposed approach.**

I. Introduction

An affinity biosensor is a device that can detect and assay specific target bio-molecules by using probe molecules immobilized on the sensor surface. The probes are designed to bind both the surface and the desired target. In the case of an impedance biosensor, binding results in electrical changes at the surface which are detected by measuring the instantaneous complex impedance magnitude signal. In a micro-array of sensors, several time-dependent signals are acquired in parallel. Details of such a biosensor system are described in [1] and [2].

Drift in such biosignals contaminates them and usually leads to poor estimation of binding and elution parameters. As shown in Fig. 1, there are 15 channels in our biosensor device and in each channel, the curve represents the impedance magnitude changing over time. Due to drift, observed binding signal can deviate from the ``true'' signal expected from the electrochemistry. Previous drift removal work has been based on polynomial fitting, linear detrending, moving average removal (MAR), and the wavelet transform [3][4]. For instance, [4] presents an approach for removing slow baseline drift components from electrocardiograph signals that uses the discrete wavelet transform. These filtering and fitting approaches mainly rely on the assumption that the drift satisfies a parametric time-domain model, or has energy concentration in a band of frequencies. However, these assumptions do not apply to the drift generated from many biosensor signals, and imposing such assumptions will limit the accuracy of drift prediction and parameter estimation.

All authors are with Sharp Laboratories of America, 5750 NW Pacific Rim Blvd, Camas, Washington, 98607, USA (Corresponding author e-mail: chenx@sharplabs.com).

As observed from the experimental data and reported in the literature [3][4], the main challenges for estimating drift components from biosensor signals are (1) the cause of drift in the impedance response is unknown, thus making it difficult to remove the drift by simply calibrating the devices; (2) As far as we know, there is no parametric model available to characterize the pattern of the drift. To this end, we propose the use of machine learning methods to predict and remove drift in biosensor signals in a non-parametric way. The proposed method is general, training-free, and the similarity of drift patterns is learned online. Therefore, it is adaptive to various drift patterns over long time periods, under different temperatures, voltages, and stimulus frequencies.

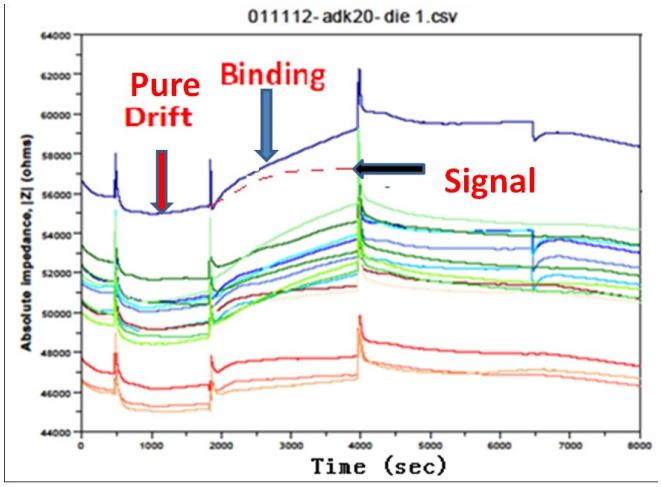

Figure 1. Visual illustration of the observed drifts (red arrow), estimated signal component (black arrow) and binding signals (blue arrow) from biosensors.

II. The Proposed Framework for Drift Removal

Based on our previous work as described in [1], the impedance magnitude in the signal component can be represented parametrically by the exponential form,

$$z(n) = B + A(1 - \exp^{-sn}) \quad (1)$$

That is, the signal component can be characterized by three constant positive parameters A, B, and s where A is the amplitude of the signal, s is the exponential time-constant, and B is a constant offset. A, B, and s are unknown and we would like to estimate these parameters after drift removal. For simplicity, we vertically shift the first sample of the drift and the binding signals to the origin so that B vanishes. The following analysis concerns only the estimation and selection of A and s. The observed signal y can be represented as

$$y(n) = z(n) + d(n) + v(n) \quad (2)$$

where *z(n)* is the sampled parametric signal component of Eq. 1, *d(n)* is the unknown non-parametric drift component, and *v(n)* is random noise. The goal is to estimate drift from the observed signal *y(n)* and then remove it, leaving a more or less uncorrupted impedance signal. Typically, leveraging the similarity of the drift patterns within a certain time period, the observed binding signals are projected onto selected drift basis and signal basis, relying on the LASSO optimization method, where the drift basis vectors can be either learned using dictionary learning or selected from the previous *pure drift* impedance responses that correspond to injections[1] of only buffer (containing no target) into the test chambers.

A. Similarity of Drift across Injections

We first investigate the similarity of pure drift data (i.e., without the signal component) by using our recently developed information theoretic measure presented in [5]. This measure accounts for the temporal structure of the data and allows us to calculate the pair-wise directed information (DI) between the drift in two channels from the second and the third injections and plot a *similarity matrix* depicted in Fig. 2.

Fig. 2 shows that in many channels, the drift shares a strong similarity with other drift in different channels in the adjacent injection. High statistical similarity in the adjacent injections allows us to estimate and remove drift component in the impedance response.

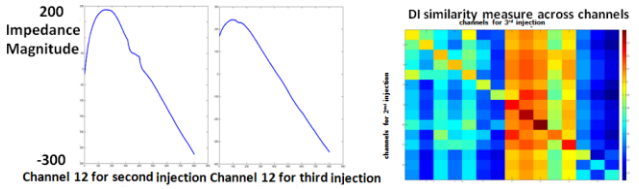

Figure 2. Visual illustration of drift similarity across injections. Left panel demonstrates an example of the impedance magnitude in channel 12 for the 2nd and the 3rd injections. Right panel plots the similarity matrix of the pair-wise directed information (DI) across channels from the 2nd injection to the 3rd injection (red color indicates high similarity).

B. Positive Lasso Optimization

The *LASSO* algorithm was originally proposed by Tibshirani et al. in [6] as a least absolute shrinkage and selection operator. As a variant of *LASSO*, *Positive LASSO* aims to solve a similar optimization problem using l_1 regularization with the additional constraint that the weights be non-negative. Here we propose to formulate the drift prediction as a l_2-l_1 optimization similar to Positive LASSO [7]. Denote the vector *y* as the observed binding signal in a single channel. This signal vector is of size $t \times 1$, where *t* is the number of samples. Define the matrix $X = [X_1, X_2]$, where X_1 is regarded as the set of drift basis vectors of size $t \times m_1$ (where m_1 is the number of drift basis vectors), and X_2 is regarded as the set of signal basis vectors of size $t \times m_2$ (where m_2 is the number of signal basis vectors). The vector $w^T = [w_1, w_2]^T$ is of size $(m_1 + m_2) \times 1$, where w_1 and w_2 represent the weights on the drift basis and the signal basis respectively. The parameter λ controls the sparsity of the weights. Drift prediction can be formulated as an optimization problem as

$$\arg\min_{w} \| y - Xw \|_2^2 + \lambda \| w \|_1 \qquad (3)$$

where $w > 0$. In other words, the optimization jointly minimizes the l_2-norm of the prediction error and l_1-norm of the weight vector. The implementation employs the interior method for l_1 regularized least squares as described in [7]. As depicted in Fig. 3, our drift prediction algorithm searches for the optimal weights on the signal and drift basis vectors so that the error between the observed binding signals and the weighted sum of signal and drift basis vectors is minimized, subject to a non-negativity constraint and a sparsity constraint on the weights. Sparsity ensures that the number of non-zero weights is very small compared to the total number of weight coefficients. Given the optimized weights $\hat{w} = [\hat{w}_1, \hat{w}_2]$, the predicted drift component can be computed as $\hat{y}_d = X_1 \hat{w}_1$ and the predicted signal component can be estimated as $\hat{y}_s = X_2 \hat{w}_2$.

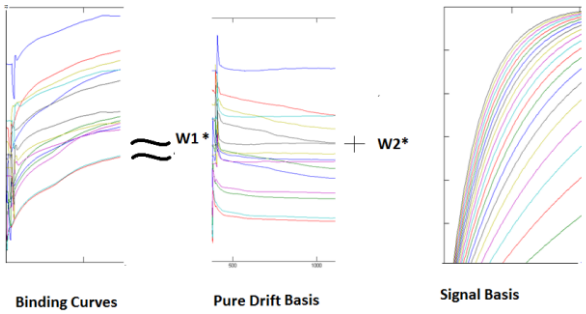

Figure 3. Decomposing binding signals into drift components and signal components, where binding curves are the data under analysis, the pure drift basis is selected using prior experimental data and the signal basis is constructed using the exponential model.

Since an approach based on *dictionary learning* is not suitable for online computation, the drift basis X_1 here is defined to be the 15 drift responses[2] from the previous buffer injection. The signal basis X_2 is constructed according to Eq. 1 by varying *s* in the interval [0.003, 0.043] with stepsize of 0.002, and *A* is selected to be 4000 or 2000. The size of the signal basis is 42. The values of *A* and *s* are chosen according to the typical ranges of *A* and *s* encountered from many prior assays using the biosystem. Theoretical analysis and experimental results show that including additional signal

[1] Solution is injected into the sensor test chambers using a pipette.

[2] As discussed in [1], our biosensor array is composed of 15 channels.

basis vectors does not necessarily increase the accuracy of prediction and will increase the computational burden. Fig. 4 and Fig. 5 show example results of drift prediction.

III. Experimental Results

Performance Evaluation: In order to demonstrate the effectiveness and efficiency of the proposed drift prediction algorithm, we first validate the algorithm using synthetic data where the ground truth of the signal component and the drift component are available. For synthetic data, we evaluate the root mean square error (RMSE) of the binding curves, RMSE of the estimated signals and RMSE of the estimated drifts. For each channel i, denote the total number of samples as N, the observed binding signal as $y(i)$ and the predicted binding signals as $\hat{y}(i)$. RMSE of the binding signal is defined as:

$$RMSE_{binding} = \sqrt{\frac{\sum_{i=1}^{N}[(\hat{y}(i)-y(i))/\max(y(i))]^2}{N}}$$

The RMSE of the estimated signals and the estimated drifts are defined in a similar manner by replacing $\hat{y}(i)$ with $\hat{y}_s(i)$ or $\hat{y}_d(i)$, and replacing $y(i)$ with $y_s(i)$ or $y_d(i)$ as the case might be. In our experimental results, N is selected to be the minimum of the lengths (in samples) of the observed binding signals and the previous drift basis. To further assess the robustness of our system in terms of the ability to separate the signal from the drift, we define the signal-to-drift ratio (SDR) in a manner similar to the signal-to-noise ratio (SNR) as $SDR = 10\log_{10}\frac{E_{signal}}{E_{drift}}$, where E_{signal} and E_{drift} characterize the energy of the signal and the drift respectively.

Validation on Synthetic Data: For real data, the ground truth of the signal component and the drift component are not available. Therefore, we first conducted two sets of validation experiments with synthesized signals to demonstrate the effectiveness of the proposed algorithm. For all the experiments, λ was selected in the interval [1.5, 2] which achieves the best performance, and the number of iterations required for convergence was set to 600. In the first set of validations, the binding signals are represented as the summation of pure drift signals [3] and pure exponential signals. We generated the ground truth binding signals by choosing *A=3500, s=0.0076* in the exponential function. Notice that the selection of s is not contained in the signal library in order to evaluate the robustness of the proposed algorithm. The ground truth drift component employed to construct the artificial binding curve was selected from a set of experimental buffer-only assays under stimulus conditions of 50mV and 960Hz, and a sensor temperature of 37 °C. The drift basis was selected from a different set of experimental buffer-only assays under operating conditions of 100mV, 24Hz and 42 °C. We evaluated the prediction performance of these synthetic data and plotted in Fig. 6 the signal weights against signal library index. As shown there, the four non-zero signal weights correspond to (*A=4000, s=0.007*), (*A=4000, s=0.009*), (*A=2000, s=0.007*) and (*A=2000, s=0.009*). Accurate estimation of the parameter values, compared to the ground truth, demonstrates the effectiveness of the proposed approach.

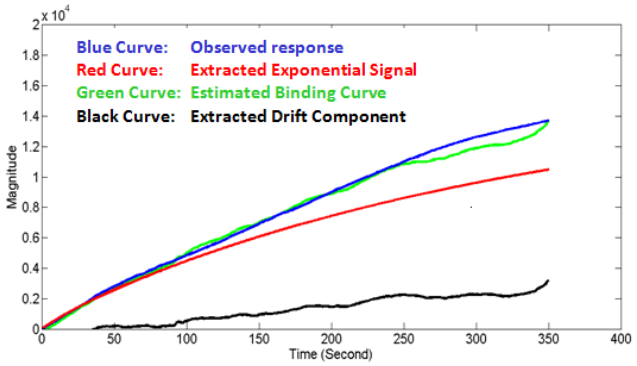

Figure 4. Drift estimation example result including the observed response of the binding signals (blue), the extracted exponential signal (red), the estimated binding curve (green) and the extracted drift component (black).

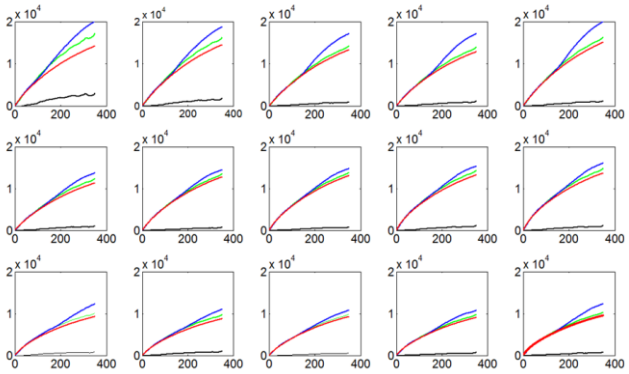

Figure 5. Drift predictions for 15 channels, demonstrating good prediction performance consistently (legend of the figure is the same as Fig. 4).

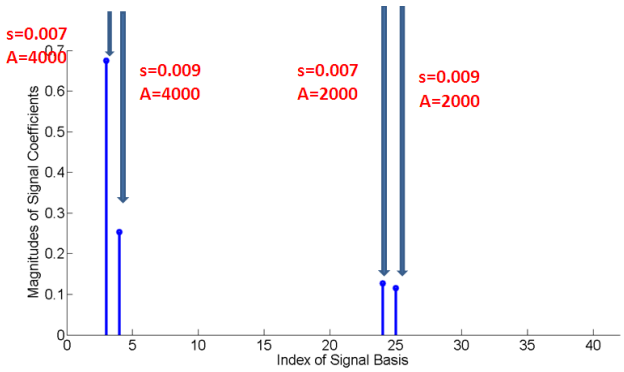

Figure 6. Plot of magnitude of signal weights versus index of signal basis.

In a second validation experiment, the binding signals are represented as the summation of a mixture of two exponential signals, a pure drift component, and additive zero mean

[3] Said signals were obtained from previous assays using buffer solution.

Gaussian noise. We varied the variance of the noise and plotted in Fig. 7 the RMSE of the signal component and the drift component versus SNR, ranging from 15dB to 35dB. Fig. 7 indicates that the prediction algorithm is robust to noise. The RMSE varied from 8% to 5.5% when SNR was varied from 15 dB to 35 dB. Moreover, the noise on the signals did not affect the prediction performance on drift. That is, the signals and drifts are well separated by the proposed prediction algorithm. The analysis of the RMSE versus the signal-to-drift ratio (SDR) shown in Fig. 8 indicates that when the SDR is larger than 15dB, the prediction algorithm performs very well for drift removal.

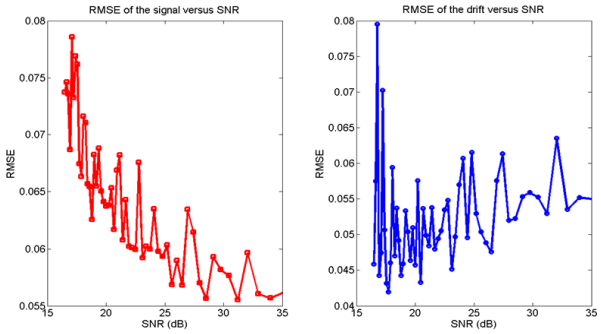

Figure 7. The RMSE versus the SNR for the estimated signal component (left) and the estimated drift component (right).

Experimental Results on Real Data: Evaluating on real data, we calculated the RMSE of the predicted binding signals over 350 trials and plotted the distribution of RMSE in Fig. 9. As shown there, for about 75% of the data, the RMSE of prediction using the proposed algorithm was below 2 %. Only about 3% of the data had an RMSE of prediction larger than 5%.We experimented with several other drift removal methods such as Discrete Wavelet and Fourier representations, and found our method to be superior. Other proposed methods such as polynomial fitting and detrending do not apply to our problem setting.

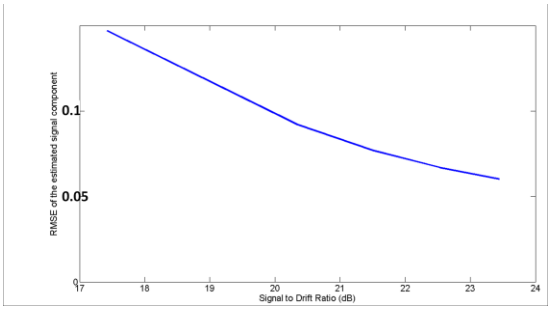

Figure 8. The RMSE of the estimated signal component versus the signal to the drift ratio (SDR).

The high accuracy of the proposed approach can be attributed to the fact that the proposed algorithm does not rely on any parametric model and is capable of choosing the optimal weights that minimize the prediction error given the high similarity among drift basis vectors. The 3% of impedance data with the RMSE of prediction larger than 5% can be interpreted as being outliers that are dissimilar to any of the linear combinations of the previous drift basis. The removal of these outliers to further enhance prediction performance is future work. The MATLAB implementation of the proposed algorithm takes only 1 minute on average to predict the drift from all 15 channels of impedance data using 400 samples for each signal.

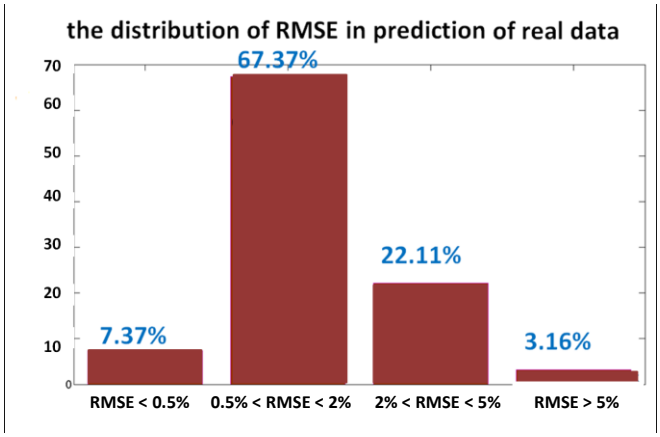

Figure 9. Distribution of prediction RMSE over 350 trials of real data.

IV. Conclusion

We have proposed a novel machine learning framework that is efficient and effective at predicting and removing the various patterns of unknown drift that often occur in affinity biosensor signals, thus improving their quantification as described in [1]. The method works by decomposing the binding curves into a drift component and a signal component. The algorithm requires no training and can therefore operate in real-time---a key feature of [1]. Validation with synthetic data demonstrates the robustness of our scheme, and results using real biosensor data show that the proposed approach achieves high prediction accuracy over a large dataset of assays.

References

[1] D. S. Messing, A. Ghindilis and K. Schwarzkopf, "Impedimetric biosignal analysis analysis and quantification in a real-time biosensor system," in *32nd Annual International Conference of IEEE EMBS, 2010* .

[2] D. S. Messing, A. Ghindilis and K. Schwarzkopf, "An improved algorithm for quantifying real-time impedance biosensor signals," in *33rd Annual International Conference of IEEE EMBS, 2011.*

[3] U. Bertocci, F. Huet, R. P.Nogueira and P.Rousseau, "Drift removal procedures in the analysis of electrochemical noise", in *Corrosion 2002, vol 58.*

[4] R.F.von Borries, J.H. Pierluissi and H.Nazeran, "Wavelet transform-based ECG baseline drift removal,"*IEEE EMBS 2005.*

[5] X. Chen, Z. Syed and A. Hero, "EEG spatial decoding with shrinkage optimized directed information assessment,"*in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012.*

[6] R. Tibshirani, "Regression shrinkage and selection via the lasso", *in Journal of Royal Statistical Society, 1994, vol 58.*

[7] S. J.Kim, K. Koh, M. Lustig, S.Boyd and D. Gorinevsky, "An interior-point method for large scale l_1 regularized least squares, *in IEEE Journal on Selected Topics in Signal Processing, 2007, vol 1.*