Bootstrapping Adaptive Linear Neuron in Near Infrared Spectroscopic Analysis

Kim Seng Chia
Faculty of Electrical and Electronic Engineering,
Universiti Tun Hussein Onn Malaysia,
86400 Parit Raja, Batu Pahat, Malaysia
kschia@uthm.edu.my

***Abstract*—Near infrared spectroscopic analysis requires a predictive model to extract relevant information from a complex near infrared spectral data so that the internal composition of products can be measured indirectly. Even though ensemble models show a better predictive performance compared to that of a single model in most cases, the computational cost will be multiplied for building multiple models. Besides, a combination of several different sub-models causes an ensemble model to be much complex than a single model. Thus, this study proposes the bootstrapping adaptive linear neuron (Adaline) that adapts the philosophy of bootstrapping aggregation approach. Without changing the architecture of an Adaline, the results indicate that the proposed the bootstrapping Adaline is promising to achieve a better performance than an Adaline with an average 18.6% improvement. This suggests that the bootstrapping algorithm is promising to enhance the predictive accuracy of the Adaline model in near infrared spectroscopic analysis.**

***Index Terms*—Adaptive Linear Neuron; Bootstrapping; Near Infrared Spectroscopic Analysis.**

I. INTRODUCTION

Near infrared spectroscopic analysis aims to extract relevant information from a highly collinear and over-lapping near infrared spectral data so that the internal composition of products can be non-destructively predicted [1, 2]. This is because the near infrared spectrum can be acquired non-destructively from a given sample that is either solid or liquid. Recent technology development has reduced the size and cost of near infrared spectrometers. However, the development of better predictive models is still an area that both researchers and industries concern about so that this technology can be applied in new applications [3].

Ensemble modelling approaches have attracted the attention of researchers as new approaches to enhance the predictive accuracy of a predictive model. Generally, there are two steps in producing an ensemble model. First, several divert sub-models are created. Second, the outputs of these sub-models are aggregated to produce a single output. For example, bootstrap aggregating (also known as Bagging) combines the outputs of multiple sub-models that are created using bootstrapping resampling strategy as its output. Bootstrapping algorithm can provide reliable point and interval estimation without any assumption about probability distributions and is applicable even for small data size [4]. Even though ensemble models show a better predictive performance compared to a single model in most cases, the computational cost will be multiplied for building multiple models [5-9]. Additionally, it is harder to understand and interpret the data using an ensemble predictive model because the architecture of the model is much more complex than that of a single model.

Adaptive linear neuron (Adaline) has been widely considered as a single layer artificial neural network. Unlike artificial neural network that contains lots of tunable parameters, Adaline only consists of two tunable parameters, i.e. the learning rate and the number of adaptation cycles. The former is to ensure the stored information is only disturbed in the smallest extent possible during training session according to the minimal disturbance principle [10]. The latter is to make sure the Adaline has been trained sufficiently without under-fitting or over-fitting problem. Using optimal learning rate and sufficient adaptation cycles, Adaline coupled with linear transfer functions and least mean square (LMS) learning algorithm has been found to be able to extract useful information for quantitative analysis and variable reduction in a near infrared spectroscopic analysis [11]. This could be due to the involvement of LMS algorithm that is capable of extracting useful information from complex signals [12].

This study aims to investigate the possibility to aggregate the basic ideas of Bagging approach and Adaline into a single model. The main idea of the former is that the strengths of a number of divert models that created using different bootstrapping sample sets can be aggregated to produce a better model. The latter, on the other hand, states that it is crucial to ensure that each learning session has a minimum effect on previous training pattern, i.e. minimal disturbance principle. After that, this paper evaluates the performance of the proposed model using an experimental near infrared spectroscopic data.

II. MATERIAL AND METHODS

In this study, firstly, the raw data was analyzed (Part A), and pre-processed (Part B). After that, the relationship between the soil organic matter (SOM) and its respective near infrared (NIR) spectral data was modeled using the proposed model by means of Adaline coupled with the Bootstrapping algorithm (Part D). Lastly, the predictive accuracy of the proposed Bootstrapping Adaline was compared with the Adaline without the proposed strategy (Part C).

A. Spectral Data and Component of Interest

A total of the 108 near infrared (NIR) spectral data of the soil samples that measured in Abisko, Northern Sweden (681210N, 181490E) were used in this study (source: http://www.models.kvl.dk/datasets/). These NIR spectral data ranged from 700 to 2498 nm, with an interval of 2 nm. 72 spectral data were acquired from soil samples with a depth

of 0 to 5cm; while the remainders were acquired from soil samples with a depth of 5 to 10 cm [13]. The component of interest, i.e. the soil organic matter (SOM) of each soil sample, on the other hand, was measured using the loss on ignition test at 550 degrees Celsius [13].

B. Data Pre-processing

MATLAB (version R2009b, win64) was used to process and analyse the data in this study. First, all the data were randomly separated into two different data sets with an equal amount, i.e. 54 each. The range of a testing data set should within the range of the training data set so that extrapolation prediction can be avoided. Thus, the data set that had soil organic matter (SOM) from 42.91 to 95.85 % was chosen as the training dataset. While another data set, in which the SOM was from 44.11 to 95.52 %, was selected as the testing data set. The mean and standard deviation of training data were 85.27 and 11.15 %, respectively. The mean and standard deviation of testing data were 85.58 and 10.59 %, respectively.

Next, second order Savitzky-Golay (SG) derivative with a filter length of 34 nm was used to improve the signal-to-noise ratio of the spectral data. After that, the spectral data of the training set were normalized into a range of -1 and 1. The return parameter of the normalization was retained to pre-process the spectral data of the testing set.

The SOM values of training data, on the other hand, were normalized into a range of -1 and 1. The return parameter was used to post-process all predicted values into their normal scale so that the accuracy of a predictive model can be analyzed in the original scale.

C. Adaptive Linear Neuron

Adaptive linear neuron (Adaline) that coupled with Widrow-Hoff delta rule or least mean square (LMS) algorithm was used to predict the soil organic matter (SOM) using the near infrared (NIR) spectral data as its input signals. LMS algorithm updates its weights and bias with a minimal effect when a training sample in this study.

In this study, the output of the Adaline is the predicted SOM. The *k*-th predicted SOM, $\hat{y}_k$ is the dot product of the *k*-th input data, X_k and the trained weight vector, W, as that stated in Equation 1. The equations of W and X_k can be represented by Equation 2 and Equation 3, respectively, and X^T is the transpose of X vector.

$$\hat{y}_k = X_k^T W \tag{1}$$

$$X_k = [x_{k,0}, x_{k,1}, x_{k,2}, \cdots, x_{k,n}] \tag{2}$$

$$W = [w_0, w_1, w_2, \cdots, w_n] \tag{3}$$

The first predictor of the *k*-th input data is the bias input that is the product of the $x_{k,0}$ (i.e. a unity signal) and the w_0. The subsequent predictors (i.e. from $x_{k,1}$ to $x_{k,n}$) are the first to *n*-th wavelengths of the *k*-th spectrum, in which, *n* is the total number of wavelengths of a spectrum.

During the training process, firstly, the residue is estimated using Equation 4. After that, the weight vector is updated according to Equation 5, in which, μ that controls the magnitude of the change in each training process. These two steps are repeated using different samples until a convergence is reached. The learning rate of the Adaline in this study was 0.0001 that was identified in the previous study [11].

$$\varepsilon_k = y_k - X_k^T W_{k-1} \tag{4}$$

$$W_k = W_{k-1} + \mu \varepsilon_k X_k \tag{5}$$

D. Bootstrapping Adaptive Linear Neuron

Figure 1 illustrates the training process of the proposed bootstrapping Adaline. For the proposed model, Adaline will be trained using different bootstrap samples in each iteration or adaptation cycle. First, 54 bootstrap samples $D^b = \{(x_1^b, y_1^b), (x_2^b, y_2^b), \cdots, (x_{54}^b, y_{54}^b)\}$ were randomly selected with replacement from the training data set, in which, b = 1. Second, these data D^1 were used to train an Adaline using Equation 4 and Equation 5 once to perform one adaptation cycle. These two steps were repeated until the predictive performance of the Adaline is satisfied, i.e., $D^2 = \{(x_1^2, y_1^2), (x_2^2, y_2^2), \cdots, (x_{54}^2, y_{54}^2)\}$ were used for the second adaptation cycle training and so on. The learning rate used was same as that used by the optimized Adaline, i.e. 0.0001.

Since the bootstrap samples were randomly selected by means of bootstrapping with replacement approach, the validation analysis was repeated for five times so that the effect of using different random samples could be investigated.

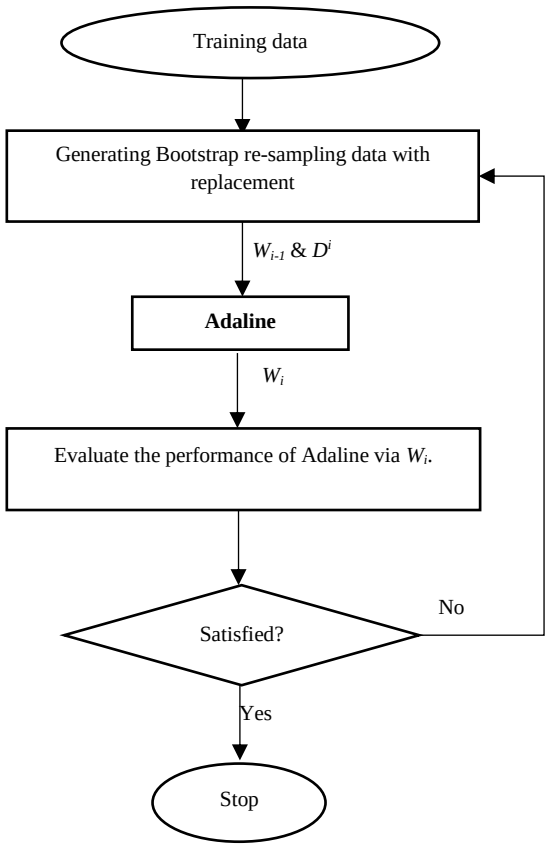

Figure 1: The learning process of the proposed bootstrapping Adaline

III. Results and Discussion

A. The Performance of the Proposed Model

Table 1 tabulates the performance of the proposed bootstrapping Adaline with five different trials. The optimal

root mean square error of prediction (RMSEP) of the bootstrapping Adaline ranges from 1.8980 to 2.0434%, with a mean and deviation of 1.9373 and 0.0600%, respectively. Although a quite similar optimal RMSEP was achieved, i.e. deviation = 0.06%, the optimal iteration varied from 366 to 1242. This suggests that the proposed model was capable of achieving similar optimal RMSEP when the model was trained optimally with sufficient adaptation cycles. For example, the proposed model was able to achieve the optimal RMSEP after it had been trained 366 times during the third trial. However, 1242 iterations were needed for the proposed model to achieve the optimal RMSEP during the fifth trial as that tabulated in Table 1. This could be due to the fact the training samples, i.e. bootstrap samples, for each adaptation cycle were different during different trials and adaptation cycles. Nevertheless, the similar predictive accuracy was achieved eventually for the five trials.

Table 1
The Performance of the Proposed Bootstrapping Adaline

Trial	Optimal Iteration	Training performance		Predictive performance	
		RMSEC (%)	r_c	RMSEP (%)	r_p
1	733	0.5205	0.9978	1.8980	0.9868
2	1077	0.4504	0.9976	2.0434	0.9868
3	366	0.8134	0.9947	1.9150	0.9843
4	887	0.5931	0.9977	1.9080	0.9865
5	1242	0.5074	0.9986	1.9220	0.9868

Figure 2(a) illustrates the RMSEP of the proposed bootstrapping Adaline during the five different training sessions for 5000 iterations. The results indicate that the RMSEP was highly dependent on the selected bootstrap samples in each iteration. Besides, the training session should be stopped immediately once the proposed model achieved its best accuracy. This is because the RMSEP of the proposed model might be worse if the training was continued with unwanted or redundant samples. This observation is in agreement with Lins et. al. (2015) who reported that the predictive accuracy may not be improved when the bootstrap replication is increased [4]. In other words, the model might be over-fitted if the training session did not stop at its optimal performance.

Figure 2(b), on the other hand, indicates the root mean square error of calibration (RMSEC) of the proposed Adaline was improving when the iteration was increasing. This observation is not in line with the trend of the RMSEP. This indicates that an over-fitting problem would happen if the training was excessive, i.e. too optimistic calibration accuracy (e.g. RMSEC) was achieved. Thus, a better indicator is needed to avoid the potential overfitting problem.

B. Bootstrapping Adaline vs. Adaline

In the previous study, the best Adaline that used the full NIR spectrum achieved RMSEC and RMSEP of 0.8859% and 2.3800%, respectively; and r_c and r_p of 0.9960 and 0.9860, respectively [11]. This shows that the proposed bootstrapping Adaline is promising in achieving an improvement of 18.6% in average. In other words, the involvement of the bootstrapping algorithm enhances the predictive accuracy of an Adaline. This could be due to the fact that only scarce samples were available in the present study, i.e. 108. The improvement may be due to the fact that the involvement of the bootstrapping algorithm could avoid suboptimal training performance that may happen during split sample or holdout validation approach [14].

Even though there were 899 input variables from each NIR spectrum, the proposed bootstrapping Adaline appears to be able to avoid over-fitting problem during a training session. This is because the proposed model achieved much better RMSEC (i.e. between 0.4504 and 0.8134%) than the Adaline (i.e. 0.8859%). This indicates that the training performance (i.e. RMSEC) of the proposed bootstrapping Adaline was much optimistic than that achieved by the Adaline. Surprisingly, this optimistic performance did not cause an over-fitting problem. This is because the proposed model was able to achieve a better predictive accuracy in terms of RMSEP. The result suggests that the proposed bootstrapping Adaline is less susceptible to the number of input variables.

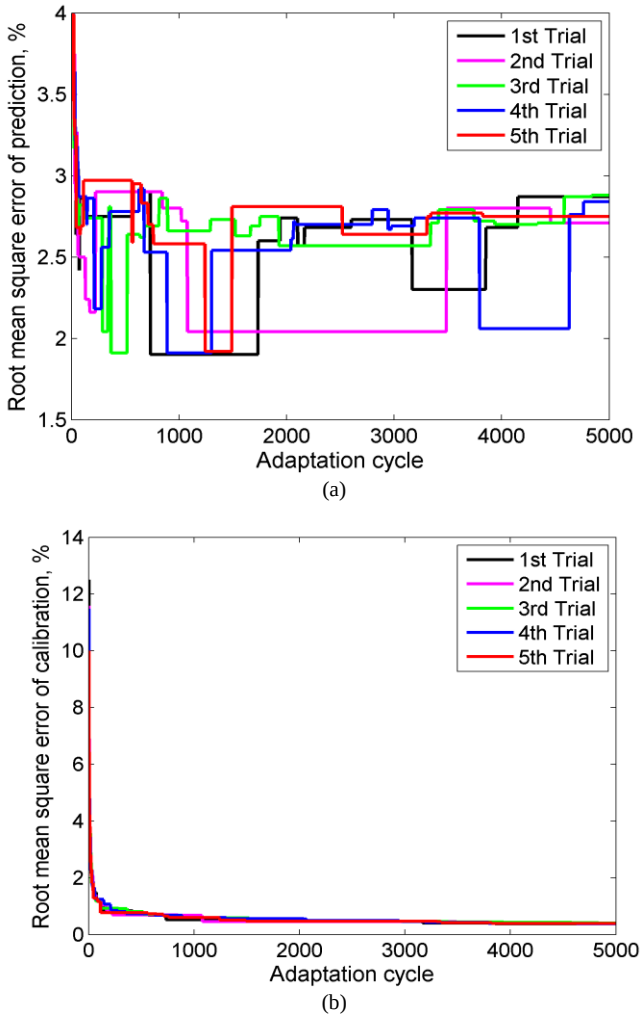

Figure 2: The Root Mean Square Error (RMSE) of the proposed bootstrapping Adaline: (a) prediction, and (b) training.

Nevertheless, the involvement of the bootstrapping algorithm causes the Adaline to be susceptible to the adaptation cycles. Since no similar sign of over-fitting from RMSEC during training process as that illustrated in Figure 2 (b), a conventional early stopping or regularization is hard to be applied to train the model optimally based on training performance. This could be due to the use of different training samples in each iteration. Thus, a better strategy is needed so that the proposed model can always achieve its best performance without over-fitting issues.

IV. CONCLUSION

The proposed bootstrapping Adaline is able to predict the SOM from near infrared spectral data with the root mean square error of prediction (RMSEP) that ranges from 1.8980 to 2.0434%, with a mean and deviation of 1.9373 and 0.0600%. This result shows that without changing the architecture of an adaptive linear neuron (Adaline), the proposed bootstrapping Adaline that adapts the philosophy of bootstrapping aggregation approach is capable of achieving a better performance than Adaline with an average of 18.6% improvement in predicting the soil organic matter (SOM) by means of near infrared spectroscopic analysis. In other words, bootstrapping algorithm is promising to enhance the predictive accuracy of the Adaline model in near infrared spectroscopic analysis.

In future, several optimization strategies will be investigated to ensure the training process of the proposed bootstrapping Adaline will be stopped at its optimal state.

ACKNOWLEDGMENT

This study was sponsored by the Research and Innovation Fund provided by the Office for Research, Innovation, Commercialization and Consultancy Management (ORICC), Universiti Tun Hussein Onn Malaysia (UTHM). Authors would like to acknowledge UTHM and Advanced Mechatronic Research Group (AdMiRe), FKEE, UTHM for providing facilities for this study.

REFERENCES

[1] L. Zhang, H. Xu, and M. Gu, "Use of signal to noise ratio and area change rate of spectra to evaluate the Visible/NIR spectral system for fruit internal quality detection," Journal of Food Engineering, vol. 139, pp. 19-23, 2014.

[2] L. Pan, Q. Zhu, R. Lu, and J. M. McGrath, "Determination of sucrose content in sugar beet by portable visible and near-infrared spectroscopy," Food Chemistry, vol. 167, pp. 264-271, 2015.

[3] K. Wiesner, K. Fuchs, A. M. Gigler, and R. Pastusiak, "Trends in Near Infrared Spectroscopy and Multivariate Data Analysis From an Industrial Perspective," Procedia Engineering, vol. 87, pp. 867-870, 2014.

[4] I. D. Lins, E. L. Droguett, M. d. C. Moura, E. Zio, and C. M. Jacinto, "Computing confidence and prediction intervals of industrial equipment degradation by bootstrapped support vector regression," Reliability Engineering & System Safety, vol. 137, pp. 120-128, 2015.

[5] K. Wang, T. Chen, and R. Lau, "Bagging for robust non-linear multivariate calibration of spectroscopy," Chemometrics and Intelligent Laboratory Systems, vol. 105, pp. 1-6, 2011.

[6] A. Ukil, J. Bernasconi, H. Braendle, H. Buijs, and S. Bonenfant, "Improved Calibration of Near-Infrared Spectra by Using Ensembles of Neural Network Models," Sensors Journal, IEEE vol. 10, p. 7, 2010

[7] Y. Hu, S. Peng, J. Peng, and J. Wei, "An improved ensemble partial least squares for analysis of near-infrared spectra," Talanta, vol. 94, pp. 301-307, 2012.

[8] X. Pan, Y. Li, Z. Wu, Q. Zhang, Z. Zheng, X. Shi, et al., "A Online NIR Sensor for the Pilot-Scale Extraction Process in Fructus Aurantii Coupled with Single and Ensemble Methods," Sensors, vol. 15, p. 8749, 2015.

[9] Z. Li, J. Lv, G. Si, Y. Zhang, Q. Wang, and S. Liu, "An improved ensemble model for the quantitative analysis of infrared spectra," Chemometrics and Intelligent Laboratory Systems, vol. 146, pp. 211-220, 2015.

[10] B. Widrow and M. A. Lehr, "30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation," Proceedings of the IEEE, vol. 78, pp. 1415-1442, 1990.

[11] K. S. Chia, "Adaptive linear neuron in visible and near infrared spectroscopic analysis: predictive model and variable selection," ARPN Journal of Engineering and Applied Sciences, vol. 10, pp. 9055-9059, 2015.

[12] P. R. Diniz, "The Least-Mean-Square (LMS) Algorithm," in Adaptive Filtering, ed: Springer US, 2008, pp. 1-54.

[13] R. Rinnan and Å. Rinnan, "Application of near infrared reflectance (NIR) and fluorescence spectroscopy to analysis of microbiological and chemical properties of arctic soil," Soil Biology and Biochemistry, vol. 39, pp. 1664-1673, 2007.

[14] E. W. Steyerberg and F. E. Harrell, Jr., "Prediction models need appropriate internal, internal-external, and external validation," Journal of Clinical Epidemiology, vol. 69, pp. 245-247, 2016.