31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Nonlinear Dimensionality Reduction of Electroencephalogram (EEG) for Brain Computer Interfaces

Mohammad Nayeem Teli and Charles Anderson

***Abstract*— Patterns in electroencephalogram (EEG) signals are analyzed for a Brain Computer Interface (BCI). An important aspect of this analysis is the work on transformations of high dimensional EEG data to low dimensional spaces in which we can classify the data according to mental tasks being performed. In this research we investigate how a Neural Network (NN) in an auto-encoder with bottleneck configuration can find such a transformation. We implemented two approximate second-order methods to optimize the weights of these networks, because the more common first-order methods are very slow to converge for networks like these with more than three layers of computational units. The resulting non-linear projections of time embedded EEG signals show interesting separations that are related to tasks. The bottleneck networks do indeed discover nonlinear transformations to low-dimensional spaces that capture much of the information present in EEG signals. However, the resulting low-dimensional representations do not improve classification rates beyond what is possible using Quadratic Discriminant Analysis (QDA) on the original time-lagged EEG.**

I. Introduction

Electroencephalogram is the measurement of the electrical activity of the brain measured by placing electrodes on the scalp. These EEG signals give the micro-voltage difference between different parts of the brain in a non-invasive manner. Interpretation of these EEG waves is important to find ways to utilize them for a BCI. People severely disabled by Amyotropic Lateral Sclerosis (ALS), brain stem stroke, cerebral palsy, and other neuromuscular disorders would greatly benefit by advances in BCI research [1].

However, Parra et al. [2] indicate that the EEG signals need to be tailored for BCI application because of their low signal to noise ratio. Gramfort et al. [3] report that a non-linear dimensionality reduction technique can provide a better understanding of the EEG signals. One of the well known non-linear methods for dimensionality reduction is a bottleneck NN [4]. Devulapalli [5] used such a network for dimensionality reduction and classification of EEG data with good success in classification accuracy.

The above research suggests that EEG data do require dimensionality reduction before they could be classified. This is particularly important because of four main reasons: low EEG signal-to-noise ratio, varying number (6 to 256) of electrodes, possible correlation as signals propagate through the brain volume, and possible correlation among channels. The hope is that dimensionality reduction might be able to remove most of the noise and the correlations between the various channels and be able to detect features that are most viable for an accurate classification. We hypothesize that a multi-layer bottleneck neural network would help classify EEG data through dimensionality reduction, with higher accuracy, than its classification without dimensionality reduction.

The EEG for this study was observed in subjects performing two mental tasks. A subject is asked to think about a task, the EEG signals are measured and these signals are used to classify the tasks.

Mohammad Nayeem Teli is a graduate student and Charles Anderson is the faculty with the Department of Computer Science, Colorado State University, Fort Collins, CO 80526,USA
{nayeem,anderson}@cs.colostate.edu
This work was supported in part by the NSF Grant No. 0208958.

EEG data classified with dimensionality reduction will henceforth be called reduced EEG. We compare the classification using NNs with other techniques like Linear Discriminant Analysis (LDA) [6], QDA and Support Vector Machine (SVM) [7]. The classification is carried out on both EEG as well as reduced EEG data.

We chose NNs for dimensionality reduction and classification because of their popularity as a non linear approach in EEG research [8]. The choice of other methods was based on simplicity of approach for LDA and proven success for SVMs [9]. QDA was primarily chosen because of the dearth of classifying EEG using it and an intuition that it might be a good classifier for EEG.

II. Approach

We use a five layer bottleneck NN, trained to approximate the identity map $f(x) = x$. Our network has three hidden layers with a middle bottleneck layer having fewer hidden units. The transfer functions of the units of hidden layers are sigmoid functions. The output layer has a linear transfer function. The NN is trained using two algorithms: Scaled Conjugate Gradient (SCG) [10] and Fast Convergence Algorithm based on Levenberg-Marquardt (FCALM) [11]. Our goal prior to classification is to reduce the dimensionality of the data at the bottleneck layer. This goal is reached when the data at the output layer of the NN matches with the input data. The objective is achieved by minimizing the Mean Square Error (MSE) between the i-th multi channel EEG sample, x_i, and the output of the NN, $f(x_i)$, represented as:

$$MSE = \frac{\sum_{i=1}^{N}(f(x_i) - x_i)^2}{N} \tag{1}$$

where $i = 1, \ldots, N$ and N is the total number of observations. The lower dimensional bottleneck layer output data, reduced EEG, is classified using a separate three layer NN classification network, LDA, QDA and SVM approaches and the results are compared.

A. Electroencephalogram(EEG) Data

In this research we use the data collected by Keirn et al. [12], using the 10-20 system of electrode placement, shown in Figure 1. This is a standardized system of measuring EEG based on the

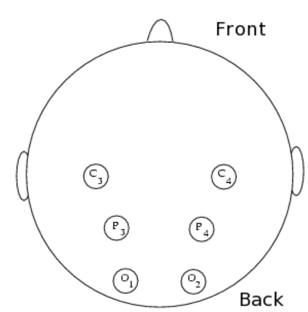


Fig. 1. Top view of the 10-20 system of electrode positions for EEG.

position of the electrodes on the scalp. Each letter represents the location of the brain with the number indicating the hemisphere. The letters C, P, and O represent central, parietal and occipital parts of the brain. The data for this study was measured at six sites: $C3$, $C4$, $P3$, $P4$, $O1$, and $O2$. The data is stored at a frequency of 250 samples per second. Sets of ten second trials were recorded for each of two mental tasks: letter writing and mental arithmetic. For the imagined letter writing task the subject is asked to compose a letter. In subsequent trials the subject is asked to resume the letter from where the previous trial had left off. In the mental arithmetic task each subject is asked to multiply two multidigit numbers, with the numbers being different in the different trials. The subjects were asked not to vocalize either task.

1) Data Partitioning: The data is observed on one subject. Each task has ten trials. Each trial consists of 2500 observations per session with each session lasting for ten seconds. A set of EEG data consisted of six dimensions (corresponding to the six channels), with 2500 samples for each dimension. A separate seventh channel, the eye-blink channel, is used to record eye-blink information. A high potential spike in the eye-blink channel ($>$100 μ Volts) lasting up to 10 milliseconds is considered as an eye-blink. The sample points corresponding to all six channels which fall in the region of eye-blinks are removed. Our training dataset has 20000 observations consisting of trials 1 to 4 of each task. The test dataset is the fifth trial for each task and therefore 5000 samples.

2) Time Embedded EEG: The 6-channel EEG data was embedded in time by augmenting each EEG sample with past samples. This enables a classifier to use fine temporal variations among channels. The embedding was carried out by combining multiple samples together, thereby increasing the dimensionality of the neural network input dataset. A time embedding dimension of 1 means that the dataset retained its dimensionality of 6. A dimension of 2 means that the dimensionality was changed from 6 to 12, a dimension of 3 implies the new dimensionality of the dataset is 18, and so on. Training as well as test datasets lag by equal amounts.

B. Bottleneck Algorithm Parameters

We varied the number of bottleneck units from 1 to 6, 10 and 20 with 30 units in the hidden layers on either side of the bottleneck layer. Using this setup we ran our final set of experiments using both SCG and FCALM training algorithms.

From pilot experiments we determined that the SCG algorithm worked well with 500 iterations. The algorithm was terminated either after 500 iterations or if the second derivatives of the location in the weight search space reached the machine precision, whichever happened sooner. The search direction was updated using Polak-Ribiere formula or we would restart in the direction of negative gradient after a fixed number of steps determined through pilot experiments. The initial search direction was chosen to be the negative of the gradient with lower bound on scale to 1.0×10^{-15} and the upper bound on scale as 1.0×10^{100}.

For FCALM, the maximum number of iterations was again chosen through pilot experiments and set to 200. The algorithm was terminated either at the end of the maximum number of iterations or when the machine precision was reached. The initial weight multiplication factor was chosen based on values reported by Wilamowski et al. [11]. If the MSE increased in an epoch, the learning factor of the algorithm would be decreased by a tenth and if the error decreased, the learning factor would be multiplied by a factor of 10 to move faster in that direction of weight change.

1) Training and Testing Classification Algorithms: Classification of EEG and reduced EEG are compared. The reduced EEG data was the output of the bottleneck layer obtained after training the whole neural network. This output was used as input data for classification. The classification network has a single hidden layer represented as a network like n-p-1, where n is the bottleneck layer output dimensionality representing the classification network input dimensionality, and p the hidden layer with 50 units. There is only one output with target values 0 and 1 for multiplication and letter writing tasks, respectively.

LDA, QDA and SVM algorithms were also used for classification of EEG and the reduced EEG data. The SVM algorithm used is a part of the e1071 library of 'R' statistical package. We used C-classification with a cost of 10 in our experiments. Rest of the algorithms were implemented by us.

III. Results and Analysis

We restricted the maximum number of bottleneck units to be 20 in all our experiments because the test MSE in the networks with more bottleneck units indicated over training. We also tested our data for different time embedding dimensions. FCALM algorithm could not handle time embedding dimensions greater than 20 (or total input dimension greater than $20 \times 6 = 120$), because it needs to compute the Hessian matrix. Therefore, we used time embedding dimension of 10 and 20 while training the NN with this algorithm. SCG was used to train the NN with time embedding dimensions of 10, 20, 50 and 100.

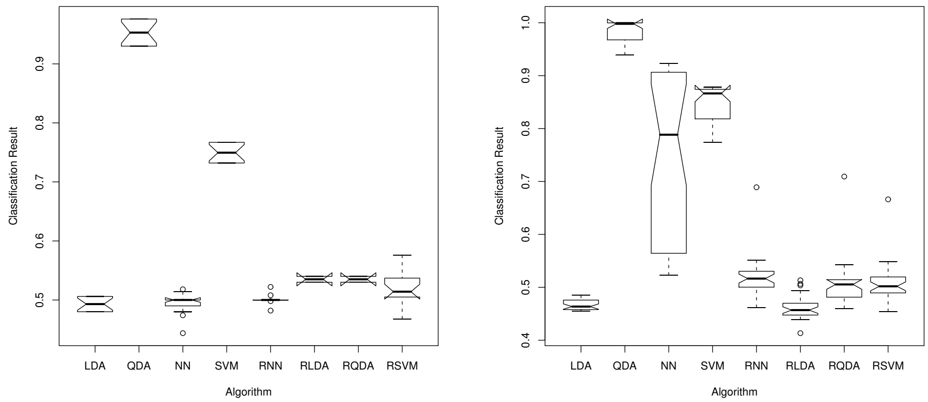
We will first report the best classification results of all the algorithms when the NN was trained using the FCALM algorithm. A boxplot comparison of these results is shown in Figure 2(a). The y-axis shows the correctly classified results on a scale of 0 to 1 with 1 being 100% correctly classified. QDA classification results on unreduced EEG data outperformed all other algorithms. The classification values shown by the boxplots are combined over time embedding dimensions of 10 and 20.

In order to understand the neural network behavior, we studied the test data MSE for different bottleneck units and the data lags after NN was trained using FCALM and SCG algorithms. The results showed that SCG was able to learn the weights between the hidden layers much better than the FCALM algorithm. It seems that although both algorithms learn equally well initially, SCG is able to fine tune the weight selection much better. Since SCG did better in terms of MSE reduction than FCALM, we investigated the data reconstruction at the output of our bottleneck NN, when trained using SCG, resulting in an MSE of 0.045.

The NN trained using SCG did better than the one trained using FCALM as shown in Figure 2(b). QDA performed the best again with the performance becoming better with the increase in the number of time embedding dimensions. The classification results for each algorithm are combined over the data lags of 10, 20, 50 and 100. Overall, the unreduced EEG data classification was better than the reduced dimensional EEG data across the algorithms.

It is important to have a closer look at the impact of time embedding dimensionality on the classification accuracy. Figure 3 represents reduced EEG data correctly classified when the time embedding dimensionality is varied. The classification accuracy is higher with lower time embedding dimensionality as shown in Figure 3(a). This figure also indicates that the higher the number of bottleneck units the better the classification accuracy of the reduced EEG. From this figure it appears that higher number of bottleneck units would get even higher accuracy, however that was untrue because the test RMSE was higher than the training RMSE for a higher number of bottleneck units, implying overtraining [13].

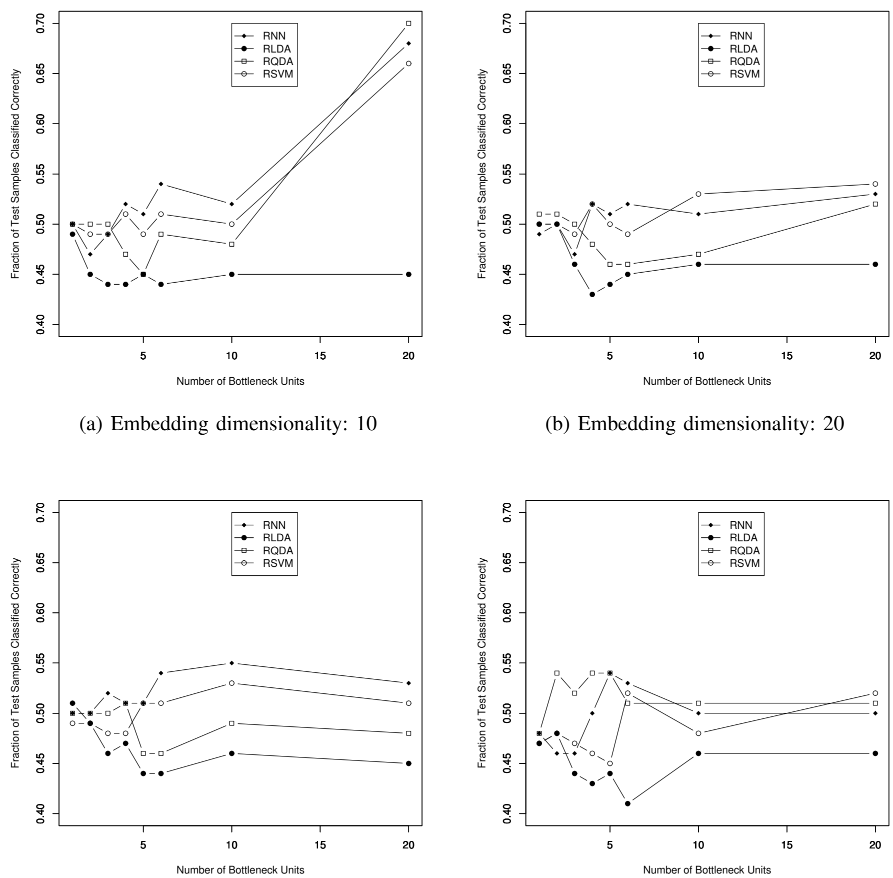
It appears that the reduced dataset can be classified with higher accuracy when the time embedding dimensionality is low. It also



(a)

(b)

Fig. 2. Classification results showing a fraction of test samples classified correctly by different algorithms. 2(a) neural network trained using FCALM algorithm; 2(b) neural network trained using SCG. R prefix implies the classification of the bottleneck output of the neural network using the corresponding algorithm. Absence of prefix means unreduced EEG classification.



(a) Embedding dimensionality: 10

(b) Embedding dimensionality: 20

(c) Embedding dimensionality: 50

(d) Embedding dimensionality: 100

Fig. 3. Each figure represents fraction of reduced EEG test samples classified correctly. Neural network is trained using SCG algorithm.

seems that dimensionality reduction and time embedding influence the classification accuracy in opposite ways. Total lack of dimensionality reduction combined with lower time embedded dimensions actually results in higher accuracy using NN, as supported by Figure 4. Neural networks' performance drops as the time embedding dimensionality increases and so does that of SVM. On the other hand the classification accuracy of QDA and LDA increases. LDA still shows the worst results, though. NNs do better than SVM

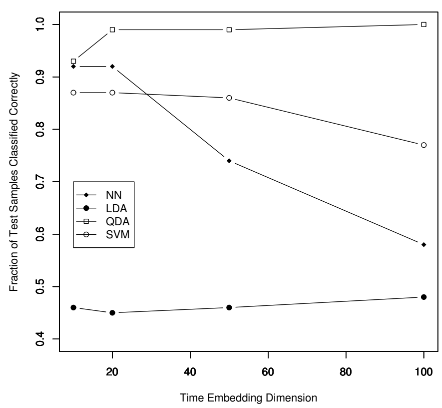


Fig. 4. Classification results of different algorithms showing fractions of test samples correctly classified for different time embedding dimensions on unreduced data. Neural network was trained using SCG algorithm.

for time embedding dimensionality below 30 with classification accuracy close to 90%. For a time embedded dimensionality of 10, NN shows approximately the same accuracy as QDA. Therefore, the key is to choose optimum time embedded dimensions.

It is pertinent to point out that despite poor classification accuracy of all the algorithms on the reduced EEG, the NN is able to mimic the input very well. This is displayed very well in Figure 5, which shows the reconstructed data output compared to the input dataset for the first channel with 20 bottleneck units. Similar plots for the

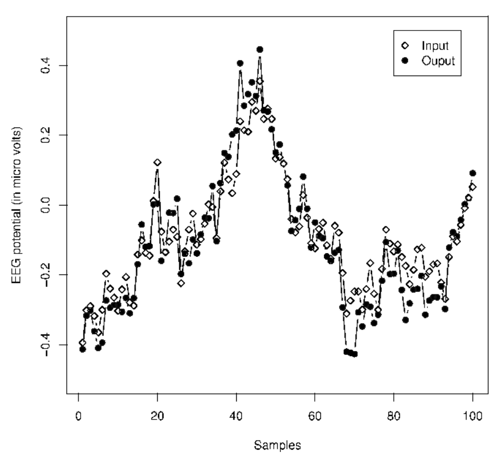


Fig. 5. The figure represents the reconstructed test data at the output of the bottleneck NN versus the input test dataset for an EEG channel.

rest of the channels support the argument that the reconstruction is fairly well. This makes it likely that the bottleneck layer is able to distinguish between the various input EEG signals. However, the success in data reconstruction is not getting translated into better classification accuracy of the reduced EEG data. On the contrary, unreduced EEG proved easier to classify efficiently, in general. It seems that the differentiability of the data is lost through dimensionality reduction. Among the algorithms, QDA successfully differentiated between the data features and classified the data much better than any other method.

IV. Conclusions

The results indicate that a bottleneck NN did not provide any advantage in classification. Overall, in all our experiment runs QDA performed the best followed by SVM. NN classifier performed slightly better than LDA and showed better accuracy in classifying unreduced EEG data than the reduced. NN was able to achieve its best classification accuracy of 92%, whereas QDA achieved 100% accuracy in classifying the test data.

This research was based on the hypothesis that a bottleneck NN would classify EEG data better than classification techniques like NN, QDA, LDA and SVM without the dimensionality reduction. However, QDA appears to have an outstanding performance on these data and does not require dimensionality reduction. Given the challenges associated with training a NN, QDA is a clear winner. We do not need to reduce the dimensionality of the EEG dataset. Temporal variations obtained by time embedding of the data, capture typically all the information that would lead to an excellent classification (consistently above 90%) using QDA. For a time embedding dimensionality below 20, NN also gives a classification accuracy of over 90% for the test dataset. For our setup of the NN and training algorithms, a bottleneck network might be a good choice for EEG classification only after more parameter tuning. Amongst the training algorithms, SCG is better in terms of the reconstruction MSE and the classification results.

References

[1] T. Vaughan and J. Wolpaw, "Guest Editorial, The Third International Meeting on Brain-Computer Interface Technology: Making a Difference," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, 2006.

[2] L. Parra, C. Christoforou, A. Gerson, M. Dyrholm, L. An, M. Wagner, M. Philiastides, and P. Sajda, "Spatiotemporal Linear Decoding of Brain State," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 107–115, 2008.

[3] A. Gramfort and M. Clerc, "Low Dimensional Representations of MEG/EEG Data Using Laplacian Eigenmaps," *Joint Meeting of the 6th International Symposium on Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging, NFSI-ICFBI.*, pp. 169–172, Oct. 2007.

[4] M. A. Kramer, "Autoassociative Neural Networks," *Computers and Chemical Engineering*, vol. 16, pp. 313–328, 1992.

[5] S. Devulapalli, "Non-Linear Principal Component Analysis and Classification of EEG During Mental Tasks," Master's thesis, Colorado State University, 1996.

[6] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009.

[7] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag New York Inc., 1995.

[8] F. Lotte, H. Mouchere, and A. Lecuyer, "Pattern Rejection Strategies for the Design of Self-Paced EEG-based Brain-Computer Interfaces," in *International Conference on Pattern Recognition, ICPR08*, Tampa, Florida, USA, 2008, pp. 1–5.

[9] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A Review of Classification Algorithms for EEG-Based BrainComputer Interfaces," *J. Neural Eng.*, vol. 4, no. 2, pp. R1–R13, June 2007. [Online]. Available: http://dx.doi.org/10.1088/1741-2560/4/2/R01

[10] M. F. Møller, "A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning," *Neural Networks*, vol. 6, pp. 525–533, 1993.

[11] B. M. Wilamowski, S. Iplikci, O. Kaynak, and M. O. Efe, "An Algorithm for Fast Convergence in Training Neural Networks," in *International Joint Conference on Neural Networks*, vol. 3, 2001, pp. 1778–1782.

[12] Z. Keirn and J. Aunon, "A New Mode of Communication Between Man and his Surroundings," *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 12, pp. 1209–1214, 1990.

[13] M. N. Teli, "Dimensionality Reduction and Classification of Time EmbeddedEEG Signals," Master's thesis, Colorado State University, 2007.