Minimizing Sensors for System Monitoring -
A Case Study with EEG Signals

Goutam Chakraborty
Department of Software & Information Science
Iwate Prefectural University, Japan
Email: goutam@iwate-pu.ac.jp

Shigeki Horie
Graduate School of Software & Information Science
Iwate Prefectural University, Japan
Email: g031j141@s.iwate-pu.ac.jp

Hikaru Yokoha
Graduate School of Software & Information Science
Iwate Prefectural University, Japan
Email: g231l034@s.iwate-pu.ac.jp

Zbigniew Kokosiński
Cracow University of Technology, Faculty of ECE,
Dept. of Automatic Control and Information Technologies,
ul. Warszawska 24, 31-155 Kraków, Poland;
Email: zk@pk.edu.pl

Abstract—For monitoring any system, be it a chemical plant,
a nuclear power station, or a human heart or brain, we need
to attach sensors and analyze the multivariate time-series data
collected by those sensors. If the system has two states, say
state 1 (good) and state 2 (bad), we need to infer which state
the system is, by classifying the collected time-series signals.
To make this work efficiently, it is important to search the least
number of probes that would give best classification result. It is
a multi-objective optimization problem. The proposed approach
works in two steps. We start with a large number of probes.
As the first step, we cluster the time-series signals, and choose
a representative one from each cluster. Next, we run pareto GA
to select the smallest set of probes (from cluster representatives),
that would give the highest classification result. Depending on
the nature of the signals, and the target application, appropriate
signal-features, clustering and classification algorithms will be
different, but the basic principle is applicable to any system.
In this paper, we tested the effectiveness of our algorithm with EEG
signals, to detect the presence or absence of ERP 300. Improved
results with less number of probes compared with previous works
validated the approach.

Index Terms—Ward’s Hierarchical Clustering, Dendrogram,
Electroencephalogram (EEG), Brain Computer Interface (BCI),
Pareto GA, Artificial Neural Network Classifier,

I. INTRODUCTION

While monitoring a system, sensors/probes are set at different
positions to collect on-line information which is analyzed
to monitor the system. For successful monitoring, the probes
have to be placed at proper locations, so that analysis of the
time-series data, received from them, will be able to detect
the target change with high accuracy. To start with, when the
proper locations are not known, a large number of probes are
needed to collect data from all possible points. As an example,
say we would like to find the speed of walking/jogging from
variation of foot-pressure at certain points of sole. Without any
a-priori knowledge, we need to put pressure sensors at many
points, and then identify the ones, signals from which would
give correct classification of walking speed. The motivation
of this work is to find the minimum number of such sensor
signals which would lead to the best target result. It is to be
noted that the problem is not identifying the time when such
change occurs. Here, we have a window of observation, we
assume the state is same during that period, and we need to
identify the state.

If signals from two probes are the same, over the window
of observation, we can use information from one probe and
discard the other, and reduce the number of required probes.
In the first step, we cluster the signals. In Fig. 1, we have
shown a synthetic example. Suppose the signals collected from
6 probes are as shown. When we cluster those signals, we
have 3 groups. We can select probe 1 or 5, and one from 2
and 6, etc., without reducing information collected by all the
six probes. In short, we select one signal from each group.

Going back to Fig. 1 again, our next question is whether
all the three signals, one from each cluster, are important?
If the aim is to look for a triangular shape of change which
is seen in cluster 1 when there is such change, and never
in other two clusters, it is sufficient to analyze signal from probe 1 only. For complex systems, it could be a subset of all clusters. If the number of clusters are large, finding the smallest subset which would give the best target result, is a multi-objective combinatorial optimization problem. We need to find the smallest subset, as well as the best classification result. These two requirements could be contradictory. There are algorithms to solve such problems. In this work we use multi-objective genetic algorithm (MOGA).

To demonstrate the effectiveness of the proposed method, we used electroencephalogram signals collected by 21 probes. In many BCI applications, the system works on the principle of detecting whether a particular positive even related potential, ERP300, is emanated from the brain or not. In other words, we need to classify whether the signals received by the probes do contain ERP300 or not - a classification task. A common application of Brain Computer Interface (BCI) is BCI Speller. Its working principle is simple. The subject chooses a letter, and it flashes on display at random intervals. When the chosen letter flashes, the subject’s brain is stimulated to generate ERP300 signal. This is because it happens occasionally, at random intervals, and the subject has to concentrate so that he does not miss it. When it happens, lot of neurons simultaneously generate a small positive potential which add up as ERP300, strong enough to be detected with probes attached to the skull. Due to noise and other instabilities, due to individual variations, to identify ERP300 signal, multiple probes are used at different positions of the skull. The robustness of the system is improved when multivariate signals are analyzed together.

In Fig. 2, we have shown the electrical potentials collected at 21 probes during a session, when the subject’s brain did respond with an ERP300 because the intended letter flashed. We also have EEG data when the other letters flashed, EEG data without ERP300. Thus, we have a supervised data set, and the task is classification. The problem here is to identify the minimum number of signals that would give highest accuracy of classification.

The rest the paper is as follows. In Section II, we briefly describe how the experimental data was acquired. In section III, we will describe the details of the algorithms used - preprocessing, clustering and selecting the signals for maximum classification result. In section IV, experimental results are described and compared with previous work. Section V is the conclusion and future plan.

II. DATA COLLECTION

To ensure the effectiveness of the proposed algorithm, to minimize the number of sensors for monitoring a system, we will use the human brain as the system and sensors are the EEG probes. This is a perfect example because EEG probes are large in number and the signals are noisy. In Brain Computer Interface (BCI) applications, EEG data of brain activities generated by controlled visual or auditory stimulation, serves as an interface between the human intention and the computer [2] [3]. Computer analyzes the EEG signals to decipher the intention of the user. Applications of this technique could be communication support to Systemic paralysis patient, such as spinal cord injury patients and Amyotrophic Lateral Sclerosis patients. This is useful for patients whose brain is working normally, but communication by speech and gestures are difficult due to paralysis of the body. Helping communication by using BCI is one important area of research and development.

We used 128 probe Net Station System 300 to collect EEG signals. The platform for the experiments was BCI speller. Four subjects participated in the experiment. The experimental setup is explained below. Our aim is to classify EEG signals in to two classes, and identify whether the signal contains ERP300 or not. The optimization criteria are

1) maximizing the percentage of correct classification and
2) minimizing the number of probes

A. Equipment used in the experiment

The BCI equipments used in our experiments is Netstation System 300 for Dense array EEG from Electrical Geodesics, Inc. (EGI) [4]. It is equipped with 128 channels. The specified noise level is 0.7 µV RMS. A/D conversion resolution is 24 bit. Maximum sampling rate is 2000 Hz. Input impedance is 200 MΩ. In our experiments, we used a sampling rate of 1000 Hz.

For our investigation, we need signals with and without ERP300. For generating such signals, we used BCI speller. The display of BCI speller is a 6 × 6 matrix, consisting of 26 characters and 10 numerals. They are organized in 6-rows and 6-columns, as shown in Fig. 3. All elements of a row or a column flash simultaneously. The subject targets one character at a time. When the array containing the character flashes, the subject needs to perform a task, like count up. Flashing of rows and columns are random. One needs to concentrate so as not to miss the target character. When the target character flashes, the brain responds strongly to the match. In BCI speller, there
is a training phase for the participants, call the "Copy Spell Phase." Next is the "free spell phase," when the subject decides a character in his mind. He then counts when the intended character flashes. The BCI speller identifies this unknown character by analyzing when P300 signals are generated.

Conventional BCI speller is a 8-probe equipment. Eight probes are placed at locations where ERP300 are strong for any shape of human skull in general. In our experiments, we used 128 channel Net Station System 300. The placement of 128 electrodes is shown in Fig. 4. Though it is possible to measure signals from all the 128 electrodes (high-density EEG), we selected 21 probes whose locations are recommended by the International 10-20 system [5] and shown in Fig 5. In international 10-20 system, electrodes are evenly spaced. Individual electrodes cover around 10% to 20% of the area of the head. Statistically, this distribution of electrodes is valid for anyone, irrespective of the head size. Thus we get data covering the whole area of the head, though with reduced spatial resolution. In future, we will use all 128 probes for a better resolution of probe location with expectation of better classification results.

B. Experimental set-up

In our experimental set-up, the subjects are provided with the word to spell, so that we get a labeled set of supervised samples. To communicate a single character or numeral, all the six rows and six columns flash for 10 times. The flashing occurs with random uniform probability distribution. The duration of the flash is 600 ms. Thus, the time required to communicate a single target character/numeral is 

\[(6+6) \times 10 \times 600 = 72\text{ Sec.}\]

Every time the target character is included in the row or column which is flashed, the subject is to count up. At the end, the total count should be 20, i.e., for a character presented we get 20 signals with ERP300 and 100 without. The subject concentrates, so that the flash of the target character is not missed.

C. Data collection

The aim of this work is to classify ERP300 and signals without, with high accuracy and less number of electrodes. We acquired EEG data from four subjects, all young males age between 20-23. Each subject’s data was separately analyzed. We compare the results with our previously reported work [8].

III. ALGORITHM DESCRIPTION

A simple flow diagram of the analysis procedure is shown in Fig. 6. EEG signals collected during the experiment are first separated into two sets, one consists of signals containing the ERP300, we call it set $\Xi$. The other set consists of signals without ERP300, we name this set as $\Phi$. Signals of set $\Xi$ is used for clustering, as a first step to reduce the number of probes. Elements of $\Xi$ are first cleaned of high frequency noise by band-pass filter, using frequency band of 0.1 Hz. to 13 Hz. As our target is to classify ERP300 which is below 13Hz., the upper band is set at 13 Hz. (which includes $\alpha$, $\theta$ and $\delta$ waves), Next, signals from a particular probe obtained during a session (20 times the target character is flashed and 100 times other than target character flashed) are averaged. This is a standard procedure to filter out short pulses due to random brain activities. What remains prominent is the relevant ERP
waveform. The averaged signals from 21 probes, when the target character flashed, are shown in Fig. 2.

The amplitude of the signals vary over probes. To normalize, we used linear pulse code modulation (LPCM), with 8 levels. A typical EEG signal, containing ERP, is shown in Fig. 7. The same signal, after band-pass filtering is as shown in Fig. 8. After normalizing, the signals from 21 channels are clustered. The distance between pre-processed normalized signals are done using dynamic time warping (DTW) algorithm. This accommodates signals of similar shape, but shifted a little, in the same group, i.e., distance measured between them are low. The result of DTW, for 21 signals collected in a session is shown in Fig. 9.

Using DTW distances, the signals are clustered by Ward’s algorithm. This is a top-down agglomerative algorithm. The purpose of choosing top-down algorithm is that, we can set a threshold of the inter cluster distance and tune the number of clusters visually, using dendrogram. The dendrogram is Fig. 10, where it is shown how the number of clusters could be chosen by adjusting the inter-cluster distance. In our experiments, we used 8 clusters. The reason is that, in that case the MOGA searching will be simple. As we increase the MOGA search space, the computation time increases exponentially.

Next, we need to select a representative from each cluster.
We select the one for which the change in potential is highest. This is explained in Fig. 11. As we see, there are 3 members in the cluster. We measure the highest difference of potential for each member of the cluster. That is the score for an individual member. The member with highest score is selected as the representative of the group.

A. Multiobjective Genetic Algorithm

The final step is to select minimum number of signals, that would give the highest classification accuracy. After selecting representatives from each cluster, we have 8 EEG signals. Now, selection of a subset from 8 is a combinatorial problem which is easy to code in GA. Chromosomes with 8 binary units are used. A gene with 1 represents that that signal is selected, whereas a 0 represents that the signal is not selected. For crossover, 2-point crossover is used, with probability 0.5. Mutation is used with 0.01 probability. Tournament selection is used.

Calculation of the fitness of a chromosome is computationally heavy. First, features from the selected signals, as described by the chromosome’s genes with 1, are extracted. From every signal we extract 2 features, the maximum difference of the potential and the difference of time for their occurrences. Suppose a chromosome has five 1s. Ten features are extracted from the five signals, for all the samples containing ERP300 and not containing ERP300. The samples are labeled. We used back-propagation training algorithm to train multilayer perceptron. The classification result is one criterion of a chromosome’s fitness. The other criterion is number of signals, which in this case is 5. Using these two criteria, rankings of all the chromosomes in the population are calculated according to what is explained in Fig. 12. This ranking, proposed by us in our previous work, is a little different from traditional MOGA. We use the summation of two criteria - the fittest sample will be ranked 1. We normalize the two criteria for equal weightage.

In the next section, we will compare the results of classification with our previous work reported in [8]. In that work, the number of clusters were fixed to 4, i.e., 21 probe signals were divided into 4 groups based on similarity. There were some differences in clustering algorithm. The classification result based on the signals from four probes. Due to fixed number of clusters, cluster index for some clusters were poor. The classification results were based on the signals from 4 probes, which were representatives of the 4 clusters. As we used 8 clusters, individual cluster members have very similar shapes. Further, in this work we search for signals which really play a role to classify the presence and absence of ERP300. As a result, we could reduce the number of probes, sometime as low as 2 probes without sacrificing classification accuracy.

IV. Result

First, we will give results of our experiment. The same data were used in [8]. We will next compare the results. As we see in Tab. I, though there are common probes selected for different subjects, not all probes are the same. It clearly shows that the relevant locations are different for different subjects. For all the subjects, data were taken over several days, but within a few months. It is interesting to see, whether the location of the important probes for a particular subject changes with time or not.

In Tab. II, we compared results with our previous study [8]. In the previous work, for all subjects, 4 probes were used. The locations were different for different subjects. The signals were clustered, and 4 representative signals from 4 clusters were used. We see that we could achieve almost similar classification results with fewer number of probes. In fact, we ran the MOGA algorithm only for 50 generations. We expect better results with longer generations for genetic search.

Effectiveness of the method was demonstrated by experiments to show the same or better accuracy, when as low as 2 probes are used, compared to 4 reported in our previous work [8]. In [8], we compared results with [12] using data...
collected at our laboratory as well as data used in Hoffmann’s work [12]. Even our previous work [8] could achieve better performance with 4 EEG probes. Comparison with the present work and Hoffmann’s work[12] is not included in this paper.

### V. CONCLUSION AND FUTURE WORK

In this paper, we proposed a method to reduce the number of probes needed to monitor a system. As a case study, we used EEG signals. The task was to classify whether the EEG signal contains an ERP300 or not. The results were evaluated by classification accuracy as well as the number of probes used.

As the clustering algorithm is fast, in our next attempt we will use all the 128 probes available with the netscan equipment. We hope, by identifying probe locations more accurately, we could get better classification result with even less number of probes. We also want to study, whether the best probe locations, for a subject changes over long time or not.

### REFERENCES


