Multi-channel Surface EMG Classification Based on a Quasi-optimal Selection of Motions and Channels

Taro SHIBANOKI*, Keisuke SHIMA†, Takeshi TAKAKI*, Yuichi KURITA*, Akira OTSUKA‡, Takaaki CHIN§ and Tosho TSUJI†

* Graduate School of Engineering, Hiroshima University, Higashi-Hiroshima, Hiroshima, JAPAN
† Graduate School of Biomedical Science, Hiroshima University, Minami-ku, Hiroshima, Hiroshima, JAPAN
‡ Faculty of Health and Welfare, Prefectural University of Hiroshima, Mihara, Hiroshima, JAPAN
§ Hyogo Rehabilitation Center, Nishi-ku, Kobe, Hyogo, JAPAN

Abstract—This paper introduces a motion and channel selection method based on a partial Kullback-Leibler (KL) information measure. In the proposed method, the probability density functions of recorded data are estimated through learning involving a probabilistic neural network based on the KL information theory. Partial KL information is defined to support evaluation of the contribution of each dimension and class for classification. Effective dimensions and classes can then be selected by eliminating ineffective choices one by one based on this information, respectively. In the experiments, effective channels for classification were first selected for each of the six subjects, and the number of channels was reduced by 32.1 ± 25.5%. After channel selection, appropriate motions for classification were chosen, and the average classification rate for the motions selected using the proposed method was found to be 91.7 ± 2.5%. These outcomes indicate that the proposed method can be used to select effective channels and motions for accurate classification.

Index Terms—Kullback-Leibler information, class selection, variable selection, pattern classification, electromyogram (EMG)

I. INTRODUCTION

To control machines using biological signals, it is necessary to determine the relationships between these signals and actual motions. Various methods to classify biological signal patterns have been proposed [1]–[3]. Tsuji et al. proposed a log-linearized Gaussian mixture network (LLGMN) [2] that includes a Gaussian mixture model (GMM), and confirmed its effectiveness in biological signal classification [2], [3].

To classify biological signal patterns correctly, optimal recording sites for such signals and motions that are accurately classified must be selected in advance. Studies on optimal recording site selection have been conducted by evaluating all combinations of EMG channels [4], [5]. In regard to selecting motions (classes) for classification, Kita et al. proposed a method for extracting motions that are clearly separated from one another [6]. With a large number of channels, however, computation takes a long time. Moreover, to realize motion classification using the selected channels’ EMGs and/or motions, a classifier needs to be constructed after selection. Additionally, related efforts to date have focused only on the selection of optimal channels or motions.

The authors have developed a channel selection method based on the Kullback-Leibler (KL) information [7]. In this method, a partial KL information measure is defined as a metric of selecting optimal channels and ineffective channels for classification are eliminated one by one. A motion selection method involving the application of partial KL information for optimal class selection has also been proposed [8]. However, selecting the appropriate channels and motions cannot be performed at the same time.

Against such a background, this paper describes a motion and channel selection method based on the use of a partial KL information. In the proposed method, recorded signals are regarded as probability variables, and their probability density functions (pdfs) are estimated using probabilistic neural network (PNN) learning based on KL information. Additionally, ineffective channels and motions that are difficult to classify are eliminated one by one based on partial KL information [7], [8]. Under the proposed method, appropriate channels and motions can be selected at the same time.

II. A MOTION AND CHANNEL SELECTION METHOD BASED ON A PARTIAL KL INFORMATION MEASURE

A. Partial KL Information Measure [7], [8]

A pattern classification problem can be regarded as an estimation problem of the pdf for a given data [2]. Let us consider a case that \( N_k \) samples belong to the class \( k (k = 1, 2, \ldots, K) \) and each sample is an \( L \)-dimensional variable vector \( \mathbf{x} = [x_1, x_2, \ldots, x_L] \in \mathbb{R}^L \). It is assumed that \( \mathbf{x}^{(k)}_n \in \mathbb{R}^L \) is the probability variable vector of the \( n \)-th...
sample in the \(k\)th class, and that the probabilities of the \(k\)th class in the true probability distribution and the estimated distribution are \(P^n_k\) and \(Q^n_k\), respectively. The KL information is calculated using

\[
I_n(P, Q) = \sum_{k=1}^{K} P^n_k \log \frac{P^n_k}{Q^n_k},
\]

(1)

where \(P = [P^n_1, P^n_2, \ldots, P^n_K]^T\) and \(Q = [Q^n_1, Q^n_2, \ldots, Q^n_K]^T\), and the first term on the right side is reduced to a constant; accordingly, if the second term reaches a minimum, then so does the KL information. This means that the estimated distribution is close to the true distribution.

Based on this concept, the authors proposed a partial KL information measure with \(E_{i}[i]\) and \(E_{i'}[i]\) as a metric for variable and class selection [7], [8]:

\[
E_{i} = \frac{I(P, Q)}{I(P[i], Q[i])} = \frac{\sum_{n=1}^{N} I_n(P, Q)}{\sum_{n=1}^{N} I_n(P[i], Q[i])},
\]

(2)

\[
E_{i'} = \frac{I(P, Q)}{I(P'[i'], Q'[i'])} = \frac{\sum_{n=1}^{N} I_n(P, Q)}{\sum_{n=1}^{N} I_n(P'[i'], Q'[i'])},
\]

(3)

Here, \(I_n(P[i], Q[i])\) and \(I_n(P'[i'], Q'[i'])\) are the KL information for the \(n\)th sample \(x_n\), where \(P[i'] = [P^n_1, \ldots, P^{i'-1}_n, P^{i'+1}_n, \ldots, P^n_K]^T\) and \(Q[i'] = [Q^n_1, \ldots, Q^{i'-1}_n, Q^{i'+1}_n, \ldots, Q^n_K]^T\). They are obtained using the probability variables \(x[i]\) and \(x[i']\), in which the \(i\)th dimension and the \(i'\)th class are eliminated, respectively. In addition, \(N\) is the total number of samples, and \(N'\) is the total after class elimination. The \(i\)th dimension or the \(i'\)th class for which \(E_{i}\) or \(E_{i'}\) is the largest can be eliminated [7], [8].

B. Partial KL Information-based Motion and Channel Selection Using PNN

This section reports on motion and channel selection method using a partial KL information measure. The structure of the selection method is shown in Fig. 1. In the proposed method, the subject’s motions are learned by the LLGMN [2], and effective channels and motions for classification are selected at the same time.

1) Feature Extraction [3], [7], [8]: First, EMG signals measured from \(L\) pairs of electrodes are digitized using an A/D converter (sampling frequency: 1 [kHz]), and are rectified and filtered out through a second-order Butterworth filter (cut-off frequency: 1 [Hz]). These sampled signals are defined as \(EMG_l(n)\) \((l = 1, \ldots, L)\). Next, the \(EMG_l(n)\) values are normalized to make the sum of \(L\) channels equal to 1, and resulted vector is defined as the feature vector \(x(n) \in \mathbb{R}^L(n = 1, \ldots, N)\). Moreover, a channel average of \(EMG_l(n)\) is defined as force information.

2) Learning and Selection of Channels and Motions [2], [3], [7], [8]: The LLGMN, which is based on the GMM and a log-linear model of the pdf [2], is used for learning. This network can calculate the \(a posteriori\) probability \(\\hat{Y}(n) = [Y_1(n), Y_2(n), \ldots, Y_K(n)]^T\) for the input pattern \(x(n)\) after learning [2].

A simple backpropagation algorithm is feasible when the teacher vector \(T(n) = [T_1(n), T_2(n), \ldots, T_K(n)]^T\) for the \(n\)th input vector \(x(n)\) is given. As the energy function \(J\) for the network,

\[
J(x^*) = I(T, Y|x^*) = \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K} T_k(n) \log \frac{T_k(n)}{Y_k(n)}
\]

(4)

is used, and learning is performed to minimize the KL information for true distribution. This means that a learned network can approximate the data distribution. Using this energy function, therefore, it is possible to select optimal dimensions (channels) and classes (motions) that can be classified at the same time as the learning network.

The initial vector set \(x^* \in \mathbb{R}^L\) is defined as \(x^* = \{x(n) | n = 1, 2, \ldots, N\}\), and ineffective dimensions are eliminated one by one from \(x^*\) based on the variable selection method [7]. Here, \(D\) and \(D[i]\) are the average classification rates for \(x^*\) and \(x[i]^*\), respectively, and \(x[i]^*\) is the feature vector for which the \(i\)th dimension is eliminated from \(x^*\). Channels are first eliminated one by one while \(D[i] \geq D\). After channel selection, motions that are easy to classify can be selected using the class selection method [8] from the vector set \(x^* \in \mathbb{R}^R\), in which \(r\) dimensions are eliminated by the channel selection \(R + r = L\). For motion selection, motions are eliminated one by one until the average classification rate for the selected motions \(F[i]\) is greater than the pre-specified threshold \(F_0\). Effective channels and motions are thus selected sequentially.

3) Motion Classification [2], [3], [7], [8]: For EMG classification, since the output of the network’s third layer provides the \(a posteriori\) probability of each class, Bayesian discrimination (in which the class with the highest \(a posteriori\) probability is selected) is used.
ori probability becomes the result) is used for classification.

III. EXPERIMENTS

A. Methods

The subjects were five healthy volunteers (A – E) and one upper-limb amputee (F). In the experiments, EMG signals were recorded using an MT-11 multitelemeter (gain: 60 [db]; NEC Sanei Co., Ltd.). A pair of sheet electrodes (Unique Medical Co., Ltd., see Fig. 2) was attached to each healthy subject’s right forearm. Since the maximum number of channels which can be measured using MT-11 is 13, 13 pairs of electrodes were selected in advance from all the electrodes in the sheets \( (L = 13) \). Two pairs of Ag/AgCl EMG electrodes (SEB120, GE Marquette) were attached to the amputation site of subject F \( (L = 2) \). Subjects A – E spent a few seconds performing each of 16 motions \( (K = 16) \); see Fig. 3) for 15 times, and Subject F performed 11 motions.

The sampling frequency was 1 [kHz], six sets of recorded data were used for learning and selection of channels and motions (referred to here as the selection data), and the others were used only for classification (referred to as evaluation data). The number of selection data was 120 samples (randomly selected from 1,000×6 samples for each motion), and the number of evaluation data for each motion was 9,000 samples.

B. Results and Discussion

Figure 4 shows an example of the recorded data, and force information. The gray zone indicate that the force information less than the predefined threshold \( F_{th} \). The threshold value was set as \( F_{th} = 0.2 \) to enable the detection of voluntary motions [3]. It can be seen that the EMG signals measured from each channel are different for each motion. Firstly, effective channels are selected for each user. Table I lists channel combinations chosen using the channel selection method, along with the average classification rates for selection data before and after channels were selected. It can be seen that the rates increased after channel selection for Subjects A–E. This is because redundant channels, which are not effective for classification, can be eliminated using this method. Additionally, since a small number of channels (two channels) were attached in advance, all two channels were selected for Subject F. It can be concluded that the channel selection method was capable for selecting effective channel combinations.

On the other hand, Fig. 5 shows the results of reducing the number of motions using selection data measured from Subject C. It is confirmed that the classification rates are lower than the threshold with 16 to 11 motions, and ten motions were selected by the termination condition of the proposed method. The average classification rate was 98.4 ± 2.1%. Table II shows motions selected and the average classification rates of selection data for all subjects. Here, since the average classification rate of the selection data is high (98.0 ± 1.2%), motions are not eliminated for Subject A, B and D. From these results, it is confirmed that classification
We would like to cordially acknowledge and express our appreciation to Mr. Motonobu Shigeto and Mr. Kazuyuki Fujita for their assistance in the implementation of this research.

ACKNOWLEDGMENT

This paper introduces selection of appropriate motions and effective channels for classification using the partial KL information measure. In order to confirm its effectiveness, the method was applied to motion selection and EMG channel selection for EMG classification. In experiments on forearm motion classification, effective channels and appropriate motions were selected for each subject; the number of channels was reduced by 32.1 ± 25.5% and the average classification rate using the selected motions was 91.7 ± 2.5% with a high accuracy. In future research, we plan to apply this channel and motion selection approach to training systems for EMG-based human-machine interfaces.

IV. CONCLUSION

This paper introduces selection of appropriate motions and effective channels for classification using the partial KL information measure. In order to confirm its effectiveness, the method was applied to motion selection and EMG channel selection for EMG classification. In experiments on forearm motion classification, effective channels and appropriate motions were selected for each subject; the number of channels was reduced by 32.1 ± 25.5% and the average classification rate using the selected motions was 91.7 ± 2.5% with a high accuracy. In future research, we plan to apply this channel and motion selection approach to training systems for EMG-based human-machine interfaces.