Current Applied Physics 11 (2011) 740–745

Contents lists available at ScienceDirect

Current Applied Physics

journal homepage: www.elsevier.com/locate/cap

Comparison of *k*-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions

Kang Soo Kim [a], Heung Ho Choi [a], Chang Soo Moon [b], Chi Woong Mun [a,c,d,*]

[a] Dept. of Biomedical Engineering, Inje University, Obang-dong, Gimhae, Gyeongnam 621-749, Republic of Korea
[b] KMG Ltd., Saha-gu, Busan, Republic of Korea
[c] U-Health Care Research Center, Inje University, Obang-dong, Gimhae, Gyeongnam 621-749, Republic of Korea
[d] Foundation of Inje Research Taskforce, Inje University, Obang-dong, Gimhae, Gyeongnam 621-749, Republic of Korea

A R T I C L E I N F O

Article history:
Received 1 August 2010
Received in revised form
16 November 2010
Accepted 17 November 2010
Available online 27 November 2010

Keywords:
EMG
LDA
QDA
k-NN
DAMV
Wrist motion

A B S T R A C T

In this study, the authors compared the *k*-Nearest Neighbor (*k*-NN), Quadratic Discriminant Analysis (QDA), and Linear Discriminant Analysis (LDA) algorithms for the classification of wrist-motion directions such as up, down, right, left, and the rest state. The forearm EMG signals for those motions were collected using a two-channel electromyogram(EMG) system. Thirty normal volunteers participated in this study. Thirty features with a time-window size of 166 ms per feature during a 5-s forearm muscle motion were extracted from the gathered EMG signals. The difference absolute mean value (DAMV) was used to construct a feature map and the LDA, QDA, and *k*-NN algorithms were used to classify the directions of the signal. The recognition rates were 84.9% for *k*-NN, 82.4% for QDA, and 81.1% for LDA. There was a statistically significant difference between the *k*-NN and LDA algorithms ($P < 0.05$).

1. Introduction

Many studies on electromyogram (EMG) systems have been reported in several fields, such as in providing assistance for the disabled, in rehabilitation training, and in device control [1]. Recently, EMG signals were used to control game devices for entertainment and the authors exhibited a “King of the muscle” game device, driven by forearm EMG signals, which was a modified version of the hammer in the “King of the hammer” game at the Korean International Medical & Hospital Equipment Show (KIMES) 2010 [2]. Additionally, evaluation of musculoskeletal system disorders using EMG devices is a technique that allows recording, analyzing, and treating the occurrences of muscular electrical signals. Recognition regarding directions of exercise is needed to use control device or game machine. Hence, the pattern recognition algorithm is a fundamental element of the system. The feasibility and performance of pattern recognition algorithms for EMG signals from the forearm muscles have been investigated in a number of studies involving normal subjects [3–6]. There are several pattern recognition algorithms, such as fuzzy logic [1,3,7], multivariate autoregressive (AR) modeling [4], discriminant analysis (DA) [5,8], and artificial neural networks [9,10]. The study that performed pattern recognition as using EMG signals confirmed recognition rates along motions as selected algorithm to their purpose. However, the pattern recognition algorithm should be quantitatively selected among various algorithms in order to find out the most reliable algorithm on motions. When made a difference only algorithm method and put all of conditions equally, we can find out the most reliable algorithm as the compared result. Because simplicity and fast speed of the classification process are essential factors for a real-time system, three relatively simple and fast algorithms were selected: The linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and *k*-nearest neighbor (*k*-NN) algorithms were implemented for the classification of muscle-motion directions, based on forearm EMG signals, and their recognition performances were compared.

* Corresponding author. Dept. of BME, Inje University, Obang-dong, Gimhae, Gyeongnam 621-749, Republic of Korea. Tel.: +82 55 320 3297; fax: +82 55 320 3797.
E-mail address: mcw@inje.ac.kr (C.W. Mun).

1567-1739/$ – see front matter Crown Copyright © 2010 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.cap.2010.11.051

1.1. k-Nearest neighbor

The k-NN classification algorithm predicts the test sample's category according to the k training samples, which are the nearest neighbors to the test sample, and classifies it to the category that has the largest category probability. Suppose that there are j training categories, as c_1, c_2, … c_j, and the sum of the training samples is N. Also, class X is the same feature vector as all the training samples. When d_i is one of the neighbors in the training set, $y(d_i,c_j) \in \{0,1\}$ indicates whether d_i belongs to class c_j, and $Sim(X, d_i)$ is the similarity function for X and d_i. Then, the probability density function $P(X, c_j)$ for the feature data X, given class c_j, can be written as equation (1) [11].

$$P(X,c_j) = \sum_{d_i \in kNN} Sim(X,d_i)\cdot y(d_i,c_j), \tag{1}$$

$Sim(X, d_i)$ can be calculated using the Euclidean distance, cosine, and correlation methods. In this study, the Euclidean distance method was selected because it is often used as the distance metric. The k-value is a user defined constant number of neighbor group elements and an unlabelled vector is classified by assigning the label that occurs most frequently among the k training samples nearest that query point. In this study, the k-value was initially fixed at 5 and it will be discussed in the Discussion. According to the label of the k neighbors and the distributions of the similarity value, the class of the input vector $\boldsymbol{X}$ was discriminated.

1.2. Linear and quadratic discriminant analysis

Linear and quadratic discriminant analyses were carried out using a simple max gate function $g(X)$ as a classification rule. The prior probability of class i is π_i, and the conditional density of X in class i is $f_i(X)$. It is assumed that the vector of feature variables $\boldsymbol{X}$ is multivariate normally distributed in the group with mean vector μ_i and common covariance matrix $\sum$ in the case of LDA or group specific covariance matrix $\sum_i$ for QDA. It is also assumed that $g_i(\boldsymbol{X}) > g_j(\boldsymbol{X})$ for $j \neq i$, $\boldsymbol{X}$ is a member of class i, and the prospect densities are Gaussian. Then, $f_i(X)$ can be computed as shown in equation (2). Using the Bayes rule and natural logs, the MAP (maximum a-posteriori) discriminant functions can be written as shown in equations (3) and (4) [12].

- Multivariate Gaussian:

$$f_i(X) = \frac{1}{(2\pi)^{P/2}\left|\sum_i\right|^{1/2}}\exp\left[-\frac{1}{2}(X-\mu_i)^T\sum_i^{-1}(X-\mu_i)\right], \tag{2}$$

- Linear discriminant function:

$$g_i(X) = X^T\sum^{-1}\mu_i - \frac{1}{2}\mu_i^T\sum^{-1}\mu_i + \log(\pi_i), \tag{3}$$

- Quadratic discriminant function:

$$g_i(X) = -\frac{1}{2}(X-\mu_i)^T\sum_i^{-1}(X-\mu_i) - \frac{1}{2}\log\left(\left|\sum_i\right|\right) + \log(\pi_i), \tag{4}$$

where P is a dimension factor (1 for LDA and 2 for QDA), and T is a transpose operator.

2. Materials and methods

2.1. Subjects

Thirty normal volunteers (15, 2127-year-old males and 15, 20–27-year-old females) participated in the study. They were well informed about the experimental procedures before starting the study.

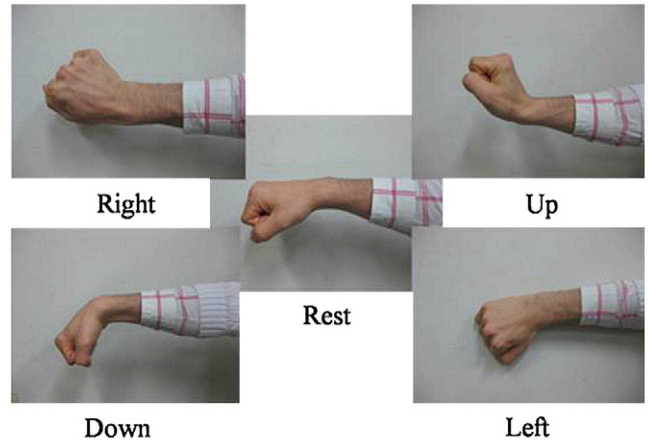

Fig. 1. Pictures of four wrist motions and the rest state.

2.2. Experiment protocol

iDAQ400 (Physiolab Co., South Korea) with Ag/AgCl (3M, MN, USA) disposable electrodes were used to acquire EMG signals. The iDAQ400 system has two channels simultaneously obtaining the EMG signals through a band-pass filter of 10–500 Hz bandwidth. In this study, wrist motions, down, left, and right, were discriminated among EMG signals as shown in Fig. 1, which were collected using two-channel EMG electrodes attached to the extensor digitorum for channel 1 and the flexor carpi for channel 2 as shown in Fig. 2. Each motion was sequentially repeated in 10-s intervals and EMG signals were collected simultaneously; 5 s of the resting state was followed by 5 s of wrist-moving action, which took 45 s of total acquisition time, by adding a 5-s resting state at the end.

The sampling frequency of the 12-bit A/D converter was 1 kHz. Thirty features with a time window size of 166 ms per feature during each wrist-moving time, 5 s, were extracted from the gathered EMG signals. Next, a feature map was generated from the difference absolute mean value (DAMV), the mathematical expression of which is shown in equation (5), of the acquired EMG signals of both channels 1 and 2. $\boldsymbol{X}$ is the EMG signal, i the sequence index of the samples, and N the total number of data points [13]. The reason of feature selection will be examined more in Discussion. Finally, the directions of the wrist motion were

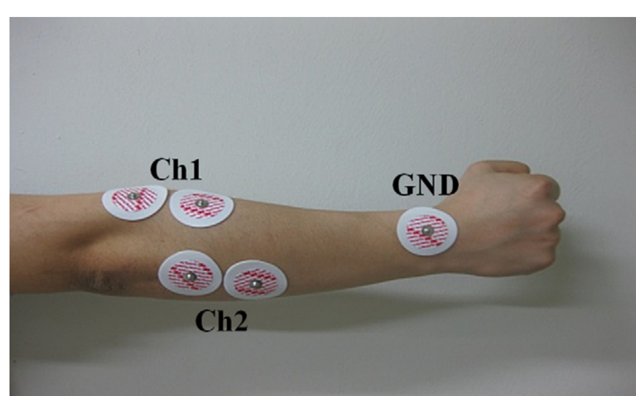

Fig. 2. The two-channel-electrodes positions of the right forearm.