Finger joint angle estimation based on sEMG signals by Attention-MLP

Zhebin Yu\textsuperscript{1*}, Hui Wang\textsuperscript{2,3} and Wenlong Yu\textsuperscript{4}

\textsuperscript{1} School of artificial intelligence, Hebei University of Technology, Tianjin, 300131, China
\textsuperscript{2} Shenzhen Key Laboratory for Innovative Technology in Orthopaedic Trauma, Department of Orthopaedics and Traumatology, The University of Hong Kong-Shenzhen Hospital, Shenzhen, China.
\textsuperscript{3} CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS), Shenzhen, China.
\textsuperscript{4} Shenzhen RunYiTaiYi Technology, Shenzhen, China.

*Corresponding author’s e-mail: yzm9696@163.com

Abstract. sEMG (Surface electromyography) signal was widely applied in human-machine interactive field, especially in robotic arm control. In this paper, we built the Attention-MLP (Multilayer Perceptron) model to implement a type of continuous joint angle estimation method based on sEMG for six grasp movements, we tested this model on Ninapro dataset and the average Pearson correlation coefficient (CC) and the average root mean square error (RMSE) of the proposed Attention-MLP method achieved 0.812±0.02 and 10.51±1.98; the average CC and RMSE of this method are better than Sparse Pseudo-input Gaussian processes (SPGP), its average CC and RMSE are 12.14±2.30 and 0.727±0.07. Compared with the traditional method SPGP, our model performed better on continuously estimation of ten main hand joint angles under 6 grip movements.

1. Introduction
sEMG signal is a kind of bioelectrical signal that records muscle activities. Many machine learning classification algorithms were proposed for recognize motions by sEMG. Recently, in order to realize a natural and precise control methods, continuous movement estimation by sEMG has been a popular research topic\cite{1}; the corresponding relationship between the sEMG signals and the real-time joint angles can be well excavated by deep learning models to realize fine-grained control systems such as myoelectric manipulator.

Deep learning was successfully used in signal processing for its great generalization ability; attention mechanism of this paper is an effective deep learning model for extracting important features information. The concept of attention mechanism was inspired by attention in biology and it was first proposed in Computer Vision to extract image features\cite{2}. After 2014, attention mechanism has been successfully applied in text classification, machine translation and other tasks\cite{3}. Recent years, attention mechanism has been used to replace Recurrent Neural Network in Sequence-to-Sequence tasks\cite{4}; because position encoding enables attention mechanism to perceived sequence timing information. In
this study, we used the Multi-Head Self-Attention mechanism processed by positional encoding to further extract the sEMG signals information; then we use Multilayer Perceptron (MLP) to generate real-time joint angles of six movements[5]. Both sEMG and joint angles comes from the public dataset, NinaPro[6]. The feature of sEMG we chose is RMS[7], for its low computational complexity and sufficient information before putting signals into the network.

2.Method

2.1 Database

Ninapro, a public dataset [6], included 7 databases. These databases include sEMG, hand kinematics when subjects doing hand motions. The requirement of each action is to continue for 5S, rest for 3S and repeat for 6 times. In this study, we selected the sEMG and 10 main joint angle sensors which is located on metacarpal and proximal of every finger before and after completing six grasping movements in the dataset for complete the training and testing of continuous joint angle estimation model, The grasp motions is shown in figure 1.

Ninapro placed 12 sEMG signal wireless sensors on the forearm and upper arm [8]. The sEMG signals with 12 channels were recorded by Trigno™ wireless sEMG recording system (Delsys, Inc, Natick, Massachusetts, USA) through twelve channels separately at the sampling frequency of 2000H.

The hand joint angles with 10 channels were recorded by CyberGlove II data-glove, shown in Fig.2. The yellow dot represents our estimate of ten angles. The sensor of glove was resampled to 2000 Hz to synchronize with the sensor of Delsys.
2.2. Data Preprocessing
Before using the data, sEMG signals and hand joint angles signals should be preprocessed. Firstly the Butterworth filter is used to remove the noise; Secondly the sEMG signals were extracted root mean square (RMS) features of every sampling point; the window of 100 ms duration and a stride length of 0.5 ms was used in features extraction[9]. Then, zero-phase lowpass filter was used to smooth the glove angle signals, which making it more consistent with the normal movement of the human body. Finally in order to speed up finding the optimal solution and improve the accuracy of the model, Min-Max normalization is used to map sEMG features and hand angles to [-1,1]. After the feature extraction, feature sequence and joint angle sequence were divided into 100 ms shifting windows with 50ms-long overlap as the inputs of the neural network[10]. In the Ninapro database, each movement have six trials; after completing the above data processing, we selected four trials of each movement as training dataset and two trials of each movement as testing dataset.

2.3. Models
2.3.1 Attention-MLP
In this study, a deep network model combining Multi-Head Attention and Multilayer Perceptron (MLP) is adopted, shown as figure 4. The multi-head attention module used in this paper is a stacked multi-head attention for further feature processing [4]; multi-head attention is a special attention model which consists of multiple Scaled Dot-Product Attention models (Single head attention) shown as figure 3.

Assuming that the input vector of sEMG feature is defined as \( X = \{x_0, x_1, \ldots, x_n\} \in \mathbb{R}^{n \times 1} \) and the output of single head attention defined as \( H = \{h_0, h_1, \ldots, h_n\} \in \mathbb{R}^{n \times d_m} \). The calculation of single head attention is:

\[
\text{Attention}(\mathbf{x}_i, \mathbf{H}) = \frac{e^{\mathbf{a}^T \mathbf{W}_k \mathbf{x}_i}}{\sum_j e^{\mathbf{a}^T \mathbf{W}_k \mathbf{x}_j}},
\]

where \( \mathbf{W}_k \) is the weight matrix, \( \mathbf{a} \) is the attention vector, and \( e \) is the exponential function.
attention for a specific point \( x_t \) in the sequence can be described as the following steps: firstly it linearly map \( x_t \) to three different spaces \( W_0 \in \mathbb{R}^{l \times d_0}, W_K \in \mathbb{R}^{l \times d_k}, W_V \in \mathbb{R}^{l \times d_v} \) and calculate attention score function \( s(x_t, x_o) \) to express weight coefficient of time \( t \) and time \( o \), the function \( S(x_t, x_o) \) is:

\[
S(x_t, x_o) = \frac{(x_t W_0 (x_o W_K))^T}{\sqrt{d_z}}
\]

then output point \( h_t \) of input point \( x_t \) can be expressed as:

\[
h_t = \text{att}(x_t) = \sum_{o=1}^{n} \frac{s(x_t, x_o) x_o W_V}{\sum_{p=1}^{n} s(x_t, x_p)}
\]

Due to the single head attention mechanism cannot directly process the position information of timing signals, we add position encoding to each input vector \( x_t \), which enabling the attention model to capture the position information of the input sequence. We use the sinusoidal Positional encoding \( P = \{p_0, p_1, ..., p_n\} \in \mathbb{R}^{n \times 1} \) proposed by Google in 2017[4].

\[
p_t(t, 2i) = \sin \left( \frac{t}{1000^i} \right)
\]

\[
p_t(t, 2i) = \cos \left( \frac{t}{1000^i} \right)
\]

For the new input \( X+P \), attention score of time \( t \) and time \( o \) is:

\[
\left((x_t + p_i) W_0 \right) \left((x_o + p_o) W_K \right)^T
\]

due to the existence of \( p_t W_0 W_K p_o^T \), the attention score function can capture the distance information between time \( o \) and time \( i \) [11]. When the actual model is used, input of this model calculates a complete sequence signal segment simultaneously, packed together into a query Vector matrix \( Q = \{q_0, q_1, ..., q_n\} \in \mathbb{R}^{n \times d_q}, k_1 = c_i W_Q \) , key vector matrix \( K = \{k_0, k_1, ..., k_n\} \in \mathbb{R}^{n \times d_k}, k_1 = c_i W_K \) and value vector matrix \( V = \{v_0, v_1, ..., v_n\} \in \mathbb{R}^{n \times d_v}, v_1 = c_i W_V \) : The output vector sequence can be described as [4]:

\[
Z = \text{Attention} = \text{softmax} \left( \frac{Q K^T}{\sqrt{d_k}} V \right)
\]

In this study, the multi head attention module consists of \( m \) single head attentions, which do not share parameters for capture different interactive information in multiple different projection spaces [12]. Concat module concatenates the single head attention’s output \( Z_1 \sim Z_m \) together, so, the output of Multi-Head Self-Attention can be described as

\[
H = \text{Concat} (Z_1, Z_2, ..., Z_m) W_0
\]

\( W_0 \in \mathbb{R}^{m d_v \times 0} \), in our network, \( m \) is three;
the output of multi-head attention module will be used as the input of MLP to complete the regression task. It is composed of three-layer fully connected feedforward neural network, the activation function enhances the fitting ability of the model through.

The formula can be expressed by the following formula:

\[
H_1 = \sigma(W_1 H + b_1)
\]

\[
H_2 = \sigma(W_2 H_1 + b_2)
\]

\[
O = W_3 H_2 + b_3
\]

\( W_1, W_2, W_3 \) and \( b_1, b_2, b_3 \) represent the weight matrixes and bias vectors of each layer respectively, and \( \sigma \) is the activation function described as:

\[
\sigma = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}
\]

MLP module completes three dimensional transformations, 128, 64 and 10 dimensions respectively. In this study, the Mean Square Error between the predicted and real angle is chosen as loss function. Adam Optimizer was applied to minimize the loss of entire networks [13].
2.3.2 Sparse Pseudo-input Gaussian processes

Gaussian regression can perform regression analysis on data through prior information. According to the assumption of Gaussian processes, if we have a set of training data \( D = \{X, Y\} \), supposing \( Y \) obeys a normal distribution with a mean of 0 and has Gaussian noise \( \varepsilon \sim N(0, \sigma^2) \), the function \( Y = f(X) \) can be described as \( Y \sim N(0, K(X, X) + \sigma^2 I_n) \).[14] \( K(X, X) \) is the covariance matrix commonly calculated by RBF kernel function. The likelihood function of a single sample point can be expressed by:

\[
P(y|x, X, Y, \theta) = \mathcal{N}\left(y \left| k(x, x) - k^T(x, X)[K(X, X) + \sigma^2 I_n]^{-1}k(X, x) + \sigma^2 \right. \right)
\]

(11)

\( \theta, \sigma^2, k(X, x), k_*(x, x) \) are hyperparameters. GPs are prohibitive for large data sets, because time complexity is \( O(N^3) \) due to the inversion of the covariance matrix. The computational complexity of predicting a sample is \( O(N^2) \), and the time required for prediction increases gradually with the increase of training set \( N \). In order to solve this problem, SPGP was presented by Edward Selsdon et al. in 2006.[15] In SPGP, a pseudo data set \( \tilde{D} = (\tilde{X}, \tilde{f}) \) replaces the real data set \( \tilde{X} = \{\tilde{x}_i\}_{i=1}^M \) is pseudo-inputs, \( \tilde{f} = \{f_m\}_{m=1}^M \) is pseudo targets, this prediction distribution obtained from the pseudo data set is taken as a parameterized likelihood function.[15] Thus, the complexity of computation in training process and predicting process will reduce to \( O(M^2 N) \) and \( O(M^2) \). In this study the pseudo-input \( M = 30 \). The code of this algorithm comes from Sheffield ML toolbox which contains MATLAB codes for Gaussian processes.

2.4 Post processing

In order to prevent excessive sudden changes in the output by the model from damaging the user and the controlled mechanical equipment, we analysed the absolute value of the maximum difference between two adjacent angles in the training data, when the output joint angle is greater than the threshold, we set the output joint angle as the threshold itself. At the same time, we take the average of the two adjacent joint angles of 50ms as the output of the primary joint angle, so as to enhance the smoothness and stability of the output hand joint angles and reduce the noise.

3. RESULTS

After filtering and extracting features of the raw sEMG signals, we constructed the Attention-MLP model for training and generating real-time hand joint angles under grasping movement. The Attention-MLP model was implemented using Pytorch, and SPGP was built on MATLAB; To evaluate the performance of Attention-MLP and SPGP, we used the correlation coefficient (CC) and root mean square error (RMSE) to evaluate the prediction’s accuracy. The experimental results were shown as in figure. 5 and figure 6. Compared with the model of SPGP, the Attention-MLP results showed that the accuracy of the Attention-MLP model was significantly better than the Traditional machine learning model.
Figure 4. The average accuracy of Attention-MLP and SPGP for 8 subjects. Subfigure (a) compared the correlation coefficients (CC) of SPGP and Attention-MLP, the CC of Attention-MLP and SPGP were $0.812 \pm 0.02$ and $0.727 \pm 0.07$, respectively; subfigure (b) compares the root mean square error (RMSE) of two models, the RMSE of two models were $10.51 \pm 1.98$ and $12.14 \pm 2.30$ respectively.

Figure 5. The predicted and actual values of Attention-MLP and SPGP. The number in the upper right corner of the subgraph represents the label corresponding to the sensor on the glove.
4. Conclusion
In this paper, we proposed an attention-MLP model to estimate finger joint angle during grasping movement simultaneously and proportionally. We can see this model has achieved good accuracy between 10 hand joint angles, because through attention mechanism and pre-processing, the model can extract the temporal and spatial characteristics of sEMG information, MLP can better fit the joint angles. Then we chose traditional pattern recognition method SPGP to compare with attention-MLP under the same task, results showed that Attention-MLP is more suitable. In the future, We will further improve the attention mechanism in order to develop the potential of attention mechanism on the estimation of multi degrees of freedom continuous motion and other human–computer cooperation method based on sEMG.

References
[1] Han J, Ding Q. (2015) A State-Space EMG Model for the Estimation of Continuous Joint Movements. J. IEEE Transactions on Industrial Electronics,62:4267-4275.
[2] D Bahdanau, K Cho, Y Bengio (2014). Neural machine translation by jointly learning to align and translate. In Computer Science. pp:1409-1473.
[3] Yang Z, Yang D. (2016) Hierarchical attention networks for document classification. In NAACL, San Diego, USA. pp 1480–1489.
[4] Vaswani A, Shazeer N, Parmar N, et al. (2017). Attention is all you need. In Advances in neural information processing systems, California, USA. pp. 5998-6008.
[5] Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. J. Pergamon,2:359-366
[6] Atzori M, Gijsberts A, Castellini C, et al. (2014) Electromyography data for non-invasive naturally-controlled robotic hand prostheses. J. Scientific Data. 1:140053
[7] Phinyomark A, Phukpattaranont P, Limssakul C. (2012) Feature reduction and selection for EMG signal classification. J. Expert Systems with Applications, 39:7420–7431.
[8] Gijsberts A, Atzori M, Castellini C. (2014) Movement error rate for evaluation of machine learning methods for sEMG-based hand movement classification. IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society. 22: 735-44
[9] Guo W, Ma C, Wang Z, Zhang H, Farina D, Jiang N. (2021) Long exposure convolutional memory network for accurate estimation of finger kinematics from surface electromyographic signals. Journal of Neural Engineering, 18.
[10] Englehart K, Hudgins B. (2003) A robust, real-time control scheme for multifunction myoelectric control. J. IEEE transactions on bio-medical engineering,50: 848-854.
[11] Su J. (2021) Transformer upgrade path Sinusoidal Explore the root of Positional Encoding. https://spaces.acn.cn/archives/8231.
[12] Bishop C (2007) Pattern recognition and machine learning[M]. Springer, Berlin.
[13] Kingma D, Ba J. (2014) Adam: A Method for Stochastic Optimization, Computer Science, pp. 1–15.
[14] Xiloyannis M, Gavriel C. (2017) Gaussian process autoregression for simultaneous proportional multi-modal prosthetic control with natural hand kinematics IEEE Transactions on Neural Systems and Rehabilitation Engineering 25: 1785-1801
[15] Ahsan, Md (2009) EMG signal classification for human computer interaction: a review. J. European, Journal of Scientific Research,33:.480-501.