Using the naive bayes classifier to classify the motion of a recorded EMG signal into palmer and lateral motions

Ali.Hossam ELdin
Biomedical engineering department
Faculty of engineering
Cairo, Egypt
hossam26644@live.com

Attia.Ahmed
Biomedical engineering department
Faculty of engineering
Cairo, Egypt

Saeed.Hazem
Biomedical engineering department
Faculty of engineering
Cairo, Egypt

Abstract—We used the set of records provided by the teaching assistant (Elbially.M) to construct a naive bayes classifier to classify EMG records into palmer and lateral motions using python as a programming language.

I. NAIVE BAYES

Naive bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. Abstractly, naive Bayes is a conditional probability model: given a problem instance to be classified, represented by a vector

representing some n features (independent variables), it assigns to this instance probabilities

for each of K possible outcomes or classes.

Using Bayes' theorem, the conditional probability can be decomposed as

II. IMPLEMENTATION

A. Importing libraries and records

First, scipy library is imported to be used in importing the file given with mat format.

Three other libraries are imported; numby library to perform some array operations, matplot library to plot sample records, and math library for some math operations.

Records are imported using scipy.io.loadmat function and then stored in two dictionaries;Raw_Lateral to store lateral dictionary and Raw_Palmer to store palmer dictionary.

Records are then overwritten in the two storage elements Raw_Lateral and Raw_Palmer; resulting in both storage element holds the 2-D array of elements with the size of the records, 150 records with 3000 value of each record for each file

B. Noise cancellation

Electric noise cancellation is an essential step to removed the induced electric noise resulting from the main power lines; 50 HZ noise is reduced by using a notch filter where a frequency band of 4HZ is canceled around the 50HZ frequency to ensure electric noise cancellation.

A function is constructed to remove noise from all records named “Implement_Notch_Filter”. this function is then called twice to remove noise from each set of records(Raw_Palmer and Raw_Lateral)

C. Creating features table

$$\text{CurveLength} = \sum x_i - x_{i-1}$$
$$\text{NonlinearEnergy} = \sum -x_i x_{i-5} + x_{i-1}{}_5$$
$$\text{4th Power} = \sum x_i{}_4$$
$$\text{Energy} = \sum x_i{}_3$$

a.features calculations

For features are selected to hold the characteristics for each record; Energy, Power, Nonlinear energy and Curve length.

A function is created to calculate the features any give table(features_table) was constructed, both energy and 4 th power are calculated using numby.sum, while nonlinear energy and curve length are calculated by element by element calculations using two loops with two pointers to select both a

row and a column and hence a specific record, non linear energy calculation starts with the third element in each record as its calculation requires the two elements before it in the record, while curve length calculation starts with the second element as the equation requires only one element before it in the array.

Function features_table is then called twice(one time per each table) and the first three records in each features table is displayed as an example.

D. Shuffling records

Shuffling records is an essential step to overcome bias;

np.random.shuffle function is used to shuffle records.

E. Slicing tables

Tables are sliced into a set of testing records and a set of training records, the ratio adopted for this slicing process is 5/7:1-(5/7) resulting in 107 values for training and 43 for testing.

The value training holds the number of training records, .shape function is used to get the number of training records then it is multiplied by the factor(5/7).

F. Mean and standered deviation calculation

Mean and Slandered deviation are calculated to get the probability of a records features to lie within the range of palmer movement or lateral movements.

np.mean is used to calculate the mean of each features from features tables, hence eight means are calculated; lateral power mean, lateral energy mean ... palmer energy mean, palmer lateral mean.. palmer curve length mean.

np.std is used to calculate the standard deviation for each features from features tables, hence eight Slandered deviations are calculated; lateral power standard deviation, lateral energy standard deviation ... palmer energy standard deviation, palmer lateral standard deviation.. palmer curve length standard deviation.

G. Probability calculation

To calculate the probability of each record belonging to the palmer movement or lateral movement; two functions to calculate probability are constructed, Record_probability takes the array and row number to calculate the multiplication of all features belonging to a certain movement, Probability is used to calculate the input feature probability using the mean and standard deviation.

H. Accuracy

Calculating accuracy is the key to deter min the success of this classifier.

Accuracy is calculated based on the ratio between correctly classified records divided by all records.

A loop is used to move along every records for both palmer and lateral tables, for each record the probability of it for being lateral and the probability of it for being palmer are calculated by calling the function Record_Probability; if the record belonging to the palmer table having a palmer probability higher than a lateral probability then the number of correct estimate is increased by one, else if a record belonging to the lateral table having a lateral probability higher than a palmer probability then the number of correct estimate is increased by one.

Accuracy is then calculated by dividing n(correct estimates) by total number of records.

REFERENCES

[1] wik pedia(https://en.wik ipedia.org/wik i/Naive_Bayes_classifier)

[2] Stack overflow com m unity(http://stackoverflow.com)

[3] Num by and scipy docum entation (https://docs.scipy.org/doc/)

[4] E lbialy.Moham ed T utorials.