The issue of pattern recognition (EEG Signals)

Abstract

This project aims to address the challenge of pattern recognition in EEG signals, particularly in response to Virtual Reality (VR) stimuli. The methodology involves implementing EEG signal processing and feature extraction using evolutionary algorithms and the Fisher method. Subsequently, a machine learning model is developed employing Multilayer Perceptron (MLP) and Radial Basis Function (RBF) networks for effective pattern recognition.

Phase 1

PART A:

Data Extraction: Loading and organizing the EEG data from the 'Project_data.mat' file.

Feature Extraction: Features were extracted from the data, encompassing both frequency domain and statistical features. These features consist of maximum frequency, average frequency, median frequency, relative energy of power spectrum bands, variance, amplitude histogram, form factor, and correlation between two channel signals. Each of these features was extracted from each of the 59 channels' data. Certain features, such as correlation or domain histogram, yield more than one value for each channel. The total number of features calculated amounts to 2832.

PART B:

Fisher Calculation: In this phase, we employ the Fisher criterion to identify optimal features. For each feature, we calculate the Fisher value based on the function introduced in the slides for the one-dimensional Fisher criterion. Subsequently, all 2832 Fisher values, along with their corresponding feature names, are aggregated into an array. Recognizing that a higher Fisher value signifies greater suitability for the respective feature, we sort the array in descending order and select the top 50 features, representing the most influential ones.

Results: Naming corresponds to the feature name along with the respective channel.

```
['band_theta_channel49', 'band_theta_channel56',
'hist2_channel10', 'mod_freq_channel42',
'hist3_channel19', 'band_gamma_channel58',
'band_theta_channel58', 'hist5_channel10',
'band_gamma_channel49', 'band_theta_channel22',
'mean_freq_channel22', 'hist3_channel48',
'max_freq_channel23', 'hist2_channel8',
'hist6_channel19', 'max_freq_channel42',
'hist5_channel19', 'max_freq_channel22',
'corr_channel38channel58', 'hist6_channel48',
'hist1_channel10', 'hist2_channel48',
'corr_channel47channel51', 'corr_channel41channel55',
'mod_freq_channel22', 'corr_channel36channel58',
'mod_freq_channel55', 'hist3_channel8',
'mod_freq_channel32', 'mod_freq_channel58',
'mod_freq_channel53', 'hist6_channel8',
'hist3_channel10', 'mean_freq_channel32',
'band_gamma_channel22', 'corr_channel12channel52',
'band_gamma_channel41', 'corr_channel34channel57',
'mean_freq_channel58', 'hist5_channel8',
'band_gamma_channel55', 'band_theta_channel51',
'max_freq_channel53', 'band_gamma_channel32',
'max_freq_channel50', 'hist4_channel24',
'corr_channel22channel24', 'corr_channel38channel51',
'mod_freq_channel46', 'band_theta_channel41']
```

PART C:

MLP: The Multilayer Perceptron (MLP) network was designed for predictive modeling, utilizing a range of parameters and distinct feature sets. To ensure consistent scaling across features and channels, the training data underwent Min-Max normalization. Additionally, 5-fold cross-validation was employed to robustly assess its performance, ensuring a thorough evaluation across different subsets of the dataset. Two categories of selected features were employed: the first comprised the 50 best features identified by the Fisher method, while the second involved the random selection of 50 features from the top 75 features according to the Fisher method. To explore different combinations, a for loop was implemented, randomly selecting 50 features from the 75 best features four times.

Two arrays, `'all_best_accuracies_MLP'` and `'all_predicted_MLP'`, were defined. The MLP network was trained iteratively for each combination of parameters and features, and the results were appended to these arrays. Subsequently, the predicted values were reported based on the highest achieved accuracy among the various configurations.

Results:

Best MLP network parameters using 50 best features:

```
Best MLP Network Configuration:
Number of Layers: 4
Neurons per Layer: [128, 63, 32, 16]
Activation Function: ['relu', 'relu', 'relu', 'relu']
Average Cross-Validation Accuracy: 92.00%
```

Best MLP network parameters using randomly selecting 50 features of 75 best features:

```
Num of random selstion:0, Number of Layers: 3, Neurons per Layer: [128, 64, 32], Activation Function: ['relu', 'relu', 'relu'], Average Cross-Validation Accuracy: 91.82
Num of random selstion:1, Number of Layers: 3, Neurons per Layer: [128, 64, 32], Activation Function: ['relu', 'relu', 'relu'], Average Cross-Validation Accuracy: 89.27
Num of random selstion:2, Number of Layers: 3, Neurons per Layer: [128, 64, 32], Activation Function: ['relu', 'relu', 'relu'], Average Cross-Validation Accuracy: 90.73
```

PART D:

RBF: Similar to the previous section, an RBF network was implemented using the Support Vector Classification (SVC) from sklearn.SVM. The dataset underwent normalization, and 5-fold cross-validation was applied to assess performance robustly. Different combinations of parameters and features were explored in a manner similar to the MLP network implementation.

Results:

Best MLP network parameters using 50 best features:

```
Best Parameters: {'C': 10, 'gamma': 1}
Average Cross-Validation Accuracy: 81.64%
```

Best MLP network parameters using randomly selecting 50 features of 75 best features:

```
Num of random selction:0, C: 10, gamma: 1, Average Cross-Validation Accuracy: 77.09
Num of random selction:1, C: 10, gamma: 1, Average Cross-Validation Accuracy: 81.64
Num of random selction:2, C: 10, gamma: 1, Average Cross-Validation Accuracy: 77.45
Num of random selction:3, C: 15, gamma: 1, Average Cross-Validation Accuracy: 80.00
```

PART E:

Comparing the RBF and MLP networks, it was observed that the MLP network consistently exhibited higher accuracies.

PART F:

Predicted values using MLP:

```
[0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 1,
1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1,
1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0,
1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0,
1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1,
1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0,
1, 1, 1, 1, 1]
```

Predicted values using RBF:

```
[0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 0, 0,
0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0,
0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1,
1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0,
1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0,
0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0,
1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0,
1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1]
```

Phase 2

Evolutionary Algorithm:

In this segment, the Particle Swarm Optimization (PSO) algorithm, were employed to identify an optimal subset of features. Initially, 1000 best features out of the 2832, as determined by the Fisher method, were selected for the PSO algorithm. The algorithm was designed with particles represented as 1000 binary arrays, where each element indicated whether a feature was included in the best features or not. Standard PSO procedures were followed for updating velocity and position.

The fitness function utilized a Random Forest classifier, with an additional penalty term for the number of features exceeding 100. This penalty term aimed to promote the selection of a more concise subset of features. The features selected by this algorithm were then employed in the implementation of both MLP and RBF networks.

Similar to the previous sections, the data was min-max normalized, and a 5-fold cross-validation approach was adopted. Different combinations of parameters were explored for both MLP and RBF networks. The results, including the best accuracies and predicted data, were appended to two arrays each, named `'all_best_accuracies_MLP_PSO'` and `'all_predicted_MLP_PSO'` for MLP, and `'all_best_accuracies_RBF_PSO'` and `'all_predicted_RBF_PSO'` for RBF. The final predicted data, chosen based on the highest achieved accuracy, was reported.

Results:

Best MLP network parameters using features selected by PSO:

```
Best MLP Network Configuration:
Number of Layers: 3
Neurons per Layer: [128, 64, 32]
Activation Function: ['relu', 'relu', 'sigmoid']
Average Cross-Validation Accuracy: 96.00%
```

Best RBF network parameters using features selected by PSO:

```
Best Parameters: {'C': 10, 'gamma': 0.1}
Average Cross-Validation Accuracy: 89.27%
```

Predicted values using MLP and PSO:

[0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1]

Predicted values using RBF and PSO:

[0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1]

Tina