CSE-535 Mobile Computing (Group 17)

Assignment 2 - Report

Gesture Recognition RESTful Application

Submitted to

Dr. Ayan Bannerjee

Ira A. Fulton School Engineering

Arizona State University.

Submitted by Group 17

Abhik Dey (1216907406)

Rohit Eppepalli (1215350630)

Sethu Manickam (1218452066)

Sree Vashini Ravichandran (1217841794)

Introduction:

Gesture Recognition project was aimed to develop an application to be deployed in the server where the users can use this application to find the gesture of each video. The main idea of this project is to get a hands on experience with video processing, machine learning and server deployment of an application using REST protocol.

Source of the data:

The practice videos recorded for 5 seconds each upon 6 chosen American Sign Language gestures (Buy, Communicate, Fun, Hope, Mother, Really) were used as the source. Image frames from each video were extracted and these frames were used to generate Key Points using PoseNet which were used as raw data to train the gesture recognition models.

Data pre-processing:

Data normalization was done with respect to the nose and eyes to avoid the effect of depth of the image as each person might place their cameras at different angles.

$$\text{Key points} = \text{Key points} - \text{Nose} / (\text{Left Eye} - \text{Right Eye})$$

Feature Extraction:

The features extracted are

1. Average of the positive and negative slopes (Slope considered between every 2 consecutive points)
2. Mean (The average that is used to calculate the central value of data among each column)
3. Min (Minimum value among the data in each column)
4. Max (Maximum value among the data in each column)
5. Variance (The average of the squared differences from the mean)

$$\text{Variance} = (\text{Value} - \text{Mean})^2$$

6. Fast Fourier Transform

   Fourier transform is used for continuous time series modeling to transform a periodic function into a set of simpler functions. The coefficients obtained from these periodic functions can be used to construct back the original data/signal against time series. Reason behind the selection of this feature is, when the coefficients of FFT are plotted against the frequency, we can infer from the graph that; maximum amplitudes at respective time frames denote the change of movements. Hence identifying the peaks which denote change of hand movements with significant changes at respective times can be used to identify different gestures.

7. Welch's Method

Welch's method is an approach for spectral density estimation, i.e. it can be used to estimate the power of the signal at different frequencies. Welch's method is based on the concept of periodogram (a common tool used for examining the amplitude vs frequency characteristics) that converts the signal from time domain to frequency domain. Welch's method reduces the noise in the power spectra in exchange for reducing the frequency resolution caused by imperfect and finite data, which makes Welch's method often desirable.

Models used:

- **Linear Discriminant Analysis:**

Linear Discriminant Analysis is a dimensionality reduction technique which is commonly used for supervised classification. It is used to project the features of higher dimension space into a lower dimension space. For multi class data that are having a large number of features for representation, the model gets confused to draw a partition line among the classes, with simple representations such as mean, variance that leads to a distribution, LDA can be used to better represent the data and draw partition lines among the classes.

The model's **accuracy** trained based on **Linear Discriminant Analysis is 69.23%**

- **Linear Regression:**

Linear Regression is used to model the relation between the dependent variable (scalar) and one or more independent variables (explanatory) by fitting a linear equation to observed data. Based on the scatter plot that was observed among the features, as there existed an association between the multiple classes, we have chosen to use Linear Regression which seems to have better accuracy compared to other models. Linear Regression seems to work better with image analysis compared to other classifiers due to the associativity nature of the images.

The **accuracy** of the model trained with **Linear Regression is 58.25%**

- **Multi-Layer Perceptron:**

Multilayer perceptron is a deep, artificial neural network. The name itself suggests that it is composed of multiple (more than one) perceptrons.

Perceptron: A perceptron is just a simple algorithm that is intended to perform binary classification; that is, it predicts whether an input belongs to that class or not by separating the input into two categories with a straight line. The mathematical representation of perceptron is:

$$y = \varphi\left(\sum_{i=1}^{n} w_i x_i + b\right) = \varphi\left(W^T X + b\right)$$

Where **ϕ** is the non-linear activation function, **W** denotes the weight of vector, **X** is the vector of inputs and **b** is the bias.

The MLP consists of an input layer that is used to receive the input signal, an output layer to provide the predicted result on the input with an arbitrary number of hidden layers that are the true computational engine of MLP. The best **accurate result** was of **Multi-Layer perceptron** which is **72.11%**

- **Support Vector Machine (Multi-class SVM):**

SVM is inherently two-class classifiers. The most traditional technique is to build one-versus rest classifiers and to choose the class which classifies the test datum with the greatest margin. SVM uses the multi-class formulation based on the following formulation,

$$\frac{1}{2}\sum_{i=1}^{k} w_i^2 + \frac{c}{n}\sum_{i=1}^{k} \xi_i$$

Show that, for all y in [1...k]: $[x_1 * w_{yi}] >= [x_1 * w_y] + 100 * \Delta(y_1, y) - \xi_1$.....
for all y in [1...k]: $[x_1 * w_{yn}] >= [x_1 * w_y] + 100 * \Delta(y_n, y) - \xi_n$

C is the usual regularization parameter that trades off margin size and training error. $\Delta(y_n,y)$ is the loss function that returns 0 if y_n equals y, and 1 otherwise.

The **accuracy** obtained by using **Multi-class SVM is 60.19%**

Web service (Heroku server):

To host the application on the server, we have used a free cloud platform offered by Heroku and pushed into Heroku GIT repository, once the application is run, we can access the application through the URL. The URL exposed to access the application is https://cse535-group17.herokuapp.com/upload. We have used an application called **Postman** to test our service, where it is enough to just expose the URL and phrase the body of request to the service, the application posts the request to the service and returns a response as designed by the application.

Contribution:

- Abhik Dey (1216907406) - **Linear Discriminant Analysis**
- Rohit Eppepalli (1215350630) - **Linear Regression**
- Sethu Manickam (1218452066) - **Multi-Layer Perceptron**

- Sree Vashini Ravichandran (1217841794) - **Support Vector Machine**

Conclusion:

Thus, based on the analysis of the data and due to the presence of multiple classes, the 4 different models chosen by us (group 17) are - Linear Discriminant Analysis, Linear Regression, Multi – Layer Perceptron and Support Vector Machine (multi-class), out of which MLP seems to show better accuracy with multi-layer perceptron that adjusts the weight based on the feedbacks of front and back propagation. The application was successfully hosted in the Heroku cloud service and it can be accessed anytime.