Project Title: Gesture recognition - The Jedi Master

Date: 19/11/2015

0458147 Denis Sedov

0458134 Sinan Kaplan

Abstract: *This paper simply introduces the developed method for Gesture recognition by testing on given data sets. Different kind of methods are applied to extract features and to classify the data. For feature extraction Principal Component Analysis is used while KNN is chosen for classification. The overall accuracy of the proposed method for classification is %85.*

1. INTRODUCTION

In the given task, it is asked to develop a classification application that simply classifies the given data. Therefore, it is important to understand the nature of the provided data to develop a method for the task [1]. Below, we simply analyzed the nature of the data according to the given set.

a) *Data Set Properties*

When examining the given data, it has been seen that each recorded move consists of three equal length vectors, which are conducted from the readings of three sensors at fixed period of time. And also it is obvious that each sample move takes different time to complete respectively each person. From the experiments it has been understood that the given data is time series data [2, 3 and 4].

When the nature of the given data set is understood, it is easy to implement any solution for the task. In next section we explained each step of design and implementation of our solution in detail.

2. DESIGN AND IMPLEMENTATION OF THE PROPOSED METHOD

The general steps of our solution is shown figure 1. It simply consists of four main steps which are data preprocessing, feature selection, feature extraction and classification.

2.1. Data Preprocessing

At this step, the data is normalized to scale each sample in the data. Normalization of the data is important to get better and reliable classification results. We applied normalization by using subtraction of mean. Snapshot of the code in matlab that is implemented for the normalization can be seen below.

```
15    s = load(file_name, '-ascii');
16    % Normalization(subtraction of mean)
17    mn = mean(s);
18    vr = var(s);
19    s = s - repmat(mn,size(s,1),1);
```
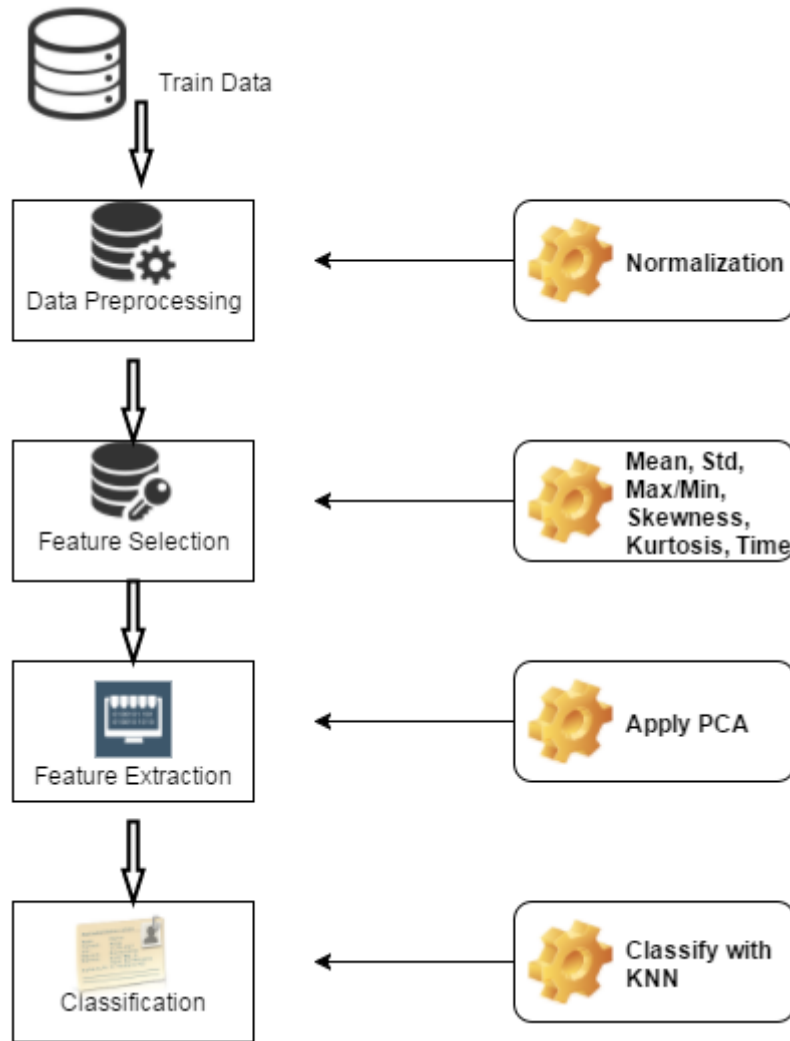
Fig.1. Flow Chart of Proposed Solution

2.2. Feature Selection

The selection of feature plays an important role on each classification tasks. Hence, it is important to identify the features according to the nature of the data. As we have studied our data and it is time series data, at this point it is more logical to use some special features of time series data rather than any feature. The study in [5] has been chosen as a reference point at this step to select proper features. Also the question, which has been asked in [6], was quite helpful to have some ideas about features in time series data.

The most important challenge is that the moves in the given data set take different amount of time because of that we need think about extracting data from them in such a way the feature vectors should have same lengths. To solve this problem, each data set is divide into 2, 3, 4 and 5 parts and then combine all the features from those sets. The main reason behind

the dividing data several times (2, 3, 4 and 5) is to get more accurate feature information. After that by taking advantage of aforementioned references for feature selection, the listed features below are selected to use in classifier.

- ➢ Time
- ➢ Mean
- ➢ Standard Deviation
- ➢ Maximum and Minimum values
- ➢ Skewness and Kurtosis

2.3. Feature Extraction

Once the feature selection is done, it is time to think about how to reduce dimensionality by applying a transformation to the data to project it into a new feature space with lower dimension. For this purpose, PCA (Principal Component Analysis) [7, 8] is used. One challenge with PCA method is how to choose k component. To decide number of k (number of components) we have examined the eigenvectors in such a way that we can retain some fraction variance %98.

2.4. Classification

As the nature of data set is identified, next step is to choose proper methods to classify the time series data. Based on the studies in [3, 4], KNN method is used to classify the given data. The method is chosen because of its simplicity. To decide the number of neighbors to use some experimental studies are conducted. More details are explained in next section.

3. EXPERIMENTAL EVALUATION

During the testing stage, we first tested how chosen features affect the classification result. After the initial tests, we found that Skewness, Kurtosis and Time have no noticeable effects on the classification task. To reduce the dimensionality of the feature vector and also computational cost of the algorithm we have decided to not to use those features.

Once the feature vector is generated, next step was to apply PCA analysis for feature extraction. To do so we choose the components whose eigenvalues divided by total eigenvalues is equal to %98. To decide this value, related tests are conducted as it can be seen below table.

Variance Fraction (%)	Classification Accuracy (%)	Number of Components
90	83,7	14
95	83,7	20
98	85,3	26

When feature selection and extraction experiments were conducted, following stage was to choose appropriate k (neighbor) value for K-NN classification algorithm. All the conducted tests k value is given below table. Classification accuracy is tested according to the *leave-one-out* method on 226 provided data samples.

Table 1. Classification Results with Different k Values

K value (KNN)	Correctly Classified Samples	Incorrectly Classified Samples	Classification Accuracy (%)
1	194	32	85,8
3	193	33	85,3
5	184	42	81,4
7	172	54	76,1
9	158	68	69,9
11	145	81	64,1

The table above shows that the more neighbors we take the worst result we get. And also when we have just 1 neighbor it causes more outliers even it provides the best accuracy. To overcome the outlier problem in the developed classifier, k value is assigned to 3 as it provides accuracy %85.

4. CONCLUSION

According to the given task, we have implemented aforementioned solution for the task by using PCA for feature extraction and kNN for classification. To solve the given problem it is quite important to understand the nature of the data. That is why we have analyzed given data first, and then try to select features based on the nature of the data. As our data is an example of time series data we have select the features based on that assumption. Also data is divided several times into different subsets (2, 3, 4 and 5) to get equal length of feature vector from each observation. Mean, standard deviation, minimum/maximum values are used as a feature

vector. Time, Skewness and Kurtosis are not used because they did not have effects on the classification results. PCA is used to reduce dimensionality of the feature vector . As a number of neighbor k is assigned to value 3. After all tests are conducted, the accuracy of the implemented solution is %85 by performing the classification using leave-one-out testing method.

References

[1]. Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "From data mining to knowledge discovery in databases." *AI magazine* 17.3 (1996): 37.

[2]. Brillinger, David R. *Time series: data analysis and theory*. Vol. 36. Siam, 1981.

[3]. [Online Available], Tarek Amr, "Survey on Time-Series Data Classification", http://tarekamr.appspot.com/pdfs/tsdm2012.pdf

[4]. Chaovalitwongse, Wanpracha Art, Ya-Ju Fan, and Rajesh C. Sachdeo. "On the time series k-nearest neighbor classification of abnormal brain activity."*Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on* 37.6 (2007): 1005-1016.

[5]. Nanopoulos, Alex, Rob Alcock, and Yannis Manolopoulos. "Feature-based classification of time-series data." *International Journal of Computer Research*10.3 (2001).

[6]. [Online Available], Features of Time Series Classification, http://stats.stackexchange.com/questions/50807/features-for-time-series-classification

[7]. [Online Available], Lindsay I Smith, "A tutorial on Principal Components Analysis", http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

[8]. [Online Available], Principal Component Analysis for Dummies, https://georgemdallas.wordpress.com/2013/10/30/principal-component-analysis-4-dummies-eigenvectors-eigenvalues-and-dimension-reduction/