**Human Activity Recognition Using Smartphones Data Set**

**Abstract**

Wearable smart devices are getting more and more popular nowadays. By wearing these devices everyday, we could easily collect data and perform further analysis. In spite of the abundance of available data from these sensors, there isn’t too much information we can tell from these raw data about human’s activity. In this project, our goal is to recognize patterns from these raw data, and extract useful information about the user’s daily activities. Activity-Based Computing [1] aims to capture the state of the user and its environment by exploiting heterogeneous sensors in order to provide adaptation to exogenous computing resources. When these sensors are attached to the subject's body, they permit continuous monitoring of numerous physiological signals. This has appealing use in health care applications, e.g. the exploitation of Ambient Intelligence in daily activity monitoring for elderly people.This report discusses various possible classifiers, and compare them to seek the best performing one. We try to find another more effective possible algorithm, SVM with different kernels, and find SVM with linear kernel is able to provide more precise performance.

**1. Introduction**

The use of smartphones has become an inalienable part of everyday life of people of almost all ages globally. 91% of the world population owns a mobile device.Besides the basic function of telephony, smartphones provides many other features. Out of all the features, multitasking and the deployment of various sensors, are currently merged to the current handsets. It is pictured that with those features, we are capable to use smartphones to keep track of our activities. We can learn from them, and later assist us to make better decisions in future. In this report, we apply various algorithms to analyze those data of accelerometer and gyroscope, and then predict and recognize six human activities (walking, walking upstairs, walking downstairs, sitting, standing, and laying).

The report is structured in the following way: The state of the art regarding previous work is depicted in Section 2. The description of the data set is given in Section 3. The description of the adopted methods and comparison of the analyses results is presented in Section 4. The explanation of the best performance to distinguish between the rest state and other states using SVM with linear kernel classifier is presented in Section 5. Lastly, the conclusions of this research project are described in Section 6.

**2. Background and Related Work**

There has been a variety of work done in this field. Very early work tries to identify human activities based on different kinds of data source [5]. Some data comes form sensors like bodywom sensors, which may not give much information about human activity [2]. Whereas some data comes from specialized devices, which are costly, uneasy for users and become a headache to the user during regular activity [3]. Thus more and more researches are focusing on using wearable smart devices’ data to classify human activity[4]. In this project, we used a dataset obtained from smartphones containing data related to both activities and postural transitions.

Bao et al. [6] developed algorithms to detect bodily activities from everyday tasks, and observed that while some activities are classified more accurately with subject-independent training data, others require subject-specific training data. This suggests that multiple sensors aid in recognition because conjunctions in acceleration feature values can help to identify many activities.

**3. Data Set**

The dataset includes all triaxial acceleration from the accelerometer and triaxial angular velocity from the gyroscope at a constant rate of 50 Hz captured integrated in the smartphones. We have 561 features. UCI link is: <https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>

The experiments have been carried out with a group of 30 volunteers within an age range of 19-48 years. Each person performed six activities (SITTING, STANDING, LAYING, WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS) wearing a smartphone (Samsung Galaxy S II) on the waist.These six activities as six classes are corresponding to each data point with 561 features.

1. **Approach**

There are several algorithms which are capable for multiclass prediction of the dataset. We’ve used LDA, Logistic Regression, K-Nearest Neighbors and SVM. We compare the confusion matrix of those classification results on the test data using the diffrent training classifiers, presented by confusion matrices.

LDA

Logistic Regression

KNN

Different kernels for SVM

1. **Results**

Confusion Matrix of the classification results on the test data using the required classifier are shown below. Rows represent the actual class and columns the predicted class. The diagonal entries show the number of test samples correctly classified.

**LDA:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Activity | Walking | Upstairs | Downstairs | Sitting | Standing | Laying | Recall % |
| Walking | **490** | 6 | 0 | 0 | 0 | 0 | 98.79 |
| Upstairs | 11 | **460** | 0 | 0 | 0 | 0 | 97.66 |
| Downstairs | 1 | 14 | **405** | 0 | 0 | 0 | 96.43 |
| Sitting | 0 | 1 | 0 | **434** | 56 | 0 | 88.39 |
| Standing | 0 | 0 | 0 | 22 | **510** | 0 | 95.86 |
| Laying | 0 | 0 | 0 | 0 | 0 | **537** | 100 |
| Precision % | 97.61 | 95.63 | 100 | 95.18 | 90.11 | 100 | **96.23** |

**Logistic Regression:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Activity | Walking | Upstairs | Downstairs | Sitting | Standing | Laying | Recall % |
| Walking | **494** | 0 | 2 | 0 | 0 | 0 | 99.6 |
| Upstairs | 23 | **448** | 0 | 0 | 0 | 0 | 95.12 |
| Downstairs | 4 | 9 | **407** | 0 | 0 | 0 | 96.9 |
| Sitting | 0 | 4 | 0 | **432** | 55 | 0 | 87.98 |
| Standing | 2 | 0 | 0 | 13 | **517** | 0 | 97.18 |
| Laying | 0 | 0 | 0 | 0 | 0 | **537** | 100 |
| Precision % | 94.46 | 97.18 | 99.51 | 97.08 | 90.38 | 100 | **96.2** |

**KNN:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Activity | Walking | Upstairs | Downstairs | Sitting | Standing | Laying | Recall % |
| Walking | **486** | 1 | 9 | 0 | 0 | 0 | 97.98 |
| Upstairs | 42 | **426** | 3 | 0 | 0 | 0 | 90.45 |
| Downstairs | 51 | 42 | **327** | 0 | 0 | 0 | 77.86 |
| Sitting | 0 | 4 | 0 | **420** | 67 | 0 | 85.54 |
| Standing | 0 | 0 | 0 | 51 | **481** | 0 | 90.41 |
| Laying | 0 | 0 | 0 | 2 | 1 | **534** | 99.44 |
| Precision % | 83.94 | 90.46 | 96.46 | 88.79 | 87.61 | 100 | **90.74** |

**SVM for rbf kernel:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Activity | Walking | Upstairs | Downstairs | Sitting | Standing | Laying | Recall % |
| Walking | **492** | 0 | 4 | 0 | 0 | 0 | 99.19 |
| Upstairs | 17 | **452** | 2 | 0 | 0 | 0 | 95.19 |
| Downstairs | 13 | 29 | **378** | 0 | 0 | 0 | 90 |
| Sitting | 0 | 2 | 0 | **424** | 65 | 0 | 86.35 |
| Standing | 0 | 0 | 0 | 44 | **488** | 0 | 91.73 |
| Laying | 0 | 0 | 0 | 0 | 0 | **537** | 100 |
| Precision % | 94.25 | 93.58 | 98.44 | 90.6 | 88.25 | 100 | **94.03** |

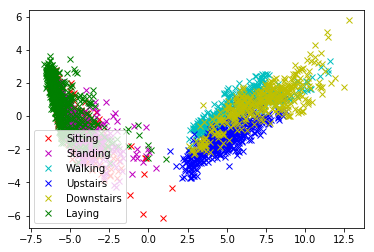
**SVM for linear kernel:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Activity | Walking | Upstairs | Downstairs | Sitting | Standing | Laying | Recall % |
| Walking | **492** | 1 | 3 | 0 | 0 | 0 | 99.19 |
| Upstairs | 18 | **451** | 2 | 0 | 0 | 0 | 95.75 |
| Downstairs | 4 | 6 | **410** | 0 | 0 | 0 | 97.62 |
| Sitting | 0 | 2 | 0 | **435** | 54 | 0 | 88.59 |
| Standing | 0 | 2 | 0 | 16 | **516** | 0 | 96.99 |
| Laying | 0 | 0 | 0 | 0 | 0 | **537** | 100 |
| Precision % | 95.72 | 98.04 | 98.8 | 96.45 | 90.53 | 100 | **96.4** |

**SVM for poly kernel:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Activity | Walking | Upstairs | Downstairs | Sitting | Standing | Laying | Recall % |
| Walking | **491** | 0 | 5 | 0 | 0 | 0 | 98.99 |
| Upstairs | 38 | **424** | 9 | 0 | 0 | 0 | 90.02 |
| Downstairs | 54 | 44 | **322** | 0 | 0 | 0 | 76.67 |
| Sitting | 0 | 4 | 0 | **410** | 77 | 0 | 83.5 |
| Standing | 0 | 1 | 0 | 41 | **490** | 0 | 92.11 |
| Laying | 0 | 0 | 0 | 0 | 0 | **537** | 100 |
| Precision % | 95.53 | 92.17 | 77.59 | 90.91 | 85.96 | 100 | **90.74** |

We first decompose the test data to 2-dimension in order to plot the classification as shown in the given figure. We can easily figure out that the dataset is more linear separable, and we assume that SVM with linear kernel function will get the highest accuracy among other two functions.



1. **Discussion**

So, we have successfully implemented LDA, Logistic Regression, K-Neighbors Classifier and SVM to our dataset. Now wecan compare the results for all the above mentioned algorithms and can find the best fit classifier for Human Activity Recognition Dataset. The table shows the comparison of the classification results on the test data using the training related classifiers.

|  |  |  |
| --- | --- | --- |
| Classifier | Accuracy | Error Rate |
| SVM with linear kernel | 96.4 | 3.60 |
| LDA | 96.23 | 3.77 |
| Logistic Regression | 96.2 | 3.80 |
| KNN | 90.74 | 9.26 |

Hence, we can say that SVM with linear kernel approach is the best fit classifier to our dataset. The SVM with linear kernel gives the best performance for all the accuracies and error rates. SVM with linear kernel being a flexible approach which is capable to reduce overfitting. And SVM with linear kernel performs astonishingly well for this dataset as linear separable data. But for other different data, it would not have such good performance.

1. **Conclusion**

In this report, we discusses and examine various classifiers (LDA, Logistic Regression and K-Nearest Neighbors) and their performance in classifying human activities using smartphones, and find that their accuracies are quite different from each other. We tested different approaches and algorithms on our dataset. We finds that on changing the kernel, SVM classifier gives different results and amazingly find SVM with linear kernel provides a much more fine accuracy than all the other classifiers used.

**References**

1. Davies, N., Siewiorek, D.P., Sukthankar, R.: Activity-based computing. IEEE Pervasive Computing 7(2) (April 2008) 20-21
2. R. Poppe. Vision-based human motion analysis: An overview Computer vision and image understanding, 108(1):4–18, 2007.
3. N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman. Activity recognition from accelerometer data. In AAAI,volume 5, pages 1541–1546, 2005
4. J.-L. Reyes-Ortiz, L. Oneto, A. Ghio, A. Sama, D. Anguita, and X. Parra. Human activity ecognition on smartphones with awareness of basic activities and postural transitions.In Artificial Neural Networks and Machine Learning–ICANN 2014, pages 177–184. Springer, 2014
5. C.-C. Yang and Y.-L. Hsu. A review of accelerometry-based wearable motion detectors for hysical activity monitoring. Sensors , 10(8):7772–7788, 2010.

6 Bao, Ling and Stephen S. Intille. Activity recognition from user-annotated acceleration data, 2004