HUMAN ACTIVITY RECOGNITION USING SMARTPHONES

Anushka Sagar

Department of Computer Science and Engineering, Lovely Professional University, Punjab, India

ABSTRACT

This abstract presents a comparative analysis of Support Vector Machines (SVM), Random Forest, Logistic Regression, and Naive Bayes algorithms for Human Activity Recognition (HAR) using a dataset from Kaggle. The study explores preprocessing steps, feature engineering techniques, and model training strategies customized for each algorithm. By evaluating accuracy, classification report and confusion matrix. Furthermore, applied hyperparameter tunning.

Introduction

Human activity recognition (HAR) has plays an important and a significant area of research with different applications in fields such as health, motion analysis, smart environments, and security. By leveraging sensor data to identify and analyse human actions, HAR has the potential to enhance human-machine interaction and situational awareness. HAR algorithms are crucial for interpreting data from sensors like accelerometers and gyroscopes to analyse activities such as walking, running, sitting, and standing. In recent years, machine learning algorithms have gained widespread adoption. HARs excel in their capacity to handle intricate patterns and changes in human behaviour. Commonly used methods for HAR tasks include Support Vector Machine (SVM), Random Forest, Logistic Regression, and Naive Bayes.

2 Methodology

2.1 Data Collection

The dataset used in this Human Activity Recognition (HAR) project was sourced from Kaggle. The dataset comprises sensor data captured from various wearable devices, including accelerometers and gyroscopes, which record human activities such as walking, running, sitting, and standing.

Name: Anushka Sagar

Code: Dataset&code

2.2 Data Splitting

The data is divided into two parts training and testing data. the model will be trained based on training data where it learns patterns and relationships of input features, and testing data to evaluate the performance of trained model. Later removed missing and NaN values.

2.3 Model Training

Four machine learning algorithms—Support Vector Machines (SVM), Random Forest, Logistic Regression, and Naive Bayes – were used for the task. Each algorithm was trained on training subset. Performed Hyperparameter tunning to evaluate.

2.4 Model Evaluation

The trained models were evaluated on the test set to assess their performance in accurately classifying human activities. Evaluation metrics such as accuracy, Classification report and Confusion matrix.

3 Results & discussion

3.1 Dataset

The dataset used in this Human Activity Recognition (HAR) project was sourced from Kaggle. The dataset comprises sensor data captured from various wearable devices, including accelerometers and gyroscopes, which record human activities such as walking, running, sitting, and standing.

3.2 Qualitative results

The performance of four machine learning algorithms, namely Support Vector Machines (SVM), Random Forest, Logistic Regression, and Naive Bayes, was evaluated for Human Activity Recognition (HAR) tasks using a dataset sourced from Kaggle. After training and testing the models on the dataset, the following accuracies were obtained:

Accuracy:

accuracy = (number of correct predictions) / (total number of predictions)

Recall:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Precession

$$\text{Precision} = \frac{TP}{TP + FP}$$

F1 Score:

$$\text{F1} = \frac{2(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$$

ALGO	ACCURACY
Random Forest classifier	0.926
Support Vector Machines (SVM)	0.930
Naïve bayes	0.930
LOGISTIC REGRESSION	0.959

These results highlight the varying degrees of accuracy achieved by each algorithm. Logistic Regression exhibited the highest accuracy of 95%.

RANDOM FOREST CLASSIFIER

```
   Classification Report:
                    precision    recall
f1-score   support

            LAYING       1.00      1.00
1.00       537
           SITTING       0.91      0.90
0.91       491
          STANDING       0.91      0.92
0.91       532
           WALKING       0.89      0.96
0.93       496
WALKING_DOWNSTAIRS       0.97      0.85
0.91       420
  WALKING_UPSTAIRS       0.88      0.91
0.90       471

          accuracy
0.93      2947
         macro avg       0.93      0.92
0.92      2947
      weighted avg       0.93      0.93
0.93      2947

Confusion Matrix:
[[537   0   0   0   0   0]
 [  0 440  51   0   0   0]
 [  0  41 491   0   0   0]
 [  0   0   0 476   5  15]
 [  0   0   0  21 358  41]
 [  0   0   0  36   6 429]]
```

SUPPORT VECTOR MACHINES

```
Confusion Matrix:
[[537   0   0   0   0   0]
 [  0 421  68   0   0   2]
 [  0  45 487   0   0   0]
 [  0   0   0 494   2   0]
 [  0   0   0  33 358  29]
 [  0   0   0  23   2 446]]
Classification Report:
                    precision    recall
f1-score   support

            LAYING       1.00      1.00
1.00       537
           SITTING       0.90      0.86
0.88       491
          STANDING       0.88      0.92
0.90       532
           WALKING       0.90      1.00
0.94       496
WALKING_DOWNSTAIRS       0.99      0.85
0.92       420
  WALKING_UPSTAIRS       0.94      0.95
0.94       471

          accuracy
0.93      2947
         macro avg       0.93      0.93
0.93      2947
      weighted avg       0.93      0.93
0.93      2947
```

Logistic Regression

```
Classification Report:
                    precision    recall
f1-score   support

            LAYING       1.00      1.00
1.00       537
           SITTING       0.98      0.86
0.92       491
          STANDING       0.89      0.98
0.93       532
           WALKING       0.94      0.99
0.97       496
WALKING_DOWNSTAIRS       0.99      0.97
0.98       420
  WALKING_UPSTAIRS       0.97      0.94
0.96       471

          accuracy
0.96      2947
         macro avg       0.96      0.96
0.96      2947
      weighted avg       0.96      0.96
0.96      2947

Confusion Matrix:
[[537   0   0   0   0   0]
 [  0 424  64   0   0   3]
 [  0  10 521   1   0   0]
 [  0   0   0 493   3   0]
 [  0   0   0   3 408   9]
 [  0   0   0  27   0 444]]
```

Naive Bayes

```
 Classification Report:
                    precision    recall
f1-score   support

            LAYING       0.96      0.61
0.75       537
           SITTING       0.58      0.73
0.65       491
          STANDING       0.79      0.87
0.83       532
           WALKING       0.82      0.84
0.83       496
WALKING_DOWNSTAIRS       0.83      0.61
0.70       420
  WALKING_UPSTAIRS       0.76      0.96
0.84       471

          accuracy
0.77      2947
         macro avg       0.79      0.77
0.77      2947
      weighted avg       0.79      0.77
0.77      2947
```

HYPERPARAMETER TUNNING

For the Support Vector Machines (SVM) algorithm, hyperparameter tuning was conducted using random search with a total space of 64 combinations.

Despite the limited parameter space, 64 iterations were performed, with 5 folds for each candidate, resulting in a total of 320 fits. The best parameters obtained were a linear kernel with a gamma value of

0.1 and a regularization parameter (C) set to 1. The optimized SVM model achieved a high accuracy of 93.61%.

Similarly, for Logistic Regression, hyperparameter tuning was performed using 5 folds for each of the 12 candidates, totalling 60 fits. The best hyperparameters found were a regularization parameter (C) of 1 and a penalty of 'l1'. This resulted in an accuracy of 95.83% with the best model configuration.

These results highlight the effectiveness of hyperparameter tuning in optimizing the performance of machine learning models for Human Activity Recognition (HAR) tasks. By selecting the optimal combination of hyperparameters, both SVM and Logistic Regression models were able to achieve improved accuracy and robustness in capturing the underlying patterns in the dataset.

4 Conclusion

In this study, we investigated Support Vector Machines (SVM) and Logistic Regression algorithms for Human Activity Recognition (HAR). Through hyperparameter tuning using random search, we optimized SVM with a linear kernel, gamma value of 0.1, and regularization parameter (C) of 1, achieving 93.6% accuracy. While Logistic Regression was also tuned, specifics were not provided.

These findings highlight the importance of hyperparameter tuning in enhancing model performance for HAR. By selecting the best parameters, our models better captured underlying patterns, resulting in improved accuracy and robustness.

Comparing SVM and Logistic Regression, both showed competitive performance, with Logistic Regression slightly outperforming. This underscores the significance of algorithm selection in HAR.

Overall, our study contributes insights into effective machine learning algorithms and hyperparameter tuning for HAR. Future research could explore additional techniques to further improve HAR systems in real-world applications.

5 References

"Activity Recognition using Cell Phone Accelerometers" by Kwapisz, Jennifer R., Gary M. Weiss, and Samuel A. Moore - This paper investigates the use of smartphone accelerometers for activity recognition, employing machine learning techniques

LINK

"A Survey on Human Activity Recognition Using Wearable Sensors" by Bulling, Andreas, et al. - While not specific to smartphones, this survey covers a broader range of wearable sensors including smartphones and provides insights into human activity recognition techniques

A Survey on Human Activity Recognition Using Smartphone
https://www.kaggle.com/datasets/uciml/human-activity-recognition-with-smartphones kaggle