Activity classification based on human sensory (Final Report)

Ayush Yadav
2017335
IIIT-Delhi, India
ayush17335@iiitd.ac.in

Piyush Aggarwal
2017356
IIIT-Delhi, India
piyush17356@iiitd.ac.in

Manish Kumar
2017245
IIIT-Delhi, India
manish17245@iiitd.ac.in

1. Abstract

Classifying human activities such as running, cycling, sitting, etc on the basis of IMU (Inertial Measurement Unit) readings and heart rate data from user's hand, chest and ankle is an important task for both patients and health organisations. In this project, we classified human activities using machine learning models such as logistic regression, decision tree, random forest, SVM and Multi-Layer Perceptron and also compared the accuracy rates of all models. To compare the classification techniques, we changed the dataset size (due to computational problems), changed the hyperparameter values to find the value at which models give the best accuracy. SVM with RBF kernel gives the best accuracy as it can classify the data irrespective of the separability of data using kernel trick.

2. Introduction

Activity Classifying human activities such as sitting, walking, running, cycling, sleeping, etc on the basis of using heart rate data and IMU readings from a user's hand, chest, and ankle. Activity recognition is an important task in several healthcare applications. By continuously monitoring and analyzing user activity it is possible to provide automated recommendations to both patients and doctors. There are also applications to consumer products such as data logging for smartwatches health apps.

3. Related Work

Due to the multiple applications, human activity classification is a popular research area. Related research articles that use similar approaches are:

Panu Korpipa Juha Parkka Miikka Ermes. "Activity Classification Using Realistic Data From Wearable Sensors". In: IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE. IEEE, 2006. In this study used an ordinary decision tree grown via cross-validation and using the Gini Loss for each split. One strength of the article is that, in addition to this "automatically generated decision tree," the researchers also created a "custom decision tree". However, the classes in the article are improperly balanced (with one class accounting for between 50 and 60 percent of all of the data), which could affect the true test performance (assuming that the test data also suffers from the same problem). The paper by Youngwook Kim and Hao Ling discusses an interesting approach of human activity recognition through data obtained by a Doppler radar. Similar to our approach, this paper incorporates SVM models that are tuned through cross-validation over a range of hyperparameter values to pick an optimal one.

Other research are: Attila Reiss and Didier Stricker German Research Center for Artificial Intelligence (DFKI). "Creating and Benchmarking a New Dataset for Physical Activity Monitoring". In this study used decision tree with following variants: normal, boosted decision tree and bagging decision tree. They also did Naive Bayes and kNN. They achieved good results but the time taken to train the models was high and time taken to test the models was higher. Compared to them, we tried more types of classifiers to see if there is a better alternative in the sense of complexity and training time and found that SVM can perform better. In addition to this, random forest also performed well.

4. Methodology

i. Dataset and Evaluation

The total number of data collected has 2850505 samples with 52 features. The data had 23 classes. It had certain transient data, NAN data, etc that needed pre-processing.

a. Pre-processing

There were useless data that were identified with '0' under the activity column i.e. These data points had no activity indicated in the data. (Basically, the data collected in between activities) Those data points were removed.

Next, there were values with NAN, these were handled by replacing them with the mean of the non-NAN values of that column.

Next, we did feature scaling to a range and scaled all the data to the range of (0,1). This was done by replacing each value by subtracting it with the mean of that column and then dividing by the standard deviation.

$$x' = \frac{x - \text{mean}(x)}{\max(x) - \min(x)}$$

After removing that total number of data becomes 1958406 and number of features remains the same. Data now has 12 classes after removing the transient data.

We divided the data into a training set and testing set in the proportion of 5:1. In order to get fair results, in both the training and the testing data, the number of samples from each class was the same i.e. same no of samples from each class. We relabeled all the required classes that we have used in the dataset.

Due to computational problems, we trained the models on a smaller set with 600000 training examples i.e. 50000 from each class and 120000 testing examples i.e. 10000 from each class.

Below is a summary of the data in tabular form:

Total no. of samples	2850505
No. of features	52
Total number of classes	23
Sample size after reduction	1958406
No. of classes left	12
Training size	600000
Testing size	120000

b. Evaluation Metrics

Since IMU data was not balanced, so in addition to accuracy, we also used the confusion matrix to find the true positive, true negative, false positives, and false negatives and try to reduce the false positive and false negatives.

In addition to confusion matrix, precision and recall are also useful to find the correctness of the model. And since both precision and recall work differently than each other, the f1 score is also a good choice to perform the evaluation.

ii. Logistic Regression:

Since the task is to classify between different activities performed by a human, first we used logistic regression with both Gradient Descent and Stochastic Gradient Descent. The learning rate chosen was 0.5 which worked best as the loss continuously decreases and at a good rate.

We tried logistic regression with both L1 and L2 regularization but as the L1 regularization works slow on a large dataset and gives an almost similar result. Therefore, we used the L2 regularization.

For regression, softmax was used to estimate the probability of classes, and cross-entropy loss is used.

iii. Support Vector Machine

Then we used SVM since it can classify the data irrespective of the separability of data using kernel trick. Since SVM uses quite a bit of parameters, we used k-fold cross-validation in order to find out the best parameters for our model.

We tried various kernels, linear kernel, RBF kernel, polynomial kernel with different degrees, and we also tried them with various hyperparameters C.

Also, we run all the folds on both one-vs-one and one-vs-all. Since it performed equally well on both of them, we used one-vs-all to perform the test.

We found that the RBF kernel with hyperparameter C=10 gives the best accuracy on the testing data of that fold. So we used it to train on the complete dataset and find the accuracy of this model.

iv. Decision Trees

For a dataset having non-linear feature sets and multiple classes, we can use decision trees as it perform greedy splits with a specific threshold on each feature of the data.

We also used pruning technique to reduce the size of the tree. Pruning technique removes the sections of the tree that contributes very less towards the classification which further reduces the complexity, overfitting and improves the accuracy.

v. Random Forest

Since Decision tree clearly overfits and gives low testing accuracy and high training accuracy, we tried to use the random forest to check how multiple decision trees would increase the testing accuracy. After trying with different no of trees (estimators), we calculated the results with 500 trees.

vi. Multi-Layer Perceptron

Finally, we decided to use a deep learning algorithm so we tried a multi-layer perceptron.

The loss function used was Cross entropy loss. After trying on various on various different variations of the number of hidden layers and the number of neurons in each of them and considering the computational complexity because of the large dataset, we used 3 hidden layers.

Input layers: 52 neurons
Hidden Layer 1: 64 neurons
Hidden Layer 2: 38 neurons
Hidden Layer 3: 25 neurons
Output Layer: 12 neurons

Similarly, through k-fold cross-validation, we found that the model worked best with a learning rate=0.005.

For the activation functions, ReLU activation function was used for all layers except the last layer because at the last layer we need to limit the value to 1 because it indicates the probability.

```
f(x)=max(0,x)
```

So for the last layer, Softmax activation function was used.

$$f_i(\vec{a}) = \frac{e^{a_i}}{\sum_k e^{a_k}}$$

To avoid overfitting, dropout was introduced at each hidden layer with dropout probability values 0.05, 0.1, 0.125 at each layer. Again these values are not random but tested and these worked out best.

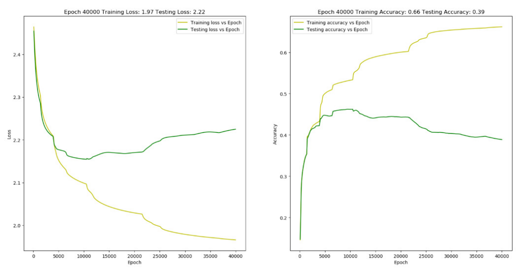
5. Result

Method	Accuracy	
	Training Set	**Testing Set**
Logistic Regression	66.17%	38.87%
Logistic Regression(SGD)	52.71%	44.84%
Linear SVM	82.3%	55.74%
SVM with RBF Kernel	92.5%	90.5%
SVM with Polynomial Kernel	90.6%	87.65%
Decision Tree	59.67%	51.81%

Random Forest	88.99%	76.71%
Multi-Layer Perceptron (MLP)	87.50%	75.36%

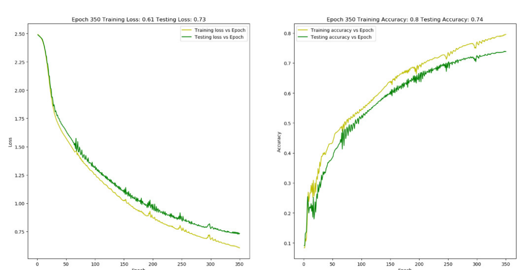
Clearly, logistic regression performed the poorest due to the inseparability of data and RBF kernel performed the best amongst all three kernels.

The graph below shows the accuracy and loss at different epoch value for the logistic model, we can clearly observe that even at higher epoch, the accuracy is very low.



For linear, RBF and polynomial SVM, we applied k-fold cross-validation with different hyperparameters and added the best results in the table. Since while pre-processing, we shuffled the data, results are varying in the close range when the model is applied to new subsamples.

The graph below shows the accuracy and loss curve for Multi-layer perceptron which performed better than random forest and logistic regression but not as efficient as SVM model.



RBF performed the best due to the fact that it generated non-linear boundaries.

6. Conclusion

This work presented the comparison of different classification models on human sensory data. The data consists of 12 different activities with 52 features. 9 subjects performed activities and their heart rate, IMU ankle, IMU chest and IMU hand was used for evaluation. Data was normalised and labels were corrected and other things were handled for further classification. 5 different classifiers with different variations were performed.

Analysing the data and results shows us that logistic regression performed worst which was expected because data was not linearly separable. Decision tree performed slightly better and did better with pruning. Multilayer perceptron outperformed both by big margin. Random forest performed slightly better than multilayer perceptron. And finally SVM performs best with rbf kernel.

References

1. https://s3.amazonaws.com/academia.edu.documents/12195796/parkka06.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1544770601&Signature=QBZmGI0nx4\%2FLEqsaBbGl6Sm1AXQ\%3D&response-content-disposition=inline\%3B\%20filename\%3DActivityClassificationUsingRealistic.pdf.
2. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4801689&tag=1
3. https://www.researchgate.net/publication/235348488_Creating_and_Benchmarking_a_New_Dataset_for_Physical_Activity_Monitoring