HUMAN ACTIVITY MONITOR VIA SMARTPHONE

Sidhant Naveria, Abhineet Mishra

Institute of Systems Science, National University of Singapore, Singapore 119615

ABSTRACT

In a specific region with current application it is very difficult to track a person's daily commute. There is no application which can tell if a person is commuting through a public transport or personal vehicle in a given region.We will be using mobile phone sensors to measure the parameters on which, system can predict the mode of transport. Using mobile sensors is a best choice as every one carries a mobile phone and all the smart phone carries sensors like gyroscope, accelerometer which can be used for our purpose.we propose to work with sensor related data acquire using phone sensors like "accelerometer","gyroscope" etc and predict the commute of the person in that region.

***Index Terms*—** One, two, three, four, five

1. INTRODUCTION

Since the appearance of the first commercial hand-held mobile phones in 1979, it has been observed an accelerated growth in the mobile phone market which has reached by 2011 near 80 percent of the world population [2]. It is clear that using mobile phone sensors is a correct option as almost everyone has a phone and everyone keep the mobile phone with them when they head out.These smarts phones are equipped with necessary hardware with high computation power, which is very important in our case

Our experiment aims at using the data collected through mobile phone sensors and train the model using the collected data to predict the commute method of the person accurately.This model can be used in multiple sectors like infrastructure improvement(improve the frequency of the public transport in rush hours), surveillance etc.In the experiment, we will be using multiple models to predict the commute mode and we will compare the findings to get the best.

2. RELATED WORK

We have found a work in github which is related to predicting human activity using mobile sensor data [1].This work is related to the predictions like if the person is walking or is the person is going upstairs or downstairs. we have used it as a base to understand processing of the signals in machine learning and we have build models to predict mode of transport using similar sensor's signal data

3. PROPOSED APPROACH

we are using mobile sensor data from kaggle website. Data contains columns like accelerometer.min, accelerometer.max, accelerometer.mean etc. we have drop the min and max columns and used mean column instead. we have done the same for rest of the sensor's columns. we have done the normalization of the data to bring the data to the common scale. As the data in the target column is categorical, we have used category coding to transform it. This is a multi-class classification problem and we have used DecisionTreeClassifier, RandomForestClassifier, SVM.SVC algorithms to train the model the model.we have used model accuracy, Precision Recall and F1 to score the model and compare the models with each other.

After the preprocessing of the data, we have applied the DecisionTreeClassifier and calculate the accuracy of the model. According to the article in the wikipedia , Random forest algorithm applies the general technique of bootstrap aggregating, or bagging, to tree learners. Given a training set X = x1, ..., xn with responses Y = y1, ..., yn, bagging repeatedly (B times) selects a random sample with replacement of the training set and fits trees to these samples.Bagging procedure leads to better model performance because it decreases the variance of the model, without increasing the bias. This means that while the predictions of a single tree are highly sensitive to noise in its training set, the average of many trees is not, as long as the trees are not correlated [3].This suggest that the accuracy achieved by the decision tree should be improved with the random forest due to the bagging.

In the end we will apply SVC with kernel rbf.SVC with kernel rbf is a popular choice for classification problem. SVC uses one-vs-one multiclass reduction and by default minimises regular hinge loss instead of squared hinge loss(in linearSVC).But SVC converges slower than the linear SVC due to the use of libsvm estimator which do not penalise the intercept.

4. EXPERIMENTAL RESULTS

For evaluating the purformance of decision tree model, dataset mentioned above kaggle has been used. Model has achieved accuracy of around 77 percent. Confusion matrix has been calculated please refer to table 1 and using confusion matrix precision, recall and f1 score has been calculated. precision, recall and f1 can be calculated using below formulas.

$$precision = \frac{TP}{TP + FP}$$

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

$$F1Score = 2 * \frac{(Recall * Precision)}{(Recall + Precision)}$$

Table 1. Confusion matrix for decision tree with precision recall and f1.

	Car	Still	Train	Walking	Bus	Precision	Recall	F1
Pred Car	291	32	16	21	11	0.78	0.79	0.39
Pred Still	32	252	15	30	10	0.74	0.70	0.36
Pred Train	7	38	290	28	7	0.78	0.76	0.38
Pred Walking	26	25	47	210	20	0.64	0.72	0.33
Pred Bus	12	9	12	1	326	0.90	0.87	0.44

So accuracy of the model is 77 percent and average recall and average f1 score is 0.76 and 0.38 respectively.Accuracy of the model is not very bad, we can use random forest to improve the performance of the current model.

Table 2. Confusion matrix for Random forest with precision recall and f1.

	Car	Still	Train	Walking	Bus	Precision	Recall	F1
Pred Car	311	19	24	7	10	0.83	0.94	0.44
Pred Still	11	300	16	10	2	0.88	0.90	0.44
Pred Train	2	2	357	6	3	0.96	0.77	0.43
Pred Walking	2	10	54	261	1	0.79	0.91	0.42
Pred Bus	4	0	8	1	347	0.96	0.95	0.47

So accuracy of the model is 89 percent and average recall and average f1 is 0.894 and 0.44.

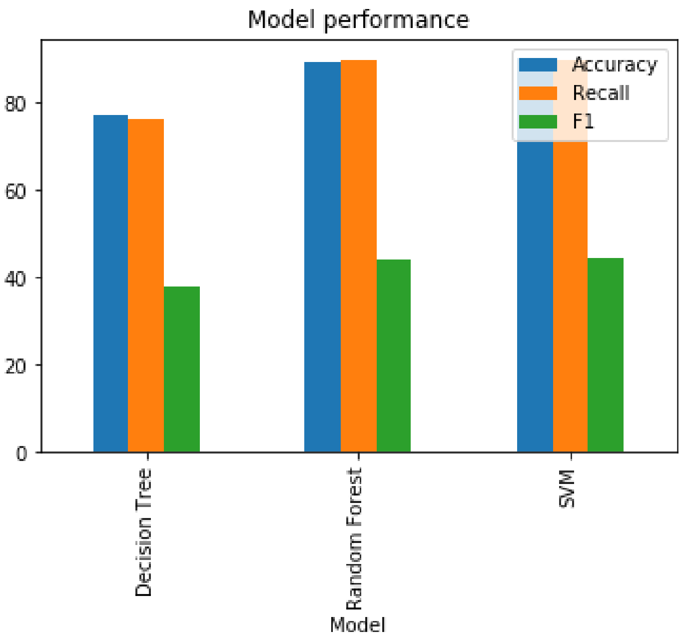
Table 3. Confusion matrix for SVM with precision recall and f1.

	Car	Still	Train	Walking	Bus	Precision	Recall	F1
Pred Car	325	13	7	16	10	0.87	0.91	0.44
Pred Still	11	307	10	6	5	0.90	0.92	0.45
Pred Train	5	4	335	18	8	0.90	0.86	0.44
Pred Walking	7	4	33	281	3	0.85	0.87	0.43
Pred Bus	7	3	1	1	348	0.96	0.93	0.47

So accuracy of the model is 90 percent and average recall and average f1 is 0.898 and 0.446. From table 2 it is evident that random forest has improved the performance of the decision tree.

5. CONCLUSIONS

To conclude which is the best model for current purpose, we have plotted a graph with model accuracy and the average values of recall and f1 score for each model in below graph.



The above graph suggests that SVM and Random Forest gives the better performance than the decision tree. But choosing between SVM and Random forest model is bit difficult as Recall and F1 score is almost same for both the models.Though the accuracy of SVM model is slightly more, but it wont impact much on the performance of the model. So in conclusion any one model can be used for the current purpose.

References

[1] Erfanul Hoque Bahadur. *Machine-Learning-Classifiers-for-Human-Activity-Recognition-with-Smartphone-sensors-data*. URL: https://github.com/ErfanulHoque/Machine-Learning-Classifiers-for-Human-Activity-Recognition-with-Smartphone-sensors-data.

[2] Jessica Ekholm and Sylvain Fabre. "Forecast: Mobile data traffic and revenue, worldwide, 2010-2015". In: *Gartner Mobile Communications Worldwide, 2011* ().

[3] wikipedia. *Random forest*. URL: https://en.wikipedia.org/wiki/Random_forest.