International Journal of Scientific & Engineering Research Volume 8, Issue 5, May-2017
ISSN 2229-5518

Mobile Location Prediction: Profound Study using Multi-class Random Forest Predictors

Mira H Gohil, Dr. S. V. Patel

Abstract— The ubiquitous connectivity of Location Based System (LBS) allows that location, situation, or event-related recommendation can be sent to targeted consumers to promote "point of purchase." Point-of-purchase promotion is important, because it can reach the potential users at the time and place where decisions are made. Apart from this many more application domain can exist whose implementation only possible using user mobile location e.g Navigation, Tracking, Geotagging, Augmented Reality, Billing, Sports, Location based Gaming, Location based Social Media, Location based News or Information, Marketing, Emergency system etc. For this type of strategy, the probabilistic information of mobile current location of user plays important role. Users next location can be materialize after protracted period of time but user proposed location at given period of time satisfies the need of the hour. Taking into consideration Anytime Prediction Model (APM) with mobility rules to extract common mobility sequence is the problem to be solve in the area of mobile location prophecy.
Many researchers has discussed this problem earlier and we also incarcerating this problem as challenge, research has considered this problem as multiclass classification machine learning (MCML) problem to get the desire accuracy in APM.

Index Terms—Mobile location prediction, Location prophecy, geolocation prediction, Next location prediction, Random Forest, Multiclass Classification Random Forest, Unsupervised Learnin, R programming

◆

1 Introduction

Mobile phones have turned into a need for individuals all through the world. The capacity to stay in contact with family, business partners, and access to email are just a couple of explanations behind the expanding impact of mobile phones. The Internet is another one of the greatest favors to man by innovation. One just can't visualize the existence without the Internet. Everybody likes/ needs to stay associated with the Internet constantly and it is just because of the cellular telephones that make it feasible. The cell telephones let clients appreciate the online networking on the go. A major section of cutting edge world is dependent on social network on internet. So, internet is one of the parameter in mobile phone revolution. The number of mobile Internet users in India is projected to double and cross the 300 million mark by 2017 from 159 million users at present, a new report by Internet and Mobile Association of India (IAMAI) and consultancy firm KPMG [1]. Your location has much to say about what you are doing. You do not do the same activities in your office as you do in your kitchen.

Knowing the location of your mobile device, and therefore your location, contributes to inferring your activity. Location-awareness is therefore a major component in context-awareness, which in turn, enables your devices to become your invisible assistants [2].There is an incredible research done upon the issue of location prediction from the innovation of the GPS and getting back consideration, because of the spread utilization of cell phones having GPS trackers.

APM is depending on, past history of mobility, current timestamp and location. Numerous algorithms were used in location prediction study but this research has focused Random Forest multiclass classification machine learning algorithm to predict the user's location.

Data for this research was provided in agreement by Nokia Research Center (NRC) by means of signed Memorandum of understanding (MoU) and agreeing to the terms and conditions. Data obtained for this research was collected by Nokia Research Center Lausanne together with its Swiss research partners (Idiap and EPFL) as discussed earlier.

The rest of the paper is organized as follows: The Section II describes literature survey details; Section III presents proposed model of this research. Section IV presents result analysis part and Section V is giving the concluding remarks and further work.

2 Literature Survey

After reviewing various literatures in the same line it has been concluded that many researchers have addressed the mobile location prediction issue and hence given different model and methodology to improve the prediction for the same. [2] [3] proposed a classification approach to predict the next place of the user. They come up with user-specific decision trees learned from each user's history. The classification tree is built based on simple, intuitive features with some mobile data-specific enhancements and [4] come up with J48 Decision Tree-based classifier (an implementation of the C4.5 algo-

- *Mira H Gohil is currently pursuing Master in phylosophy program in Computer science in Veer Narmad South Gujrat University, Gujarat, India, PH-8652857218. E-mail: miragohil260178@gmail.com*
- *Dr. S.V. Patel is retired professor in computer science in Veer Narmad South Gujarat University,, Gujarat, India, PH-9825740103. E-mail:patelsv@gmail.com*

rithm) as one of the classifier for the same. [5] used genetic algorithm for mobile current place prediction as one of the solution where they written genetic program to extract decision tree from the dataset. For applying GP and to learn decision trees the nominal data set was used. All the features are turned to nominal and used as nominal. [6] has made use of Hoeffding Trees as the base learner parameter for mobile next location prediction and compared with other classification algorithm using Hoeffding tree classifier variations and they obtained similar results. [7] discussed different approaches to attack the MDC's dedicated task 2. After experimenting with different prediction algorithms such as Decision Tree, k-NN, SVM and DBN, they present a soft classifier fusion technique used for predicting user's future location. Instead of using one complex model for capturing user's mobility patterns, they use multiple models and each one of them is focusing only on a certain aspect of location prediction such as time-location dependency or likelihood of location transitions. [8] used more sophisticated method which is require to redesign the predictive model with integration of multiple types of observation, such as the spatio-temporal decision tree. They used statistical prediction method called random forest combined with other models collectively to get the optimum results. They used Lausanne Data Collection Campaign (LDCC) dataset as input for the model and get mobility prediction accuracy in random forest 0.635. [9] further extend their study combining all individual features in two supervised learning models named linear regression and M5 model trees to get higher overall prediction accuracy. They found that the supervised methodology based on the combination of multiple features offers the highest levels of prediction accuracy and M5 model trees are able to rank in the top fifty venues one in two user check-ins, amongst thousands of candidate items in the prediction list.

In addition to this, the papers which are published in Nokia Mobile Data Challenge Workshop based on Nokia MDC dataset also referred. These papers are majorly focused on semantic place prediction, location prediction, human behavior and interactions prediction etc [11].

It has been seen from above survey that, more information can be exploited from the MDC spatial temporal database. In addition to this, there is more stress on next location prediction rather than anytime prediction. Anytime mobile location prediction is important because companies are beginning to expertise offers based on where a customer is at any given point of time. It has been also observed that common sequence of the user mobility not addressed in any of the above classification research problems which is the subject needs to address. Therefore, constructing "Anytime Prediction Model" (APM) incorporation with extraction of common mobility rule has been a topic of interest in the area of mobile location prophecy. According to the nature of the problem it falls under the category of classification problem.

3 Proposed Model

To solve the given problem, we used the following steps to achieve highest accuracy.

- Data Collection and preprocessing
- Data Cleaning
- Proposed Framework

3.1 Data Collection

Real time mobile data collection part is very difficult because of various reasons like:

- Users don't want to reveal their current location easily. Data privacy is the biggest issue.
- Internet connectivity is required 24 hours
- If users want to use internet facility 24 hours then mobile battery and internet connectivity are the biggest problems in India.
- Even though we tried to collect the location data of the 10 students of TIMSCDR College, Mumbai using mobile application. The application developed in Apache Cordova (formerly PhoneGap 1.4.1), IDE Eclipse 4.3.2 and Android SDK 19 API platform but the data is showing highly regular pattern. In this case location can be predicted without any algorithm implementation. so research has used MDC dataset as explain earlier.

In data preprocessing, data was available in the form of timestamp. A variety of information is extracted from this timestamp as well as new field are included like working/week day field.

3.2 Data Cleaning

Some of the records are marked as "trusted start visit", "trusted transition" and "trusted end visit," perhaps because of the user's mobile lost the signal for some period of time or because of Handoff mechanism or due to impairments. Eliminating all such trusted visits provides more assured visit information, but also rejects a considerable amount of potentially valuable information. Thus, at some stage in data preprocessing, we eliminate such visits only if there is a huge gap with their successive visits.

Other various data cleaning done as follow:

1. Used only trusted and last one year data.
2. At a time used only one user data to understand the moving pattern
3. Remove duplicate records
4. User data less then threshold value eliminated from the research
5. Throughout this research first top 10 users data has been calculated and these uses are as given in the table 1.

TABLE 1
Top 10 User (Total 129097 rows in the table)

ID	USERID	Total Record

International Journal of Scientific & Engineering Research, Volume 8, Issue 1, January-2017
ISSN 2229-5518

1	5947	1993
2	5964	1877
3	6177	1872
4	5927	1745
5	5967	1700
6	5953	1674
7	5973	1674
8	5928	1652
9	5966	1645
10	5925	1631

3.2 Proposed Framework

Research has taken the APM problem as supervised learning problem and performed following task to solve the problem and to get the model correct:

- **Feature Selection** will identify the important feature from the dataset.
- **Parameter Tuning** will tune the model parameter to get the highest accuracy.
- **Model implementation** will implement the model with the tuned parameter to get the highest accuracy
- **Cross validation** will validate the performance of the model
- **Prediction experiments** will use to predict the new data from the tuned implemented model
- **Prediction results** will give the outcome of the performed experiment
- **Prediction accuracy** will give the accuracy of the prediction for the selected model

3.2.1 Feature Selection

This research is using random forest multiclass classification algorithm where Random Forests grows many classification trees. To classify a new object from an input vector, put the input vector down each of the trees in the forest. Each tree gives a classification, and we say the tree "votes" for that class. The forest chooses the classification having the most votes (over all the trees in the forest).

To extract the best performing features from each feature group, we applied feature selection algorithm called MeanDecreaseAccuracy and MeanDecreaseGini in R. This algorithm evaluates features with respect to their individual predictive ability along with the degree of redundancy between the features.

In figure 1 the first graph shows us that how much our accuracy decrease, should we remove each of the non performing predictors (while others in the tree) and it is called MeanDecreaseAccuracy. The second plot shows us the same thing for Gini index (node purity) and it is called MeanDecreaseGini.

Finally out of 17 independent variables, 11 independent variables are selected for the model to implement the model correctly

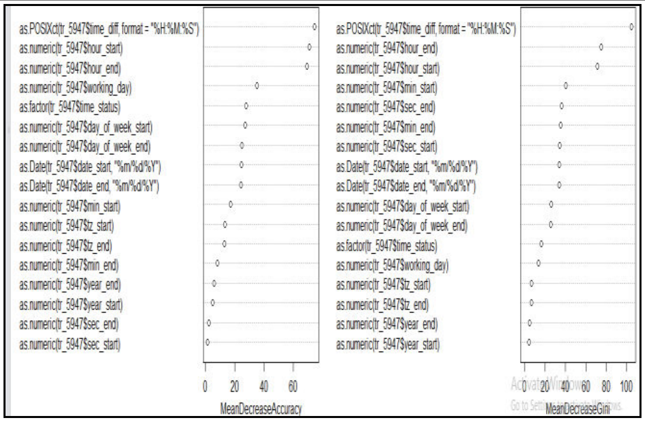

Fig 1. Variable importance indexes

3.2.2 Parameter Tuning

The important parameters for the model are (1) no of feature selected at the time of tree generation (mtry) (2) no of trees generate (ntree). These parameters required tuning to get higher accuracy in the model.

mtry: mtry parameter tuning done using the rf method of the random forest in R programming which gives value of mtry is 5 as output as shown in figure 2.

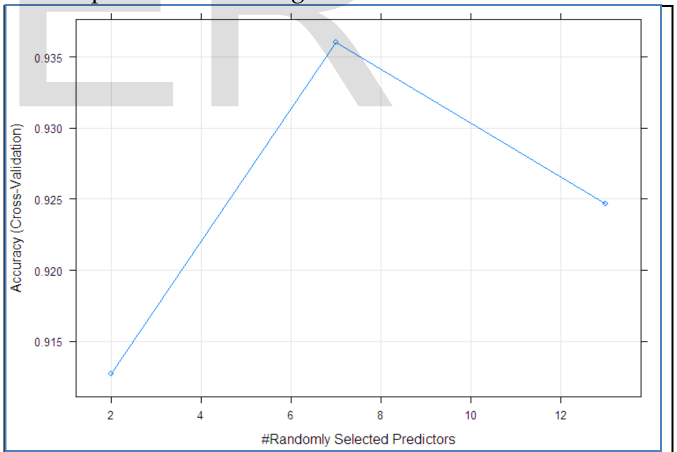

Fig 2. Value of mtry with respect to accuracy of the model

ntree: how many node to be used to get the accurate result is decided by the parameter ntree and to tune this parameter, we require to do trial and error method for the different values of the ntree and find out the minimum OOB error as shown in figure 3 and get the value of ntree. We have used algorithm in R programming to get the optimum results in the model. The result got optimum values of the parameters which are 5 and 500 for mtry and ntree respectively.

International Journal of Scientific & Engineering Research, Volume 8, Issue 1, January-2017
ISSN 2229-5518

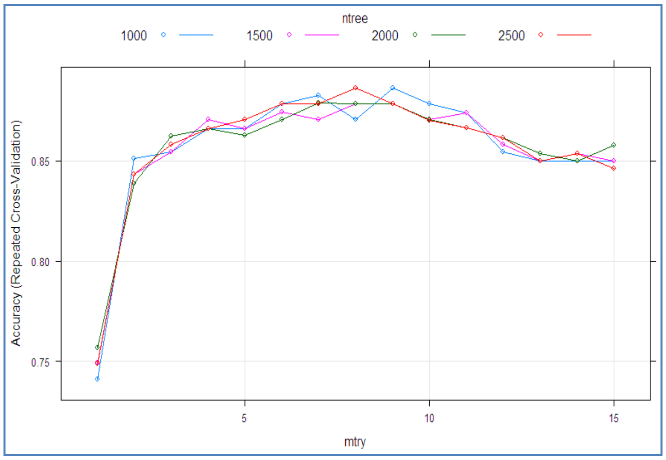

Fig 3. mtry/ntree with respect to accuracy

3.2.3 Model implementation

Using the tuned parameter for the model i.e. mtry = 5 and ntree = 500, we implemented the model and checked for the accuracy using OOB error value for user 5947. We used RandomForest package in R programming for the model. When we use tuned parameter than we get even less OOB estimation of error in our case it is approximately 13.30%.

3.2.4 Cross Validation

For cross validation, we perform confusion matrix interpretation and also used other cross validation method like cforest, rpart, rf method of R programming and validate the model.

3.2.4.1 Confusion matrix interpretation:

The confusion matrix itself is relatively simple to understand, but the related terminology can be confusing. Here we have explained the confusion matrix for the user 5947 with random forest parameter mtry=5 and ntree=500.

In multiclass classification, the accuracy rate for estimating the quality of the classifier. Let (rf) be a classifier and dataset with instances ai together with N classes where every instance ai belongs exactly to one class bi and ai and bi be respectively an example in the data base and its class.

The accuracy (AC) is the proportion of the total number of predictions that were correct. It is determined using the equation:

$$\mathrm{accuracy}(rf) = \frac{1}{n}\sum_{i\leq n}^{1} rf(ai) = (bi)$$

In our case, as data given in the confusion matrix,
AC=(TP)/TotalPrediction
=176/203 = 0.866995 = 86.69%

In multiclass classification, the empirical error rate for estimating the quality of the classifier. Let (rf) be a classifier and dataset with instances ai together with N classes where every instance ai belongs exactly to one class bi and ai and bi be respectively an example in the data base and its class.

$$\mathrm{err}(rf) = \frac{1}{n}\sum_{i\leq n}^{1} rf(ai) \neq (bi)$$

In our case based on the above confusion matrix,
= (27)/203 = 0.1330 = 13.30%
Similarly we calculated for all top ten users as shown in table2.

TABLE 2
Top Ten users Accuracy according to confusion matrix

Sr.no	User	OOB estimated error rate	Accuracy
1	5947	13.3%	86.70%
2	5964	11.73%	88.27%
3	6177	24.29%	75.71%
4	5927	15.23%	84.77%
5	5967	11.81%	88.19%
6	5953	6.9%	93.10%
7	5973	9.61%	90.39%
8	5928	27.27%	72.73%
9	5966	15.11%	84.89%
10	5925	9.4%	90.60%
		14.465%	**85.53%**

3.2.4.1 10 fold Cross validation

It is not always feasible to gather an independent set of observations for testing the models' performance, because assembling statistics is typically an expensive activity.
A feasible way out is to use **cross-validation** (CV). In its basic version, the so called kk-fold cross-validation, the samples are randomly partitioned into kk sets (called folds) of roughly equal size

TABLE 3
Average Accuracy for Top 10 users using 10 fold validation

Method Used	ACCURACY	KAPPA VALUE
Method ="rf"	**0.89508**	**0.8601**
Mehod="cfprest"	**0.6181**	**0.4994**
Method="rpart"	**0.52344**	**0.3831**

So, from above table results of 10 fold cross validation methods shows that Random forest algorithm is working good among all.

3.2.5 Predictions

Model is ready for the prediction. So we can apply it to solve the problem. To solve the problem, we divide the problem in two phases and solve it one than the other. The two phases of the problem are:

- Prediction of the user location based on time
- Prediction of the user location based on the daily routine

International Journal of Scientific & Engineering Research, Volume 8, Issue 1, January-2017
ISSN 2229-5518

3.2.5.1 Prediction of the user location based on the given time:

Following steps are performed to prediction experiment of APM:

- For the prediction purpose, we divide the data into training and testing sets. So, total data of user 5947 divided into 70% training and 30% testing set data (Appendix (vi-1)).

- So, we have two datasets tr_5957 as training and te_5947 as testing set. Implement the random forest on training set first and get the model variable train_rf_5947 (Appendix (vi-2)).

- Prediction performs on the testing set but before testing we need to convert the table column data type as per the data type of the Model train_rf_5947. The R commands are used for data type conversion as per the model (Appendix (vi-3)).

- The table te_5947_mody is the dataset with the same datatype used in the random forest training model. We need to delete the placeid_char column because this column will be predicted by the model.

- To predict the testing data based on trained model, we use the predict algorithm of the Random forest and get the results.

There are two types of prediction done in the Random forest which are 1) probabilistic and 2) response type. In probabilistic, the prediction done in the form of probability and in the response, prediction gives the direct value of the dependent variable as prediction.

As per the prediction result where type=”prob”, model will calculate the probability of the entire placeid for the given record as shown below.

>te_5947_mody_pre<-predict (train_rf_5947, data=te_5947_mody, type="prob")
> te_5947_mody_pre[5,]

P1 P10 P105 P11 P111 P12 P14 P15 P2 P27 P41 P5 P53 P7 P98

75730.072 0.000 0.000 0.060 0.064 0.000 0.000 0.074 0.024 0.000 0.054 **0.632** 0.000 0.000 0.020

7628 **0.980** 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.004 0.000 0.000 0.008 0.000 0.000 0.008

So, predictions are in the form of probability where out of 15 classes the maximum probability location class is P5 with probability 0.632 for first data input. 0.980 probabilities for place id P1 for second record and so on.

Response Algorithm also shows the direct response of the prediction using R programming command shown as follows:

>te_5947_mody_pre_res<-predict(train_rf_5947, data=te_5947_mody, type="response")
> te_5947_mody_pre_res[2,]

7573 7628
P5 P1

From the above R programming output direct response value received i.e. P5 place is predicted for the first record by the model and P1 for second record.

Prediction Accuracy:

The ROC curve is a fundamental tool for diagnostic test evaluation. In a ROC curve the true positive rate (Sensitivity) is plotted in function of the false positive rate (100-Specificity) for different cut-off points of a parameter.

For the performance checking, we can calculate the accuracy of this model using Multiclass.roc functions in R programming using **RORC package** of R library which gives multiclass area under the curve result 0.8904

So for the testing set values the prediction accuracy is 89.04% in the random forest for user 5947 figure 4 shows computed and plotted the roc object.

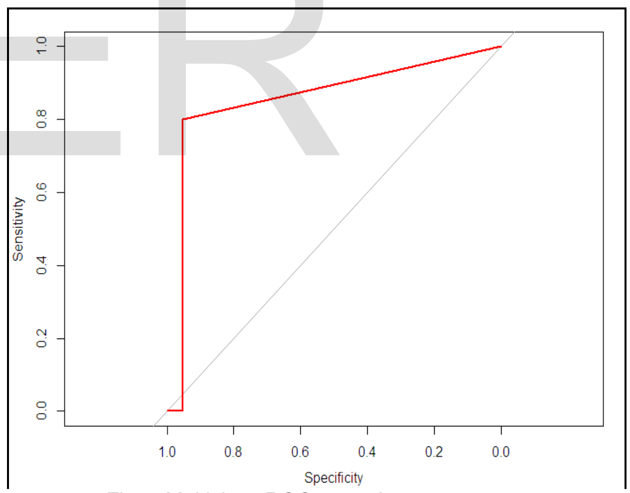

Fig 4. Multiclass ROC curve for user 5947

3.2.5.2 Prediction of the user location based on daily routine:

To solve the second phase of the problem i.e. prediction of the user location can be done using daily mobility routine. To analyze the pattern, we have two methods to perform:

1. Rules extraction using pattern analysis
2. Pattern extraction using sequence analysis

➢ **Rule extraction using pattern analysis**

Prediction Experiment:

For user specific generalized rules extraction, research uses the methods available in package **inTree**. The inTree algorithm gave the most significant common generalized rules of the user based on the user common behavior pattern in the mobility as shown in the following table. Initially algorithm extracted 2553 rules with 6 condition but to extract the prominent rules research used the inTree method **getFrePattern** and got the 19 prominent rules for the user mobility.

TABLE 4
Sample output of inTree package (getFreqPattern Method)

	Len	p	Conf	Condition	Pred
[1,]	"1"	"0.076"	"0.647"	"hour_start>10"	"P1"
[2,]	"1"	"0.042"	"0.733"	"time_status %in% c('A','E','N')"	"P1"
[3,]	"1"	"0.042"	"0.545"	"hour_start>15.5"	"P1"
[4,]	"1"	"0.032"	"0.737"	"hour_end<=18.5"	"P1"
[5,]	"1"	"0.023"	"0.577"	"hour_end<=15.5"	"P1"
[6,]	"2"	"0.015"	"0.514"	"working_day<=0.5 & hour_start<=10"	"P5"
[7,]	"2"	"0.013"	"0.654"	"working_day>0.5 & hour_start<=10"	"P2"
[8,]	"2"	"0.012"	"0.842"	"hour_start>15.5 & hour_end<=18.5"	"P1"
[9,]	"1"	"0.011"	"0.583"	"hour_end<=19.5"	"P1"

Prediction Result and Accuracy:
In the above table 9, we can see the output of the common rules applied for the user 5947 mobility prediction. So if we want to predict the location between the hours 15 to 18 than it will be always place p1 that is the generalized rule for user 5947 and for this prediction the confidence level is **0.842** is also calculated from the method getFreqPattern.

➢ Pattern extraction using sequence analysis

Prediction experiment:
TraMineR package of R is free and open source library which is having different function based on the type of database format to render state sequences. Using this package's function, we can easily get the common sequence pattern of the user mobility [23].

Sequence data
To generate the day wise sequence from the data, TraMineR package used which is giving sequence for the data:
Current MDC research data is in the form of SPELL, so it needs to convert into the STS for sequence generation and analysis. TraMineR command sqeformat used to convert SPELL in STS format and seqdef used to retrieve the daily mobility sequence of the user as shown in the following output.
>sts.seq[5,]

101 *-*-*-*-*-*-*-*-P2-P2-P2-P2-P2-P2-P2-P2-P1-P1-P1-P1-P1-P1-P1-P1
102 P1-P1-P1-P1-P1-P1-P1-P1-P1-P1-P1-P2-P2-P2-P2-P5-P5-P5-P5-P5-P1-P1-P1-P1
103 P1-P1-P1-P1-P1-P1-P1-P1-P2-P2-P2-P2-P2-P2-P2-P2
104 *-*-*-*-*-*-*-*-P2-P2-P2-P2-P2-P2-P2-P1-*-*-*-P1-P1
105 *-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-P1-P1-P1-P1-P1-P1-P1

Visualize the first 10 days data which are in the form of STS using the graph as shown in the following figure.

>seqiplot(sts.seq, title = "Index plot (first 10 sequences)", cex.legend=0.7)

First 10 sequence of the data displays place id with respect to hour start (1 to 24). As shown in the figure 5, the second day user location between 1 to 11:30 hour is P1, between 11:30 to 16 location is P2, between 16 to 20 use spent his time at location P5 and than last visited location is again p1. This type of observation is quite logical where the first and last visited location is home location of the user. As per the record in the file 5947, the 10th day is Sunday and that is why most of the location is showing home location P1 in 10th day pattern in the graph.

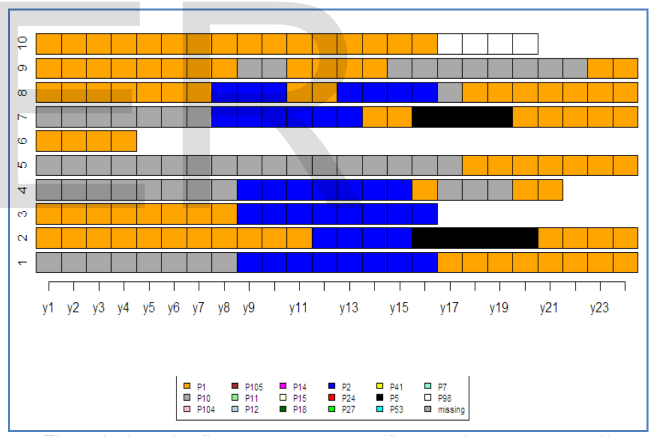

Fig 5. Index plot first 10 sequences (first 10 days sequence))

State distribution chart
All the place id distribution in 1 to 24 hour slot plotted in the following figure 6. Again the logical observation from the distribution chart is that the p1 place is home location and p2 is working location for the user.

>seqdplot(sts.seq, title = "State distribution plot", cex.legend=0.7)

International Journal of Scientific & Engineering Research, Volume 8, Issue 1, January-2017
ISSN 2229-5518

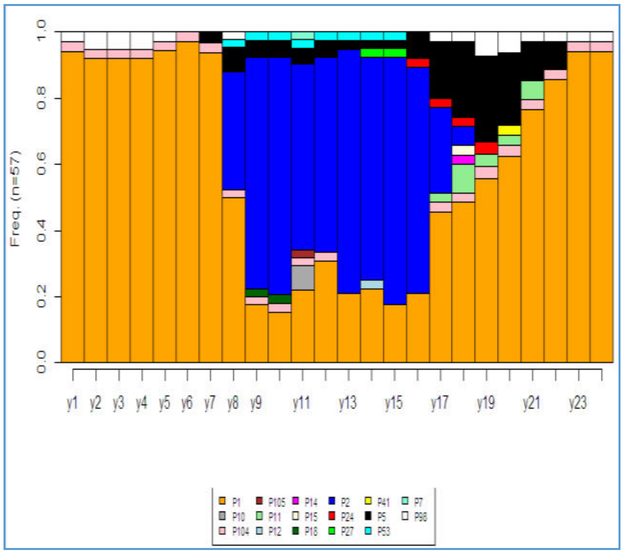

Fig 6. State distribution plot

It is necessary to convert data in the compressed form of SPS format for analysis. Following command will be used to get data in compressed SPS format.

```
sps.data <- seqformat(user_5947_seq,var=c("id","hour_start","hour_end","placeid_char"),from="SPELL",to="SPS",compressed = TRUE,process=FALSE)
 [>] time axis: 1 -> 24
 [>] SPELL data converted into 57 STS sequences
 [>] preparing 57 sequences
 [>] coding void elements with '%' and missing values with '*'
 [>] STS sequences converted to 57 SPS seq./rows
```

❖ **Prediction Results**

```
> head(sps.data)
   Sequence
[1] "(*,8)-(P2,8)-(P1,8)"
[2] "(P1,11)-(P2,4)-(P5,5)-(P1,4)"
[3] "(P1,8)-(P2,8)"
[4] "(*,8)-(P2,7)-(P1,1)-(*,3)-(P1,2)"
[5] "(*,17)-(P1,7)"
[6] "(P1,4)"
```

Result retrieve the most common sequences from the pattern and they are P2 to P1 and P1 to P2 because as per the MDC dataset the P1 is the home location and P2 is the work location of the user. Other patterns for user 5947 are P2 to P5, P5 to P1, P1-P2-P5-P1 and P1 to P1.

Pattern P1 to P1 says that the maximum time users spent on location p1. There are multiple consecutive records are available in the dataset and as per dataset details, P1 is the home location of the user.

4. Result Analysis

- From the above methodology and finding, we can say that all users are having their own unique mobility pattern. As we have seen in the Table 3, where initial random forest algorithm implemented in R with all user data, there we have reported maximum error rate 66.32%. When we changed our model and used user specific data then significantly error rate decreased and accounted 38.09%. So, we used this model as user specific model for other findings.
- We have used the feature selection algorithms MeanDecreaseAccuracy and MeanDecreaseGini to get the important feature for the model and avoid the unimportant feature from the model.
- Also we tuned and trained our model with parameter tuning ntree and mtry and get the best values for the ntree and mtry for the model.
- To check accuracy of the model we have used 10 fold cross validation method in different packages of R and reported accuracy accordingly but the best accuracy got in the random forest package by using method rf reported average of top ten users 89.51% while other methods like cforest reported 61.81% and rpart reported 52.34% for the same data.
- In the prediction phase, we got the accuracy or area under the curve for multiclass problem about 89.04% for the user place prediction at a given time.
- We also used the inTree package of R studio to get the best common general rules of the user which we can easily predict without any model execution.
- In R programming, the package called TranMineR is responsible to get the sequence data or common path way of the user in his/her daily life through which we can predict the user approximate next location without any model execution

Comparing with other best results

- However, we compared our best result with the, following table shows the best prediction accuracy of result of other methods used in the classification tree based approach:

TABLE 4
Comparition with other best results

Srno	Approach	Users	Data Set	Accuracy
[2]	Classification with several enhancements. Accuracy with regular pattern data	User specific model	MDC	70%
[6]	Classification algo-	User	MDC	59.6%

	rithm in J48 Weka tool where private testing set used for validation	specific model		
[11]	Multiclass classification approach Random forest	User specific approach	MDC	63.5%
[12]	Data mining	User specific Approach	Personal database	80%
	Multiclass classification random forest algorithm (APM)	Users specific approach	MDC-(top 10 users data)	85.5%

- And finally we conclude that using Random Forest algorithm in R programming we can get all the details about user mobility easily with highest accuracy.

5. Concluding Remark

According to the historical information of mobile nodes trajectories, the state transition matrix is constructed by the location as the transition state and Anytime Prediction mode (APM) is used to predict the mobile node location with the certain duration. We present how random forest machine learning is applied to given data for predicting future location. The model is evaluated with the 10-fold cross validation method. Our findings show that, the voting classification scheme is appropriate for location prediction since it exhibits satisfactory prediction results for diverse user mobility behavior. Moreover, we compare the performance of the proposed APM model with the others and it is justifying the importance of the machine learning classification in prediction.

In future, we can analyze the correlation between the different users and research can be extended as social networking analysis. Meanwhile, social relationship between nodes is exploited for optimization and amendment of the prediction model.

Acknowledgment

This research is partial support by Nokia Research (Mobile Data Challenge 2012). I would like to extend my sincere gratitude to my guide Professor S. V. Patel for his valuable instructions and suggestions in my research. Without his patient help and illuminating instruction, it would be impossible for me to complete this research on time.

References

[1] http://www.iamai.in/research/reports_details/3647

[2] Le-Hung Tran, Michele Catasta, Luke K. McDowell, Karl Aberer. 2012. Next Place Prediction using Mobile Data. In Mobile Data Challenge 2012 (by Nokia), Newcastle

[3] Anna Monreale , Fabio Pinelli , Roberto Trasarti , Fosca Giannotti, WhereNext: a location predictor on trajectory pattern mining, Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, June 28-July 01, 2009, Paris, France [doi>10.1145/1557019.1557091]

[4] T. Anagnostopoulos, C. Anagnostopoulos, S. Hadjiefthyrniades, M. Kyriakakos and A. Kalousis, "Predicting the Location of Mobile Users: A Machine Learning Approach", Proc. ACM International Conference on Pervasive Services (ICPS)

[5] A. Antonic, I. P. Zarko, and D. Jakobovic, "Inferring presence status on smartphones: the big data perspective," inProceedings of the 18th IEEE Symposium on Computers and Communications (ISCC '13), pp. 600-605, July 2013.

[6] J.B. Gomes, C. Phua, S. Krishnaswamy,"Where will you go? Mobile data mining for next place prediction", Data Warehousing and Knowledge Discovery, Springer, Prague (2013), pp. 146–158

[7] Le T. Nguyen, Heng-Tze Cheng, Pang Wu, Senaka Buthpitiya, and Ying Zhang, "PnLUM: System for Prediction of Next Location for Users with Mobility, " In Proceedings of Mobile Data Challenge by Nokia Workshop, 2012

[8] A. Monreale, F. Pinelli, R. Trasarti, F. Giannotti, Wherenext: a location predictor on trajectory pattern mining, in: Proc. KDD, 2009.

[9] T.M.T. Do, D. Gatica-Pereza, " Where and what: Using smartphones to predict next locations and applications in daily life Pervasive and Mobile Computing (2013)

[10] GABADINHO, Alexis, et al. Analyzing and Visualizing State Sequences in R with TraMineR. Journal of Statistical Software, 2011, vol. 40, no. 4, p. 1-37

[11] T.M.T. Do, D. Gatica-Pereza, " Where and what: Using smartphones to predict next locations and applications in daily life Pervasive and Mobile Computing (2013)

[12] M. Morzy, "Mining frequent trajectories of moving objects for location prediction," in Proc. of MLDM, 2007, pp. 667-680.