International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 11 (2015) pp. 30167-30175

http://www.ripublication.com

Analysis of Students Behavior Using Moving Objects by Rpart and Rattle - R Software

Pallavi M S and Mr Vinayak Hegde
Dept. of Computer Science
Amrita vishwa vidyapeetham, Mysuru Campus
Mysuru, Karnataka, India
palls.ms@gmail.com, Mobile No - 8884559752

Abstract

Moving objects can be defined as an object that changes its geographical position as time changes; these are called spatio-temporal data. They are represented as set of points, located in space and time. Moving objects data of students can be used to analyze the student's behavior like regularity of visit to college, places they visit in a period of time, so that if they get deviated from their regular college visit, then corrective measures can be taken, to counsel by the teacher and concern authorities of college and parents. This paper focuses on analysis of student's behavior by taking data considering from morning 7am to evening 8pm. The methodologies used here is recursive portioning and regressive tree as classification tree. The result shows graphical representation of the maximum place visited and also shows new placed visited by the student.

Index Terms: Rattle, rpart , classification tree, moving objects, spatio-temporal data, R Software.

Introduction

Now a day's lot of effort has been devoted to the development of efficient software for mobile devices i.e. palm tops, GPS , cell phone, sensor network etc .Very little has been done to analyses the data generated by mobiles. Every work should be done in very little time about getting the nearest petrol bunks, restaurants, nearest ATM center, navigation, etc.[2]

In this paper, Student's data on time bases is taken and represented in terms of longitude and latitude of the particular place. i.e. the areas in which they concentrate more and their areas of spending time .These data will be pre-processed and consistent data will be taken. Missing points will be detected and filled in pre-processing step. These data will be segmented into different groups with the help of

algorithm and detecting the points. After the segmentation, data will be passed to the rpart package and then final graph will be presented based on the classification of data. This graph will help the user (student/parents/teachers) in what area they are lacking in, and effective measures can be taken to overcome the drawback.

The optimal representation of student moving object is to provide graphical representation of the required data which in turn help the users to overcome the drawback based on the graph and also to improve the area of concentration by the students. It also provides awareness to the students, parents as well as teachers. It will also help them to know if the student is regular to college or if student is spending time more outside and effective measures can be taken based on the result provided. The method is helpful for the head of the institution, teachers, parents or any other user from the college who would like to know whether the student is regular or irregular.

The rest of the paper is organized in five sections. Section II – Related Work gives information to related works carried out by various people. Section III – Methodologies used are rpart and rattle software package of R software. Section IV – Experimental results and V gives the conclusion.

Related Work

"Mining Periodic behavior for Moving Objects" by Zhenheri Li and Jiawei Han is a work carried out to find animal movement daily or yearly periodic behaviors. The author captures reference locations. Fourier transformation and autocorrelation is used for retrieving periodic movements. A probabilistic model is proposed to characterize periodic behavior and the through hierarchical clustering these movements are clustered.[1] In Tutorial on spatial and spatio-temporal data mining – Introduction to Moving Data and Moving object Databases by Vania Bogorny. [2] In "Querying Moving Objects in SECONDO" the author uses DBMS environment for moving objects using software SECONDO. Here the author explain the demo of moving objects by a prototype implementing a data model and query language also explain about algebra of moving objects.[3] The author has explained the basis of moving object and its representation. Also author explains how these moving objects can be used. Author also explains SECONDO and HERMES software for using in moving objects. Data Mining II(Mobility Data Mining) by Micro Nanni, ISTI-CNR author explains in details about moving object clusters.[4][8]. In Continuous Clustering of Moving Objects by Christian explains how moving data can be clustered.[5]. In Clustering Moving Objects by Jiawei Han and jiong Yang explains about micro clustering and dynamically maintained moving object.[6]

Methodologies Used

Initially moving objects data will be stored in moving objects databases. Representation of moving objects in the database[7] .i.e. data will be collected on the basis of time spent by the student for every half an hour in terms of x, y & t. Longitude, latitude, location and time will be taken for every half an hour and these

data will be stored in database. For every week in a month or on monthly basis theses data will be taken for pre-processing. Pre-processing is done in order to remove the noise that is suppose there is a time gap while taking up the data we can find the missing points and fill it with appropriate method.

After pre-processing the data, these data will segmented into different groups based on the data provided and we can even find relationship between the data provided. Segmentation will be done based on the relationship of moving objects and similar groups. Then the data is classified using rpart and rattle software package. Here a Classification tree i.e. feature tree will be constructed based on the data. A graph is generated based on the Classification Tree. Based on the graph, summary will be given. This graph will help the user (student/parents/teachers) to identify in what area they are lacking in, and effective measures can be taken to overcome the drawback. This graph will also let us to know if the student is regular to class and active student.

1. Rattle: Graphical User Interface For Data Mining In R

Rattle (the R Analytic Tool To Learn Easily) provides a Gnome (RGtk2) based interface to R functionality for data mining. The aim is to provide a simple and intuitive interface that allows a user to quickly load data from a CSV file (or via ODBC), transform and explore the data, build and evaluate models, and export models as PMML (predictive modelling markup language) or as scores. All of this with knowing little about R[9,10]

2. Rpart: Recursive Partitioning And Regression Trees

Recursive partitioning for classification, regression and survival trees. An implementation of most of the functionality of the 1984 book by Breiman, Friedman, Olshen and Stone.[10]

3. Rpart. Plot: Plot Rpart Models. An Enhanced Version of Plot. Rpart

Plot rpart models. Extends plot.rpart and text.rpart in the rpart package.[10]

Experimental Results

The data that has been considered is a single student day wise longitude and latitude data that is from morning 7.00 am to 7.00pm. For every half an hour the longitude and latitude will be taken. Here is a sample data given in the Figure 1.

	log	lat	Time	Location
1	1	1	7-00 AM	Home
2	1	1	7-30 AM	Home
3	1	1	8-00 AM	Home
4	2	2	8-30 AM	Road1
5	3	3	9-00 AM	College
6	3	3	9-30 AM	College
7	3	3	10-00 AM	College
8	3	3	10-30 AM	College
9	3	3	11-00 AM	College
10	4	4	11-30 AM	Library
11	3	3	12-00 PM	College
12	3	3	12-30 PM	College
13	3	3	1-00 PM	College
14	5	5	1-30 PM	Road2
15	5	5	2-00 PM	Road2
16	6	6	2-30 PM	Road3
17	7	7	3-00 PM	Road4
18	8	8	3-30 PM	Road5
19	9	9	4-00 PM	Road6
20	1	1	4-30 PM	Home
21	1	1	5-00 PM	Home
22	1	1	5-30 PM	Market
23	1	1	6-00 PM	Market
24	1	1	6-30 PM	Home
25	1	1	7-00 PM	Home

Figure 1: Day-wise Student Data

Figure 1 show the data set which we have considered for analysis. The data is in CSV format, with the name Day1.csv. This data set has longitude, latitude and time values. This data is further compared with the Data Base and by identifying the longitude and latitude value corresponding Location is replaced.

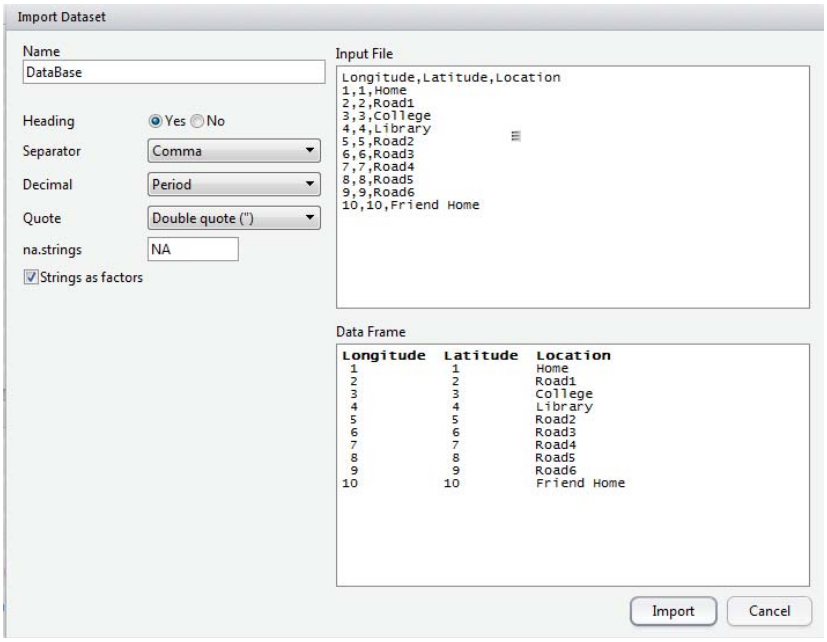

Figure 2: Show The Database Which Has Been Imported Into R Studio

A training database is created which stores all the places where the student may visit. If a new Longitude and Latitude is found then it is recorded in the Database as Location "New1". It helps the analyzer to identify, whether the student has visited new Location in the Day/ Week/ Month.

	Longitude	Latitude	Location
1	1	1	Home
2	2	2	Road1
3	3	3	College
4	4	4	Library
5	5	5	Road2
6	6	6	Road3
7	7	7	Road4
8	8	8	Road5
9	9	9	Road6
10	10	10	Friend Home

Figure 3: A Database which shows the Location for corresponding Longitude and Latitude

//Code for Generating the Tree.

```
>install.packages('rattle')
>install.packages('rpart.plot')
>library(rattle)
```

```
>library(rpart.plot)
>frmla = Location ~.
>Day1$Location = as.factor(Day1$Location)
>library(rpart)
>fit = rpart(Location ~., data=Day1, method="class", control=rpart.control(minsplit=8, cp=0))
>fancyRpartPlot(fit)
```

Figure – 4 shows the classification tree. It summarizes and shows that the student has spent his maximum time in the College and then In Home. He has not visited any new place in this day.

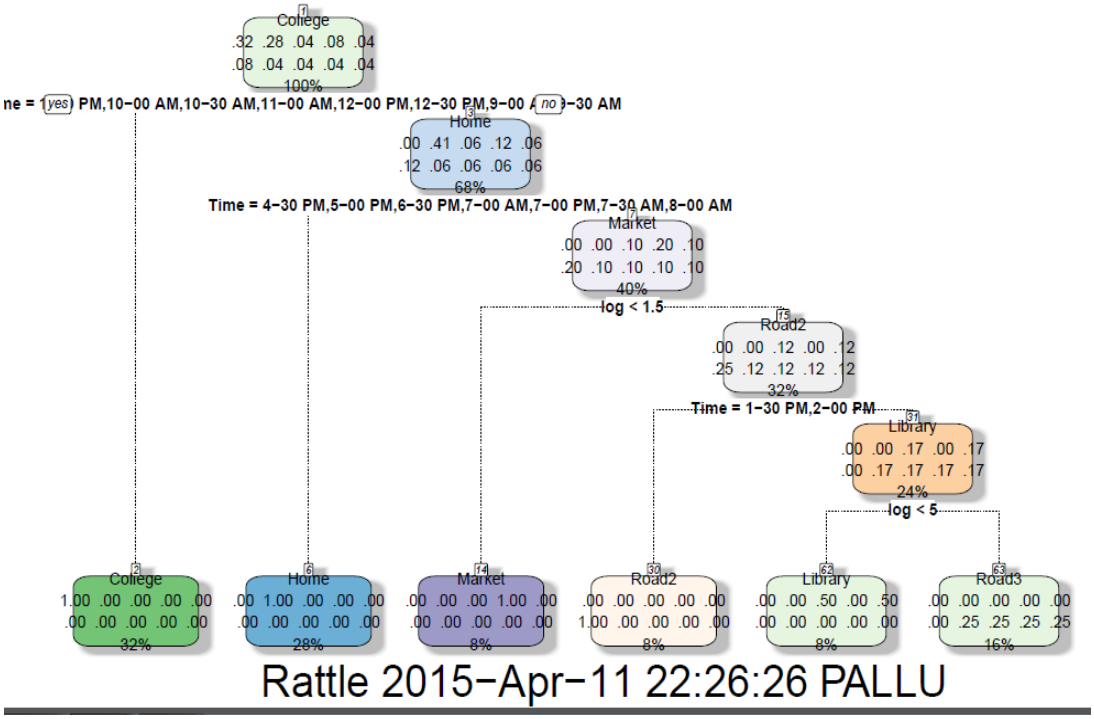

Figure 4: Classification Tree

A Day-wise Classification tree can be generated and data regarding each Day is recorded.

	Days	MaxVisited	NextMaxPlaceVisted	NewPlaceVisted	NewPlaceVisted.1
1	Day1	College	Home	NA	NA
2	Day2	College	Home	New1	NA
3	Day3	Home	Friend Home	New2	New3
4	Day4	College	Home	NA	NA
5	Day5	Home	College	New2	NA

Figure 5: Summary of five Days Moving Data

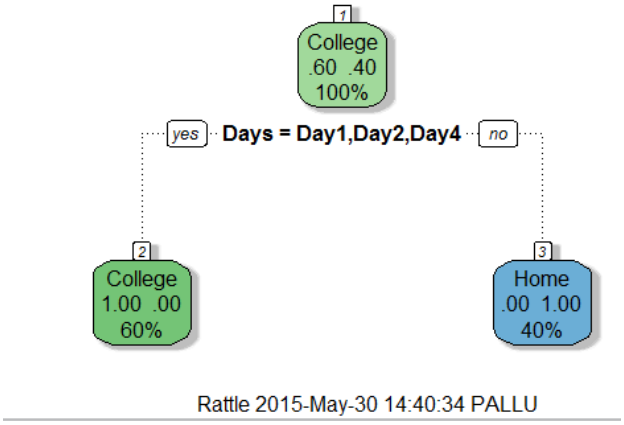

Figure 6: Summary of Classification Tree Day Wise

Figure – 6 shows that he has spend more time in College and then in Home in five days.

Every Week-wise Classification Tree is generated and summary of Maximum time spent and new places visited are recorded.

	Days	MaxVisited	NextMaxPlaceVisted	NewPlaceVisted	NewPlaceVisted.1	NewPlaceVisted.2
1	Week1	College	Home	New1	New2	New3
2	Week2	Home	College	New3	New4	New5
3	Week3	Home	Friend Home	New1	NA	NA
4	Week4	Home	College	New1	New4	New2

Figure 7: Summary four Week Moving data

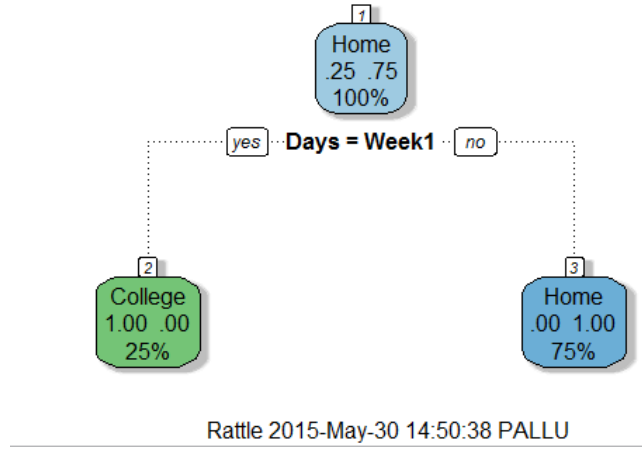

Figure 8: Summary of Classification Tree Week Wise

Figure – 8 shows that he has spend more time in Home and then in College in four weeks.

Summarizing the week, a Month-wise data is generated.

```
//Code for Generating the Tree where the student has visited new Places.
>library(rattle)
>library(rpart.plot)
>frmla = NewPlaceVisted~  NewPlaceVisted.1~ NewPlaceVisted.2~.
>Week.wise$NewPlaceVisted = as.factor(Week.wise$NewPlaceVisted)
>library(rpart)
>fit = rpart(NewPlaceVisted ~., data=Week.wise, method="class", control=rpart.control(minsplit=4, cp=0))
>fancyRpartPlot(fit)
```

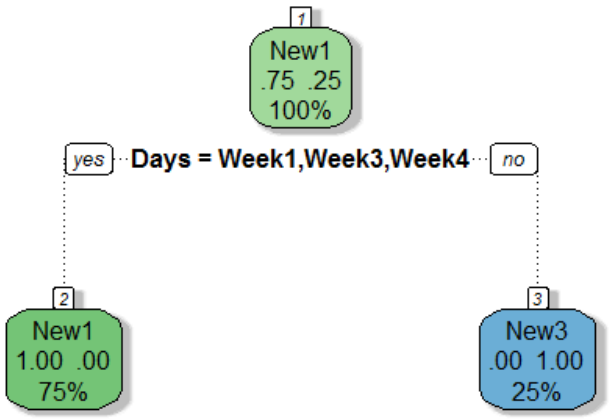

Figure 9: Summary of Classification Tree for New Places Visited By the Student.

The Figure -9 shows that the Student has visited New1 place for more time in a month and by the Figure – 8 shows that the student has spend maximum time in his Home and then in College in a month.

Conclusion

The moving object of student data is represented in a graphical format by the Rattle Classification tree. The graphical representation gives the overall summary of the student maximum time spent in a month and all the new placed visited by him in a month. If a student does not visit college daily and suppose he is visiting the new place for maximum time, then the parents and college authorities can identify this and counsel him as necessary.

References

[1] "Mining Periodic behavior for Moving Objects" by Zhenheri Li, Bolin Ding, Jiawei Han, Roland Kays, Peter Nye. University of Illinois at Urbana-Champaign, Illinois, US. New York State Museum, New York, US

[2] Tutorial on Spatial and Spatio-Temporal Data Mining (SBBD-2008) *Introduction to Moving Data and Moving Object Databases* Vania Bogorny Universidade Federal do Rio Grande do Sul. Shashi Shekhar University of Minnesota

[3] "Querying Moving Objects in SECONDO" Victor Teixeira de Almeida, Ralf Hartmut Güting, and Thomas Behr LG Datenbanksysteme für neue Anwendungen Fachbereich Informatik, Fernuniversität Hagen D-58084 Hagen, Germany

[4] "Data Mining II (Mobility Data Mining) by Micro Nanni, ISTI – CNR.

[5] "Continuous Clustering of Moving Objects" by Christian S Jensen, Dan Lin, Beng Chin Ooi.

[6] "Clustering Moving Objects", Jiawei Han, Yijan Li and Jiong yang. Department of Computer Science, University of Illinoies, USA

[7] Moving Objects Databases. Morgan Kaufmann Publishers, 2005 by R H Guting and M Schneider.

[8] Data Mining Concepts and Techniques by Jiawei Han, Micheline Kamber and Jian Pei.

[9] Rattle: A Data Mining GUI for R by by Graham J Williams

[10] http://cran.r-project.org/web/packages/rpart.plot/index.html