Webology (ISSN: 1735-188X)
Volume 18, Number 6, 2021

Modeling The Data Analytic Approach For Efficient Activity Classification And Clustering

Syed Umar[1] , Cherukuri Roopa Chowdary [2] , Bezawada Neha Chowdary [3] , Talasila Sri Harsha [4] , Kondepudi Srikar Sai [5]

[1] Professor, Department Of Computer Science, College of Engineering, Wollega university, Nekemte, Ethiopia

[2] student, Department Of Computer Science& Engineering, Vellore Institute Of Technology, Amaravati, India,

[3] student Department Of Computer Science& Engineering, Vellore Institute Of Technology, Amaravati ,India

[4] student, Department Of Computer Science& Engineering, Vellore Institute Of Technology, Amaravati, India

[5] student, Department Of Computer Science& Engineering, Vellore Institute Of Technology, Amaravati, India

---

ABSTRACT

The human activity context awareness techniques have been involved in multiple sectors like sports security and health monitoring applications and in order to acquire activity information we have some embedded approaches like accelerometer data and external approaches like cameras data. To collect this data, we may use multiple sensors simultaneously for the efficient support to the real time acquisition of data. Also, most of the researchers are focusing accuracy and efficient classification after activity based learning and large amount of data processing. Now, we came up with the unique architecture where we can classify and cluster the data instantaneously. this been incorporated to one of the open source tool call WEKA. we've been processed the data into year four stages one is data collection and feature extraction activity recognition and pattern mining concentrated how efficiently we can organize these input and output data with the efficient processing speeds the data organizations can be manipulated through multiple international and open source classification and clustering algorithms. They've been working with set up people and gathering accelerometers data R which already defined in the form of public datasets to build their architectures.

Keywords: Data Aggregation, Interface Mechanism, Classification, Clustering, Smart Home, Behavior Analysis.

INTRODUCTION

General data mining procedures can be involving multiple attribute selection training and building the datasets performing multiple validation tests and generating resultant confusion matrix determining accuracy rates and processing times. Our main goal is to compare multiple different classification methodology's using different training and building models, data specification matches and works enables performance assessment with accuracy rate predictions. Here we already used it based network which already implemented with neural networks With excellent results and it also helps too investigate further on multiple validation methodology's and Configuration techniques. We also used K fold cross validation techniques for best results and to process large amount of data processing in multiple times initially the users can able to access historical action database and non sequential procedure manipulations, then it creates sequences based on actions under the action rules which been already generated.

$$[a_x, a_y, a_z, \theta] \quad (1)$$

To eliminate rotation interference, accelerator data is normalized:

$$[a_x, a_y, a_z, \theta] = [\sqrt{a_x^2 + a_y^2 + a_z^2}, \theta] = [\bar{a}, \theta] \quad (2)$$

With the addition of first order difference, input vector for feature extraction is:

$$[\bar{a}, \bar{a}', \theta, \theta'] \quad (3)$$

Sequential mining method is been implemented and figured out the frequent sequential patterns and personal habits operating device manager so nonsequential rules are being checked for users activity recognition and following sequential patterns.

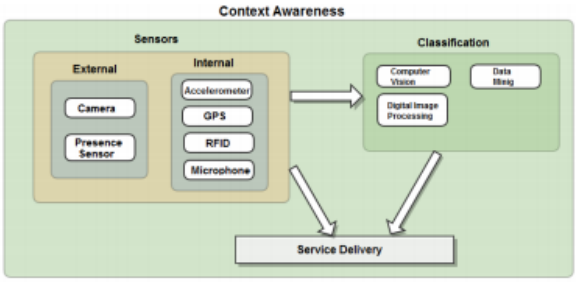

Figure 1 Classification matrix

Wireless technologies also rapidly developing in nowadays and the competing techniques has been implementing distributed methodologies to perform efficient computing off data. Here in this human activity analysis we may retrieve contrast information off locations objects devices and environment variable statuses for this wireless sensors are portable devices could be more

helpful to collect the data from the surrounding environments. a large amount of context awares systems being developed and designed to deal with large amount of context data. here we obtain the data from multiple sensors on devices in several topics has been interlinked with quality context data collection been proposed predefined policy rules being helpful to generate raw data from implicit context information to gather inconsistent data and high level of contact information in francs with multiple set of policy rules.

From the knowledge base we are mapping setoff relationships between different transitions and their statuses and combinedly quality as policy rule. in most of the cases the domain experts are system designers can we build this policy rules for the context aware systems these people are part of the system and not the customers but one who really uses this architectures. These rules can be changed once we've been done the modifications after the system being designed. so we need to build additional. rules in a personalized manner to facilitate multiple requirements of end users. Activity patterns were bit dependent on user habits and based on this user's activity the context aware systems can provide services. in generally these systems can be analyzed by using image or a video. Multiple devices can be provide the information of human activities are human actions in this environment.

Rather than capturing the movements are poses from the images are videos we are analyzing human activity behaviors through context information which is being provided by the these systems will give accurate human habits and we can able to recognize the real behavior.

Figure 2 Architectural flow

Nowadays the usage of this wearable computing technology services are applications has been increased and major gadgets manufacturers like Samsung and Apple has been launched their heartbeat sensors and smart watches 2 aid their customers to monitor customer health and customer behaviors these devices are very intelligent and more near real time services based on individual user behaviors. the term human activity recognition where generally focuses the usability and near real time information classification of human activities based on the data which been collected by multiple users in a home are single action single activity user.

We compared multiple devices multiple user information's with respect to location and heartbeats and we've been concluded multiple scenarios based on the user activities and provider services.

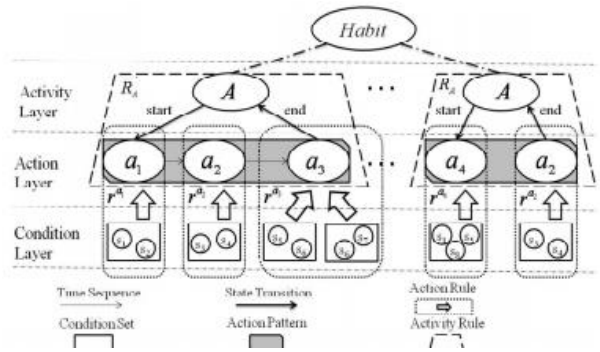

Figure 3 Layer dependency model for activity,action and conditions

The example applications where we've been considered the usage of human activity recognition in applications like false detection in elderly care, location based mobile services in security, Physical activities monitoring in assisted health care applications, etc. The computational process can be performed with multiple patterns based on the large amount of data to find out the intelligence in an understandable manner and multiple data techniques and data mining algorithms been utilizing for this retrieval of useful information from big data. We are going to classify our performing clustering processes to train the database are analyze the individual set of rules which can be created the environment for human activity recognition and automatically it can able to analyze knew incoming data as well. we can also use multiple classification machine learning algorithms not only trying the new data to generate most accurate results as much as possible.

Our data disaggregation parameters structures also we are using this data mining approaches commonly to collect the data from sensors such as accelerometers then to classify these results to determine the activity recognition. Our context awareness diagrams can be explained as part of context awareness applications and these were little bit complex and covers most of the stages off data collections from different sensors and generate accurate and dynamic services based on user inputs.

In other area to perform the behavior analytics of a human the researchers are performing multiple research use cases on social relationships shopping lists and commentary cards off consuming goods and comments so on. these capturing interests based on the taste and the characteristic of the users we could be able to identify their behavior of particular aspects and their involvement towards the use cases.

We've been provided a great system architecture to address these problems of activity predictions and human activity recognition and also addresses how the data collection can be performed and analyzed in nowadays activity location sequences parameters. In other yes team or the system design architecture we have 5 individual stages where it can be maintain a flow of system analysis and prepares the data to the customers.

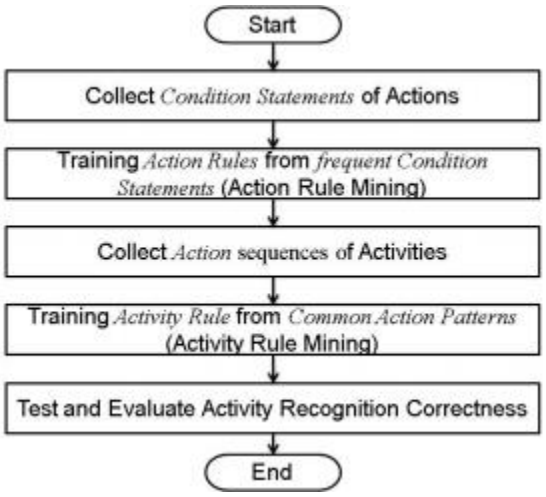

Figure 4 data flow stages from collecting to testing

Stage 1: Data Collection

Here we are using data acquisition yes DEQ task for running and executing applications on smart phone which can able to initial sensor and GPS Global Positioning System information. the general frequency of sensor data is 40 hedges. Whatever the data we're collecting from this sensors will go into upload to a centralized web server for transmitting and low power consumption

Stage 2: Feature Extraction

Here extraction of feature can be happened as important module which can able to direct and accelerate here human directions can be calculated and mapped with the help of GPS and compass, campus easy internal application can be helpful within the mobile, and GPS is a application which can able to detect the location details of outside environment. multiple vectors and exhilarating information can be used to eliminate rotation interference and making this accelerator data much more normalized with respect to the data sequences the segments which can be have more than 10 seconds will be act as a stable activities for the future extraction the sample features among 49 regular features have been selected and working on the expertise the domain features like mean variance K peak value and K list value skewness and etc

Stage 3: Activity Recognition here we are going to be label and predicted the unlabeled activities along with training with the activity classifier whenever new data arrives to the architecture this data will be compared with the training data in an aspect of one versus rest to label this features and extracting their activities from detrained models. Multiple scenarios has been distinguished to label these classifiers the data could be single arise are multiple level data integrations the optimization can be performed in efficient manner and recognizing these human activities with the help of labeling concept.

Stage 4: Pattern Mining

Whenever we are ready with the activity sequences retrieved from different objects of data manipulation techniques based on the level of feature extraction we are clustering and analyzing this activity sequences to perform pattern mining methodology.

The generalized architecture where user can able to feed the data to the sensors and the sensor information can also process the human actions and perform basic operations related to categorization techniques the time analysis can be helpful here to categorize the human actions based on multiple set of time periods morning evening afternoon night , and also changes with day off week.

1. Planning. a) The connections between the info instances information and the classes that are objective seemed for, isolating a lot of data for planning. I) The preparation information could be isolated utilizing procedures which can be various because it really well might be seen later in this paper. b) The classifier is constructed reliant on a dataset whoever classes are now actually known - Supervised earning.

2. Approval Tests: a) With the model that is assembled lot of rules), for example the classifier itself, it utilizes another arrangement of information for the tests. b) by making use of the classifier, positive results are contrasted plus the classes which can be ideal what's the class to which has a place each dataset case.

3. outcomes: a) the outcomes that are overall examined: precision rates, Confusion Matrices, the preparation and model development execution and approval tests results.

the interesting point the human actions can also vary with the seasons for example the usage of colors can be more in summer by comparing with winter. The mapping of these activities can also need lot of concentration towards the time day and situations we couldn’t be performed the accurate to 100%, because the data manipulations can be changed with any of different kind of human intentions. we also tried to perform location analysis where we can able to track the humans and the device locations to track much accuracy but it has a lot more challenges to deal with high frequency and less efficient processes. We have been targeted these evaluation models with respect to accuracy and recall rates the results of training data set and test datasets have been compared with High architecture models.

The WEKA workbench is an assortment of AI calculations and information preprocessing apparatuses that incorporates essentially all of the calculations portrayed in our guide. It is prepared with the target that one may rapidly assess methods that are existing new datasets in adaptable manners. It gives uphold that is broad the entire period of trial information mining, including setting up the knowledge information, assessing learning plans factually, and imagining the knowledge and the consequence of learning. Just as a assortment that is wide of calculations, it includes a wide scope of preprocessing instruments. This assorted and toolbox

that is exhaustive gotten to through a typical interface so its clients can look at changed strategies and recognize those that are suitable for the issue at hand. WEKA was created at the University of Waikato in New Zealand; the true name represents Waikato Environment for Knowledge Analysis.

The WEKA, articulated to rhyme with Mecca, is a flightless fledgling with a curious nature discovered distinctly on the Hawaii an islands of the latest Zealand outside the college. The framework is written in Java and disseminated under the conditions regarding the GNU General Open License. It operates on virtually any stage and happens to be tried under Linux, Windows, and Mac frameworks that are working. WEKA gives executions of learning calculations you can undoubtedly apply to your dataset. It additionally incorporates a variety of instruments for changing datasets, for example, the calculations for examining and discretization. It is possible to preprocess a dataset, feed it into a learning plan, also, digest the classifier that is subsequent its exhibition—all without creating any program rule by any stretch of the imagination. The workbench incorporates strategies for the principle information mining issues: relapse, characterization, grouping, affiliation rule mining, and trait dedication.

Becoming acquainted with the info is an vital aspect of the work, and information that is numerous offices and information preprocessing devices are given. All calculations take their share to the type of a solitary table that is social is perused from a record or created by an information base question. One method of utilizing WEKA would be to apply a learning method to a dataset and dissect its yield to review the info. Another is to utilize learned models to produce forecasts on new cases. A third is to use a few students which can be unique look at their presentation completely to choose one for forecast. In the WEKA that is intelligent screen select the educational strategy you will need from a menu. Numerous techniques have actually tunable boundaries, which you access through a property product or sheet proofreader. A evaluation that is typical is useful to quantify the presentation things considered.

Evaluation & Results:

Usage of real learning plans will be the most asset that is significant WEKA gives. Nevertheless, instruments for preprocessing the information, called channels, come a nearby second. Like classifiers, you select channels from a tailor and menu them to your prerequisites. The most approach that is effortless utilize WEKA is via a graphical UI called the Explorer. This offers admittance to the entirety of its offices menu that is utilizing and structure filling. As an example, you can instantly read in a dataset from the document and construct a choice tree from this.

We can also consider the data retrieval systems with action rule mining and the activity rule mining the generalized idea of categorizing these into two techniques To manage the efficiency between activities and actions. the human can able to perform multiple actions under these actions can able to reach activity sequences, there are lot more challenges we've been faced while retrieving these data and we've been managed to consider a large data set which are being

able to deliver in public environment the input data set we used is united kingdom's largest home appliances usage statistiques which has been maintained from last five years of database as a sample we've been taking the data into my SQL database and perform multiple actions and redecoration procedures on my scale database and we've been integrated the entire environment using graphical user interface which being developed by Java programming language. Java is the advanced programming language where we can able to write logical sequences of applications to perform multiple sequences and behavior of software application. We have a lot more advantages with the Java where it's been object oriented and having major concepts of in capsulation abstraction inheritance polymorphism along with key attributes of class and object. The fundamental of this application to develop are to service the end users who wants to know their individual bills based on recorded household information.. the statistiques application can only perform in a different manner but categorization can also be requested on the scope.

```
Time taken to build model (full training data) : 1.25 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       174437 ( 46%)
1       202832 ( 54%)


Class attribute: timeslot
Classes to Clusters:

     0      1  <-- assigned to cluster
 65439 76112 | NIGHT
 29164 33923 | EVENING
 28954 33623 | AFTERNOON
 50880 59174 | MORNING

Cluster 0 <-- MORNING
Cluster 1 <-- NIGHT

Incorrectly clustered instances :       250277.0         66.3391 %
```

Figure 5 Clustered values of timeslot data

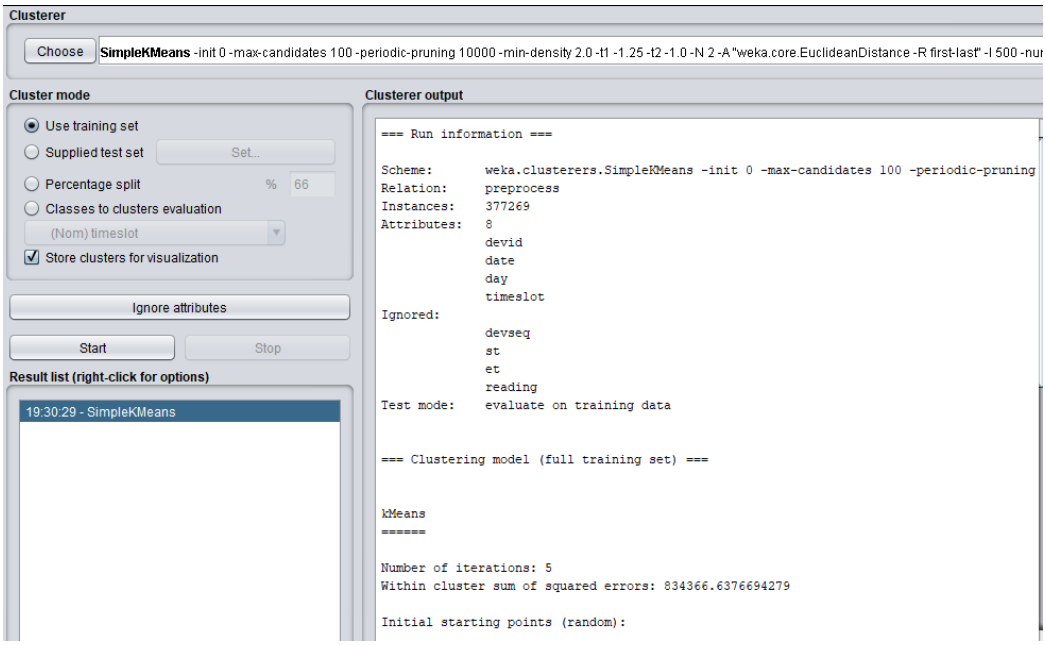

Figure 6 K means clustering of input data

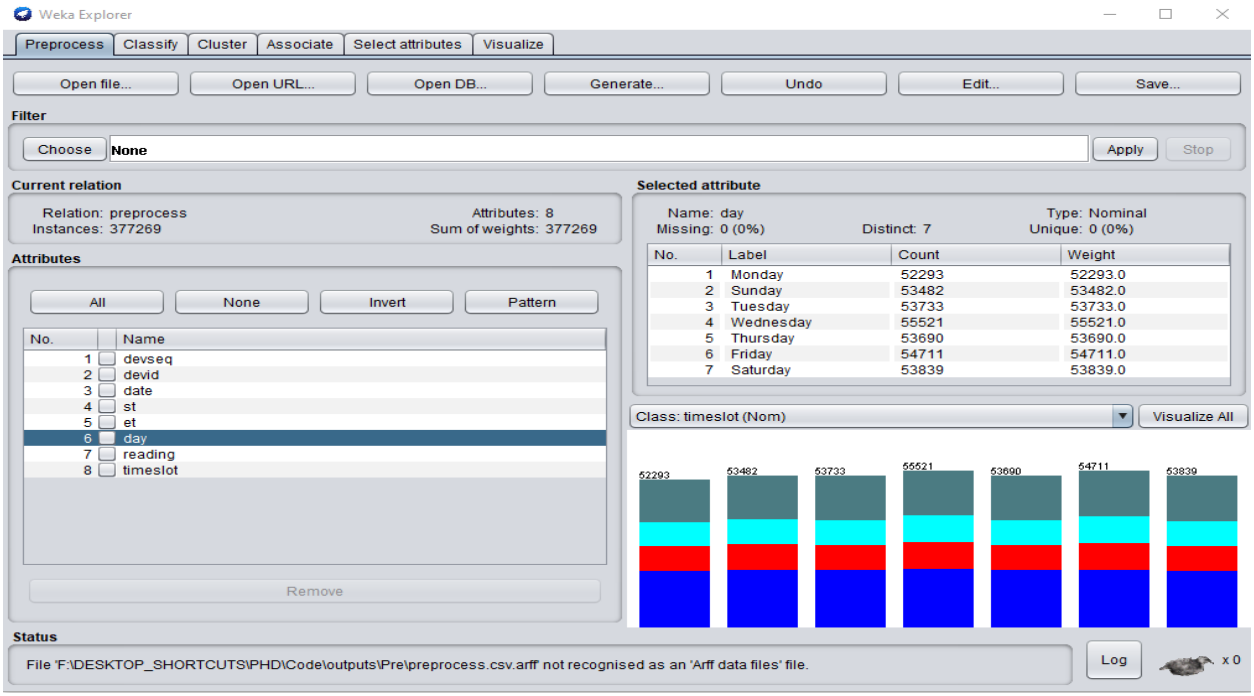

Figure 7 day wise classification towards timeslot data

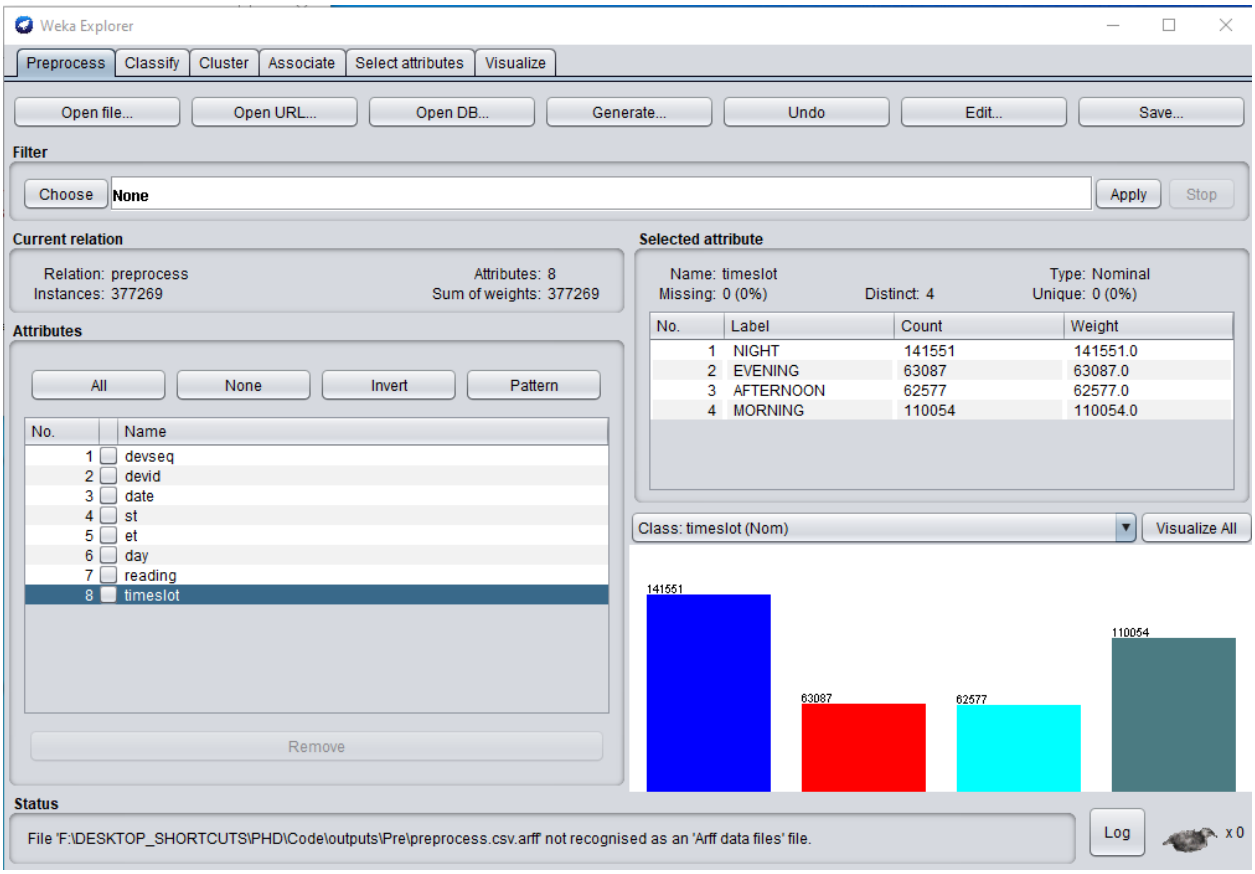

Figure 8 timeslot data classification in WEKA tool

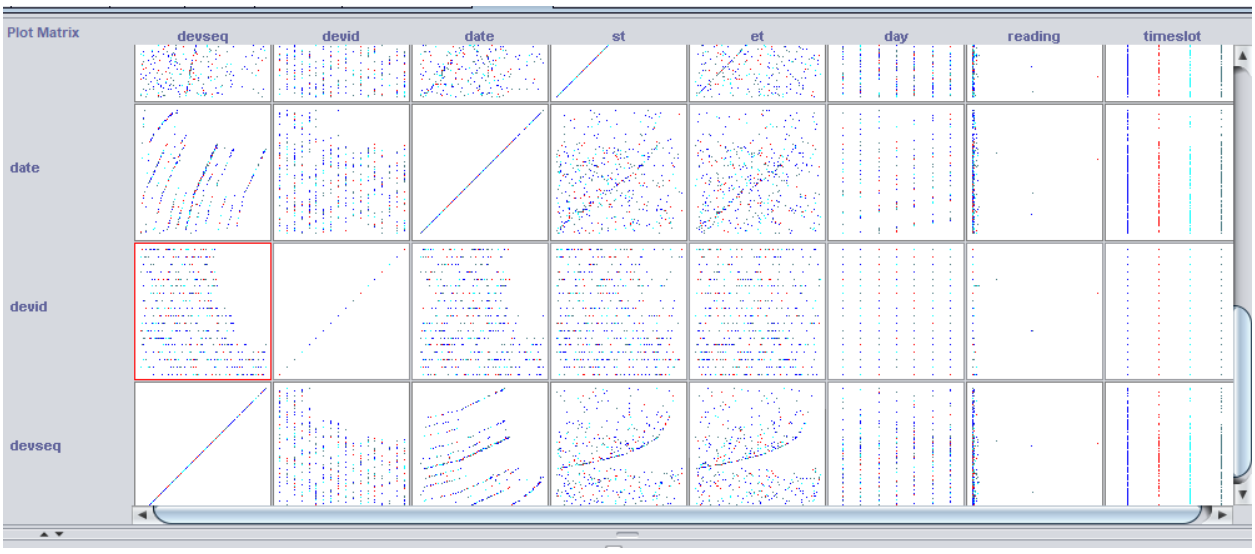

Figure 9 Data visualization after data processing

CONCLUSION:

Human activity patterns are much crazier not before most of the researchers nowadays want to explore huge amount of human activities and predict the human behaviors we also done multiple research mentions on this behavioral activities. Now we came up with the unique architecture where we can classify and cluster the data instantaneously. This been incorporated to one of the open source tool call WEKA. we've been processed the data into year four stages one is data collection and feature extraction activity recognition and pattern mining concentrated how efficiently we can organize these input and output data with the efficient processing speeds the data organizations can be manipulated through multiple international and open source classification and clustering algorithms. These data results can also occupy the huge amount of preprocessing data which being generated through Hadoop distributed clusters we also projected this data to one Java based customized application and generated graphs based on appliance too usage statistics. The normalized data can also occupy these generic experiments and involved into a multiple parameter. The experiments were shown the efficient results not only with respect to the classifiers and the data which we've been given as input.

Our future work wanted to be building a robust application where even in non-technical person also can able to categorize these classifiers and cluster huge amount of disaggregated data.

REFERENCES:

1. L. M. Rodrigues and M. Mestria, "Classification methods based on bayes and neural networks for human activity recognition," 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, 2016, pp. 1141-1146, doi: 10.1109/FSKD.2016.7603339.
2. J. Li, P. Huang, Y. Kuo and K. Lee, "A Hierarchical Rule-Based Activity Recognition System with Frequency Pattern Mining," 2009 Fourth International Conference on Innovative Computing, Information and Control (ICICIC), Kaohsiung, 2009, pp. 1554-1557, doi: 10.1109/ICICIC.2009.19.
3. H. Gong, K. Xing and W. Du, "A user activity pattern mining system based on human activity recognition and location service," IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Honolulu, HI, 2018, pp. 1-2, doi: 10.1109/INFCOMW.2018.8406918.
4. W. Ugulino, M. Ferreira, E. Velloso e H. Fuks. Virtual Caregiver: Colaboração de Parentes no Acompanhamento de Idosos. Anais do SBSC 2012, IX SimpósioBrasileiro de Sistemas Colaborativos, pp. 43- 48. São Paulo, SP: IEEE. ISBN 978-0-7695-4890-6, 2012, (in Portuguese).
5. H. Gjoreski, M. Lustrek e M. Gams. Accelerometer Placement for Posture Recognition and Fall Detection. In: 7th International Conference on Intelligent Environments, IE, 2011
6. Weka 3: Data Mining Software in Java. Available on: https://www.cs.waikato.ac.nz/ml/weka/ . Access on: 03 November 2014.

7. Y. Bu, T. Gu, X. Tao, J. Li, S. Chen, and J. Lu, "Managing quality of context in pervasive computing," Proc. Quality Software International Conference (QSIC), Oct. 2006, pp. 193-200, doi: 10.1109/QSIC.2006.38.
8. X. Lin, S. Li, J. Xu, W. Shi, and Q. Gao, "An efficient context modeling and reasoning system in pervasive environment: using absolute and relative context filtering technology," International Conference on Advances in Web-age Information Management, Nov. 2005, doi: 10.1007/11563952_32.
9. S. Hongeng, F. Brémond, and R. Nevatia, "Bayesian framework for video surveillance application," Proc. IEEE International Conference on Pattern Recognition (ICPR), IEEE Press, vol. 1, Sep. 2000, pp. 164-170, doi: 10.1109/ICPR.2000.905296.
10. M.S. Ryoo and J.K. Aggarwal, "Recognition of composite human activities through context-free grammar based representation," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Press, Jun. 2006, pp. 1709-1718, doi: 10.1109/CVPR.2006.242.
11. S.S. Li, "Human Activity Pattern Mining and Recognition by Sequence Alignment algorithm," Master Thesis, University of National Cheng Kung University, 2008, pp. 10-19, unpublished.
12. Xing Su, Hanghang Tong, Ping Ji, "Activity Recognition with Smartphone Sensors," TsingHua Science And Technology, pp235-249, Volume 19, June 2014.
13. A. M. Khan, Y.-K. Lee, S. Y. Lee, T.-S.Kim, "Human Activity Recognition via An Accelerometer-Enabled-Smartphone Using Kernel Discriminant Analysis," FutureTech Conference, Busan. South Korea, ISBN: 978-1-4244-6948-2, June 2010.