Vol.3 No.1 2019

Universidad de Oviedo

CMR Institute of Technology
EXPLORE TO INVENT

ISSN: 2581-3242

International Journal of

Machine Learning and Networked Collaborative Engineering

Vol.3 No.1 2019

International Journal of
Machine Learning and Networked Collaborative Engineering

PREFACE

The International Journal of Machine Learning and Networked Collaborative Engineering (IJMLNCE) with ISSN: 2581-3242 continues its growth. The journal is becoming more and better indexed in platforms such as BASE (Bielefeld Academic Search Engine), CNKI Scholar, CrossRef, CiteFactor, Dimensions, DRJI, Google Scholar, Index Copernicus, JournalTOCs, J-Gate, Microsoft Academic, PKP-Index, Portico, ROAD, Scilit, Semantic Scholar, Socolar, WorldCat-OCLC. After more than two years of intense work, we are proud to present the seventh volume of the journal, Volume No-03 Issue No-01, which introduces five high quality works written by recognized authors that deal with different aspects within the scope of the journal.

Nigar[1] published a work entitled "A Study on Internet of Things in Women and Children Healthcare". Author focuses on the importance of the new internet of things related- technologies, which can be applied to healthcare. In fact, its integration with electronic health and telemedicine is gaining attention. The paper describes some methods, practices and prototypes based on the internet of things in the field of healthcare focusing on women and children.

Gunagweare and Kiani [2]published a work entitled "Ultimate Indoor Navigation: A Low Cost Indoor Positioning and Intelligent Path Finding". Authors deal with the drawbacks of the global positioning system (GPS), which is not useful in indoor environments or places where some buildings can interfere with the satellite signal. In this paper, authors present a simple, low-cost, context-aware and user-friendly indoor navigation system based on a common smart phone.

Jaidev et al.[3] published a work entitled "Artificial Intelligence to Prevent Road Accidents". Authors focus on the traffic congestion that in turn can lead to more car accidents. The idea of the authors is to study and review the literature related to approaches for detecting unsafe driving patterns to predict accidents with the help of artificial intelligence. Two apparently similar but different examples could be drivers under the influence of drugs or drivers under the influence of alcohol.

Rimal [4] published a work entitled "Deterministic Machine Learning Cluster Analysis of Research Data: using R Programming". The paper discusses various types of cluster analysis of different data sets with large number of dimensions (iris, utilities, mclust and dbscan). The main goal is to explain the simplest way for clustering analysis whose data structure is wide scattered. The work of the author is based on the R programming language and several specific packages.

Gunjal and Shaik [5] published a work entitled "A Robust Decomposition Based Algorithm for Removal of Pattern Noise from Images". Authors work on a melting pool of complex vectors to present a technique that requires less computer resources and less time for any image removal of pattern noise compared to other previously stated strategies. The work is based on the idea a picture includes components that can be described separately.

REFERENCES

[1] Nigar, Nishargo (2019). A Study on Internet of Things in Women and Children Healthcare. *International Journal of Machine Learning and Networked Collaborative Engineering*, 3(01) pp 1-15.

https://doi.org/10.30991/IJMLNCE.2019v03i01.001

[2] Gunagwera, Alex & Kiani, Farzad (2019). Ultimate Indoor Navigation: A Low Cost Indoor Positioning and Intelligent Path Finding. *International Journal of Machine Learning and Networked Collaborative Engineering*, 3(01) pp 16-34.

https://doi.org/10.30991/IJMLNCE.2019v03i01.002

[3] B Jaidev, Garg Sonakshi, Makkar Sandhya (2019). Artificial Intelligence to Prevent Road Accidents. *International Journal of Machine Learning and Networked Collaborative Engineering*, 3(01) pp 35-45.

https://doi.org/10.30991/IJMLNCE.2019v03i01.003

[4] Rimal, Yagyanath (2019). Deterministic Machine Learning Cluster Analysis of Research Data: using R Programming. *International Journal of Machine Learning and Networked Collaborative Engineering*, 3(01) pp 46-65.

https://doi.org/10.30991/IJMLNCE.2019v03i01.004

[5] Gunjan, Vinit Kumar & Shaik, Fahimuddin (2019). A Robust Decomposition based Algorithm for Removal of Pattern Noise from Images. *International Journal of Machine Learning and Networked Collaborative Engineering*, 3(01) pp 66-75.

https://doi.org/10.30991/IJMLNCE.2019v03i01.005

Editor-in-Chief's

International Journal of Machine Learning and Networked Collaborative

Vijender Kumar Solanki, CMR Institute of Technology, Hdyerabad, TS, Insia

https://orcid.org/0000-0001-5784-1052

Vicente García-Díaz, University of Oviedo, Spain

https://orcid.org/0000-0003-2037-8548

DOI : https://doi.org/10.30991/IJMLNCE.2019v03i01

EDITORIAL BOARD MEMBER

EDITORIAL BOARD MEMBER

INDEX

EDITORIAL OFFICE

Vicente Garcia Diaz, Ph.D.
Associate Professor
Department of Computer Science
University of Oviedo, Spain
garciavicente@uniovi.es

Vijender Kumar Solanki, Ph.D
Associate Professor
Department of Computer Science & Engineering
CMR Institute of Technology (Autonomous), Hyderabad, TS, India
vijendersolanki@cmritonline.ac.in,spesinfo@yahoo.com

CALL FOR PAPER

International Journal of Machine Learning and Networked Collaborative Engineering (IJMLNCE) with ISSN **2581-3242**, is a quarterly published, open access, peer-reviewed, international journal, focuses on publishing authentic and unpublished quality research papers in the emerging field of Machine Learning and Collaborative Engineering.

This is a scientific journal aims to promote highly qualitative and innovative research works, not limited to but focuses on the area of machine learning, collaborative engineering, and allied areas. We publish only original research articles, review articles, and technical notes.

The journal is registered with CrossRef **10.30991/IJMLNCE**. Each accepted manuscript shall be assigned a unique digital object identifier for identification. The abbreviated key title for our journal allocated by ISSN is **Int. j. mach. learn. Networkedcollab.eng**.

The journal appears in popular indexes like BASE (Bielefeld Academic Search Engine), CNKI Scholar, CrossRef, CiteFactor, Dimensions, DRJI, Google Scholar, Index Copernicus, J-Gate, PKP-Index, Portico , ROAD, Scilit, Semantic Scholar, Socolar, WorldCat. Our journals manuscripts can also be found in the directory of ReserachGate & Kudos.

In this present technological era, the areas like machine intelligence, machine learning, and its associated domains are one of the most popular and demanding choices for the researchers as well as the industry personnel.

In last few years, numerous uses of machine learning and its related domain, has drawn ample attention of the people, that has generated a large number of applications in this field, making machine learning and collaborative engineering highly admired one.

Machine intelligence or machine learning is not a new concept. In terms of Artificial Intelligence, we were familiar with several aspects of the field, but nowadays with the introduction of machine learning, the use of this has been highly evolving, especially for improving the lifestyle of the human being.

There are numerous application areas of machine learning or machine intelligence, irrespective of any famous sector, from Healthcare, Space, Automation, Aviation industries etc. to entertainment industry and even in academia.

With this insight, IJMLNCE invites quality and authentic research works in form of manuscripts, for our forthcoming version. The Journal solicits latest research work carried out by the researchers, academicians and industry practitioners, in the domain of machine learning and collaborative engineering.

Plagiarism is considered as a serious ethical and professional malpractice in case of IJMLNCE. So similarity check should be taken care at its highest level. The originality of the work will be highly appreciated. Each paper will be peer-reviewed by the expert reviewers and in a multilayer approach. The editorial board of the journal believes in high moral and ethical practices.

The authors are invited to submit their authentic and unpublished manuscripts online only. They will get final decision about the acceptance or rejection of the manuscript within 06-12 weeks from the date of submission of manuscripts. Once the manuscript gets accepted, author will be asked to complete the copyright agreement. An e-mail communication will also be sent to the author's registered mail ID, after verification of incorporation of all reviewers' comments, in term of their final camera-ready-paper; it will be published to the upcoming journal's volumes.

Please note that there is **no article processing cost** for publishing articles in this journal and the journal is open access.

Vol. 03, No. 1 ,(2019), 001–015 , **ISSN:2581-3242**

IJMLNCE JOURNAL

International Journal of Machine Learning and Networked Collaborative Engineering

Journal Homepage: http://www.mlnce.net/home/index.html

DOI : https://doi.org/10.30991/IJMLNCE.2019v03i01.001

A Study on Internet of Things in Women and Children Healthcare

Nishargo Nigar

IEEE Member

nishargo@ieee.org, https://orcid.org/0000-0003-2639-7213

Abstract

Individual entities are being connected every day with the advancement of Internet of Things (IoT). IoT contains various application domains and healthcare is one of them indeed. It is receiving a lot of attention recently because of its seamless integration with electronic health (eHealth) and telemedicine. IoT has the capability of collecting patient data incessantly which surely helps in preventive care. Doctors can diagnose their patients early to avoid complications and they can suggest further modifications if needed. As the whole process is automated, risk of errors is reduced. Administrative paperwork and data entry tasks will be automated due to tracking and connectivity. As a result, healthcare providers can engage themselves more in patient care. In traditional healthcare services, an individual used to have access to minimal insights into his own health. Hence, they were less conscious about themselves and depended wholly on the healthcare facilities for unfortunate events. But they can track their vitals, activities and fitness with the aid of connected devices now. Furthermore, they can suggest their preferred user interfaces. This paper describes several methods, practices and prototypes regarding IoT in the field of healthcare for women and children.

Keywords

Internet of Things

Healthcare,

eHealth,

mHealth

1. Introduction

The usage of IoT in healthcare services has been increased to a great extent across a range of IoT use cases. As human lives depend on healthcare, it is certainly a crucial area. Digitizing process is quite challenging for the researchers. They are constantly creating new and novel solutions on digital medical services. IoT is playing a major role in this sector. According to Cisco, about 50 billion devices will be connected to internet by the year 2020. We can trace IoT's success in the use of medical device integration, telemonitoring, smart sensors, glucose monitors, medication dispensers, activity trackers and wearable biometric sensors. Special applications like smart pills, smart beds, personal healthcare, smart home care and Real-Time Health Systems (RTHS) are latest additions to digital healthcare. Remote patient monitoring is a new way to introduce smart healthcare to mass people [1].

In underdeveloped and developing countries, women and children healthcare is an important agenda of government because of high maternal and children mortality rate. 44% decline in global maternal mortality ratio is observed between 1990 and 2015 according to Sustainable Development Goal (SDG)

Report in 2016. World Health Organization (WHO) highly puts an emphasis on the quality of antenatal care [2]. eHealth and mobile health (mHealth) system developers are designing specific healthcare applications for women and children to create awareness regarding malnutrition, obesity and mortality issues. Children healthcare faces a lot of troubles because of various factors like insurance coverage, limited resources and insufficient medicinal resources [3]. In order to polish the current healthcare facilities available to women and children, their active participation rate is mandatory. Attributes of a healthy lifestyle consists of several activities, proper knowledge of nutrition, appropriate diet plan and awareness of risky diseases. Implementation of such healthcare applications [4] [5] assists in active involvement of users. In addition, the mHealth and eHealth solution developers come to learn about user interactions, faults present in traditional systems and useful feedback.

User experience must be taken into consideration while making an innovative approach. The dynamic functionalities of IoT is making it possible to ensure user experience and it has been possible after a lot of research on security, protocols, embedded devices, applications, interoperability and network architectures. To deploy the technology, a lot of countries are applying relevant policies and guidelines [6]. Women and children healthcare is considered as a sensitive case because of its vulnerable and ill-fated statistics. It is necessary to implement the most appropriate and efficient systems to uphold their health status. But the pressure of rising costs in healthcare systems can decelerate the whole process. It is a matter of elation that various software and connected devices are available to mitigate some financial problems that the healthcare providers are facing at present. This is being possible by the recent advancements of IoT.

In this paper, we made an effort to highlight the digital healthcare solutions based on IoT available for women and children and how they are reshaping the current scenario of conventional healthcare management systems. For the most part, an extensive set of prototypes available for women and children are demonstrated throughout our paper. We also described the challenges and ongoing trends of present digital healthcare systems for the target users.

2. Applications & Services

Applications of IoT are rising rapidly due to our needs and demands. Radio frequency Identification (RFID), machine-to-machine (M2M) and Near-field communication (NFC) are being applied and creating a revolution in the field of IoT. Although the concept of M2M might sound complex, the idea behind the execution is pretty simple. M2M networks function as Wide Area Network (WAN) or Local Area Network (LAN) but primarily used for permitting controls, sensors and machines to communicate among each other. The devices feed information acquired to other devices present in the same network. Such system allows a person to assess what is happening in the entire network environment and issue necessary instructions to member devices.

Various researches have been conducted until the present relating to IoT targeting women and children. We have created an exclusive category of works and prototypes related to IoT in women and children healthcare in figure 1. From time to time, researchers have built smart devices, proposed new methods or suggested enhanced systems for the welfare of the target group. We have also discussed about their drawbacks so that in future, researches can be conducted solving the issues.

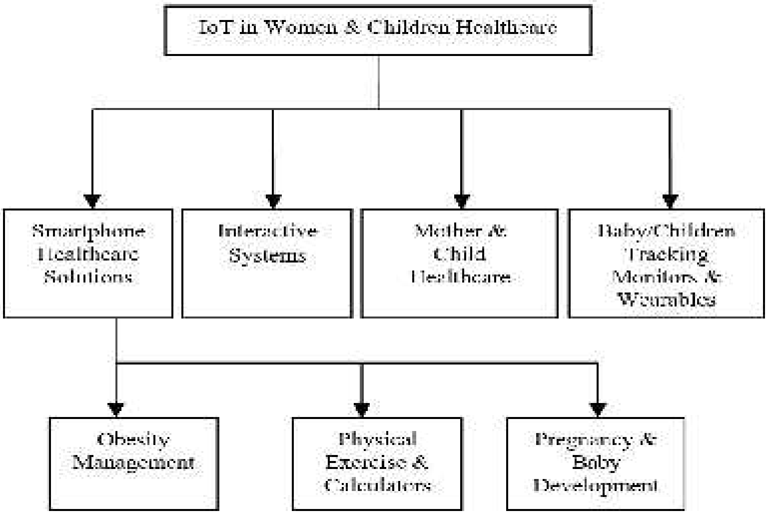


Figure 1. IoT applications and services targeting women and children

2.1. Smartphone Healthcare Solutions

Nowadays, smartphones are capable of doing more than making phone calls or sending texts. They are equipped with Global Positioning System (GPS), Wireless Fidelity (Wi-Fi), NFC, voice recognition and sensors which make them so productive [8]. Today's smartphones consist of almost 14 sensors that generate raw data of location, motion and surrounding environment. With the aid of micro-electromechanical systems (MEMS), this revolution has been possible. MEMS structure is composed of a number of four elements: microsensors, microactuators, microstructures and microelectronics (shown in fig. 2). Microsensors and microactuators are classified as transducers that exchange energy from one type to another. The device turns a mechanical signal into an electrical one when it comes to microsensors. But the main purpose of MEMS is to combining such tiny structures into a regular silicon substrate with integrated circuits. Micromechanical elements are fabricated with micromachining process where sections of silicon wafer are incised or new layers are added to obtain electromechanical devices.

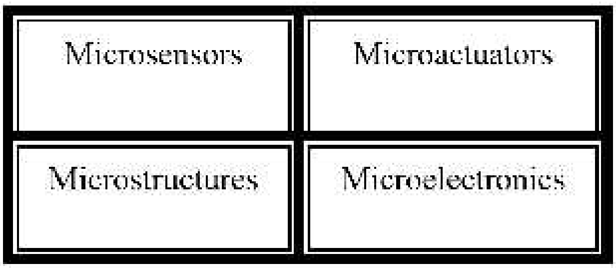


Figure 2. Components of micro-electromechanical systems (MEMS)

The major sensors in a smartphone include: magnetometer, GPS, gyroscope, accelerometer, barometer, proximity sensors and so on. Magnetometer is actually known as a compass and with GPS, it can detect one's current location. In cell phones, GPS receivers communicate with available units from an array of 30 global positioning satellites within the system. Trilateration (process of detecting absolute location points) is done with built-in receiver from minimum three satellites and the receiver. GPS can measure one's location by an equation based on the GPS receiver of cell phone and intersection point of overlying spheres gained from satellites. On the other hand, accelerometer in a smartphone is comprised of a circuit with seismic mass which means it is made from Silicon. In fact, it is a circuit with the basis of MEMS. The circuit senses a force of acceleration due to tilting or movement of the phone.

Several smartphone based healthcare solutions were created previously targeting women and children. They are of different subjects and offer distinct services towards healthcare management of an individual.

They are mostly based on nutrition facts, obesity management, exercise, diet plan, calorie counter, step counter, period tracker and pregnancy management.

2.1.1. Obesity Management

Concept of IoT is utilized to spread knowledge of various food items and obesity management. In the past, Automatic Identification and Data Capture Techniques (AIDC) has been used in mHealth platform to create health awareness among children. In AIDC, three major elements are necessary (fig. 3); the first one is a data encoder. A code represents symbols that are generally alphanumeric characters. While encoding data, characters are turned into machine readable format. A tag including encoded is attached to an item that needs to be detected. A scanner assists in reading encoded data and converts them to an electrical signal. Finally, a data decoder converts the signal to digital data and returns to the initial alphanumerical characters. RFID uses radio waves to implement AIDC technique. An RFID tag holds an antenna and integrated circuit to send data to RFID reader. Reader transforms the radio waves to a usable form and the data gained from tags are transmitted with the aid of a communication interface to a particular host system so that data can be recorded in the database and examined if necessary.

Figure 3. Three Steps of AIDC Technique

In the system, notifications are delivered to the tutors after scanning the Quick Response (QR) code or RFID tag stored in the food item. This tag includes essential information about the food like its nutritive elements: sugar, saturated fat, calories and sodium. The paper followed two implementation approaches; one with using RFID and another one is QR code. RFID tag is readable with a mobile device that is NFC-enabled. But RFID has a disadvantage; while utilizing RFID tags, maximum users need to alternate their handheld devices to an NFC-enabled version. That is why Bluetooth can be a great alternative. On the other hand, QR codes are easily decoded by a scanner and a camera of a mobile phone [9]. The downside of the research is- it is based on a really old technology. For example, it works with HTC Magic (Operating System: Android 1.6 Donut) and Nokia N-95 (Operating System: Symbian OS v9.2, S60 3rd edition). In addition, iOS platform can also be tested for such mHealth technology.

At this time, plenty of eHealth services are accessible to general people in order to maintain proper health condition. Mobile applications can be very constructive when it comes to managing healthy lifestyles. The research works reported in [10] [11] seem to have positive outcomes concerning the subject. The best thing about mobile applications is they are developed in a personalized manner for an individual. For example, data are being added by a user everyday and the user can monitor the activities anytime in future.

It has been proved that games are useful in case of a child's memory and critical thinking skills. They can also have positive impacts such as enhancing analytical thinking skills, conceptual learning capabilities, problem solving skills, creativity, linguistic improvement and so on. Designing a game or mobile application for a child consists of several important components like prototyping, design, production, testing, fixing bugs, maintenance etc. shown in figure 4.

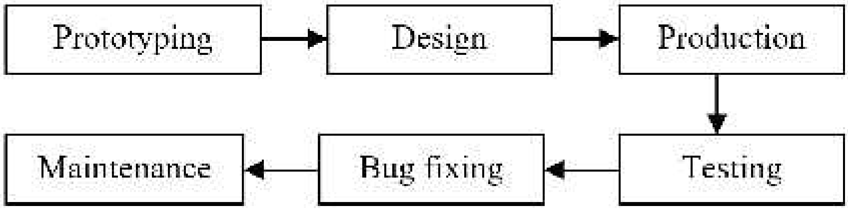


Figure 4.Development process of mobile applications and games

In [12], the authors proposed a mobile application as an informative tool to prevent increased obesity rates. The application is based on IoT and it includes remote capturing, tracking food intake and continuous monitoring system for children. The application creates awareness among children regarding healthcare and also explores nutrition benefits. The mobile application is certainly great for preventing obesity rates, but it could suggest a methodology for children suffering from malnutrition or undernutrition as well. A mHealth platform [13] using transfer learning and IoT has been proposed to improve children's health consciousness and behavior. In the paper, a balanced and categorized food chart according to children's meal time and nutrition has been prepared. The task is done using transfer learning and You only look once (YOLO) algorithm. Transfer learning has two approaches; a. Pre-trained model approach and b. Develop model approach where the former one is chosen for the platform. In pre-trained approach, a source model is chosen at first and then the model is reused as the initial point for other model on another task. Tuning might be necessary to modify data of input-output pair. In YOLO algorithm, a neural network is applied to a full image. After that, the image is partitioned into few regions. Thus, bounding boxes for each region are found. Height, weight, (x, y) coordinates and confidence scores are received from the bounding box. To find out confidence scores for the boxes, the following equation is used;

$$Pr(Classi|Object) * Pr(Object) * IOU = Pr(Classi) * IOU \quad (1)$$

where, IOU = Intersection over union.

In the module, two subsystems are present for child user and health instructor. They are both connected for communication purpose. The platform holds a practical value for children as whenever a child user is capturing a photo of the meal he is going to take through the camera, he is receiving the nutrition value through the module. Although the platform is beneficial for children, it lacks suggestion on how to avoid junk foods. *Health Attack* [14] is a mobile application based on iPhone platform targeting afro-american children who are aged between 7 and 11 years old focused on informing them about nutritional value of diverse food items. There is a total of 7 different nutrients; water, vitamins, fats, proteins, minerals, carbohydrates and fiber. Children should gain the idea of them when they are developing their mental and physical abilities. The downside of the app is it is static in nature and more functionalities regarding practical approaches should be added.

2.1.2. Physical Exercise & Calculators

The works presented in [15] [16] inspired children and teenagers to perform physical activities more often. Mobile games have been created with the purpose of fulfilling the requirements. Women can utilize the menstrual cycle calculator apps like Maya Apa [17], which is a popular app in Bangladesh. The app allows users to ask questions anonymously so that they never feel uncomfortable to know about any sensitive health information. But a big disadvantage is, as the user can enter the system without providing any personal information, many unpleasant posts appear in the app anonymously written by ill-natured users. There are many apps online which present integrated Body Mass Index (BMI) and Total Daily Energy Expenditure (TDEE) calculators. To determine one's BMI and TDEE, following equations are followed:

$$BMI = \frac{Weight\ (lbs)}{Height(in)^2} * 703 \quad (2)$$

Or, $BMI = \frac{Weight(kg)}{Height(m)*Height(m)}$

TDEE formula applied for women:

1. Measure Basal Metabolic Rate (BMR) using;

$$BMR = (height\ in\ cm * 6.25) + (weight\ in\ kg * 9.99) - (age * 4.92) - 161 \quad (3)$$

2. To calculate TDEE;

$$TDEE = BMR * activity\ level \quad (4)$$

For sedentary lifestyle; $TDEE = BMR * 1.1 \quad (5)$

Lightly active lifestyle; $TDEE = BMR * 1.275 \quad (6)$

Moderately active lifestyle; $TDEE = BMR * 1.35 \quad (7)$

Very active lifestyle; $TDEE = BMR * 1.525 \quad (8)$

After taking inputs from user, the calculators show the results and recommend a standard score and suggest how to achieve the goal. The shortcoming of these calculators is that a new kind of calculator is available now which is more accurate and simpler than the previous ones. This calculator is based on relative fat mass index (RFM). In this case, one only requires a tape measure rather than scales. To determine RFM of a woman;

$$76 - \left(20 * \frac{height}{waist\ circumference}\right) = RFM(women) \quad (9)$$

On the other hand, trending smartphone applications consist of nutrition facts for children that are available on Apple App Store and Google Play Store. These facts are taught in an entertaining manner. Recently, edutainment has taken the concept of education to next level. These applications strongly motivate a child's access to the understanding of learning, health and nutrition through games.

2.1.3. Pregnancy & Baby Development

For mothers, apps pertaining to pregnancy are available containing tips of morning sickness, prenatal vitamins and workout tips to avoid pregnancy weight gain and so on. They also proffer features like pregnancy calendar, baby growth detector, baby due date calculator and ovulation calculator. An example is *Pregnancy Tracker & Baby App* [18]. Pregnancy is a long process where a mother must be careful about her health and should take nutrient-rich foods. Such apps are becoming very popular among pregnant mothers because of their wide-ranging set of features. The biggest drawback of these apps is it does not include any real time services or features. A pregnant woman might need an urgent support during the months of pregnancy. But such features are still not available yet.

We have conducted an exclusive survey with 23 women who uses mobile applications for various healthcare services. Among these 23 users, 16 women use Android platform, 5 women use iOS platform and 2 of them uses Windows platform. In addition, the types of health apps and the problems they face using such apps are demonstrated in figure 5 and 6 respectively. Most of the women uses multiple apps to maintain their health. The survey illustrates 14 users use health apps for exercise and fitness, 9 of them use health apps for counting calories, 12 of them use such apps for maintaining diet and weight, 10 of the users keeps track of their period or menstrual cycle, 3 of them use them for pregnancy purpose, 8 of them check blood pressure with the aid of such apps, 4 of them use apps for sleep and meditation purpose and 5 of them also use the health apps for other tasks.

In response to the difficulties, 9 of the women users stated that such health apps often contain bad user interface, 3 of them mentioned minor bug issues, 8 of them said they hold no practical values, 12 of them pointed out that they only have static features and 4 of them stated they need to be updated more often.

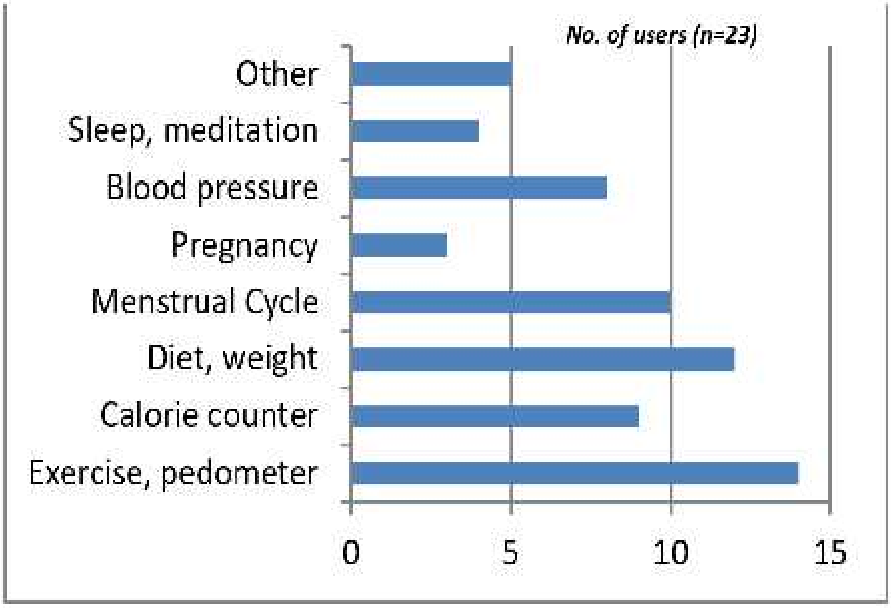


Figure 5.Health apps and their users (women)

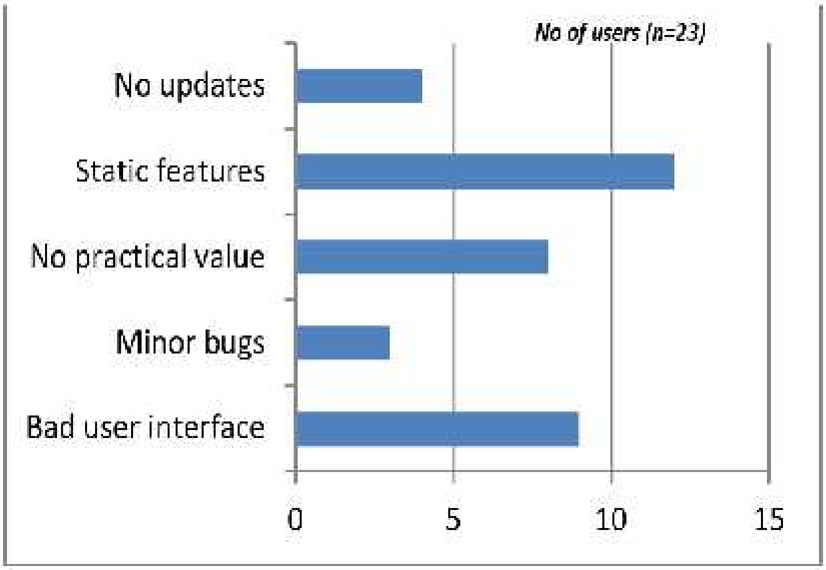


Figure 6.Health apps and their problems (women)

We believe major changes can be applied for the welfare of these health apps and their users. According to our survey, there are lots of opportunities in health apps, as well as challenges to overcome.

2.2. Interactive Systems

An IoT enabled interactive totem [19] was developed in a paediatric ward of a hospital to improvise hospitalization experience for children. With the help of the totem, children can get friendly with the unknown environment surrounded by doctors, nurses and tons of medical equipment. The totem involves the feature of self-ordering meals where a child is recognized by the system and his favorite meal is suggested. It contains webcam, stereo speakers, 26’’ LCD touch screen and internet. The system mentioned in the paper lacks important features like emergency support, food detection and nutrition facts. In addition, what happens after a child patient is discharged is not discussed.

Usually when applying recognition feature in an IoT system, it is also really important to activate triggers for each actions. For such purposes, various tools are available now. Combining ThingSpeak and If this, then that (IFTTT) platforms can result in a number of useful services for children and women. There are also other software tools that have potential to generate improved solutions for this target group. Table 1 discusses the collection of software tools that can be used for this purpose.

Table 1. Software tools for IoT environment

Software tool	**Function/Purpose**
Windows IoT	Operating system
AllJoyn	Open source software framework
Node-Red	Flow-Chart based development tool
openHab	Open source automation software for home
MindSphere	"IoT operating system"/open cloud platform
IOTivity	Open source software framework
Android Things	Android-based embedded operating system

An intelligent children healthcare system is developed in [20] where the system tracks records, monitors progress and promotes nutrition through mHealth and edutainment features. After analyzing the data, the results were recorded with big data analytics tool called Hadoop. Finally, they were compared with Logistic Regression and Naïve Bayes algorithms where Logistic Regression performed better than the other. The system only detects if a child is in good health condition or not. It lacks recommendation or suggestion feature based on a child's current health state. In [21], a smart system was designed by the researchers with the help of IoT and MapReduce framework of Hadoop to detect anxiety disorders, disruptive behavior disorders and Attention-deficit/hyperactivity disorder (ADHD). MapReduce framework has two classes; a. Mapper and b. Reducer. Mapper class reads data blocks and create key-value pairs as outputs. The output is taken as an input to reducer class and it combines key-value pairs to smaller group of key-value pairs. To classify the child behavior, C4.5 decision tree algorithm has been used. In [22], the authors proposed an intelligent system with a smart nutrition card that holds a child's daily activity, calorie intake, emotional state and performance after collecting the data through smartphone sensors, recommendation methods and artificial intelligence. The interactive system collects the child's attendance rate, emotional and physical state, BMI & TDEE updates, activities, nutrition info and real time health issues. The negative aspect of the research is that the smart nutrition card needs to be practically implemented with actual users in a hospital environment.

2.3. Mother & Child Healthcare

To reduce infant and maternal mortality rate, Mother and Child Tracking System (MCTS) was developed in India [23]. The system engages caregivers with patients through web-based application and SMS. A kind of similar prototype was also done in Nigeria [24]. In Indonesia, a multiplatform system [25] was designed for mother and children care to detect diseases early through remote monitoring service. The system includes multiple sensors, mobile application and a portal. All of it is maintained by a cloud so that caregivers and patients can obtain easy access. The system offers positive healthcare experience in terms of mother and children. The biggest drawback of such multiplatform systems is security vulnerability due to cloud management system depicted in figure 7. A lot of devices are using cloud services for various purposes and it is indisputably unsafe because of hackers. These smart devices and applications need strong cryptographic support. Otherwise, the private information can be breached unintentionally.

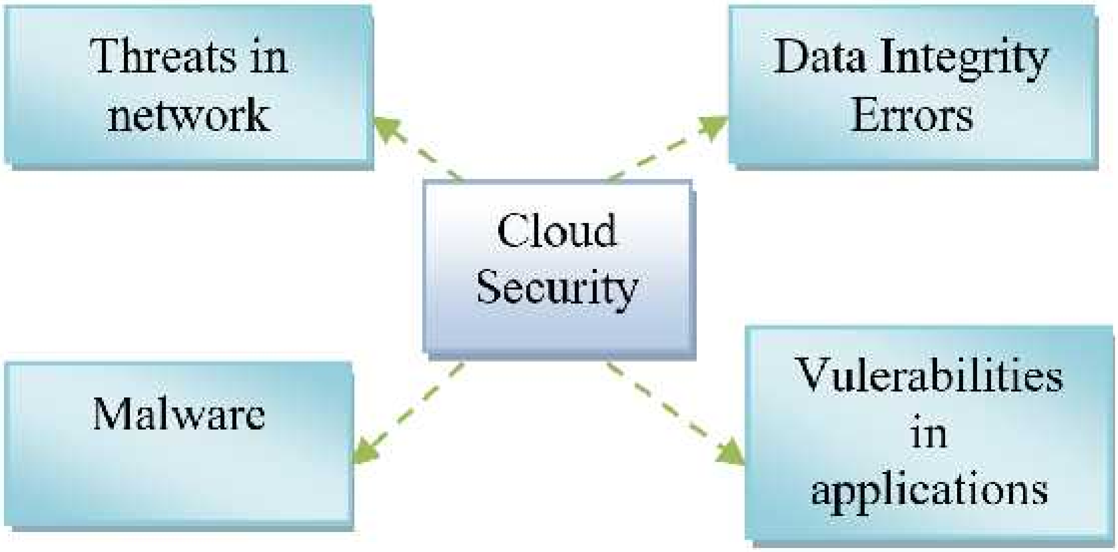


Figure 7. Cloud Security Issues

2.4. Baby/Children Tracking Monitors & Wearables

A real-time location infant monitoring system has been developed in [26] with wristband active RFID tags. Using this, the newborn’s important data like current location and body temperature can be tracked without any difficulty. The baby’s location is recognized by decision tree classifiers. Although the infant monitoring system is quite constructive, it may contain potential health hazards especially for children. A safety system for school going children [27] on a bus has been designed so that the parents can make sure that their children are on the right track. This research could be quite simpler if they could apply GPS for the bus riders into their system. A child safety wearable device [28] using Short Message Service (SMS) and Global System for Mobile (GSM) communication is created for parents to know about their child’s current location and temperature. The system utilizes Arduino Uno, a visual distress signal known as SOS light, temperature sensor, Ultraviolet (UV) sensor, GPS sensor, alarm buzzer, GSM shield and base shield. In [29], a wearable device called *Raksh* was built. The aim of the paper was to fight against pneumonia from which children might suffer a lot. Special IoT based wearable devices are presented in [30] [31] [32] [33] targeting women and girl children for ensuring their safety in adverse and dangerous situations. These papers reflect how girls can be protected in unsafe situations with the help of smart devices. But unfortunately, smart gadgets like wearables hold major privacy concerns and unwanted health issues due to high radiation levels.

Table 2 discusses the collection of software tools that can be used for this purpose.

Table 2. Comparative analysis among used methods or platforms for women & children

Used methods/platforms	Merits	Demerits
RFID	Real-time tracking, efficient, reduces errors	Sometimes alternating to NFC enabled versions is mandatory
Games (children)	Holds constructive and practical values, a great source of edutainment	Traditional approach, common ideas
Mobile Applications (children)	Simple and easy	Static features, raw data only
Calculators (women)	Helps to keep track	Accuracy is questionable
Interactive Systems	Face recognition, personalized choices	Lacks emergency support
Mobile Applications related to pregnancy and baby development	Important tips, calendars	No real-time services available, no emergency support
Multiplatform Systems	Better healthcare experience, getting connected to caregivers	Vulnerable cloud security issues

In addition, we have applied Analytic Hierarchy Process (AHP) in order to find out the desired user interface for such systems. Common user interface types include form, menu, graphical user interface (GUI), command line and natural language. We have considered three major options: form based, menu driven and GUI in case of comfort in usability.

A pair-wise comparison is a must to select preferences for criteria and choices. For each matrix that includes pair-wise elements,

$$\begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix}$$

In the table shown below (Table 3), intensity judgment is done with pairwise comparison matrix. A = Form based, B = GUI and C = menu driven where GUI is more preferable than form based and also more preferable than menu driven approach. So,

Table 3. Comparison of matrix of Form based, GUI and menu driven approach

Choosing UI	A	B	C
A	1	1/3	2
B	3	1	5
C	½	1/5	1

Now, the values are summed up in each of the column of pair-wise matrix in (10),

$$C_{ij} = \sum_{i=1}^{n} C_{ij} \quad (10)$$

Thus, to find out the normalized matrix, addition of columns must be done. The process is shown in table 4 and 5.

Table 4. Addition of columns

Choosing UI	**A**	**B**	**C**
A	1.000	0.333	2.000
B	3.000	1.000	5.000
C	0.500	0.200	1.000
Total	4.500	1.533	8.000

Table 5. Normalized matrix & sum of rows

Choosing UI	**A**	**B**	**C**	**Total**
A	0.22	0.217	0.25	0.687
B	0.67	0.652	0.625	1.947
C	0.111	0.130	0.125	0.366

To generate a normalized pair-wise matrix, each element should be divided by its column in the matrix like shown in equation (11).

$$X_{ij} = \frac{C_{ij}}{\sum_{i=1}^{n} C_{ij}} \begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \end{bmatrix} \quad (11)$$

To create the weighted matrix, equation (12) is used where n = no. of criteria,

$$W_{ij} = \frac{\sum_{j=1}^{n} X_{ij}}{n} \begin{bmatrix} W_{11} \\ W_{12} \\ W_{13} \end{bmatrix} \quad (12)$$

Normalizing sum of rows produce,

$$\begin{bmatrix} 0.687/3 \\ 1.947/3 \\ 0.366/3 \end{bmatrix} = \begin{bmatrix} 0.229 \\ 0.649 \\ 0.122 \end{bmatrix} \quad (13)$$

According to the priority matrix, GUI has the maximum priority. Finally we need to measure the consistency ratio (CR) which should be <=10% in accordance with Saaty, who developed the original method. To get consistency ratio, we have to calculate consistency index (CI) at first which is based on equation (14),

$$\begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} * \begin{bmatrix} W_{11} \\ W_{21} \\ W_{31} \end{bmatrix} = \begin{bmatrix} Cv_{11} \\ Cv_{21} \\ Cv_{31} \end{bmatrix} \quad (14)$$

After that, division of weighted sum vector is executed with criterion weight.

$$Cv_{11} = \frac{1}{W_{11}} [C_{11}W_{11} + C_{12}W_{21} \;\; C_{13}W_{21}] \quad (15)$$

$$Cv_{21} = \frac{1}{W_{21}} [C_{21}W_{11} + C_{22}W_{21} \;\; C_{23}W_{31}] \quad (16)$$

$$Cv_{31} = \frac{1}{W_{31}} [C_{31}W_{11} + C_{32}W_{21} \;\; C_{33}W_{31}] \quad (17)$$

To determine λ, average value of consistency vector is measured. CI is used to perceive deviation.

$$\lambda = \sum_{i=1}^{n} Cv_{ij} \quad (18)$$

$$CI = \frac{\lambda - n}{n - 1} \quad (19)$$

Therefore, we follow the aforementioned steps to get our desired values,

$$\begin{bmatrix} 1 & 1/3 & 2 \\ 3 & 1 & 5 \\ 1/2 & 1/5 & 1 \end{bmatrix} * \begin{bmatrix} 0.229 \\ 0.649 \\ 0.122 \end{bmatrix} = \begin{bmatrix} 0.69 \\ 1.95 \\ 0.37 \end{bmatrix}$$

$$\begin{bmatrix} 0.69/0.229 \\ 1.95/0.649 \\ 0.37/0.122 \end{bmatrix} = \begin{bmatrix} 3.01 \\ 3.00 \\ 3.03 \end{bmatrix}$$

$$\lambda = \frac{3.01 + 3.00 + 3.03}{3} = 3.01$$

$$CI = \frac{3.01 - 3}{3 - 1} = 0.005$$

Finally, to obtain CR, equation (20) is used;

$$CR = \frac{Consistency\,Index(CI)}{Random\,Index\,(RI)} \quad (20)$$

$CR = \frac{0.005}{0.90} = 0.0055$ (acceptable)

Therefore, our calculation proves that GUI is the most preferred option as the user interface than form based or menu driven interfaces. To make solutions better, one should put an emphasis on generating creative, simple and comfortable GUI for children and women users.

3. Current Trends & Challenges

The annual growth rate of internet of medical things is estimated to reach around $72 billion by the year 2021. IoT is being adopted by the healthcare community for its extensive features and potentials. Implantable devices, wearables and monitors will carry on sending real time data to hospitals. For mothers, a smart toy called ‘Teddy the Guardian’ can be very helpful as it tests baby’s temperature and heart rate. Not only that, the smart stuffed toy checks the oxygen saturation level when it receives a hug from the baby. Care-plan-specific mobile applications are brought into play to decrease readmission rates. The concept of smart bag (with microcontroller ATmega16) and smart shirt (enabled with NFC chip) for children are on the rise. Virtual assistants and mobile health applications have become popular already. Healthcare in IoT will sooner or later result in better patient experience, improved outcomes, superior disease management and lower treatment costs.

Devices that are connected to internet can come in a variety of appearances. Data can be collected from temperature monitors, wearables, healthcare apps or interactive systems. Most of these calculations need a follow-up communication with a doctor or a professional in the area. It generates an opportunity for smart devices to distribute precious data. IoT is surely making a difference in the industry of healthcare. Data management in hospitals has been improved in an overwhelming manner caused by IoT devices like monitors and scanners. Diseases like diabetes are being detected by such digital healthcare services. A transformation is taking place in healthcare industry and business organizations should come forward to invest in latest technologies with the aim to enhance medical care. They are already grabbing the opportunity and creating personalized solutions for individuals and particular entities. Smart hospitals are allowing RFID enabled pharmacy inventory system, automated drug delivery, smart patient flow, real-time location data sharing and smart consoles with the intention of treating patients more precisely. Researchers and developers across the earth are exploring vivid solutions that complement previous services by activating the possibilities of IoT.

IoT devices are not restricted to a small network. The devices receive and send personal health information (PHI). Such information must be kept protected. IoT health monitoring devices are directly connected to the internet. As a result, privacy is the most vulnerable security issue of smart healthcare services. Once upon a time, we only considered the privacy issues of our personal computers. Then the smartphones joined the market. And finally we have the next big thing here, which happens to connect all the gadgets in a single platform – wearables, home appliances, car, phone, remote control and so on.

4. Conclusion

Numerous applications and services have been developed yet for the welfare of women and children. Active participation of this target group is mandatory to ensure the success of the eHealth, mHealth and IoT services built for them. In the paper, we discussed about their possibilities and benefits, as well as the challenges to overcome. The paper demonstrates a new perspective to make use of technology for women and children healthcare. IoT has promising economic and technological prospect in spite of its security issues. We believe that IoT will change the traditional healthcare system eventually for the greater good, especially for women and children.

References

[1] Malasinghe, L.P., Ramzan, N. & Dahal, K, “Remote patient monitoring: a comprehensive study”, *J Ambient Intell Human Comput*, pp. 1–20, 2017. [Online]. Available: https://doi.org/10.1007/s12652-017-0555-8

[2] World Health Organization, “WHO recommendations on antenatal care for a positive pregnancy experience”, World Health Organization, 2016.

[3] Stoddard, Jeffrey J.; St. Peter, Robert F; and Newacheck, Paul W., "Health Insurance Status and Ambulatory Care for Children", *New England Journal of Medicine*, Vol. 330, pp. 1421–1425, 1994.

[4] Wang, MH., Lee, CS., Hsieh, KL. et al., “Ontology-based multi-agents for intelligent healthcare applications”, *J Ambient Intell Human Computing*, Vol. 1, No. 2, pp 111–131, 2010.

[5] Verma, S.K. Sood, Cloud-centric IoT based disease diagnosis healthcare framework, *J. Parallel Distrib. Comput. (*2017), https://doi.org/10.1016/j.jpdc.2017.11.018.

[6] S.M.R. Islam, D. Kwak, M.D.H. Kabir, M. Hossain, and K. Kwak, "The internet of things for health care: a comprehensive survey," *IEEE Access* 3, 2015, pp. 678-708.

[7] Business Insider, 2018. Available Online: http://www.businessinsider.com/iot-solutions-expected-to-save-us-healthcare-system-billions-2015-6.

[8] A. Siddiqui, O. Koch, A. Rabie, and U. Handmann, "Personalized & Adaptable mHealth Architecture," 2014

[9] M. Vazquez-Briseno, C. Navarro-Cota, J. I. Nieto-Hipolito, E. Jimenez-Garcia and J. D. Sanchez-Lopez, "A proposal for using the internet of things concept to increase children's health awareness," CONIELECOMP 2012, 22nd International Conference on Electrical Communications and Computers, Cholula, Puebla, pp. 168-172, 2012.

[10] Atienza, A.A., King, A.C., Oliveira, B.M., Ahn, D.K. and Gardner, C.D. "Using handheld computer technologies to improve dietary intake". *American Journal of Preventive Medicine*, vol. 34 no. 6, pp. 514–518, June 2008.

[11] Gasser, R., Brodbeck, D., Degen, M., Luthiger, J., Wyss, R. and Reichlin, S. "Persuasiveness of a mobile lifestyle coaching application using social facilitation". *Lecture Notes in Computer Science*, vol. 3962, pp. 27–38, 2006.

[12] M. Alloghani, A. Hussain, D. Al-Jumeily, P. Fergus, O. Abuelma'atti and H. Hamden, "A mobile health monitoring application for obesity management and control using the internet-of-things," 2017 Sixth International Conference on Digital Information Processing and Communications (ICDIPC), Beirut, 2016, pp. 19-24. doi: 10.1109/ICDIPC.2016.7470785

[13] N. Nigar, M.N. uddin, "A mHealth Platform using Transfer Learning and Internet of Things to Improve Children's Health Consciousness and Behaviour", unpublished.

[14] Hill,D., Blunt, J., Pugh, T.,Monk, M., Kim, J., Winchester, W., McCrickard, D., Estabrooks, P., and Doswell, F. "Mobile technologies for promoting health and wellness among african american youth". *6th international conference on Universal access in human-computer interaction: context diversity - Volume Part III (UAHCI'11)*, Constantine Stephanidis (Ed.), Vol. Part III. Springer-Verlag, Berlin, Heidelberg, 36-45.

[15] Arteaga, S.M., Kudeki, M. and Woodworth, A. "Combating obesity trends in teenagers through persuasive mobile technology". *SIGACCESS Access.* Comput. 94 (June 2009), 17-25.

[16] Cheok, A.D., Goh, K.H., Liu W., Farbiz F., Fong, S.W., Teo, S.L, Li, Y., and Yang X., "Human Pacman: a mobile, wide-area enterteinment system based on physical, social and ubiquitous computing. *Personal and Ubiquitous Computing*, vol. 8, pp. 71-81, 2004.

[17] Maya Apa, 2018. Available on Google Play Store: https://play.google.com/store/apps/details?id=com.maya.mayaapaapp&hl=en

[18] Pregnancy Tracker & Baby App, 2018. Available on Apple App Store: https://itunes.apple.com/us/app/pregnancy-tracker-baby-app/id386022579?mt=8

[19] S. Vicini, S. Bellini, A. Rosi and A. Sanna, "An internet of things enabled interactive totem for children in a living lab setting," 18th International ICE Conference on Engineering, Technology and Innovation, Munich, pp. 1-10, 2012.

[20] N. Nigar, M.N. uddin, "An Intelligent Children Healthcare System in the Context of Internet of Things", 2nd International Conference on Innovative Research in Science, Technology & Management, NUS, Singapore, 2018, in press.

[21] P. K. Binu, V. Akhil and V. Mohan, "Smart and secure IOT based child behaviour and health monitoring system using hadoop," *2017 International Conference on Advances in Computing, Communications and Informatics(ICACCI)*,Udupi,2017,pp.418-423. doi: 10.1109/ICACCI.2017.8125876

[22] N. Nigar, M.N. uddin, "An Internet of Things Enabled Intelligent System and Smart Nutrition Card to Enhance Children's Health Consciousness", IEEE-International Conference on New Trends in Engineering & Technology (ICNTET), GRT College, Chennai, Tamil Nadu, India, Singapore, 2018, in press.

[23] Roadmap for Telemedicine, Innovative Working Group (IWG), Asia, Technical Report, 2014, pp. 1-48.

[24] Geidam, M. Ajagana, R. Prasad, and I.A. Bello, "Child and maternal health care using telemedicine: a case study of Yobe State, Nigeria," International Journal of Computer Engineering and Applications 7, No. 1, 2014.

[25] T. M. Kadarina and R. Priambodo, "Preliminary design of Internet of Things (IoT) application for supporting mother and child health program in Indonesia," *2017 International Conference on Broadband Communication, Wireless Sensors and Powering (BCWSP)*, Jakarta, 2017,pp.1-6. doi: 10.1109/BCWSP.2017.8272576

[26] Shou-Hsiung Cheng, Jui-Chen Huang and Chun-Jung Lin, "A real-time location and infant monitoring system based on active RFID," *2012 International Conference on Machine Learning and Cybernetics*, Xian, 2012,pp.1844-1849. doi: 10.1109/ICMLC.2012.6359656

[27] L. D'Errico, F. Franchi, F. Graziosi, C. Rinaldi and F. Tarquini, "Design and implementation of a children safety system based on IoT technologies," *2017 2nd International Multidisciplinary Conference on Computer and Energy Science (SpliTech)*, Split, 2017, pp. 1-6.

[28] A. Moodbidri and H. Shahnasser, "Child safety wearable device," *2017 International Conference on Information Networking (ICOIN)*, Da Nang,2017,pp.438-444. doi: 10.1109/ICOIN.2017.7899531

[29] K. Mala, B. M. Kumar, R. Vignesh and K. M. Kumar, "A wearable diagnostic device to combat children's pneumonia," *2016 IEEE Global Humanitarian Technology Conference (GHTC)*, Seattle, WA, 2016, pp. 654-659. doi: 10.1109/GHTC.2016.7857348

[30] A. Jatti, M. Kannan, R. M. Alisha, P. Vijayalakshmi and S. Sinha, "Design and development of an IOT based wearable device for the safety and security of women and girl children," *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Bangalore, 2016, pp. 1108-1112. doi: 10.1109/RTEICT.2016.7808003

[31] G. C. Harikiran, K. Menasinkai and S. Shirol, "Smart security solution for women based on Internet Of Things(IOT)," *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*,Chennai,2016,pp.3551-3554.
doi: 10.1109/ICEEOT.2016.7755365

[32] A. Helen, M. F. Fathila, R. Rijwana and Kalaiselvi V. K. G., "A smart watch for women security based on iot concept 'watch me'," *2017 2nd International Conference on Computing and Communications Technologies(ICCCT)*,Chennai,2017,pp.190-194.
doi: 10.1109/ICCCT2.2017.7972266

[33] N. Viswanath, N. V. Pakyala and G. Muneeswari, "Smart foot device for women safety," *2016 IEEE Region 10 Symposium (TENSYMP)*, Bali, 2016,pp.130-134.doi: 10.1109/TENCONSpring.2016.7519391

[34] J. Ahamed and A. V. Rajan, "Internet of Things (IoT): Application systems and security vulnerabilities," *2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA)*, Ras Al Khaimah,2016,pp.1-5. doi: 10.1109/ICEDSA.2016.7818534

Author's Biography

Nishargo Nigar holds a Bachelor of Science in Computer Science & Engineering from East Delta University, Chittagong, Bangladesh. She is an entrepreneur and she is running two IT startups currently. She is a student member of IEEE. Her research interests are Internet of Things, mHealth and machine learning.

How to Cite

Nigar, Nishargo (2019). A Study on Internet of Things in Women and Children Healthcare. *International Journal of Machine Learning and Networked Collaborative Engineering*, 3(01) pp 1-15.

doi : https://doi.org/10.30991/IJMLNCE.2019v03i01.001

Vol. 03, No. 1,(2019), 16-34, ***ISSN:2581-3242***

International Journal of Machine Learning and Networked Collaborative Engineering

Journal Homepage: http://www.mlnce.net/home/index.html

DOI : https://doi.org/10.30991/IJMLNCE.2019v03i01.002

Ultimate Indoor Navigation: A Low Cost Indoor Positioning and Intelligent Path Finding

[a]Alex Gunagwera[*], [b]Farzad Kiani

Department of Computer Engineering, Engineering and Natural Sciences Faculty, Istanbul Sabahattin Zaim University, 34303, Kucukcekmece, Istanbul/Turkey

[a]alexantosh@gmail.com, https://orcid.org/0000-0002-0143-3743

[b]Farzad.kiani@izu.edu.tr, https://orcid.org/0000-0002-0354-9344

Abstract

Despite the rapid improvement in mobile devices, overall gradual growth in the ubiquitous computing field, the wide applicability, more usefulness of location based services in general and indoor navigation. The Global Positioning System (GPS) has undergone tremendous improvement since the 1900s and it, indeed is considered one of the most successful navigation systems known to date. However, it is still inefficient for sufficiently accurate positioning in both indoor environments and environments with many tall buildings such as skyscrapers since such buildings block or interfere with its signal transmissions. In particular, building a sufficiently accurate, efficient and relatively cheap indoor navigation system in a GPS-free environment is still a challenging task with a lot of tradeoffs and constraints to put into consideration. In this paper, a simple yet robust, low-cost, context-aware user-interactive, user-friendly hybrid of fingerprinting and dead reckoning indoor navigation system suitable for both the visually and the physically disabled as well that takes advantage of the results yielded by sensor fusion is proposed. The presented system is also designed to allow for efficient evacuation of users in cases of emergences. The prototype is made majorly of the following parts; user tracking, optimal, context-aware and dynamic route calculation and planning and dynamic route representation with an upper bound of 2m and an average of 0.8-1.3m accuracy. All that is required from the user is a smart phone without installation of extra hardware.

Keywords

Indoor Navigation,

Tracking Algorithm,

Fingerprinting,

Visually Impaired,

Dead Reckoning

1. Introduction

The Global Positioning System (GPS) has undergone tremendous improvement since the 1900s and it, indeed is considered one of the most successful navigation systems known to date. However, it is still inefficient for sufficiently accurate positioning in both indoor environments and environments with many tall buildings such as skyscrapers since such buildings block or interfere with its signal transmissions. It is a

[*] Corresponding Author

Alex Gunagwera

Email: alexantosh@gmail.com

satellite navigation and navigation system with a network of at least 24 satellites. These satellites were commissioned by the US Department of Defense and placed in Earth's orbit. The poles of the earth's magnetic field will gradually give way to GPS receivers; GPS is a system that is navigated by a group of satellites. The satellites that circle around Earth in their orbits; these satellites are in contact with special stations on the ground, and their position in space is always clear. Your GPS receiver, by communicating with a number of these satellites, sets the distance to them, and then your exact position is obtained on the ground. The GPS Global Positioning System consists of 24 satellites that are located at an altitude of 20,000 km from the ground, and in 6 circuits, each of which has four satellites, and has a circle of 55 degrees and a 12-hour period. Each GPS satellite sends two waves with two frequencies in the electromagnetic waves (L1, L2). The L1 wave is powered by a frequency of 1575 MHz and a L2 wave with a frequency of 1227 MHz.

The ubiquity of mobile devices (such as cell phones) has led to the introduction of Location Based Services (LBS), or Location-Aware Services. LBS aim at providing information/services relevant to the current location and context of a mobile user. One of the first several LBS applications, named Active Badge Location System, was introduced in R. Want et al. [1]. This system employed infrared technology for tracking a user's current location and used this location to forward phone calls to a telephone close to the user. Since then, many researchers have studied this topic and as such many LB applications have emerged over time because of the increased market of both outdoor and indoor location based services and applications.

Recent technological advance such as the gradual maturation of ubiquitous computing M. Weiser series [2], or pervasive computing, and the evolution of mobile devices (such as PDAs, cell phones, etc.) and wireless communication (3G, Wireless LAN, Wireless Sensor Network, IoT [37], etc.) has further increased the pace of progress A. Butz et al [3]. They also make an overview about map-based mobile guides using the dimensions of Positioning (either GPS, WiFi, UMTS, or other), Situational factors (user or context-related), Adaptation capabilities, interface/use interaction (multi-model or others), Use of maps (2D vector, 2D bitmap, 3D model etc.), and Architecture (client-server, interacting, multi-blackboard or multi-agent system). These dimensions are roughly defined, and some of these dimensions need to be further subdivided. Raper et al. [4] developed a much more complete classification which used the axes of Application (tourism, recreation, transport, and museum), Positioning, Architecture, Presentation, Context relevance, Delivery (pull or push), Use case, and Adaptively (resource adapted, resource adaptive, resource adapting), and then made a research about LBS applications in the published literature. However, these researches are mainly for outdoor applications cases in indoor cases have various requirements that they can be solved by different positioning algorithms and technologies. As a result, more detailed dimensions on positioning, such as signal (infrared, ultrasonic, radio signals, etc.) and signal metrics (Cell of Origin, Time of Arrival, Time Difference of Arrival, Angle of Arrival, to mention but a few), are needed to evaluate the various positioning technologies. Because context-aware is very important for LBS systems, there should be some dimensions that assess the context awareness of these indoor applications due to the fact that the correct user is not easy to monitor [5].

Still the serious participation of well-known firms' sectors such as Bing maps and Google maps in [6] and [7] as well as the steady show of interest of companies such as [8] and many more, further stresses the importance and potential of indoor navigation. However, even for such powerful companies accurately mapping all buildings and provided precise navigation is still very hard, if not impossible due to the continuous growth of the construction and business industries. It is simply hard to keep up. As the range of LBS applications is vast, it is practically impossible to introduce all of them here due to brevity. Mobile navigation systems, which aim at providing wayfinding services and tracking to the user, are the most important applications of LBS. However, in indoor navigation and localization, accurately pinpointing the location of a user and correctly and optimally guiding the user to the desired destination is pretty challenging. Till to date, available techniques and attempts at indoor mapping and/or localization can be distinctly categorized in two major groups; those that employ fingerprinting [9] and those that do not [10, 11, 12]. Humans to generating discriminative signatures for the access points (APs) face most fingerprinting based techniques with the problems ranging from susceptibility to intervention. Also, the Received Signal Strength Indicator (RSSI) varies over time due to noise, multipath, reflection off and absorption by surfaces and other objects.

The fingerprinting [9, 34] techniques include Wi-Fi RSSI and FM broadcast fingerprints to help map the indoor environment in question. Indoor locations are attached to signal strength values thereby mapping the environment. Techniques that do not involve finger printing generally the dead reckoning technique to calculate the user's current position in reference to pre-known values. There is also the issue of presenting the route to the user in the most efficient way possible. So, the users do not understand signal strengths in decibels (say -45dB) or coordinates in form of (x, y) or (x, y, z) and instead objective descriptions such as "to the left of the canteen right in front you" or suitable graphical presentation and so on and so forth are much easier to understand and hence make more sense. All these and many more problems are considered and addressed in our paper.

The rest of this paper is organized as follows. Section 2 presents related works, section 3 introduces the methodology used in this work and the model of the proposed prototype, section 4 shows the field experiments and results performed and discussions concerning the results. Section 5 presents the future works, possible improvements to the proposed system and conclusion.

2. Related Work

In our lives, the importance of machine learning and intelligent systems are increasing day by day. In recent years, many valuable studies have emerged in the literature and in practice such as [35, 36]. Indoor navigation has become a serious and popular field of interest owing to its wide applicability, usefulness and the improvement in ubiquitous computing and the development of mobile devices in general and hand held smart phones in particular. As mentioned above, previous attempts to map indoor environments can be generalized in two main categories; those that employ Wi-Fi-fingerprinting and those that do not. Those that do not use fingerprinting exploit techniques such as crowdsourcing [13], triangulation and trilateration. This can be more broadly categorized in two again those that use wireless and sensor technologies (such as most of those discussed later on) and those that use image recognition and processing [14].

In [15] is proposed an enhanced form of GPS for indoors and urban area usage. However, the accuracy attained was still not impressive as GPS signals were still being lost through absorption or reflection from walls and other objects – the so-called multipath problem. This coupled with other GPS related issues on its application indoors render it not the best of choices. The complex nature and structure of the indoor environment, however, still poses a practical solution to the indoor navigation problem using GPS – even with the Chinese Satellite system [16] cooperating with the recent Galileo Positioning System, despite the potential they present. We bypass GPS related issues by not relying on the GPS system for positioning or user tracking. Instead, we use Wi-Fi fingerprinting coupled with dead reckoning techniques.

In [17] is suggest a Kalman filter based Wi-Fi fingerprinting and dead reckoning indoor navigation system as well. The major differences between their work and ours includes is first, they highlight their usage of a Kalman filter. As pointed out by most studies, a Kalman filter is not suitable for this kind of work; we use a low pass filter. We also use a text based feedback relay approach coupled with audio feedback. This is in order to make the application helpful during emergence situations and emergence evacuation scenarios and furthermore take the visually impaired and blind into consideration.

In [18], they proposed pressure sensitive floors whereas [19] proposed pressure sensitive shoes. These systems may be able to solve the indoor positioning and tracking problem but not only do they not offer a means for navigation, they are also extremely expensive to setup especially for mass deployment, not to mention the necessity to track the weight of each of person in case of pressure sensitive floors. Setting up these requirements for pressure sensitive floors and shoes can barely be categorized as low-cost. Besides, this system can only provide tracking, not navigation in best scenario.

In [13] is suggested an approach that does not require the use of Wi-Fi signals at all. They simply use a crowdsourcing approach based on sensor data and a step detection algorithm to determine the position of the user. This different approach is not suited for emergence situations.

In [20], the steps to deal with varying position-tracking accuracy in mobile augmented reality systems are discussed. They argue that neither the problem nor the solution is limited to only augmented reality problems but to all systems that require relatively accurate position tracking – which our system qualifies to be.

In [21] a time dependent optimal routing model for emergency evacuations in which the route is made basing on the position of the sensors and not the architecture of the structure itself is proposed. In the model, the geometry of the structure in question is ignored and all operations and algorithms developed depending on the available sensors. Hence a change or misread from the sensors could easily lead to significantly flawed navigation and routing service results. As it was designed to handle emergence situations, such potential sources of errors should best be avoided. Furthermore, their system is specifically designed for micro-scale temporary emergence evacuation. Therefore, their simulations included human detection with 5-10 second intervals. Our proposed system is a multi-purpose prototype that provides navigation and optimal path routing no matter the situation.

One of the difficulties faced throughout recent research is the accurate prediction of signal propagation. This can be approached using a technique called Location fingerprinting, that is based on measuring actual signal strengths from surrounding access points [22, 23, 24]. Furthermore, [24] provides a purely Wi-Fi fingerprinting based approach to localization and indoor navigation. They directly perform scans and store the RSSI with the corresponding MAC addresses into the databases without further processing and simply read and compare the obtained RSSI from the user with the pre-stored values. If the RSSI is within a given range (in their case a $+4 \geq RSS \geq -4$), then the user is at a correct location. In addition, they thus let the users know so. Else, they are at a wrong location. Major drawbacks to this approach, however, include; requirement of constant database update due to temporal variation in RSS over time and incase of change of the environment. In addition, sometimes the variation n RSS is much greater or less than the (+4, -4) given interval. Under such circumstances (especially for very large or very small roomed buildings), erroneous navigation is inevitable. We apply dead reckoning using filtered data to overcome and minimize the effect of temporal variation of RSS. There is also the issue of latency caused by such variations and multipath. Since calculations are performed much faster, the dead reckoning branch of the algorithm also minimizes this.

An indoor navigation model that supports optimal length-dependent routing is suggested in [26]. It, however, is limited to PCs, uses a server client model and necessitates extra plugin installation thereby barely qualifying as cost efficient. One needs to have at least a laptop to use the application. This is very inconvenient for navigating indoors especially nowadays that mobile phones provide pretty much all the basic functionality sufficient to facilitate navigation.

The CricketNav project [5] proposed the design and implementation of an indoor mobile navigation system using the cricket infrastructure developed in MIT labs. Their project requires installation of hardware developed in MIT labs. They also developed special Cricket beacons and Cricket listeners to aid their cause.

In [25] is proposed a resource-adaptive mobile navigation system (REAL). Their complete project had three major components; an information booth that had a 3D graphics workstation, an indoor navigation system based on strong infrared signal transmitters planted into ceilings and PDAs used for presentation and finally a head-set laptop combination for outdoor navigation. The routing information is presented depending on the kind of device used. Route optimization is also dependent on the resources available on the device being used.

Cyberguide [26] is among the first systems that employed location aware information to aid tourists. The project was designed to help tourists both indoors and outdoors. It comprised of two major components; an indoor and outdoor component. This project’s indoor component depended on beacons that broadcast unique ID using infrared signals from infrared beacons. On the other hand, the outdoor system used GPS. Both components functioned independently from each other.

In [27] a probabilistic navigation system for pedestrians based on mainly inertial sensors found in a specially made device [28] is presented whereas The NEXUS system also tries to provide a general framework for mobile and location aware computing. The concept of an augmented world is used to keep information necessary for a user’s location. It is the basis model for the virtual information towers that it can connect information objects.

In this paper, we present a user-interactive context-aware hybrid system of dead reckoning and fingerprinting that provides navigation services to users. It also puts into consideration visually impaired and physically incapacitated users by provided audio directives and a list of options that are aimed to make the user’s navigation experience less troublesome and onto the point. For example, a user on a wheel chair can set the system such that stairs are excluded from path/route presented to the user.

3. Ultimate Indoor Navigation: A Low Cost Indoor Positioning and Intelligent Path Finding (UIN)

In this section, the methodology and architecture of Ultimate Indoor Navigation (UIN) are descripted.

3.1. Methodology in UIN

Given the attention Indoor navigation has attracted, so many studies about and around the topic have been carried out. Alongside these studies, a plethora of approaches to achieve this feat has been presented in previous literature. In this section, we present the complete methodology that we propose, accompanied with the more important algorithms involved.

3.1.1. Tools Used

Wi-Fi fingerprinting and smartphone's inbuilt inertial sensors are used in our paper. This choice is made because Wi-Fi Access Points (APs) are readily available and are already installed throughout not only the school campus (the test case in this paper) but also in most indoor facilities nowadays such as industries, airports, hospitals to mention but a few and pretty much everyone owns a smartphone. Given these conditions, a low cost functioning, robust navigation system can be built.

The smartphone's accelerometer provides x, y and z coordinate values that, with data processing, can not only provide a lot of information but also provide pretty amazing results. Information it can provide ranges from the relative position of a user in general and the user's device in particular, the distance travelled over time and so forth. However, the accelerometer values are not used to determine the distance travelled by the user in this application because that would require double integration of the achieved values whose error bounds increase tremendously with the operation. We instead use fingerprinting, dead reckoning whilst employing the classic Euclidean distance formula in equation (3) below to estimate the distance travelled by the user to a sufficiently estimate the pose of the navigator. Using the magnetometer and gyroscope, the user's orientation and direction can be and are determined and it is much easier to determine whether the user is on or off track as smartphones have Wi-Fi receivers and scanners. It should also be pointed out at this point that smartphone technology nowadays has improved significantly. Therefore, smartphones are capable of carrying out computation tasks of this scale without negatively affecting the phone's daily/expected tasks.

3.1.2. Sensors

In this section, a brief overview of the sensors most relevant to the paper will be made.

<u>Accelerometer</u>

The acceleration in smartest phone devices is basically related to the phenomenon of weight experienced by a test mass that's found on the reference frame of the accelerometer (device). An accelerometer is thus a device that measures this proper acceleration and hence, this acceleration is not necessarily the change of velocity of the smart phone in space/coordinate acceleration. Given the general structure of this accelerometer, a device at rest relative to the earth's surface would show roughly 1g upwards due to its weight. Where g is the gravitational force, whose unit is m/s2. The accelerometer of a smartphone measures both dynamic motions such as movement, phone tilting in the x and y-axes and static forces such as the gravitation force (in the z-axis).

<u>Gyroscope</u>

A gyroscope is a device that, according to the principles of angular momentum, sensors, measures orientation. A mechanical gyroscope which consists of three two gimbals onto which a spinning a wheel that resists changes in orientation is attached. Unlike it, conventional gyroscopes in smart phones generally are made of Micro Electro-Mechanical Systems that measure angular rate, hence the name rate-gyros, while mechanical gyroscopes observe the change in the angle of adjacent gimbals as the spinning wheel remains at a constant global orientation.

<u>Magnetometer</u>

Magnetometer is a device designed to measure changes in the strength of the Earth's magnetic field. Some of the spacecraft carry magnetometers to measure the magnetic field of other planets. The magnetometer shows that the shape and strength of the Earth's magnetic field is constantly changing.

3.1.3. Steps Involved

The system's fingerprinting mechanism set up was made in phases, an online phase and an offline. During the online phase, fingerprints were recorded and stored for reference. It is of paramount importance that all stored fingerprints have distinct values from each other. Fingerprints have signal strength value at a given location. This is coupled with a short vivid description of the location (referred to as node during route planning and calculation) for assistance in routing. For example; to the left of the canteen right ahead of you, since indoors mere signal strength values (say -75dB) or coordinates such as (x, y, z) carry no comprehensible meaning to the user who may be a temporary guest at the institution in question. An appropriate floor identification (floor id) value is also added to the AP's identifiers. This is very useful for multi floor buildings. It helps to distinguish one floor from another with ease. Assigning a value of say 11 to the first floor, 22 to the second floor, 33 to the third floor and so on and so forth, helps distinguish floors that receive sufficiently strong signals from identical APs since the height from one floor to the other is significantly.

After data collection, then we can go onto navigating the user(s). An instance of the node's coordinates is also stored – this is mainly for usage by the dead-reckoning algorithm as explained later. The dead reckoning service runs in background and, basing on the user's concurrent location, the user's entered parameters (main source and destination), aids in the timely warning of the user in case the user is off route. In addition, furthermore, triggers route dynamic recalculation thereby reducing the delay these operations would cause in real time.

3.1.4. Data Acquisition

One of the most challenging and time consuming tasks in any fingerprint based approach of indoor mapping is the generation of discriminate fingerprints for given points of interests (POIs/nodes). This process is not only time consuming but also requires that empirical results be obtained a couple of times due to signal instability of the Wi-Fi networks resulting from interference, multipath and so on and so forth. As such, we felt the need to develop, as part of the paper, an application that will aid to ease the process. This we believe will not only benefit us, but also future researchers in the field.

Fig. 1. Data Acquisition and Configuration Assisting Application.

The Figure 1 is part of the smart system that helps in the data acquisition process. The system keeps scanning the environment at a constant pre-specified interval – the default is 2 seconds. The user only needs to walk through the indoor environment while stopping at significant POIs (nodes). If the user stops at a place for a period greater or equal to 3 seconds, the system will automatically save the place in the application database. Saved will be the relevant information for both the dead reckoning and fingerprinting algorithms. Such information includes filtered sensor data, room name, ranked RSS (Received Signal Strength) from nearby Access Points (APs) and a short description of the node. This data is displayed to the user before it is actually used in the auto Graph (to be described later) generation phase. This gives the user time to make necessary changes to the data before it is saved. The user (one doing the mapping) then presses the saved button if he/she is satisfied with the current information. The saved information is then used in the generation of an undirected connected graph [29] used in route calculation and path presentation.

3.2. Route Calculation and Planning

A modified and optimized form of Dijkstra's shortest path algorithm is used to generate an optimal route from the user's current position to the desired destination. We modified Dijkstra's algorithm to suit our needs in various ways, ranging from enabling physically challenged users to dodge staircases in tall buildings to being able to evacuate as fast and safely as possible through emergency exits. Most path planning algorithms and applications use mainly either graph-based approaches or cell-based approaches [30]. Our system belongs to the former. Dijkstra algorithm basically finds the short path, given a graph G comprising of a set of Vertices, V and a set of Edges, E so that the graph can mathematically be represented as G = {V, E}. Attached to each edge, E is a weight w. A special vertex s can be fixed and considered as the source. The algorithm then finds the shortest path from s to each v Є V. We modified the algorithm such that the shortest path is found from a specified source to a specified destination. Extra parameters are also fed to the algorithm so that directions such as left & right are catered for depending on the user's current location and the desired destination. If the user digresses from the right path, the system automatically notifies the user and recalculates the route/path updating the source as the user's current position whilst keeping the destination fixed.

3.2.1. Route Representation

After successfully calculating an optimal route from the user's current location to the desired destination, it is very crucial that this information be clearly presented to the user in the simplest way possible. Visually impaired and physically challenged users should also be put into consideration while choosing the display mechanism of the route information to the user. Either a dynamically updated visual 2D or 3D map-model or simple precise list directives in form of text instructions coupled with audio assistance would do just fine. This is because we would not want the user to always be locked on to the screen while travelling, but the updated information should be always available when needed, say for consultation. We therefore integrate text instructions in our system with audio assistance to the users. The instructions/directives are kept as precise, accurate and understandable as possible (Figure 2.).

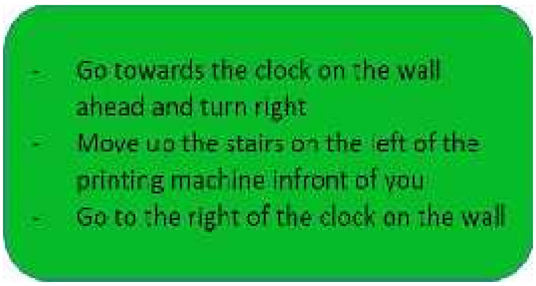


Fig. 2. Route representation as simple precise textual directives.

In this paper, we choose the later because it is easier to comprehend and follow. Some people are biased towards map reading. Furthermore, we think it is easier to listen to directives or quickly comprehend written directives during an emergence than try to understand and follow a map in an unfamiliar environment.

3.2.2. Position and Tracking

Signal strength received from the APs (RSS) and data received from the sensor fusion operations are used for user positioning and tracking. Whichever set yields the least error compared to the values stored in our database is considered as the current user position. This is done in order to compensate for temporal variations in signal strength or random noise received from the smartphone's inertial sensors – though most of the time the two are used to support each other.

Detecting User's Steps (The algorithm)

Recent versions android (4.4 and later) support the step-counting feature that can easily be used to detect user steps. However, the accuracy of the API used varies greatly from device to device. During the course of this paper, it was found to be more accurate while using Google's nexus 5 phone than with the HTC M8S phone. Occasionally the step detection was delayed in the nexus device as well. This delay was more pronounced in the HTC device rendering the current API alone unusable for our purposes. Need for a more consistent, faster algorithm arose. The approach for step detection in this paper is similar to that employed in [13] and [31]. By applying a few operations on the accelerometer readings, we can get the user's steps.

The algorithm to detect the user's steps is as follows. Empirically determine a suitable window to pass over filtered sensor readings to categorize them as either a step or not. When a user takes a step, the accelerometer magnitude values show a peak. This window's value is of paramount importance to the accuracy of the results yielded as a very small value gives too many steps since noisy data or even slighter motions can be detected as steps and, conversely, an extra-large window will give less steps than those actually taken by the user.

Let R_a represent accelerometer readings, W be the chosen window, $M(R_a)$ be the median accelerometer reading, T_{elap} be the elapsed time value, $D(W)_{cur}$ be the current standard deviation of the window and D_{thres} be the standard deviation threshold value of the window.

We say the window represents a step if;

$Max|R_a| = M(R_a)\forall R_a \in W$ and

$D(W)_{cur} \geq D_{thres}$ (1)

Over a given time T_{elap} for any two successive steps. With most Points of Interest (POI) being fingerprinted studies, find it sufficient to use the fingerprints to estimate the user's current location. In this paper, we further supplement this practice with a dead-reckoning service to improve accuracy and performance of the application.

For example assuming a user moves from a start position to a point $p(x_1, y_1)$, his position can be defined in terms of the distance (d_1) travelled and the direction (α) can be easily determined using the Magnetometer sensor. Now, if the user moves to point $q(x_2, y_2)$. The successive position can be obtained using the dead-reckoning technique as follows:

$y_1 = d_1 \sin\alpha$

$x_1 = d_1 \cos\alpha$

$y_2 = y_1 + d_1 \sin\beta$

$x_2 = x_1 + d_1 \cos\beta$ ……(2)

The Euclidean distance, d, between the two points can be obtained using the equation;

$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ (3)

Dead reckoning requires the starting position, the change in direction/orientation and the actual or an estimate of the distance travelled by the navigator. If, however, any of these parameters is not sufficiently accurate, dead reckoning yields accumulative errors [32].

The gained results from the dead reckoning algorithm are always check and compare with the results of the fingerprinting based algorithm and the least error is considered the true result. Thus correcting and updating the algorithm of the more erratic algorithm – assuming that the result return by the more erratic algorithm is catastrophic, else no need to update the parameters. Say if there is a difference greater than 2 meters between the returned results.

Filtering Sensor Data

Raw sensor readings can yield disastrous results. This is because the data obtained by these sensors contains noise. It is thus necessary to filter the data before usage in the algorithms to avoid errors.

Below formula is the first-order discrete low-pass filter applied to the sensor data. It is a recursive filter and can easily be implemented (in Java for this paper):

$$y(t) = y(t_{-1}) + \mu(x(t) - y(t_{-1}))$$

s.t: $$\mu = \frac{dt}{TC+dt} \quad (4)$$

Where, dt: is the period of sampling the sensor. TC: cutoff frequency or time constant. Doing this yields relatively smoother less noise readings as shown in Figure 3 and Figure 4.

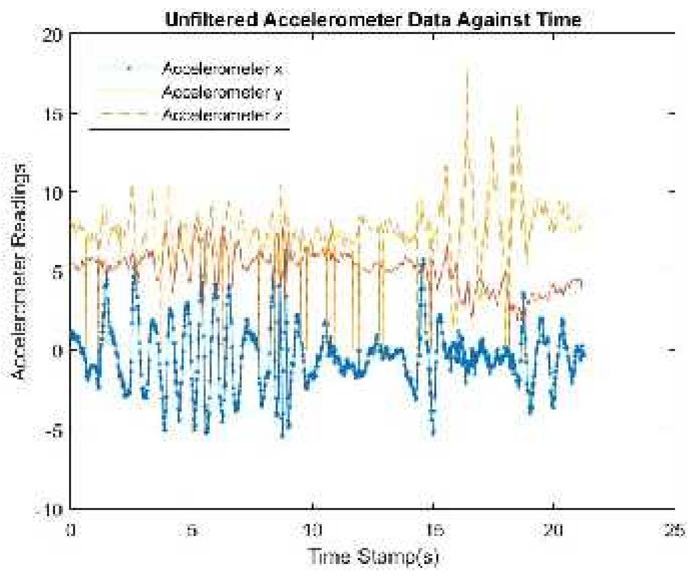


Fig. 3. Unfiltered Accelerometer Readings

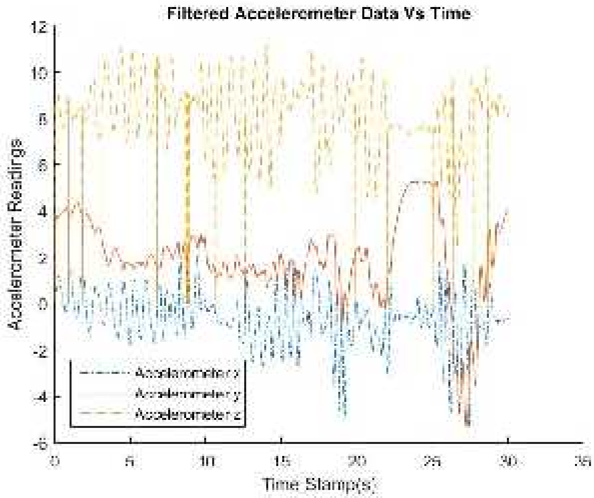


Fig. 4. Filtered Accelerometer Readings using a low pass filter.

The figures 3 and 4 are shown the behavior of accelerometer data against time in unfiltered and filtered reading method, respectively.

3.2.3. Route Confirmation and Context Awareness during Navigation

Indoor navigation can be tricky sometimes as the app user can either easily miss the destination or go off track. It is, thus of paramount importance to let the user know once they are off the right track. As soon as the route has been calculated, the user is constantly notified whether they are on or off the right track. In this paper, fingerprinted nodes and the results from dead-reckoning are used to determine whether the user is on or off the right track. If the user is not on the right track, an appropriate notice is issued and the route is re-calculated. A few of the advantages of the fingerprinting, dead reckoning hybrid algorithm are briefly described below;

1. If an AP breaks down or is changed/replaced, the fingerprinting algorithm will give erroneous results. The dead-reckoning algorithm compensates for this.
2. RSS varies over time. This could lead to errors in the routing algorithm. In such cases, the dead reckoning algorithm backs up the fingerprinting algorithm. Likewise, dead-reckoning results are prone to error accumulation as stated above. Periodic comparison of results from both algorithms helps address this issue. Since some rooms/nodes use one AP, if there happens to be a temporary significant variation in RSS, the navigator may easily be routed to the wrong location.

For example, our assuming is the navigator was wishing to move from TD006 with AP E to TD005 with AP A. Where A and E are the base nodes for the Wi-Fi fingerprints stored in the knowledge base. If AP B happens to be disabled, or replaced or breakdown -depicted by the Red Cross in the Figure 5- the situation described above would arise in which case the results from dead reckoning would be more than sufficient to seamlessly rectify the issue without troubling the navigator.

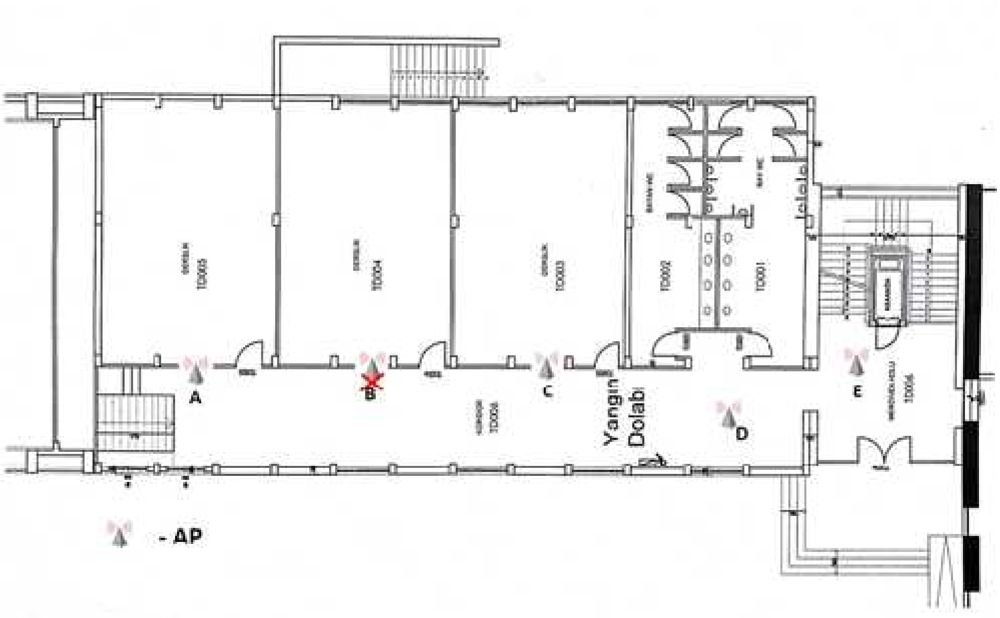


Fig. 5. A demonstration of algorithm compensation.

4. Architecture of the UIN

In this section, the architecture of the proposed system is presented. It explains how the major components of the entire system are linked to and interact with one another. The system comprises mainly of receivers, transmitters, a knowledge base (database) and a processing unit. The navigator's smart phone works as a transmitter, receiver, processing unit and contains the knowledge base. The APs are receivers as well. They receive signals transmitted by the smart phone. Then they in turn also transmit data to the smart phones. As mentioned section 3, the architecture set up is in two main phases; the offline and online phases of the Wi-Fi fingerprinting. During the offline phase, reference data is acquired, processed and added to the knowledge base. During the online phase, a user unfamiliar with the environment or wishing to do navigation to a certain location within the environment simply queries the system for a route from a given location to a desired destination.

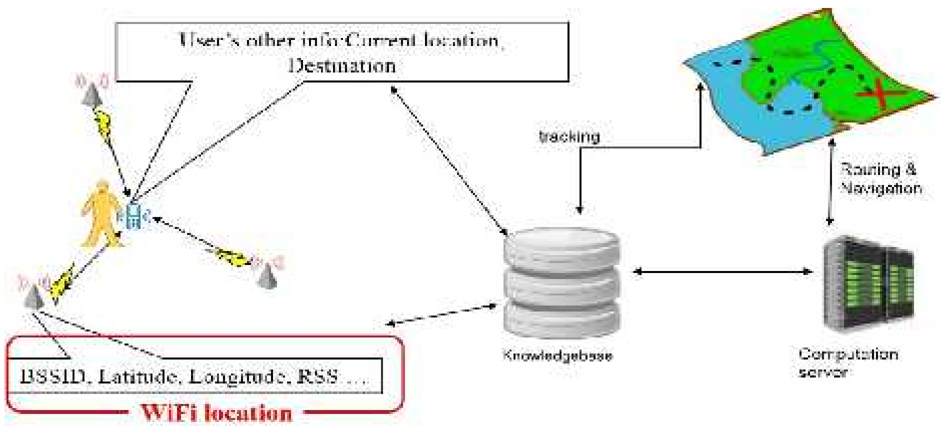


Fig. 6. Model of the proposed system.

Generally, as shown in the Figure 6, the navigator first provides the current position of the location from which they wish to start the navigation guidance to their desired destination and provides the optimal route according to the navigator's configurations. Among the options the navigator has are:

- Whether they wish to use stairs or elevators.
- The rate (in seconds) over which the audio directions are given. This is set to 10 seconds by default.
- Whether they are physically disabled …

Then as the navigator moves, the system keeps on periodically calculating and updating the route according to the user's current location in real-time whilst providing notifications whether or not they are on the right track or not. Figure 7 shows the simple choice screen of the prototype.

Fig. 7. Simplified Choice Screen of prototype.

The message in the toast is of no significant importance. It is simply a message meant for debug purposes and it displays the strongest received signal now the screenshot was taken.

5. Experimental Results

In this section, the experiment tests carried out to test the suggested prototype are explained. Empirical results obtained when testing the system described above are displayed. A Google Nexus 5 using android version 6.0.1 was used throughout the testing procedure at the Istanbul Sabahattin Zaim Campus (Figure 8). These results include the accuracy level, success and the failures with corresponding failure levels.

Satisfying as the results may be, there is still a lot of room for improvement and development. Both as far as performance and providing a much better user experience are concerned. This section covers such areas that can be improved or enhanced are presented. Android Studio [33] was used as the Integrated Development Environment.

5.1. Tests and results

The following scenarios were tested;

1. User at the very entrance of the floor provided both source and destination. In addition, the user can follows instructions from the application in an unbiased way.

2. User provided both source and destination, and then followed the instructions while intentionally getting off the suggested route.
3. User provided wrong position as the current location (Check whether and, if so, how fast the program will update the current location).

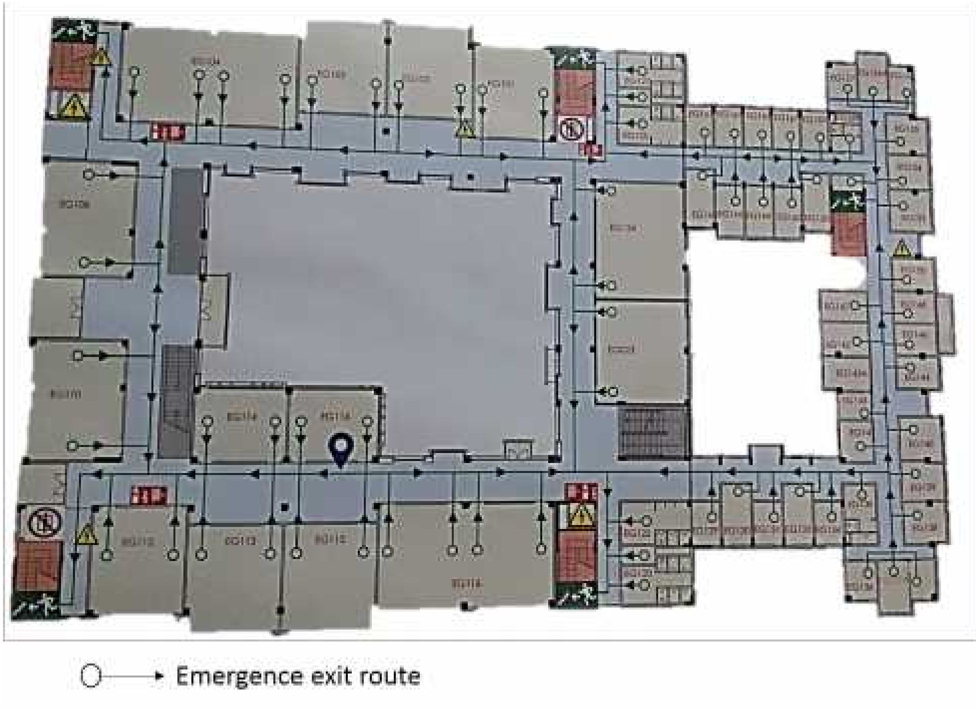


Fig.8. A simplified version of the used map in the paper prototype of the first floor of the Education blocks left wing.

As shown in Table 1, some results denote failures. These errors are obtained when the paper prototype gives a wrong node at a given point/node, say if Class EG117 is shown as EG116. Every correctly identified significant node on the path calculated and presented to the navigator is considered as success. The values in the brackets show how far off the yielded results actually are from the intended location. Using the analogy above, this corresponds to how far the user was taken from EG117. As for the case of the route from EG110-EG108, it is because both room EG110 and room EG108 share the same node and were assigned the same node value. Hence an empty list of results is returned since it was, for test purposes, was configured to do so. The same behavior will also be exhibited in case we try to navigate towards a disconnected node in a disconnected graph. There are so many choices for handling these situations, in this case since the two rooms are right adjacent to one another so we chose to simply let the user know that they have already arrived at their desired destination and add a direction they should turn to depending on their orientation. Given that, all the failures are with ≤2m range or less – with 2m being the upper bound. The descriptive navigation instructions accurately and very efficiently rectify this offset since all the intended targets are within eye sight so considering the whole system has an average accuracy of about 0.8-1.3m. This is sufficiently accurate for indoor building even those with small rooms since room level (~2m) accuracy is just sufficient. The average of ~0.8m was observed for standard sized rooms (~4m3) such as student laboratories or even larger whilst that of 1.3m was observed for relatively small rooms (~2.1m3) for example, the rest rooms.

Table 1. Sample results table.

Current Location	**Input Location**	**Destination**	**Time to calculate route(ms)**	**Average Update Time If required(ms)**	**Success**	**Failure (Deviation in meters)**
EG123	EG123	EG110	16	Null	5	0
EG110	EG110	EG104	27	Null	4	0
EG110	EG110	EG108	5	Null	1	1(~1.2m)
EG113	EG113	EG101	27	Null	6	2(~≤1.33m, *0.4m*)
EG124	EG117	EG104	20	46	5	1(≤*1.5m*)
EG104	EG101	EG122	25	37	7	1(≤*1.28m*)

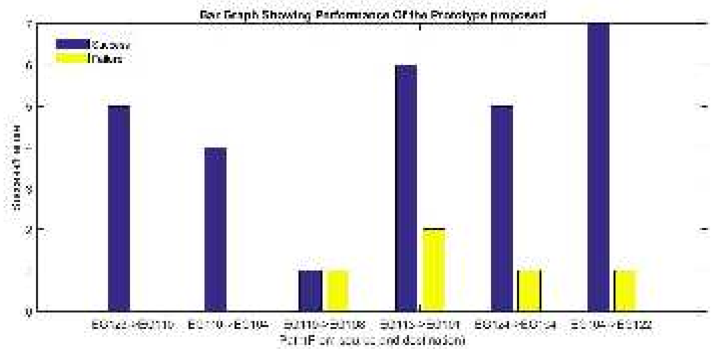


Fig. 9. Bar graph reflecting the performance of the suggested prototype.

Figure 9 shows the performance of the proposed prototype. The overall accuracy of the system, let alone the fast response time, is sufficiently accurate at room level - which is approximately less than (in most cases) or equal to 2m (~ ≤ 2m). This accuracy, however, drops drastically if there happens to be a disconnected node in the graph. Hence, paramount attention must be paid whilst creating the graph. The graph should not be disconnected. The effects of a disconnected graph on the performance and accuracy of the proposed prototype are demonstrated in Figure 10 and Figure 11.

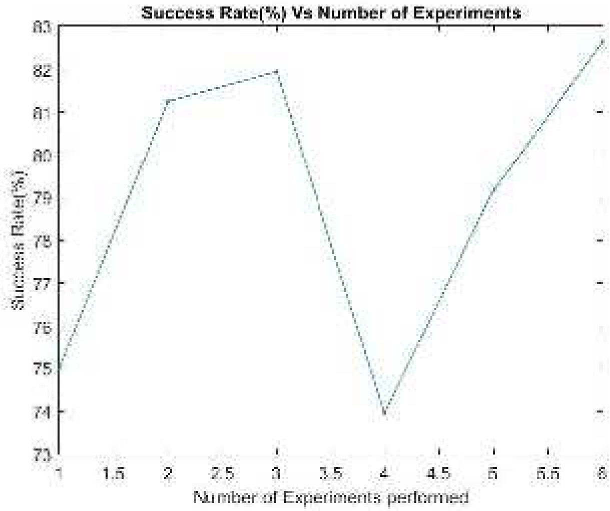


Fig. 10. Average success rate of a disconnected graph

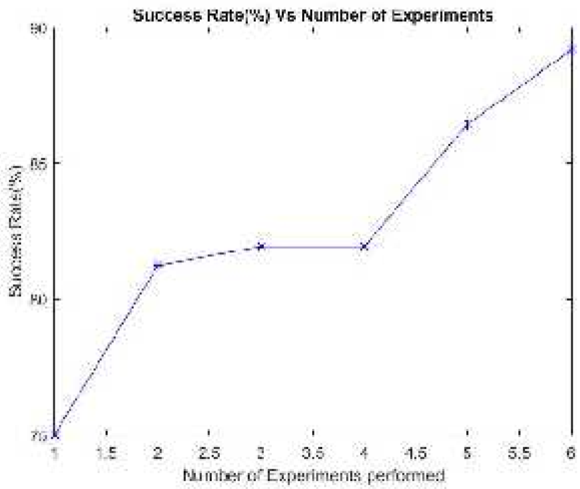


Fig. 11. Average success rate of a connected graph.

In the disconnected graph that is shown in Figure 10, a single classroom was not connected to any other room/node in the graph whilst in Figure 11. This node was not put into consideration.

Furthermore, the prototype scales with building room or node number size. Latency or calculation delays do not get worse with increasing number of node points. This is clearer shown in Figure 12. This Figure shows the scalability of the suggested prototype. The number of nodes in the map does not have a significant negative effect on the time take to calculate a route.

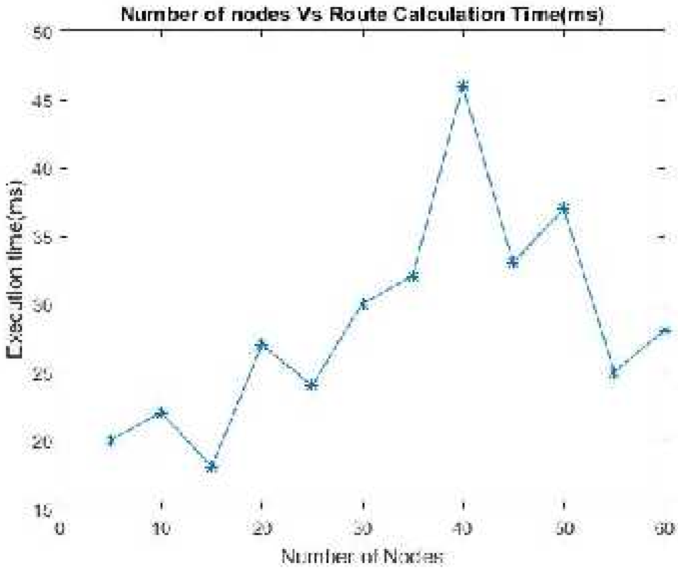


Fig. 12. Number of nodes Against Route Calculation time (ms).

5.2. General Comparison with trending approaches

In this section, we briefly show how UIN fairs with other trending approaches from related works performance wise. On top of providing an optimal route and navigation services for users – including the visually impaired and the physically disabled, UIN performs rather well performance wise as shown in Figure 13.

Method / Approach	Average Accuracy (m)
KNN	4.37
Probabilistic	2.78
Fingerprinting	2-3
Probabilistic and particle filter	1.96
ALPCIPIPF	0.8 – 1.3

Fig. 13. General Performance Comparison.

6. Conclusion and Future Work

In this paper, a cost efficient user interactive indoor navigation system using Wi-Fi fingerprinting and sensor fusion that requires no installation of extra hardware or third-party software is proposed. The system offers optimal navigation to the user with an upper bound of 2m and an average of 0.8-1.3m accuracy. The proposed prototype is also suitable for the visually impaired and physically challenged as it enables the user to choose the kind of configuration they prefer. In any case, the most efficient route is always suggested and presented to the navigator. The system implementation is a hybrid of Wi-Fi fingerprinting and dead reckoning performed using resultant data from the application of a filter on sensor data so as to reduce noise. Prototype test cases were carried out on the Education block of IZU campus. This prototype is of an indoor navigation system that runs on a user's smart phone device and requires nothing else. The suggested prototype is suitable for navigation and emergence evacuation for both normal, visually impaired and physically challenged users. Audio support is included into the system so users are able to get voice directions as they navigate throughout an unfamiliar indoor environment.

In this section, future works for this paper are presented and possible improvements are suggested. The system is likely to suffer serious accuracy issues if any AP is taken down, removed or changed especially, assuming the node(s) associated with the AP have not multiple AP access but single or two AP access are connected to many other nodes. Worst-case scenario is if the other APs are also experiencing non-insignificant RSS variations. This would throw both the fingerprinting and the DR algorithms off track. This means that periodic checkup of the APs would be in our best interest or constant communication with the ones responsible with AP setup in the facility in case of any changes. Furthermore, environmental changes such as infrastructural renovations also have an impact on the accuracy, but not the performance of the application. We therefore plan to implement automatic AP update or leave it up as a significant future work.

This prototype was designed basing mainly on only one building or infrastructure - from the building/infrastructure's management's point of view (say a university campus such as ours). However, we would also like to think from the users /navigators' point of view. If the users install this software from a given building, it also implies that they would have to re-install the software when they go to another infrastructure (again say, university) that uses the same software. Hence, with a little finance or support we would like to set the system up such that users install the software once and can use the finished paper in any place that uses the same software. Seamless integration with google maps so that the navigator seamlessly travels not only from building block to building block, but can also easily and comfortably use the app outdoors. Multi-language support. In this prototype, due to brevity, only the English language is supported. However, in the later versions, a wider range of language choices for example French, Turkish, Arabic, Spanish, to mention but a few will be implemented. A more locale language based implementation is not a big challenge. We plan to give the user a choice of the route/path representation in the future. A 3D/2D dynamic map such the one used by Google maps or the textual representation as shown here.

References

[1] R. Want, A. Hopper, V. Falc˜ao, and J. Gibbons. "The active badge location system", *ACM Transactions on Information Systems (TOIS)*, 10(1), 91-102, 1992. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.36.123&rep=rep1&type=pdf

[2] M. Weiser. "The computer for the 21st century", *ACM SIGMOBILE Mobile Computing and Communications Review*, 3(3), 3-11, 1999. https://www.ics.uci.edu/~corps/phaseii/Weiser-Computer21stCentury-SciAm.pdf

[3] A. Butz, J. Baus, A. Krüger, M. Lohse, "Hybrid Indoor Navigation System", *International Conference on Intelligent User Interfaces, ACM*, 25-32, 2001. https://dl.acm.org/citation.cfm?id=359832

[4] J. Raper, G. Gartner, H. Karimi, C. Rizos, "A critical evaluation of location based services and their potential", *Journal of Location Based Services*, 1(1), 5-45, 2007. https://dl.acm.org/citation.cfm?id=1392058

[5] A. Miu, "Design and Implementation of an Indoor Mobile Navigation System", *Master Thesis of MIT*, 2002. http://inat.lcs.mit.edu/papers/cricketnav-thesis.pdf

[6] "Bing Venue Maps", http://www.bing.com/maps/?vmb=1

[7] "Google Indoor Maps Availability", http://support.google.com/gmm/bin/answer.py?hl=en&answer=1685827

[8] "Nokia, Samsung, Sony, Qualcomm, Others Team Up for Indoor Positioning", http://www.xbitlabs.com/news/mobile/display/ 20120824002723 Nokia Samsung Sony Qualcomm Others Team Up for Indoor Positioning.html

[9] Y. Chen, D. Lymberopoulos, J. Liu, B. Priyantha, "Fm-based indoor localization", *In Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services, MobiSys '12*, 169-182, 2012. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/sys029fp-Chen.pdf

[10] N. Priyantha, A. Chakraborty, H. Balakrishnan, "The Cricket Location Support System", *Proceeding of the Sixth Annual ACM International Comference on Mobile Computing and Networking (MOBICOM)*, 32-43, 2010. https://dl.acm.org/citation.cfm?id=345917

[11] D. Gusenbauer, C. Isert, J. Krosche, "Self-contained indoor positioning on off-the-shelf mobile devices", *International Conference on In Indoor Positioning and Indoor Navigation (IPIN)*, 1-9, 2010. https://ieeexplore.ieee.org/document/5646681

[12] J. Link, P. Smith, N. Viol, K. Wehrle, “Footpath: Accurate map-based indoor navigation using smartphones”, *International Conference on In Indoor Positioning and Indoor Navigation (IPIN)*, 1–8, 2011. https://ieeexplore.ieee.org/document/6071934

[13] G. Trein, N. Singh, P. Maddila, “Simple approach for indoor mapping using low-cost accelerometer and gyroscope sensors”, *Computer Science, Toronto University*, 1-25, 2013. http://www.cs.toronto.edu/~trein/csc2228/files/report.pdf

[14] N. Ravi, P. Shankar, A. Frankel, “Indoor Localization Using Camera Phones”, *7th IEEE Workshop on Mobile Computing Systems and Applications, WMCSA'06*, 19-19, 2016. https://ieeexplore.ieee.org/document/4625206

[15] A. Brown, “Urban/Indoor Navigation using Network Assisted GPS”, ION 61st Annual Meeting, 1-6, 2016. https://pdfs.semanticscholar.org/8f5b/167364f02dd5352d3e4cd3f4840882742eb3.pdf

[16] R. Abdalla, “Introduction to Geospatial Information and Communication Technology”, Springer *International Publishing Switzerland*, 21-40, 2016. https://www.springer.com/gp/book/9783319336022

[17] A. Panyov, A. Golovan, A. Smirnov, “Indoor positioning using Wi-Fi fingerprinting pedestrian dead reckoning and aided INS”, *IEEE International Conference In Inertial Sensors and Systems (ISISS)*, 1-2, 2014. https://ieeexplore.ieee.org/abstract/document/6782540

[18] T. Mori, Y. Suemasu, H. Noguchi, “Multiple people tracking by integrating distributed floor pressure sensors and RFID system”, *IEEE International Conference on Systems, Man and Cybermetics*, 6(10), 5271-5278, 2014. https://ieeexplore.ieee.org/document/1401031

[19] J. Saarinen, H.Suomela, M. Elomaa, A. Halme, “Personal navigation system”, IEE/RSJ *International on intelligent Robots and systems*, 212-217, 2004.

[20] T. Höllerer, D. Hallaway, N. Tinna, S. Feiner, “Steps toward accommodating variable position tracking accuracy in a mobile augmented reality system”, *International Workshop on Artificial Intelligence in Mobile Systems (AIMS’01)*, 31-37, 2011. http://www.cs.columbia.edu/~drexel/research/Hollerer-2001-AIMS.pdf

[21] I. Park, G. U. Jang, S. Park, J. Lee, “Time-Dependent Optimal Routing in Micro-scale Emergency Situation”, *International Conference on Mobile Data Management: Systems, Services and Middleware*, 714-719, 2009. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5089029

[22] M. Brunato, K. Csaba,”Transparent location fingerprinting for wireless services”, *Toronto University, Technical Report*, 1-4, 2002. http://eprints.biblio.unitn.it/227/

[23] J. Stook, “Planning an indoor navigation service for a smartphone with Wi-Fi fingerprinting localization”, *Master's thesis in OTB Research Institute for the Built Environment Faculty, TuDelft Univ.*, 2012. https://dspace.library.uu.nl/handle/1874/234909

[24] B. Kannan, F. Meneguzzi, M. Dias, K. Sycara, “Predictive indoor navigation using commercial smart-phones”, *In Proceedings of the 28th Annual ACM Symposium on Applied Computing*, 519-525, 2013. https://www.ri.cmu.edu/pub_files/2011/12/percom-navigation-main.pdf

[25] J. Baus, A. Krueger, W. Wahlster, “A resource-adaptive mobile navigation system”, *In Proceedings of the 7th International Conference on Intelligent User Interfaces*, 15-22, 2012. https://dl.acm.org/citation.cfm?id=502723

[26] G. Abowd, C. Atkeson, J. Hong, S. Long, R. Kooper, M. Pinkerton, “CyberGuide: A Mobile Context-Aware Tour”, *Wireless Networks*, 3(5), 421-433, 1997. https://smartech.gatech.edu/bitstream/handle/1853/3513/96-27.pdf?sequence=1&isAllowed=y

[27] VTI Technologies, “SCC1300-D04 combined gyroscope and 3-axis accelerometer with digital SPI interfaces”, rev. 1.0.3. Doc.Nr. 82 1131 00 A, Data sheet, 2010.

[28] F. Hohl, U. Kubach, A. Leonhardi, K. Rothermel, M. Schwehm, “Next century challenges: Nexus-an open global infrastructure for spatial aware applications”, *International Conference on Mobile Computing and Networking (MobiCom ’99), ACM Press*, 249-255, 1999. https://dl.acm.org/citation.cfm?id=313549&dl=ACM&coll=DL

[29] E. Dijkstra, “A note on two problems in connexion with graphs”, *Numerische Mathematik*, 1, 269-271, 1959. https://dl.acm.org/citation.cfm?id=2722945

[30] J. Link, P. Smith, N. Viol, K. Wehrle, “Footpath: Accurate map-based indoor navigation using smartphones”, *International Conference on In Indoor Positioning and Indoor Navigation (IPIN)*, 1-8, 2011. https://ieeexplore.ieee.org/document/6071934

[31] P. Bahl, V. Padmanabhan, “RADAR: an in-building RF-based user location and tracking system”, *19th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM ’00)*, 2(1), 775-784, 2010. https://ieeexplore.ieee.org/document/832252

[32] H. Huang, G. Gartner, "A survey of mobile indoor navigation systems", *In Cartography in Central and Eastern Europe, Springer*, 305-319, 2009. https://link.springer.com/chapter/10.1007/978-3-642-03294-3_20
[33] Android Studio, https://developer.android.com/studio/index.html
[34] A. Dubey et al., "FM based indoor localization and mapping system with real-time implementation on FPGA", *IEEE 2015 International Conference on Information and Communication Technology Convergence (ICTC)*, 176-181, 2015. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7354522
[35] V. Kadam, Sh. Tamane, and V. Solanki, "Smart and Connected Cities through Technologies", *IGI-Global*, 1-24. 2018. https://www.igi-global.com/chapter/smart-and-connected-cities-throughtechnologies/211738
[36] O. Kucuk, F. Kiani, "Smart Advisor: An Intelligent Inventory Prediction based on Regression Model", *International Journal of Machine Learning and Networked Collaborative Engineering*, 2(3), 86-94. https://www.mlnce.net/index.php/Home/article/view/41
[37] Sharma, S., Chauhan, V., & Khari, M. IoT Based Intelligent Vehicle Parking Solution System. *International Journal of Machine Learning and Networked Collaborative Engineering*, 2(04), 143-150. 2018. https://doi.org/10.30991/IJMLNCE.2018v02i04.00

Author's Biography

Alex Gunagwera got undergraduate and master degree in computer engineering from Fatih and Istanbul Sabahattin Zaim Universities in 2015 and 2017 respectively. He is a PhD candidate in Istanbul Sabahattin Zaim University. His interest research area is robotics, machine learning and ubiquitous. He is corresponding author of this paper.

Farzad Kiani has PhD degree and he is Assistant Prof. in computer engineering dept. at Istanbul Sabahattin Zaim University (IZU) from 2014 to now. In addition, he is deputy of head of computer engineering and head of wireless sensor networks and IoT Labs at IZU from 2016 and 2017 to now, respectively. His current research interests include wireless sensor networks, Vehicle Ad-hoc networks, flying Ad-hoc networks, machine learning, game theory and IoT. He is also working on optimization algorithms and security.

How to Cite

Gunagwera, Alex & Kiani, Farzad (2019). Ultimate Indoor Navigation: A Low Cost Indoor Positioning and Intelligent Path Finding. *International Journal of Machine Learning and Networked Collaborative Engineering*, 3(01) pp 16-34.

doi : https://doi.org/10.30991/IJMLNCE.2019v03i01.002.

Vol. 03, No. 1 ,(2019), 035–045 , **ISSN:2581-3242**

IJMLNCE JOURNAL

International Journal of Machine Learning and Networked Collaborative Engineering

Journal Homepage: http://www.mlnce.net/home/index.html

DOI : https://doi.org/10.30991/IJMLNCE.2019v03i01.003

Artificial Intelligence to Prevent Road Accidents

[a]Jaidev B, [b]Sonakshi Garg, [c]Sandhya Makkar

[a,b,c]*Lal Bahadur Shastri Institute of Managenent, Delhi, India.*

[a] *jaidev_b@lbsim.ac.in, https://orcid.org/0000-0002-1896-1198*

[b] *sonakshi_garg@lbsim.ac.in, https://orcid.org/0000-0002-3912-2289*

[c]*bajajsandhya@gmail.com, https://orcid.org/0000-0003-3442-2730*

Abstract

Due to increasing demand in urban mobility and modern logistics sector, the vehicle population has been growing progressively over the past several decades. A natural consequence of the vehicle population growth is the increase in traffic congestion which in turn will lead to more accidents. Accident prediction is one of the most vital aspects of road safety. An accident can be predicted before it occurs, and precautionary measures can be taken to avoid it. Artificial Intelligence (AI) can help in improved awareness of road conditions, driving behaviour of the people and can avoid accidents with the help of improved active safety and improved traffic condition. Drug impaired driving is becoming a serious cause of accidents as the days go by. Moreover, it is more difficult to detect drivers who are under the influence of drugs than drivers who are the influence of alcohol. So the purpose of this research is to study and review the literature & industry reports and put in the approaches for detecting the unsafe driving pattern and also maintaining the health of the car to avoid accidents.

Keywords

Artificial Intelligence, Accident prediction, Road safety, Drug impaired driving, Connected cars, Detection of driver behaviour, Prognostics, Health Management

1. Introduction

Road Traffic Accidents are the major cause of death and constitutes one-tenth major cause of all deaths worldwide. An estimation is reported [17] that 1.2 million people lose their lives in road accidents, 50 million are injured and about 30-70% people suffer from orthopedic illness. And if this issue continues, then road accident will become third- major contributor to the worldwide burden of disease and injury by 2020.The major bearers of the road accidents are developing countries. Also, Road Traffic Accidents create tremendous hardship due to loss of family's primary source of support for earnings, since 73% of total road accidents affect mainly males and those aging from 15-45 years old [17].

Artificial Intelligence (AI) is a study of Computer Science that focuses on building intelligent machines and developing algorithms and so to make them ready to act like a human. The automotive industry is among the industries of greatest advancement using AI to imitate and support the actions of humans.AI can be used to make life in car more convenient and safer, beyond the shadow of self- driving vehicles. AI can help in improved awareness of road conditions, driving behavior of the people and can avoid accidents with the help of improved active safety and enhanced traffic condition. From manufacturing or management point of view, AI can also help in making processes more efficient and digital [5].

These days companies are using Internet Of Things (IOT) techniques in every field. In context of automotive industry, field of interest are 'Connected Cars'. Connected Cars are vehicles that are provided with internet access and wireless local network and use communication technologies to communicate with the driver about the outside car environment, this helps in improving the road safety. Since, road safety is of prime importance to both the driver and the pedestrians, connected vehicles helps in connecting the driver to different networks i.e. car to infrastructure, car to other cars, car to cloud, car to pedestrian, in short, car to everything [6]. Preventing accidents include major aspects: First is, prognostics and health management, then second is detecting the behavior of the driver and third is intelligent driver assistance. These are defined as under:

1.1. Prognostics and Health Management (PHM)

Prognostic is related to prediction of an asset and is based on prediction about the reliability, availability and health state of an asset. The asset prediction is made in order to reduce the maintenance cost and it also prevents the asset breakdown and further promotes the safety reducing accidents. Prognostics can be used in various applications like aerospace vehicles, civil infrastructures, mining machinery [10]

1.2. Behavior Detection

Behavior detection is a study of paying attention to human signals, both relating to behavior as well as human normal physical functioning. Towards traffic accidents, detection of human behavior is an important area of interest to improve global road safety problem [11].

1.3. Intelligent Driver Assistance

Advanced Driver Assistance System (ADAS) assists the driver of vehicle by providing safety warning messages during critical and life-threatening situations. It not only ensures safety of drivers, but also ensures the safety of cars and outside environment (pedestrians, animals etc.). For instance, the driver drowsiness detection system in which the artificial intelligence tool detects whether the driver is falling asleep during driving and alert the driver so that he can stop driving and take rest. Any of the driving styles can be measured through ADAS in real time and it also provides necessary feedback [7], [10].

2. Background

In any industry, the damage of asset can cause social as well as economic damage due to partial defaults and degradation in the asset. Likewise, in automotive industry, the damage to automotives may lead to car accidents and casualities (social loss). Reason being, in order to avoid accidents, most machines and systems (automotives), depend upon routinal preventive maintainence on regular interval basis. Routinal preventive measures were having their own limitations in preventing failure and incur high cost in case of unnecessary replacement of undamaged parts. So, to prevent this unnecessary cost, PHM came into light gathering real time data about condition of vehicle from wireless sensor network and analysing it**.** PHM offer bundle of servies (a) health monitoring for defaults or degradation (b) diagnosis of abnormal physical condition (c) forecasting of remaining useful life (RUL) (d) accusation when maintainence is necessary.All these services help in predicting the potential failure in the operating device and provide information required for risk mitigation and management [8].

Furthermore, understanding the behavior of Vehicle Driver to predict the accident has become an important topic of research these days. In Vehicle driver environment system driver behavior is an important factor. In increasing driving safety real driving monitoring system has a big role [10],[11].

2.1. Prognostics in automotive industry

Prognostics in automotive industry is a systematic way of observing the reliability of the vehicle on real time basis. In the automotive industry, this technology has been adopted in order to render advance word of advice about safety-related failures, to reduce the unscheduled maintenance of vehicle. Vehicle is a complex combination of mechanical, electronic and computer engineering structure. It comprises of various subsystems such as gearbox, engine, brakes etc., associated with Engine Control Unit (ECU), which ensures the optimum performance of the engine and other parts of the vehicles as their performance is dependent in performance of the engine. ECU is connected with Controller Area Network (CAN), through which different subsystems of vehicle and controller of vehicle communicate with each other. Protocols On Board Diagnostics (OBD) and Unified Diagnostic Services (UDS) are needed communicate with ECU. OBD model provide the vehicle owner or to a repair professional right to obtain the data about current condition of different subsystems of vehicle whereas UDS provide details about the condition. OBD compares real time data received from subsystems like battery, engine, gear with results obtained from exploratory pattern of the conditions of those subsystems. The current situation of system is obtained through diagnostics and prognostics where diagnostics is related to current state of the vehicle subsystem and prognostics is concerned with future state of subsystem [8], [10]. Prognostic maintenance copes with OBD, which collects the data from different sensors when vehicle is on the move. Then, signals generated from data streams are continuously sent to smart devices (smartphones, laptops) connected to the vehicle via wireless communication. Patents related to PHM in automotive have been taken by various research institutes and car manufacturers. Ford car manufacturer owns an automotive patent that presents a vehicle structure consist of prognostic module; On Board Diagnostic. It is used to determine the device characteristics of degradation [14].

There is large cost associated with usage of sensor data as it requires large memory space and high processor space. So various solution for data reduction were given. One of them, was introduced by [9] various techniques using machine learning for engine. [4] introduced an android based application for vehicle health monitoring system in which engine condition, battery condition, emission system were observed, and notification about the same was delivered to the driver of the vehicle via android phone. A real time Vehicle Monitoring and Maintenance system was introduced (VMMS), under this subsystems of vehicle, ignition system, exhaust system, fuel system, and cooling system are examined in detail. Data collected through sensor is used for default prediction using different machine learning; Decision tree, K-NN, Random forest. Through On Board Diagnostics scanner and smart phones, the data is generated, while vehicle is in move. Data is collected in the form of Diagnostic Trouble Codes (DTC) is communicated to smart phone via Bluetooth and then send to back-end using algorithms patterns are learned which can cause failure to system and in abnormal condition user of vehicle is notified through smart phone or email notification [8], [14].

2.2. Driver Behaviour Detection and Intelligent System

Road safety is of prime importance to both the driver and the pedestrians. AI can help in improved awareness of road conditions, driving behaviour of the people and can avoid accidents with the help of improved active safety and enhanced traffic condition. From manufacturing or management point of view, AI can also help in making processes more efficient and digital. One of the major ways that AI has contributed to road safety is radar based communication. This can be employed by using an algorithm that will predict when a person/object is about to come in front of the car and thereby warning the driver to stop. This can be taken one step further when the car is able to use brakes all by itself. Tesla had recently tested this algorithm. It was able to predict an accident seconds before it happened, and the car was able to brake on its own. However, for Indian roads, the algorithms to be developed are slightly more complex as compared to the West. This is because Indian roads have more objects that can block a self-driving car such as cyclists and animals [3]

[1]The **Driver Alcohol Detection System for Safety (DADSS)** helps in detecting alcohol levels by two mechanisms – the first mechanism is a Breathalyzer system which is located on the steering wheel or the driver side door. It is capable of detecting the alcohol in the air particles around it. This will help in determining how many drinks the driver had. The engine will not start if the system detects the presence of alcohol. The second mechanism is a touch sensor either on the ignition button or on the gear shift lever. The sensor employs near-infrared tissue spectroscopy to determine driver's blood alcohol content. Just like the Breathalyzer system, if it detects the alcohol content beyond the legal limit, a running engine would simply stop or will not start at all. [2]Detecting motorists who are Driving Under the Influence of Drug (DUID) is more complicated than detecting drivers who are driving under the influence of alcohol and also difficult to perform tests for medications and prescriptions in DUID case. This is due to presence of large amount of constituents which may lead to impaired driving and increase the risk of accident, the impact of different type of drugs on driving, lack of information about certain drugs, how drugs can affect body and behaviour. As DUID is a scenario which continues to rise, the ability to identify these cases and apply suitable measures correspondingly is still a challenge for researchers to predict and prevent road accidents. [3]Aggressive driving style may be defined as the driving behaviour that can intentionally increase the risk of collisions. This would include irregular, instantaneous and abrupt changes in vehicle speed, inconsistent or excessive acceleration or deceleration. About 56% of deadly crashes occurred between 2003 and 2007 are associated with aggressive driving (according to American Automobile Association Foundation for Traffic Safety).

[4]Inattentive driving style can be observed as an instantaneous deviation from normal driving behaviour with the following of an immediate reaction of the driver to rectify it (i.e. trying to get back into attention). It is instantaneous and sporadic in nature as compared to aggressive driving style in which there is a pattern of misbehaviour over a period of time [11], [13].

3. Approaches

Approach 1 For Health monitoring a new system has been proposed which shows real time monitoring known as Vehicle Monitoring and Maintenance System (VMMS). The schematic diagram has been shown below

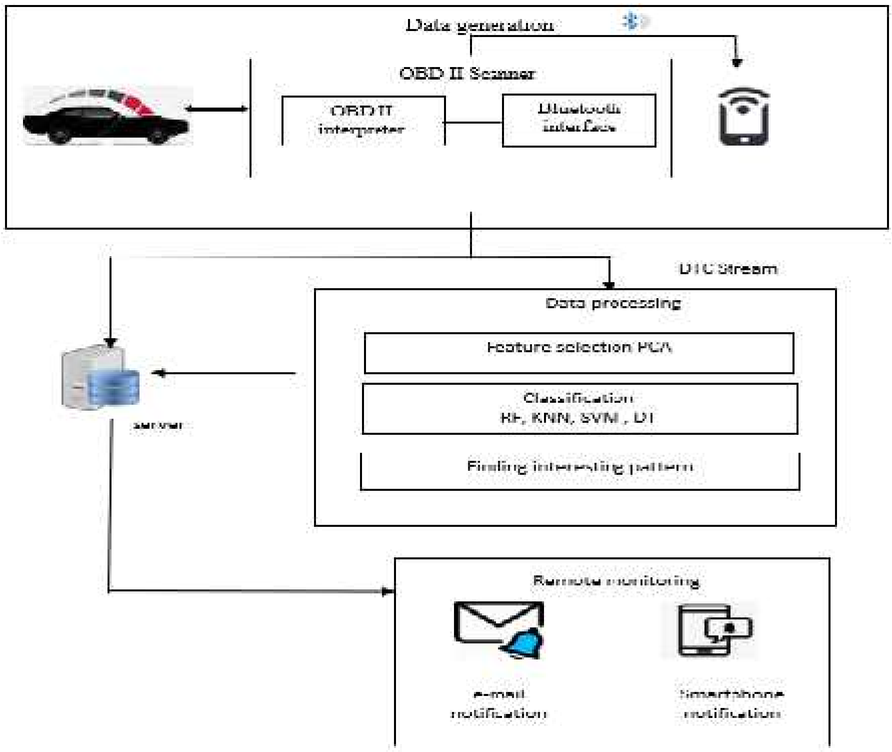


Fig. 1 VMMS Architecture (Source: Ufera Shafi, *et al* Jan 2018 [14])

VMMS Architecture - VMMS Architecture is constituted of three main levels as shown in Figure 1; in first level the data generated. An OBD (On Board Diagnostics) II scanner is wired in the vehicle through OBD II port. There are various car scanner in the market, for say car scanner klavkarr-100 (was developed for only users of Mac or Pc). These scanners are act a link between vehicle and portable device, relevant information from car can be connected to directly laptop screens or smartphones which is supported by Bluetooth or USB cable. Data generated is in form of DTC when vehicle is in move and sent to portable device (laptop or smartphone). Data keeps on generating throughout the vehicle move and transmitted to portable device via Bluetooth. The output of the system is in binary form. The value is set to 1 if DTC is generated (if fault occurs or system breaks down and 0 if system is working perfectly. For this particular case the dataset consisted of 150 examples. Also, the sensor data was used for Corolla Cars owned by Toyota

The algorithms that were selected perform with high accuracy on binary data. The algorithms are as follows:

1) Decision Tree (DT): It is a graphical representation of possible solutions to a decision based on certain conditions. The basic algorithm is applied using Gini Diversity Index (GDI) which shows the measure of node impurity.

   GDI = 1- $\sum p^2 (i)$, where

p(i) is the probability of an instance belonging to class Ci where i can be either 0 or 1 depending on the situation.

2) Support Vector Machine (SVM): It is a supervised machine learning algorithm which can be used for classification or regression problems. SVM separates the instances of two classes and by maximizing the margin between both sides of line, it classifies test instance. It uses a function known as Kernel trick (K) to transform the data and finding an optimal boundary between possible outputs based on the transformations.

   $K(x, x') = \exp(-\eta |x-x'|^2)$, where

   η is a parameter to handle nonlinear classification

3) K-Nearest Neighbour (K-NN): This method is applied when Euclidean distance (d) is used to measure similarity.

   $d(X, Y) = \sqrt{\left(\sum_{i=1}^{n}(Xi - Yi)^2\right)}$

4) Random Forest (RF): To improve the accuracy of Decision Trees, RF is employed. For each sample, a decision tree is learned. If data is sampled n number of times with replacement, we get n decision trees. Such multiple number of decision trees constructed is known as Random Forest.

The performance of the above mentioned algorithms are evaluated on the basis of precision, recall, accuracy and F1 scores measures. The equations used for calculating each are as follows.

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

$$RECALL \quad \frac{TP}{TP+FN}$$

$$F1\ SCORE \quad \frac{2 * P * R}{P + R}$$

$$ACCURACY = \frac{TP + TN}{TP + TN + FN + FP}$$

Where TP → True Positive (detects a condition when the condition is actually present)

TN→ True Negative (does not detect a condition when the condition is actually absent)

FP → False Positive (detects a condition when the condition is actually absent)

FN → False Negative (does not detect a condition when the condition is actually present)

Approach 2 To find out whether drug has a serious impact on increasing amount of accidents, two types of data were collected: Structured and Narrative [13].

Structured data: It refers to information or data that has a defined format. It is organized into what is known as database so that its elements can be used for effective processing and analysis. Examples include numbers, dates or even group of words and numbers termed as strings. In Borba's study, he has used data regarding incidence rates in automobile accidents and percentage of accidents with injury for various conditions such as Time of day, Environment, Nature of Accidents and Driver condition (Fatigued or Alcoholic)

Narrative data: This approach focuses on gathering information for the purpose of research through storytelling. The researchers then interpret stories that are told within the context of research. In the case of Borba's study, data regarding Incidence rates in automobile accidents and percent of accidents with injury due to presence of Medication, Prescription, drug or illegal narcotic have been employed.

Approach 3 iPhones have an app known as Drivesafe which deduces drowsy and aggressive driving behaviours and gives corresponding feedback to the drivers and scores to their driving. From [12], they have taken help of UAH-Driveset which is recorded by Drivesafe app. The data was collected of 6 different drivers and cars and simulated 3 different behaviours (normal, drowsy and aggressive).

In normal driving, the driver was told to drive in the usual way. In aggressive driving, the driver was told to push to his limit (pedal hard on the accelerator, abrupt braking etc). In drowsiness case, the driver was told to act like he was slightly sleepy (slightly unaware of the road ahead). The processed data also contains maneuvers recognition which are acceleration, braking, turning, lane weaving, drifting, overspeeding, car following. Each driver performed on motorways as well as secondary roads.

4. Findings

4.1. Health Monitoring and Prognostic Maintenance (Based on Approach 1)

Table 1: Ignition systems VMMS

Classifiers	Precision	Recall	Accuracy	F1 score
DT	.73	.5	72.5	.68
SVM	.94	.98	96.6	.96
kNN	.82	.78	81.9	.77
RF	.79	.75	79.2	.76

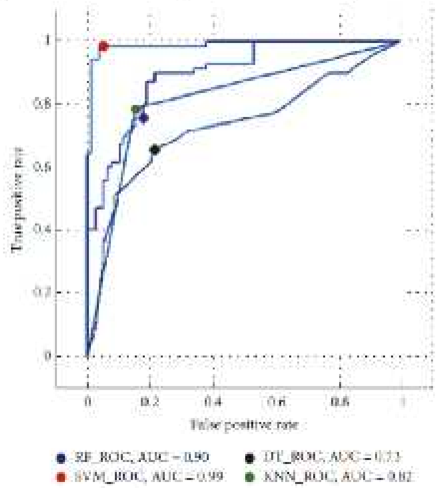


Fig 2: ROC curve for Accuracy **(Courtesy**: Ufera Shafi, *et al* Jan 2018: Vehicle Remote Health Monitoring and Prognostic Maintenance system [14])

For failure prediction of ignition system, selected four classifiers Decision Tree (DT), SVM, K-NN and Random Forest are applied on the dataset as shown in Table 1. ROC curve is basically a graphical representation which is used to compare accuracy of different classifiers as can be seen from Figure 2. The area covered by ROC curve shows the measure of accuracy.

From Table1 and Figure 2, we can see that SVM trumps the other classifiers with 96.6% accuracy.

Similarly, ROC comparison was performed for predicting failure for other parameters like fuel system, exhaust system and cooling system in Table 2. The comparison between accuracies of different classifiers of various parameters are as follows:

Table 2: Classifier's Comparison (**Courtesy**: Ufera Shafi, *et al* Jan 2018: Vehicle Remote Health Monitoring and Prognostic Maintenance system [14])

System VMMS	DT %	SVM %	kNN%	RF%
Ignition system	72.5	96.6	81.9	79.2
Fuel System	76.5	98.5	94.6	90
Exhaust System	78.5	98	89.9	88.6
Cooling System	75.8	96.6	94.6	89.3

It can be clearly observed that SVM shows much better results in every department in comparison to the other classifiers. SVM performs well on smaller observations (150 in this case). Decision Tree considers all dataset and classifies each instances stepwise. The limitation of Decision Tree is that it depends on training data set and therefore, it shows poor performance in test data set.

4.1 Drug Impaired Driving Behaviour (Based on Approach 2)

Table 3: Incidence among automobile accidents and percentage of accidents with injury for various conditions (Structured Data) (**Courtesy**: Predictive Analytics, Text mining and drug impaired driving in automobile accidents by Dr Philip S Borba, 03 Apr 2013 [13])

Condition	Incidence Among Accidents	Percent With Injury	Injury Frequency Compared to "All Accidents"
All accidents	100%	73%	
Time of day/week			
Night	22%	69%	-
Weekend	22%	73%	+
Environment			
Weather	24%	71%	-
Wet roads	16%	67%	-
Nature of accident			
Multiple vehicles	74%	76%	+
Rear-end	18%	70%	-
Head-on	2%	86%	+
Turned into path	16%	81%	+
Driver condition			
Driver fatigued	13%	76%	+
Alcohol (police report)	6%	82%	+

Table 4: Percentage of Accidents with injury due to Medication, Prescription, Drug and Illegal Narcotic (Narrative Data) (Courtesy: Predictive Analytics, Text mining and drug impaired driving in automobile accidents by Dr Philip S Borba, 03 Apr 2013 [13])

Condition	Incidence Among Accidents	Percent With Injury	Injury Frequency Compared to "All Accidents"
All accidents	100.0%	73%	
Medication	15.7%	82%	+
Prescription	6.4%	80%	+
Drug	6.6%	80%	+
Illegal narcotic	2.4%	89%	+

According to Table 3, out of all accidents 73% were injurious. The injury frequency of accidents occurred in different conditions are compared with all accidents. If the value is greater than or equal to 73%, the injury frequency is taken as '+', else '-' sign is provided.

The Narrative data of Incidence rates in automobile accidents and percent of accidents with injury due to presence of Medication, Prescription, drug or illegal narcotic is as follows

From Table 4, it can be observed that level of injury that occurred during accidents under the influence of drugs, medication etc. are much higher as compared to overall accidents. Using Table 3 and Table 4, i.e. combining the structural as well as narrative data, we can determine whether the influence of drugs really have an impact on accidents which are commonly occurring and if so by how much.

4.2 Analysis of Driver Behaviour (Based on Approach 3)

Drivesafe app performs behaviour analysis, real time maneuver detection, scoring. From the table, the maneuver scores for each driver part provide the scores provided by Drivesafe. The minimum of the attributes are marked in bold. The Behaviour part contains the ratios provided by Drivesafe for normal, aggressive and drowsiness behaviour. All scores provided are in base 10. The scores for acceleration, braking and turning usually depends on driver profile and road conditions. However aggressive driving behaviour will result in abruptness. Therefore, the scores are usually lower as compared to the other driving behaviours, as can be observed from the table. Lane weaving analysis can help us identify whether the driver is changing lanes slowly and carefully or abruptly and with lack of awareness. Drifting analysis is beneficial to understand the capacity of the driver to continue centred on its own lane. Both the scores of Weaving and drifting are observed to be lower in drowsy driving behaviour compared to others. Car following is a very important. The driver has to maintain a safe distance behind the other vehicle in case the vehicle brakes abruptly he/she should have time to react. For aggressive driving behaviour, both over speeding and car following scores are low.

From Table 5, it can be clearly seen that the behaviour detected by Drivesafe was 100% correct on secondary routes, while it was approximate in case of Motorways. E.g. : In case of D2 on Drowsy (Motorway), the score for Normal Behaviour was given 4.2 while for Drowsy Behaviour, score of 4.1 was obtained (which was supposed to be the actual one).

Table 5: Scores of Drivers for Simulated Driving Behaviour in Different Routes **(Courtesy**: Eduardo Romera *et al*: Need Data for Driver Behavior Analysis [12])

State	Driver	Duration		Speed (Km/h)		Maneuver scores							Behavior		
		Time	Km	Avg	Max	Acc	Bra	Tur	Weav	Drift	Overs	Carfoll	Nor	Drow	Agg
Normal (Motorway)	D1	14m.	25	107	131	10	9.7	8.7	9.3	**7.9**	9.4	9.8	**6.8**	1.4	1.8
	D2	15m.	26	98	127	9.9	9.9	**7.2**	10	7.5	9.6	9.3	**6.8**	1.5	1.7
	D3	15m.	26	101	122	10	9.9	9.4	9.4	**8.1**	9.7	9.8	**7.3**	1.3	1.4
	D4	16m.	25	91	120	9.9	9.9	9.7	10	**8.9**	9.9	9.9	**8.2**	0.6	1.2
	D5	15m.	25	99	120	9.0	9.4	**7.8**	10	8.0	9.3	9.1	**6.8**	1.2	2.0
	D6	17m.	25	89	104	9.7	9.7	**3.5**	10	8.7	9.8	9.7	**8.0**	0.8	1.2
Drowsy (Motorway)	D1	15m.	25	97	113	10	3.8	6.9	**2.6**	4.3	9.7	9.7	3.2	**5.6**	1.2
	D2	15m.	25	98	122	9.4	4.8	7.8	5.2	**4.7**	9.7	9.4	**4.2**	4.1	1.6
	D3	16m.	26	91	129	9.8	10	7.9	**1.5**	5.2	9.7	9.9	2.6	**6.0**	1.4
	D4	17m.	25	88	106	9.9	9.8	8.7	**4.1**	4.6	9.0	9.9	3.8	**4.6**	1.6
	D5	18m.	25	83	96	8.6	4.2	8.2	**0.9**	3.1	9.5	9.9	1.8	**6.8**	1.3
	D6	17m.	25	84	99	9.6	9.2	**1.8**	3.9	4.8	7.1	9.9	2.5	**4.7**	2.8
Aggressive (Motorway)	D1	12m.	24	120	148	10	7.0	8.1	10	8.5	**6.1**	9.1	**5.1**	0.9	4.0
	D2	14m.	26	107	147	6.6	5.9	6.6	9.2	5.7	6.7	**2.1**	1.2	2.7	**6.1**
	D3	13m.	26	110	146	9.1	**0.0**	9.4	10	8.0	6.9	6.5	**5.4**	1.2	3.4
	D4	15m.	25	97	130	6.8	**2.7**	8.5	9.0	8.6	8.3	3.3	3.7	1.0	**5.3**
	D5	13m.	25	114	147	7.8	2.4	1.3	10	7.7	6.1	**0.3**	1.3	1.4	**7.3**
	D6	15m.	25	101	127	6.4	5.3	**0.0**	10	8.9	8.4	4.4	**4.8**	0.6	4.6
Normal[1] (Secondary)	D1	10m.	16	96	116	10	10	8.7	10	**6.5**	7.5	9.8	**6.4**	1.5	2.1
	D2	10m.	16	91	103	9.9	10	10	10	**7.4**	7.8	9.9	**6.2**	1.5	2.3
	D3	11m.	16	85	97	9.9	10	10	10	**6.9**	9.6	9.8	**6.9**	1.9	1.2
	D4	11m.	16	82	101	10	10	9.5	10	**8.8**	9.6	10	**9.1**	0.7	0.2
	D5	11m.	16	84	102	9.4	9.9	9.5	10	**7.3**	9.4	8.9	**7.6**	1.6	0.8
	D6	13m.	16	75	90	9.9	9.7	**4.5**	10	9.2	9.9	10	**9.5**	0.4	0.0
Drowsy (Secondary)	D1	8m.	13	94	107	10	4.9	6.6	10	**2.8**	7.7	10	3.3	**4.3**	2.4
	D2	10m.	16	91	110	8.8	3.8	8.1	**0.0**	4.1	8.5	9.6	0.9	**7.2**	1.9
	D3	10m.	17	91	118	10	9.4	9.5	**0.0**	4.0	8.1	9.9	0.7	**7.2**	2.1
	D4	11m.	17	87	102	9.9	9.1	8.1	**2.0**	3.9	9.4	9.9	1.8	**6.0**	2.2
	D5	11m.	16	84	100	10	9.7	4.8	10	**1.4**	9.8	9.2	3.4	**5.1**	1.5
	D6	12m.	16	80	94	8.7	8.8	2.5	**0.0**	4.6	9.9	10	1.4	**7.1**	1.5
Aggressive (Secondary)	D1	8m.	16	112	132	10	2.9	5.7	10	5.9	**0.0**	9.5	0.5	2.4	**7.1**
	D2	10m.	16	96	119	7.2	3.7	10	10	5.8	**0.2**	0.7	0.0	2.5	**8.7**
	D3	11m.	16	87	119	8.4	8.2	8.6	10	6.4	7.3	**1.5**	1.5	2.1	**6.4**
	D4	10m.	16	89	113	6.8	8.0	10	10	6.9	8.0	**2.3**	1.8	1.9	**6.3**
	D5	7m.	12	100	147	9.0	0.1	6.2	10	5.0	**0.0**	4.6	0.0	3.0	**8.0**

5. Conclusion

As vehicles are increasing manifold throughout the year, the risk of road accidents occurring will also be on the rise. Driver behaviour plays a massive role in determining the risk of accidents that can occur. The study of different types of behaviour like normal, aggressive and drowsiness shows the proof. In addition to that Driving Under the Influence of drugs is a huge matter of concern regarding the severity of road accidents. The approaches adopted in this research paper has shown that from the accidents that occurred during different conditions (structured data) and the accidents that occurred due to intake of drugs, medication etc (narrative data), drugs have a huge impact on the rise in amount of accidents and are more fatal. Four classifiers Decision Tree, SVM, K-NN and Random Forest have been used for fault prediction. Even though the objective is to reduce the fault frequency of systems in vehicles, by doing so we can reduce the amount of road accidents that can occur due to malfunction of the components. Researches are still going on in these areas as required algorithms are not easy to implement practically.

References

[1] AbuAli Najah, Abou-Zeid Hatem, 2016, Driver Behaviour Modelling: Developments and Future Directions

[2] Akbar Khushanoor, Baseer K.K., Anil Kalaga, 2018: Improving the Efficiency of Automotive Service with Recovery Analytics.

[3] Alluhaibi Sarah Kadhim, Munaf S. Najim Al-Din, Moyaid Aiman, 2018, *et al*: Driver Behaviour Detection Techniques: A survey.

[4] Babu, G. S., Zhao, P., & Li, X.-L. (2016): Deep convolutional neural network based regression approach for estimation of remaining useful life in *International conference on database systems for advanced applications* (pp. 214–228).

[5] Byttner S., R¨ognvaldsson T., and Svensson M.: "Consensus self-organized models for fault detection (COSMO)," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 5, pp. 833–839, 2011.

[6] Centre for advanced Automotive Technology, 2018

[7] De Winter J.C.F., Dodou D., 26 Nov 2010, *et al:* The Driver Behaviour Questionaire as a predictor of accidents: A Meta-Analysis.

[8] Gu Jie, Vichare Nikhil, Tracy Terry, Michael Pecht, 2007, *et al*: Prognostics Implementation Methods for Electronics.

[9] K.Choi,J.Luo, K.R.Pattipati, S.M.Namburu, L.Qiao,and S.Chigusa, "Data reduction techniques for intelligent fault diagnosis in automotive systems," in Proceedings of the 2006 IEEEAUTOTESTCON - IEEE Systems Readiness Technology Conference, pp., USA, September.

[10] Kim Jonghyuk, Hwangbo Hyunwoo and Kim Soyean, 03 Jan 2018: An empirical study on real-time data analytics for connected cars: Sensor based applications for smart cars.

[11] Meiring Gys Albertus Marthinus and Myburgh Hermanus Carel, Dec 2015: A Review of Intelligent Driving Style Analysis Systems and Related Artificial Intelligence Algorithms

[12] Romera Eduardo, Bergasa M. Luis, Arroyo Roberto, Nov 2016, *et al:* Need Data for Driver Behaviour Analysis? Presenting the Public UAH - Driveset.

[13] S. Borba Philip, 03 Apr 2013, Predictive Analytics, Text mining and Drug-impaired Driving in Automotive Analytics

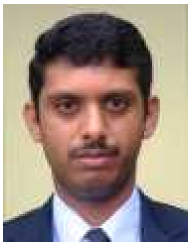
[14] Shafi Ufera, Safi Asad, Shahid Ahmad Raza, Ziauddin Sheik, Saleem Muhammad Qaiser, 18 Jan 2018: Vehicle Remote Health Monitoring and Prognostic Maintenance System. (https://www.researchgate.net/publication/307628008)

[15] Wahab D.A., Amelia L, Hooi N.K., Che Haron C.H., Azhari C.H., 01 Dec 2008: The Application of Artificial Intelligence in Optimisation of Automotive Components for Reuse.

[16] West Robert, Kemp Richard, Elander James, 1993, *et al*: Direct Observation of Driving, Self-Reports of Driver Behaviour, and Accident Involvement.

[17] Worley Heidi, Population change – Road Traffic Accidents Increase Dramatically Worldwide, March 1, 2006

Author's Biography



Jaidev B is currently pursuing Post Graduate Diploma in Management in Research and Business Analytics at Lal Bahadur Shastri Institute of Management, Delhi, India. He is also a Mechanical Engineering graduate from Sree Narayana Gurukulam College of Engineering, Kadayiruppu, Kerala. He has done academic project on the topic "Welding Processes used in Shipbuilding and Development for Welding Procedure Specification for Duplex Stainless Steel" from Cochin Shipyard Ltd., Kerala.

Sonakshi Garg is currently pursuing Post Graduate Diploma in Research and Business Analytics at Lal Bahadur shastri Institute of Management , Dwarka, New Delhi. Degree in Bachelor of commerce from University of Delhi. Earlier, presented a paper on green purchasing behavior of young consumers and opportunities for marketers at 5th Annual International Commerce conference, 2016 in Vallabhbhai Patel Chest Institute, University of Delhi.

Sandhya Makkar is Sr. Assistant Professor(operations) at Lal Bahadur Shastri Institute of Management, Delhi. She has more than 12 years of teaching and research experience. She is editor of many international books and journals. She has published umpteen number of research articles and papers. Her areas of interest are machine Learning, Optimization and Supply Chain Management.

How to Cite

B Jaidev, Garg Sonakshi, Makkar Sandhya (2019). Artificial Intelligence to Prevent Road Accidents. *International Journal of Machine Learning and Networked Collaborative Engineering*, 3(01) pp 35-45.

doi : https://doi.org/10.30991/IJMLNCE.2019v03i01.003.

Vol. 03, No. 1,(2019), 46-65, **ISSN:2581-3242**

IJMLNCE JOURNAL

International Journal of Machine Learning and Networked Collaborative Engineering

Journal Homepage: http://www.mlnce.net/home/index.html

DOI : https://doi.org/10.30991/IJMLNCE.2019v03i01.004

Deterministic Machine Learning Cluster Analysis of Research Data: using R Programming

Yagyanath Rimal

Faculty of Science and Technology

Pokhara University, Nepal

Email: rimal.yagya@gmail.com

https://orcid.org/0000-0003-1045-7728

Abstract

This review paper clearly discusses the compression between various types of cluster analysis of different data sets were explained sufficiently. Although there is large gap between the way of analysis of collected data and its cluster categorization research data using r programming. Its primary purpose is to explain the simplest way of clustering analysis whose data structure were wide scattered using R software whose outputs were sufficiently explain with various inter-mediate output and graphical interpretation to reach the conclusion of analysis. Therefore, this paper presents easiest way of clustering when data sets with large dimensions with multivariate analysis of iris , utilities, mclust and dbscan data sets from internet and its strengths for data analysis using R programming.

Keywords

Data Analytic Machine Learning, K mean, Hierarchical, Dbscan

1. Introduction

Data science is not only data collection as a database, but it is an interdisciplinary science of statistics, data analysis and automatic learning of the scientific methods of process algorithms. The data collection, data processing and data analysis with different algorithms and, visualization of data which require the knowledge of calculation, mathematics, statistics and interpretation in each step to analyze research data. However, the data available in today's world in any structured, unstructured and semi-structured format are raw data. Raw data includes the integration of data and the selection of required data. The author [1] claimed that data cleaning takes 50 percent to 80 percent of a data scientist's work is significant which needs some methods and algorithms to analyzed and then a visual report is produced from data base. The data column is known as data structured. The data scientist creates many unstructured data formats. At one point, the data scientist integrated data from different sources [2]. Then selection of part for analysis, so that the data scientist finally makes the normalization of data with various transformation records. On many cases the data analyzer omits missing data, has been cleaned up while some computer scientists do so. This process needs machine learning for establishing the statistical requirements of the system for automatic processing. Therefore, modern data scientists need four pillars of statistical, computer skills, communication and visualization and competencies of many domain [3]. R programming with supervised learning has the best tools for data analysis; The data scientist can use many other data analysis processes. Machine learning can be applied in the analysis of neural networks, deep learning and artificial intelligence networks. The machine learning always requires information of requirement, since formal entry into the system produces the program and is defined as output according to the requirements of the system. Therefore, machine learning is the system that takes input and output as input, so the computer system automatically produces the model based on some previous parameters. For the simplest example, x entry and y is exit of given table, so find that the prediction at 4 is 40 with a relation having y = x * 10. Based on static observation, machine learning predicts its relationship. Similarly, example when x is input z is another output, so it predicts that its relation

```
> fitk=kmeans(irs,3)# here we are using 3 cluster of species
> fitk
```

Table 1: Data

X	Y	Z
1	10	14
2	20	18
3	20	22
4	?	?

(14 + 4) between x and z is 26. So, the machine learning will be predicting the value of z when x is 500. This could be easily calculated with y = mx + c where m is slope 4, which is calculated with the relation y2 y1=x2 x1 and c is the intercept with the constant value when the line of mean intersection of the slope on y is 10. Therefore, the prediction when x becomes 500, the value of z becomes 2010 (4X + 10) relationship.

The data science work flow includes the process of defining cyclical problems, collecting data, developing models, implementing models, improving and monitoring performance. According to [7] the R, Python and Weka are the most important technologies used in science in modern world. Therefore, machine learning strategies first, learns the relationship and then predicts the relationship [4]. First data processing phase is data processing and then the processing data were translated into relationship for other data prediction. This process is known as the conversion of data into a statistical analysis design. The second step is to send the model, the third step is to evaluate the performance that is always compared with the output data in a rigorous process, which is accurately evaluated against errors. There were some techniques to improve performance are classification, grouping and regression in machine learning [1]. Machine learning technologies use R an open source programming language developed in 1993 by Ross Ilha and Robert Gentleman with dynamic programming [5]. which supports large complete functional package is installed in the r environment for a special package. R is a binary package, which does not require compilation, but must be connected to the r console with a different library for special packages. Clustering is one of the most important methods for extracting data to discover knowledge in multidimensional data. The purpose of grouping is to identify models or groups of similar objects. Cluster analysis is a group of data objects based on the information found in objects and data groups. The grouping means increasing the grouping distance among research data sets with various another field relationship. The big data analysis, supply chains network analysis, data interpretation predicts the future are the key application of data clustering. In many times running organization can change plans when the results are changed and success is achieved after data analysis [6]. Hierarchical grouping is also known as nested groupings that are groupings to form a tree, which uses local heuristics to form a nesting hierarchy of nested datasets. The most notable exception of the hierarchical grouping of a single link is distorted into convex and hyper spherical groupings [7]. The advantages of hierarchical clustering are low efficiency, as it has a time complexity of O(n), unlike the linear complexity of K-Means and GMM.

Partitioned cluster is simply the division of grouped data objects that do not overlap, so that each object is exactly in a subset. The exclusive grouping assigns each object to a single grouping. The stacking grouping used to reflect the fact that grouping widespread in this grouping. The hierarchical cluster analysis uses a set of differences in the object being grouped with the objectives of compact groups of appropriately the same diameter. The well ranked cluster mapped with distance between any two points within a group. The prototype cluster is often measuring with continuous data attributes, since clutters tend to be globular.

The density-based cluster is used when groups are irregular and with noise in data sets values is grouped up. Therefore, grouping, is the process of identifying the model in groups for effective segregation into groups of large data groups with various relationship among them. The most significant clustering is K means which group the data based on prototypes that try to define the number of groups (k) that are represented as centroids when the samples were more than 100. The agglomerated hierarchical grouping is closed groupings until a single and global grouping is maintained. Similarly, Dbscan is a density-based clustering algorithm that produces a particular grouping, in which the algorithm determines the number of clusters [7].

```
> data("iris")
Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
1          5.1        3.5          1.4         0.2     setosa
2          4.9        3.0          1.4         0.2     setosa
3          4.7        3.2          1.3         0.2     setosa
> plot(iris)
```

2. Using R Programing

For detail analysis here researcher tries to the explanation with iris data sets with records of 150 observations in 5the categories.

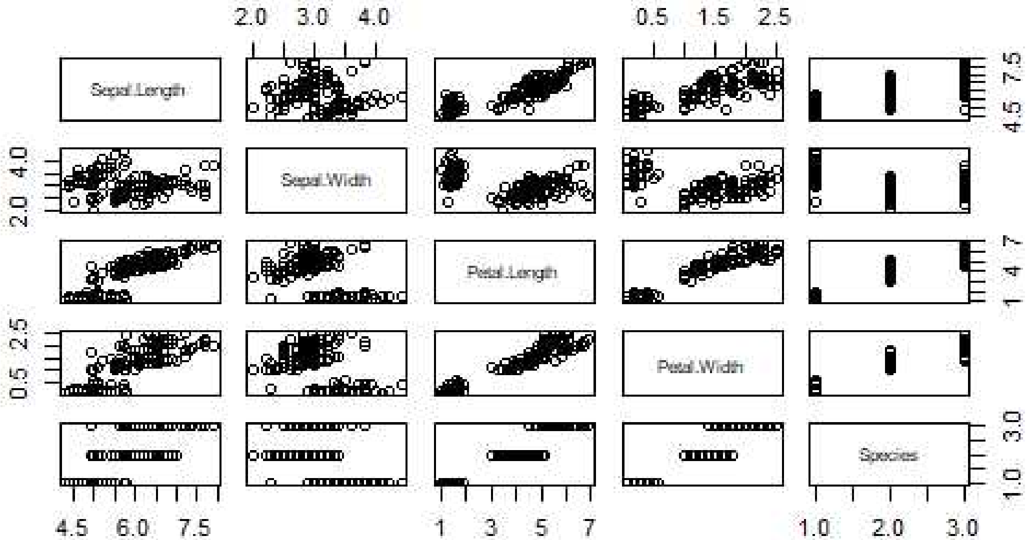


Fig1: Edgar Adelson's Pair Scatter Plot

Fig1 shows the multidimensional figure of five variables of all the recorders were shown. mostly the last field species clearly demonstrates three categories most. Speal length, sepal width petal length petal width were numeric continuous variable records. Where there were some records were scattered in some cluster but how many clusters were fit those variables is primary of researcher judgement. The scaling and normalization are always required when the data set have large variation in data structure. If in case some data were in cm and some data were in m or km. in such case, we need mean and standard deviation of those data should be calculated. But there the last datasets species have three categories of species name should be excluding categorical variable form the rest of data sets using scale function.

```
> irs=scale(iris[,-5])  all row and columns except 5th
> irs
  Sepal.Length Sepal.Width Petal.Length   Petal.Width
  [1,]  -0.89767388  1.01560199  -1.33575163 -1.3110521482
   ……………………………………………………
  [150,]  -1.13920048 -0.13153881  -1.33575163 -1.3110521482
```

The kmean clustering is the process of finding k means center of data where each data is centered from k mean using kmean function. Each center searches the distance between observed value were grouped into based on data relationship to each cluster. All the data points were falls into k means of clusters. If we set out 4 k means it developed 4 clusters data groups from research data.

```
> fitk=kmeans(irs,3)# here we are using 3 cluster of species
> fitk
```

K-means clustering with 3 clusters of sizes 47, 53, 50 is the number of observations within each cluster. Cluster means:

```
Sepal.Length Sepal.Width Petal.Length Petal.Width
1   1.13217737  0.08812645    0.9928284   1.0141287
2  -0.05005221 -0.88042696    0.3465767   0.2805873
3  -1.01119138  0.85041372   -1.3006301  -1.2507035
```

```
Clustering vector:
[1] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 33 3 1
…..
1 1 1 2 1 1 1 1 1 1 2 2 1 1 1 1 2  1 2 1 2 1 1 2 1 1 1 1 1 1 2 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 2
Within cluster sum of squares by cluster:
[1] 47.45019                  44.08754             47.35062
(between_SS / total_SS =  76.7 %)
```


```
> irs=scale(iris[,-5])  all row and columns except 5th
> irs
  Sepal.Length Sepal.Width Petal.Length   Petal.Width
  [1,]  -0.89767388  1.01560199  -1.33575163 -1.3110521482
   ……………………………………………………
  [150,]  -1.13920048 -0.13153881  -1.33575163 -1.3110521482
```


Which implies how well the cluster dispersed among and within the data sets the 76.7% higher the percentage means higher the dispersed among data sets. Available components:

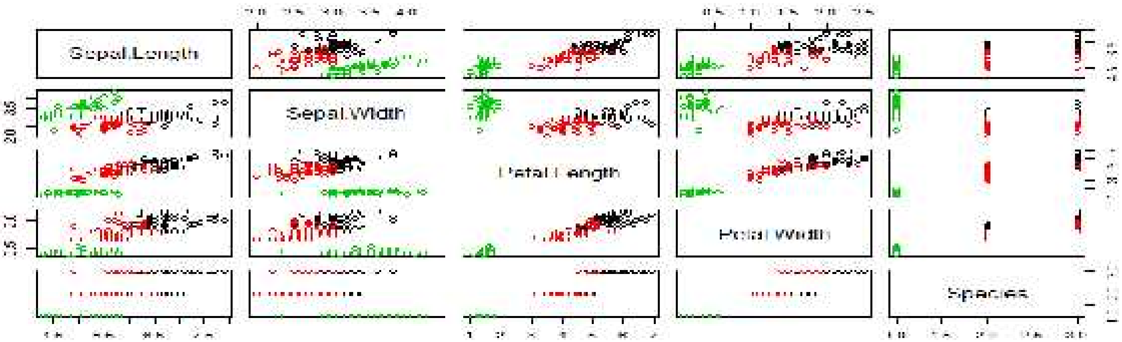


Fig2: Scatter Plot With All Cluster

```
[1] "cluster" "centers"  "totss" "withinss" "tot.withinss" "betweenss" "size" "iter"  "ifault"
> str(fitk) display the structure.    List of 9
 $ cluster     : int [1:150] 3 3 3 3 3 3 3 3 3 3
 $ centers     : num [1:3, 1:4] 1.1322 -0.0501 -1.0112 0.0881 -0.8804
attr(*, "dimnames")=List of 2
$ : chr [1:3] "1" "2" "3"
$ : chr [1:4] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width"
$ totss        : num 596
$ withinss     : num [1:3] 47.5 44.1 47.4
$ tot.withinss: num 139
$ betweenss    : num 457
$ size         : int [1:3] 47 53 50
$ iter         : int 3
$ ifault       : int 0
attr(*, "class")= chr "kmeans"
>  plot(iris,col=fitk$cluster)
```

The above table demonstrates the three cluster with different colors the nearest data were in same colors than others data structure which separate from another. Here is great question how many clusters does it optimum fits for those records. To do so the empty list of k is designed and repeat the same procedure of kmean into list process is carried out.

```
k=list()
for(i in 1 :10){
k[[i]]=kmeans(irs,i)}
k
```

Which produces the 10 integration and it produced the sum of square and total sum of square of each reputation as below the k mean fit where there was gentle slope of data flow.

```
> plot(1:10,betweenss_totss,type="b",ylab="between ss/total SS",xlab="center(k)")
```

```
[[1]]
K-means clustering with 1 clusters of sizes 150.Cluster means:
Sepal.Length  Sepal.Width Petal.Length   Petal.Width
-9.793092e-16       4.503805e-16       5.107026e-17       -6.217249e-17
Clustering vector:
[1] 1 1 1 1 1 1 ……… 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Within cluster sum of squares by cluster:
[1] 596
(between_SS / total_SS =   0.0 %)
Available components:
[1] "cluster" "centers"  "totss" "withinss" "tot.withinss" "betweenss"  "size"  "iter" "ifault"
[[10]]
K-means clustering with 10 clusters of sizes 21, 16, 12, 17, 13, 17, 16, 12, 11, 15
Cluster means:
  Sepal.Length Sepal.Width Petal.Length  Petal.Width
1    -0.9666815  0.92820079  -1.29259152 -1.217343093
………………
10    0.9942845  0.35790793   1.04344975  1.452740239
Clustering vector:
[1]  1  4  4  4  1  8  1  1  4  4  8  1  4  4  8  8  8  1  8  8  1  1  1  1  1  4  1  1  1  4  4  1   4  4  1 1
  ………..
 9  6 10  2  9  2 10  9  2  7  2  9  9  9  2  3  2  9 10  7  7 10 10 10  2 10 10 10  2  7 10  7
Within cluster sum of squares by cluster:
[1]  3.397867  5.606072  1.858918  5.163861  2.347768  9.671551  5.324952  3.954505 10.2
03534  4.597857
(between_SS / total_SS =  91.3 %)
Available components:
[1] "cluster"   "centers"   "totss" "withinss"  "tot.withinss" "betweenss" "size" "iter"  "ifault"
> betweenss_totss=list()
> for(i in 1:10){
betweenss_totss[[i]]=k[[i]]$betweenss/k[[i]]$totss }
```

This process calculates the 10 ratios of total sum of square in between_totss of k index whose output is further plot and produce with type is b means both dots line type in between y lab and x lab.

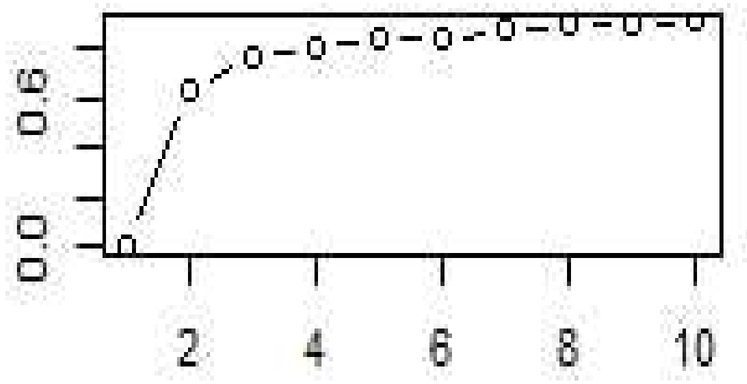


Fig:3 Best fit Cluster Plot

This picture shows the best fit of cluster in k means on x axis and between sum of total sum of square is in y axis, when first sharp drop-down point list between two points were neglected when there is normality between adjoin two points is the best fitted line area of k group. Here the two to four k mean considered as the best cluster groups is best fitted. Similarly, the 4 k mean cluster plot could be easily plot as.

```
>for(i in 1:4){ plot(iris,col=k[[i]]$cluster) }
```

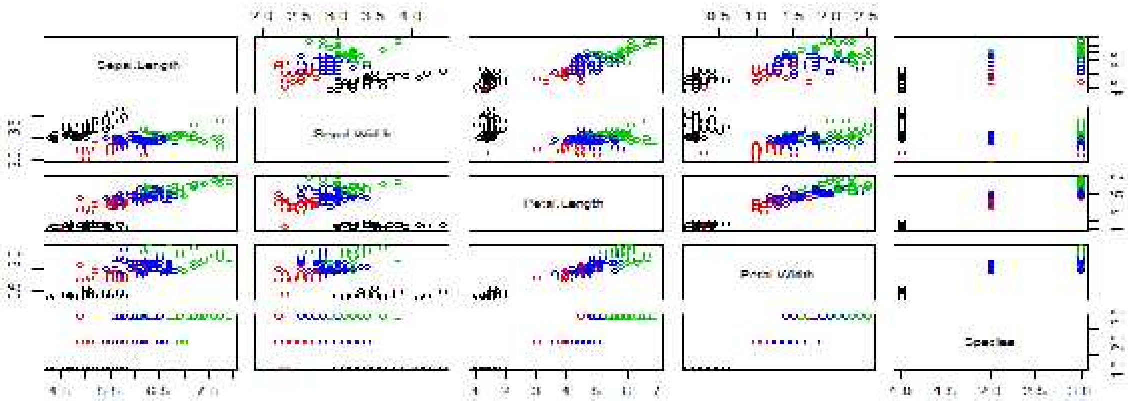


Fig 4: K Means Four Fitted Cluster Plot

```
>for(i in 1:3){  plot(iris,col=k[[i]]$cluster) }
```

This demonstrate four k means fitted model cluster produced but there is highly overlapped one another which is suite for the iris data sets where red, black and green points were mixed so this is not best fitted cluster of this data sets.

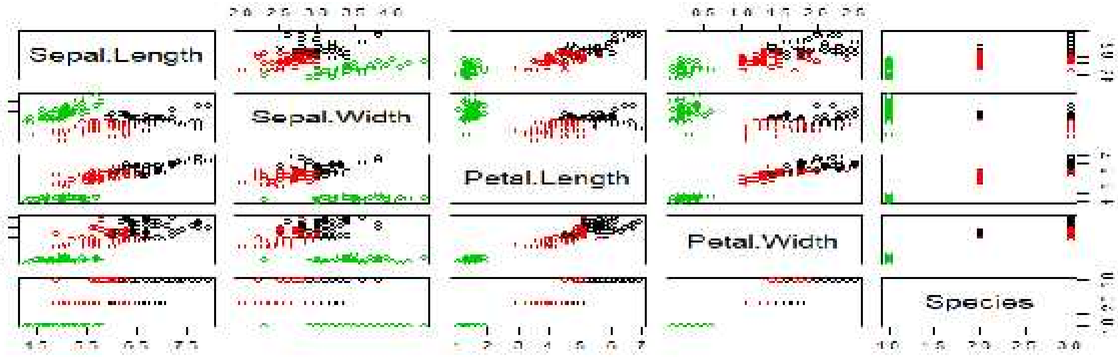


Fig 5: K Mean Three Cluster Scatter Plot

This cluster is quite best but there was less overlap than above figure when clustering the principle behind that the intra cluster dependency should consider as less one another in many cases the red and black cluster groups were found mixture in fixed side although.

```
>for(i in 1:2){ plot(iris,col=k[[i]]$cluster) }
```

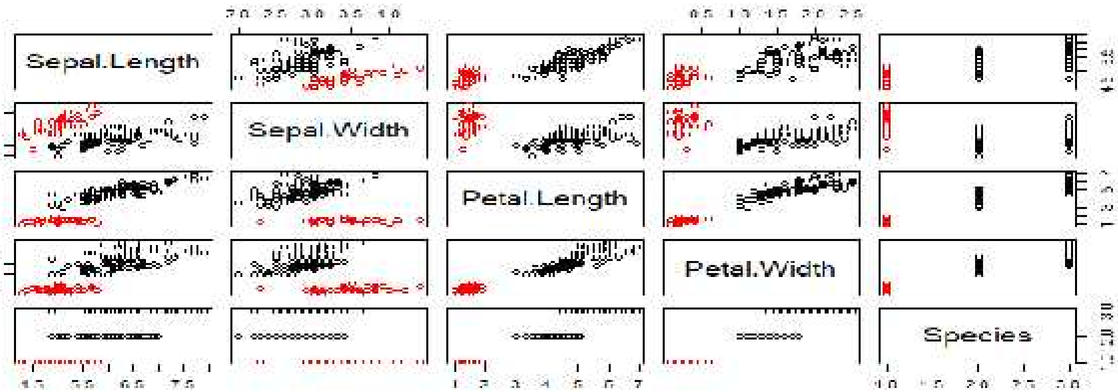


Fig: 6 K Mean Two Cluster of Scatter Plot

Figure conclude that the most fitted k mean is 2 which is generalized the figure and data structure of specific species data correlation among various data types on the data sets. Where each group are completely separated another group is best perfect than other 4 and three k mean.

3. Hierarchal Clustering

The hierarchal clustering places each data point in hierarchal order of its weight which measures each item in hierarchal **pattern [8].** It also measures how each point are similar to each other in tree format. The distance measures of every point of data sets using dist. function are merged typically to formed hierarchal tree pattern rather than one cluster to another cluster.

```
>d=dist(irs)
>fith=hclust(d,"ward.D2")
>plot(fith)
```

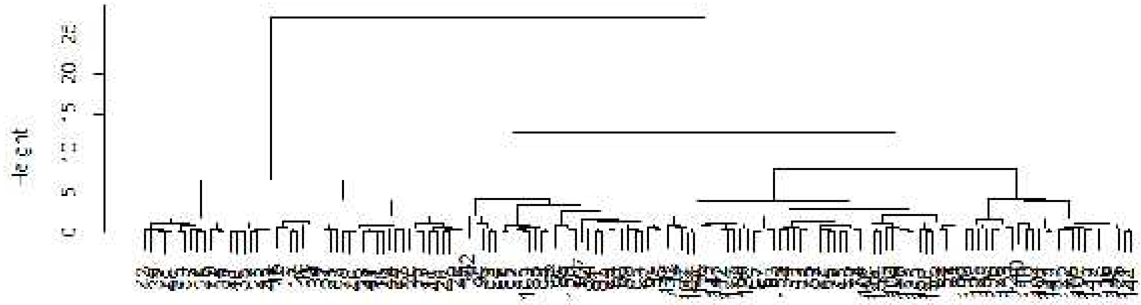


Fig: 7 Cluster Hierarchal Plot

From the above figure each similar observation is in lower cluster then the higher cluster was again connected with above cluster finally formed largest at top. Demonstrate higher cluster in top as parent cluster of another smaller child. The data scientist has to design where to cut to get appropriate data proportion. The lower down the cluster cut the more the cluster we get from the iris data sets. To fit the data structure in fixed cluster cut tree function is used.>rect.hclust(fith,k=3,border="red")

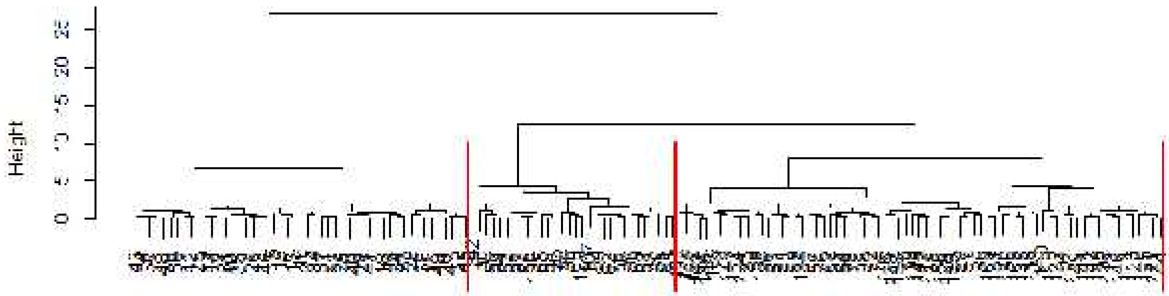


Fig: 8 Hierarchal cluster with Border Plot Dendrogram

From the above figure the three clusters will be easily selected when we cut at the red line point.

```
> cluster=cutree(fith,3)
> cluster
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1
1 1 1 1 3 3 3 2 3 2 3 2 3 2 2 3 2 3 2 3 2 2  2 2 3 3 3 3 3 3 3 3 3 2 2 2 2 3 2 3 3 2 2 2 2 2 3 2 2
2 2 2 3 2 2 3 3  3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
> plot(iris,col=cluster)
```

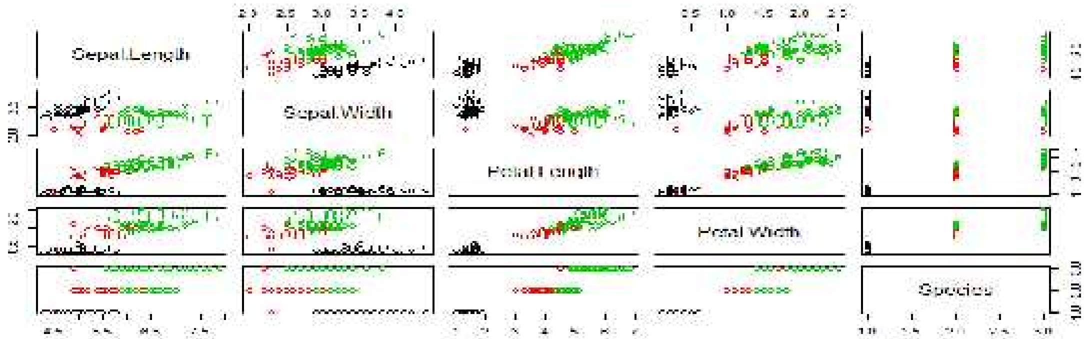


Fig: 9 Scattered Plot of Three Cluster


```
> cluster=cutree(fith,3)
> cluster
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1
1 1 1 1 3 3 3 2 3 2 3 2 3 2 2 3 2 3 2 3 2 2  2 2 3 3 3 3 3 3 3 3 3 2 2 2 2 3 2 3 3 2 2 2 2 2 3 2 2
2 2 2 3 2 2 3 3  3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
> plot(iris,col=cluster)
```


Fig: 9(a) Scattered Plot of Three Cluster


```
> cluster=cutree(fith,3)
> cluster
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1
1 1 1 1 3 3 3 2 3 2 3 2 3 2 2 3 2 3 2 3 2 2  2 2 3 3 3 3 3 3 3 3 3 2 2 2 2 3 2 3 3 2 2 2 2 2 3 2 2
2 2 2 3 2 2 3 3  3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
> plot(iris,col=cluster)
```


Fig: 9(b) Scattered Plot of Three Cluster

```
> cluster=cutree(fith,3)
> cluster
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1
 1 1 1 1 3 3 3 2 3 2 3 2 3 2 2 3 2 3 2 3 2 2  2 2 3 3 3 3 3 3 3 3 2 2 2 2 3 2 3 3 2 2 2 2 3 2 2
 2 2 2 3 2 2 3 3  3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
> plot(iris,col=cluster)
```

Fig: 9(c) Scattered Plot of Three Cluster

```
> cluster=cutree(fith,3)
> cluster
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1
 1 1 1 1 3 3 3 2 3 2 3 2 3 2 2 3 2 3 2 3 2 2  2 2 3 3 3 3 3 3 3 3 2 2 2 2 3 2 3 3 2 2 2 2 3 2 2
 2 2 2 3 2 2 3 3  3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
> plot(iris,col=cluster)
```

Fig: 9(d) Scattered Plot of Three Cluster

```
> cluster=cutree(fith,3)
> cluster
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1
 1 1 1 1 3 3 3 2 3 2 3 2 3 2 2 3 2 3 2 3 2 2  2 2 3 3 3 3 3 3 3 3 2 2 2 2 3 2 3 3 2 2 2 2 3 2 2
 2 2 2 3 2 2 3 3  3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
> plot(iris,col=cluster)
```

Fig: 9(e) Scattered Plot of Three Cluster

```
> cluster=cutree(fith,3)
> cluster
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1
 1 1 1 1 3 3 3 2 3 2 3 2 3 2 2 3 2 3 2 3 2 2  2 2 3 3 3 3 3 3 3 3 2 2 2 2 3 2 3 3 2 2 2 2 3 2 2
 2 2 2 3 2 2 3 3  3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
> plot(iris,col=cluster)
```

Fig: 9(f) Scattered Plot of Three Cluster

This figure demonstrates the three-clustering using hierarchal method is exactly similar in above k mean figure however there were some mixed overlap too when we set with three clusters

3.1. Model Based Clustering

Model based clustering always use various best models in various ellipsoidal, varying volume shape and orientation with mclust library.

```
> library(mclust)
> fitm=mclust(irs)
fitting =====================| 100%
>plot(fitm)
Model-based clustering plots:  1: BIC   2: classification  3: uncertainty 4: density Selection: 1
```

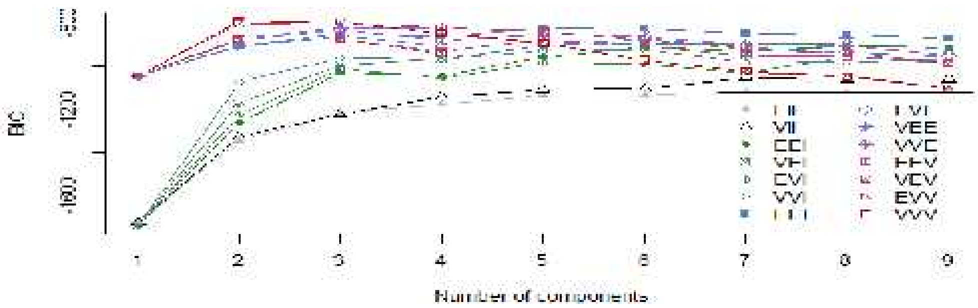


Fig: 10 Baysiean Cluster:

The Baysiean information cluster indicates the plot describes in x axis there were number of clusters where as in y axis based on criteria measurement. It further explained the how much variability of model fits he data the higher the value is the best fit the model. The above red cluster is the best model fitted and measured the at two with VVV cluster, describes that the adding the additional data points with cluster doesn't affects the cluster relationship therefore the best two cluster mostly describes the fitted in the data sets.

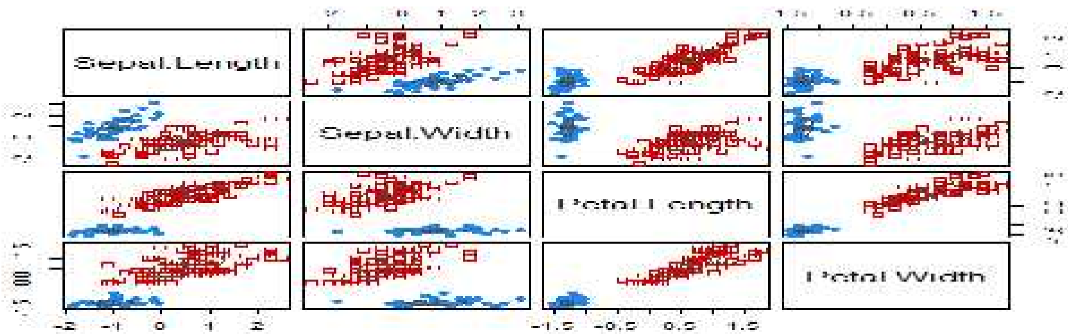


Fig: 11 Model-based Clustering Plots

The classification cluster describes the two different cluster is best fitted where two species where the significant on data sets is the best fitted.

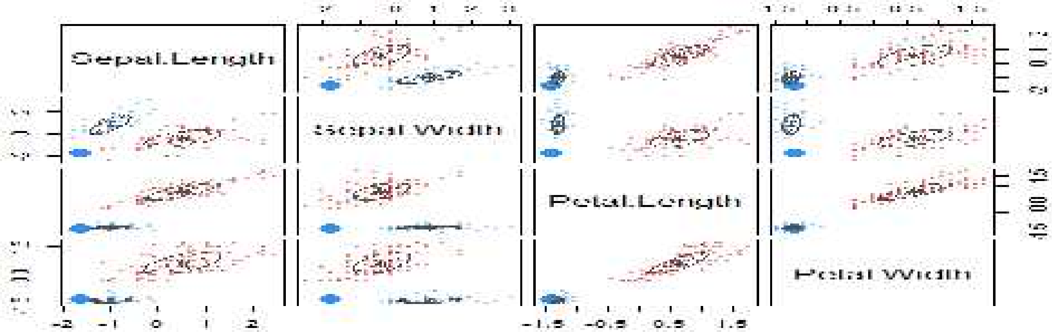


Fig: 12 Uncertainty Cluster Plot

This plot describes the uncertainty plot of subject of data structure the blue points does not fall inside the large circle where there were some cluster does no exactly fit the cluster were many uncertainty points values.

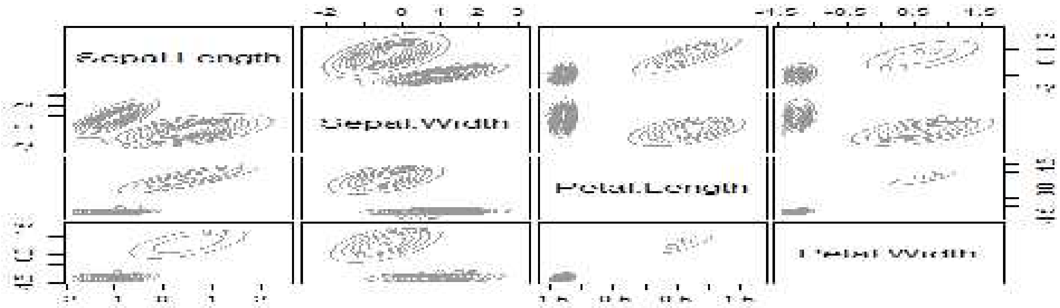


Fig: 13 Density Based Cluster Plot

The density-based cluster is final model of cluster which describes the two bivariate normal distribution of data to formed its circle.

4. Density Based Clustering

Density based clustering is another way of grouping which measures the cluster in which date were mostly falls in region where mostly data values lies [9]. When another cluster again formed when there were mostly likely values lies which can easily be calculated with package dbscan.

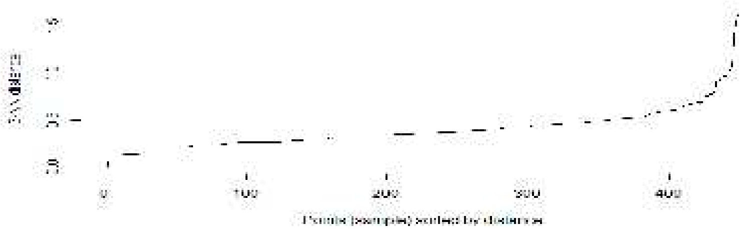


Fig: 14 The Line Graph

```
> install.packages("dbscan")
> library(dbscan)
> kNNdistplot(iris, k=3)
```

This figure describes the knndist distance with three groups of species.
> abline(h=0.7,col="red",lty=2)

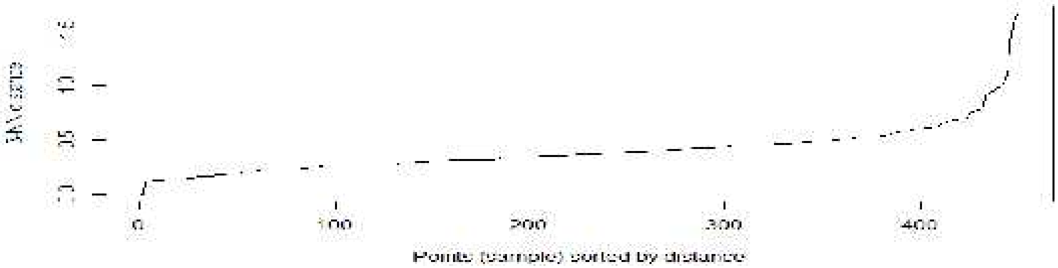


Fig: 15 Abline Plot with Cuts

This figure cuts abline at 0.7 points with.

```
> fitd=dbscan(iris,eps=0.7,minPts = 5)
> fitd
DBSCAN clustering for 150 objects. Parameters: eps = 0.7, minPts = 5. The clustering cont
ains 2 cluster(s) and 6 noise points.   0  1  2
                                        6 48 96
```

Available fields: cluster, eps, minPts. This dbscan nicely groups 2 different cluster of 48 and 96 cases were completely similarities but there were 6 noise cases with dissimilarities out of 150 records of numerical data sets. > plot(iris,col=fitd$cluster)

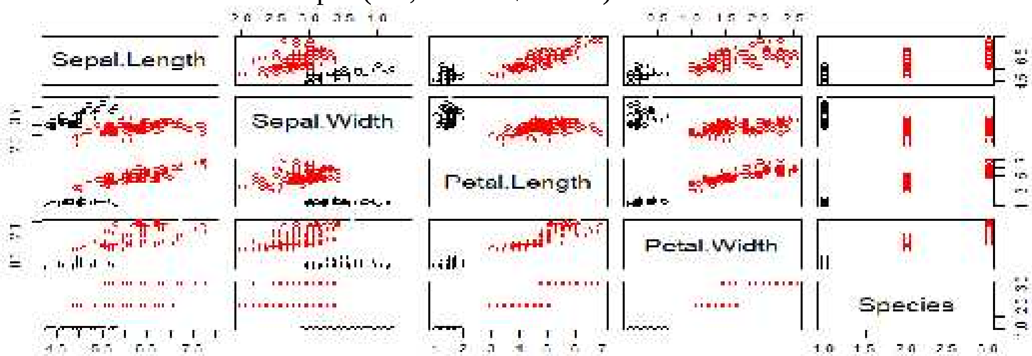


Fig: 16 Hierarchal Clustering with 2Ccluster of Iris

This figure also further same results of kmean and hierarchal clustering when we set out 2 cluster design of iris data sets. There were two groups of data having best clustering. Thus, when we add more variables makes its values with similar with group after elbow is necessary for data analysist. Therefore, principle component reduction techniques support clustering of data sets before data analysing.

Using R Programming

Another data sets of 27 records with 8 variables could analysed other types clustering.

```
>library(readxl)
> Utilities <- read_excel("C:/Users/Rimal Sir/Desktop/Utilities.xls")
> View(Utilities)
> zz=Utilities[,-c(1,1)] # exclude name field
> zz
# A tibble: 27 x 8
  Charge Salary  load demand  fuel gasbill elecbill totalbill
 1     26    680   194    5.5  210   113.      68.2      4.82
 2     18    600   164    5.6  139.   95.9      43.3      4.64
# ... with 25 more rows
> plot(Salary~fuel,data=Utilities) with(Utilities,text(Salary~fuel,labels
=company,pos=1, cex=.5))
```

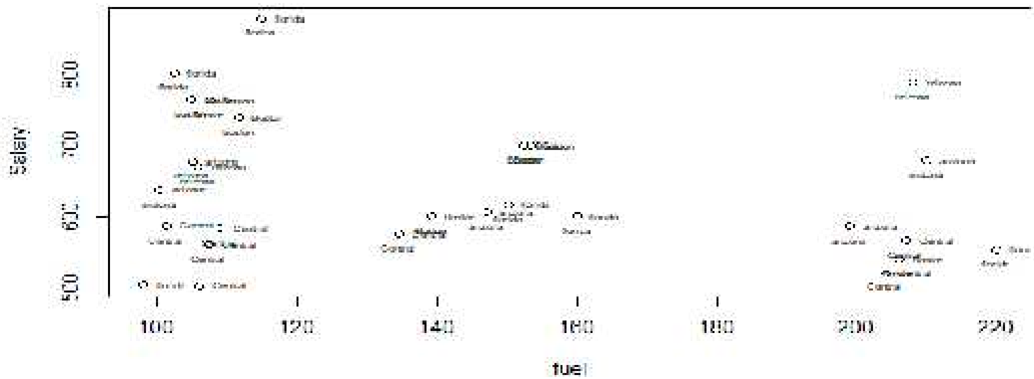


Fig:17 Sales Vs Fuel Consumptions of Utilities

This plot demonstrate sales vs fuel consumpton data of utilities database with name of country of each cluster.pos means position of points and cex is size of text in plot. T here were some company which has high fuel cost but low salary and the middle has averge in both fuel and salary where as one cluster has low to high salary will be one cluster. The normalization is sometime very required if data has high ranges of values this could be achieve using substracting mean and dividing standard deviation of all data sets where company name field should be removed first.

```
> m=apply(zz,2,mean)# 2 for data are in columns
> s=apply(zz,2,sd)# 2 for column
> z=scale(zz,m,s)#normalized dataset of utilities table
> print(z,digit=2)# 2 dugits data only
> print(z,digit=3)
      Charge  Salary    load demand    fuel gasbill elecbill totalbill
 [1,] -1.1773  0.3913  1.4084  1.069  1.5477   1.050   0.3123     0.651
 [2,] -1.5643 -0.4144  1.0187  1.107 -0.0920   0.722  -0.8581     0.505
[27,]  1.0482 -0.5453 -0.9171 -0.832 -0.9669  -0.802   1.0431    -0.401
attr(,"scaled:center")
```

```
> hc.c=hclust(d)
> hc.c
Call: hclust(d = d)
Cluster method   : complete  Distance        : euclidean  Number of objects: 27
> plot(hc.c,hang=1,labels=Utilities$company)
```

```
Charge    Salary     load    demand     fuel   gasbill  elecbill totalbill
50.33    641.15     85.59     2.69    143.15     58.77     61.59      4.00
attr(,"scaled:scale")
Charge    Salary     load    demand     fuel   gasbill  elecbill totalbill
20.67     99.30     76.97     2.63     43.19     51.41     21.32      1.26
> d=dist(z)# eucalidan distance calculation
> print(d,digits=2)# making more compact
1
2  2.27
3  2.52 2.54
4  4.48 4.30 5.62
5  3.98 2.93 5.15 3.20
6  2.39 2.59 3.78 4.04 2.55
7  4.74 3.98 5.79 3.09 1.57 2.95
8  5.41 5.44 7.25 4.67 3.16 3.65 2.81
…………………………………………………………….
27 4.64 4.96 6.01 2.46 3.42 3.51 2.55 3.47 2.27 3.17 2.75 3.93 7.66 2.95
```

The Euclidean distance will be calculated with distance function while print function shows only digits specified so that data become normal looks of all records. The hierarchal cluster dendrogram can be easily calculated with hclust function the default is complete linkage with eculidan distance whose plot can be easily calculated with hang.

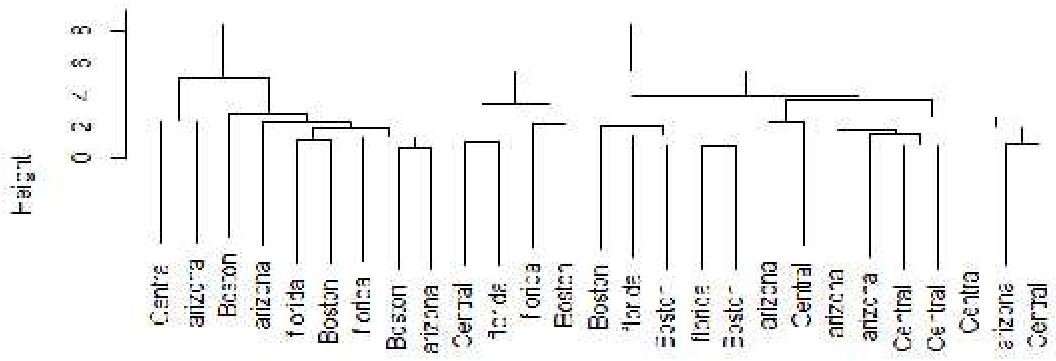


Fig: 18 Cluster Dendrogram of Utilities

The hierarchal dendogram display each company of its variable wastage with its name by its when certs and amazon are in similar data value formed one cluster which similarly goes all data values formed hierarchal clustering.

```
> hc.a=hclust(d,method="average")
> hc.a
Call: hclust(d = d, method = "average")
Cluster method   : average
Distance         : euclidean
Number of objects: 27
> plot(hc.a,hang=1)
```

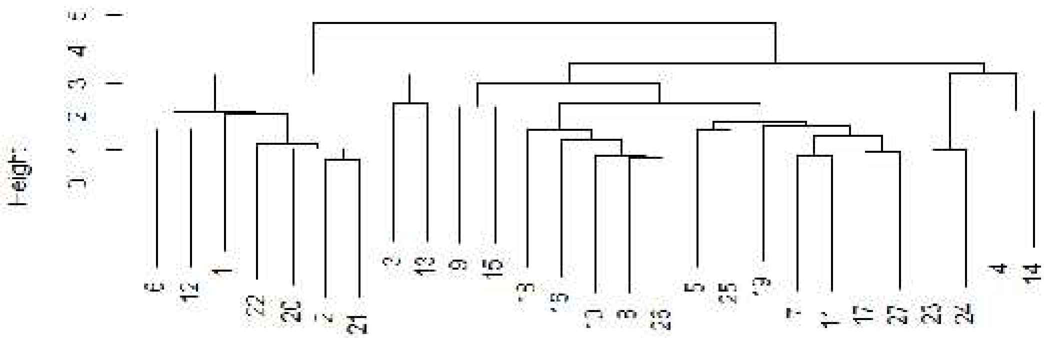


Fig: 19 Cluster Dendrogram of Hang1

From the above figure the company 22 and 23 are formed one cluster then goes subsequently union single cluster tree formed.

```
> m.c=cutree(hc.c,3)
> m.a=cutree(hc.a,3)
> table(m.c,m.a)
      m.a
m.c    1  2  3
     1 9  0  0
     2 0  4  0
     3 0  0 14
```

The cutree function cuts its height at specified location and the table command shows two-dimensional table of data sets presents below clusters cuts. The table show that the average linkage there were 9, 4 and 14 cluster whereas complete cluster there were 9, 4 and 14 clusters in each label, however there were not any company presents both companies. The average of two cluster can calculated with aggregate function.

```
> aggregate(z,list(m.c),mean)
Group.1     Charge      Salary      load      demand       fuel     gasbill    elecbill   totalbill
1     0.1019961  -0.179411  1.271287  1.263528  0.582989  1.214592 -0.587976  1.133909
2    0.0685389 -1.167201 -0.374722 -0.452166  0.332376 -0.472027 -1.275203 -0.879289
3     0.6888436  0.448822 -0.710193 -0.683079 -0.469743 -0.645945  0.742328 -0.477716
```

The above data shows that the fuel has significant among three company varying .33 to -0.46 is lower than average consumption and total bill also has high variation mean with in cluster.

```
> aggregate(Utilities[,-c(1,1)],list(m.c),mean) # in o riginal units
  Group1   Charge   Salary load   demand  fuel   gasbill elecbill totalbill
1. 127.55556 623.3333 183.44444 6.0111111 168.3344 121.20556 49.05778  5.43325
2. 2 51.75000 525.2500  56.75000 1.5000000 157.5100  34.50000 34.40750  2.893158
3. 3 64.57143 685.7143  30.92857 0.8928571 122.8650  25.55929 77.41714  3.399837
```

Similarly, the load 183 is the highest mean of company 1 to 30 lowest mean of company 3. Silhoueplot demonstrate the bar diagram of all cluster members.

```
> library(cluster)
> plot(silhouette(cutree(hc.c,4),d))
```

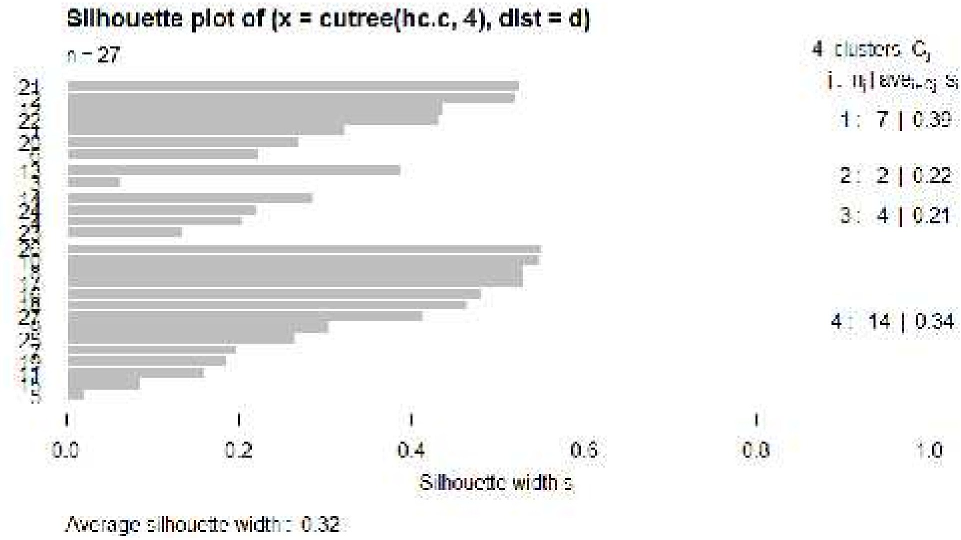


Fig: 20 Sllhouette Plot with 4 Cutee

The scree plot calculates sum of square within group.

```
> wss=(nrow(z)-1)*sum(apply(z,2,var))
> for(i in 1:20)wss[i]=sum(kmeans(z,centers=i)$withinss)
> plot(1:20,wss,type="b",xlab="Number Cluster",ylab="Within Groups")
```

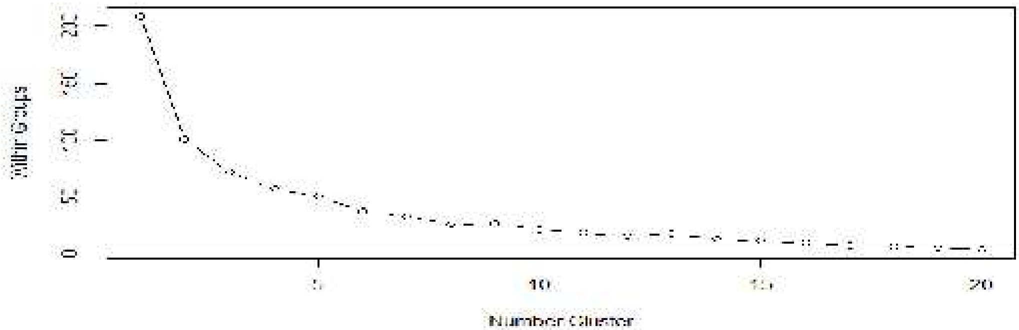


Fig: 21 Line Plot of Number of Cluster within groups

The scree plot shows that the three is high falls of data in 1 to 2 cluster when it increases it will not such significant when it reaches 5 cluster there is no such significant within group.

5. Conclusion

Cluster analysis is a technique used to classify objects or cases into related groups, is sometime called conglomerates. Cluster analysis is also classification analysis or numerical. if there is no group information or cluster membership for any of the objects is generally known as unsupervised clustering. Clustering procedures in cluster analysis can be hierarchical, non-hierarchical, or twostep procedures. A hierarchical procedure in the analysis of conglomerates is characterized by the development of a structure similar to a tree. A hierarchical procedure can be agglomerating. Agglomeration methods in cluster analysis consist of linking methods, variance methods, and centroid methods. Non-hierarchical methods in cluster analysis are often referred to as clustering. The two-step procedure can automatically determine the optimal number of groupings by comparing the model values. The choice of grouping procedures and the choice of distance measurement are correlated. The correlated dimensions of the groupings in the cluster analysis must be significant. Groups should be interpreted in terms of group centroids describes which variables or samples belong to which groups. Cluster analysis is popular in many fields like cancer research for the group expression profile to identify the molecular profile of patients with good or bad. The business segmentation marketing by identifying sub-groups of customers with similar profiles that could be included in a particular form of business promotion. agglomerated hierarchical grouping is deterministic only for the bounded distances when a single link is not used. DBSCAN is deterministic, except for the data set permutation in rare cases. Kmeans is deterministic except for initialization. It is possible to initialize with the first k objects, so it is also deterministic. The variables in which the cluster analysis will be performed. The hypothesis to be tested and the judgment of the experimenter should also be selected by the theory. It is necessary to select an appropriate measure of distance or similarity; The most used measure is the Euclidean distance or its square.

References

[1]. A. Ruiz, “The cognitive coder infoworld.” 2017.

[2]. S. Kakati, “What is data science. attend coursera, udemy and stanford online lagunita.” 2018.

[3]. G. Fiona, “Overfitting regression analysis,” www.bmj.com, 2018.

[4]. A. Nasridinov, “The third international conference on ieee, visual analytics for big data using R.” In Cloud and Green Computing (CGC), 2013.

[5]. R. Ihaka, “Mining big data: Current status andforecast to the future,” 1996.

[6]. M. RIJMENAM, “How does big data analytics help in decision making. p. https://datafloq.com/about/.” 2018.

[7]. M. Hahsler, “dbscan: Fast density-based clustering with r. volume wright state university.” Version Febu, 2013.

[8]. J. Leskovec and Anand Rajaraman Data Mining, Stanford University 2018

[9]. Joasang Lim1, Joongjin Kook2 and Jinman Kim DBSCAN-D: A Density-Based Clustering Method of Directionality, 2017

[10]. N. Golubovic Centaurus: *A Cloud Service for K-means Clustering*, 2015

Author’s Biography

Yagyanath Rimal is lecturer of Computer Science Information Technology at Pokhara University, Nepal. He received his Master of Science in Information Technology from SMU, India with Information Technology specialization and programming, he had wide area of interest like teaching student, writing books, doing

How to Cite

Rimal, Yagyanath (2019). Deterministic Machine Learning Cluster Analysis of Research Data: using R Programming. *International Journal of Machine Learning and Networked Collaborative Engineering*, 3(01) pp 46-65.

doi : https://doi.org/10.30991/IJMLNCE.2019v03i01.004.

Vol. 03, No. 1 ,(2019), 066–075 , **ISSN:2581-3242**

International Journal of Machine Learning and Networked Collaborative Engineering

Journal Homepage: http://www.mlnce.net/home/index.html

DOI : https://doi.org/10.30991/IJMLNCE.2019v03i01.005

A Robust Decomposition based Algorithm for removal of Pattern Noise from images

[a]Vinit Kumar Gunjan, [b]Fahimuddin Shaik

[a]Associate Professor , Department of Computer Science & Engineering ,

CMR Institute of Technology, Hyderabad, TS, India

vinitkumargunjan@gmail.com

https://orcid.org/0000-0002-3222-4186

[b]Assistant professor, Department of Electronics & Communication Engineering ,

Annamacharya Institute of Technology & Sciences, Rajampet Andhra Pradesh, India

fahimuddin.shaik.in@ieee.org

http://orcid.org/0000-0001-6852-7851

Abstract

This article aims a melting pool of complex vectors, that is, the aggregation and the minimization problem of sufficiency spectra. A mixture of this blended standard and image, decline issue works admirably to reduce and deteriorate the example of concussion which occurs when old pictures are filtered with granular surfaces. In most cases, the appealing appropriation of regular photos easily reduces from low repetition to the high repetition band, while the episode of concussion is scarcely circulating. We agree along these lines that a picture viewed includes an idle image and an example clamor, describing them separately by using the full range and capacity work. This enables the two parts to decompose sensibly. In contrast to the comparative strategies of deterioration, for instance, robust PCA, our technique is decent, less computer expenditure, and moreover less time suited for any image organization.

Keywords

Decomposition, Gaussian, Image, Noise, Pixel

PCA,

RPCA

1. Introduction

We are aiming to reduce example commotion based on the product and propose a precondition show to isolate the first image and the example clamor. Our divisions demonstrate and compute depend on the Wright et al. proposal (RPCA) for heartfelt primary part study[1] and on the Li et al. proposal for reflecting part evacuation[2]. These techniques accept that a watched picture comprises of two parts, and separate it into a perfect picture segment (in the future called "dormant" picture) and an antiquity picture segment unique in relation to shot commotion. While communicating the model as a scientific enhancement issue, notwithstanding the information fidelity communicating the perception procedure, these strategies use a few regularizations communicating the element of every segment, and separate picture parts by using the distinctions of the regularizations.

We regard design as an antique part in this paper. We propose a condition display to isolate an inactive frame and an example clamor image using a previously mentioned segment division and the vector standard of complex numbers.In the model, the dormant picture is portrayed by absolute variety minimization, while the example commotion is described by ℓ1 standard minimization of ghastly qualities with a zero meanrequirement. This model can be transmitted as a curved improvement problem, so that proper calculations are taken into account throughout the world. Subproblems that arise in the iterative calculation are also efficiently determined.

2. Literature Review

M. Belkin and P.Niyogi [2003] presented Laplacian eigenmaps for dimensionality decrease and information portrayal. Neural Computation.They proposed a geometrically roused calculation for speaking to the high-dimensional information. The calculation gives a computationally proficient way to deal with nonlinear dimensionality decrease that has area protecting propertie and a characteristic association with bunching.

E.J.Candes' [2006] proposed Robust vulnerability standards: definite flag recreation from exceptionally fragmented recurrence data. This considers the model issue of reproducing an article from deficient recurrence tests.

Z. Lin, A. Ganesh, J. Wright [2009] presented a Fast raised advancement calculations for careful recuperation of a tainted low-position network .Adaptive processing in multi - sensor co-mputer advances.

Wright et al, Y. Peng, Y. Mama, A. Ganesh, and S. Rao [2009] Proposed "Vigorous main segment investigation: accurate recuperation of ruined low-position lattices by raised advancement". The RPCA respects luminance change (reflection) brought about by reflected light as an antique segment, and describe it in order to limit the entirety of outright luminance esteems over the entire picture.

Y. Li and M. Darker [2014] proposed "Single picture layer detachment utilizing relative smoothness". The reflection segment expulsion manages a swoon and obscured reflection picture segment, and describe it in order to limit the entirety of luminance varieties over the entire picture.

2.1. Robust Principle Component Analysis

PCA may be the most commonly used tool for data analysis and dimensional reduction. However, its sensitivity to severely corrupted observations is often jeopardized by the validity of a single largely damaged entry in the M range..Net blurring is currently widespread in today's applications, such as image development, web information research and bioinformatics, where some estimates may disappear discretionarily or simply unimportant for the small structure that we want to recognize (because of impediments, harmful alterations or sensors disappointments). More than a few years ago, various common ways of dealing with robustifying PCA have been investigated and proposed in writing.

Fig. 1: Background modelling from video:

Three frames from a sequence of 250 frames taken in a lobby with different lighting. (a) Video original M. (b)-(c) Low-rank (L)+ sparse (S+ PCP. (d)-(e) low-ranking and sparse components obtained by means of a competing approach based on an alternating minimization of m-estimation.

In spite of less prior information, convex programming produces a more attractive outcome. Delegate approaches include influential work plans, multi-variant cutting, minimization substitutes and irregular examination systems. Unfortunately, none of these current methods provides a solid performance calculation of polynomial times. The new issue that we are looking into can also be regarded as an appreciated form of robust PCA in which we plan to reclaim a low position L0 from very poor estimates of M= L0 + S0. The passages in S0 cannot be considerably considered as a minor clamor term N0 in the established PCA, and their support is considered scarce but dark.

2.1.1. Limitations

In the RPCA, while we expect a low-ranked data matrix from a pattern image, the latent image is also low, so that when we tried to extract the pattern component the low frequency component was extracted incorrectly.

2.2. Reflection component removal

The elimination of the reflective component[2] addresses a weak and blurred reflective image component and defines it as the sum of light variations over the whole picture (Tikhonov regulation).

(a)RCR

Fig. 2: comparison of RCR with original .Observed image(top),pattern image(bottom).

2.2.1. Limitations

Although in the reflection retrieval fig. 2(a) we tried to extract the reflective component as a pattern, the low-frequency component was also extracted as a reflecting component (bottom line).

3. Proposed Methodology

This area presents a detail as a curved progress problem in order to isolate a depleted photo into a idle image and an example image. First, we present the picture perception display, and then portray picture strategies (regularization terms) for the isolation of an image into each segment.

Development model of image segments and the term information fidelity: We categorize a viewed picture as a gray N pixel image and as a vector segment of the RN and pixel calculations. Every shadow canal in this paper is shaped freely by our strategy and does not think about the relationship between the hues.

We defined the inactive image l to RN and the example clamor image e to the RN and the connection between it as:= l+e1. At that moment we defined the inactive imagel In addition, we make the associated assumptions: the mean estimates of example concussion is acknowledged as 0. The shot clamor in the shooting is small enough, lost from the model.

The model is commonly communicated as a minimum of the l2 perception error when taking care of the above perception as an advancement problem Our technique then again shows that this is the 2nd ball limitation for tracking: −(l+e) the whole thing 2 for each other, 1 folder=0, where −(alternatively), −(l+e) the client defines resilience. The mean estimate (whole) is 0 for 1 to 1N is a vector of one.

Characterization for the latent image: TV regularization: In many places, luminance estimates of an image are consistent and flat, and luminance estimates change vigorously around the elementary edges. The sum of the light varieties (all kinds: television) turned out to be small, thanks to clear pictures. We also think of this TV minimum, and use the mixed l2,1 standard and the differential filter, define the character as "compared to Dl," where D:=[D v^T, D h^T] a mixture between R2N and Dn and Dd RNN, and the differential filters. Note that the differential filter is general with 2-tap coefficients [−1,1] but a round filter is used so that the figures shown below can be reorganised. The filter framework transposed D diam is also related to a filter with 180 diameters [1,−1].

Characterization of the noise pattern: Spectral regularization: The spectrum over frequency coordinates, corresponding to the design, tends to grow in a pattern in which the same pattern of the artifacts occurs iteratively. When the same pattern appears strongly, the sum of spectral values (total spectrums) is small. We focus on that character and use a mixed lC,1 norm and the Fourier transform, define spectra reduction as» Transformation of the Spectrans« 1, where F al RN alternatively denotes the rapidly discrete transformation of Fourier. Note that we defineF as a unitary matrix, so the sum of a function square is equal to the sum of its transform square. In this case, FH and the inverse Fourier transform F−1 become equivalent FH= F−1 which makes the preceeding equations easier to manage.

Formulation for pattern noise decomposition: Combining the aforementioned data fidelity term, regularization terms, and constraint, we propose the formulation for noise pattern separation as

$$\min_{l,e} \|Dl\|_{2,1} + \lambda \|Fe\|_{C,1},$$

$$\text{s.t.} \ \|y - (l + e)\|_2 \leq \eta, 1^T e = 0 \qquad (1)$$

Where the first and second terms of the parameter are balanced in the objective function. In the experimental results V the parameters of α and Ț are described. Note that constraint 1Te=0 can be added to spectrum minimization and processed with only some algorithm alterations described below, so do not make it easy.

3.1. Algorithm using ADMM

In order to find a solution to the proposed (1) equation, some algorithms can be used and an algorithm can be displayed using ADMM. We do not draw the algorithm and calculations in each step due to the page limitations.

First, we prepare to add restrictions to the function of the object using the Lagrange multiplier method. To direct the ℓ_2 ball constraint $B_{y,\eta} := \{x \mid \|x-y\|_2 \leq \eta\}$ as a regularizer, we familiarize an indicator function $\iota B_{y,\eta}: \min_{1,e} \|Dl\|_{2,1} + \lambda \|Fe\|_{C,1} + \iota B_{y,\eta}(l+e)$. The indicator function is defined as $\iota_C(x) := \begin{cases} 0 & x \in C \\ \infty & \text{otherwise} \end{cases}$.

Where the set C is a set convex. In the case of the l2 ball, the indicator function becomes a convex function with an easy calculation of the proximity operator.These functions are then non- differentiable convex functions, so that the variables are replaced to make the functions more tractable,z_l := Dl, z_e := Fe, z := l+e and add these restrictions on equality to the objective function as

$$L(l,e,\{z\},\{u\}) := \|z_l\|_{2,1} + (\rho/2)\|z_l-(Dl + u_l)\|_2^2 + \lambda\|z_e\|_{c,1} + (\rho/2)\|z_e-(Fe + u_e)\|_2^2 + \iota_{B_{y,\eta}}(z) + (\rho/2)\|z-(l + e + u)\|_2^2 \qquad (2)$$

Where u_l,u_e u are multipliers of Lagrange.¨is the size of the step to control the ADMM convergence and set in this paper to ρ= 1.

Algorithm of ADMM: When solving (2), each variable is solved iteratively by the following minimization subproblems w.r.t. and each variable is solved while other variables are fixed. The calculation is performed by t-th iteration

$$l^{t+1},e^{t+1} := \min_x L\,(z_l^t, z_e^t, z^t, u_l^t, u_e^t, u^t),$$

$$z_l^{t+1} := \min_{z_l} L(X^{t+1}, u_l^t) = \mathrm{prox}_{1/\rho\,\|\cdot\|_{2,1}}(Dl^{t+1} + u_l^t),$$

$$z_e^{t+1} := \min_{z_e} L(X^{t+1}, u_e^t) = \mathrm{prox}_{\lambda/\rho\,\|\cdot\|_{c,1}}(Fe^{t+1} + u_e^t),$$

$$z^{t+1} := \min_z L(X^{t+1}, u^t) = \mathrm{prox}_{1/\rho\,\iota_{B_{y,\eta}}}(l^{t+1}+e^{t+1}+u^t),$$

$$u_l^{t+1} := \min_{u_l} L(X^{t+1}, z_l^{t+1}) = u_l^t + Dl^{t+1} - z_l^{t+1},$$

$$u_e^{t+1} := \min_{u_e} L(X^{t+1}, z_e^{t+1}) = u_e^t + Fe^{t+1} - z_e^{t+1},$$

$$u^{t+1} := \min_u L(X^{t+1}, z^{t+1}) = u^t + l^{t+1} + e^{t+1} - z^{t+1} \qquad (3)$$

The details on how l, e and proximity operators can be solved are described below. The number of iterations in Sec. V is displayed. As for the initial values, we set $l^{t=0}$ = y and zero vectors for other variables.

4. Result and Discussion

4.1. For Different Input Images

This segment appears, firstly, reenactment results utilizing the first inactive and design pictures, and afterward genuine outcomes utilizing checked and shot pictures. All considerations utilized in this trial are fixed: with respect to our model, λ = 3 and η = 0; concerning the ADMM, ρ = 1 and the quantity of emphasis is 100.

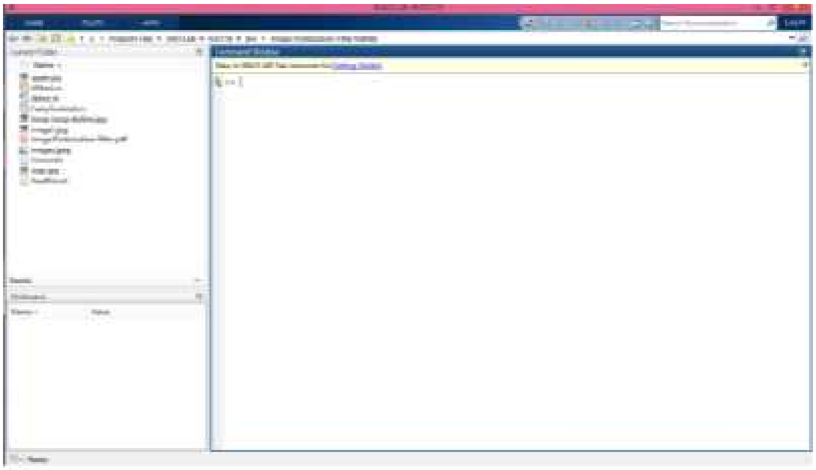


Fig. 3. : Reading an image

In the MATLAB programming first we have to peruse the picture from the present organizer as appeared underneath fig. 3.

Simulation Experiment: Preparing unique inert and design pictures as appeared in Fig.4, we produce perception pictures by straightly blended them, and afterward isolating the perception pictures in to inactive and design pictures again , finally look at PSNRs (crest flag and clamor proportion) between the first pictures and resulting pictures. We utilized the "Lena" and "Mandrill" as the idle pictures, though free resources3 as the example pictures. The picture sizes are every one of the 512 × 512 and its luminance is standardized inside the range [0,1]. With respect to the example pictures , halfway 86×86 locales are appeared for showing the subtleties. Reminder that, in creating perception pictures, the luminance estimations of each example picture are focused with the goal that the mean an incentive to be 0, and afterward mounted to half (×0.5), finally added to the dormant pictures.

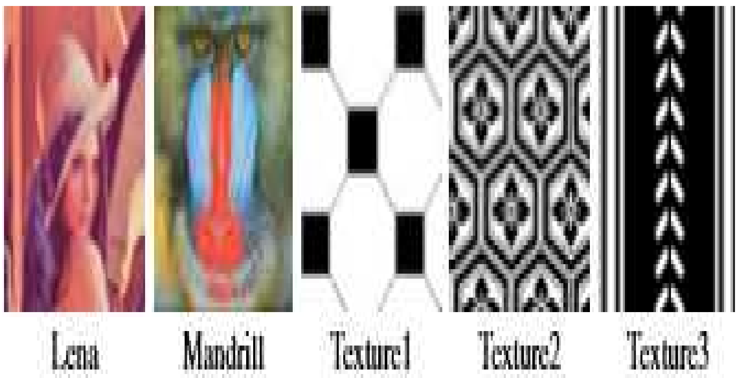
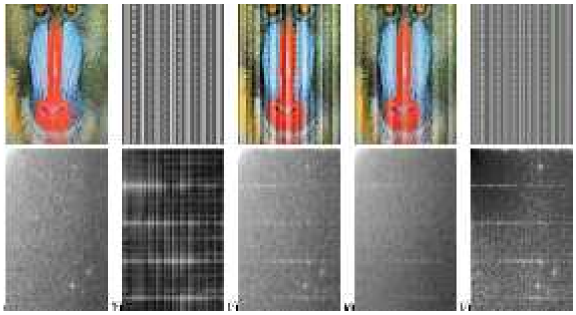


Fig. 4: Original latent images and patterns used in the experiment with simulation.

Qualitative assessment of the resulting images: The resulting images are shown in Fig. 5 and in Fig. 8(d), 8 (b) the original pattern picture (gray color is null light),8 (c) a mixed image, 8(d) a latent picture, and (e) the pattern picture obtained are shown from Fig.5, (a) a latent original picture. In the end their spectral images are shown. Here at Fig. 8 (d), constant ring at the image boundary and isotropic design with identical texture in both horizontal and vertical directions. In Fig.5 instead, we used a difficult pattern which, in the horizontal direction, is anisotropic and non-continuous.



(a) (b) (c) (d) (e)

Fig. 4:. Simulation results using an anisotropic pattern that is non-circular.

The intensity of spectral images is normalized to display the specifics. The image as input is composed of the latent and model components in Fig. 5 below**.**

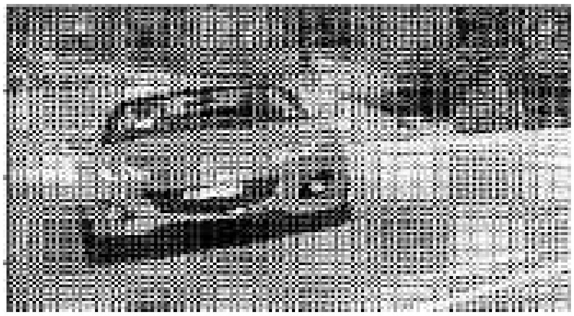


Fig. 5: Input image consists of latent and pattern components.

The latent image which is separated from input image is shown below in fig. 6.

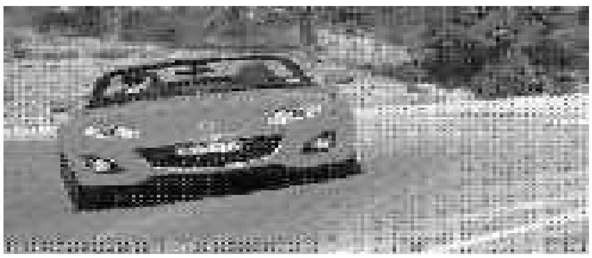


Fig. 6: Latent component separated from input image.

The pattern noise component separated from input image is shown below in fig. 7.

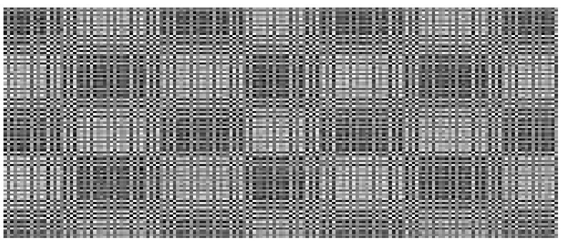


Fig. 7: Pattern noise component separated from input image.

The results of simulation are based on an uncircular anisotropic pattern. The images below are their spectrum. The spectral image intensity is standardized to display the detail. The pictorial components as the input in Figure4.4 below are the latent and model parts.

4.1.1. Comparison with related methods:

The outcome of this comparison and of the deletion of reflection[2] are shown in Fig. 8. Fig. 8. To extract components from the pattern, each method's factors are adjusted. We blur the pictures in this picture so that they appear as reflections in[2]. The standard 3 (pixel) difference is used to blur the spread-points Gaussian type function. Moreover, we change the pixel-specific multiplication model into the additive model in our method[2].In the RPCA (Fig.3.7(b)) the latent image is also low-ranking, although we expect that an image from the data matrix is low-ranking, so that the low-frequency component was wrongly extracted from the top row when we tried to extract the pattern part (b). Although in the reflective removal (c), the low frequency of the latent image was extracted, also as a reflection component (c) at the bottom of the line, as we attempted to extract the reflective component as a pattern. On the other hand, our technique can make a clear distinction between latent and pattern images.

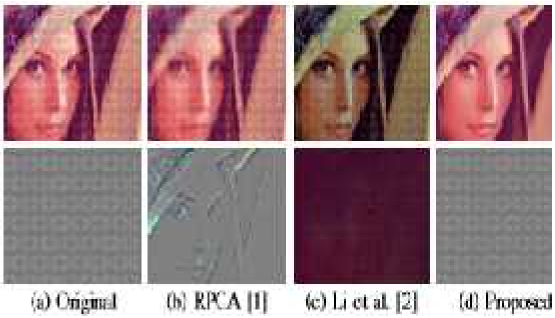


Fig. 8: Comparison with the methods involved. (a) is the image observed (top) and the pattern of the ground truth (bottom).

5. Conclusion and Future Scope

In this paper, we examined a technique for removing from a degraded picture a model component object. Since the spectrum of the model is sparse, we distinguish the image with the mixed Standard containing the complexly appreciated standard and the L1 as a problem of mathematical optimization. While this method cannot remove any noise from patterns, we have shown that it does better than conventional methods such as RPCA and removal of reflective components. In conclusion, we note that the problem of the pattern considered for eliminating noise is linked to the image break-down methods of cartoon texture that distinguish the texture with a regularization of local low-ranks, for example[11],[12]. The noise pattern problem can be managed by these methods. We also note, however, that in each iteration these methods require an unparalleled decomposition to solve the associated optimizing problems that is costly than our method.

References

[1] J. Wright, Y. Peng, Y. Ma, A. Ganesh, and S. Rao(2009), Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization, in *Proc. NIPS*, 2009, pp. 2080–2088.

[2] Y. Li and M. Brown(2014), Single image layer separation using relative smoothness, *in Proc. IEEE CVPR*, 2014, pp. 2752–2759.

[3] T.-S. T. Chan and Y. H. Yang(2016), Complex and Quaternionic principal component pursuit and its application to audio separation , *Signal Proce.Letters*, vol.23, no. 2, pp. 287–291

[4] S. Samarah, S. Obeidat, and R. Salman(2005) ,A Schur test for weighted mixed-norm spaces, *Anal. Math.*, vol.31, pp. 277–289

[5]A. Benedek and R. Panzone(1961), The space ℓp with mixed norm,*Duke Math. J.*, vol. 28, pp. 301–324

[6]M. Kowalski(2009) ,Sparse regression using mixed norms,Applied Comput. *Harmonic Anal*., vol. 27, no. 3, pp.303–324

[7]L. Rudin, S. Osher, and E. Fatemi(1992), Nonlinear total variation based noise removal algorithms, *Physica D*, vol. 60, no. 1-4, pp. 259–268

[8] D. Gabay and B. Mercier(1976),A dual algorithm for the solution of nonlinear variational problems via finite elements approximations, Comput. *Math.Apps*., vol. 2, pp. 17–40

[9]S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein(2011), Distributed optimization and statistical learning viathe alternating direction method of multipliers," *Founda. & Trends in Machine Learn*., vol. 3, no. 1, pp. 1–122

[10]K. Shirai and M. Okuda(2014), Fft based solution for multivariable l2 equations using kkt system via fft andefficient pixel-wise inverse calculation, in *Proc. IEEE ICASSP*, 2014, pp. 2648–2652.

[11] H. Schaeffer and S. Osher, "A low patch-rank interpretation of texture," *SIAM J. Imag. Sci*., vol. 6, no. 1, pp. 226–262, 2013.

[12] S. Ono, T. Miyata, and I. Yamada(2014), Cartoon- texture image decomposition using blockwise low-rank texture characterization, *IEEE T. Image Proce*., vol. 23, no. 3, pp. 1128–1142

[13] M. Belkin and P. Niyogi(2014), Laplacian eigenmaps for dimensionality reduction and data that the representation.*Neural Computation*,15(6):137 3-1396,

[14] D.P. Bertsekas(2008), *Constrained Optimization and Lagrange Multiplier Method.* Academic Press,1982. for matrix completion.preprint,

[15] E. J. Cand`es and Y. Plan(2010), Matrix completion with noise. *Proceedings of the IEEE,* Volume: 98 , Issue: 6 , June 2010.

[16] S.Chen, D.Donoho, and M.Saunders(2001), Atomic decomposition by basis pursuit. *SIAM Review*, volume 43,number 1, pp:129– 159

[17] S. Dewester, S. Dumains, T. Landauer, G. Furnas, and R. Harshman(1990), Indexing by latent semantic analysis. *Journal of the Society for Information Science*, volume 41 , number 6, pp:391–407

[18] C. Eckart and G. Young(1936), The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–218

[19] M. Fazel, H. Hindi, and S. Boyd(June 2013), Log-det heuristic for matrix rank minimization with applicationsto Hankel and Euclidean distance matrices. In *Proceedings of the American Control Conference*,pp: 2156–2162

[20] M. Fischler and R. Bolles(1981), Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24, pp :381–385

[21] A. Georghiades, P. Belhumeur, D. Kriegman(2001), From few to many: Illumination cone models for face recognition under variable lighting and pose *IEEE Trans. on Pattern Analysis and Machine Intelligence*, volume 23, number 6.

[22] R. Gnanadesikan and J.Kettenring(1971), Robust estimates,residuals ,and outlier detection with multi-response data. *Biometrics*,28, pp:81– 124

Authors' Biography

Vinit Kumar Gunjan : Associate Professor in Department of Computer Science & Engineering at CMR Institute of Technology Hyderabad (Affiliated to Jawaharlal Nehru Technological University, Hyderabad). An active researcher; published research papers in IEEE, Elsevier & Springer Conferences, authored several books and edited volumes of Springer series, most of which are indexed in SCOPUS database. Awarded with the prestigious Early Career Research Award in the year 2016 by Science Engineering Research Board, Department of Science & Technology, Government of India. An active Volunteer of IEEE Hyderabad section; volunteered in the capacity of Treasurer, Secretary & Chairman of IEEE Young Professionals Affinity Group & IEEE Computer Society. Was involved as organizer in many technical & non-technical workshops, seminars & conferences of IEEE & Springer. During the tenure I had an honour of working with top leaders of IEEE and awarded with best IEEE Young Professional award in 2017 by IEEE Hyderabad Section.

Fahimuddin Shaik :He holds a PhD in Electronics and Communication Engineering and currently serves as an Associate Professor in Department of Electronics & Communication Engineering at Annamacharya Institute of Technology & Sciences (AITS) (an Autonomous Institute), India. He is BOS Member of the Department and also held a position as the Academic Council Member of the Institute.

His research interests include Signal Processing, Time Series Analysis and Biomedical Image Processing. He is a member of professional bodies such as IEEE, BMESI and ISTE. He is the Vice-Chair for IEEE YP AG (Young Professionals Affinity Group), IEEE Hyderabad Section for the year 2019 and Executive Committee Member and Treasurer for IEEE YP AG in 2017 and 2018respectively. He acted as Publication Chair of IEEE Conference ICRTEECT-2017organized by SREC, Warangal in 2017 and as Organizing Committee member of 15th IEEE International Conference, ICACT-2013 held at AITS, Rajampet in 2013. He has authored books by titles Medical Imaging in Diabetes with Cinnamonteal Publishers in 2011, Image Processing in Diabetic Related Causes with Springer Publications in 2015, Signal and Image Processing in Medical Applications with Springer Publications in 2016 and Computational Methods in Molecular Imaging Technologies with Springer publications in 2017. He filed 5 Patents and four of them published in IP India Journal, in the area of Health care and Image Processing. He is one of the Co-Principal Investigators for a project funded by the Science Engineering Research Board, Department of Science & Technology, Government of India.

How to Cite

Gunjan, Vinit Kumar & Shaik, Fahimuddin (2019). A Robust Decomposition based Algorithm for Removal of Pattern Noise from Images. *International Journal of Machine Learning and Networked Collaborative Engineering*, 3(01) pp 66-75.

doi : https://doi.org/10.30991/IJMLNCE.2019v03i01.005