Data Replication for the Distributed Database Using Decision Support Systems

P.Elango

Research Scholar

Allagappa University

Karaikudi, India

Dr.K.Kuppusamy

Professor and Head

Allagappa University

Karaikudi, India

N.Prabhu

Assistant Professor

Gobi Arts and Science College

Gobichettipalayam, India

Dr.N.M.Mallika

Dean and Assistant Professor

Sri Vasavi college

Erode, India

Abstract: Replication is a subject of enthusiasm for the conveyed figuring, disseminated frameworks, and database networks. Choice emotionally supportive networks ended up pragmatic with the advancement of minicomputer, timeshare working frameworks and appropriated figuring. Reproduced information may get lacking because of framework disappointment, adaptation to internal failure, and unwavering quality. A halfway replication is quantized in the replication framework will expand the non repeated framework. Adaptation to internal failure is the property that empowers a framework (regularly PC based) to keep working legitimately. Exchange Processing Replication (TP-R) and Decision-bolster replication diagram (DDS-R) will clear the non copy and it is utilized to clear the server issues and framework blunder. This procedure is top notch in disseminated frameworks and it doesn't neglect to distinguish the framework blunders when different accesses are multiplexed.

Keywords: Cloud computing, Data mining, Data replication, Partial replication, cascades replication.

1. INTRODUCTION

Replication is the way toward sharing data in order to guarantee consistency between repetitive assets, for example, programming or equipment parts, to improve dependability, adaptation to non-critical failure, or availability. It could be information replication if similar information is put away on various capacity gadgets or calculation replication if a similar figuring task is executed commonly. It is the procedure of naturally conveying duplicates of information and database objects among SQL Server examples, and keeping the dispersed data synchronized. Replication is the way toward sharing data in order to guarantee consistency between excess assets, for example, programming or equipment segments, to improve dependability, adaptation to internal failure, or openness. It could be information replication if similar information is put away on numerous capacity gadgets or calculation replication if a similar processing task is executed ordinarily. The safe sharing of data in this kind of condition is an unpredictable issue. The proprietors of the distinctive information sources will have diverse strategies on access to and the scattering of the information that they hold. There are two principle sorts of replication conventions: dynamic replication, in which all imitations forms simultaneously all info messages. Aloof replication, in which just a solitary one of the reproductions forms all information messages and intermittently transmits its present state to different copies so as to keep up consistency from the previous years, Distributed Databases have taken consideration in the database look into network. Information dispersion and replication offer open doors for improving execution through parallel inquiry execution and burden adjusting just as expanding the accessibility of information. In an appropriated database framework, information are frequently recreated to improve unwavering quality and accessibility, subsequently expanding its trustworthiness.

2. DATA REPLICATION

Replication is the way toward sharing data to guarantee consistency between excess assets, for example, programming or equipment segments to improve dependability, adaptation to internal failure, or availability. It could be information replication if similar information is put away on numerous capacity gadgets or calculation replication if a similar registering task is executed ordinarily. Replication has been contemplated in numerous regions, particularly in appropriated frameworks (essentially for adaptation to internal failure purposes) and in databases (principally for execution reasons). Replication is one of the most seasoned and most critical subjects in the general territory of disseminated framework. A critical issue in circulated frameworks is the replication of information. Information are for the most part recreated to upgrade dependability or improve execution. Replication is the way toward duplicating information from an information store or document framework to various PCs to synchronize the information. Database replication is ending up progressively imperative job for database applications. Duplicated information are ending up increasingly more of intrigue recently. Replication is a financially savvy approach to build accessibility and utilized for both execution and blame tolerant purposes in this way presenting a steady tradeoff among consistency and productivity. Replication gives reinforcement database extensive endeavors as a rule have locales where it is basic to get to information consistently. In the event that a server breakdown it is critical to approach similar information on an alternate server and this typically requires authoritative mediation . Database replication is rapidly turning into a basic device for giving high accessibility, survivability, and elite for database applications. Be that as it may, to give helpful replication one needs to take care of the non-inconsequential issue of keeping up information consistency between every one of the reproductions. Replication utilizes the accompanying three servers to be specific Publisher, Distributor, and Subscriber. The essential issue with information replication is that an update to some random intelligent item should be spread to all put away duplicates of that object. A trouble that emerges quickly is that a few locales holding a duplicate of the item may be inaccessible (on account of a site or system disappointment) at the season of the update. The conspicuous procedure of spreading refreshes quickly to all duplicates is in

this manner likely inadmissible, in light of the fact that it infers that the update and along these lines the exchange will fall flat if any of those duplicates is at present inaccessible. It might be said, truth be told, information is less accessible under this technique than it would be in the non-imitated case.
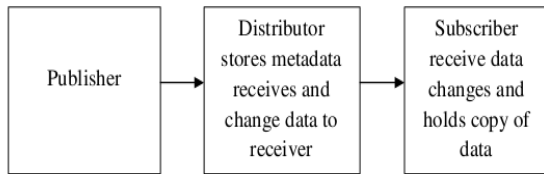

Figure. 1 Replication Process

3. ANALYTICAL MODEL FOR PARTIAL REPLICATION

Give us a chance to accept that a non recreated database has a preparing limit, that a non-reproduced database can execute exchanges per time unit. All destinations have a similar limit, a non-recreated database in which the whole handling limit is utilized for executing neighborhood exchanges, however locales in the repeated database need to utilize a portion of its preparing limit with respect to coordination with different destinations. We term the coordination fill in as remote work. Along these lines, each site in a recreated database utilizes a small amount of its handling limit with regards to nearby work and the rest of the limit with respect to remote work, for example at that point, the nearby work performed in a site is

$$Li = c - Ri \tag{1}$$

The scale out is the sum of the amount of local work executed at each site, divided by the processing capacity of a non-replicated database $Scaleout = \dfrac{\sum^{n} i = 1 \ Li}{C}$ i.e. how many times the capacity of a non-replicated system is increased when it is replicated. The more local work each site executes, the better the scalability of the system. The total amount of local work at site is the sum of accesses to objects stored at (Eq. 2). The objects stored at site are defined by the function

$$r(i,k) \quad L_i = \sum_{K=1}^{o} C.r(i,k).a_i k, \ \forall = 1...n \tag{2}$$

4. TRANSACTION PROCESSING REPLICATION (TP-R)

Value-based replication is a sort of replication that enables information adjustments to be proliferated steadily between servers in a dispersed situation. Value-based replication can be utilized for a wide range of uses, from revealing servers and information warehousing situations to Web servers and web based business applications. Value-based replication is utilized at a significant number of the prevalent Web locales

on the Internet that run SQL Server, including MSN.com, Passport.com, Barnes and Noble.com, and Buy.com. Value-based replication is an adaptable and solid answer for dispersing information in elite situations. An exchange handling replication (TP-R) approach that can keep up close ongoing exchange respectability at information duplicate locales is fundamental. In languid replication, exchanges are executed first at one imitation. Any updates are spread to different reproductions simply after exchange submits, therefore giving quick reaction times. The idea of exchange was first presented in database frameworks, with the goal of supporting the steady execution of simultaneous tasks over shared information However, from that point forward exchanges have been connected substantially more extensively, e.g., in dispersed frameworks in numerous application situations, in which they improve unwavering quality and certification information consistency. This is about the exchange submit for recreated databases. There is work here for certain years, with creators proposing the usage of deliberations, regularly used to indicate dependable conveyed frameworks (e.g., agreement, all out request multicast) to help crash blame tolerant database replication or, all the more conventionally, exchange handling. TP-R replication is essentially worried about making a solitary picture of a database crosswise over appropriated self-governing destinations and saving database honesty in close continuous preparing. The general uprightness of databases is safeguarded by sending information changes coming about because of single client exchanges.

5. CASCADING REPLICATES

Course conduct is a Pareto improvement over conduct in which person's base choices just on their private data, since a course mirrors a coordination of more private data than any single individual has.
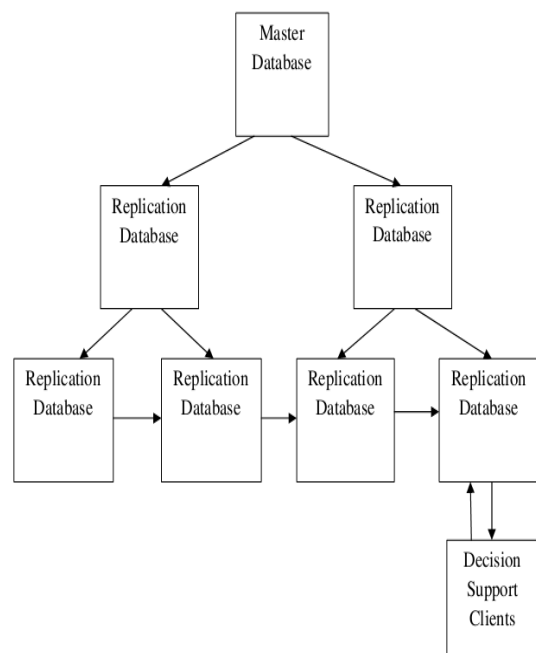
Figure. 2 cascading database replication

Table 1: Comparison of timeout on bandwidth

6. DECISION-SUPPORT REPLICATION SCHEMA (DDS-R)

DSS-R ways to deal with replication generally are based on different innovation varieties of table duplicating. The common choice help application has a prerequisite for reliable period information sources and not really for information that is regularly updated current. DSS-R approaches, at that point, don't commonly stress over keeping the information current. Predictable, stable information for a given period is the most astounding necessity for these sorts of utilizations. The choice emotionally supportive networks are tuned for inquiry preparing, commonly by including more lists. For this situation, at that point, persistent proliferation of updates would meddle with the capacity of the inquiry apparatus to give sensible execution. The replication server ought to give different planning choices which can make duplicates dependent on coordinated occasions (clock or interim), on application occasions (for example end of day compromise finished), or on manual solicitation. Other vital prerequisites for choice help incorporate the capacity to get to inheritance generation framework information from sources, for example, IMS, RMS, VSAM, and level documents and to give complex information control/improvement to that information.

7. DSS-R SCHEMA

The esteem added to the information by control or improvement is imperative in DSS-R conditions. Sources are commonly inheritance frameworks and the replication arrangement ought to give the capacity to rebuild the information from heritage groups into the social model. Apparatuses ought to offer help for joining information from various sources, for computing new qualities, for accumulating information and for changing encoded information into spellbinding structures. A vital side point to remember is that one of the key advantages of DSS-R, accumulation of information or de standardization, is something that ought not to be done when the duplicate is updatable.

8. RESULT AND DISCUSSION

In our proposed strategy, the information replication for choice supporting framework is portrayed. The usage was done in JAVA. Here in our proposed part, the manufactured database is utilized. Timeouts are best when a huge level of the transient disappointments can be overlooked, which is reliant on the downtime appropriation. Be that as it may, for strength to stay high, the normal hub lifetime should be fundamentally more prominent than the timeout. To assess this situation where timeouts ought to have sway, we played out a trial utilizing an engineered follow where we fluctuated the fix edge and the hub timeout. Since the framework would perceive hubs returning after a perpetual disappointment and promptly terminate all pending timeouts for these hubs, we allotted new characters to such hubs to permit long timeouts to lapse ordinarily. Table1 demonstrates the consequences of this reenactment the all out bytes sent as a component of timeout

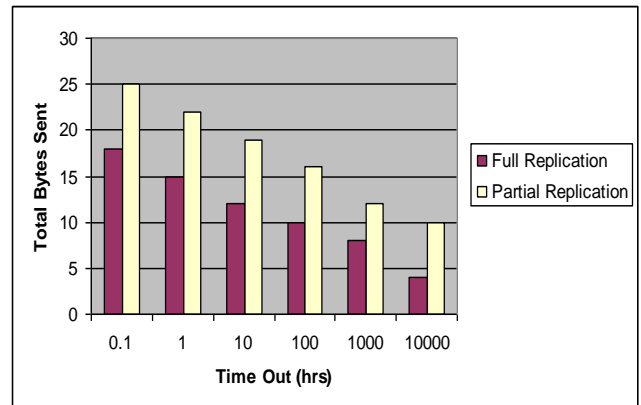| Time Out (hours)	Full Replication	Partial Replication
0.1	18	25
1	15	22
10	12	19
100	10	16
1000	8	12
10000	4	10

Figure 3 Comparison of time out on bandwidth

The impact of timeouts on bandwidth is compared with full and partial replication using decision support system. Figure3 shows the number of copies created for various timeout values.

Table2 Comparison of Full and Partial Replication based on Fault occurrence rate and System error

Methods	Fault Occurrence Rate	System error
Full Replication	75%	25%
Partial Replication	45%	10%

9. ACKNOWLEDGMENTS

10. REFERENCES

[1] Reza Akbarinia, Mounir Tlili, Esther Pacitti, Patrick Valduriez and Alexandre A. B. lima," Replication in DHTs using Dynamic Groups," Lecture Notes in Computer Science, Vol. 6790, No.3, pp. 1-19, 2011.

[2] Reza Akbarinia, Mounir Tlili, Esther Pacitti, Patrick Valduriez and Alexandre A. B. lima," Continuous Time stamping for Efficient Replication Management in DHTs," Lecture Notes in Computer Science, Vol. 6265, pp. 38-49, 2010.

[3] Prasanna Padmanabhan, Le Gruenwald, Anita Vallur and Mohammed Atiquzzaman," A survey of data replication techniques for mobile ad hoc network databases, International journal on very large database," Vol. 17, pp. 1143–1164, 2008.

[4] Dutta, H., Kamil, A., Pooleery, M., Sethumadhavan, S., Demme, J.: Distributed Storage of Large-Scale Multidimensional Electroencephalogram Data Using Hadoop and HBase," Grid and Cloud Database Management, pp. 331-347, 2011.

[5] Hannes Muhleisen, Tilman Walther and Robert Tolksdorf," Data Location Optimization for a Self-Organized Distributed Storage System," In Proc. of the Third World Congress on Nature and Biologically Inspired Computing (NaBIC), pp. 176 - 182, Oct 2011.

[6] Yi Lin, Bettina Kemme, Marta Patino Martinez, Ricardo Jimenez Peris," Consistent Data Replication: Is it feasible in WANs," Information Science and Computer communication, 2004.

[7] Zahia Guessoum, Jean-Pierre Briot, Nora Faci and Olivier Marin," Towards reliable multi-agent systems: An adaptive replication mechanism," Journal of Multi agent and Grid Systems, Vol.6, No. 1, pp. 1-24, 2010.

[8] Byung-Gon Chun, Frank Dabek, Andreas Haeberlen, Emil Sit, Hakim Weatherspoon, M. Frans Kaashoek, John Kubiatowicz and Robert Morris," Efficient Replica Maintenance for Distributed Storage Systems," In Proc. of the Symposium on Networked Systems Design and Implementation, 2006.

[9] Michael H. Zack," The role of decision support systems in an indeterminate world," Decision Support Systems, Vol. 43, No. 4, pp. 1664–1674, 2007.

[10] Sanjay Kumar Tiwari and A.K.Sharma," Management Issues in Replicated Distributed Real Time Database," International Journal of Advance in Science and Technology, Vol.3, No1, pp. 75-89, 2011.

[11] Maris G. Martinsons and Robert M. Davison," Strategic decision making and support systems: Comparing American, Japanese and Chinese management," Decision Support Systems, Vol. 43, pp. 284 – 300, 2007.

[12] G. Bent, D. Vyvyan, David Wood, Petros Zerfos and Seraphin Calo," Distributed Policy Based Access to Networked Heterogeneous ISR Data Sources," In Proc. of the conference on International society of optics and photonics, Vol. 7694, 2010.

[13] H. T. Barney and G. C. Low," Object Allocation with Replication in Distributed Systems," World Academy of Science, Engineering and Technology, Vol. 48, pp. 558-566, 2008.

[14] Costel Ciuchi, Dan Picu and Gyorgy Todoran," Managing Knowledge and Data for a Better Decision in Public Administration," Administration and Public Management, Vol. 2011, No. 17, pp. 64-81, 2011.

[15] Carlo Curino, Evan Jones, Yang Zhang and Sam Madden," Schism: a Workload-Driven Approach to Database Replication and Partitioning," In Proc of the Conference on very large database Endowment, Vol.3, No.1, Sep 2010.

Use and Analysis on Cyclomatic Complexity in Software Development

Chandrakant R. Gujar
Assistant Professor,
MCA WKBS Mandals
MCA College for Women's
Deopur Dhule, Maharashtra, India

Dr. Suryakanta R. Ajmera
MCA WKBS Mandals
MCA College for Women's
Deopur Dhule, Maharashtra, India

Abstract: Cyclomatic complexity is software metric used in software developments as White box testing and structural testing. The purpose of the paper is to describe the Use and Analysis on Cyclomatic complexity in Software development with an example. The Cyclomatic complexity is computed using the flow graph of the program: the nodes of the graph correspond to one or more code statement and the edges connect two nodes. Based on the flow graph how to find Cyclomatic complexity is described here.

Keywords: Cyclomatic Complexity, Flow graph, Predicate node, connected components, developers, testers.

1. INTRODUCTION

Cyclomatic complexity is software metric (measurement). It was developed by Thomas J. McCabe, Sr. in 1976 and is used to indicate the complexity of a program. It is a quantitative measure of the complexity of programming instructions. It directly measures the number of linearly independent paths through a program's source code. It is one of the metric based on not program size but more on information/control flow.

The Cyclomatic Complexity is software metric that provides quantitative measures of logical complexity of a program.

Basic Concepts

The Cyclomatic complexity of a section of source code is the count of the number of linearly independent paths through the source code. For instance, if the source code contained no decision points such as IF through the code. If the code had a single IF statement containing a single condition, there would be two paths through the code: one path where the IF statement is evaluated as TRUE and one path where the IF statement is evaluated as FALSE.

1.1 What is Cyclomatic Complexity?

Cyclomatic complexity is a software metric used to measure the complexity of a program. These metric, measures independent paths through program source code. Independent path is defined as a path that has at least one edge which has not been traversed before in any other paths. Cyclomatic complexity can be calculated with respect to functions, modules, methods or classes within a program.

This metric was developed by Thomas J. McCabe in 1976 and it is based on a control flow representation of the program. Control flow depicts a program as a graph which consists of Nodes and Edges.

In the graph, Nodes represent processing tasks while edges represent control flow between the nodes. fig 1 [1]
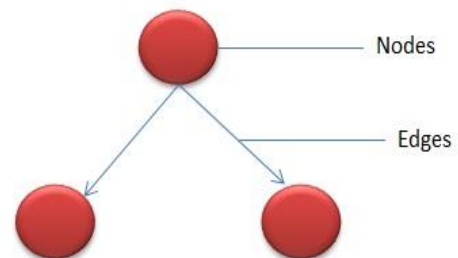
1.2 Flow graph notation for a program:

Flow Graph notation for a program is defines. Several nodes connected through the edges. Below are Flow diagrams for statements like if-else, While, until and normal sequence of flow. fig.2 Reference [1]
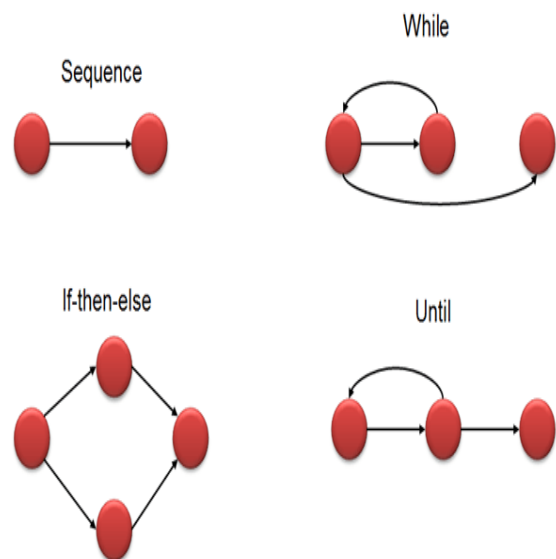
1.3 How to Calculate Cyclomatic Complexity?

The Cyclomatic Complexity is computed in one of five ways:

- The number of regions of the flow graph corresponds to the Cyclomatic complexity.
- The Cyclomatic complexity, V(G), for a graph G is defined as **V(G) = E – N + 2**
- Where E is the number of flow graph edges and N is the number of flow graph nodes.
- The Cyclomatic complexity, V(G), for a graph G is defined as **V(G) = E – N + 2P**
- Where E is the number of flow graph edges, N is the number of flow graph nodes and P is connected components.
- The Cyclomatic complexity, V(G), for a graph G is also defined as **V(G) = P + 1**
- Where P is the number of predicate nodes contained in the flow graph G. The predicate node is a node that has of out degree two i.e. Binary node.
- The Cyclomatic complexity, V (G), for a graph G is also defined as total number of independent path of flow graph.

For Example:

```
i = 0;
n=4; //N-Number of nodes present in the graph
while (i<n-1) do
j = i + 1;
while (j<n) do
if A[i]<A[j] then
swap (A[i], A[j]);

end do;
i=i+1;

end do;
```

Flow graph for this program will be fig3 [1]
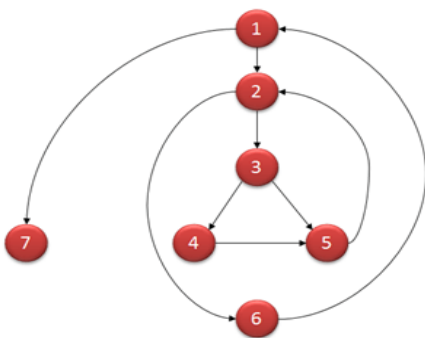
1.4 Computing mathematically,

- . $V(G) = 9 - 7 + 2 = 4$
- $V(G) = 3 + 1 = 4$ (Condition nodes are 1,2 and 3 nodes)
- Basis Set - A set of possible execution path of a program
- 1, 7
- 1, 2, 6, 1, 7
- 1, 2, 3, 4, 5, 2, 6, 1, 7
- 1, 2, 3, 5, 2, 6, 1, 7

2. PROPERTIES OF CYCLOMATIC COMPLEXITY:

Following are the properties of Cyclomatic complexity:

1. V (G) is the maximum number of independent paths in the graph
2. V (G) >=1
3. G will have one path if V (G) = 1
4. Minimize complexity to 10

2.1 How this metric is useful for software testing?

Basis Path testing is one of White box technique and it guarantees to execute at least one statement during testing. It checks each linearly independent path through the program, which means number test cases, will be equivalent to the Cyclomatic complexity of the program.

This metric is useful because of properties of Cyclomatic complexity (M) -

1. M can be number of test cases to achieve branch coverage (Upper Bound)
2. M can be number of paths through the graphs. (Lower Bound)

Consider this example -

```
If (Condition 1)
Statement 1


Else
Statement 2


If (Condition 2)
Statement 3

Else
```

Statement 4

Cyclomatic Complexity for this program will be 9-7+2=4.

As complexity has calculated as 4, four test cases are necessary to the complete path coverage for the above example.

Steps to be followed:

The following steps should be followed for computing Cyclomatic complexity and test cases design.

Step 1 - Construction of graph with nodes and edges from the code

Step 2 - Identification of independent paths

Step 3 - Cyclomatic Complexity Calculation

Step 4 - Design of Test Cases

Once the basic set is formed, TEST CASES should be written to execute all the paths.

More on V (G):

Cyclomatic complexity can be calculated manually if the program is small. Automated tools need to be used if the program is very complex as this involves more flow graphs. Based on complexity number, team can conclude on the actions that need to be taken for measure.

Table 1 : Overview on the complexity number and corresponding meaning of v (G):

1.4 Complexity Number	1.5 Meaning
1.6 1-10	1.7 Structured and well written code High Testability Cost and Effort is less
1.8 10-20	1.9 Complex Code Medium Testability Cost and effort is Medium
1.10 20-40	1.11 Very complex Code Low Testability Cost and Effort are high
1.12 >40	1.13 Not at all testable Very high Cost and Effort

3 TOOLS FOR CYCLOMATIC COMPLEXITY CALCULATION:

Many tools are available for determining the complexity of the application. Some complexity calculation tools are used for specific technologies. Complexity can be found by the number of decision points in a program. The decision points are if, for, for-each, while, do, catch, case statements in a source code.

Examples of tools are

- **OCLint** - Static code analyzer for C and Related Languages
- **devMetrics** - Analyzing metrics for C# projects
- **Reflector Add In** - Code metrics for .NET assemblies
- **GMetrics** - Find metrics in Java related applications
- **NDepends** - Metrics in Java applications

4 USES OF CYCLOMATIC COMPLEXITY:

Cyclomatic Complexity is a very common buzz word in the Development community. This technique is mainly used to determine the complexity of a piece of code or functionality [1].

The technique was developed by MaCabe and helps to identify the below 3 questions for the programs / features

- Is the feature / program testable?
- Is the feature/ program understood by every one?
- Is the feature / program reliable enough?

As a QA we can use this technique to identify the "level" of our testing. It is a practice that if the result of Cyclomatic complexity is more or a bigger number, we consider that piece of functionality to be of complex nature and hence we conclude as a tester; that the piece of code / functionality requires an in-depth testing. On the other hand if the result of the Cyclomatic Complexity is a smaller number, we conclude as QA that the functionality is of less complexity and decide the scope accordingly.

Cyclomatic Complexity can prove to be very helpful in

- Helps developers and testers to determine independent path executions
- Developers can assure that all the paths have been tested atleast once
- Helps us to focus more on the uncovered paths
- Improve code coverage in Software Engineering
- Evaluate the risk associated with the application or program
- Using these metrics early in the cycle reduces more risk of the program

5. CONCLUSION:

Cyclomatic Complexity is software metric useful for White Box and structured testing in Software development. It is mainly used to evaluate complexity of a program. If the decision points are more, then complexity of the program is more. If program has high complexity number, then probability of error is high with increased time for maintenance and trouble shoot.

6. ACKNOWLEDGMENTS:

7. REFERENCES:

[1] https://www.guru99.com

[2] Pressman Roger, Software, Engineering: A Practitioner's Approach Tata McGraw Hill, New Delhi.

[3]   A Review and Analysis on Cyclomatic Complexity by Ramesh Patelia and Shipan vyas.

[4] https://www.softwaretestinghelp.com/ Cyclomatic-complexity/

[5] Software Engineering by K K Agrawal New Age Publication

[6] Somerville, Software Engineering, Addison-Wesley.

Student Performance Prediction

Isha D Shetty
Department of Information
Technology
VESIT
Chembur, India

Dipshi   Shetty
Department of Information
Technology
VESIT
Chembur, India

Sneha Roundhal
Department of Information
Technology
VESIT
Chembur, India

Abstract:Predicting Students performance beforehand can be very beneficial for educational institutions to improve their teaching quality. This paper proposes to predict students performance by considering their academic details. For classification, decision tree, Random forest, Support vector machine, Linear regression, Naive Bayes are used. This paper also focuses on the use of Artificial Neural network algorithms such as MLP and Convolutional Neural network for predicting the performance.

1.   INTRODUCTION

Today every educational institution handles and deals with large amount of student data which can be beneficial for a number of reasons. One of the important application of such data is predicting student performance. Such a prediction can be useful not only for the students but also for teachers/mentors. Mentors can provide special assistance to the students who are on the verge of failing. In order to determine which category a student lies, such data can be quite helpful. This application can be used by any prominent school or colleges. It can be used to predict the pointer ranges or percentage range for future semester exams. These ranges can be predicted using a number of data mining algorithms such as classification algorithms, rule-based algorithms, ensemble methods, and neural networks. The main aim of this project is the selection of features that show a strong relationship with a target attribute that is to be predicted from a high dimensional dataset. We have evaluated and compared the number of algorithms such as decision tree, random forest, support vector machine, naive Bayes and neural networks by applying them on the dataset The rest of the paper provides an explanation on nature of neural networks along with the results of our evaluation.

1.1   Objective

Today every educational institution handles and deals with large amount of student data which can be beneficial for a number of reasons. One of the important application of such data is predicting student perfo

2.   OVERVIEW OF DATASET

This dataset includes actual academic details of the students of one of the prominent engineering institute of Mumbai University. Engineering students of mumbai university are classified as Regular students (who join engineering colleges after their Higher Secondary education) and Direct Second Year (DSE-who join after completing their diploma degree). The SGPI(Student Grade Performance Index) of a particular student varies from .01 to .99 and can be difficult to analyze and predict, hence to get efficient prediction the academic performance data is categorized.

It has 6050 instances and 55 attributes. The dataset contains students Personal details like parents name, date of birth, address etc and academic features like marks of all semesters, HSC percentages, SSC percentages etc. There is a mix of the numeric and nominal attribute.

Some of the attributes considered while training the model are :

Table -1: Dataset

Attributes
Student Type (Regular or DSE)
SSC Board
SSC Percentage
HSC Board
HSC Percentage

DSE Board
DSE Percentage
Semester Marks (Graduation)
Number of KT's
Education Gap
Branch
Admission Type(CAP or Minority)

3. DATA PREPROCESSING

The dataset is prepared to apply data mining techniques. Preprocessing methods such as data cleaning, transformation of variables, data balancing and data partitioning had to be applied.

There are 17 numeric attributes with null values. To deal with null values, semester pointers and 10th percentage columns are replaced with their respective column medians. On the other hand, the columns such as the number of KTs were replaced with 0 as in the dataset there is proper mention of whether the person has a kt. So in case if the value of this attribute is NAN, then it means the student has no KT. The SSC percentages which are in terms of CGPI were converted into percentages.

The dataset was split into the training and testing data. The ratio of the training and testing data is 8: 2.

3.1 Data Balancing

Unbalanced data typically refers to a problem with classification problems where the classes are not represented equally. The unbalanced data is solved using SMOTE also called as synthetic sampling. In both 3 and 5 category section, the SGPI range<=4 i.e class 0 has less number of instances comparatively. Since this class is important to identify weak students, it cannot be neglected. This technique is similar to oversampling where the instances of class 0 are increased to handle imbalance.

3.2 Feature Selection

The attributes such as name,email-id, phone-number, mother's name, father's name, address are unique and are dropped as it shows the least relationship with the target attribute. The Semester pointers and 12th board results and 10th board results showed a linear relationship with the target variable. Correlation states how the features are related to each other or the target variable.Thus, the correlation matrix was calculated for finding the relationship. For instance, the semester 5 SGPI range was to be predicted. So, it showed a strong relationship with the following features.

Fig -1: Result of correlation matrix using heatmap
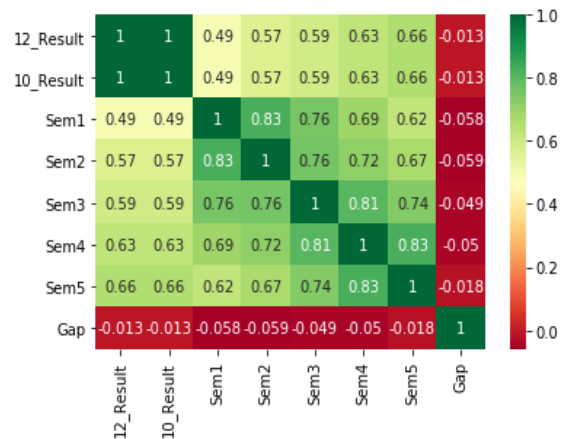
The features such as 12_Result,10_Result,Sem1,Sem2, Sem3, Sem4 were selected.Thus, the dataset was trained using these features. This is applicable for the dataset of both regular and Diploma students.

4. METHODOLOGY

3 category

The student's SGPI is converted into 3 categories. These categories are:

if sgpi<=4 the 0
If sgpi<=7 then 1
else 2

The categories for semester wise SGPI are predicted by applying ensemble algorithms -Random forest, Bagging and Boosting, rule-based algorithms- Naive Bayes, classification algorithms - decision tree and artificial neural networks- Convolutional neural network and Multilayer Perceptron.

5 category

The student's SGPI is converted into 3 categories. These categories are:

if sgpi<=4 the 0
If sgpi<=5 then 1
If sgpi<=6 then 2
If sgpi<=7 then 3
else 4

The categories for semester wise s are predicted by applying ensemble algorithms -Random forest, Bagging and Boosting, rule-based algorithms- Naive Bayes, classification algorithms - decision tree and artificial neural networks- Convolutional neural network and Multilayer Perceptron.

5. RESULTS

For Regular Students

The results of these algorithms were compared and the results are:

Table -1: Results of Algorithms

Algorithms	Accuracy	
	3 Category(*100)	**5 Category(*100)**
Random Forest	0.99	0.99
Gaussian Naive Bayes	0.78	0.87
Support vector machine	0.70	0.81
Bagging	0.99	0.99
Boosting	0.98	0.99
Decision Tree	0.98	0.98
Linear Regression	0.73	0.78
MLP	0.85	0.77
CNN	0.92	0.80

Chart -1: For 3 Category
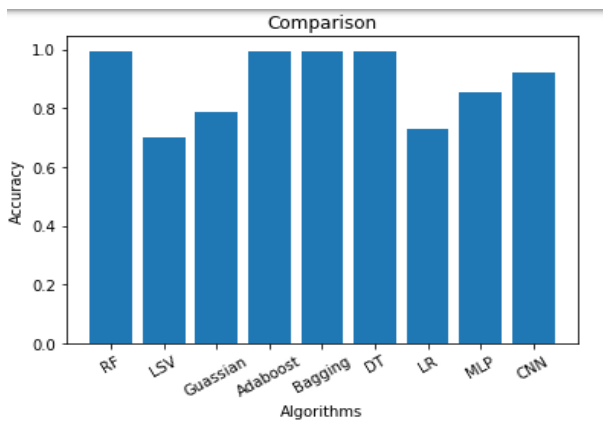
Chart -2: For 5 Category
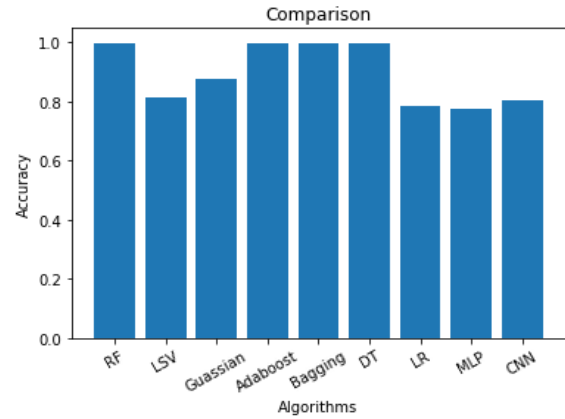
For Diploma Students

The results of these algorithms were compared and the results are:

Table -1: Results of Algorithm

Algorithms	Accuracy	
	3 Category(*100)	**5 Category(*100)**
Random Forest	0.99	0.99
Gaussian Naive Bayes	0.97	0.95
Support vector machine	0.93	0.88
Bagging	0.99	0.99
Boosting	0.98	0.99
Decision Tree	0.98	0.99
Linear Regression	0.73	0.79
MLP	0.62	0.77
CNN	0.88	0.78

Chart -1: For 3 Category

Chart -2: For 5 Category

6.   CONCLUSION

In this paper, the classification, rule-based learning, ensemble methods, and neural network based algorithms are employed in student information to predict the students' division on the premise of previous information. The accuracy was calculated using a confusion matrix. The confusion matrix proved to be a good metric for unbalanced data.
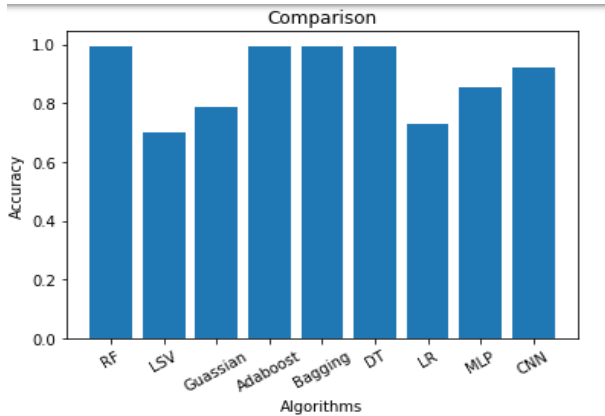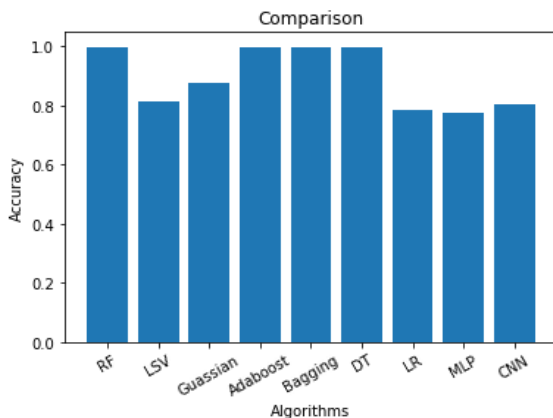
Ensemble methods work best with such structured data as compared to its counterpart i.e neural networks. The neural networks require a large amount of data for training.Since the dataset used is less complex and is in a structured format, ensemble methods work best as compared to Convolutional neural networks and Multilayer perceptron for both 3 categories and 5 categories prediction.This study may be quite helpful for both students and teachers for improving the performance and improving the future results of the weak students.

7.   ACKNOWLEDGEMENT

Our sincere thanks to Dr.Mrs. Shalu Chopra ,HOD of Department of Information Technology ,VESIT for providing the data about academic details of our college students .We are grateful to her for helping and guiding us throughout the project.

8.   REFERENCES

[1] Prediction of student performance using educational mining by Ms.Tismy Devasia, Ms.Vinushree T, Mr.Vinayak Hegde Department of Computer Science Amrita Vishwa Vidyapeetham University, Mysuru Campus -2016 International Conference on Data Mining and Advanced Computing.

[2] Proposed academic students' performance prediction model: A Malaysian case study.

[3] Performance Prediction of Engineering students using Decision trees .-R. R. Kabra, R. S. Bichkar -International Journal of Computer Application.

[4] Student performance analysis using classification techniques -International Journal of Pure and Applied Mathematics

[5] Data Mining Techniques for Predicting Student Performance -2015 IEEE International Conference on Engineering and Technology (ICETECH), 20th March 2015, Coimbatore, TN, India

[6] GritNet: Student Performance Prediction with Deep Learning -2018 Educational Data Mining Organisation,Mountain View , California

[7] Students' Performance Prediction Using Deep Neural Network -International Journal of Applied Engineering Research ,Volume 13, Number 2 (2018)

[8] Student Performance Prediction using Machine Learning -International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 IJERTV4IS030127 www.ijert.org (This work is licensed under a Creative Commons Attribution 4.0 International License.) Vol. 4 Issue 03, March-2015

[9] Early Prediction of Students Performance using Machine Learning Techniques-International Journal of Computer Applications (0975 – 8887) Volume 107 – No. 1, December 2014

[10] Using Machine Learning to Predict Student Performance-University of Tampere Faculty of Natural Sciences Software Development M. Sc. Thesis Supervisor: Jorma Laurikkala,June 17

[11] Machine Learning Approach to Predict Student Academic Performance-International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887 Volume 6 Issue IV, April 2018

[12] Predicting Students' Academic Performance in the University Using Meta Decision Tree Classifiers -Authors : Shanthini, A., G. Vinodhini and R.M. Chandrasekaran

[13] Machine Learning Based Student Grade Prediction: A Case Study -Authors : Zafar Iqbal , Junaid Qadir, Adnan Noor Mian and Faisal Kamiran

Analysis of Student Feedback using Deep Learning

Shejwal Asmita S
Jaihind College Of
Engineering, Kuran,
Pune

Deokar Anuja T
Jaihind College Of
Engineering, Kuran,
Pune

Dumbre Ashwini B.
Jaihind College Of
Engineering, Kuran,
Pune

Abstract: Feedback plays a key role in improving quality. To ensure improvement in teaching method and facilities provided by college, opinion of the students should be properly analysed and used. Text Sentiment analysis method are used to carry out such analysis. It can be performed in two ways - Machine Learning approach and Lexicon based approach. Presently, the teacher evaluation and feedback analysis are based on identifying student's opinion. Methods used for such classification are Naive Bayes, Voting ensemble method. Along with determining polarity, classifying feedback as strength, weakness and suggestions can improve to be more beneficial. Success of deep learning inspires us to propose a better and efficient system. The System that will use Word2Vec for text processing, Convolution Neural Network for automatic feature extraction. Supervised Support Vector Machine will be used for final classification. The Proposed system will result in classification of feedback as Strength , Weakness and Suggestions to faculty.

Keywords: Supervised learning, Convolution neural network, feature extraction, Sentiment Analysis, Deep learning.

1. INTRODUCTION

Youth is the hope and future of nations. Today's youth are students. Students play a vital role in society. It is primary necessity that education given to students needs to be quality education and must consider their views regarding it. Technology has allowed students to explore new fields. Hence considering student point of view to improve educational system can be beneficial. To keep track of performance and help improve abilities of faculty, opinions of students can be helpful. Finding out the subjective meaning from opinions is the major task. Sentiment analysis can be one way to do it. Sentiment analysis is contextual mining of text which identifies and extracts subjective information in source material, and helps to understand the social sentiment when monitoring an organization. Natural language processing (NLP) have to do with building of computational algorithms to automatically analyze and represent human language. The branch of natural language processing is shifting from statistical methods to neural network methods. Deep learning methods are achieving state-of-the-art results on challenging machine learning problems for example describing photos and translating text from one language to another. Our proposed system will be useful in educational sector. The feedback of students, collected in the form of text, will be analyzed using word2vec and CNN. The proposed system will result in classification of feedback as strength, weakness and suggestions to faculty.

2. LITERATURE SURVEY

Most existing sentiment analysis algorithms were designed for binary classification, meaning that they assign opinions or reviews to bipolar classes such as Positive or Negative. A series of experiments with convolutional neural networks built on top of word2vec are described in [2]. The results of experiment show that unsupervised pre-training of word vectors is an important ingredient in deep learning for NLP. In [7] the paragraph of sentences given by the customer is accepted and after extracting each and every word, they are checked with the stored parts of speech, articles and negative words. After checking against the database, Context free Grammar (CFG) is used to validate proper formation of the sentences.

In [3] Automatic evaluation system based on sentiments to overcome drawback of traditional questionnaire system. Feedback is collected in the form of running text and sentiment analysis is performed to identify important aspects along with the orientations using supervised and semi supervised machine learning techniques. It focuses more on subjective sentences and not on objective sentences. The scores are collected and aggregated to calculate final result. Term Frequency – Inverse Document Frequency (TF-IDF) and Naïve Bayes (Unigram, Bigrams) methods is used. It does not use advanced machine learning techniques and so the results were not accurate. Author of [4] states Text Mining techniques are broadly extended to classify the effective improvement of sentiment polarity analysis. Different techniques like Support Vector Machine (SVM), KNN and Decision tree are generally used but they are not always effective. Reducing the feature in data pre-processing stage and teaching sentiment analysis using voting ensemble method of machine learning are proposed and compared with existing typical machine learning for sentiment analysis.

The system achieves accuracy improvement of subjective polarity in sentiment analysis. Lack of weight assignment for feature extraction is observed. Methods like Naïve Bayes, ID3, J48 Decision Tree are used. The system described in [6] evaluates faculty and rates them with certain specified parameter to improve academic and education standard. The system is based on attribute and uses multipoint rating system. System uses text mining for deriving high quality information. Academic performance of students is considered when using the feedback given by them. Weights are assigned to feedback based on academic performance and sincerity. Multipoint rating is provided. The number of comparisons can be reduced in the system for more effective system. Naive Bayes method is used for text mining.

In [1] Pre-trained Word2Vec for text pre-processing and to gain vector representations of words which will be the input for suitable Convolutional Neural Network (CNN) architecture for deep features extraction is applied. Rectified Linear Unit and Dropout functions is used to improve the accuracy. Support Vector Machine classifier was used to predict the final classification. Author of [9] combines the advantages of CNNs and SVM, and constructs a text sentiment analysis model based on CNNs and SVM. The pre-trained word vector is used as input, and CNNs is used as an automatic feature learner, and SVM is the final text classifier. It is found that the accuracy of using CNN model results in better other models of depth learning, which shows that CNN model is more suitable to deal with text affective classification problem [8].

3. ARCHITECTURE

Students will give the feedback in the form of text and opinions regarding the faculty members and facilities provided by educational institute. The system will make use of Word2Vec for text processing. In word2vec, a distributed representation of a word is used. Take a vector with several hundred dimensions (say 1000). Each word is represented by a distribution of weights across those elements. So instead of a one-to-one mapping between an element in the vector and a word, the representation of a word is spread across all of the elements in the vector, and each element in the vector contributes to the definition of many words. Convolution Neural Network will the perform automatic feature extraction. Supervised Support Vector Machine will be used for final classification. Finally, the faculties will be notified about their strength and weakness and suggestion given by students [11].
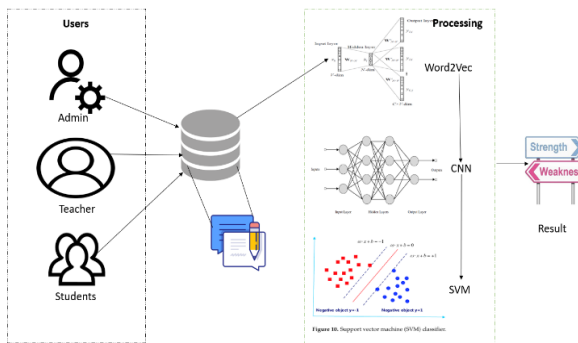

Figure 1. System Architecture

3.1 Word2Vec

Humans are good at alphanumeric but to process and analyse the huge amount of text devices like computer (processing unit) is necessary. But computers cannot understand human language. In order to process, sentiments they must be converted to numeric form. Word2vec is a particular machine learning model that produces word embedding. Word Embedding mean a word can be associated to a number. One can perform "math" with numeric values to find similarity between words. In neural network or word embeddings we define a model that aims to predict between a centre word Wt and Context words in terms of Word Vectors

The idea behind Word2vec is to predict words and its context word. The Word2Vec is based on two algorithms Skip gram and Continuous Bag of Words (CBOW). It is represented as P(context | Wt). The Skip gram predict context words given a target word whereas CBOW predict target word from bag of word context[10].

3.2 Convolutional Neural Network

CNN takes input data then performs convolutions on the data and apply pooling to data. To better understand what actually the CNN does consider a vector for sentence given as input to CNN. The vector is output of the word2vec text processing. A binary operation is performed to extract feature. Output of word2vec is first operand for operation and filter second. The sequence generated after operation is the feature map. The convolution process can be explained using diagrammatic representation as shown below. Example: "He is experienced and knowledgeable.". The vector for given sentence is considered as input to the CNN model. A 2-word filter is applied to the sentence[9].

Step 1: The filter convolves over the first to words and feature map value is calculates. That is if words in sentence are Wt , Wt+1……., WT. The convolution operation is carried out on Wt, Wt+1 as shown in figure.
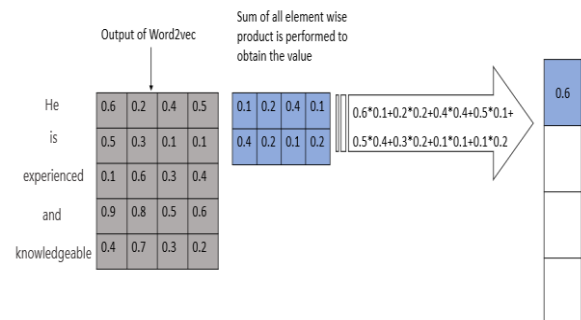

Figure 2. CNN Example step 1

Step2: In second step the sliding window shifts on place forward and calculate the feature map value for next two words i.e. Wt+1 and the next word.
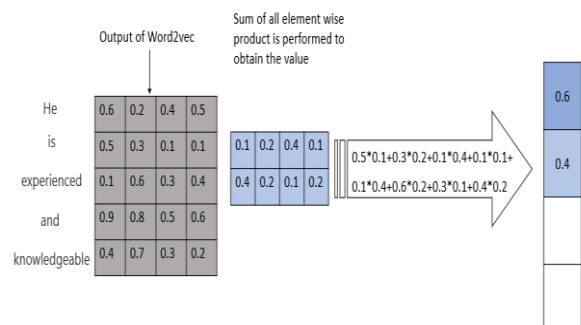

Figure 3. CNN Example step 2

The process continues till the last two words. The feature map for "He is experienced and knowledgeable" is shown in figure
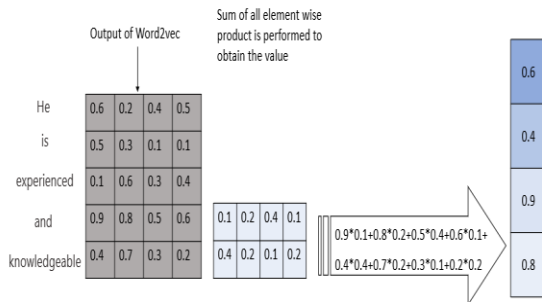

Figure 4. CNN Example step 1

Layer for automatically features extraction using three convolution kernels (convolution filters) of different sizes, ReLu Layer, Pooling Layer with nonlinear sampling method in order to decrease the number of characteristic parameters and prevent overfitting, and Fully-Connected Layer[2].

3.3 Support Vector Machine

SVM is a type of classification algorithm which classifies data based on its feature. It is type of binary classifier. Convolution neural network can extract meaningful feature representation from input samples effectively, but the classification ability of fully connected classification layer is weak for nonlinear separable data. SVM is a supervised machine learning model, which is two-classification model. The SVM method is based on the theory of statistical learning theory and the principle of minimum structural risk. According to the limited data information, SVM try to find the best compromise between the complexity and the learning ability, in order to get the best generalization ability. SVM can find the optimal classification surface for the characteristics. As shown in figure it considers the best line of separation between data points. In figure the blue circles represent strengths and orange block represent weakness. When a new point is detected it is checked across hyperplane of SVM. The hyperplane determines whether it is strength and weakness[12].
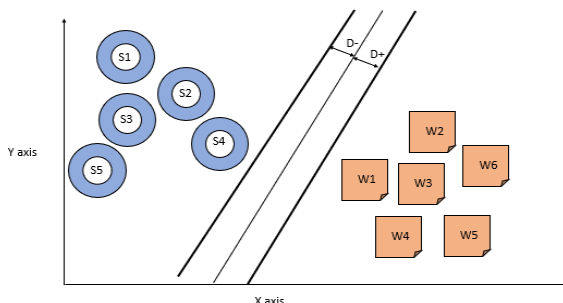
Figure 5. SVM Classification

4. MATHMATICAL MODEL

System Description:

Let S be a closed graph system that draws the object; such that S= I,P,R,O ?s where I represents the set of Input; I =i0 ?i and P is a set of Process; P = p0,p1,p2,?p and R is set of rules R = r0,r1 and O is set of Outputs; O= o0,o1

Where,

- I=Input

  I= I1

  Where,

  I1= Student Feedback

- P= Process

  P= P1, P2, P3

  P1= Student Process

  P11= Student Registration

  P12= Student Login

  P13= For particular subject i.e. for respective teacher Fill Feedback Form

  P14= logout

- P2= Admin Process

  P21= Admin Login

  P22= Manage Feedback

  P23 = Manage User and Full Application

  P24= Analysis Feedback Report

  P25= Analysis Faculty Report

- P26= Logout

  P3= Teacher Process

  P31= Teacher login

  P32= View Feedback (Strength/ Weakness/ Suggestions)

  P33= View Ranking

- R=rules

  R1=Login Validations

  R2= Validate type of Access (Student/ Admin/ Faculty).

- O=Output

  O1= Generate Feedback Type report

  O2 = Generate Class Report

Activity1 Let fi be a rule of I into P such that student will feed the data into the system. it returns

fI(i0) ¿ p0,p1, p2 P.

Activity2 Let fp be a rule of P into R such that the

processed data will go through particular

set of rules it returns

fp(p0) ¿ r0,r1 R.

Activity3 Let fr be a rule of R into O such that the

processed data will give the respective

output. it returns

for(R0) ¿ O0,O1 O. As described in activity 1 following vein diagram is drawn: (Analysis: The system will give the result of feedback based on the data feed by the registered student. The student will select the respective subject/teacher then will answer respective questions and also give suggestion. The process will analysis this feeds and generate the results. Admin and teacher will be able to see the response of the students. Text of State Diagram: State Diagram of the system is shown below:

Where,

Q0=Student Feedback

Q1= System Analysis

Q2= Admin Login

Q3= Teacher Login

Q4= Background System Data Processing

(Tokenization, Pre-processing, Filtering,

Classification, prediction, comparison)

Q5= Result Display

5. RESULT AND CONCLUSION

Students Feedback analysis using deep learning can be useful in educational sector. The expected result is a classification of the students feedback in strength and weakness of teachers.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] A bubakr H. Om babi, O nsaLazeez, Wael Ouarda, Adel M. Alimi "Deep Learning Framework based on Word2Vec and CNN for Users Interests Classification" 978-1-5386-0667-4/17/$31.00 ©2017.

[2] Yoon Kim "Convolutional Neural Networks for Sentence Classification" Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1746–1751,October 25-29, 2014, Doha, Qatar. ©2014.

[3] Alok Kumar, Renu Jain "Sentiment Analysis and Feedback Evaluation" 978-1-4673-6747-9/15/$31.00 2015.

[4] Chakrit Pong-Inwong, KonpusitKaewmak "Improved Sentiment Analysis for Teaching Evaluation Using Feature Selection and Voting Ensemble Learning Integration" 978-1-4673-9026-2/16/$31.00 ©2016.

[5] Tanzim Mahmud, K. M. Azharul Hasan, Mahtab Ahmed, Thwoi Hla Ching Chak "A Rule Based Approch for NLP Based Query Processing" 978-1-4673-9257-0/15/$31.00 ©2015 IEEE.

[6] Krishnaveni K S, Rohit R Pai, Vignesh Iyer "Faculty Rating System Based on Student Feedbacks Using Sentimental Analysis" 978-1-5090-6367-3/17/$31.00 ©2017.

[7] Biswarup Nandi, MousumiGhanti, Souvik Paul "TEXT BASED SENTIMENT ANALYSIS" 978-1-5386-4031-9/17/$31.00 ©2017.

[8] Fan Xia, Zhi Zhang "Study of Text Emotion Analysis Based on Deep Learning"978-1-5386-3758-6/18/$31.00 c 2018.

[9] Yuling Chen, Zhi Zhang "Research on text sentiment analysis based on CNNs and SVM" 978-1-5386-3758-6/18/$31.00 c 2018.

[10] Tomas Mikolov, Kai Chen, Greg Corrado, and Je_rey Dean. E_cient "Estimation of word representations in vector space." 01 2013.

[11] Shejwal Asmita S, Deokar Anuja T, Dumbre Ashwini B, Prof.S.S.Gore, "REVIEW ON STUDENT FEEDBACK ANAYSIS USING DEEP LEARNING" IJRAR1904692 © 2018 IJRAR November 2018, Volume 5, Issue 4.

[12] Walaa Medhat a,*, Ahmed Hassan b, Hoda Korashy b " Sentiment Analysis Algorithms and applications: A survey " Ain Shams Engineering Journal (2014) 5, 1093–1113

Segmentation of Brain MR Images with ITK Toolkit

P. Narendran
Associate Professor
Head, Department of Computer Science
Gobi Arts and Science College Gobi, India

Dr V Thiagarasu
Associate Professor and Principal
Department of Computer Science
Gobi Arts and Science College Gobi, India

Abstract: Segmentation of nontrivial images is one of the most difficult tasks in image processing. Segmentation accuracy determines the success or failure of computerized analysis procedures. Medical image segmentation has become an important diagnostic tool in the practice of modern medicine. Segmentation of MRI Brain images is the delineation of neuro-anatomical structures such as the Cerebrum, Cerebellum, Hippocampus, etc. as well as abnormalities such as tumors. Although numerous methods have been proposed during the past two decades for brain MRI segmentation, it remains a challenging task. In this paper brain MRI segmentation is demonstrated for delineation of a tumor, ventricles and other anatomical structures using Insight Segmentation and Registration Toolkit (ITK) routines as the foundation. This public software provides callable routines only. It does not provide graphics, visualization, or other interactive tools specific to a given task, such as segmentation. Segmentation of medical images is an important step in forming realistic tissue models. To assist in classifying the relevant literature, three generations of development are identified where each generation adds an additional level of algorithmic complexity. The first generation is composed of the simplest forms of image analysis, the second is characterized by the application of uncertainty models and optimization methods, and the third incorporates knowledge into the segmentation process. The progress toward fully automatic segmentation is discussed and sources of segmentation software from industry and academia are identified, along with databases for segmentation validation. We have implemented a variety of semiautomatic and automatic segmentation routines as well as preprocessing algorithms provided by ITK. This implementation facilitates interactive adjustments of algorithm parameters and their consequences on the segmentation results via an intuitive GUI incorporating the ITK segmentation routines.

Keywords: Brain, Image Segmentation, MRI, ITK, CMake, Tumor and Filter

1. INTRODUCTION

Image segmentation is a process to identify and classify regions of interest from digital images. Typically, digital images are acquired from medical instrumentation as CT (Computer Tomography) or MRI (Magnetic Resonance Image) scanners, digital mammograms etc. Image segmentation plays an important role in medical image processing which enables a variety of clinical applications [10]. It is also a tool to facilitate the detection of abnormalities such as cancerous lesions in the brain. Segmentation of medical images is a challenging task. A variety of different methods have been proposed and implemented in recent years. Although numerous efforts have advanced this technique, there is no single approach that can generally solve the problem of segmentation for the large variety of image modalities existing today [4].

Various segmentation techniques can be classified, for example, as classical techniques such as thresholding, boundary based technique, region based technique, or statistical technique. Depending on the level of interactivity, segmentation can be classified as manual, semiautomatic, or automatic. Manual segmentation is time consuming [2, 6], costly, and non-repeatable. Inconsistencies in the segmented extent of various structures are common among qualified experts. Moreover, it does not use the full multi-dimensional image data. Automatic segmentation methods are sensitive to noise and unexpected situations, leading to errors. Their advantages are a minimum time commitment from the user and results that are highly reproducible, albeit potentially erroneous. These methods are usually problem-specific, and an image-processing expert is needed to determine which image-processing functions are best suited for a given segmentation task.

Semi-automatic image segmentation combines manual interaction with automated sub-components to solve segmentation problems. Semi-automatic segmentation is faster than a manual strategy with more reproducible results compared to both manual and fully automatic. It uses the complete image data set, and is minimally affected by human inconsistency and error. In addition, the user can harness the power of automated image-processing algorithms without being an image-processing expert

2. BACKGROUND OF THE STUDY

There are multiple medical imaging strategies such as MRI (Magnetic Resonance Imaging), PET (Positron emission Tomography), CT (Computer Tomography), Ultrasound, SPECT (Single Photon Emission Computed Tomography and many more. This thesis has concentrated on MRI imaging due to the research focus of our laboratory, the Center for Comparative Neuro Imaging.

Magnetic resonance imaging (MRI) is an imaging technique used primarily in medical settings to produce high quality images of soft tissue structures. Unlike conventional X-ray imaging or Computed Tomography, which produce images that show the X-ray attenuation of tissues, MRI measures the amount of hydrogenous materials (water and lipids) in tissues. The nucleus of the hydrogen atom is a spinning charged proton with magnetic properties which the MR imaging strategy utilizes.

2. Two magnetic fields are used in MRI. The first being a strong static magnetic field which causes the hydrogen atoms

in the body to align in a direction parallel to the field. A second magnetic field (radio-frequency pulse) is applied at right angle to the first field causing the hydrogen atoms to change their alignments. When radio-frequency pulse is turned off, the hydrogen atoms return to their alignment along the static magnetic field direction. The time rate of recovery differs with tissue properties. Longitudinal relaxation (T1) depends on the recovery time of the hydrogen atoms to return to the axis of the primary magnetic field. The transverse relaxation time (T2) measures the rate of decay of the hydrogen proton alignment due to the RF pulse.

3. APPROACH

Image segmentation is a process to identify regions of interest from digital images. Image segmentation plays an important role in medical image processing which enables a variety of clinical applications. It is also a tool to facilitate the detection of abnormalities such as cancerous lesions in the brain. Although numerous efforts in recent years have advanced this technique, no single approach solves the problem of segmentation for the large variety of image modalities existing today. Consequently, brain MRI segmentation remains a challenging task. The purpose of this paper is to demonstrate brain MRI segmentation for delineation of tumors, ventricles and other anatomical structures using Insight Segmentation and Registration Toolkit (ITK) routines as the foundation. ITK is an open-source software system to support the Visible Human Project. Visible Human Project is the creation of complete, anatomically detailed, three-dimensional representations of the normal male and female human bodies. Currently under active development, ITK employs leading-edge segmentation and registration algorithms in two, three, and more dimensions. A goal of this thesis is to implement those algorithms to facilitate brain segmentation for a brain cancer research scientist.

4. SEGMENTATION METHODS

Several methods of image segmentation have been proposed in literature [1]. The following figure 1 arranges these methods.

4.1 SINGLE CONTRAST METHODS

4.1.1 Thresholding: Thresholding is the most intuitive approach to segmentation where values of upper and lower thresholds are provided by the user. This method is limited and successful application for clinical use is hindered by the variability of anatomy and image artifacts.

4.1.2 Edge Based Segmentation: Edges are usually detected by rapid changes in intensity values observed while traversing a set of pixels. Operators such as first and second derivatives play significant role in edge detection. Sometimes, insufficient intensity gradients, presence of noise or artifacts, or poor guess of threshold may cause over or under segmentation.

4.1.3 Boundary Tracing: In this method, user selects the pixel on the boundary of the region for outlining purposes. Then the method follows the boundary from this user selected seed point. This method is useful for cases where good definition of an edge boundary exists. This method can cause problems for images with tissue variance.

4.1.4 Seed Growing Segmentation: In this method, user selects a seed or multiple seeds and specifies threshold value.

Using this data, the method examines neighboring pixels of the selected seed(s) and includes them in the region if those neighbors also satisfy the criteria.

4.2 MULTI-SPECTRAL SEGMENTATION

These methods are classified as supervised and un-supervised. Supervised methods require operator input for segmentation. This is done by selecting training pixels or training regions in the images. Un-supervised methods define regions in the image without operator input.

4.3 SUPERVISED METHODS

4.3.1 Pattern Recognition Methods: This is the most common approach for multi-spectral segmentation. There are several pattern recognition techniques and many of them assume particular distribution of the features based upon parametric models. Non-parametric models do not rely on predefined distributions but on the actual distribution of the training samples [9].

4.3.2 Algebraic Methods: For images that clearly identify signature vectors, these methods provide an elegant solution to deal with the partial volume effect, which might have influence on volume measurements. Algebraic approaches can become impractical for images showing complex pathology.

4.4 UNSUPERVISED METHODS

Unsupervised methods, also called "clustering" find the structure in data automatically. A cluster is an area in feature space with high density and can be promising for tumor volume determination. However, the initialization is very important for meaningful clustering and reasonable computation time. These methods are reproducible but may not necessarily arrive at meaningful segmentation and often require computation time.

ITK (Insight Segmentation and Registration Toolkit) has implemented some segmentation methods discussed. ITK is an open-source software system to support Visible Human Project.

5. ITK

Insight Segmentation and Registration Toolkit (ITK) is an open-source software system to support the Visible Human Project. ITK is sponsored by National Library of Medicine at the National Institutes of Health and developed by six principal organizations, three commercial (Kitware, GE Corporate R&D, and Insightful) and three academic (UNC Chapel Hill, University of Utah, and University of Pennsylvania). ITK is currently advancing leading-edge segmentation and registration algorithms in two, three, and more dimensions. ITK does not provide GUI or Visualization. ITK is cross-platform package. A common build environment called CMake is used to manage the compilation process [8].

CMake is a cross-platform open source software system. It was developed by Kitware, as part of the ITK project. CMake is a build process which is independent of the operating system and compiler. Currently it supports UNIX and Windows platforms and produces native build files appropriate to these OS. On Unix CMake produces makefiles and on Windows, it generates projects and workspaces. CMake allows the user to control the software compilation process using simple platform and configuration files (contain pre-defined CMake commands as well as user-defined commands) that are compiler independent. In addition, an automated wrapping process generates interfaces between C++ and interpreted programming languages such as Tcl, Java, and Python. Wrapping process is achieved using

CableSwig. This ITK utility enables developers to create software using a variety of programming languages.

6. ITK SEGMENTATION

Image segmentation plays an important role in medical image processing. The goal of segmentation is to extract one or several regions of interest in an image. Depending on the on the context, a region of interest can be characterized based on a variety of attributes, such as gray scale level contrast, texture, shape, size etc. Selection of good features is the key to successful segmentation. There are a variety of techniques for segmentation, ranging from simple ones such as thresholding, to more elaborate strategies including region-growing, edge-detection, morphological methods, artificial neural networks and much more. Although numerous methods have been proposed in recent years, there is no single approach that can solve segmentation problems for many existing imaging modalities today [7].

The Insight Toolkit provides a basic set of algorithms that can be used to develop and customize a full segmentation application. The most effective segmentation algorithms can be obtained by carefully customizing combinations of the basic segmentation algorithms and pre-processing and post-processing algorithms implemented in ITK for images of specific modality. The parameters of these algorithms can be tuned for the characteristics of the image modality used as input and the features of the anatomical structure or abnormalities to be segmented [3].

7. THRESHOLD SEGMENTATION

Following section presents the results of applying Connected Threshold Segmentation on Brain MRI to segment tumor and white matter.

7.1. Segmenting Tumor

In following figures, image in figure 2 (a) represents original brain MRI. Figure 2 (b) shows the result of applying curvature flow smoothing with Time Step = 0.125, Number of Iterations = 5. In second row figure 2 (c) was obtained by applying Connected Threshold Segmentation on smoothed image with Lower Threshold = 169, Upper Threshold = 182 and Seed Index (76, 91). Second result in figure 2 (d) was obtained by applying Curvature Flow Smoothing twice, with Time Step = 0.125 each time and Number of Iterations = 5 and 10 respectively. Then Connected Threshold Segmentation was applied with Lower Threshold = 169, Upper Threshold = 182. Dilation was applied to this above result as shown in figure 2 (e). Figure 2 (f) was obtained by applying Edge preserving smoothing (Curvature Anisotropic diffusion with time step = 0.125, conductance = 1, number of iterations = 5) and then applying Connected Threshold to this smoothed result with lower threshold = 50, upper threshold = 90 and Seed Index (76, 91). Figure 2 (g) shows the result of applying dilation to the segmented result in figure 2 (f).

Now we apply Normalization to the image smoothed with curvature anisotropic diffusion, as shown in Figure 2 (h). But as seen from result figure 2 (i), normalization does not contribute to any significant improvement in the result of connected Threshold Segmentation. Hence, hereafter, we have used Normalization in some cases for display purpose only [6].

As we can see here, segmented result does not cross the external boundary of the tumor. Gray matter is not being completely segmented. This illustrates the vulnerability of the region growing methods when the anatomical structures to be

segmented do not have a homogeneous statistical distribution over the image space.

One solution will be to apply level set segmentation to the above result. This result can be dilated so as to cross the internal boundary of the tumor, bring it close enough to the external boundary of the tumor and then apply the level set segmentation such as laplacian or canny edge detection to refine the result.

(a) (b)

(c) (d) (e)

(f) (g) (h)

(i)
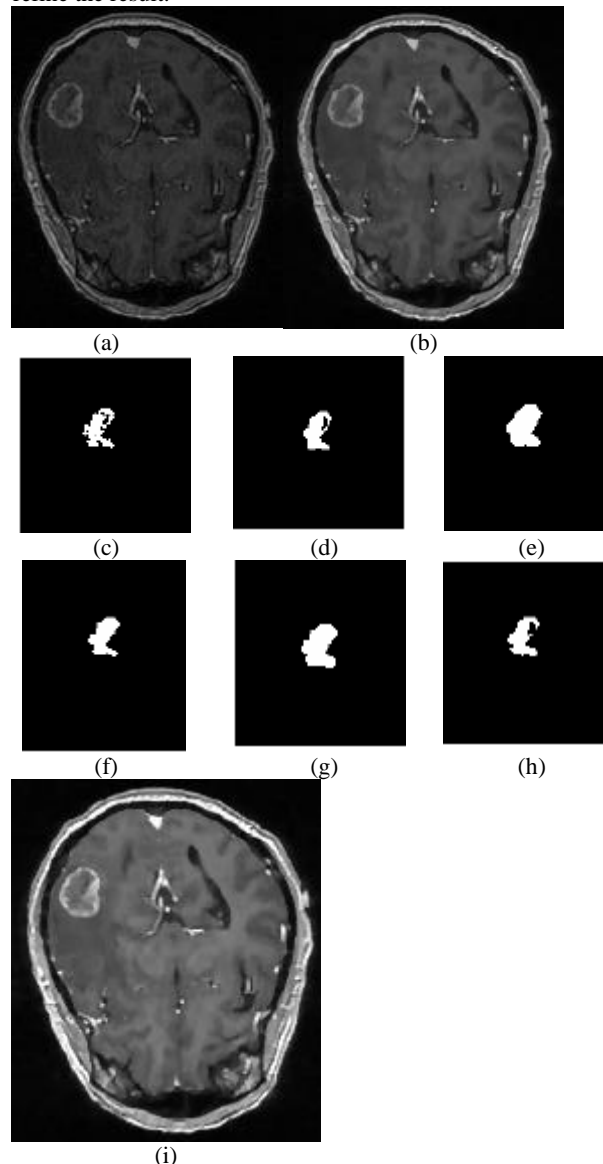
Figure 2 (a) – (h) Segmenting tumor with connected threshold filter

Other level set segmentation algorithms such as shape detection, geodesic active contour may be applied to obtain the desired segmentation.

7.2. Segmenting Brain (White Matter)

Figure 3 (a) shows image smoothed original brain MRI of 256*256.. Figure 3 (b) shows brain area segmented with Connected Threshold with lower threshold = 45 and upper threshold = 55 and Seed Index (160,115). Here, smoothing was not applied.

As seen in the results, this filter produces the output with holes. There are different approaches that we can use to fill above holes [5]:

    1. Use dilation operation.

    2. We can apply flood fill operation using itk::GrayscaleFillholeImageFilter

3. If size is enough for characterizing the holes we can use the connected components filter (itk::ConnectedComponentImageFilter) and then collect the number of pixels per component in order to identify the labeled regions those need to be set ON to fill the holes.

4. We can use the Connected Threshold Image Filter and use the first pixel of the image as seed point (assuming it is in the exterior of the shape) and segment the "exterior" of brain area.

Then make a logical OR between the exterior and the shape, which will give the negative map of the holes. Negating this map and do an OR between the map and the original shape, you will end up with the shape without holes.

Figure 3 (c) shows the result after applying dilation with X-radius=1, Y-radius=1. Here, most of the holes disappear, but still some holes are left and it changes the segmentation result Figure 3 (d) shows the result of applying GrayScaleFillHoleImagefilter to 3 (b). With this filter, all holes are filled into ON region, including ventricles area.

Figure 3 (e) shows the result of Connected Threshold segmentation with lower threshold = 39 and upper threshold = 56 and figure 3 (f) is the result of applying GrayScaleFillHole filter to the result in figure 3 (e).

In all above cases, smoothing was not used. Noise present in the image can reduce the capacity of this filter to grow large regions. When faced with noisy images, it is usually convenient to pre-process the image by using edge-preserving smoothing filters such as Curvature Anisotropic Diffusion Filter, Gradient Anisotropic Diffusion Filter [2].

(a)   (b)

(c)   (d)

(e)   (f)

(g)   (h)

(i)   (j)
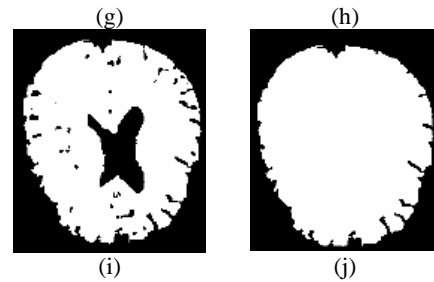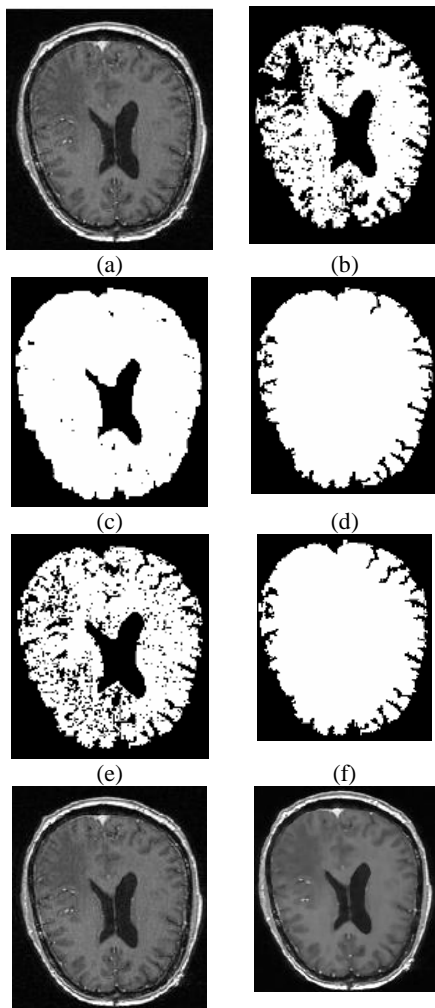
Figure 3 (a) - (j) Segmenting brain with connected threshold filter with smoothing

Here, figure 3 (g) represents original image, figure 3 (h) represents image smoothed with curvature anisotropic Diffusion with No. of Iterations = 5, Time step = 0.125 and Conductance = 1. Figure 3 (i) is the image segmented with Connected Threshold filter with Seed Index (160, 115) and Lower Threshold = 80 and Upper Threshold = 125. This result can be improved with GrayScaleFillHole filter as shown in figure 3 (j).

8. IMPLEMENTATION

This section describes the programs developed within this paper and the ITK routines used within the programs. All programs were written in C and C++ compatible on Windows operating systems.

Configuration files (Cmakelists.txt) were written and programs compiled and built using CMake 2.0 and Microsoft Visual C++ 6.0. Typical inputs for these programs were 2-D brain slices with the following data characteristics:

- Endian Type = Big Endian
- Element Spacing=(X=1.0156,Y=1.0156)
- Dimension Size = 256 * 256
- Element Type = 16-bit Signed Short

Original brain MRI slices, with details mentioned previously were converted into MetaImage format. This format uses an ASCII text MetaImage header file and a binary raw data file for each slice. This format was directly compatible with the ITK Input Reader (itk::ImageFileReader). Results (at intermediary times) were written to a file using the ITK Output Writer (itk::ImageFileWriter). Both ITK routines handle multiple file I/O file formats.

The actual low level task of reading and writing specific file formats was done by super class itk::ImageIO. Input images to the programs were specified with .mhd or .mha extension. MetaImageIOFactory was registered with objectFactoryBase. This allowed run-time instantiation of the classes that supported MetaImage format, based on the extension specified for an input file.

The IO architecture of the ITK makes it possible to avoid explicit specification of the file format used to read or write images. The object factory mechanism enables the ImageFileReader and ImageFileWriter to determine the file format at run-time. Typically, file formats are chosen based on the filename extension, but the architecture supports arbitrarily complex processes to determine whether a file can be read or written. Alternatively, the user can specify the data file format by explicit instantiation and assignment of the appropriate itk::ImageIO subclass. Following example file shows the header files associated with brain MRIs as input.

- ObjectType = Image
- NDims = 2
- BinaryData = True
- BinaryDataByteOrderMSB = True
- Color = 1 0 0 1
- ElementSpacing = 1.01563 1.01563
- DimSize = 256 256
- ElementType = MET_SHORT

➢ ElementDataFile = tumour75.raw

All output files were written as RAW files along with their MetaImage header files that consisted all the information

9. CONCLUSION

Image segmentation plays an important role to facilitate the detection of abnormalities such as cancerous lesions in the brain MRI. Although numerous efforts in recent years have advanced this technique, there is no single approach that can generally solve the problem of segmentation for the large variety of image modalities existing today. We have demonstrated the segmentation of brain MRI for detection of abnormalities as well as ventricles and white matter in the brain. Different basic segmentation filters implemented in ITK were used in pipeline for brain MRI segmentation.

In general, presented results showed that the application of region growing, and combination of region-growing and level set techniques implemented in ITK proved to be efficient for the task of tumor, ventricles and white matter segmentation. Setting the seed points for the fast marching segmentations for initial contour generation provided good results. There was a large improvement concerning quality with level set methods, compared to using only basic segmentation techniques like region growing.

In conclusion, ITK segmentation filters offered a large number of state of the art algorithms implemented in a very consistent and meaningful coding style. With the combination of techniques discussed above, it is important to get some experience in setting the parameters correct in order to get the best results for brain MRI segmentation using the ITK filters.

10. REFERENCES

[1] Pham, DL, Xu, CY, and Prince, JL, "Current methods in medical image segmentation," *Annual Review of Biomedical Engineering,* Vol. 2, pp. 1, 2000.

[2] Tomi Heinonena, Prasun Dastidarb, Harry Freyc, Hannu Eskolaa, "Applications of MR Image Segmentation", *International Journal of Bioelectromagnetism*, Vol. 1, No. 1, 35-46, 1999.

[3] Luis Lbanez and William Schroeder with Lydia Ng and Josh Cates, *The ITK software Guide ITK 1.4*, Kitware Inc 2003, pg 315.

[4] Rafael Gonzalez and Richard E. Woods, *Digital Image processing*, second edition, Pearson Education Inc. 2003, pgs 613.

[5] Tomasi, C. and Manduchi, R., "Bilateral Filtering for gray and color images", *Proc. of the IEEE International Conference on Computer Vision*, 1998.

[6] D. Adalsteinsson and J.A. Sethian, *A fast level set method for propagating interfaces, J. Comput. Phys.*, 118 (1995), pp. 269{277.

[7] F. Meyer and S. Beucher, "Morphological segmentation", *Journal of Visual Communication and Image Representation,* 11, No 1:21{46, 1990}.

[8] F. Meyer, A. Oliveras, P. Salembier, and C. Vachier "Morphological tools for segmentations: connected Filters and watershed", *Annals of telecommunications,* 1997.

[9] J.A Sethian, "Fast Marching Method"*,* SIAM Review, *1999 Society for Industrial and Applied Mathematics*, Vol. 41, No. 2, pp. 199–235.

[10] Maintz, J. and Viergever, M. (1998). A survey of medical image registration. Medical Image Analysis, 2(1):1–36.

Application of Knowledge Management System Using Influence of Inukshuk and Kano Model (Case Study : Palembang Private Higher Education)

Orissa Octaria
Departement of Informatics
Faculty of Computer Science
Universitas Sriwijaya
Palembang, Indonesia

Ermatita
Departement of Informatics
Faculty of Computer Science
Universitas Sriwijaya
Palembang, Indonesia

Sukemi
Departement of Informatics
Faculty of Computer Science
Universitas Sriwijaya
Palembang, Indonesia

Abstract: Knowledge management is needed in any field without the exception of education. This research developed the Knowledge Management process by combining the Inukshuk model and the Kano model. The purpose of this study is to facilitate knowledge management control and combine the two models to produce a KMS prototype that is well suited to the needs of the user. Inushuk is an expansion of KMS with the addition of Leadership, Culture, and Technology components to measure KMS needs to be implemented into a KMS prototype feature. The Kano model serves to filter features into the Kano category namely Must Be, performance, and Delighter. The data analysis process of the Inukshuk model uses index data analysis while the Kano model in analyzing the data uses the Kano calculation method. The results of the study of combining the two models resulted in a KMS model that could be used as a reference for the development of KMS for STMIK MDP, CANDRADIMUKA STISIPOL, and UIGM in managing knowledge. Combining the two methods above produces Chat features, Discussion Forums, Q n A and Document Search.

Keywords: Knowledge Management, Inukshuk, Kano, Combining, Technology

1. INTRODUCTION

In this time knowledge management is needed in any field without exception in the world of education. Sometimes the problem faced is the difficulty of getting a knowledge that was previously known. This is due to the lack of knowledge management of an organization to summarize the problems that have been resolved. Knowledge management is needed to summarize existing knowledge or new knowledge acquired in the present, and can be used in the future. Understanding of knowledge here is knowledge, experience, factual information and opinions of experts [1]. As said before, an organization must be skilled in managing existing knowledge, so that later it can be used in the future and remain useful for a company or organization.

Understanding and experience are built on knowledge, whether manifested in an individual or inherent in the actual processes and applications of an organization. Knowledge management can also be a decision-making system. The decisions that will be taken can be summarized from the management of knowledge that has been made.

In this paper, knowledge management will be made in the field of education. Starting from how the teaching staff can continue to teach well and can save the knowledge of teaching to young teachers, so as to achieve customer satisfaction in this case students. Until the change of position or structural in a college that makes it easier for new officials to understand what authority and responsibility must be considered?

This study designed the KMS of the Inukshuk and Kano models. The Inukshuk method itself is often used because the method in knowledge management is an extension of the SECI method, namely by adding several components such as Leadership, Culture, and Technology. The Kano diagram itself is a diagram that divides customer specifications into

three types, must be, performance and delighter and compares the existence of a specification [2]. Both of these models have their respective advantages, therefore, this study combines the two models in managing knowledge in the Palembang private university. The formulation of the generated KMS model will be formed into a prototype that can be evaluated according to the needs of the knowledge management process at the high level. To be easily accessed and used by teaching staff from web based. The purpose of this study was to create a KMS model at the private tertiary institution of the city of Palembang along with the making of a KMS prototype to facilitate knowledge management control with the Inukshuk and Kano approaches.

2. LITERATURE REVIEW

2.1 Knowledge Management System

According to [3], a knowledge management system (KMS) is a system that can be used to facilitate knowledge management processes, so that KMS can be divided into each knowledge management process, namely knowledge discovery system, knowledge capture system, knowledge sharing system and knowledge application system.

Starting from Knowledge Management, Knowledge Management (KM) is an expression that describes a series of strategies, systems and techniques used by individuals, teams and corporations to manage knowledge. There are various definitions of knowledge management and also the definition of knowledge that is developing but has not yet reached a global agreement. Knowledge management is a topic that continues to look for forms among researchers. KM in Indonesian can also be called *Management Pengetahuan* (MP) and is discussed directly by academics in knowledge management, and to answer the needs of practitioners who want a simple picture of knowledge management. Knowledge

management is defined as the concept of knowledge management which includes the following knowledge management processes [3].
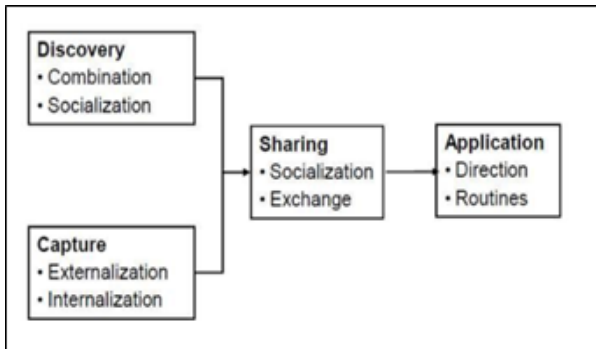

Figure. 2.1 Knowledge Management Process [3]

a. *Knowledge Discovery*

Defined as the development of new tacit knowledge or explicit knowledge, from data and information or from prior knowledge integration. New explicit knowledge, found through a combination of prior knowledge of knowledge. While new tacit knowledge is found through socialization between individuals.

b. *Knowledge Capture*

Defined as the process of receiving explicit and tacit knowledge. The knowledge capture process comes from two Knowledge Management sub processes, namely externalization and internalization.

c. *Knowledge Sharing*

it is the process of communication with explicit or tacit knowledge between individuals. This process involves two sub-processes, namely socialization and exchange. Socialization focuses on tacit knowledge sharing, while exchange focuses on explicit knowledge.

d. *Knowledge Application*

it Depends on the knowledge available in the previous knowledge process, namely knowledge of discovery, capture, and sharing. In the Knowledge Management process there are sub-processes of direction and routines. Direction refers to the process by which individuals have direct knowledge of the actions of other individuals without sharing with the individual, the knowledge underlying the direction of the action. Routines are related to the use and understanding of explicit knowledge that has been established in the organization.

2.2 INUKSHUK

Nonaka and Takeuchi (1995) [4] had proposed the SECI model which had become the foundation of knowledge creation and transferring theory. This model distinguishes two dimensions of knowledge as tacit and explicit knowledge, and proposes a process of knowledge creation through social interaction to convert knowledge between two dimensions.

Knowledge was divided into tacit knowledge and explicit knowledge as two main types of human knowledge [4]. There are also 4 processes of knowledge conversion, namely socialization, externalization, combination, and internalization or often referred to as SECI. The knowledge conversion can be drawn in the following spiral knowledge [4].
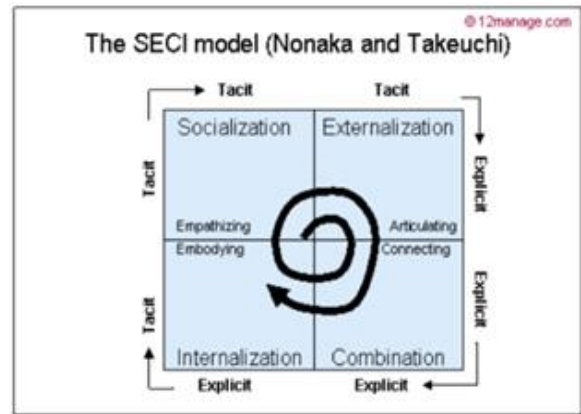
Figure. 2.2 SECI Model [4]

a. Socialization

Conversion from Tacit Knowledge to Tacit Knowledge. Tacit Knowledge is delivered to others through the socialization process in organizations. The socialization process can be done through social interaction and sharing experiences between members of the organization.

b. Externalization

Conversion from Tacit Knowledge to Explicit Knowledge. Externalization is the process of converting hidden knowledge into actualized (explicit) knowledge. For example experience, ideas or desires become a source of knowledge from prototypes, models, books, blogs, and others.

c. Combination

Explicit Knowledge conversion to Explicit Knowledge. Combination is the concept of creating explicit knowledge by combining, categorizing, and collecting two existing or more explicit knowledge.

d. Internalization

Explicit Knowledge Conversion to Tacit Knowledge. Explicit Knowledge is learned to then proceed with new ideas or actions.

The Inukshuk Knowledge Management model is an enhanced framework of the SECI model with the addition of components such as Leadership, Culture and Technology. The connection with Knowledge Management is that it can provide information about Tacit and Explicit Knowledge in the organization [5].

Figure. 2.3 Inukshuk Knowledge Management Model

Inukshuk Knowledge Management model stages in detail per stage as follows:
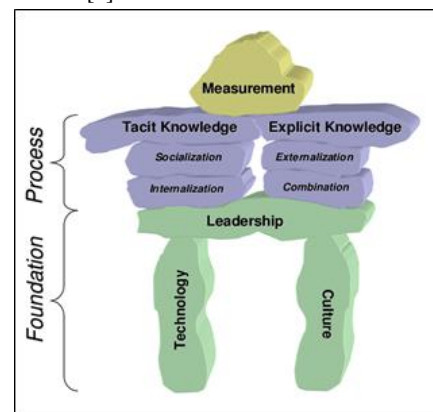
a. Leadership

In this process, an Explicit Knowledge is needed so that decisions and leadership styles of Top Managers can be shared with other employees through Knowledge Management. he leader must also be able to motivate employees in implementing Knowledge Management. One way that can be done is to provide concrete evidence of involvement in the use of Knowledge Management [6].

b. Culture

To create a culture of sharing between employees, company management can provide intensive or reward systems. reward system here can be in the form of appreciation to motivate employees to share knowledge with other employees, or can use a punishment system for employees who do not want to share [6].

c. Technology

At this stage is an explanation of what computing technology is used. Can be discussed through software and hardware used.

2.3 Kano

The Kano diagram was discovered by Professor Noriyaki Kano from the Tokyo University of Science. Professor Kano made an extraordinary theory about the factors that influence customer satisfaction with what is given by a product or service. This theory is known as the Kano Model or Kano Diagram. Kano Model or Kano Diagram is a diagram that divides the specifications of the products given to customers into three types, they are Must be, Performance, and Delighter. below :

a. Must Be

These are things that must exist in a product. He must exist. But its existence will not increase customer satisfaction. Customers will act normal. If it's not there, then get ready you will be complained by the customer. In Kano this model is described as a turtle that is located at the bottom. It is an attribute, function or basic feature that is mandatory for a product / service.

b. Performance

These are factors that are also known as more is better, more are better. If we fail to fulfill it, the customer will be disappointed. If we can fulfill it, then the customer will be mediocre. However, if we exaggerate it, the customer will be happy. The better the performance of this attribute, the higher the level of customer satisfaction. Conversely, poor performance of this attribute can also decrease. Customer satisfaction. An example when reading a brochure is a laptop that has the i5 processor, but it turns out that when the item is received i3. But if it turns out that the laptop purchased has an i7 processor, then the customer will be very satisfied.

c. Delighter

Delighter is more to satisfy customers. If we don't give, customers will not be disappointed, but if we provide this specification, customers will be very satisfied. Example: giving a high end portable sound system when buying a laptop will make customers very satisfied. It should be noted that this will shift over time and the level of competition. Today the Delighter, tomorrow may be Performance and could be the next year Must Be. The example of a webcam on a laptop follows the diagram of a kano or a Kano Model.
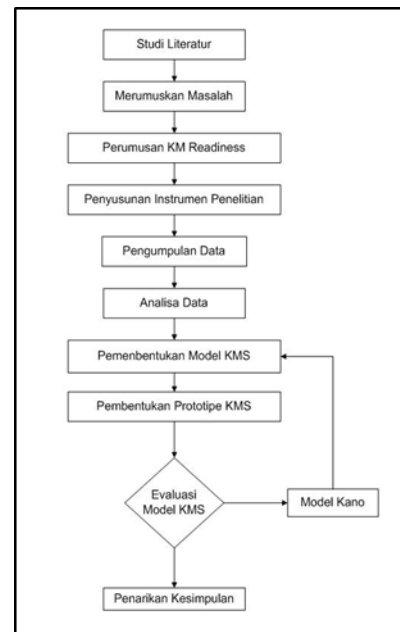
3. RESEARCH METHODOLOGY

3.1 Research Design

Figure. 3.1 Research Design

3.2 Collecting Data Method

The method of data collection in this study was divided into two types, namely primary and secondary data collection methods. Both methods support each other. Primary data is collected through interview processes and questionnaires while Secondary Data is collected from organizational documentation in Higher Education.

3.3 Data Analysis Method

3.3.1 Index Analysis Method

The analysis method used in the research questionnaire analysis process is the index analysis method. Questionnaires that have been distributed and filled out by responder are followed up with measurements of their respective values, the measurement results will be analyzed by the index analysis method so that it will produce conclusions from the research.

3.3.2 Kano Calculation Method

The Kano Diagram is a diagram that divides the specifications of the products given to customers into three types, namely Must Be, Performance, and Delighter. In this study the questionnaire will play an important role in the research process. Kano calculations are carried out in the prototype evaluation process. So that method will produce conclusions from the research.

4. RESULTS AND DISCUSSION

The process of preparing research instruments in the form of a questionnaire design. The design of this questionnaire was made based on the results of the literature studies that have been conducted. The questionnaire in this study was divided into three questionnaire needs.

a. KM Readiness

Knowing the readiness of Higher Education in implementing knowledge management system is very important, so that measurement of knowledge management Readiness is very necessary because a good system must have a good place too.

Knowledge management rreadiness can be interpreted as the ability to adopt, use and benefit from that knowledge management.

b. KM Process Identification Questionnaire

This stage is an identification process that is taken from the second questionnaire data, namely the questionnaire identification of the KM process. Questionnaire questions have represented the KM and INUKSHUK processes, namely socialization, externalization, combination, internalization, leadership, culture and technology.

c. Kano Questionnaire

To achieve the objectives of this research, it is important to spread the questionnaire sequentially because each questionnaire has its own function. The KM readiness questionnaire is used to see the extent to which Universities are ready to implement Knowledge Management so that the implementation of Knowledge Management can run according to what has been designed.

KM Process Identification Questionnaire which is a questionnaire containing questions about Socialization, Externalization, Combination, Internalization, Leadership, Culture, and Technology, each question representing the criteria so that calculations can be made. Calculation using index data analysis.

The Kano Questionnaire aims to measure the requirements and specifications of the system to be made. Making this questionnaire to combine SECI, INUKSHUK, and KANO methods simultaneously.

4.1 Data Analysis
4.1.1 Identify the Current KM Process

This stage is an identification process that is taken from the second questionnaire data, namely the questionnaire identification of the knowledge management process. Questionnaire questions have represented the KM and INUKSHUK processes, namely socialization, externalization, combination, internalization, leadership, culture and technology respectively.

Table 4.1. Current KM Process

KM Process	Index Value	Level
Socialization	130,2	High
Externalization	127,6	Medium
Combination	123,8	Low
Internalization	122	Low
Leadership	127,8	Medium
Culture	127,6	Medium
Technology	132,4	High

4.1.2 Identify of the Current KM Process Facilities

After identifying the current KM process the next step is to measure the current KM process facility. Retrieving the index value comes from table 4.1 above because the calculation includes existing facilities.

Table 4.2. Current KM Facilities

Proses KM	Nilai Indeks	Tingkat
Socialization	125,4	Medium
Externalization	132,4	High
Combination	125,6	Medium
Internalization	119,2	Low
Leadership	127	Medium
Culture	126,4	Medium
Technology	118,4	Low
Jumlah	874,4	

In Externalization facilities that support are very high so implementation will be very easy to do with index value 132.4. Internalization and technology have a low index value.

4.1.3 Identify Of KM Process and Supporting Facilities

Based on the results of the identification of the level of processes and supporting facilities for KM processes that occur at the current Private Universities, this stage categorizes the priority needs of the current KM process development.

Table 4.3. Mapping of Categories Needs Now

KM Process	Current KM Process Level	Curren Level of Supporting Facilities KM	Current KM Development Need Priorities
Socialization	High	Medium	2
Externalization	Medium	High	4
Combination	Low	Medium	8
Internalization	Low	Rendah	9
Leadership	Medium	Medium	5
Culture	Medium	Medium	5
Technology	High	Rendah	3

Mapping the current category of knowledge management development needs if in a graph can be seen in graph 4.9. The chart below is based on needs so that the shorter the graph the greater the priority. Internalization and Combination show top priority with priority needs 1.
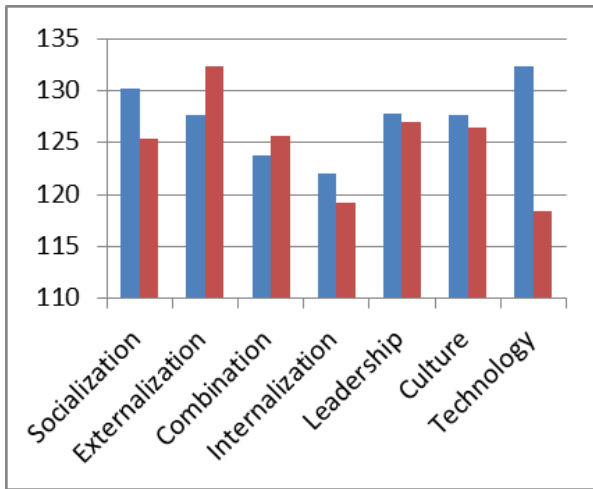

Chart. 4.1 KM Current Needs

Socialization has the highest priority, namely with a value of 2 followed by technology that has a value of 2. The calculation of the second value can be seen in the graph above socialization has 2 bars that are balanced so that it has a high value. Based on the graph above, the creation of features will prioritize features that help socialization and technology.

4.1.4 Calculation of The Kano Model

In table 4.3 the knowledge management process is mapped into technology in the form of supporting features in the Knowledge Management system process. These features are then grouped so that the same features are united, namely Chat, Discussion Forum, Chat box, Q n A, Document Management, Article Management, Document Search, Article Search, Knowledge Monitoring, Value Giving, FAQ, Suggestions and Input, Rewarding, Management Rating Knowledge, Responsive. These features will be calculated by Kano so that the group features Must Be, Perfomance or Delighter can be known

Table 4.4. Mapping Of Categories Needs Now

KM Process	Features
Socilization	- *Chattting* (M) - Discuss Forum (M) - *ChatBox* (Q) - *Q n A* (M)
Technology	- Article Search (O) - Documen Search (M) - *Responsive* (I)
Externalization	- Chatting (M) - Q n A (M) - Discuss Forum (M) - Document Management (A) - Article Management (O)

Leadership	- Monitoring Knowledge (A) - Giving Score(A) - FAQ (R) - Sugestions and Input (O)
Culture	- Giving *Reward* (M) - Knowledge Rank Management (M)
Combination	- Discuss Forum (M) - Q n A (M) - Document Management (A) - Article Management (O)
Internalization	- Discuss Forum (M) - Q n A (M) - Article Search (O) - Documen Search (M)

Socialization, Technology, and Externalization are the highest priorities after calculating questionnaires with the INUKSHUK model. The three knowledge management processes have features that support the process of knowledge management, but not all features will be taken and implemented in making knowledge management system prototyping. Researchers only take features that have a Must Be (M) value or features that must exist in a knowledge management system. If we map between the results of the INUKSHUK model and the Kano model with features as objects, the features that will be made into the system are as follows:

Table 4.5. Features of Questionary Calculations

KM Process	Features
Socilization Externalizationi	1. *Chattting* (M) 2. Discuss Forum (M) 3. *Q n A* (M)
Technology	4. Document Search (M)

There are four features that will be made into the Knowledge Management system, namely Chat Features, Discussion Forums, Q n A, and Document Search. The four features are the results of the questionnaire inukshuk kano model and are the current knowledge management process needs.

4.2 Making KMS Prototypes

The stages of forming this knowledge management system model utilizing the results of data analysis that have been carried out are then formed into the knowledge management system model by looking at the priorities of the knowledge management process that has been recommended with the approach in the previous stage. The formation of knowledge management system prototyping uses an international modeling language namely Unified Modeling Language

(UML) which consists of Use Case Diagrams and Activity Diagrams. The design results are implemented into a prototyping so as to form a system that suits your needs.

4.2.1 *Use Case Diagram*

Use Diagram in Figure 4.1 has 2 actors namely admin and user, admin is tasked to manage existing data while the user is to interact, exchange and access information.
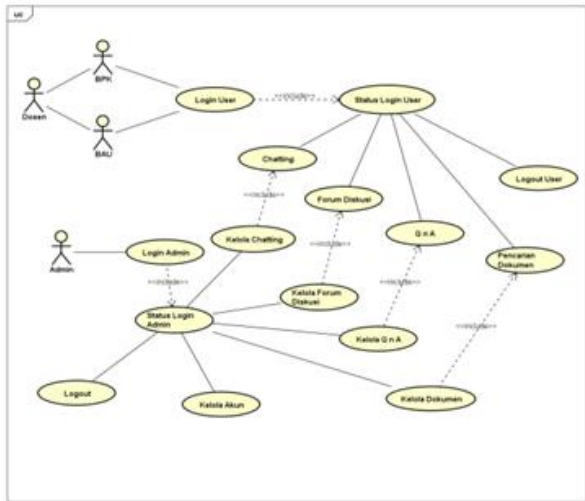

Figure. 4.1 Use Case Diagram

4.2.2 *Home Page*

When the user or admin has passed the log in page, the next page will display the home page of the system. On this page display menus which are features that have been calculated. These features are chat, discussion forums, Q n A and Documents.
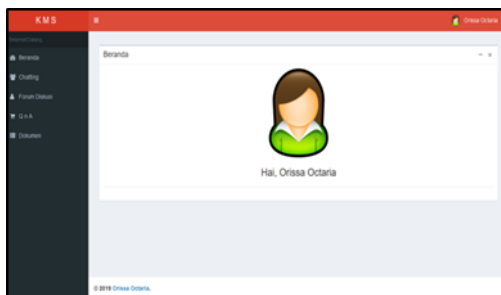

Figure. 4.2 Home Page

5. CONCLUSION AND SUGESTION

5.1 Conclusion

The conclusions generated in this study are as follows:

a. The KMS design developed in this study used the Inukshuk model and the Kano model.

b. The KM process was used in the process of making the Inukshuk questionnaire namely Socialization, Externalization, Combination, Internalization, Leadership, Culture, and Technology. As for the Kano model, it uses the Must Be, Performance, and Delighter product specifications.

c. KMS designed to produce features that are in accordance with the needs of users and existing facilities so that the objectives of KMS will be easily achieved and useful for structural position changes at STMIK MDP, STISTIPOL Candradimuka, and UIGM.

d. Knowledge Management features after Inukshuk and Kano calculations include Chat, Discussion Forum, Question and Answer, and Document Search.

5.2 Sugestion

For the development of further research there are several suggestions related to KM system, including:

a. The scope of the research was extended not only to lecturers, BAU, and BPK, but also included all students.

b. The application of KMS that has been studied only uses web-based technology. Further development is carried out using Android-based technology.

REFERENCES

[1] Setiarso, B., Harjanto, N,. Triono., dan Subagyo, H (2009): *Penerapan Knowledge Management Pada Organisasi*. Edisi pertama. Graha Ilmu. Yogyakarta.

[2] Widiawan, Kriswanto., Irianty. (2004): *Jurnal Pemetaan Preferensi Konsumen Supermarket Dengan Metode Kano Berdasarkan Dimensi Servqual*. Universitas Kristen Petra

[3] Becerra-Fernandes, I., dan Ravij, Sabherwal. (2010): *Knowledge Management System and processes*. M.E. Sharpe, Inc. New York: M.E. Sharpe, Inc.

[4] Nonaka, I., & Takeuchi, H. (1995): *The Knowledge - Creating Company*. New York: Oxford University Press.

[5] Dalkir, Kimiz. (2011): *Knowledge Management in Thery and Practice*. Massachusetts Institute of Technology.

[6] Susilo, S.O., Erwin, A., Liempah, C., dan Soerapto, N.K. *Jurnal Perancangan Knowledge Management System pada PT. Total Prime Engineering*.

[7] Elias, M. A., & Ghaziri, H. (2003): *Knowledge Management*. United States: Prentice Hall.

[8] Girard, Jhon, P, *The Inukshuk: A Canadian Knowledge Managemen Model,* 2014.

[9] Nawawi, Ismail. (2012): *Manajemen Pengetahuan (Knowledge Management).* Ghalia Indonesia. Bogor.

[10] Saputra, Ariansyah. "*Perancangan Knowledge Management System dengan Gabungan Pendekatan Inukshuk dan Kano Pada Rumah Sakit Ernaldi Bahar Palembang*", Universitas Indonesia, Depok, 2018.

[11] Tiwana, A. (1999): *Knowledge Management Toolkit.* United States: Prentice Hall.

[12] Walden, David. *Center Quality of Management Journal,* vol. 2. 873-8950. 1993.

Frequent Itemset in Sequential Pattern Matching Using Bigdata

M. Ilakkiya
M.Tech Computer Science
PRIST University
Thanjavur, India

D. Vinotha
Head and Assistant Professor, Department of CSE
PRIST University
Thanjavur, India

Abstract: A novel frequent item set mining algorithm, namely Horizontal parallel-Apriori (HP-Apriori), is proposed that divides database both horizontally and vertically with partitioning mining process into four sub-processes so that all four tasks are performed in parallel way. In addition, the HPApriori tries to speed up the mining process by an index file that is generated in the first step of algorithm. The proposed algorithm has been compared with Count Distribution (CD) in terms of execution time and speedup criteria on the four real datasets. Experimental results demonstrated that the HPApriori outperforms over CD in terms of minimizing execution time and maximizing speedup in high scalability. We deal with the problem of detecting frequent items in a stream under the constraint that items are weighted, and recent items must be weighted more than older ones. This kind of problem naturally arises in a wide class of applications in which recent data is considered more useful and valuable with regard to older, stale data. The weight assigned to an item is, therefore, a function of its arrival timestamp.

1. INTRODUCTION

With the fast development of networking, data storage, and the data collection capacity, the size of databases is rapidly growing in all domains. According to a report from International Data Corporation (IDC), in 2011, the overall created and replicated data volume in the world was 1.8ZB ($\approx$ 1021B), which increased by nearly nine times within five years. Big data typically includes masses of unstructured data that requires more real-time analysis. Big data analytic by machine learning and data mining techniques has become an important research problem. Mining with big data is very difficult problem when the current data mining methodologies tools with a single personal computer are used to deal with very large datasets due to their large size and complexity. Big Data has special characteristics that make it an extreme challenge for discovering useful knowledge. These characteristics including, large-volume, heterogeneous, autonomous sources with distributed and decentralized control, and seeks to explore complex and evolving relationships among data. The Association Rule Mining (ARM) finds regularities in the transaction data that can lead to develop and implement better strategic business decisions. The ARM contains two phases, (1) mining the Frequent Item sets (FIs), and (2) Association Rule (AR) extraction. The Apriori algorithm is considered as one of the most influential data mining algorithms to extract knowledge in the forms of ARs or FIs. Others well-known AR mining algorithms include FPGrowth, Eclat, and D-CLUB. When the size of datasets becomes very large, the traditional ARM algorithms cannot deliver the results in a reasonable amount of time. Especially the Apriori algorithm that uses a breadth-first search to count the support count of item sets; therefore, it is far less efficient than the other algorithms. In the ARM from big data, the complexity of FI mining is more than the AR extraction. This

is due to the fact the one database scanning is needed to compute the support count of each item set; on the other hand, the calculating the confidence of the strong ARs is performed simply. Therefore, frequent item set mining step requires very more execution time than AR mining step. The parallel paradigm can be considered to tackle this big data problem in an efficient way.

2. METHODS AND MATERIAL

Literature Survey

Data Mining With Big Data

Big Data concern large-volume, complex, growing data sets with multiple, autonomous sources. With the fast development of networking, data storage, and the data collection capacity, Big Data are now rapidly expanding in all science and engineering domains, including physical, biological and biomedical sciences. This paper presents a HACE theorem that characterizes the features of the Big Data revolution, and proposes a Big Data processing model, from the data mining perspective. This data-driven model involves demand-driven aggregation of information sources, mining and analysis, user interest modeling, and security and privacy considerations. We analyze the challenging issues in the data-driven model and also in the Big Data revolution.

Mining Ars in Big Data with Ngep

Big data is a data with large size means it has large volume, velocity and variety. Now a day's big data is expanding in a various science and engineering fields. And so there are many challenges to manage and analyse big data using various tools. This paper introduces the big data and its Characteristic concepts and Next section elaborates about the Challenges in Big data. In Particular, wed discuss about the technologies used in big data Analysis and Which Tools are mainly used to analyze the data. As big data is growing day by day there are lot of application areas where we need to use any of the

technology and tools discussed in paper. Mainly this paper focuses on the Challenges, Technologies, Tools and Applications used for big data Analysis.

Big Data Mining With Parallel Computing: A Comparison of Distributed and Map Reduce Methodologies

Applications of machine learning are widely used in the real world with either supervised or unsupervised learning process. Recently emerged domain in the information technologies is Big Data which refers to data with characteristics such as volume, velocity and variety. The existing machine learning approaches cannot cope with Big Data. The processing of big data has to be done in an environment where distributed programming is supported. In such environment like Hadoop, a distributed file system like Hadoop Distributed File System (HDFS) is required to support scalable and efficient access to data. Distributed environments are often associated with cloud computing and data centres. Naturally such environments are equipped with GPUs (Graphical Processing Units) that support parallel processing. Thus the environment is suitable for processing huge amount of data in short span of time. In this paper we propose a framework that can have generic operations that support processing of big data. Our framework provides building blocks to support clustering of unstructured data which is in the form of documents. We proposed an algorithm that works in scheduling jobs of multiple users. We built a prototype application to demonstrate the proof of concept. The empirical results revealed that the proposed framework shows 95% accuracy when the results are compared with the ground truth.

Parallel Mining of ARS

One of the important and well-researched problems in data mining is mining association rules from transactional databases, where each transaction consists of a set of items. The main operation in this discovery process is computing the occurrence frequency of the interesting set of items. i.e., Association Rule mining algorithms search for the set of all subsets of items that frequently occur in many database transactions. In practice, we are usually faced with large data warehouses, which contain a large number of transactions and an exponentially large space of candidate itemsets, which have to be verified. A potential solution to the computation complexity is to parallelize the mining algorithm. In this paper, four parallel versions of a novel sequential mining algorithm for discovery of frequent item sets are proposed. The parallelized solutions are compared analytically and experimentally, by considering some important factors, such as time complexity, communication rate, and load balancing

Large-Scale Parallel Data Mining

The explosive growth in data collection in business and scientific fields has literally forced upon us the need to analyze and mine useful knowledge from it. Data mining refers to the entire process of extracting useful and novel patterns/models from large datasets. Due to the huge size of data and amount of computation involved in data mining, high-performance computing is an essential component for any successful large-scale data mining application. This chapter presents a survey on large-scale parallel and distributed data mining algorithms and systems, serving as an

introduction to the rest of this volume. It also discusses the issues and challenges that must be overcome for designing and implementing successful tools for large-scale data mining.

Existing Process

As a consequence, whilst in traditional frequent item mining applications we need to estimate frequency counts, we are instead required to estimate decayed counts. These applications are said to work in the time fading model. Two sketch-based algorithms for processing time-decayed streams have been recently published independently near the end of 2016. The Filtered Space Saving with Quasi-Heap (FSSQ) algorithm, besides a sketch, also uses an additional data structure called quasi-heap to maintain frequent items.

Proposed Methodology

Forward Decay Count-Min Space Saving (FDCMSS), our algorithm, cleverly combines key ideas borrowed from forward decay, the Count-Min sketch and the Space Saving algorithm. Therefore, it makes sense to compare and contrast the two algorithms in order to fully understand their strengths and weaknesses. We show, through extensive experimental results, that FSSQ is better for detecting frequent items than for frequency estimation. The use of the quasi-heap data structure slows down the algorithm owing to the huge number of maintenance operations. Therefore, FSSQ may not be able to cope with high-speed data streams. FDCMSS is better suitable for frequency estimation; moreover, it is extremely fast and can be used in the context of high-speed data streams and for the detection of frequent items as well, since its recall is always greater than 99%, even when using an extremely tiny amount of space. Therefore, FDCMSS proves to be an overall good choice when considering jointly the recall, precision, average relative error and the speed.

3. RESULTS AND DISCUSSION

Users

When users access the system through Portal Direct Entry, they are considered guests until they log in. The Login Module is a portal module that allows users to type a user name and password to log in. This module can be placed on any module tab to allow users to log in to the system.

Search Engine

Search engine optimization ("SEO") ensures that the website can be properly indexed by popular search engines, such as Google, Bing, and Yahoo. Search engine optimization can provide higher levels of website traffic and reduced digital advertising costs with programs such as Google Ad words. For these reasons, search engine optimization can be a key success factor for many projects.

Frequent Items Measure

In Find item sets by you can set criteria for item set search: Minimal support: a minimal ratio of data instances that must support (contain) the item set for it to be generated. For large data sets it is normal to set a lower minimal support (e.g. between 2%-0.01%). Max. Number of item sets: limits the upward quantity of generated item sets. Item sets are generated in no particular order.

Filter item sets:

If you're looking for a specific item or itemsets, filter the results by regular expressions. Separate regular expressions by comma to filter by more than one word.

Forward Decay

It is a different model of decay satisfying the forward decay is computed on the amount of time between the arrival of an item and a fixed point L, known as the landmark.

Server

A server is a computer program or a device that provides functionality for other programs or devices, called "clients". This architecture is called the client–server model, and a single overall computation is distributed across multiple processes or devices. Servers can provide various functionalities

Ranking

The ranks of the modules in an exact sequence. Satisfy the equality. The rank of a free module over an arbitrary ring (cf. free module) is defined as the number of its free generators.

Conclusion

To overcome the drawbacks of traditional algorithms, we proposed an algorithm, namely HP-Apriori, which mines knowledge in the form of FIs using performing four sub-tasks in the parallel way. The HP-Apriori performs similarly as the CD with the only difference that the HP-Apriori partitions the data horizontally and uses two processes in each partition for calculating the support counts of local itemsets instead of one process, so that the execution time is reduced significantly. We examined the feasibility and effectiveness of HP-Apriori to solving the ARs mining problems in the big data. The experimental results showed that computational complexity for HP-Apriori is reduced than CD.

Future Enhancement

Priority Based Selection: In Cognitive Radio network the users are classified into Licensed Primary Users and Unlicensed Secondary Users and there is no dedicated channel to send data, sensors need to negotiate with the neighbors and select a channel for data communication in CR-WSNs. This is a very challenging issue, because there is no cooperation between the PUs and SUs. PUs may arrive on the channel any time. If the PU claims the channel, the SUs have to leave the channel immediately. Therefore, data channels should be selected intelligently considering the PU's behavior on the channel and using some Priority Based Selection algorithms. Therefore USFR has been shown to effectively improve self-coexistence jointly in spectrum utilization, power consumption, and intra-cell fairness.

4. REFERENCES

[1] X. Wu, X. Zhu, G. Q. Wu, and W. Ding, "Data mining with big data", IEEE Transactions on Knowledge and Data Engineering, vol. 26, no.1, pp. 97-107, 2014.

[2] M. Chen, S. Mao, and Y. Liu, "Big data: a survey", Mobile Network Application, vol. 19, pp.171-209, 2014.

[3] Y. Chen, F. Li, and J. Fan, "Mining ARs in big data with NGEP", Cluster Computing, vol. 18, no. 2, pp. 577-585, 2015.

[4] C. F. Tsai, W. C. Lin, and S. W. Ke, "Big data mining with parallel computing: a comparison of distributed and MapReduce methodologies", Journal of Systems and Software, vol. 122, pp. 83-92, 2016.

[5] W. Fan, and A. Bifet, "Mining big data: current status, and forecast to the future", ACM SIGKDD Exploration, vol. 14, no. 2, pp. 1-5, 2012.

[6] R. Agrawal, T. Imielinski, and A. Swami, "Mining ARs between sets of items in large databases", The International Conference on Management of Data, p. 207-216, 1993.

[7] Y. Le Bras, P. Lenca, and S. Lallich, "Opt monotone measures for optimal rule discovery", Computational Intelligence, vol. 28, no. 4, pp. 475-504, 2012.

[8] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation", ACM SIGMOD International Conference on Management of Data, p. 1-12, 2000.

[9] M. J. Zaki, S. Parthasarathy, M. Ogihara, and W. Li, "New algorithms for fast discovery of ARs", Technical Report TR651, University of Rochester, 1997.

[10] J. Li, A. Choudhary, N. Jiang, and W. K. Liao, "Mining frequent patterns by differential refinement of clustered bitmaps", International Conference on Data Mining, 2006.

Search Engine Development to Enhance User Communication

Ms.K. Ramya
M.Tech Computer Science
PRIST University
Thanjavur, India

Mrs.D. Vinotha
Head and Assistant Professor, Department of CSE
PRIST University
Thanjavur, India

Abstract: Search Engine Optimization (SEO) is important for websites to improve the rank for search results and get more page views requested by the user. A search engine ranks provide the better and optimized result to user, which will help them to view the popular page among the number of pages available in the search results. Apart from this search engine ranking, it also enables the websites to compete with other rival's website as each website owner expects to see their own website on the list before other's websites. This paper puts forward idea about SEO principles and basic strategies. It also expresses different techniques that are employed by search engines to improve its results. In addition, it presents the observation section, which gives the comparative analysis of SEO techniques. Update concepts sent to concern SEO search member.

Keywords: Search Engine Optimization (SEO), Search engine ranking.

1. INTRODUCTION

Today Internet has become a vast source of information with many websites being added every day. To manage the access and data integration of this vast resource has been a challenge these days. Since search engines handle 93% of internet traffic, so exploring the potentials of search engines is crucial. Search Engine Optimization is a tool that is used by website owners to defeat its competitors by placing its link before others website links in search engine's results. Since internet is full of raw data, the job of directing towards relevant and exact data lies in the hands of search engine. Due to this reason, Search Engine Optimization Techniques has become an important topic for researchers and academicians and business organizations. SEO techniques make use of original results, which are analyzed by search engines to increase the popularity of website. These techniques help government and business organizations to promote their websites and increase the movement of traffic in and out of their websites. Through the discovered search results, SEO improves the ranking of a website or a web page in search engine websites list. Also, to increase the ranking of website, SEO considers various parameters such as, the response time of search engine for a particular website, frequently being visited websites by users, type of search terms entered into URL and type of search engines preferred by users. Through SEO one can ensure that a particular site is accessible through search engine, which improves the chances of quickly discover a site by the search engine. Search Engine Optimization is typically a set of best practices that adopted by webmasters, web developers and web content producers to achieve a better ranking in search engine scenario.

2. METHODS AND MATERIAL

Literature Survey

Page Rank Algorithm Page Rank

Algorithm was proposed by founders of Google Larry Page and Sergey Brin in 1996. Page rank algorithm calculates the rank of the web pages based on its importance. Page importance is depends on the number of occurrences of the web page. In most of the cases, the page rank algorithm follows the link structure of web page to calculate the rank.

Link structure of web page is mainly depends on the number of incoming and outgoing links for that web page. Consider the pages A and B. A is having incoming link for B and B is having outgoing link for A. So to calculate the A's rank we require rank of B and to calculate B's rank we require rank of A. If the page is having more number of incoming and outgoing links, then it is having highest vote or rank. Page rank algorithm is easy to understand as it considers the parameters like the rank of the web page, which depends on the occurrence of the web page as well as the number of incoming links to the webpage. But the calculation doesn't give exact answer, if they performed only once. Accurate ranks are obtained through multiple iterations. A concept called 'Page Dependency', which is used in Page Rank algorithm to calculate the page rank. This page rank depends on the rank of other pages that are linked to it. For calculating the rank of page, the algorithm considers the probability of web page's rank, which is already stored in the database instead of taking the current rank status, which is required by the query. It always divides the page rank of incoming links equally among the web page, which is not important or relevant

Weighted Page Rank Algorithm

This algorithm was introduced by Wenpu Xing and Ali Ghorbani. This algorithm also follows the link structure of web pages that is the incoming and outgoing links to or from web pages. Weighted Page Rank algorithm (WPR) calculates the rank of the pages based on both incoming and outgoing links. The page has a highest rank, if it is having more number of outgoing links associated with it. The popularity of page depends on both incoming and outgoing links. Weighted Page Rank algorithm (WPR) considers the current status of page rank at the time of user query as against the probability of page rank, which is considered in case of Page Rank algorithm.

HITS Algorithm

Hypertext Induced Topic Search (HITS) algorithm was proposed by Jon Kleinberg to rate a webpage based on link

structure of the web pages [4]. This algorithm analyses the structure of web pages links based on the concept called Authority and Hubs. An authority page contains the useful information based on the user query whereas hubs are the pages that provides links to the authority pages. HITS algorithm mainly has two steps. 1) In sampling step the algorithm collects the number of relevant web pages based on user query. 2) In iterative step the authority and hub pages are calculated in order to serve user query in efficient time. These steps performed in HITS algorithm to calculate the page rank of the webpage. Authority and Hub pages collects the samples of the pages from ranking model and based on the incoming and outgoing links of webpage page rank is calculated.

Query Dependent Algorithm

Query Dependent Algorithm was proposed by Lee, Jiang, et al.. This algorithm mainly focuses on links (incoming and outgoing links) as well as the contents of user query. This algorithm measures the similarity between the user queries. Based on the similarity between queries, this algorithm builds the ranking model also called training set. Depending on the contextual features of user query, this algorithm extracts the relevant documents to the ranking model. Then the "probability function" of the document is calculated to find the rank score of the web page. This algorithm sometimes considers the occurrences of documents in the ranking model and assigns the page rank to the pages as in Time ranking algorithm. Once the user query is fired to find the equivalent content on the web page, the query evaluation engine extracts the query from web page and does the query optimization to meet the desired results.

Study on Website Search Engine Optimization

With the rapid development of information technology, search engine optimization (SEO) technology has attracted more and more attentions. In order to improve their website visit quantity, SEO techniques can make a better ranking in the search result using the keyword selection and deployment, high quality back links, rational website constitution, and rich content, etc. This paper discusses in detail the technical process of website search engine optimization in terms of the search engine work principle, factors affecting search ranking, and website search engine optimization method.

Existing Process

Search Engine Optimization (SEO) is important for websites to improve the rank for search results and get more page views requested by the user. This search engine ranks provide the better and optimized result to user, which will help them to view the popular page among the number of pages available in the search results. Apart from this search engine ranking, it also enables the websites to compete with other rival's website as each website owner expects to see their own website on the list before other's websites.

Proposed Methodology

This paper puts forward idea about SEO principles and basic strategies. It also expresses different techniques that are employed by search engines to improve its results. Also it presents the observation section, which gives the comparative analysis of SEO techniques. SEO techniques make use of original results, which are analyzed by search engines to increase the popularity of website. These techniques help government and business organizations to promote their websites and increase the movement of traffic in and out of their websites. Also, to increase the ranking of website, SEO

considers various parameters such as, the response time of search engine for a particular website, frequently being visited websites by users, type of search terms entered into URL and type of search engines preferred by users.

3. RESULTS AND DISCUSSION

Research

There are many keyword finders available, which will take input in the form of keywords and gives the output in the form of how many times that keyword was entered by users i.e. the number of occurrences of that keyword. However, if a particular keyword has the highest number of occurrences in the search results, that keyword will become the greatest competitor within the search results.

After the keyword analysis phase, competitive analysis of the subject website is performed against its competitors websites. To perform competitive analysis of websites, different parameters are considered such as, incoming and outgoing links, ranking of website on search results, number of visitors for the website, view rate or bounce rate, appearance of web contents on webpage etc.

Reporting & Goal Setting

Most important step in Reporting and Goal Setting is to analyze the sites traffic. Site's traffic mainly depends on type of search engine, number of occurrences of keywords in search engine, popularity of web contents, number of visitors etc. Sometimes the parameters from SEO metrics are also used to calculate the site's traffic. Most important factor considered while calculating site's traffic is the number of incoming and outgoing links. If the site has more number of outgoing links to other website as well as incoming links from another websites, the number of visitors will be more for that website. This will ultimately increase the bounce rate, which will result in high traffic rate.

Content Building

Content is the important factor of search engine optimization. The website with the high-quality content will provide the better competitive environment in search engine optimization. Content building phase is divided into two steps. First, a site is loaded with high-quality content, which gives reason to user to stay on the website and to come back to the website. If user finds most relevant information on particular website, then instead of visiting other sites, he/she always stick to that particular website. The main aim of the user to stay or to come back on site is just to find or search information.

Webpage Optimization

This phase considers all the parameters, which are related to Graphical User Interface (GUI) of the website. Along with the Content building phase, the appearance of the media components on the webpage is equally important. The webpage optimization is carried out in four steps, namely, webpage titles, web content exploration, prominence of targeted keyword phrases and site outline.

Social & Link Building

Now a day's billion number of users are using social media for various purposes. In marketing, sharing of web contents through various social media provides number of options for organizations to establish a best customer relationship and also emphasize on other clients to use those social media for

sharing of site contents and improves connectivity with other potential customers

Maintaining Progress Report of SEO Plan

At the start and end of each SEO phase, progress report is generated to analyze the activities performed in particular SEO phase. The outcome of the phase is recorded for further analysis. Rankings, site's traffic levels and some other parameters are considered to measure the ranking results.

Conclusion

Search Engine Optimization come from more advanced technologies in traditional web search engines such as Google search engine. One of the important aspects of search engine is to improve the performance and usability of search engines through various techniques so that user interaction will be increased to that particular web site. In this paper, we analysed search engine results depend on the various algorithms, which improves web page rank. Web pages are displayed according to their rank, which is calculated by using factor like content, number of incoming and outgoing link etc. This analysis will help to improve the overall performance of search engine in competitive world.

Future Enhancement

Google's all algorithms are focused on improving the user experience. If you love your users, Google will love your website. If you have ever cared to closely look at the Google's algorithms, you already know that they will rank a website that takes care of the user's needs first.

Initially, SEO was considered to be a profile for IT department. As it was believed that SEO involves too much of technicalities. However, if you understand SEO or have worked as an SEO professional, you would know it is an art to its core. The SEO professionals today understand that mere technical knowledge is not enough in this user-friendly virtual world. Creativity is required for an SEO professional to catch the interest of a user.

4. REFERENCES

[1] Search Engine Optimization [online] Avaliable: http://www.denningesolutions.com/website-design/search-engineoptimization/seo-process/

[2] Gokhan Egria, Coskun Bayrakb, "The Role of Search Engine Optimization on Keeping the User on the Site," Complex Adaptive Systems, Publication 4, Missouri University of Science and Technology 2014.

[3] Dilip Kumar Sharma, A.K. Sharma, "A Comparative Analysis of Web Page Ranking Algorithms," (IJCSE) International Journal on Computer Science and Engineering, 2010.

[4] Rashmi Rani, Vinod Jain, "Weighted Page Rank using the Rank Improvement," International Journal of Scientific and Research Publications, July 2013 .

[5] Archana Kurian, Jayasree M, "Analyzing and Classifying User Search Histories for Web Search Engine Optimization," 2014 3rd International Conference on Eco-friendly Computing and Communication Systems, Dept. of Computer Science & Engineering, Govt. Engineering College, Thrissur.

[6] Jyoti Yadav, Anil Kumar, "A Comparative Study of Ranking Techniques for Hidden Web and Surface Web, Seema Rani," International Journal of Innovative Research in Computer and Communication Engineering, May 2015.

[7] Mridula Batra , Sachin Sharma, "Comparative Study Of Page Rank Algorithm with different Ranking Algorithms adopted by Search Engine for Website Ranking", Jan-Feb 2013.

[8] Taruna Kumari1, Ashlesha Gupta, Ashutosh Dixit, "Comparative Study of Page Rank and Weighted Page Rank Algorithm", International Journal of Innovative Research in Computer and Communication Engineering February 2014.

[9] Pooja Devi1, Ashlesha Gupta, Ashutosh Dixit, "Comparative Study of HITS and Page Rank Link based Ranking Algorithms", International Journal of Advanced Research in Computer and Communication Engineering , February 2014

[10] Ashish Kumar Kushwaha , Prof. Nitin Chopde, "A Comparative study of Algorithms in SEO & approach for Optimizing the search engine results using Hybrid of Query Recommendation and Document clustering, Genetic algorithm," International Journal of Computer Science and Information Technologies, 2014.

A Learnability Model for Children Mobile Applications

Stephen Kioko
Jomo Kenyatta University of
Agriculture and Technology
Nairobi, Kenya

Michael Kimwele
Jomo Kenyatta University of
Agriculture and Technology
Nairobi, Kenya

Lawrence Nderu
Jomo Kenyatta University of
Agriculture and Technology
Nairobi, Kenya

Abstract: Nowadays, mobile device use among young children is on the rise. A kid using a smartphone or a tablet is a common sight today. This has been triggered by the worldwide advancement of mobile culture & technology. As a result, the number of mobile applications targeting the young children has tremendously increased. Studies, however, reveal that most of the applications are far from perfect. They are less learnable and hardly usable by the young children, as they do not match their mental models. Usability and learnability issues usually have a profound impact on the success of a mobile application. Nevertheless, there has been limited research on these issues for mobile applications designed for children. In regard to this, this paper aims to bridge the gap by reviewing usability guidelines suitable for designing 'near perfect' children mobile applications. We then propose a learnability model to help mobile developers and evaluators in designing and evaluating mobile application for this target population (5-10 year olds). Our discussion concludes that, though there are many mobile applications designed for young children, for any app to stand out and be successful, it needs to be designed with different technological skills, learning capacities and preferences tailored to young children.

Keywords: Learnability Model; Children Mobile Applications; General Interaction; Graphical User Interface; Multimedia and Text.

1. INTRODUCTION

Mobile devices and applications have become an integral part of many children's lives. The market of their mobile applications is now rich and very diverse. The Google play store and the Apple app store, for instance, contains lots of applications targeted for this group of users. There are educational mobile applications, apps for entertainment, communication, and even others developed for children with special needs. The children usually use these devices and applications to play games, learn, communicate, or engage in a plethora of other activities offered.

As a consequence of the rapid advancement of the mobile technology, parents, elementary schools and other learning facilities have started accepting and adopting the devices as educational tools [1],[2], teaching aids [3], and for entertainment [4]. It is believed that educational possibilities for using mobile applications to learn, exist [5], and there is a growing view that they offer a new and exciting era in children's learning phase [6]. Despite this, it should be noted that developing mobile apps suitable for children is different from designing for adults [7],[8]. Children apps require distinct usability approaches, that even include targeting content narrowly for children of different ages [9]. According to [10], "the needs, skills, and expectations of this population segment are drastically different from those of adults".

The interface of the mobile applications should thus be usable and compatible with the cognitive skills of children in order to provide an effective learning experience [11]. Usually, according to [8], "the applications should be funny, intuitive and interesting, so as to keep children's attention for a long period of time".

Though there is a large number of new mobile applications designed for children, [12] cites that many of them are difficult to use, some studies, like by [10] & [11] point out that most of the learnability and usability guidelines used in designing mobile apps have been developed with adults in mind. Many developers, as a result, tend to scale down these guidelines to fit children when designing their applications. This is a critical design issue because children are not just mini-adults. According to [4] "they have their own needs and goals which are not necessarily met by adult tools". When designing their applications, it is important to aim for the best possible usability for them [13]. The usability and learnability principles are fundamental when designing any software application as they play a crucial role in the success of mobile applications [14]. Not only do they lead to better productivity quickly, but are also a key to the initial adoption or rejection of an application.

Away from the usability design guidelines, We also noted that, most of the existing usability and learnability evaluation models are also designed for general mobile applications [15]. They do not capture the limited conceptions of time and understanding that affect children as they interact with mobile applications. [14] Notes that "They lack one or the other criteria important for this area". This makes it challenging to test and evaluate the effectiveness of mobile applications for children.

To address the raised issues, the rest of this paper is organized as follows: Section 2 presents a discussion of related work in literature. Section 3 presents the proposed model that includes Guidelines, Goals, Questions and Metrics. Section 4 is a Case study followed by Results and Discussion in Section 5. Section 6 covers conclusion & future work.

2. RELATED STUDY

The literature on how to measure and evaluate usability and learnability is limited in the area of mobile applications [16], [13]. It is even more scarce when the applications are made for young children [17]. There are fewer guidelines on how various definitions of the usability factor, rules, and criteria are related, and even on how to measure the usability of mobile applications [18]. According to [15], the usability and learnability evaluation of mobile applications has not yet touched the accuracy level of other web-based applications. Nevertheless, a number of mobile evaluation models and methods have been introduced in literature to measure and evaluate the mobile apps. These models are, however, more general meaning that they may not be directly applied to some specific mobile applications.

[19] Introduced a usability measurement model based on a review of empirical mobile usability studies. The usability dimensions proposed in the model are quite comprehensive, though they lack appropriate descriptions on which usability

dimension is to be chosen for a specific mobile application. The model also lacks supporting metrics for each dimension and guidelines for choosing suitable dimensions for a particular mobile application.

While attempting to address the challenges facing the evaluation of mobile application [13] developed a usability model that considers four contextual factors: user, environment, technology and task/activity. This usability evaluation model contains ten dimensions for measuring the usability of mobile applications. However, though the model is well equipped with relevant dimensions, it lacks adequate descriptions on how it can be used to evaluate a specific mobile application, like an app for children, since it was designed for general mobile applications. The model did not also provide supported metrics for each dimension to support convenient and effective usability evaluation approach particularly for a defined mobile application.

Another common model in the literature is the PACMAD usability model (People At the Center of Mobile Application Development) which has two versions PACMAD 2013 by [20] and PACMAD 2015 by [21]- an extension of the former.

The aim of PACMAD 2013 was to extend the existing usability models, such as Nielsen's or the ISO to the context of mobile applications. The model has seven components- Effectiveness, efficiency, satisfaction, learnability, memorability, errors and cognitive load. This model, however, just like the others lacks guidelines and metrics related to chosen dimensions. The 2013 version was also not evaluated to examine its accuracy for mobile applications.

The extended version, PACMAD 2015, filled the gaps above by extending PACMAD 2013 to include relatively low-level metrics in addition to usability attributes. This extended version comprises 21 metrics. It used GQM guide to develop usability metrics matching those metrics yielded from literature. Two evaluation instrument task list and user satisfaction questionnaire are used to collect objective and subjective data for complete usability evaluation of the extended PACMAD. The model is however still general, not touching on anything regarding the evaluation of mobile applications for children.

A more recent model that is relevant to children applications was developed by [17],"A measurement model based on usability metrics for mobile learning user interface for children". The model consists of guidelines, usability characteristics, goals (interface design criteria), questions, usability metrics and two evaluation instruments (A task list and a satisfaction questionnaire). It was validated by applying the proposed metrics and evaluation instruments in a usability study conducted on two android educational apps for children. Though the model is more relevant to children compared to the others, it focuses exclusively on evaluating the interface of mobile educational applications for children. It is not broad enough to accommodate guidelines and metrics suitable for other types of children mobile applications like gaming and entertainment apps. Moreover, the goals presented in the model only focus on interface design.

From these related works, the general observation is that, most of the existing design guidelines and evaluation models cannot be relied upon when designing and evaluating applications for young children as these apps are different both in terms of features and functions when compared to any other general app.

3. THE PROPOSED LEARNABILITY MODEL

Learnability is often a complex quality concept. A model is, therefore, necessary to specify its quality requirements, identify its components and understand them better. To extend and address some of the shortcomings of the existing mobile evaluation models, this study presents a comprehensive learnability evaluation model for children mobile applications. It consists of mainly two sections. The first section presents a review of synthesized design guidelines suitable for apps designed for the target users. A thorough analysis has been done to select only the relevant guidelines, resolve the conflicts and rephrase the unclear ones.

In the second section, the Goal Question Metric (GQM) approach is used to develop questions and metrics for learnability evaluation of the applications. The GQM is a goal-driven method for developing and maintaining a meaningful metrics program. It was developed by [22] and is based on three levels; Conceptual, Operational and Quantitative levels. The approach brings success for the reason that it is adaptable to many different environments [18]. It can possibly be extended to measure usability and learnability guidelines by providing metrics for the guidelines.

The conceptual level entails identifying the goal. In our case, the goal represents the overall aim of the learnability evaluation. The design criteria under consideration in this research act as goals for the GQM. They are General Interaction, Graphical User Interface, Navigation & Search, Multimedia & Text, Content & Engagement, Feedback Adequacy, and Consistency. At the second step, operational level, the goals, and guidelines are used to formulate a list of questions which when answered will indicate if the goal has been achieved. The questions developed constitute the basis for quantitative metrics definition. This study ensures that the questions formulated can easily be answered by the targeted user group.

The last step, quantitative level, involves the development of a set of metrics in order to collect data to answer each question in a quantitative way. Metrics usually represent some sort of measurement as to whether or not we have achieved a certain criterion. The metrics developed are both objective and subjective as not all created questions can be answered objectively. As such, some of the questions will be answered subjectively using a questionnaire to assess user satisfaction.
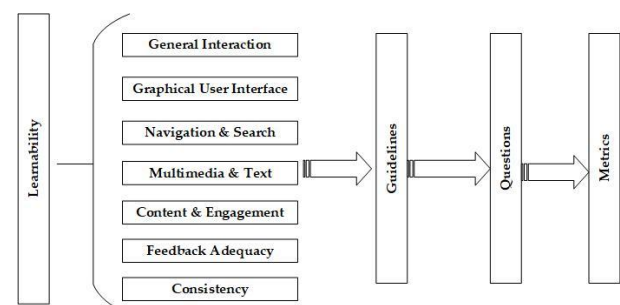

Figure 1. Learnability Model

3.1 General Interaction

When designing and developing mobile applications for young children, it is important for the app interfaces to take into consideration the fact that, these users might not yet have completely understood conventional concepts. The actions of any of the features on the app should at-least map directly to their actions in real world. If some styles and features are not intuitive, some of the children may require extra training with others unable to grasp how the interaction works. The application should thus be developed in a manner that gives children the ability to define their experiences while being in control of the interaction.

3.1.1 Guidelines

1. Make sure every feature and interaction tool is visible and works.

   Many features in an app interface do not necessarily qualify it to be usable. An app can be more usable while having fewer features that all work perfectly, rather than including many of them, with some not working or a level above the intended user group. Avoid screen notifications like "The page is under construction" when designing for children as this group will expect everything to work and work well.

2. Ensure that the interaction tools such as images/icons and buttons reflect if they have been used; using for example different color, size, shape or even pop up messages. This feature may help children to understand what relationship exists between their actions on the screen and the expected outcome.

3. Strive to ensure that the need to type in the mobile app is as minimum as possible. Adults find it uncomfortable to type in a mobile app, meaning it can be more challenging for young children. It is often error prone.

4. Show current interaction state (Users next action should be visible right on the screen). Ensure that the next required action is quite clear to the user. This can be done with effective visual hinting using intuitive icons or graphics.

5. Children learn from examples. Show them a demo: What they learn, they retain. With this quality in mind show them how to perform tasks on the app once and they will most likely not need help again throughout.

6. Create immediate success.

   Children demand fast applications. They normally have little patience for any application that takes a lot of time to load or accomplish tasks.

7. Let the children be in control of the application by allowing them to use it at their own convenient pace. Do not force them to follow a strict process. For instance, let it possible for them to use shortcuts. Allow them also to skip the app on-boarding phase if they are aware of the introductions and instructions. Additionally, ensure that they are aware of the likelihood or possibility to skip them.

8. Make the application interactive; Children love an application that they can interact with by talking to it and it talking back to them. If it is an educational application, the talk-back aspect can enhance the learning process by changing the traditional classroom setting into a friendlier learning environment.

9. Make use rewards efficiently to promote continuity in use and learning.

   Rewards may be simple "Great job" messages, audio cheers, a badge or a sticker, etc. Points can as well be used for doing various things on the app, e.g. showing a leaderboard in terms of where they stand in the overall rankings etc.

10. If the application is purposely educational, provide occasional entertainment diversions that may keep children more engaged and motivated during learning of tasks. For instance, they find humorous multimedia diversions very enjoyable.

11. Be cautious of promotional elements which attract too much attention. Generally, avoid much advertisements in children applications. Any On-screen interventions should be supportive rather than distracting. Enable Ad-blocks so children won't get distracted with inappropriate information.

    Distinguishing between content and adverts may not be easy and straightforward for children. They regard adverts as important and relevant application elements. For instance, they will especially tap on adverts where the banner contains popular characters or where it is seemingly cool.

12. Avoid making registration a must for these applications to be used; but if it cannot be avoided completely, allow the user to skip sign up and let them be able to sign up later.

Table 1. General Interaction Questions and Metrics

Goal: General Interaction	
Questions	**Metrics**
Is the application UI easy and intuitive to use?	Time taken on the first attempt Number of tasks successful in the first attempt Number of mistakes while learning to use the application
How much time does the application take to load tasks?	Time taken to load tasks Rating scale for loading time
Does every feature an interaction element work perfectly as expected?	Number of features not working correctly
Is it simple to key in-in data?	Time taken to key in data Number of errors while keying data Time taken to recover from errors
Does the UI provide useful demonstrations to help children understand an activity?	Number of users who can complete tasks without help Rating scale for usefulness of help demos
Does the application protect	Number of accidental errors

against errors in user actions?	during interaction with the application
How much time and effort is required for completing a given task?	Time taken to complete a task Rating scale for task effort
Does the application obstruct children with promotional content?	Number of distracting adverts during task performance

The metrics in red are Objective, The rest are Subjective

3.2 Graphical User Interface (GUI)

The GUI plays a key role in how children and users, in general, perceive an application. Children, notably, like colorful animated designs with simple text and clear navigations. The metrics on GUI measure to what extent a given user interface can be usable, attractive, and satisfactory for the specified users to effectively achieve particular objectives.

3.2.1 Guidelines

1. Reduce cognitive load by making use of more graphics and less Text.

   Cognitive load in an application may be characterized by too many features, too many options to choose from, too many questions to answer, or even too much information to read and understand. Therefore include only the most essential elements.

2. Design Finger-friendly Tap-targets.

   Due to their small and tender fingers, young children have difficulty targeting small objects on the screen. Thus make the objects simple and big enough so that they're easy for them to tap. Also, ensure that elements aren't located too close to each other so as to prevent false input.

3. Allow customization and personalization.

   Let the application be configurable as per children's interests. For example, colors, background images, size of the font, type of the font etc. should be modifiable based on their needs.

4. Provide aesthetic, intuitive, attractive and organized design.

   This can be achieved by ensuring that; the font style and size are appropriate; use of bright colors for children, colorful and animated buttons, and a clear consistent screen layout.

5. Children prefer animated colorful characters, icons or images as opposed to static ones. The animations help to illustrate the effect of choosing that particular icon, while also bringing about entertainment. These users also like it when interactive elements can "come alive" when tapped.

6. Use colorful graphics and intuitive UI to make the app engaging and easy-to-understand.

   Graphics need to be recognizable, appropriate and relevant to the layout or intended objective. The colors used should be pleasing to children. For example, gaming apps can have bright, lively colors with more graphics, whereas for educational apps, design to allow easy reading, matching and search capabilities.

7. Add good visual affordance (indications of elements use and how they are to be used), on all interactive elements like images and buttons. Show also task related clues/ hints.

   Allow children to be able to correctly predict how an interface element will behave by just looking at it. Lack of perceived affordance, such as overly flat graphics may cause the users to overlook them and thus miss important features, they would otherwise have followed.

8. Use language, gestures, and concepts that are appropriate and familiar to children. The icons and symbols used should be familiar and matching to their meaning in the physical world.

   They must be relevant and content-specific to the information or function they represent. Resist from using images of outdated objects such as floppy disk, typewriter etc. that may not be recognizable to today's five-year-olds.

9. The arrangement of components must be consistent and simple for the whole application.

10. Allow integration with 3D touch technology to make the experience even more interactive

Table 2. GUI Questions and Metrics

Goal: GUI	
Questions	**Metrics**
Do buttons and icons show clickability affordance so that children can easily recognize them?	Number of icons/buttons recognizable on first attempt Number of task related clues/hints
Does the UI provide finger friendly tap targets?	Number of mistakes committed by tapping the wrong button/icon
Can the application UI be personalized or customized per child interests?	Number of options for personalization/customization Failure for using personaliza-tion/customization options
Is it clear what different features of the application stand for?	Number of unclear and ambiguous features
Is the UI design aesthetic and attractive to children?	Rating scale for attractive interface design
Are children satisfied with the graphics and animations used in the UI?	Rating scale for icons/images, interface color and graphics

The metrics in red are Objective, The rest are Subjective

3.3 Navigation and Search

The navigation aspect, is one of the basic elements of a great user experience. To achieve it, according to [23] "every action or piece of information should be easily communicated and executed".

The navigation must be discoverable, accessible and take little screen space. It should also be able to complement the search technique.

Concerning search, children should be able to access and search an element of interest in the application with as little effort as possible. That means few taps, little scrolling, and zero page loads if possible.

3.3.1 Guidelines

1. Provide a clear and consistent main menu. Don't hide menus on individual pages because it might confuse or disorient young users.

2. Make it possible to use the application in either portrait or landscape orientation.

3. Show the users where they are in the application, where they can go, how to get there, and how to go back to their starting point. Use intuitive graphics and icons to achieve this.

4. Avoid designing an interface that requires a lot of scrolling.

   Children interact with what is visible on the screen, therefore, strive to provide a proper interface that shows just everything the child needs so as to interact and learn.

5. Use standard navigation and search schemes such as Tab bar (for iOS) and Navigation drawer (for Android). Many children are familiar with these common conventions and use them easily. Avoid hidden navigation such as gesture-driven because children will have a hard time finding it or figuring it out.

6. The search icon and back icon are very important navigation options. Place them where they are prominent and easily accessible. Don't make kids look for these options.

7. Create excitement about searching and make the search results easy to comprehend.

   Children need interfaces that inspire them to search, otherwise, they will never bother to. Many may be unfamiliar with the tools available, but more importantly; many may not understand the possible content that awaits them if it isn't easy to comprehend.

8. While implementing search use the auto-complete feature to allow fast searching.

Table 3. Navigation and Search Questions and Metrics

Goal: Navigation and Search	
Questions	**Metrics**
Does the UI provide a consistent navigational menu suitable for children?	Rate of success/failure when navigating
Is it easy for children to navigate across the UI?	Number of mistakes resulting due to unclear navigation
Is the screen orientation of the application effective when performing all tasks?	Rating scale for screen orientation
Does the application indicate easy scrolling if a lot of information is present?	Mistakes committed as result of children's inability to scroll Rating scale for easy scrolling
Can children be able to easily locate and use the search/ help icon?	Time taken to locate and use the search icon Number of mistakes while searching
Can children be able to comprehend the search results?	Rating scale for comprehending search results

The metrics in red are Objective, The rest are Subjective

3.4 Multimedia and Text Usage

Multimedia aspects like sounds, animations, graphics, and other art based attributes play a significant role in the success of children's software applications

When used effectively, they enhance their learning experience and speed the interaction process.

3.4.1 Guidelines

1. Use motion and sound on important elements on the screen to attract child attention and engage them.

   Proper tunes and animation make applications even better perceived by children as they find them extremely engaging. The animations and sound effects often excites children and encourages them to stay longer with an application.

2. Make introduction animations short and interesting.

   Animations that are short (10 to 20 sec) and meaningful (funny, surprising, intriguing and enjoying) work best for children.

3. Allow different modes of communication between the user and the application. For example audio and text modes can be used simultaneously. This can end up making the application accessible to both reading and pre-reading children.

4. When interacting with a multimedia clip, show the users the status of multimedia playtimes. Present progress to completion and how long it takes to play the entire multimedia clip.

5. Allow the children to control the multimedia clip. Provide an explicit skip feature such as skip button for all clips/movies.

6. Use only understandable and brief to the point text. Minimize the amount of text on the screen and consistently maintain a high level of readability.

7. Use simple and relatively large fonts.

Avoid decorative wording that makes it difficult to understand different choices. The decorative or fancy wording in interfaces confuse children and prevent them from understanding the available choices.

8. Avoid animated texts.

   Special effects on texts may cause problems to children. For example when the animated texts are too slow, too fast or when the texts and objects are superimposed.

9. Provide a good contrast between the text and the background. Contrast is very important for children who are beginning readers.

10. Avoid providing instructions in blocks of text. Place easy to find links to brief text explanations whenever the user might need or want instructions.

Table 4. Multimedia and Text Usage

Goal: Multimedia and Text	
Questions	**Metrics**
Does the application provide adequate visual and audio assistance?	Number of voice assistance options in a task
Are the multimedia elements (e.g. animations, graphics, sound and video) appropriate for children?	Number of multimedia elements the user can identify on first attempt Rating scale for multimedia elements usage effectiveness
Can children be able to control the pace of multimedia clips?	Rating scale for satisfaction while playing clips
Does the application provide easy text readability for children?	Rating scale for text readability
Is the text usage (font styles, size and color) appropriate for children?	Rating scale for text usage satisfaction

The metrics in red are Objective, The rest are Subjective

3.5 Content and Engagement

For an application to garner the attention of children, the content should be engaging. Effective content and Mobile app engagement are two metrics that provide genuine insight into the success of an application. Ineffective content and Low app engagement are a recipe for failure, while high engagement and retention equal to the opposite. The success of an application can be determined by the user engagement and activity on it.

3.5.1 Guidelines

1. Use appropriate content, language and familiar concepts.

   Young children are excited by content that is funny and entertaining. Additionally, the app interface should be unobtrusive and let children get to the content as simply as possible.

2. Design interaction elements that children can understand and identify with. Children love elements that appear to be animated.

3. To make children more engaged in the application, let them be in full control of the interface. Young children are exploratory and will enjoy influencing various elements while seeking out all the possible ways they can interact with them. The more they interact with the design elements, the more engaged they become, regardless of how minimal the interaction is.

4. Tasks on the app should be easy for children to handle. If they are too difficult, the children may feel bored and avoid using the application. Provide in the app an option to adjust the level of difficulty for children of different ages. Include an option allowing them to skip certain challenges if they get stuck.

5. Avoid unnecessary information. Unnecessary information in the application can decrease children's focus and attention and also confuse them.

6. Motivate children by providing interesting rewards for each correctly solved task. Motivation can be done using stars, points or any other type of image accompanied by a pleasant melody or a congratulatory message. Points or grades collection add the sense of competition to the process of learning

7. On top of motivating them, Praise the children for the value of their actions. This improves their confidence and may help them to perform subsequent tasks better.

8. Reduce the learning curve of using the system by including onboarding if necessary.

Table 5. Content and Engagement

Goal: Content and Engagement	
Questions	Metrics
How much time do children spent on the application by themselves?	Time spent on the app(session length) when child is alone
Does the application provide exciting rewards to keep children engaged?	Frequency of rewards Rating scale for engagement through rewards
Is the app content satisfactory, pleasing and appropriate for children?	Rating scale for appropriate content
Does the app provide useful and interesting learning activities for children?	Task completion rate

Does the application provide different task difficulty levels?	Number of tasks with different difficult levels
	Rating scale for engagement through difficulty levels
Are concepts used familiar and matching to children mental models?	

The metrics in red are Objective, The rest are Subjective

3.6 Feedback Appropriateness

When interacting with an application, users rely on feedback to inform them that they are progressing towards the desired objective. The immediacy of the mobile devices and the short-time period interactions which users have with these devices means that feedback must be delivered faster and more regularly in order to keep users engaged.

3.6.1 Guidelines

1. Task response time must be quick. Children are going to become frustrated if they don't receive some feedback on their action as soon as it's possible or relevant to do so.

2. In case it is not possible to offer instant response, the application should provide a clear indication that the task is in progress. This can be done using a clever and entertaining animation to retain the attention of the children while waiting. The Progress bar or spinning wheel may not work well with children as they sometimes feel as though they've been waiting for so long, even if it has only been for a few seconds.

3. To avoid confusing the children, response times for identical or similar tasks should be comparable.

4. Provide visual and auditory feedback to children whenever they do anything on the application. The feedback content should be easy for the children to understand. The way or format in which the same is given should also be suitable to the intended group and to the context under which the application is used.

5. Provide feedback in such a way that it does not interrupt the user's current work-flow. Notifications should not interrupt the user's current activity unless absolutely necessary and it should be easy to dismiss if the user wants to ignore them.

6. Make it possible to manage audio feedback by proving an option to turn it on/off

7. Keep the notifications and alerts to a minimum and use them only where they add genuine value to the user experience. They should always include valuable data and prompt meaningful actions. Alerts

and feedback should be kept as simple as possible with clear and easy to understand choices generated from them. (Pay particular attention to button labels).

8. Use feedback to validate and explain why an action was invalid. Telling a child that an input is not correct is not a high-value activity. Showing them an example of a correct input is.

9. Strive to use feedback format that recognizes or mentions the child's name. This captures their attention. Children get thrilled when they realize that the application knows some details about them. Therefore use the information you know about them to offer surprising feedback.

Table 6: Feedback Appropriateness

Goal: Feedback Adequacy	
Questions	**Metrics**
Does the application provide visual or auditory feedback to children whenever they perform a task?	Rating scale for appropriate feedback
Is task response time appropriate?	Task response time Rating scale for appropriate response time
Is the way the application responds to user actions consistent at all times?	Difference in response times
Does feedback provided interrupt the user's current workflow?	Number of times user is interrupted while performing a task.
Can children be able to control audio feedback by turning on/off?	Rating scale for child freedom to control feedback

The metrics in red are Objective, The rest are Subjective

3.7 Consistency

Consistency refers to having similar operations and similar elements for achieving similar tasks. According to [24] "it is one of the most important aspects of measuring user interface design principles". Children rely on a consistent experience. And like adults, they would be annoyed if unexpected or random elements ruined their experience.

A consistent representation enables a user to recognize elements and anticipate the next action thereby speeding up the learning curve for new products and services that a user has yet to explore.

3.7.1 Guidelines

1. The UI elements of the design interface (such as graphics, fonts, colors etc.) should be consistent at each point, as should the interaction and content. The content should be consistent in tone, mood, quality, and quantity.

2. "Make it very clear to the user what the next required action is. The more visible the available functions are, the better users will be able to perform their task.

    Interaction and input-output behavior should be consistent within a system as well as across systems. The user should be able to extend knowledge of specific interaction within and across the application to other similar situations" source[25]

3. "Changes to the internal state of the system must be visible so that users can associate with operations that caused them" source [25].

4. "Metaphors should draw on children's existing knowledge so that they can easily see what to do and predict the outcome of their actions" source [25].

5. Limit the number of ways actions and operations are represented, ensuring that users do not have to learn new representations for each task. Follow platform conventions that allow children to complete new tasks without having to learn a new tool-set.

6. Do not deviate a lot from design conventions and standards as children just like adults tend to apply rules they've experienced outside the application, bringing in a set of own expectations.

7. Make input-output behavior easy and understandable. It should be consistent within the application and across other similar applications. "Children should be able to determine the effect of future action based on past interaction history" source [26].

Table 7: Consistency

Goal: Consistency	
Questions	**Metrics**
Are the UI elements of the design interface (graphics, fonts, colors) consistent at each point?	Rating scale for consistency of design elements
Can children be able to tell what is likely to happen on the app based on their past interaction?	Rating scale for familiarity of interface
Can children be able to link metaphors used, to what they represent on the physical world?	Number of metaphors that children cannot identify with
Does the application UI deviate from conventional design standards?	Rating scale for deviation from other design standards

4. CASE STUDY

A case study was conducted involving young children (class 2 and 3 pupils in Kenya) using a learning application on the tablets provided by the government under the OTPC initiative in Kenya. The actual study involved twenty children and one teacher who acted as an expert during the data collection period.

It was conducted in three phases.

1. In Phase 1; the participants were required to answer some demographic questions. From the questions, the aim was to get information about their gender, age, class and also their level of experience while using mobile phones and tablets. To maintain their privacy, their names were not recorded.

2. In Phase 2; the participants were required to perform a series of tasks using the tablets with the aim of collecting quantitative data to answer the objective metrics. The tasks carried out are listed in the next section below

3. In Phase 3; Interviews were conducted shortly after finishing up on the tasks so as to gauge the participants' level of satisfaction with the application. This was done to obtain qualitative data which would help in answering the subjective metrics. The responses were measured on a 5-point Likert scale with responses varying from strongly disagree to strongly agree scale labels.

Task List

Each participant (child) was asked to perform each of the following four simple tasks.

1. Identify the icons in the main menu that they were familiar with

2. Match a picture to its name through drag and drop

3. Type in the plural of a set of words

4. Choose and type the correct word in a pool of words to complete a sentence

For task 2, 3, and 4 the following objective measures were recorded.

- Task Completion time

- Data Entry time

- Response Time

- Number of Errors

- Task Completion rate

- Number of voice support.

User satisfaction questions, in form of a study questionnaire, were then used right after, to interview the children so as to get their opinion about the usability of the application.

5. RESULTS AND DISCUSSION

5.1 Objective Measures

Table 1 below presents the objective measures results that have been derived from the recordings during the execution of tasks.

Table 8: Objective Measures

Objective Metric	Adult	Children		
		Minimum	Maximum	Mean
Task completion time(sec)	120	283	345	310.75
Data entry time(sec)	55	111	127	121.10
Task response time(sec)	5	7	10	8.10
Number of errors	0	0	3	1.45
Task completion rate (%)	100	100	100	100
Number of voice support	0	0	0	0.00

The table gives a summary of the minimum, maximum and means of the data collected for each objective measure from the children participants. The data takes into consideration the average of the three tasks under examination collectively. The data from the teacher (adult/Expert) is also presented to help in making comparisons.

5.1.1 Interpretation

This interpretation is based on how the participants fared on their tasks collectively. The task completion time and data entry time(s) have a considerable difference to the adult's data. The children averaged 310.75 sec(over five minutes) to complete the tasks and 2min for the task that required data entry. In contrast, the teacher averaged two minutes to complete the tasks and less than a minute to enter data.

The children also committed at least an error on average and received no voice support for all the tasks. The task response time is varying because feedback was not instant and participants had to scroll to locate and tap a "CHECK" button after which the results of the tasks would be revealed. Difficulty in scrolling led to the time lapse between completing a task and getting feedback on whether one is correct or wrong.

The task completion rate was at 100%. This has an indication that the children have the ability to perform tasks using the

application to completion when given ample time regardless of whether they will get all the tasks right or wrong.

5.2 Subjective Measures

This section presents the results obtained after analyzing the responses of participants in the study questionnaire.

Table 2 below shows the mean and standard deviation of the subjective measures. The responses were on a scale of 1-5; with 5 meaning Strongly Agree, 4- Agree, 3- Neutral, 2 - Disagree and 1 - Strongly disagree

Table 9: Subjective Measures Results

Satisfaction Measure	Mean	Std. Deviation
General Interaction	2.66	0.90
GUI	3.50	0.46
Navigation & Search	3.04	0.51
Multimedia & Text	3.28	0.60
Content & Engagement	2.83	0.56
Feedback Appropriateness	2.55	0.40
Consistency	2.41	0.28

5.2.1 Interpretation

The mean or average of the highlighted design criteria is recorded. A higher mean score implies that users were highly satisfied whereas a low mean score implies the opposite.

The results from the table indicate that participants were more satisfied with the graphical user interface of the application than on the other criteria, though it only averaged a mean of 3.50 which translates to a Neutral score. Navigation and search completeness and Multimedia and Text also average a neutral score implying that the participants had an ambivalent opinion on these design criteria on the application.

The general interaction, Content & engagement, feedback appropriateness, and consistency scored an equivalent of "disagree" score, meaning that participants probably felt not quite satisfied the application on these areas. Therefore there is a need for improvement of these design aspects.

5.2.2 Graphical Representation of Participants Rating per question

In addition to the analysis from the table, this section presents a graphical representation for each design criteria and satisfaction question on how the participants gave responses for each question.

5.2.2.1 General iinteraction

The graph on general interaction below shows the responses to questions on how the participants were satisfied when generally interacting with the application. The actual

questions which in the graph are coded as C1.1-C1.6 are as follows.

C1.1 - I found it easy to use the application

C1.2 - It is easy to complete tasks without any help

C1.3 - It was simple to key in data

C1.4 -The app provides task demos that show how a task is done

C1.5 - I understand what all the icons in this app stand for
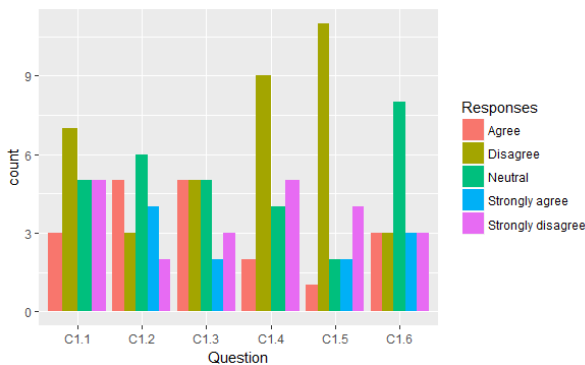
C1.6 - The application takes a lot of time loading tasks.

Figure 2. General Interaction rating

The results on general interaction point out that children did not find it easy to the application on their own. Data entry was not easy for most of them. Additionally, the app did not provide any task demos to help children understand how to perform tasks. Many of them still had problems in identifying what some of the icons represented.

These results are comparable to what [4]observed while also evaluating the interface of mobile application with children. Their observation indicated that "most of the children found it difficult to interact with menu buttons and to access the next screen. They were not sure what to do on the main screen. They had to seek help from a facilitator. There was no indication that the screen was interactive as the icons and buttons lack affordance that they are click-able or audible".

Moreover "there was no help to demonstrate the practices or interactions that children were required to do in each screen". It would be helpful for children to learn on their own if there were proper examples on each screen for the children to follow. This speeds children's learning as they need not go through trial and error nor self-explore.

5.2.2.2 Graphical User Interface
This graph shows the respondents satisfaction with the graphical user interface of the application. The question codes' statements are as follows:-

C2.1 - The main menu for this application is attractive

C2.1 - The organization of information on the screen is clear and consistent

C2.3 - The main menu contains a lot of information(crowded/cluttered)

C2.4 - The buttons and icons are easily recognizable

C2.5 - The buttons and icons are large and finger friendly

C2.6 - The colors used in this application are very attractive.
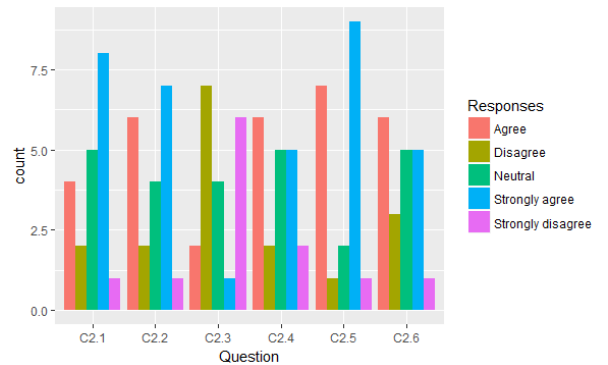

Figure 3. Graphical User Interface rating

From these results, we make observations that most of the children found the main menu of the application being attractive clear and consistent. The icons and buttons used were easily recognizable and finger friendly. The colors used were also found to be attractive. From this we can infer that the apps graphical user interface elements are appropriate to the young children.

The application GUI elements are consistent with the findings of [1] who indicated that to capture children's attention, "use colors and backgrounds that are pleasing to the age group. For kids make use of bright, lovely colors with more graphics". Additionally [27] observed that children prefer seeing many patterns and colors on screen and more interested with animated buttons.

5.2.2.3 Navigation and Search
The graph shows the respondents satisfaction with the navigation and search attributes of the application. The questions codes statements are shown as follows.

C3.1 - The orientation of the screen is comfortable while performing a task.

C3.2 - It was easy to navigate back and forward across the application

C3.3 - Scrolling through the application is easy

C3.4 - It was easy to locate and use the search and help icons

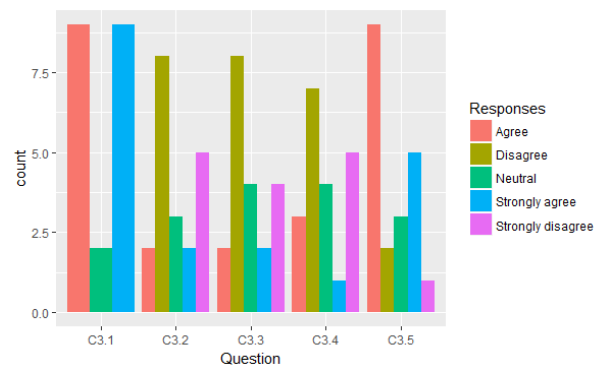C3.5 - It was difficult to understand the search and help results.

Figure 4. Navigation and Search Rating

From this graph, we can deduce the following:- That the participants were comfortable with the screen orientation of the application while performing tasks. However, it is observed that it was not easy to navigate and scroll across the application for most of the participants. They also could neither identify the search and help icons nor understand how to search effectively.

Similar to these findings,[17] while evaluating two educational mobile applications observed that both apps showed poor usability regarding help, tutorials, navigation, voice instructions and error messages. This trend indicates that there is need to improve the navigation, search and help attributes in the context of mobile applications.

5.2.2.4 Multimedia and Text

The graph shows the respondents satisfaction with the multimedia and Text elements as used in the application. The questions code statements are revealed as follows:-

C4.1 - I found the pictures and animations very interesting

C4.2 - It was easy to read the texts and numbers on this application

C4.3 - The size of text and numbers on this application were large enough

C4.4 - The music and sounds in the application were too disturbing

C4.5 - The application provides useful voice instructions

C4.6 - I was able to turn voice instructions On/Off.
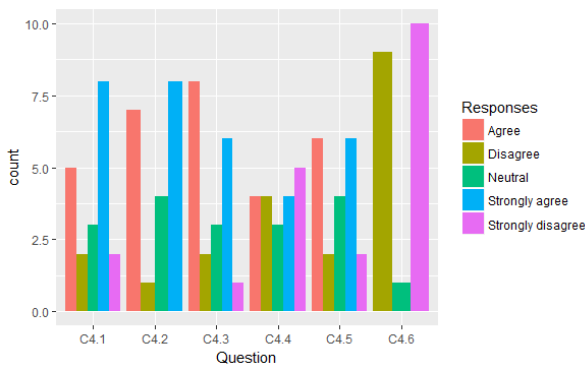
Figure 5. Multimedia and Text Rating

It can be reported that the participants found the pictures and animations interesting to them as shown from the responses. The size of text and numbers in the application can also be said to be friendly to them as it can be observed that they were large enough and easy to read. However, there are mixed reactions concerning the music. Initially it captures their attention towards the tasks, but still somehow disturbing as they could not be able to turn the music on/off.

Multimedia attributes play an important role in children's interaction with mobile applications. As reported by [28], the

musical experience is very important in the context of IT design as

1. It is a natural part of children's life and education

2. It is comprehensive and has an influence on all sensations and imaginations

3. It enriches and supports children creations and

4. It inspires and supports children participations.

5.2.2.5 Content and Engagement

The graph below is an indication of how the content in the application was satisfactory to the participants and how engaged they were while performing tasks in the application.

The questions codes statements are shown below:-

C5.1 - The learning content in this application is satisfactory

C5.2 - The learning activities are interesting and enjoyable

C5.3 - The concepts and learning activities are familiar

C5.4 - I received exciting rewards after completing a task

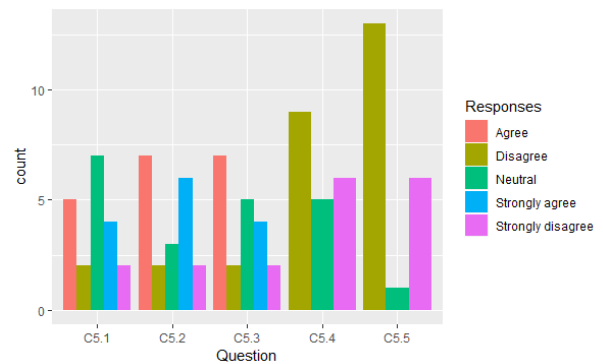C5.5 - The tasks have different difficulty levels.

Figure 6. Content and Engagement rating

The participants felt that the learning content for the tasks they performed was satisfactory to them. The tasks were generally interesting and the concepts familiar to what they are accustomed to. However, they received no rewards from the application upon completing the tasks which is an area of improvement given that children need to be congratulated for every task they perform.

This findings concerning rewards are in a way similar to ones by [8] who observed that 83% of the children in a survey they conducted were of the opinion that "it would be a good idea to receive a prize after every success in the application they use to play or study, whether in the form of stickers, asterisks, or any other kind of bonus points".

5.2.2.6 Feedback Adequacy

The graph shows how the participants were satisfied with the feedback they received from the application as they performed tasks. The question codes statements used in the graph are given as follows.

C6.1 - I receive feedback after completing tasks on whether I am right or wrong

C6.2 - When wrong the app gives error messages that clearly tell how to correctly perform the task

C6.3 - The response time after tasks is slow

C6.4 - Some error messages interrupt me as I perform tasks

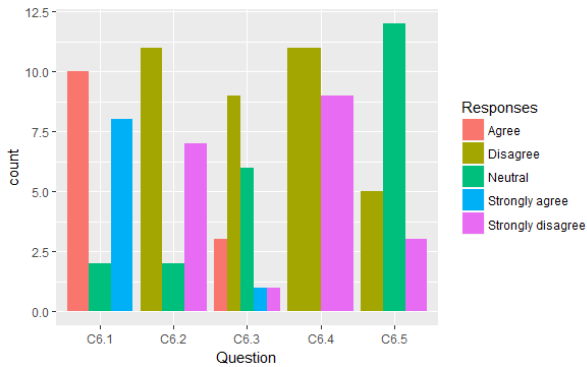C6.5 - I get responses faster in some tasks compared to others.

Figure 7. Feedback Adequacy rating

Most respondents agree that they receive feedback after completing a task on whether they are right or wrong. However, the feedback does not clearly show how to perform the task if one gets it wrong in the first attempt. All participants said that they didn't get interrupted by error messages as they performed tasks.

The inference from this observation is that the participants were not satisfied with the way in which they received feedback while performing tasks. The applications mode of relaying feedback on incorrect answers is inconsistent with what is suggested by [29] that "a wrong answer should be an opportunity for a learning moment." The use of audio and visual feedback should be encouraging and incremental. This article suggests that wrong answer feedback should typically be given within 3 scaffolded levels".

1. First Wrong Answer: Identify a wrong choice and offer encouragement. Example: "That's not it. Try again!" !

2. Second Wrong Answer: Identify a wrong choice, restate the objective, offer a hint, and provide encouragement. Example: "That's not right. You need to find a triangle. It has 3 sides and 3 angles! Try again!"!

3. Third Wrong Answer: Identify a wrong choice, restate the objective, offer a hint, and highlight the correct answer. Example: "That's not right. You need to find a triangle. It has 3 sides and 3 angles! <Triangle highlights> Tap on the triangle!"

"This should be repeated until the correct answer is selected. In some cases, we may suggest moving a child forward if he/she is struggling for a determined period of time".

5.2.2.7 Consistency
The graph shows the rating of how participants felt regarding consistency of various elements of the application.

Question code statements:-

C7.1 - The texts fonts, colors, background color, and animations are consistent across the applications

C7.2 - I am familiar with all the images used in the application in the physical world

C7.3 - The main menu of this application looks similar to others I have seen before

C7.4 - I was able to know how the application would respond based on tasks I have done before.
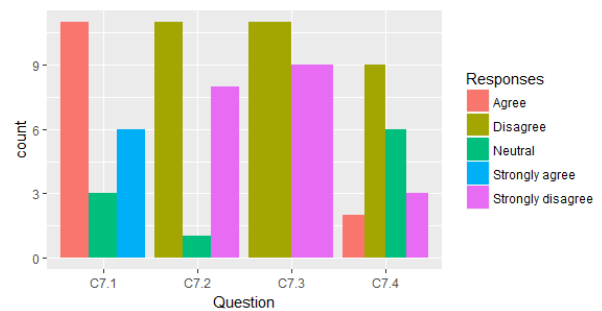

Figure 8. Consistency rating

According to the graph, most elements of the application seem to be consistent across the application. However, most of the respondents are not conversant with all that icons used in the application in the real world. This call for an improvement in this area, to make sure that images and icons used the application are familiar to children in the real world making sure that they can connect their use in the physical world to what they represent in the application.

These finding affirms what [7] suggests that "it is very important to keep design patterns consistent as both kids and adults get annoyed by design elements that seem random and unnecessary". The interactions and feedback should also be consistent so that users will be able to learn how to use the app quickly.

6. CONCLUSION AND FUTURE WORK
This paper focused on reviewing and presenting usability guidelines for children mobile applications and proposing a theoretical evaluation model for the said apps. Most of these guidelines were synthesized from previous research reporting on usability & learnability issues for children mobile applications. They can be helpful to other researchers, practitioners, designers and developers in designing usable mobile applications for young children.

The model serves as a basis for comprehensive learnability evaluation for mobile applications for children. To test the effectiveness of the model, usability testing involving children

was conducted using an educational mobile application used for learning in Kenyan primary schools. The results of the study and the inferences made have been presented in the paper. Not included in the paper but in line with the study is the automation of the Model which may be published later on as future work. Further contributions to this research may include conducting more usability testing using different kinds of applications targeting young children.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1]   A. Al-sa, D. Parry, and P. Carter, "Usability Considerations for Educational Tablet Applications Using an Arabic Interface," *5th Int. Conf. Inf. Commun. Syst.*, 2014.

[2]   A. A. Ali, "A Framework for Measuring the Usability Issues and Criteria of Mobile Learning Applications," no. April, 2013.

[3]   B. Haßler, L. Major, and S. Hennessy, "Tablet use in schools : A critical review of the evidence for learning outcomes," *J. Comput. Assist. Learn.*, vol. 32, pp. 139–156, 2016.

[4]   M. Masood and M. Thigambaram, "The Usability of Mobile Applications for Pre-schoolers," *7th World Conf. Educ. Sci. (WCES-2015), 05-07 Febr. 2015, Novotel Athens Conv. Center, Athens, Greece*, vol. 197, no. February, pp. 1818–1826, 2015.

[5]   R. Viriyapong and A. Harfield, "Facing the challenges of the One-Tablet-Per-Child policy in Thai primary school education," vol. 4, no. 9, pp. 176–184, 2013.

[6]   P. Cardellino and R. Leiringer, "Facilitating change in primary education : the role of existing school facilities in ICT initiatives," vol. 32, no. 13/14, pp. 845–855, 2014.

[7]   D. Gelman, "Design for Kids: Digital Products for Playing and Learning," *UXmatters*, 2014. [Online]. Available: https://www.uxmatters.com/mt/archives/2014/07/design-for-kids-digital-products-for-playing-and-learning.php. [Accessed: 28-Apr-2018].

[8]   R. Kraleva, "Designing an Interface For a Mobile Application Based on Children's Opinion," *Int. J. Interact. Mob. Technol.*, vol. 11, no. 1, pp. 53–70, 2017.

[9]   J. Nielsen, "Children's Websites : Usability Issues in Designing for Young People," *Nielsen Norman Group.Evidence-Based User Experience Research, Training, and Consulting*, 2010. [Online]. Available: https://www.nngroup.com/articles/childrens-websites-usability-issues/. [Accessed: 28-Apr-2018].

[10]   B. K. Okelo, S. M. Mbuguah, and S. M. Karume, "Evaluation of the Learnability Principles of Mobile-Wireless Information Systems ' Application Software for Children in Kenya," *J. Emerg. Trends Comput. Inf. Sci.*, vol. 6, no. 6, pp. 323–333, 2015.

[11]   R. Tahir and F. Arif, "Framework for Evaluating the Usability of Mobile Educational Applications Children," *Int. J. E-Learning Educ. Technol. Digit. Media*, 2015.

[12]   L. Al-wakeel, A. Al-ghanim, S. Al-zeer, and K. Al-nafjan, "A Usability Evaluation of Arabic Mobile Applications Designed for Children with Special Needs -Autism," *Lect. Notes Softw. Eng.*, vol. 3, no. 3, 2015.

[13]   R. Baharuddin, D. Singh, and R. Razali, "Usability Dimensions for Mobile Applications-A Review," *Res. J. Appl. Sci. Eng. Technol.*, vol. 5, no. 6, pp. 2225–2231, 2013.

[14]   I. Rafique, W. Jingnong, W. Yunhong, M. Q. Abbasi, P. Lew, and X. Wang, "Evaluating Software Learnability A Learnability Attributes Model," *Int. Conf. Syst. Informatics (ICSAI 2012)*, pp. 2443–2447, 2012.

[15]   F. Zahra, A. Hussain, and H. Mohd, "Usability Evaluation of Mobile Applications ; Where Do We Stand ?," *2nd Int. Conf. Appl. Sci. Technol. 2017 AIP*, vol. 020056, 2017.

[16]   A. Hussain and M. Kutar, "Usability Metric Framework for Mobile Phone Application," 2009.

[17]   T. Rabail and F. Arif, "A Measurement Model Based on Usability Metrics for Mobile Learning User Interface for Children," *Int. J. E-Learning Educ. Technol. Digit. Media 1(1) 16-31 Soc. Digit. Inf. Wirel. Commun. 2014*, no. August, 2015.

[18]   M. Kutar and A. Hussain, "Usability Metric Framework for Mobile Phone Application," 2009.

[19]   C. Coursaris and D. Kim, "A Meta-Analytical Review of Empirical Mobile Usability Studies," *Proc. Twelfth Am. Conf. Inf. Syst. Acapulco, Mex. August 04th-06th 2006*, vol. 6, no. 3, pp. 117–171, 2011.

[20]   R. Harrison, D. Flood, and D. Duce, "Usability of mobile applications : literature review and rationale for a new usability model," *J. Interact. Sci. 2013*, vol. 1, no. 1, pp. 1–16, 2013.

[21]   R. Ismail, N. Fabil, and A. M. Saleh, "Extension of PACMAD Model for Usability Evaluation Metrics Using Goal Question Metrics (GQM) Approach," *J. Theor. Appl. Inf. Technol.*, vol. 79, no. 1, pp. 90–100, 2015.

[22]   V. R. Basili, *The Goal/Question/Metric Method: a practical guide for quality improvement of software development*. McGraw- Hill Publishing Company,

1999.

[23]   R. Budiu, "Basic Patterns for Mobile Navigation : A Primer," *Nielsen Norman Group. Evidence-Based User Experience Research, Training, and Consulting*, 2015. [Online]. Available: https://www.nngroup.com/articles/mobile-navigation-patterns/. [Accessed: 29-May-2018].

[24]   J. Nielsen, "10 Usability Heuristics for User Interface Design," *Nielsen Norman Group. Evidence-Based User Experience Research, Training, and Consulting*, 1995. [Online]. Available: https://www.nngroup.com/articles/ten-usability-heuristics/. [Accessed: 04-Mar-2018].

[25]   A. Dix, G. Abowd, J. Finlay, and R. Beale, *Human-Computer Interaction*, no. January. 2004.

[26]   J. H. GELDERBLOM, "Designing Technology for Young Children: Guidelines Grounded in a Literature Investigation on Child Development and Children's Technology," no. June, 2008.

[27]   C. Tsai and C. Shih, "An analysis of preferences in smartphone interface design," *ISSN 1992-2248 ©2011 Acad. Journals*, vol. 6, no. 20, pp. 4195–4206, 2011.

[28]   R. Tikkanen and N. Iivari, "The Role of Music in the Design Process with Children," *INTERACT 2011 13th IFIP TC 13 Int. Conf. Lisabon, Port.*, no. PART 3, pp. 288–305, 2011.

[29]   Sesame-Workshop, "Best Practices: Designing Touch Tablet Experiences for Preschoolers," *Sesame Work.*, 2012.

A Model for Code Restructuring, A Tool for Improving Systems Quality In Compliance With Object Oriented Coding Practice

Moses Kibet Yegon Ngetich
Student
School of Computing and Information Technology
JKUAT
Nairobi, Kenya

Dr Collins Otieno
Lecturer
School of Computing and Information Technology
JKUAT
Nairobi, Kenya

Dr Michael Kimwele
Senior Lecturer
School of Computing and Information Technology
JKUAT
Nairobi, Kenya

Abstract: A major goal of software restructuring is to preserve or increase the value of a piece of software. Restructuring a system may make it possible to add more features to the existing system or make the software more reusable in other systems. Software restructuring approaches have become increasingly attractive as the cost of programmer time relative to computer time has increased. Various code restructuring models are often used during software maintenance, where the lack of software structure often is most evident and expensive. However, some models can also be applied in the earlier design and development phases. This research presents a code restructuring model and its associated architecture for improving the quality of object-oriented legacy system and existing ones to a new target system structure. This research reviewed existing literature on code restructuring models and their limitations, this helped in the identification of research gap. Data is collected through outsourcing codes with 'bad smell' from dreamin.net website and observable behaviour of the sample model is recorded after subjecting it to these codes. Data collected is validated, edited and coded then analysed using observable behaviours. The literature on existing restructuring models, techniques and algorithm, frameworks and tools were reviewed and used to determine the nature of the model. Findings revealed that the existing models did not effectively take care of proper restructuring. Finally, the proposed model was developed and validated. The validation process revealed that the model would assist greatly in achieving effective restructuring and therefore the research recommended a restructuring model described in this report.

Keywords: (Restructuring, software maintenance, code restructuring, object-oriented legacy, observable behaviour)

1. INTRODUCTION

Many existing software systems can benefit from code restructuring models to reduce maintenance cost and improve reusability. Yet, intuition-based, ad hoc restructuring can be difficult and expensive, and can even make software structure worse.   Code restructuring is one of the software reengineering activities. It is where the source code is analyzed and violations of structured programming practices are noted and repaired, the revised code also needs to be reviewed and tested. A wide variety of models have been proposed and used to deal with restructuring and restructuring. These include the various techniques and methods for code restructuring processes that have been applied in the development of code restructuring models that can be applied to specific code or to group legacy software. Although various code restructuring models and frameworks that have been proposed before can be used to perform restructuring of various software paradigm, most of these models are limited to a specific restructuring methods and techniques, language paradigm or specific part of the software code and did not meet the intended restructuring objectives. Most authors of these developments also leave an open window for future research of their work

2. RESEARCH OBJECTIVES

The main objective of this study was to develop a code restructuring model to improve the quality of systems in compliance with object-oriented systems. Other specific objectives are;

1   To review related work with regard to existing code restructuring models in Object Oriented Programming
2   To identify weaknesses and challenges of existing code restructuring models in Object Oriented Programming
3   To propose a model for code restructuring based on Object Oriented Programming.
4   To automate the model for code restructuring in compliance with Object Oriented Programming best practice
5   To validate the proposed code restructuring model

3. RESEARCH QUESTIONS

The study was guided by the following research questions;

RQ1 What are the existing code restructuring Models in Object Oriented Programming?
RQ2 What are the weaknesses and challenges of existing code restructuring models in Object Oriented Programming?
RQ3 How can we automate the model for code restructuring in Object Oriented Programming
RQ4 How does the proposed code restructuring model perform in comparison to existing code restructuring models?

4. RESEARCH METHODOLOGY

4.1 Research Design

The study used experimental design which involved a series of model experiments during the research work. The study employed a quantitative research approach using primary data collected during the experiments and observations.

4.2 Target Population

The target population should fit a certain specification which the researcher is studying. For the purpose of this study, the target population will be the object oriented systems and the users. Users are programmers who are involved in the day to day coding of the systems, and are therefore able to provide answers to the research questions.

4.3 Sampling Design and Sample Size

The study will use 10 object oriented systems for the purpose of this study. This is because the greater the sample size, the smaller the sampling error and the more representative the sample becomes (Mugenda & Mugenda, 2003) a sample of 30% is representative.

4.4 Data Collection Method

This research study used primary data. Primary data was collected by use of experiments and observable behavior of the sample systems. The experiments will be conducted using sampled systems in a controlled environment so that the researcher will have ample time to record all results and note down any observable behavior of the system under study at their own convenient time.

Both primary and secondary data will be used. The secondary data about code restructuring models will be collected from external sources, such as websites and books

4.5 Data Analysis and Presentation

The collected data was thoroughly examined and checked for completeness and comprehensibility. Data collected was validated, edited and coded then analyzed using Poison Distribution Model. This distribution is used quite frequently in reliability analysis. It can be considered an extension of the binomial distribution when n is infinite. It can be used to approximate the binomial distribution when n > 20 and p < 0.05.

If events are Poisson distributed, they occur at a constant average rate and the number of events occurring in any time interval are independent of the number of events occurring in any other time interval. For example, the number of failures in a given time would be given by:

$$f(x) = \frac{a^x e^{-a}}{x!}$$

Where x is the number of failures and a is the expected number of failures. For the purpose of reliability analysis, this becomes:

$$f(x; \lambda, t) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

Where:
λ = failure rate
t = length of time being considered
x = number of failures

The reliability function, R(t), or the probability of zero failures in time t is given by:

$$R(t) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}$$

or the exponential distribution.

In the case of redundant equipment, the R(t) might be desired in terms of the probability of r or fewer failures in time t. For that case

$$R(t) = \sum_{x=0}^{r} \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

5. FRAMEWORK DEVELOPMENT

Proposed Framework Architecture and Control Flow. The classes of the input Java project are parsed through the AST Parser. The detection process is done in two phases: During the initial phase, ROOC tool parses each class to gather statistical data by visiting each AST node and creates an array list of the method and variable names for each class. ROOC tool also creates a list of all the class names used during the detection for "Data Class" smell. During the second phase, the ROOC tool uses the gathered statistical data and the AST to identify the code smells requested by the user. The detected code smells are then presented to the user. ROOC tool also provides the option of applying restructuring technique(s) step by step. The user can choose to accept or discard the restructuring suggestions.

5.1 Restructuring Object-Oriented Code TOOL

The proposed tool parses the source code and categorizes those into low, moderate or highly restructured using the metrics

Figure 1. Architecture of ROOC Tool
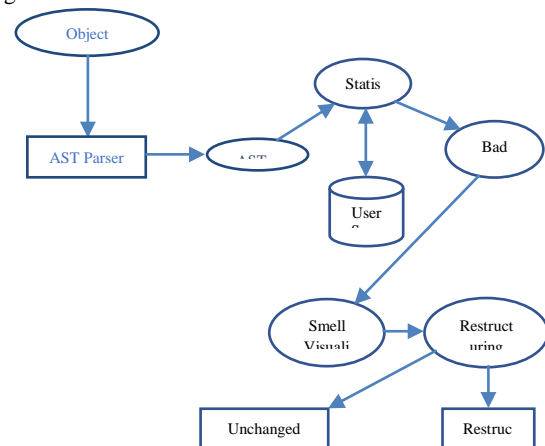
Defined in Table 2 of Section 4.1. The system consists of four components:
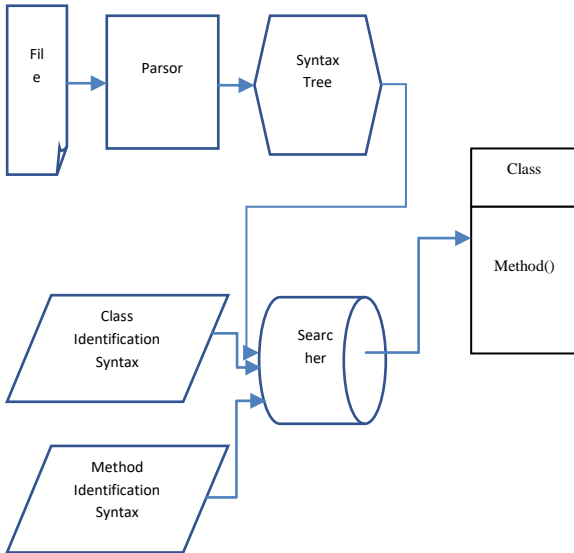1. Parser
2. Analyzer

Figure 1. Illustrates its components

5.2 Proposed Framework Coding

The architectural design of the proposed framework as depicted in Figure 1 consists a number of components with simple interface and with a *pipe and filter* architectural style. Each component (filter) processes its input data in the form of a file (pipe) and stores the results in another file for the next component.

i) Pre-Process Components
ii) Analysis Components
iii) Post-Process Components

6. Framework Implementation
6.3 Implementation Platform

ROOC tool is implemented in Java and uses the Abstract Syntax Tree (AST) parser.

- Abstract syntax tree is the tree structure representation of the source code in any programming language.
- Each node of the syntax tree represents a part of the abstract syntactic structure of the source code.
- The IDE used for the development is Eclipse SDK 3.4.0.
- For refactoring, ROOC tool uses the built in refactoring API of Eclipse, which is a part of Language Toolkit (LTK). The input of the ROOC tool is a Java project folder.

6.4 Proposed Model Validation and Test Results

ROOC tool was tested against 10 projects sourced from the internet. The selected java codes seem to have been developed by experienced Java developers, so the complexities of these codes are considerable.

Each project has an average of 13 classes. These test codes have a good level of complexity. During the design phase, ROOC tool interface was provided to different users from the

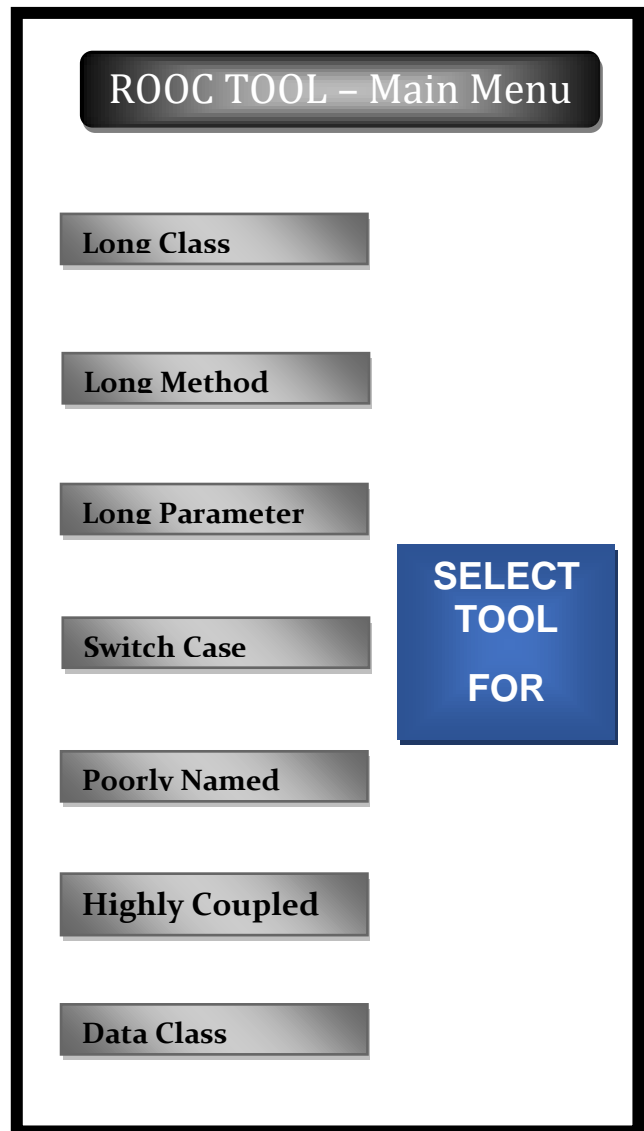technical as well as non-technical background to access the user-friendliness of GUI.

Figure 3. ROOC tool Main Interface menu

The feedbacks were used to improvise the GUI. To test the usability, performance and the code optimization feature of ROOC tool, three different tests were conducted.
1. Identify smells present in each project.
2. Time taken to understand code logic before and after restructuring.
3. Time taken to add functionality in the code before and after restructuring.

6.5 Identify Smells Present in Each Project

During this test, the ROOC tool was run across each of the project and the output was recorded (whether the project contains the specific smell or not). Later we crosschecked to verify correctness of the smell identified by the tool. Even other classes of the projects were skimmed through to identify other cases which the tool might have missed. The smells identified by ROOC tool in individual projects are represented in the tabular format in Table 3. The table cell marked "Yes"

represents the detected code smell in the project enlisted in column 1.

7. Proposed framework User Validation and Tests

Time Taken to Understand Code Logic before and After Refactoring. For this test, four Java developers were chosen ranging from two to three years of experience. The experience of the users ensured that they had sufficient background knowledge of Java to understand the logic. Three projects (named Project 1, Project 2, and Project 3) from the 10 of the above projects were selected having different difficulty level. The details of each of the three projects are shown in Table 4.

8. CONCLUSIONS AND RECOMMENDATIONS

The notion of a "finished product" is rare because existing software constantly evolve. In practice, new features, modifications and adaptations are permanently requested. A consequence is that no initial design, however good, can accommodate all the possible future changes in a real-world project. The agile methodology takes this fact as granted and proposes tools that aim in coping with change rather than defending against it. One category of these tools is restructuring, or changing an existing design. Restructuring has led to restructuring tools, which helps in adapting the existing code automatically in order to be kept synchronized with a change of the design. Restructuring tools, like any software, also evolve over the years. Hence they need to be restructured themselves. This paper discussed an evolution of restructuring tools, namely the evolution toward more complex transformations and consequently presented a generic code restructuring tool for object oriented systems. The need for an evolution was motivated by a complex restructuring: forming a template method. Exploring this restructuring model showed that the existing models and techniques were not suitable to solve some of the restructuring problem. New models and algorithms had to be introduced, such as the code differentiation process. Hence it was necessary to restructure existing algorithms toward more complex ones. Other processes on the other hand had to be restructured toward simpler versions that are more suitable to the restructuring process, such as the data and control flow analyses used for method extraction. In Chapter 1, we motivated the need for restructuring and explained the broader context in which it is used. Restructuring is not a new process, and the state of the art regarding automated implementations was presented in the literature review. Extensions of existing approaches as well as new approaches have been presented, discussing the restructuring model development, testing and validation in chapter four. Here we proposed an interoperable code restructuring implementations on object oriented codes and demonstrated how to automatically solve minor problems rather than systematically reporting them as errors to the user. Analyses, models and transformations have been used to implement our case study. However they also have other applications in many other areas. Conversely, other areas, such as web programming, still seek for additional research. We continue by summarizing our contributions, making a critical analysis of our work, and highlighting future work.

8.1 Recommendations

We have explored a more generic restructuring as a case study: forming a restructuring model. This model is further decomposed into other, smaller restructuring activities discussed in chapter Four, section 4.3. There are of course many other restructuring techniques and methods to explore that have not yet been implemented, and this could be a future research direction. In particular, our case study has shown that existing techniques are not sufficient to implement the interoperable code restructuring implementations on object oriented codes, and required extensions. It is however too early to say whether and to what extent the new introduced approaches can or cannot be reused for other, even more complex restructuring. Finally, we took a pragmatic approach to the problem, which allowed the project to get a working solution. On the other hand, a more formal approach would be necessary to discuss our algorithm in terms of its properties (preconditions and post conditions) and correctness. Formal approaches may eventually find out that parts of our algorithms are wrong or suboptimal, or may just need adjustments and extensions to cope with future programming languages. Following the Agile development philosophy, this would not be a problem: as with any real-world application, in such a case this thesis and the underlying research would just need to be researched on further.

8.2 Acknowledgement

9. REFERENCES

[1] Gligoric, M., Behrang, F., Li, Y., Overbey, J., Hafiz, M., Marinov, D.: Systematic Testing of Restructuring Engines on Real Software Projects. In: Castagna, G. (ed.): Proceedings of the 27th European Conference on Object-Oriented Programming (ECOOP'13). LNCS, Vol. 7920. Springer-Verlag, Berlin Heidelberg, 629-653. (July 2013)

[2] Massoni T., Gheyi R., Borba P. (2008) Formal Model-Driven Program Restructuring. In: Fiadeiro J.L., Inverardi P. (eds) Fundamental Approaches to Software Engineering. FASE 2008. Lecture Notes in Computer Science, vol 4961. Springer, Berlin, Heidelberg

[3] HAMIOUD, Sohaib. "UneApprochedirigée par les Modèles pour les Architectures Logicielles." (2016).

[4] Arendt, Thorsten, Enrico Biermann, Stefan Jurack, Christian Krause, and Gabriele Taentzer. "Henshin: advanced concepts and tools for in-place EMF model transformations." Model Driven Engineering Languages and Systems (2010): 121-135.

[5] Schaefer, M. and De Moor, O., 2010, October. Specifying and implementing restructurings. In ACM Sigplan Notices (Vol. 45, No. 10, pp. 286-301). ACM.

[6] Moha, N., Mahé, V., Barais, O. and Jézéquel, J.M., 2009, October. Generic model restructurings. In International Conference on Model Driven Engineering Languages and Systems (pp. 628-643). Springer Berlin Heidelberg.

[7] Shonle, M., Griswold, W.G. and Lerner, S., 2007, September. Beyond restructuring: a framework for modular maintenance of crosscutting design idioms. In Proceedings of the the 6th joint meeting of the European software engineering conference and the ACM SIGSOFT symposium on The foundations of software engineering (pp. 175-184). ACM.

[8] S. Tichelaar, FAMIX Java language plug-in 1.0, Technical, Report, University of Berne, September 1999.

[9] Raul Marticorena, "Analysis and Definition of a Language Independent Restructuring Catalog", 17th Conference on Advanced Information SystemsEngineering (CAiSE 05). Portugal., page 8, jun 2005.

[10] Pieter Van Gorp, Hans Stenten, Tom Mens, and Serge Demeyer. "Enabling and using the UML for model driven restructuring". 4th International Workshop onObject-Oriented Reengineering (WOOR), (Germany), July 21st, 2003.

[11] Technical Report 2003-07 of the University of Antwerp (Belgium), Department of Mathematics & Computer Science, 2003.

[12] Tom Mens, Tom Tourwe,"A Survey of Software Restructuring", IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 30, NO. 2, FEBRUARY 2004

[13] 31Days of Restructuring, "Useful restructuring techniques you have to know" October 2009Sean Chambers, Simone ChiarettaAnshu et al., International Journal of Advanced Research in Computer Science and Software Engineering 2 (12), December - 2012, pp. 256-260

A Review of Fuzzy Logic Model for Analysis of Computer Network Quality of Experience

Walter .B. Kihuya

Institute of Computer Science
and Information Technology,
Jomo Kenyatta University of
Agriculture and Technology,
Mombasa, Kenya.

Dr. Calvins Otieno

Institute of Computer Science
and Information Technology,
Jomo Kenyatta University of
Agriculture and Technology,
Nairobi, Kenya.

Dr. Richard Rimiru

Institute of Computer Science
and Information Technology,
Jomo Kenyatta University of
Agriculture and Technology,
Nairobi, Kenya.

Abstract: The estimation of the QoE provides valuable input in order to measure the user satisfaction of a particular service/application. Network QoE estimation is challenging as it tries to measure a subjective metric where the user experience depends on a number of factors that cannot easily be measured. All the Network analysis models can be divided into two major groups: qualitative and quantitative. In recent years many quantitative models have been developed in terms of quantitative measures i.e. use of scale of numbers between 1 to 5 to represent user perception of QoS. The challenge with this model is where user perception is subjective and not precise thus cannot be clearly measured using quantitative methods. On the other side qualitative models are in early stages of exploration. Little has been done on qualitative methods. Basing on previous studies, few models exists that measure qualitative analysis of computer network quality of experience. However none incorporated all the four parameters of integrity of service; throughput, delay, packet loss and jitter as parameters of network QoE. In this work, a literature survey is done on qualitative performance by use of a variety of variables, input and output linguistic terms. After a broad survey of the literature, we tend to propose a fuzzy logic model for analysis of computer network QoE. Likewise, the model combines all the four parameters of network integrity of service parameters since they are the primary factor for QoS quantification of any network [1]. Moreover, by using the fuzzy logic concept, the output linguistic terms shows the user perception about a product or a service (QoE) to certain levels by use of membership functions, in this case triangular membership function which shows the mapping of each linguistic term to certain range of values rather being precise to a particular value. By such means, the developed fuzzy logic model tends to accommodate some degree of uncertainty and vague network values to be used for analysis purposes. The concern is to advance the analysis and evaluation of quality of experience in computer networks by use of fuzzy logic concept. The target population for this model is the ISPs' clients. This will enable ISPs to have the best responsive measures to deal with clients' QOE parameters so as to meet the QOS as per SLAs.

Keywords: fuzzy logic, ISPs (Internet Service Providers), quality of experience (QoE), Quality of service (QoS), SLAs (Service Level Agreement)

I. INTRODUCTION

As the service requirements of network applications shift from high throughput to high media quality, interactivity, and responsiveness, the definition of Quality of Experience (QoE) has become multidimensional.

QoE in the context of telecommunications networks is defined as the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user's personality and current state [2]. QoE is defined by International Telecommunication Union (ITU) as the overall acceptability of an application or service, as perceived subjectively by the end-user. [3]

The information technology (IT) and electronics industries apply the QoE model to businesses and services since QoE depends on customer experience; assessments are compiled from large user group polls. QoE models are in two broad categories: Qualitative and Quantitative models. Among the most commonly used quantitative model is mean opinion score (MOS). The MOS is expressed as a single rational number, typically in the range 1–5, where 1 is lowest perceived quality and 5 is the highest perceived quality. Other MOS ranges are correspondingly possible, depending on the rating scale that has been used in the underlying test. This model is thus quantitative in nature while user perception is subjective and not precise thus cannot be clearly measured using quantitative methods as it tends not to accommodate uncertainty [3]

Fuzzy logic is a problem solving methodology that provides a simple way of definite conclusions from vague and

imprecise information. Fuzzy set theory was first introduced by Zadeh in 1965. He was motivated by observing that human reasoning can utilize concepts and knowledge that don't has well defined boundaries [4].

In recent years, the number and variety of applications of fuzzy logic have increased significantly. The applications range from consumer products such as cameras, camcorders, washing machines, and microwave ovens to industrial process control, medical instrumentation, decision-support systems, portfolio selection and network analysis.

In the case of Network analysis, all Network analysis models can be divided into two major groups: qualitative and quantitative. Qualitative metrics do not possess quantitative values and cannot be measured by numerical numbers. In that case, linguistic terms are used to evaluate performance of qualitative metrics [5]

QoE estimation is challenging as it tries to measure a subjective metric where the user experience depends on a number of factors that cannot easily be measured. In subjective assessment, quality is judged by users. Vagueness occurs when users have different interpretation of the same word (linguistic values) like poor, better.

II. THEORETICAL REVIEW

Network Quality of experience (QoE)

Quality of experience (QoE) is defined as the overall acceptability of an application or service, as perceived subjectively by the end-user [3]

The concept of QoE is used to measure user satisfaction level as shown in Figure 2.1. QoE includes complete end-to-end system ranging from users, terminal, customer premises network & core network and access network to service infrastructure.

Therefore based on the definition of QoE, **network QoE** can be defined as the overall acceptability of the network service(s) as perceived subjectively by the end-user.
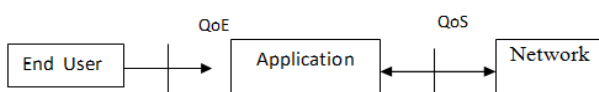
Figure2. 1: Relationship between QoS and QoE

Fuzzy logic

Fuzzy logic is a natural, continuous logic patterned after the approximate reasoning of human beings [6]. As a theory mathematical discipline, fuzzy logic reacts to constantly changing variables [6].It challenges traditional logic by not being restricted to the conventional binary computer values of zero and one. Instead, fuzzy logic allows for partial truths and multivalued truths [6].Fuzzy logic is especially advantageous for problems that cannot be easily represented by mathematical modeling because data is either unavailable or incomplete or the process is too complex [6].The real world language used in fuzzy control enables engineers to incorporate ambiguous, approximate human logic into computers using linguistic modeling, as opposed to mathematical modeling, greatly simplifies the design and modification of a fuzzy logic system [6].

Fuzzy set theory [7] was developed to address contexts in which decision Makers need to accurately analyze and process information that is imprecise in nature.

Fuzzy sets provide a conceptual framework, as well as an analytical tool to solve real World problems where there is a lack of specific facts and precision. Human semantics are embedded in the meaning of fuzziness and comparison [8] On the other hand; the usage of multi granularity linguistic information can eliminate the difference from evaluators [9]

An objective of fuzzy logic has been to make computers think like people [6]. Fuzzy logic can deal with the vagueness intrinsic to human thinking & natural language and recognizes that its nature is different from randomness [6].Using fuzzy logic algorithms could enable machines to understand and respond to vague human concepts such as hot, cold, large, small, etc. It also could provide a relatively simple approach to reach definite conclusions from imprecise information [6].

Integrity of service

Integrity of service involves maintaining the consistency, accuracy, and trustworthiness of data over its entire life cycle. Data must not be changed in transit, and steps must be taken to ensure that data cannot be altered by unauthorized people for example, in a breach of confidentiality. These measures include file permissions and user access controls. Some means must be in place to detect any changes in data that might

occur as a result of non-human-caused events such as an electromagnetic pulse (EMP) or server crash. Some data might include checksums, even cryptographic checksums, for verification of integrity. Backups or redundancies must be available to restore the affected data to its correct state [10].

ISPs (Internet Service Provider)

Internet Service Provider is a company that provides customers with Internet access. Data may be transmitted using several technologies, including dial-up, DSL, cable modem, wireless or dedicated high-speed interconnects. [3]

Among the largest national and regional ISPs are AT&T WorldNet, IBM Global Network, MCI, Netcom, UUNet, and PSINet. Examples of ISPs in Kenya include Zuku, Safaricom, Airtel, Orange, Faiba internet, Access Kenya, Internet Solution etc.

Quality of Service (QoS)

Quality of Service (QoS) is the degree of conformance of the service delivered to a user by a provider in accordance with an agreement between them [3]

In order to assess QoS of any network efficiently, network and service related performance metrics should be identified carefully. The available literatures in this context analyze the diverse characteristics of such metrics. Each service has some crucial parameters that greatly influence its performance. This has been studied in detail in [11]

These studies suggest that for QoS evaluation of any network, it is useful to choose metrics, which are interrelated to QoE parameters. As a result, it is possible to assess the network QoE by simply applying the values of the network QoS parameters.

Traditional Approaches to analysis of network QoE

Mean Opinion Score (MOS) is a numerical value which is used as an index for expressing the QoE. The standard scores for MOS are available in ITU-T Recommendation P.800 [3]. These ranking grades, which are shown in Table below, were originally applied for the quality assessment of telephone transmission. Today, these judgment scales are used to evaluate the QoE for different applications and services. For

instance, MOS was used for evaluation of users' experience of web browsing [12]

Table2. 1 : Opinion scores recommended by ITU-T.

Score	Sequence quality
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Today, compared to the objective assessment of QoS parameters, the subjective assessment of QoE has attracted an increasing attention. Traditional mechanism commonly used in assessments is a survey in which users' opinions are asked on a five-point scale and later on average the results collected to calculate the Mean Opinion Score (MOS) value to be used for analysis/evaluation. Each user rates the perceived quality based on different quality scales ranged from 1 to 5, where 5 refers to the excellent perceived quality and 1 indicates the lowest experienced quality. The results from MOS will be an indicator for service providers to understand the user's perception and define proper thresholds for each quality scale.

This approach still comes with its share of challenges. For example, the MOS scores 1, 2, 3, 4, and 5 correspond to Bad, Poor, Fair, Good, and Excellent experiences, respectively. It is not surprising that people may have different interpretations of a similar linguistic term for instance "Bad" and give different ratings even though they have the same experience in a test. Likewise, each linguistic term is mapped to a particular value thus not accommodating imprecision.

These observations prompted scholars to further investigate on the causal relationship between smart technology and QoE measurement. Fuzzy logic technique is one of the smart technologies in use. Others include neural network, Support Vector Machine, Naive Bayes, Decision Tree etc. Several researches have been done on fuzzy logic in relation to QoE though little has been done on fuzzy logic model for analysis of computer networks quality of experience.

III.METHODOLOGICAL REVIEW

Fuzzy logic models for analysis of Qualitative performance:

Several researches have been done on fuzzy logic in relation to quality performance though little has been done on fuzzy logic model for analysis of quality of experience.

The study in [13] proposed a Fuzzy logic aggregation of wireless sensor network data for smart traffic light control. This approach uses smart traffic control systems (STCS) to make traffic routing decisions. STCS use real time data and mimic human reasoning thus prove promising in vehicle traffic control. This presents a smart traffic light controller using fuzzy logic and wireless sensor network (WSN). The approach is designed for an isolated four way roundabout. It employed fuzzy logic to control the lights and determine how the green light will be assigned for each approach. The WSN collected the traffic data in real time. This data is aggregated and fed into a fuzzy logic controller (FLC) in form of two inputs – traffic quantity (TQ) and waiting time (WT) for each approach. Based on the inputs, the FLC then computes an output priority degree (PD) that controls green light assignment. Using the PD, an algorithm is formulated that assigns green light to the lane with highest PD. The cycle continues until all approaches get green.

In [14] a research study on a Fuzzy Logic System for Evaluating Quality of Experience of Haptic-based Applications was proposed. The proposed taxonomy was modeled with a fuzzy logic system and finally was tested by a Mamdani fuzzy inference system. In the mentioned study, by making some assumption like rule selection and membership function selection, the effect of different perception measures parameters such as rendering quality, physiological and psychological was studied. Here, fuzzy logic system was applied for objective measuring of QoE parameters.

The research work in [15] exhibited QoE estimation for web service selection using a Fuzzy-Rough hybrid expert system. A methodology to estimate the quality of web services based on a fuzzy-rough hybrid algorithm is proposed. The estimated web QoE is used to select the most performing service among different web services. Fuzzy expert systems are good at making decision with imprecise information; however, they cannot automatically formulate rules that they require for making the decisions. Therefore, a fuzzy-rough hybrid expert system is proposed in this study where rough set theory is used to define the rules necessary for the fuzzy expert system. Three QoS parameters: reliability, execution time (in seconds), and availability (in seconds) were measured during the performance of the tests. Input linguistic terms were: Low, Medium and High. The output linguistic terms in use were: Bad, Poor, Fair, Good and Excellent.

The research work in [16] proposed analysis of Quality of Experience by applying Fuzzy logic: A study on response time. In this work, with a fuzzy perspective, the effect of response time variation in a network on the quality perceived by users is shown. Later, shows how by applying fuzzy techniques the linguistic terms and the users' perception can be translated into quantitative values. The main objective of this project was to analyze the fuzziness of QoE in order to provide more understandable user perception. This included proposing response time performance criteria that correlate well with QoE measurement result presented by fuzzy concepts. The proposed methodology provides a fuzzy relationship between QoE and Quality of Service (QoS) parameters. To identify this fuzzy relationship a new term called Fuzzi ed Opinion Score (FOS) representing a fuzzy quality scale is introduced. A fuzzy data mining method is applied to construct the required number of fuzzy sets. Then, the appropriate membership functions describing fuzzy sets are modeled and compared with each other. The proposed methodology intended to assist service providers for better decision-making and resource management [16].

In [17], an efficient algorithm for transmitting packet for better quality of service in adhoc mobile network was proposed. In this study, Fuzzy Self Organizing Map (FSOM) provide very efficient algorithmic tools for transmitting packet in an efficient manner by taking the most efficient route, the bandwidth, latency and range network parameters are considered to determine how good is the data delivered. The results indicated that fuzzy logic can guarantee QoS of every packet in the network. Incorporation of fuzziness in the input and output of the proposed model was seen to result in better performance. Input variables were only three properties: low, normal, and high. The output variables were poor, good and excellent.

In [18] a fuzzy logic based approach is in use for maintaining VoIP Quality in a network which is affected by many network factors (packet loss, packet delay, and jitter).In this case, Resource Reservation Protocol application was configured to

control Token Bucket Algorithm and the simulation experiments were carried out with Opnet. In addition, comparison between Token Bucket with and without Quality of Service aimed at measuring network factors was performed. In this paper, building Fuzzy Token Bucket System consisted of three variables (Bandwidth Rate, Buffer Size, and New Token) in order to improve Token Bucket Shaper output variable (New Token) by Fuzzy Stability model for Voice over IP quality maintaining. The linguistic values in use for each variable were: Buffer Size {VL, L, M, H, and VH}, Bandwidth Rate {VL, L, AL, BA, AV, AA, BH, H, and VH} and New Token {VL, L, BA, AV, AA, H, and VH}

The study in [19] revealed the analysis of the impact of different network QoS parameters on users perceived video QoE for VoD (Video-on-Demand) services. Network parameters in use included: Packet loss rate, Burst packet loss and Jitter. The input linguistic terms involved were Very annoying, slightly annoying, Imperceptible, Annoying and perceptible but not annoying. The output linguistic terms in use were Very annoying, slightly annoying, Imperceptible, Annoying and perceptible but not annoying. This study proposed a methodology based on a fuzzy expert system to objectively estimate the video QoE. To validate the methodology, the developed system was integrated as part of a monitoring tool in an industrial IPTV (Internet Protocol Television) test bed and compared its output with standard Video Quality Monitoring (VQM). The evaluation results show that the proposed video quality estimation method based on fuzzy expert system can effectively measure the network impact on the QoE.

In [19], a methodology and a system based on fuzzy expert system to estimate the impact of network conditions (QoS) on the QoE of video traffic were presented. At first, subjective tests to correlate network QoS metrics with participants' perceived QoE of video traffic was conducted. Secondly, a No Reference method based on fuzzy expert system to estimate the network impact on the video QoE was proposed. The membership functions of the proposed fuzzy system were derived from normalized probability distributions correlating the QoS metrics with QoE. A simple methodology to build the fuzzy inference rules was proposed. The system was evaluated in two different sets of experiments. The estimated video quality showed high correlation with the subjective QoE obtained from the participants in a controlled test. The evaluation results show that the proposed video quality

estimation method based on fuzzy expert system can effectively measure the network impact on the QoE. [19]

In this study, six video clips of different types (sports, movie, animation, and interview) were used and generated 228 sample video clips which were constructed with different network level perturbations.

These video clips were constructed by streaming from a server to client and correspondingly introducing perturbation through emulated network. Three QoS parameters for perturbation were selected; packet loss, jitter and packet loss burstiness, which were considered promising for the mapping of QoS to QoE for video traffic. [19].

The variation of these QoS metrics impacts the quality of the delivered video and, consequently, the user satisfaction level. The objective was to design and implement a method to estimate the variation of the user satisfaction level in function of the network QoS conditions [19].

In the subjective test, different video clips were presented to the 25 participants who rated each video clip according to the perceived impairment giving one of the following scores:

_ Imperceptible (score 5): Excellent conditions.

_ Perceptible but not annoying (score 4): Good conditions.

_ Slightly annoying (score 3): Fair conditions.

_ Annoying, and (score 2): Poor conditions.

_ Very annoying (score 1): Bad conditions.

From the subjective test, a learning set that consisted of the mapping between the participants' scores and the QoS metrics for each of the considered video clips was built. A probabilistic approach to correlate QoS metrics to the participants' scores was used. Therefore, for every QoS metric, five different probability distribution functions (pdf) was built, one function per QoE score that provide the variation of the participants' ratio (%) with the QoS metric for a specific QoE score. This probabilistic information was changed into a fuzzy set by dividing the pdf by its peak value i.e. normalized pdf [20].

The fuzzy set, which has the same form as that of the original pdf, was converted into an equivalent triangular or trapezoidal fuzzy set by using a curve fitting method as demonstrated in [21].The triangular or trapezoidal fuzzy set represents the membership functions for the different QoS metrics as illustrated below.
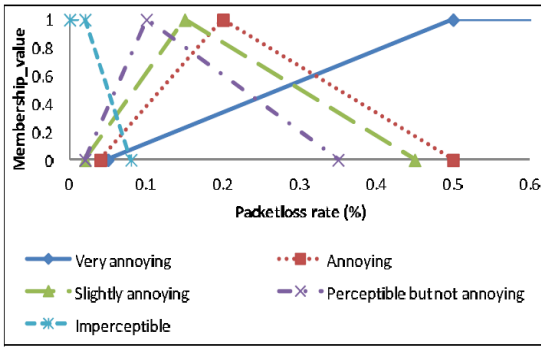
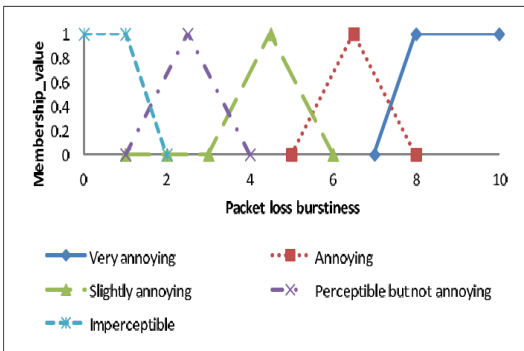Figure2. 2: Membership functions for packet loss rate metric. **[19]**

Figure2. 3: Membership functions for packet loss burstiness metric. **[19]**
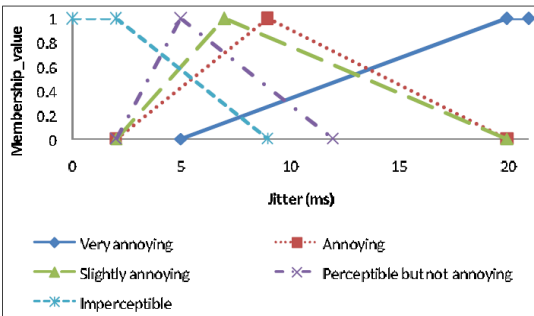

Figure2. 4: Membership functions for jitter metric. **[19]**

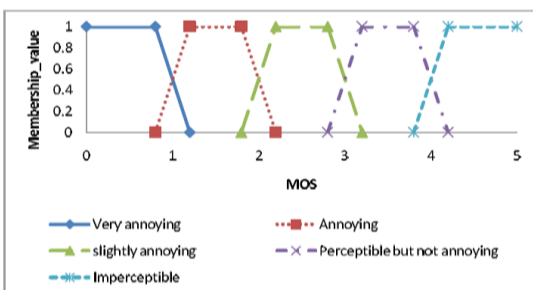
Figure2. 5: Memberships function for the estimated MOS: eMOS. [19]

Critiques of the existing Literature relevant to the study:

Based on the literature of the work cited, it's a clear indication that each research work used different types of parameters and linguistic terms.

For QoS evaluation of any network, it is useful to choose metrics, which are interrelated to QoE parameters. As a result, it is possible to assess the network QoE by simply applying the values of the network QoS parameters. The QoE parameters include Accessibility, Retain_ability and Integrity of Service each of them having its corresponding QoS parameters under consideration [10].

Few of the research work considered the underlying QoS-related parameters, which are linked to the integrity of service QoE parameters but none of them considered all the four parameters (Throughput, delay, jitter and packet loss).

Research Gap

Few models exist that analyze qualitative analysis of network QoE though most of them have limited network parameters such as [19] used packet loss, packet loss burstiness /delay and jitter metrics. None of the models incorporated all the four Network QoE metrics for Integrity of Service (throughput, delay, packet loss and jitter) as major parameters of network QoE. Therefore, this review is inspired to address this gap by presenting an alternative approach of analyzing underlying QoS related parameters under integrity of service QoE parameters incorporating all the four Network QoE metrics for Integrity of Service i.e. throughput, delay, jitter and packet loss by use of fuzzy logic concept. These four parameters are considered to be the primary factors which affect any computer networks [10]. The concern is to advance the analysis and evaluation of quality of experience in computer networks by use of fuzzy logic concept.

Fuzzy Logic Model for Analysis of computer Networks Quality of Experience.

Based on the literature work covered on fuzzy logic technique for analysis of qualitative performance by use of various variables,input and output linguistic terms,this survey tends to propose a fuzzy logic model for analysis of computer networks QoE.

Five input linguistic terms are identified: Very High, High, Medium, Low and Very Low. This approach is in reference to [11] , whereby exhausting the concept of fuzzy membership function, the working ranges of network parameters are set using the term set {Very Low (VL), Low (L), Medium (M), High (H), and Very High (VH)} rather than simply stating an acceptable range. This enables the integration of network and/or service uncertainty and dynamics together.

Likewise, centering on [22] indicates that the permissiveness of fuzziness in the human thought process suggests that much of the logic behind thought processing is not traditional two valued logic or even multivalued logic, but logic with fuzzy truths. Middle values for all the terms can be introduced and the concept of linguistic hedges can be used to identify for each fuzzy linguistic variable such as very low, very high and so on.

Moreover, in relation to [23] ,its clearly illustrated that linguistic hedge (linguistic modifier) is a function that alters the membership function of the fuzzy set associated to the linguistic label, obtaining a definition with a higher or lower precision depending on the case. Two of the most well-known modifiers are the concentration linguistic hedge "very" and the dilation linguistic hedge "more-or-less." Grounding on these facts, it prompted for the use of the stated input linguistic terms.

Five output linguistic terms are defined to describe the opinion scores: Excellent, Good, Fair, Poor and Bad. These values are based on International Telecommunication Union (ITU-T) recommendation for evaluation of QoS.

Four QoS parameters are used: delay, jitter, packet loss and throughput.

QoS parameters are factors that can affect the quality of service. These parameters are very important in Service Level Agreement (SLA) monitoring. Moreover, recent research shows the effect of variation of QoS parameters on the level of user satisfaction. Packet loss, delay (latency) and jitter are some of the most important parameters [24].

Therefore, determining a unified network QoS value can simplify the process of network QoE evaluation. The mappings between these QoS and QoE parameters are presented below:
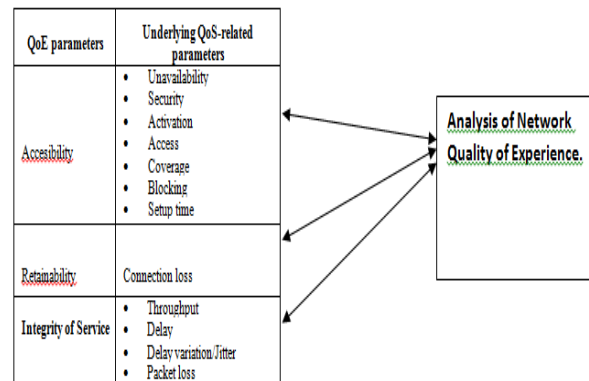
Table5. 1: Mapping between QoE and QoS Related parameters [25]

QoE parameters	Underlying QoS-related parameters
Accesibility	• Unavailability • Security • Activation • Access • Coverage • Blocking • Setup time
Retainability	Connection loss
Integrity of Service	• Throughput • Delay • Delay variation/Jitter • Packet loss

Conceptual Framework

The QoE parameter under considerations is Integrity of Service having four underlying QoS-related parameters: Throughput, Delay, Jitter and packet loss. These are the primary factors for QoS quantification of any network as indicated by [1]

Table2. 2: Conceptual Framework derived from [10]

The above mappings between Network Integrity of Service QoE and QoS related parameters when manipulated in matlab environment effects in to below Conceptual Framework model:
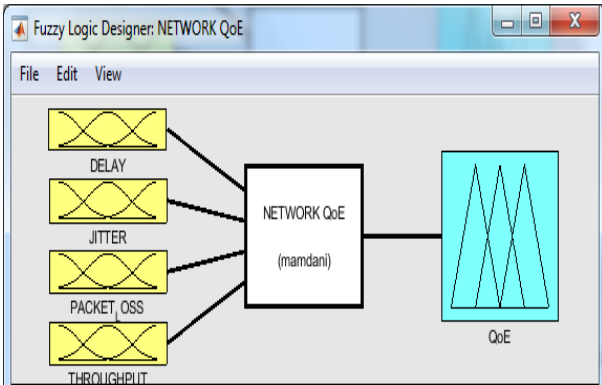

Figure2. 6: Conceptual Framework model for mappings between Network Integrity of Service QoE and QoS related parameters.

The model presents the following four attributes under network Integrity of service QoE:

Delay: Refers to an average time needed for a packet to reach from source to destination [19]**.** This parameter is intrinsic to communications, since the end points are distant and the information will consume some time to reach the other side. Delay is also referred as to latency. Delay time can be increased if the packet face long queues in the network (congestion), or crosses a less direct route to avoid congestion.

Jitter: Jitter is the variation in the packet inter-arrival delay [19]

It involves the delay variation and is introduced by the variable transmission of delay of the packets over the network. This can occur because of routers' internal queues behavior in certain circumstances for instance; flow congestion, routing changes, etc. This parameter can seriously affect the quality of streaming audio and/or video.

To handle jitter, it is needed to collect packets and hold them long enough until the slowest packets arrive in time, rearranging them to be played in the correct sequence.

Packet Loss: happens when one or more packets of data being transported across the internet or a computer network fail to reach their destination [19]. Wireless and IP networks cannot provide a guarantee that packets will be delivered at all, and will fail to deliver (drop) some packets if they arrive when their buffers are already full. This loss of packets can be caused by other factors like signal degradation, high loads on network links, packets that are corrupted being discarded or defect in network elements. Some transport protocols such as Transfer Control Protocol (TCP) make delivery control by receiving acknowledgements of packet receipt from the receiver. If packets are lost during transfer, TCP will automatically resend the segments which were not acknowledged at the cost of decreasing the overall throughput of the connection.

Throughput: refers to the number of bits received during a time unit [19]

Throughput is the amount of data which a network or entity sends or receives data, or the amount of data processed in one determined time space. It's basic unit of measures is bits per second (bit/s or bps). The throughput can be lower than the input strain due to losses and delays in the system. Throughput is a good measure of the channel capacity of a communications link.

IV. METHODOLOGY:

Fuzzy logic methodology in use involves the following steps:

a. Defining the linguistic variables and terms (Initialization).
b. Constructing the membership functions (Initialization).
c. Constructing the rule base (Initialization).
d. Converting the crisp input data to fuzzy values using the membership functions (Fuzzification).
e. Evaluating the rules in the rule base (Inference)
f. Combining the results of each rule (Inference).
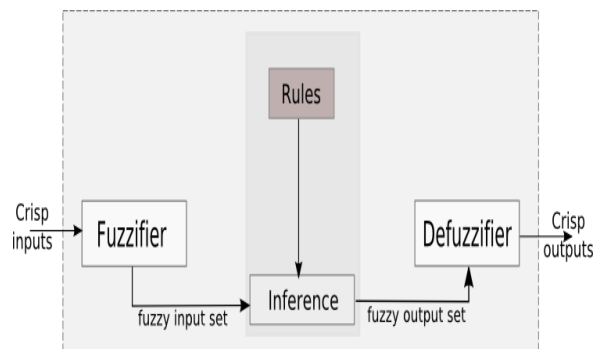g. Converting the output data to non-fuzzy values (Defuzzification).

Figure3. 1: Fuzzy Logic System

Initialization

This process involves Defining the linguistic variables and terms.

Constructing the membership functions (MF) is done at this stage. The process involves determining a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. The input space is sometimes referred to as the universe of discourse. A triangular membership function is used in this work to obtain the degree of membership for each linguistic term. According to [26], Triangular membership functions are simple and therefore facilitate easy computation. As they are also piecewise linear they provide a linear mapping of the universe of discourse in the case of a look-up table method. For instance:
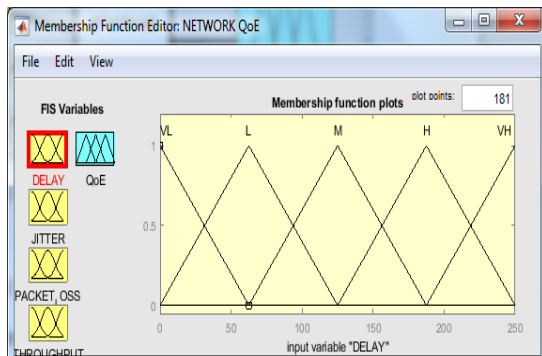
Figure4. 1: Developed membership function plots for Delay input linguistic term.
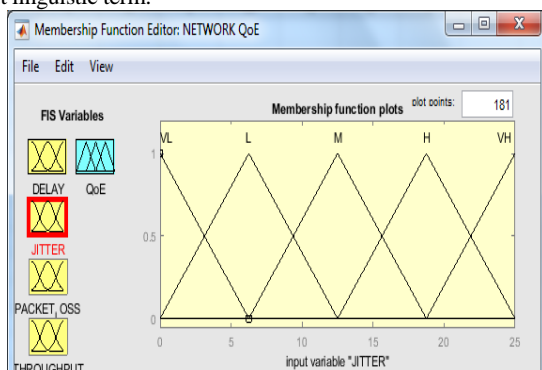

Figure4. 2: Developed membership function plots for Jitter input linguistic term.
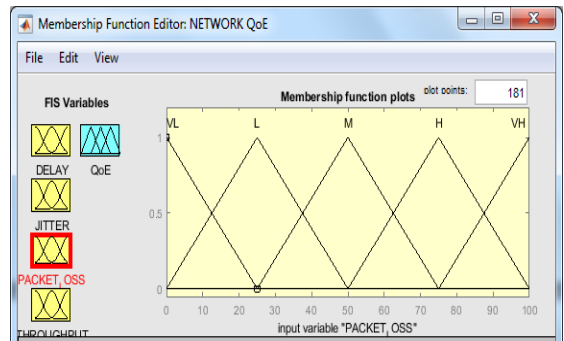

Figure4. 3: Developed membership function plots for Packetloss input linguistic term.
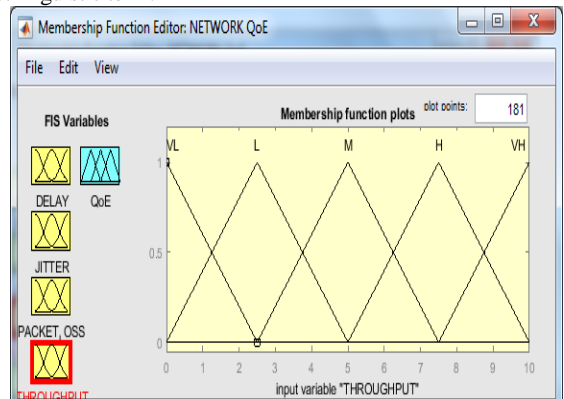

Figure4. 4: Developed membership function plots for Throughput input linguistic term.
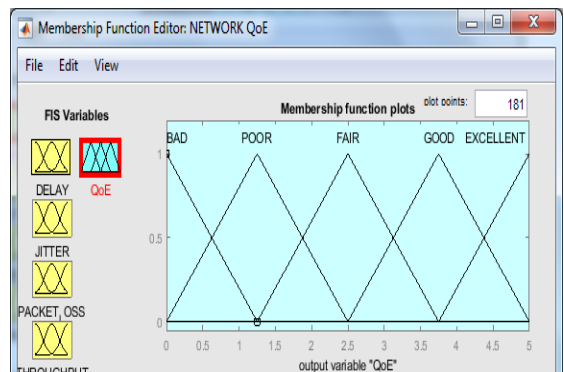

Figure4. 5: Developed membership function plots for different output linguistic terms.

Moreover, the initialization phase involves constructing the rule base. The identified Five linguistic terms for use i.e. "Excellent", "Good", "Fair", "Poor", and "bad" inclusive of the Four variables for network integrity of service QoE parameters i.e. Throughput, Delay, Delay variation/Jitter and packet loss results into 625 rules(5^4).

The rules were further dropped to 240 rules basing on expert knowledge by discarding the illogical rules thus remaining with logical rules to make rational decisions.

The illogical is as a result whereby some conditions cannot

exist at the same time for instance in rule 1 of the 625 rules indicates:

> 1. If delay is very low, jitter is very low, packet loss is very low and throughput is very low then User Satisfaction N/A.

This rule is N/A thus illogical since when delay, jitter and packet loss are very low then throughput is supposed to be high or very high in ideal network situation as these three variables which are supposed to make the throughput very low, their existence too are very low not to certain levels to affect the network throughput to match being very low.

Fuzzification:

This phase involves converting the crisp input data to fuzzy values using the membership functions. This is achieved by Fuzzifier component of the fuzzy Logic System.

Firstly, a crisp set (subset elements of the set, definitely do belong to the set), of input data are gathered and converted to a fuzzy set (sets whose elements have degrees of membership) by using fuzzy linguistic variables, fuzzy linguistic terms and membership functions. This step is known as **fuzzification** [6].

Inference:

This stage involves evaluating the rules in the rule base. Each rule follows the order to fulfill certain condition. The logical 240 rules are interpreted one after the other. This is achieved by Fuzzy Inference system component of the Fuzzy Logic System. In this work, Mamdani fuzzy inference system is used to achieve the inferencing in the developed framework. The Fuzzy set operator "AND" is used to aggregate the output of each rule.

The results of each rule are combined at this phase. The matched fuzzy rules are then used in the defuzzification process.

The logical operator "AND" is selected for connecting the inputs in this experiment since the operator returns logical 0 (false) if even a single condition in the expression is false in an ideal situation [27].For instance in the rule below:

Example 1:

If delay is very low, jitter is very low, packet loss is very low and throughput is very high then User Satisfaction EXCELLENT:

In an ideal situation, when delay, jitter and packet loss are very low then throughput is very high as the network suffers no hitches thus resulting to user satisfaction being excellent.

Defuzzification of the Output:

The linguistic variables and terms are matched, fuzzy rules generated and output results obtained for each parameter are aggregated into one crisp value through **defuzzification**.

This process involves producing a quantifiable result in Crisp logic, given fuzzy sets and corresponding membership degrees.

Moreover this process maps a fuzzy set to a crisp set. It is typically needed in fuzzy control systems. These will have a number of rules that transform a number of variables into a fuzzy result, that is, the result is described in terms of membership in fuzzy sets [28]

After the inference step, the overall result is a fuzzy value. This result should be defuzzified to obtain a final crisp output. This is the purpose of the defuzzifier component of a FLS.

Weighted average method technique is applied in this work because of its computational efficiency. It's formed by weighting each function in the output by its respective maximum membership value.

V. RECOMMENDATION.

Based on the review, it is greatly recommended to adopt Frameworks that have capabilities to accept vague and subjective values for analysis and decision making based on certain concepts or methodology for instance Fuzzy logic as user satisfaction is subjective in nature.

Moreover, in order to assess QoS of any network efficiently, network and service related performance metrics should be identified carefully.

In summary, there is a necessity to include all the four parameters which are linked to the integrity of service since

they are considered to be the primary factors affecting any computer networks [10].

BIBLIOGRAPHY

[1]   Y. Chen, T. Farley and N. Ye, "QoS Requirements of Network Applications on the Internet.," Information ● Knowledge ● Systems Management Vol4, p. 55–76, 2004.

[2]   K. Brunnstrom, S. A. Beker, K. D. Moor, A. Dooms, S. Egger and e. al., "Qualinet White Paper on Definitions of Quality of Experience," in Qualinet White Paper on Definitions of Quality of Experience Output from the fifth Qualinet meeting, Novi Sad,, 2013.

[3]   ITU-T, "Methods for Subjective determination of transmission quality," p.800, 1996.

[4]   M. Mehrdad and N. Abbas, "Supplier Performance Evaluation Based On Fuzzy Logic.," International Journal of Applied Science and Technology.1(5), pp. 257-265, 2011.

[5]   U. Ezutah and Y. Kuan, "Supply Chain Performance Evaluation and challenges," American Journal of Engineering and Applied Sciences, pp. 202-211, 2010.

[6]   L. Zadeh, "Fuzzy Sets," Information and Control Vol 8, pp. 338-353, 1965.

[7]   S. Shruti and R. Mudholkar R, "Evaluation of teacher 's performance using Fuzzy Logic Techniques.," International Journal of Soft Computing and Engineering, 3(2), pp. 306-320, 2013.

[8]   Zadeh, "A Computational approach to fuzzy quantifiers in natural language.," Computers and Mathematics with Applications 9(1), pp. 149-84, 1983.

[9]   M. Mehrdad and N. A. Abbas, "Supplier Performance Evaluation Based On Fuzzy Logic.," International Journal of Applied Science and Technology.1(5), pp. 257 -265., 2011.

[10]   F. Farid, S. Shahrestani and C. Ruan, "A Fuzzy Logic Approach for Quality of Service Quantification in Wireless and Mobile Networks.," in 10th IEEE International Workshop on Performance and Management of Wireless and Mobile Networks., Edmonton, Canada, 2014.

[11]   F. Farnaz, S. Seyed and C. Ruan, "A Fuzzy Logic Approach for Quality of Service Quantification in Wireless and Mobile Networks.," in 10th IEEE International Workshop on Performance and Management of Wireless and Mobile Networks, Edmonton, Canada, 2014.

[12]   J. Shaikh, M. Fiedler and D. Collange, " Quality of Experience from user and network perspectives.,"

https://doi.org/10.1007/s12243-009-0142-x, no. December 2009, p. 47–57, 2010.

[13]   R. Hawi, G. Okeyo and M. Kimwele, "Techniques for Smart Traffic Control : An In-depth Review," Techniques for Smart Traffic Control : An In-depth Review. 4(7), p. 566–573., 2015.

[14]   A. Hamam, M. Eid, A. E. Saddik and N. D. Georganas, " A Fuzzy Logic System for Evaluating Quality of Experience of Haptic-Based Applications.," A Fuzzy Logic System for Evaluating Quality of Experience of Haptic-Based Applications. (October 2016). https://doi.org/10.1007/978-3-540-69057-3, pp. 1-9, 2008.

[15]   J. Pokhrel, ". Intelligent quality of experience (QoE) analysis of network served multimedia and webcontents.," Analyse intelligente de la qualité d ' expérience (QoE) dans les réseaux de diffusion de contenu Web et Multimédia., pp. 49-155, 2015.

[16]   S. M. Ataeian and M. J. Darbandi, "Analysis of Quality of Experience by applying Fuzzy logic A study on response time," no. June, 2011.

[17]   O. F. W. Onifade, "Better Quality of Service Management With Fuzzy Logic In Mobile Adhoc Network," vol. 6, no. 1, pp. 59-68, 2013.

[18]   M. E. A. Ebrahim and H. A. Hefny, "Fuzzy Logic based Approach for VoIP Quality Maintaining," vol. 9, no. 1, pp. 537-542, 2018.

[19]   P. Jeevan, "Intelligent quality of experience (QoE) analysis of network served multimedia and webcontents.," Analyse intelligente de la qualité d ' expérience (QoE) dans les réseaux de diffusion de contenu Web et Multimédia., pp. 49-155, 2015.

[20]   B. S. De-lima and N. F. Ebecken, "A comparison of models for uncertainty analysis by the finite element method, Finite Elements in Analysis and Design,Vol 34,no.2," in Finite Elements in Analysis and Design, Las Vegas, USA, ScienceDirect, 2000, pp. 211-232.

[21]   J. Pokhrel, F. Lalanne, A. Cavalli and W. Mallouli, "QoE estimation for web service selection using a Fuzzy-Rough hybrid expert system".

[22]   L. Zadeh, "Some reflections on the anniversary of Fuzzy sets and systems.," Journal of Fuzzy sets and systems, 100(2), pp. 1-3, 1999.

[23]   A. S. Omar, M. Waweru and R. Rimiru, "Fuzzy Logic and Qualitative Performance Evaluation of Supply Chain Management," ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805, pp. 56-63, 2015.

[24]   M. Fiedler, "A Generic Quantitative Relationship between Quality of Experience and Quality of Service," no. April, pp. 36-41, 2010.

[25]   F. Farid, S. Shahrestani and C. Ruan, "A Fuzzy Logic Approach for Quality of Service Quantification in Wireless and Mobile Networks," pp. 629-636, 2014.

[26]   E. Herbert, K. Thomas and M. Manfred, "A 12b General Purpose fuzzy logic controller chip," IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL 4, NO 4, pp. 460-475, 1996.

[27]   mathworks, "www.mathworks.com/help/matlab/ref/logicaloperatorsshortci rcuit.html?searchHighlight=logical%20operators&s_tid=doc_ srchtitle," 19 3 2019. [Online]. Available: https://www.mathworks.com/help/matlab/ref/logicaloperators shortcircuit.html?searchHighlight=logical%20operators&s_tid =doc_srchtitle.

[28]   W. V. Leekwijck and E. Kerre, "Fuzzy Sets and Systems," Defuzzification: criteria and classification., pp. 159 - 178, 1999.

Density Based Traffic Signalling System using Image Processing

Adwait Sharma
Computer Department
St.John College of
Engineering
Mumbai, India

Aayush Soni
Computer Department
St.John College of
Engineering
Mumbai, India

Rahul Rajpurohit
Computer Department
St.John College of
Engineering
Mumbai, India

Angelin Florence
Computer Department
St.John College of
Engineering
Mumbai, India

ABSTRACT

The number of vehicles is increasing rapidly in urban areas and metropolitan cities with in crease in population. To manage such a huge amount of traffic in an efficient way is the need of the hour. The conventional traffic signalling system are not adaptable to change in the amount of traffic. This paper solves the problem by making traffic signals smart enough to distinguish between the lanes which have different densities of traffic present in them. This paper makes use of image processing with the help of stored data set of respective lane images and a microprocessor to compute an output.

Keywords

Raspberry Pi, Canny Edge Detection, OpenCV, Image Processing, Python.

1. INTRODUCTION

This paper focuses on the solution to the problem generated due to increase in the traffic volume in urban areas. This paper makes traffic signalling smart and efficient for urban areas and metropolitan cities. The complete project is divided into five modules (i.e. Image acquisition from data-set, Image pre-processing, Image processing, Decision making, Output). The data-set image will be sourced from the storage and converted to greyscale image, later this image will be processed by the microprocessor (Raspberry Pi) with the help of OpenCV, a library of Python for Image Processing. The project will make use of binary images captured at real-time and a reference image which will be stored in the system. Thereafter, capturing the image the comparison will be made between the two binary images (i.e. captured image and reference image). The similar comparison will be performed on each lane and then the suitable output will be obtained from the process. If two or more lanes have same amount of traffic then the conventional time interval method will be used to clear the traffic on those lanes.

Different techniques had been proposed in the past such as infra-red sensor, induction loop etc. to control traffic but these techniques had their drawbacks. In recent years, image processing has shown excellent outcomes in acquiring real time traffic information using camera module installed along the traffic light.

2. LITERATURE SURVEY

Prof. Vikram Deshmukh, Shruti Pantawane, Sonali Hajare, Anushree Kale proposed a system in which the conventional traffic light system has a fixed time pattern which doesn't vary as per the current traffic situation at the junction. This project helped sorting the problem by capturing and converting the lanes images into greyscale images. Threshold value is calculated which will indicate the number of vehicles present at the particular lane. The project uses Image Processing using MATLAB software and microcontroller AT89S52.
The advantages are the use of microcontroller preventing the error within machine and make things conditional and smart.

The next important advantage is use of MATLAB which is mobile and efficient. The drawbacks that were encountered were the cost of implementation regarding the hardware, use of interpreted language MATLAB which is slow and the chances of microcontroller getting crashed due to excessive load.

D. Prakash, B. Sandhya Devi, R. Naveen Kumar, S. Thiyagarajan, P. Shabarinath proposed a system in which the number of vehicles was counted by the microcontroller. The microcontroller takes decision based on different vehicles count and updates the traffic light delays as a result. The processor used is the ARM architecture. The main advantage of the system is the information like number of vehicles on each lane will be known because of the sensors. The drawbacks that can be marked were the expensive sensors which incur lot of maintenance and the system is scalable.

K. Vidhya, A. Bazila Banu proposed a system, in which the image captured in the traffic signal is processed and converted into greyscale followed by the calculation of threshold value indicating the number of vehicles present on that particular road. After this calculation, it came with the result to understand in which side the density is high. Raspberry Pi is used as a microcontroller for controlling the signals. The main advantages that are evident is firstly the use of OpenCV, secondly the usage of fast and low RAM usage. Also, the hardware level is feasible because use of Raspberry Pi. The disadvantages that can be figured out are that the OpenCV is not that flexible as compared to MATLAB. Also, the Raspberry Pi has a limited amount of memory.

3. EXISTING SYSTEM

In the existing system, the conventional traffic system the green light time is assigned to every lane of the road which is pre-defined defined and fixed. It doesn't vary as per the amount of traffic present on the road at that time. Priority factor is missed and the signals are meant for just controlling the assumed amount of traffic rather than computing the real-time traffic scenario.

It is not an efficient way to handle the varying amount of traffic present on the road in metropolitan, smart cities. Equal green light time or predefined amount of time can lead to more congestion on the busy lanes whereas the lanes with the less amount of traffic are given excess green light time which goes in vain. This also leads to frustration among the drivers and the rush leads to accidents sometimes.

This existing system does not even consider the fact that the number of vehicles is directly proportional to congestion of roads where the priority factor is missing and not considered.

4. PROPOSED SYSTEM

The proposed system aims to make the traffic signal more efficient and feasible. Therefore, reducing traffic as well as utilizing the time.

The proposed system will be density based that means it will give priority to the lane which has comparatively a greater number of vehicles. Image processing is the method that will be used for calculating the density. Raspberry Pi will be used as a processor. Images will be acquired using the data set. Acquired images will go through the steps of image processing. Now, the differences will be compared and priorities will be assigned.

The project can be divided into the following modules:

4.1 Image Acquisition from Data-Set

An image is taken from the data set of images of road having various amount of vehicle density. Images from the data-set demonstrate the real-time images of traffic.

Figure 1: Input acquired from data set

4.2 Image Pre-Processing

Pre-processing is basically used for removing unwanted noise and objects from image. The following are the steps that are to be taken in this phase:

Step 1: Greyscale Conversion is the first step in which the RGB image is converted into greyscale image for improving performance. The grey colour image consists of pixel intensities between 0-255 where 0 represents black pixel and 255 represents white pixel.

Figure 2: Greyscale Image

Step 2: Thresholding converts the greyscale image into the binary image. If pixel value is greater than a threshold value it is assigned one value (white), else it is assigned the other one. (black).

Figure 3: Image after pre-processing

4.3 Image Processing

The main purpose of this phase is to serve the processing part on the pre-processed image. The following steps are taken:

Edge detection focuses on the discontinuities in an image. All the edges within the image are detected using Canny Edge Detection algorithm. It provides:

    a. Noise Reduction
    b. Finding Intensity Gradient

Figure 4: Edge detected image

4.4 Counting the Non-Zero pixels:

This step includes counting the number of non-zero pixels i.e. the white pixel count of an acquired images of four lanes from the data-set and the reference image of respective image.

4.5 Decision Making:

Here, the value of difference is sorted and priorities are assigned to each lane as per the difference count obtained in the previous step.

The table below tabulates an example of such priority distribution :

Lane	Difference	Priority	Green Signal Period
Lane 1	103405	1	60 seconds
Lane 2	32391	4	15 seconds
Lane 3	47975	3	30 seconds
Lane 4	67484	2	45 seconds

5. SYSTEM ARCHITECTURE

5.1 Raspberry Pi

The system uses raspberry Pi 3B+ as it is small, portable and powerful enough to perform all the necessary tasks that are Image Processing, decision making, etc. Raspberry pi 3B+ has a 1.4GHz 64-bit quad-core processor, dual-band wireless LAN, Bluetooth 4.2/BLE in it.
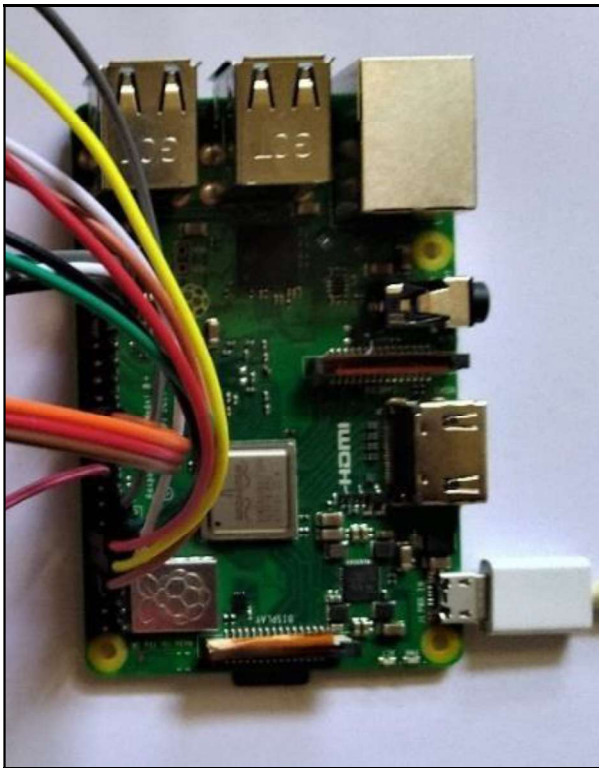
Figure8: Raspberry P3B+

5.2 Traffic Lights

LED lights are being used in this system for traffic signal because it consumes less power, have increased light output, last significantly longer, and in the event of an individual LED failure, still operate albeit with a reduced light output. Figure shows the four set of three LEDs which represent four lanes.
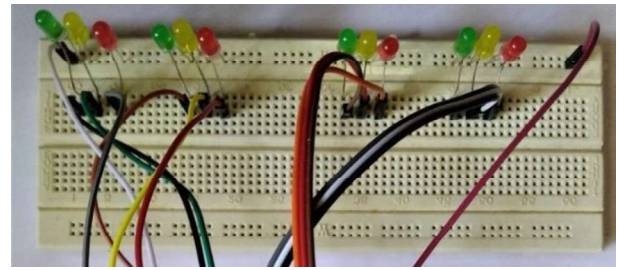

Figure 9: LEDs representing four lanes

5.3 Power Supply

Raspberry Pi uses a special 5V DC supply.

6. RESULT

When the first set of four images of four lanes at a junction point are acquired from the data set (Viz shown in Figure 6.1), it can be seen that Lane Number 1 has the highest priority as it has highest number of vehicles present on it. The second priority is given to Lane number 4 and so on priority 3 to Lane Number 3 and priority 4 to Lane Number 2.
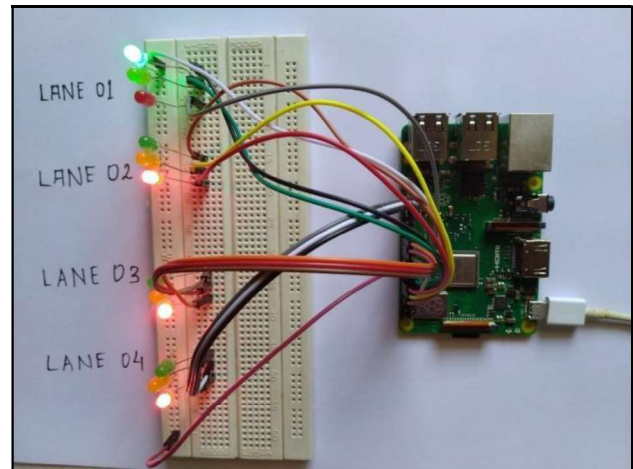

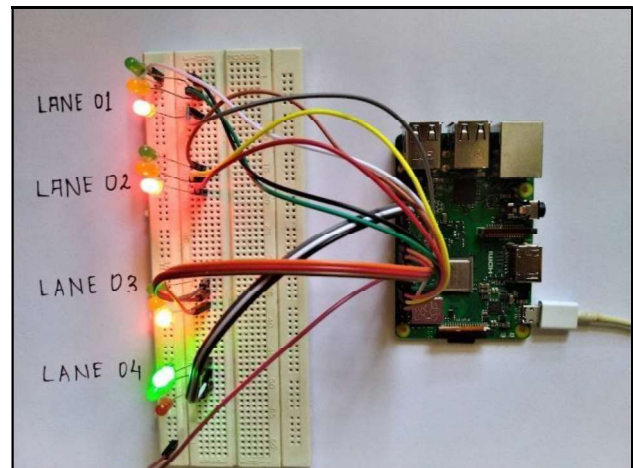**Figure 10: Lane Number 1 Green For 45sec. (Priority 1)**

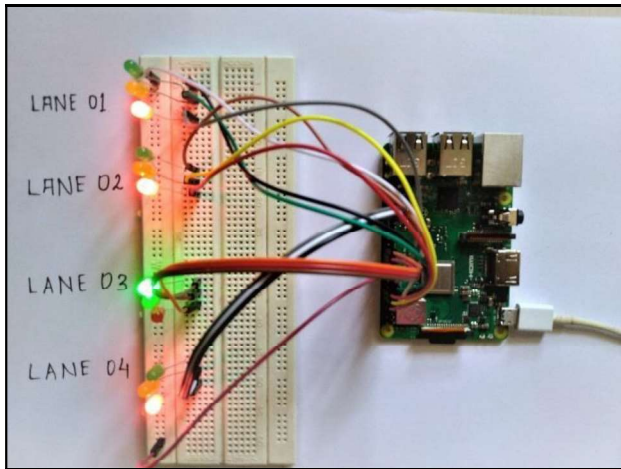**Figure 11: Lane Number 4 Green for 25sec. (Priority 2)**

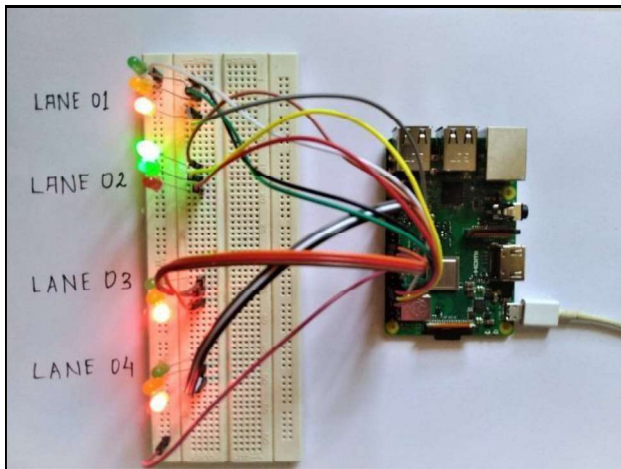**Figure 12: Lane Number 3 Green for 15sec. (Priority 3)**

Figure 13: Lane Number 2 Green for 10sec. (Priority 4)

Later after all the four lanes at the junction are given a green light time, all the signals are turned red for pedestrians to cross the road.
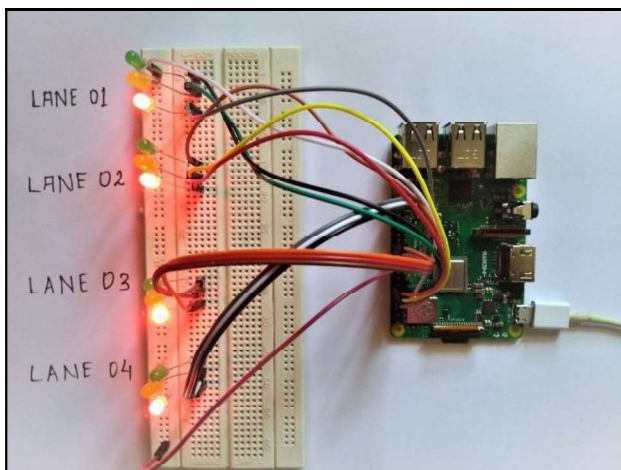

Figure 14: All signals are made red for pedestrians.

Later another set of images are acquired from the data set and the next iteration is continued.

7. DISCUSSION

The existing systems makes use of high cost sensors and/or hardware components that require high maintenance and high budget. whereas the proposed system is cost efficient and stand alone as it does not rely on any special hardware and requires less maintenance as compared to existing system.

8. CONCLUSION

Density Based Traffic Signalling System provides a more efficient replacement to the current conventional traffic system. The system uses Image processing to compute the densities and accordingly provide the green light signal to the lanes. The system is an advantage to the denser traffic lanes providing ease of movement, avoiding congestion, also the pedestrians can safely cross the roads and the system is available 24/7. The system is a standalone design and meant to maintain its computation as long as it is connected to the power supply. Also, the design of the system considers the cost factor so as to provide efficient system in way less cost. Using this method at each crossway could help in a continuous journey of the people. This project focuses on making the standard traffic signal density based. This will make the traffic signal more feasible by giving green light time to the lane which has more density.

9. FUTURE WORK

Future work can include the feature where we can monitor the signals and the status can be updated at the server. This will help in future reference for generating patterns in the traffic congestion in terms of specific days, holidays, time, etc. It can also be used to prioritize emergency vehicles distinguished by a siren at the top. This may provide an extra priority factor to ambulances, fire brigades, etc.

10. REFERENCES

[1]   Ashwini D. Bharade, Surabhi S. Gaopande, Robust and Adaptive Traffic Surveillance System for Urban Intersections on Embedded Platform, 2014 Annual IEEE India Conference
(INDICON)

[2]   Cyrel O.Manlises, Jesus M. Martinez Jr. , Jackson L. Belenzo,Czarleine K. Perez,Maria Khristina Theresa A. Postrero, Real-Time Integrated CCTV Using Face and Pedestrian Detection Image Processing Algorithm For Automatic Traffic Light Transitions, 8thIEEE International Conference Humanoid, Nanotechnology, Information Technology Communication and Control, Environment and Management (HNICEM) The Institute of Electrical and Electronics Engineers Inc. (IEEE) – Philippine Section

[3]   Heba A. Kurdi, Review of Closed Circuit Television (CCTV) Techniques for Vehicles Traffic