International Journal of Engineering Research and General Science Volume 6, Issue 1, January-February, 2018
ISSN 2091-2730

ODM: Modelling Optimization of Datamining in e-Governnance Framework

Sangeetha G
Research Scholar
Computer Science, Bharathiar University
Coimbatore, India
E-Mail: sangithagovind@gmail.com

Dr. L. Manjunatha Rao
Professor and Head
Department of MCA
Dr.AIT
Bangalore,India

Dr.Hanumanthappa.M
Professor and Head
Department of Computer Science and Applications
Bangalore University

Abstract— Implications on datamining techniques has been more than a decade old in area of e-governance. However, at present there are various tradeoffs in the existing research work with the evolving modernization of the technologies. Performing optimization of teh data mining techniques in presence of heterogeneous data to has never being explored in past. The paper also showcase that performing datamining technique cost effectively on distributed storage system is computationally complex phenomenon. Therefore, this paper presents a model called as ODM or Optimized Datamining in e-governance, which put a special emphasis of the modern technique of storing and accessing the data in distributed storage system e.g. cloud. Adopting empirical and experimental technique, a prototype model is experiment, which shows that ODM is highly cost-effective datamining technique on existing as well futuristic distributed data storage system.
Keywords-component; e-Government, datamining, knowledge discovery, Data warehousing

INTRODUCTION

The area of e-governance is slowly gaining the pace owing to the potential advantages it renders to the citizens of the nation [1]. Such applications usually call for using ICT (Information & Communication Technology) by various government bodies for delivering the public services. It is also termed as a method of disseminating the significant information to the public using various electronic mediums. Adoption of e-governance is of significant interest to various individuals that also includes researchers owing to following merit factors viz. i) enhanced efficacies on various processes run by government, ii) allows transparency as well as minimize corruption in various government transactions, and iii) motivate the citizen for participating in the governmental processes. In a nutshell, e-governance tends to convert conventional government organization to e-organization for achieving e.g. i) better and transparent communication, ii) well structured processing of information, iii) precise ability for taking decision, iv) better flexibility on government work processes, v) efficient coordination on various citizen and government activities, vi) normalizing the complex business association, and vii) generation a pleasant platform of customer relationship management. The applications of e-governance; however is in very infancy stage and bears various tradeoffs with the existing modernization of the technological advancements [2]. There is an increasing inclination towards the implementation of various IT resources to be integrated on existing portals that are offering the public services. At present, the users are more interested to use the public services sitting at the comfort of their house or in roaming using mobile computing and mobile networks [3]. The emergence of cloud computing also offers ubiquitousness of data availability and really acted as a boon for the massive data storage. There are already various sort of public cloud applications that run on various smartphones [4]. Unfortunately, there is no such application or services in cloud that offers the public services on mobile phone as dedicated applications, which is the prime trade-off. Some of the notable obstructions towards successful implementation of e-governance framework are i) lack of online work execution and support to various government departments, ii) extracting the useful data, iii) aggregation of massive data, iv) lack of technical capabilities and its respect support to government run activities, and v) summarization and processing of useful as well as meaningful data for enabling faster decision during emergency situation or some high level complex situation in nation.

At present, various forms of organization are evaluating the existing as well as old data for recognizing the meaning patterns from the massive dataset [5]. This is done for understanding the knowledge required for confirming complex business strategies. The prime

attention is always given on interactive charecteristics of data, complexity involved in data processing, and exploratory evaluation of the massive and heterogeneous data generated from various e-governance framework. In order to solve such issues, at present, the research community is more or less working on i) datamining techniques [6], ii) data warehousing techniques [7], and iii) OLAP (Online Analytical Processing) [8]. From the last 10 years, majority of the research work are found to adopt these techniques for extracting the meaningful information to be used for e-governance model. Still, there is no standard work on datamining optimization being witnessed. This paper reviews some of the significant research work and discusses about the problems being encountered in performing datamining by introducing a novel model. Section 2 discusses about the existing literatures followed by problem identification of them in Section 3. Section 4 discusses about the proposed model followed by research methodology in Section 5. Section 6 discusses about the algorithm implementation followed by result discussion in Section 7. Finally, Section 8 summarizes the paper.

RELATED WORK

Our prior study [9] has reviewed some of the existing datamining techniques for enhancing the capabilities of e-government framework. Our study has investigated various forms of e-government frameworks on different countries and studied their applicability. The work done by Rao [10] have also significantly discussed about the software architectures of e-governance with good focus on existing theoretical frameworks. The study has also investigated about the future trends of e-government frameworks. The same author has proposed an architecture that ensures optimal security of e-government framework [11].

Rao and Dey [12] have presented a text-mining method of performing mining of the data. The authors have also studied various corpus processing using local languages of India. Study on e-governance application was also carried out by Trivedi and Dutta [13], where the authors have put higher emphasis on the adoption of datamining techniques. Desai [14] have presented a prediction based technique using datamining algorithm on e-governance considering case study of birth registration problem. The outcome of the study has claimed that the technique could be possibly used in any higher management level for planning and formulating various critical decisions. Same author has also presented a different case study of vehicle registration under e-governance framework using both mining and warehousing approach [15]. The author has adopted the deployment of clustering technique as well as conventional association rule mining technique to extract knowledge. The outcome of the study was claimed to be helpful for assisting the automobile firms to maximize their sales targets. Suresh and Mahale [16] have presented an efficient technique of datamining for enhancing the education system in e-governance. The authors have adopted decision support system to formulate the design of analytics using warehousing approach. Rajput et al. [17] has used WEKA (a conventional open source datamining tool) for exploring the best classification protocols as well as for generation of decision tree. The author has adopted qualitative methodology to collect the real-time data and process it using WEKA tool. The study was also found with adoption of j48 classifier for the purpose of training the data. JRIP rule classifier algorithm was also adopted for in-depth analysis of the data. Shen et al. [18] have presented a data extraction framework that integrates meta-search engine with content mining. Mampilli and Meenakumari [19] have also discussed the issues and scope of e-governance framework. Bhanderi [20] have exclusively investigated OLAP, data warehousing, and data mining. Garg et al. [21] has also emphasized on the implementation aspect of e-governance in education system of India. Bidgoli and Akhondzadeh [22] have presented an approach of datamining on e-government framework. The authors have adopted association rule mining on the samples of the data being generated from customer's complaints. The study has also emphasized the clustering techniques and how it can assists in performing analysis of data being generated from any e-governance framework. However, benchmarking is missing in all the studies discussed till date.

PROBLEM IDENTIFICATION

Performing error-free analysis of the data being generated in massive order has always been the great concern in distributed storage system. The existing service providers for distributed storage system are not found to provide the optimal extraction of knowledge as well as good Quality-of-Service which is originally promised to the user. This section will discuss about the problems that have been identified as follows:

- **Massive Data Generation**: Majority of the existing system is found to be carried out using a smaller size of dataset. Sometimes the datasets are downloaded from internet resources or else it is collected manually. Availability of such size of data becomes a major tradeoff as in existing system, majority of the application has become pervasive, for which purpose, user can access and share the data from multiple terminals. This phenomenon is also application on futuristic e-government application that will generate around petabytes of data in a matter of minutes. Hence, it is quite a challenging situation to perform datamining techniques on such complex and unstructured data in less period of time.

- **Conventional Storage System**: Majority of the existing e-governance application used in India and also some other countries uses a single server storage system. Adoption of such storage system frequently crashes when the number of users increases, which is unpredictable. However, at present, there is also a proliferation of distributed storage system e.g. cloud, where the data is stored in distributive manner for ensuring cost-effective storage. However, this principle may sound good for same type of data but it poses a significant problem, when the data are of heterogeneous types. Hence, if such data could be stored in multiple locations by splitting, it becomes quite a challenging task to ensure the retrieval process when the user wants to access. Moreover, performing datamining over distributed storage will lead to anomalies and also result in maximum time consumption if the data are higher heterogeneous.
- **Conventional Mining Approach**: Applying the concept of mining over distributed dataset, there is a significant correlation between local and global Frequent Itemsets (FI). However, the problem of uncertainty arises when global FI is not able to confirm about its local FI. In order to solve such issues, majority of the existing system attempts to extract global FI by searching local FI in every repository. Such actions calls for higher expenditure of functional cost as well as transmission cost too. On the other hand, in absence of proper indexing mechanism of distributed data, it is almost absurd to analyze global FI without initially considering local FI.

Therefore, the problem statement can be given as – "*It is a computationally challenging task to optimize a datamining technique on distributed storage system retaining Quality-of-Service*" The proposed system mainly attempts to investigate the solution towards this problem statement and has evolved up with a technique that can ensure adherence to the maximum standards of the data storage much complexities being involved in it.

PROPOSED MODEL

The prime purpose of the proposed system is to design an optimized model of datamining that can be incorporated over e-government framework. Our prior study [23] has introduced a e-governance framework for public redressal data using a typical text-mining approach. With the increasing usage of the e-governance application, it is expected that it will generate a massive set of heterogeneous data, which poses a problem of both data heterogeneity as well as scalability. Hence, the present work introduces a model termed as ODM i.e. Optimized Datamining Model, which is designed for addressing the issues of heterogeneity and scalability of data in existing e-governance framework.

The prime contribution of the research work to formulate a multi-storage and highly distributed datamining architecture which can be developed to maintain the data accessibility as well as knowledge discovery of various distributed data storage applications (e.g. cloud) and to authenticate the users from performing any types of illegal activity. The objectives of the proposed system are e.g. i) *Validation of data*: The system should allow online users to get authenticated and validated by the e-government interface, ii) *Distributed Data Storage system*: The system should be able to get the user's data and be able to perform splitting of the data to chunks and store it randomly in distributed storage servers, iii) *Data indexing Mechanism*: the system will provide a indexing of distributed data for addressing the problem of data heterogeneity and retain scalability for the uploaded data. The system will provide distributed indexing mechanism for performing encoding and decoding of the uploaded and downloaded data respectively. The prime contributions of the proposed framework are as follows:

- To design a universal *e-Governance Interface* that interfaces global users with existing data mining cores.
- To develop a *File Management Server* that can assist in data management system and simultaneously acts as a bridge between the user and the cloud systems.
- To permit the system to perform *Data Ripping Process* that can split the user's data and reposit it in multiple and anonymous distributed storage.
- To develop a *Data Encoding and Decoding* mechanism using hash function that can perform encoding of the ripped data of user and reposit it arbitrarily over the distributed storage system.
- To design an *Optimized Datamining* technique that can perform knowledge discovery on larger and highly distributed samples of data.

The proposed system highlights an efficient distributed storage management for repositing the data generated from e-government application which are uploaded by the users in the pervasive environment. Efficient modelling of datamining architecture in our proposed system defines a multi storage structure where an authenticated user only can upload the data in a distributed storage system and keep it safe for his future requirement. As now a day's huge amount of data requires lots of space and that requested amount of space cannot be provided by any of a single Personal Computer. Moreover, the computing resources also have very limited amount of processing power which would be responsible for disk overhead. The uniqueness of this proposed optimization model is that it uses a very distinctive encoding and decoding mechanism with data indexing technique to address the issues of data heterogeneity and

scalability in existing e-governance framework. The application of proposed system can also enable the user to share a large amount of information that is also easy to archive, access, and perform knowledge discovery. As there are various proposed datamining and implication of association rules for knowledge discovery process has been implemented till date but it has been observed, there are very less attempted techniques to perform datamining on distributed database and on multi-storage system. Hence, we achieve it by using our optimization model.

RESEARCH METHODOLOGY

The proposed study consider mix-mode standard for research methodology by considering both empirical and experimental approach. Empirical approach [24] was deployed as the implementation of datamining technique usually follows a standard scientific and mathematical process. Experimental [25] approach was considered as the proposed model was required to be analyzed and tested in real-time for understanding the effectiveness of its outcome. The system architecture of the proposed model is shown in Fig.3. The design principle of the proposed model is mainly deputed to two significant user viz. i) user module and ii) supervisor module. The user seeks to exercise the system for archival of their data. The data to be reposited on distributed storage system could be of any forms. The supervisor module is responsible for the carrying out the file management system for incoming data. In order to clarify the operation of the proposed model, we discuss the core modules used for optimizing our datamining technique:

- *E-Governance Interface* (GI): This module generates a universal interface designed on Java environment for maximum supportability of users owing to its platform independence features. It also permits all the essential users to upload the raw data as well as access their information and see the outcomes of knowledge discovery process. Fig.1 shows the schema of e-Government interface that unite the online users with the distributed storage system.

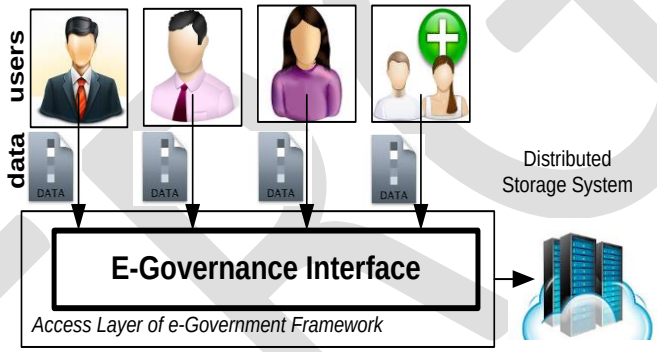

Figure 1 Representation of E-Governance Interface

The e-Governance interface permits the user to undergo a secure validation policy by checking the user ID and their respective passwords. Once the validation of the user records are accomplished than the system will permit the accessibility of the priviledge account of the user by allowing uploading their data on distributed storage. It is to be noted that e-Governance Interface runs on the access layer of the standard e-government framework. However, for incorporating the features of optimization, it is essential that every modules of ODM are highly connected and dependent with respect to core attributes. Hence, e-governance interface only allows users to upload their data, but in order to give them access to their stored data as well as knowledge (after performing our datamining technique), this module is highly dependent on next module called as File Management Server.

- *File Management Servers*: This module add more value to the data storage system when the data is of high volume and have higher degree of heterogeneity. The ODM introduces a system called as File Management Servers that runs on storage servers to create its own file types and reposit the multiple user's data in a typical order. The functionality is pictorially shown in Fig.2. The File Management Server is responsible for the self-governing data management system for multiple users on a given distributed storage system. The File Management Server is also required to perform an indexing of the user's data to normalize the heterogeneity of incoming data. Once the file is processed through this module, ODM will further perform two simultaneous operations e.g. i) performing encoding on the user's data and ii) transmitting the indexed data to the different storage systems. As direct data transmission is complex owing to large number of distributed servers in many geographical regions, hence, the system chooses to take the assistance of this module.

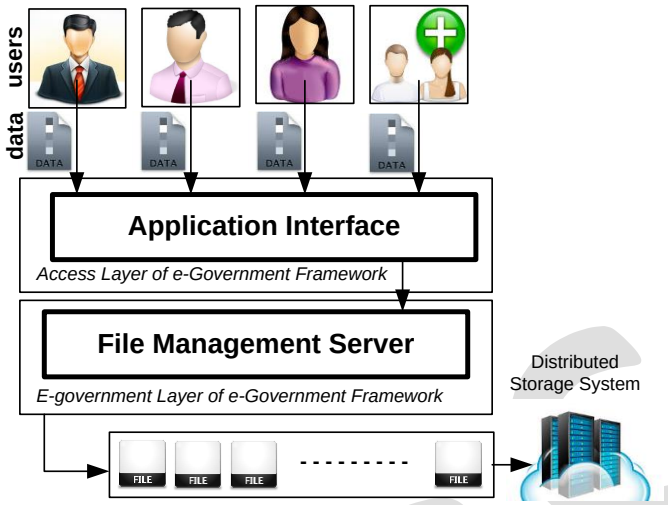

Figure 2 Schematics of Synchronous Servers

The outcome of the File Management Server module will be the indexed file with a specific file formats with a storage matrix with an ordering of the data uploaded or requested by the online users. After the data is processed by this module, it is subjected to encoding using lightweight hash function.

- *File Encoding and Decoding Process*: This module is mainly meant for performing encoding of the files that are being generated by the File Management Server. The prime objective of this module is to perform faster encoding and decoding process without occupying much memory. Interestingly, ODM doesn't directly let the indexed data generated from File Management Server for encoding. The indexed data from File Management Server is binarized to reduce the time and space complexity while applying datamining technique, which will lead to generation of an encoded key. The next step is to perform encoding of the binarized data with hash function. We choose to implement hash as i) it supports faster computation while performing the data indexing and ii) it is less vulnerable to majority of the lethal attacks on storage system. After carrying out the encoding using hash function, the system leads to the encoded binary data, which will be further subjected to Data Ripping Process.
- *Data Ripping Process*: This module is responsible for ripping the data to the different storage systems based on the availability of the memory. The core system connecting to the e-government interface also maintains a record for the distributed storage system with the availability of the memory on the cloud. The system then rips the indexed data of all the gross active users based on the number of the availability of the storage servers. This process of ripping the data will mean that if there are 20 petabytes of the processed data, the data will be splitted as 5-petabytes, 7 petabytes, 2 petabyte, 3 petabytes, and 3 petabyte, if there is availability of 5 storage servers for a particular instance of time. One of the uniqueness in the proposed system is that the generated key that is encoded with hash function is stored randomly on the designated storage servers that are found to be available on that time instance. The system also providers better service by storing the key in the network, which is completely unknown to the user as well as supervisor. Hence, the proposed system performs storage of the user data in highly distributed and secured manner.
- *Optimizing Datamining Technique*: The proposed system introduces an optimization technique that will be subjected for every data that is stored in distributed storage system. The optimization of the datamining technique is done by incorporating a novel sampling procedure for processing the indexed and encoded data stored in storage servers. In order to perform optimization, ODM initially extracts the arbitrary sample from the given dataset which is followed by extraction of minimized set of samples using proposed sampling approach. The system then filters the larger dimensions of data based on its volume and size by using entropy method. Finally, the system generates cumulative samples in storage where it is subjected to a technique of exploring the significant relationship among the variable of samples in distributed dataset. Interestingly, this module reduces the processing time by trimming less-significant knowledge from the samples and keeps on adding only strongly significant knowledge in every cycle of mining. It is mainly meant for optimizing the cost of datamining over distributed storage system.

IMPLEMENTATION

The implementation of the proposed system is carried out on 34 machines with multiple configurations. All the machines are of different operating system, processor speed, and memory with pre-installed JDK environment for running the application. choose to develop it in Java environment as it will be feasible to check the real-time evaluation of the proposed system and more over multiple machines of heterogeneous configurations can be checked at same time. The prototype development of the ODM has been carried out in highly distributed manner in MyEclipse. Out of 34 machines, 5 machines were considered for Supervisor access, ten machines for running the distributed storage server, and remaining 19 machines were for distributed users. All the users underwent a registration phase. A database has been set up in MySQL which maintains the user ID and password of the user and once the customer attempts to

access their account, the system uses standard hash function to secure the password. The implementation of the proposed study is graphically highlighted in Fig.3

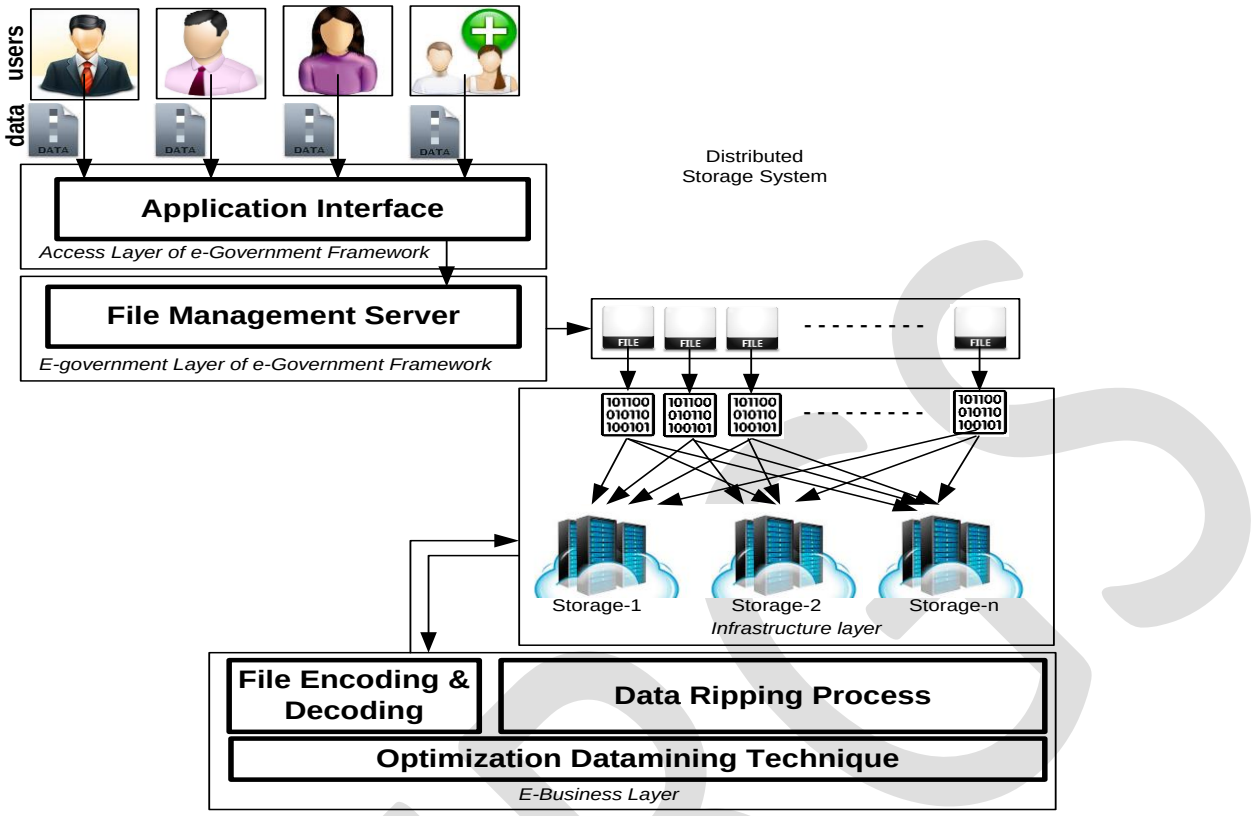

Fig.3 Schema Architecture of ODM for Implementation

The above algorithm takes the input of request of a user to store a specific set of data and checks the connectivity of the e-government interface with the storage servers (Line-1). After validating the user ID and initial password (Line-2), the system binarized the user data (Line-3). This task is also accompanied by generation of key (Line-3). The system applies hash function to further encode the generated key (Line 5-6). Therefore, the algorithm posses the capability to furnish hash encoding for both static as well as dynamic data that gives a wide supportability for the users to maintain their private data, keys, and most important, it offers ultimate data scalability. According to the algorithm, it stores the chunks of encoded data randomly on the available distributed storage server. The system also finds the availability of the storage of data based on the availability of memory (Line 7-9). Hence, ODM offers higher scalability as it is completely independent of any file types. The algorithm implementation can be seen in Fig.1 for the schematic architecture.

Algorithm for Optimized Data Mining

Input: Request of user for Storage

Output: i) Successful storage of Encrypted data, ii) Extracted knowledge

START

1. Initiate *db* connection from user to e-government interface.

2. Use MD5 to secure the password

3. Validate the user ID and password of users.

4. Binarized the incoming data→generate Key

5. Encode the data with key.

6. Apply Hash(key)→store(rand(server))

7. Estimate size (S) of indexed data

8. Size of data-remainder/cloud no.

9. Samples =Split_data (S_1) + S%Num_Storage_Server

10. Estimate minimum Support (S_P), confidence (C), Final Sample (S_F)

11. While ($|S_P|>n$) && ($|S_F|<n$) do

12. G(S_P, S_F)→disjoint(min(k, $|S_P-S_F|$)

13. If d(S_F+{t}, S_P)>d_{max} do

14. d_{max}=d(S_F+{t}, S_P), α=t

15. If d(S_P-{t},S_P)<d_{min} do

16. d_{min}=d(S_P-{t}, S_P), β=t

17. S_F=S_F+(α), S_P=S_P-β

END

Hence, the algorithm can encode any forms of data of user with less complexity of the management of key or index attributes. The decryption stage of the algorithm is just inverse of the encoding stage. Hence, it can be said that ODM resists even the supervisor, webmasters, root or any illegitimate member to the generated data samples (Line-9). After extracting the samples, the system estimates the minimum support (S_P), confidence (C), as well as quantity of the final samples (S_F) to generate the global rule for mining (Line-10). Considering if the initial samples bearing S_P are found more than total chunks of bitstreams and final samples are within the limit of size of bitstreams (where n represents total number of distributed storage) (Line-11), the system divides both the minimum support (S_P) and final sample (S_F) into number of disjoint group G (Line-12). Based on the two conditions in Line-13 and Line-15, ODM estimates the minimum and maximum distance (d_{min} and d_{max}). The estimation of distance d is done by subtracting support of final sample (S_F) to minimum support in initial level (S_P). Along with this, the system also monitors the values of transaction t (α for d_{max} and β for d_{min}) being done in each condition (Line-14 and Line-16). Therefore, ODM offers a cost-efficient datamining with lightweight indexing technique that can perform extraction of knowledge from distributed database and also furnishes a faster accessibility of data without any significant impact on the networking performance of distributed database.

RESULT ANALYSIS

The outcome of the proposed system ODM has been evaluated for its effectiveness with respect to transmission rate, encoding time, decoding time, and delay. All the outcomes were observed for test-data size of 1-10 GB. For better effectiveness of the study, the outcomes were also observed for existing studies of Mujawar [26] and Milic [27] who have also performed the similar type of research work most recently considering their own technique. Similar environments were selected for performing the data analysis.

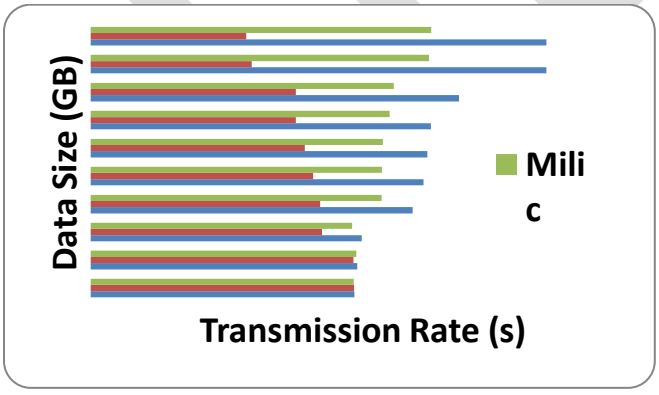

Figure 4 Analysis of Transmission Rate

The transmission rate were observed for test-data size of 1-10GB using network protocol analyzer called as WireShark. The outcome of Fig.4 shows that the proposed system is capable of sustaining the increasing load of the data request of the user. The curve for Mujawar [26] is found with lower peaks as it performs recursive fuzzy processing on every data request, whereas Milic [27] curve is

found with minimal peaks as compared to Mujawar [26] owing to less complicated structure using association rule mining. The proposed system adopts the process of ripping the data where the number of generated keys are highly dependent on the ripped data. Hence, the rate of transmission for proposed system can process higher datasets in shortest ranges of time.

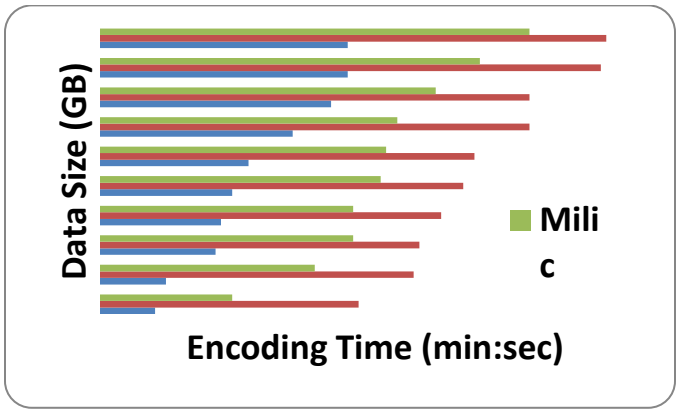

Figure 5 Analysis of Encoding Time

Fig.5 highlights the analysis of encoding time with respect to the increasing data size. It is quite obvious that after performing encoding, the size of the encoded data do increase in size to some extent which also affects an encoding time. The outcome shows that the encoding time for conventional scheme e.g. Mujawar [26] and Milic [27] is quite higher compared to proposed ODM scheme. One more thing to note that the proposed system performs the datamining on the random samples considering conventional support and confidence level of the data. The adoption of this process also makes the process to operate faster not only for the storage of the date but also for the knowledge discovery process. The entire complexity of the heterogenity of the massive data is also significantly minimized owing to the typical indexing policies incorporated by File Management Server and Data Ripping process

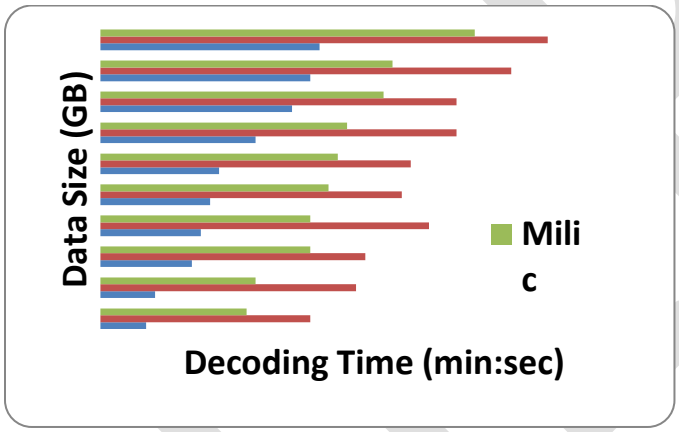

Figure 6 Analysis of Decoding Time

Fig.6 highlights the decoding time of the ODM system along with the conventional system of Mujawar [26] and Milic [27]. Decoding time is required to be analyzed for measuring the effectiveness of accessing the knowledge from the encoded file from the distributed storage system. It is also required to understand the success rate in accessibility of the uploaded file from the storage servers. A closer look into the curve will shows that decoding time is approximately minimized compared to the encoding time for almost all the techniques. However, proposed ODM scheme is witnessed with reduced decoding time, which is also in aggrement with the transmission rate that is increasing order.

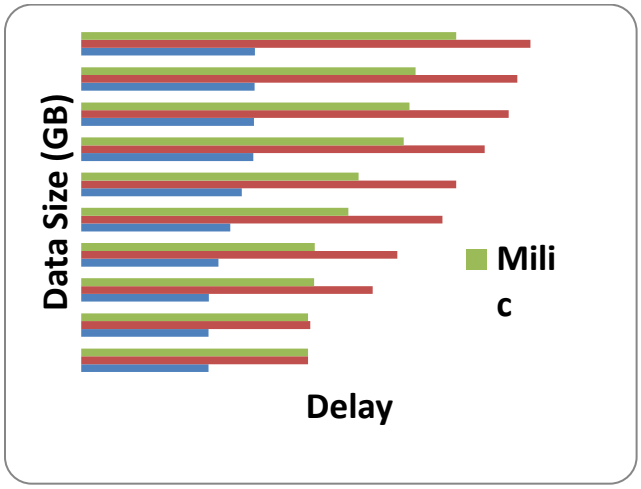

Figure 7 Analysis of Delay

Fig.7 shows the outcome of analyzed data propagation delay in case of large number of data rips on multiple storage systems. It could be expected that more the ripping process of user's data takes place on multiple storage system; more delay could be expected in the peak hours of operation over distributed storage environment. However, the outcome shows that Mujawar [26] as well as Milic [27] approaches are found with increasing trend of delay curve, which is due to increasing steps of fuzzy processing and iterative pattern analysis of data during the request or response of the user. However, the proposed ODM scheme performs lightweight indexing mechanism for which reason the decoding time is found to be quite lower as compared to encoding time. Moreover, the adoption of File Management Servers and Supervisor module assists in proper indexing mechanism of the massive chunks of data along with evaluation of the availability of the server. Therefore, although the proposed system exhibit increasing delay with increasing data size, but still it outperforms the conventional techniques of datamining found most recently in context of e-governance framework.

CONCLUSION

This paper has discussed about the problems that are associated with the existing data that are being generated from the e-government applications. Although, the adoption of e-government applications is poor in some countries, but still it can be expected that there will be a drastic change in this sector with the frequently launching new technologies. This paper has highlighted that it is not easy to perform datamining especially of the data are stored in distributed storage servers (like in cloud). As in such scenario, the data will be growing in volume as well as the data will be of different types. Hence, the existing datamining algorithm fails in extracting knowledge in this case. We show in this paper that heterogeneity problem in data can be minimized if a precise data indexing policy can be introduced. Hence, by adopting hash function, the size of the data doesn't only become less but also solves the problems of data discretization. Hence, our system can identify the data heterogeneity as well as it can store the data in its own file format. Our optimization technique is also highly cost effective as it is designed based on simple support and confidence level that can be found in association rule mining. But inspite of using association rule mining, our model shows better discovery of the rule set thereby solving the optimization problem in performing datamining over heterogeneous samples of data. By comparing the outcomes of the study with the most significant and recent work in similar area, we find that our model ensures highly reduced processing time and delay that is in compliance of the massive data processing or futuristic data analytics application on e-governance framework.

REFERENCES:

[1] C.S.R. Prabhu, E-Governance: Concepts And Case Studies, PHI Learning Pvt. Ltd., Business & Economics, 2013
[2] M. Castells, The Rise of the Network Society, Wiley-Blackwell, 2010
[3] A. K. Talukdar, Mobile Computing, 2E, Tata McGraw-Hill Education, Mobile computing, 2010
[4] P. Zheng, L. Ni, Smart Phone and Next Generation Mobile Computing, Morgan Kaufmann, Technology & Engineering, 2010
[5] F. Xhafa, L. Barolli, A. Barolli, P. Papajorgji, Modeling and Processing for Next-Generation Big-Data Technologies: With Applications and Case Studies, Springer, Computers, 2014
[6] I. H. Witten, E. Frank, M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques: Practical Machine Learning Tools and Techniques, Elsevier, Computers, 2011
[7] R. Kimball, M. Ross, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, John Wiley & Sons, Computers, 2013

[8] R. Wrembel, C. Koncilia, Data Warehouses and OLAP: Concepts, Architectures, and Solutions, Idea Group Inc (IGI), Computers, 2007

[9] S. Govind, L.M. Rao, "A Review on Contribution of Data mining in e-Governance Framework", *International Journal of Engineering Research and General Science*, Vol.3, Iss.2, March-April, 2015

[10] L.M. Rao, S.R. Krishna, "A technical insight on software architectures deployed in e-governance: A study," *IEEE-Fourth International Conference on Computing, Communications and Networking Technologies*, pp.1,6, 2013

[11] L.M. Rao, S.R. Krishna, "Efficient and Ubiquitous Software Architecture of e-Governance for Indian Administrative Services", *International Journal of Advanced research in Computer Science*, pp. 66–74, vol.4, Iss.11, 2010

[12] G. K. Rao and S. Dey, "Decision Support For E-Governance: A Text Mining Approach", *International Journal of Managing Information Technology*, Vol.3, No.3, August 2011

[13] H. Trivedi, A. Dutta, "Parametric Analysis of Water Resource Data (E-Governance Projects) Using Data Mining Techniques", *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 3 , No.2, Pages : 04 – 07, 2014

[14] P. Desai, "Data Mining Prediction Using Data Mining Extensions (DMX): A case study on E-Governance Birth Registration Data Mining Model", *International Journal of Advanced Research in Engineering and Technology*, Vol.5, Iss.6, pp. 23-26, 2014

[15] P. Desai, "Knowledge Discovery From Vehicle E-Governance Data Using Data Warehousing And Data Mining", *International Journal of Information technology & management Information System*, Vol.5, Iss.2, pp. 40-50, 2014

[16] S S Suresh, M. Mahale, "Student Performance Analytics using Data Warehouse in E-Governance System", *International Journal of Computer Applications*, Vol.20, No.6, April 2011

[17] A. Rajput, R. P. Aharwal, R. P. Aharwal, S.P. Saxena, M. Raghuvanshi, "J48 and JRIP Rules for E-Governance Data", *International Journal of Computer Science and Security*, Vol.5, Iss.2, 2011

[18] Y.Shen, Z. Liu, S. Luo ,H. Fu, Y. Li, "Empirical Research on E-Government Based on Content Mining", *International Conference on Management of e-Commerce and e-Government*, 2009

[19] B. S. Mampilli, J. Meenakumari, "A Study On Enhancing E-Governance Applications Through Semantic Web Technologies", *International Journal of Web Technology*, Vol 01, Iss.02, December 2012

[20] B. Bhanderi, "MROLE of data wareshousing & data minning in E-Governance", *International Multidisciplinary Journal of Applied Research*, Vol.1, Iss.4, ISSN 2321-7073, 2013

[21] J. Garg, S. B. Garg and N. Choudhary, "Effective Implementation of E-Governance in Technical Institutions in India using ICT to Make them World Class", *An International Journal of Engineering Sciences*, ISSN: 2229-6913, Vol. 4, 2011

[22] B. M.Bidgoli, E. Akhondzadeh, "A New Approach of Using Association Rule Mining in Customer Complaint Management", *International Journal of Computer Science Issues*, Vol. 7, Iss.5, September 2010

[23] S. Govind, L. M. Rao, "Efficient Framework of e-Government for Mining Knowledge from Massive Grievance Redressal Data", *International Conference in association with Stevenson University*, USA, ISBN NO : 978-1-63452-307-3, 2014

[24] D. Yanow, P. S.Shea, Interpretation and Method: Empirical Research Methods and the Interpretive Turn, Routledge, 2015

[25] P. K. Sahu, Research Methodology: A Guide for Researchers In Agricultural Science, Social Science and Other Related Fields, Springer Science & Business Media, Technology & Engineering, 2013

[26] A. Mujawar and V. Patil, "Fuzzy based data mining system for E-government", *Elixir-Computer Science and Engineering*, 2012

[27] P. Milic, N. Veljkovic, "Framework for Open Data Mining in e-Government", *ACM-Proceedings of 5th Balkan Conference in Informatics, pp.255-258*, 2012