

Modelling of E-Governance Framework for Mining Knowledge from Massive Grievance Redressal Data

Sangeetha G*, L. Manjunatha Rao**

* Computer Science, Bharathiar University, Coimbatore, India

** Department of MCA, Dr. Ambedkar Institute of Technology, Bangalore, India

Article Info

Article history:

Received Sep 14, 2015

Revised Nov 20, 2015

Accepted Dec 6, 2015

Keyword:

Data management

Datamining

E-Government

Grievance redressal

Knowledge discovery

ABSTRACT

With the massive proliferation of online applications for the citizens with abundant resources, there is a tremendous hike in usage of e-governance platforms. Right from entrepreneur, players, politicians, students, or anyone who are highly depending on web-based grievance redressal networking sites, which generates loads of massive grievance data that are not only challenging but also highly impossible to understand. The prime reason behind this is grievance data is massive in size and they are highly unstructured. Because of this fact, the proposed system attempts to understand the possibility of performing knowledge discovery process from grievance Data using conventional data mining algorithms. Designed in Java considering massive number of online e-governance framework from civilian's grievance discussion forums, the proposed system evaluates the effectiveness of performing datamining for Big data.

*Copyright © 2016 Institute of Advanced Engineering and Science.
All rights reserved.*

Corresponding Author:

Sangeetha G,

Research Scholar,

Computer Science,

Bharathiar University,

E-Mail: sangithagovind@gmail.com

1. INTRODUCTION

The National e-Governance Plan of Indian Government seeks to lay the foundation and provide the impetus for long-term growth of e-Governance within the country. This section provides information on creation of the right governance and institutional mechanisms, setting up the core infrastructure and policies and implementation of a number of Mission Mode Projects at the Center, State and integrated service levels. Several dimensions and factors influence the definition of e-governance or electronic governance. The word "electronic" in the term e-governance implies technology driven governance. E-governance is the application of information and communication technology (ICT) for delivering government services, exchange of information communication transactions, integration of various stand-alone systems and services between government-to-customer (G2C), government-to-business (G2B), government-to-government (G2G) as well as back office processes and interactions within the entire government framework [1]. Through e-governance, government services will be made available to citizens in a convenient, efficient and transparent manner. The three main target groups that can be distinguished in governance concepts are government, citizens and businesses/interest groups. In e-governance there are no distinct boundaries. Generally four basic models are available – government-to-citizen (customer), government-to-employees, government-to-government and government-to-business. The primary purpose of online information centre is to develop and maintain a community information network, which provides open and free access to online information for the citizens. The hallmark of this community information network is the ability of the general public to obtain information that may not have been previously, or easily, accessible to them. The main task of the grievance handling module is to maintain the details grievances received from citizens of the city.

1.1. Background

This section discusses about the existing literatures that has been carried out till date in the area of e-governance found relevant to proposed study. Mohammed and Hasson [2] demonstrated a framework that uses data warehouse techniques such as metadata common warehouse to support the universities' e-government. Renuse et al. [3] have highlighted the importance of data mining technology to design proactive services to reduce crime incidences in the police stations jurisdiction. Crime investigation has very significant role of police system in any country. Mampilli et al. [4] studied reveals that users and government agencies alike are coming to slowly realize that keyword-based search is not enough and Semantic web-based applications need to be designed. Karthika and Rangaraj [5] used to receive the mores/numerous of feedback results from user /students related to improve the educational as well as performance of educational. ALAjmi et al. [6] provides importance to the combination of Web Services on the e-Learning application domain, because Web Service is the most complex choice for distance education during these days. Moharana et al. [7] discussed different issues and challenges and suggests a framework to be adopted along with various technologies needed for successful implementation of E-Governance projects and to overcome the barriers. Das and Patra [8] presented a design approach based on the service oriented paradigm for building E-governance systems. Gudavalli et al. [9] discussed the role of biometric authentication in e-governance environment to provide services efficiently and securely over the internet.

Desai [10] demonstrates the use DMX query for making prediction from existing data mining models. Elia et al. [11] developed LR for Natural Language Processing (NLP) applications, composed by electronic dictionaries made of terminological multiword-expressions (Machine-Readable Form) and by local grammars (in the form of finite-state automata and transducers – FSA/FST. Rao and Dey[12] demonstrated how text-mining techniques can help in retrieval of information and relationships from textual data sources, thereby assisting policy makers in discovering associations between policies and citizens' opinions expressed in electronic public forums and blogs etc. Bhanti et al. [13] proposed E-governance implementation for higher education system with the use of data warehousing and data mining techniques.

Datamining in E-Governance plays an important role to analyze data. Treatment records of millions of patients can be stored and computerized and data mining techniques may help in answering several important and critical questions related to organization [14]. Without data mining it is difficult to realize the full potential of data collected within healthcare organization as data under analysis is massive, highly dimensional, distributed and uncertain [15].

Many organizations struggle with the utilization of data collected through an organization online transaction processing (OLTP) [16] system that is not integrated for decision making and pattern analysis. Critical case study of cloud for datamining has been considered in the work of radha et al. [17]. Elaachak et al. [18] have carried out study for desiging analytics on games. For successful E-governance organization it is important to empower the management and staff with data warehousing based on critical thinking and knowledge management tools for strategic decision making.

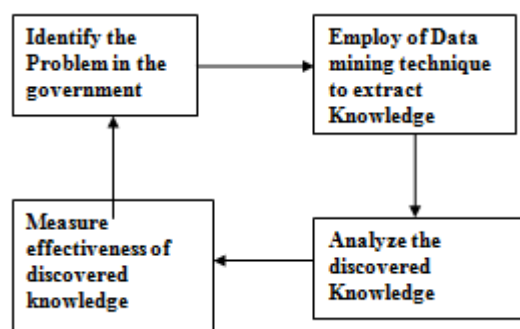


Figure 1. Data Mining Cycle

Data warehousing can be supported by decision support tools such as data mart, OLAP and data mining tools. A data mart is a subset of data warehouse. It focuses on selected subjects. Online analytical processing (OLAP) solution provides a multi-dimensional view of the data found in relational databases. With stored data in two dimensional formats OLAP makes it possible to analyze potentially large amount of data with very fast response times and provides the ability for users to go through the data and drill down or

roll up through various dimensions as defined by the data structure. The traditional manual data analysis has become insufficient and methods for efficient computer assisted analysis indispensable. A Data Warehouse is a semantically consistent data store that serves as a physical implementation of a decision support data model and stores the information on which an enterprise needs to make strategic decisions.

1.2. The Problem

The government of India, like all over the world, has began investing large amounts in Information and Communication Technology (ICT). The object behind these investment is to improve the efficiency of government function by, especially enabling citizen centric services. There are some technical issue which need to be discussed apart from above mentioned issue. The Above mentioned issue can be resolved by the government but as far as technical issues are concerned they need more focus to resolve the issue. Some of technical issues related to e-governance are:

- Technical Infrastructure support by the government
- Collection of Large amount of data
- Analysis of the data So that accurate Decision can be made
- Online Support to all department of Government organization
- Retrieval of meaningful Data
- Presentation of meaningful data so fast decision can be made

E-governance, meaning the electronic-governance, has evolved as an information age model of governance that seeks to realize process and structure for harnessing the potentialities of information & communication technologies at various level of government and public sector. E-governance is the commitment to utilize appropriate technologies to enhance governmental relationships in order to encourage the fair & efficient delivery of services. The ICT model uses the new technologies to maintain the data in government organization. Some of these are discussed in this paper which is very popular technologies now-a-days. Increasingly, government organization, are analyzing current and historic data to identify useful patterns from the large database so that they can support their business strategy Their main emphasis is on complex, interactive, exploratory analysis of very large dataset created by the integration of data from across all the part of the organization and that data is fairly static Three complementary trends are their Data warehouse, OLAP, Data Mining.

1.3. The Proposed Solution

The prime aim of the proposed system is to create a framework for grievance redressal board in existing e-governance framework where the evaluation of conventional datamining algorithm is carried out to check the efficiency of knowledge discovery of large data of grievances among the civilians in e-governance framework. An e-governance framework is designed that performs semantic evaluation for visualizing the bottleneck of current approach and need of future development of large streams of grievance data by analytic approach on the local machine. An architecture has been developed (as shown in Figure 2), where a possible scenario of grievance redressal data generation is shown. The architecture also represents the educational data generated from the civilian's community using various online grievance forums and thereby giving birth to larger size of files. The grievance data discussion forums are frequently used by various policy makers from various domain and expertise and hence different types of unstructured data are captured. The feedback system incorporates the basic source of data generation as the civilians like to share various perceptions about different social issues using various types of data. The data may be in text format or in image format or in other multimedia formats. However, for easiness in computation, we consider that the data is in text format only. Obviously, such data are highly unstructured in size which is almost impossible to perform any sorts of analysis on it. Moreover, performing conventional datamining techniques over large data is highly computational challenging task. Hence, in this paper, we try to build a computational cost efficient model using novel datamining algorithm. Document clustering is an enabling technique for many other machine learning applications, such as information classification, filtering, routing, topic tracking, and new event detection. Today, dynamic data stream clustering poses significant challenges to traditional methods. Typically, clustering algorithms use the Vector Space Model (VSM) to encode documents

2. RESEARCH METHODOLOGY

The VSM relates terms to documents, and since different terms have different importance in a given document, a term weight is associated with every term. These term weights are often derived from the frequency of a term within a document or set of documents. Much term weighting schemes have been proposed. Most of these existing methods work under the assumption that the whole data set is available and static. For instance, in order to use the popular Term Frequency – Inverse Document Frequency (TF-IDF)

approach and its variants, one needs to know the number of documents in which a term occurred at least once (document frequency). This requires a priori knowledge of the data, and that the data set does not change during the calculation of term weights. The need for knowledge of the entire data set significantly limits the use of these schemes in applications where continuous data streams must be analyzed in real-time. For each new document, this limitation leads to the update of the document frequency of many terms and therefore, all previously generated term weights needs recalibration. The schematic architecture of the proposed study is as shown as below

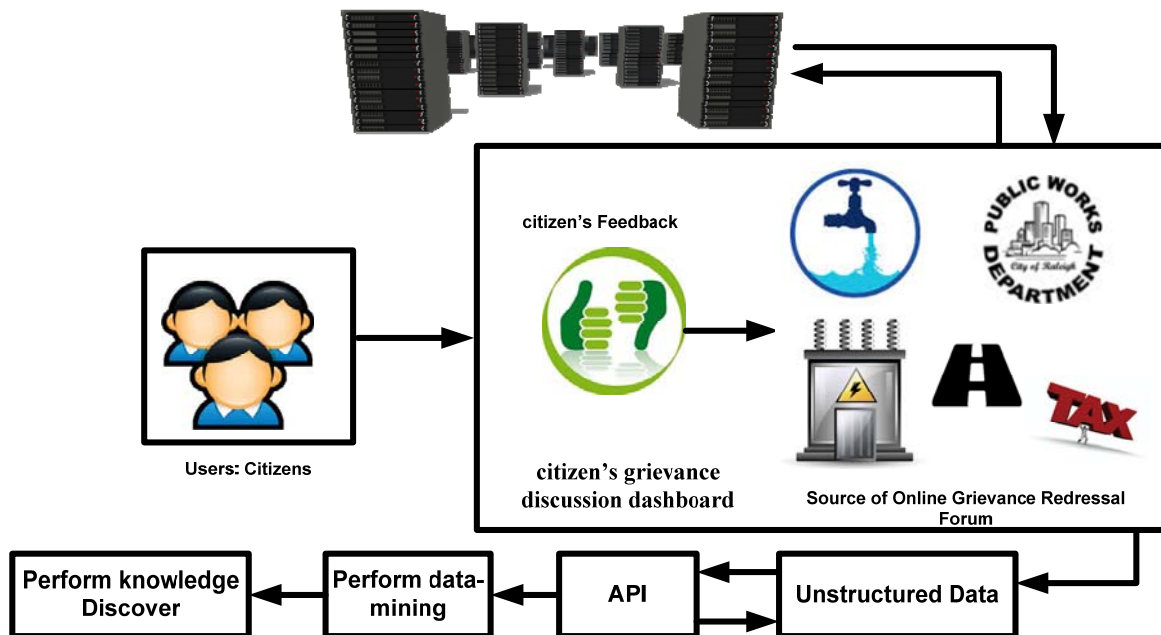


Figure 2. Schematic Architecture of Study

For N documents in a data stream, the computational complexity is $O(N^2)$, assuming that the term space M per document is much less than the number of documents. Otherwise, the computational complexity is $O(N^2 M \log M)$, where $O(M \log M)$ computations are needed to update a document.

The proposed system considers that different online user's gives feeds related to social grievance issues from multiple online civilian networking forums. In order to consider the challenges, the study considers all online civilian networking forums which are on the network. As civilian's feedback pertaining to grievance will differ highly from one to another, so proposed system is considered to have high number of missing data, noisy data, or unambiguous data, which are pre-processed by cleaning operation in conventional datamining technique. The unstructured data being collected is subjected to open source APIs for extracting the knowledge from unstructured data. The anticipated issues in the proposed system are highly likely to occur as the data is massive and highly unstructured. Moreover, the study eases the computation by not considering other file format and only considered text file with unstructured data. The framework captures the data from one row and check for noisy data ending up performing data cleaning process. The open source API is designed using java that performs extraction of the term frequency as well as inverse document frequency along with computation of simulation time. Also, it should be noted that the data are highly distributed type, where the system is developed focusing on faster processing of the datamining algorithms. The outcome of the results highlights that proposed system is found with increasing simulation time with the increase of dataset, and less linearity is found in the simulation time.

3. RESULTS AND DISCUSSION

The proposed system is designed in Java on 32 bit machine. The proposed system considers designing a database captured from multiple online civilian's grievances applications. The dataset consists of 1,15,000 online civilians networking users (Figure 3) who are exclusively found to use e-governance grievances streams information occupying around 150 Petabytes of data. The application design considering

from existing online civilian's grievances networking forums are huge and largely heterogeneous in type. The vast number of the users are exclusive considers who are found to write only text. The system however doesn't considers any other format of data e.g. PDF, audio, video etc. As the study is for its first kind, where we are attempting to perform data mining operation on grievance data using conventional datamining techniques, hence, the focus is more on the effective operation of the model without incorporating much of the complexities using different file formats. As the data size is pretty large enough and moreover it is collected from multiple online e-governance grievance redressal cell sites, hence the obtained data are quite massive and highly unstructured. The conventional datamining algorithm is expected to find a huge range of difficulties or computational challenges in doing so. Hence, the proposed system considers performing the data mining using conventional techniques on large data sets of e-governance data.

f3rmion	Nisha56977
Gabru_hsn	Elfrink_70330
Casanovakjp	Jerryzorau
Simister_79118	Tanya_6384
Serita96389	SouthernEclct
NelsonFGomez	Laurene_2969

Figure 3. Visualization of online citizen IDs

A closer look into the visual outcomes exhibited in Figure3 will show that the dataset is mainly semi-structured and possess all the challenges before applying the conventional datamining algorithm on the top of it. At present, the analysis is carried out only considering the textual data which consist of alphabets, numerals, special characters, as well as noisy and missing texts too. Therefore, a preprocessing is applied to identify the missing text using string-threshold based approach, while noisy data are identified and eliminated from the list. This step assists in making the data more reliable and compatible to be processed under proposed datamining approach for the purpose of knowledge discovery.

The next step is to perform cleaning operation. Figure 4 shows the computation being performed by proposed system that finally process the big data and compute total term, term to check, TF (Term Frequency) value, and IDF (Inverse Document Frequency) value for all the users considered from the online grievance redressal networking forum. It is interesting to know that even with large dataset, the proposed system is able to perform conventional datamining operations only on the smaller chunks of data, whereas when it comes to larger set of data, the simulation time is extensively increased, showing the need of an efficient protocol that can perform the faster computation or the knowledge discovery of the data generated from the e-governance grievance's online discussion forum.



Figure 4. Graphical Visualization of performance

The proposed system is mainly based on the feedback based analysis over the textual data and therefore analysis is being performed to identify the percentage of positive and negative feedback or opinion found in the dataset. Figure 5 highlights that the proposed system has got less negative reviews and more positive reviews. A classification scheme was constructed to quantify the extent to which positive and negative emotions were expressed in each comment. Although the conventional approaches (e.g., usability test and user interview) are useful for qualitatively knowing the interactive problems and user expectations of an online grievance network forum, they are less helpful to identify who may churn in the future. A number of research issues and challenges facing the realization of utilizing data mining techniques in online student network analysis could be identified as follows:

3.1. Linkage-Based and Structural Analysis

This is an analysis of the linkage behaviour of the online grievance redressal network so as to ascertain relevant nodes, links, communities and imminent are as of the network. The study is found with less links associated with more than 100 users, evidently proving ineffective datamining process (Figure 5).

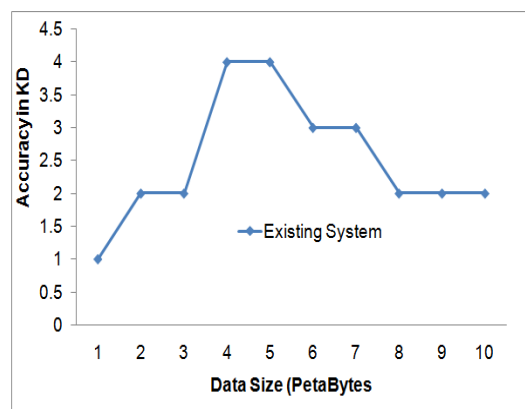


Figure 5. Analysis of Accuracy in Knowledge Discovery

3.2. Dynamic Analysis and Static Analysis

Static analysis such as in bibliographic networks is presumed to be easier to carry out than those in streaming networks. In static analysis, it is presumed that online e-governance student network changes gradually over time and analysis on the entire network can be done in batch mode. Conversely, dynamic analyses of streaming networks are very difficult to carry out. Data on these networks are generated at high speed and capacity. Dynamic analyses of these networks are often in the area of interactions between entities. The outcome of the study shows more simulation time leading to overhead.

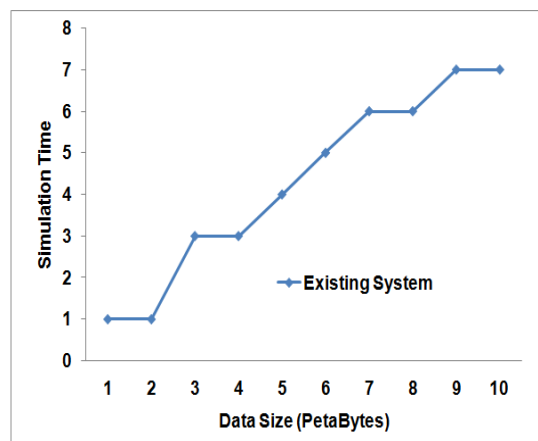


Figure 6. Analysis of Simulation Time

The outcome shows that adoption of conventional datamining technique can be definitely used for extracting unique knowledge. The results shown in Figure 6 also highlights that accuracy in analysis process is quite improving with increase of dataset from e-governance applications. Hence, it can be seen that identification of resources, technology infrastructure is there in conventional datamining algorithm. However, owing to massive size of data the accuracy is around 57%, which can be further more enhanced in future. There is a good scope of optimization principle over conventional database to perform effective mining operation.

4. CONCLUSION

The proposed system discusses about the framework that evaluates the extent of effectiveness of conventional datamining algorithms on large data captured from e-governance grievance redressal data in multiple online resources available. The outcome of the study shows higher simulation time, more overhead, and inaccuracy in knowledge discovery process. Therefore, we are successfully exhibiting the fact that conventional datamining algorithms cannot be directly applicable to Big Data for performing knowledge discovered process. Hence, the core findings of the study are as follows e.g. i) the existing and upcoming applications of e-governance framework will lead to generation of massive volume of the data that requires data analytics, ii) the existing data analytics tool (or datamining approaches) are not directly applicable to such massive database owing to unstructured or semi-structured format of the database, iii) the existing datamining technique can be applicable to massive data also provided if accurate preprocessing is done to convert unstructured or semi-structured data to structured one, iv) datamining based on feedback system could be highly enhanced using optimization techniques in future. Our future work will be in the direction of presenting a design of collaborative network that can share grievance information on a cloud. The proposed system has following benefit for the decision makers and civilians:

- They do not have to deal with the heterogeneous and sporadic information generated by various state-level computerization projects as they can access current data with a high granularity from the information warehouse.
- They can take micro-level decisions in a timely manner without the need to depend on their IT staff.
- They can obtain easily decipherable and comprehensive information without the need to use sophisticated tools.
- They can perform extensive analysis of stored data to provide answers to the exhaustive queries to the administrative cadre. This helps them to formulate more effective strategies and policies for citizen facilitation.
- They are the ultimate beneficiaries of the new policies formulated by the decision makers and policy planner's extensive analysis on person and land-related data.
- They can view frequently asked queries whose results will already be there in the database and will be immediately shown to the user saving the time required for processing.
- They can have easy access to the Government policies of the state.

The web access to Information Warehouse enables them to access the public domain data from anywhere.

REFERENCES

- [1] E. Estevez, P. Fillotrani and T. Janowski, "From e-government to seamless government", In Proceedings of the Conference on Collaborative Electronic Commerce Technology and Research, COLLECTeR Iberoamerica 2007, Cordoba, Argentina, Vol. 1. 2007.
- [2] M. Mohammed and A. Hasson, "Metadata Technique with E-government for Malaysian Universities", *IJCSI International Journal of Computer Science Issues*, Vol. 9, Issue 4, No 3, July 2012.
- [3] H.N. Renushe, P.R. Rasal, A.S. Desai, "Data Mining Practices for Effective Investigation of Crime", Prof. Hanmant N Renushe et al, *Int. J. Computer Technology & Applications*, Vol 3 (3), 865-870, 2012.
- [4] B.S. Mampilli, J. Meenakumari, "A Study on Enhancing E-Governance Applications through Semantic Web Technologies", *International Journal of Web Technology*, Vol. 01, Issue. 02, December 2012.
- [5] D. Karthika, R. Ranga Raj, "Survey of E-Governance based Higher Education System in India with Data Mining", *International Journal of Science and Research (IJSR)*, Vol. 2, Issue. 12, 2013.
- [6] M.F. AlAjmi, S. Khan, A. Sharma, "Studying Data Mining and Data Warehousing with Different E-Learning System", *International Journal of Advanced Computer Science and Applications*, Vol. 4, No.1, 2013.
- [7] Chitta Ranjan Moharana, Debasish Rout, "Barriers in Good Governance: A Study of Various E-Governance Projects in India", Volume 3, No.3.1 Quarter I, 2013.
- [8] R.K. Das and M.R. Patra, "A Service Oriented Design Approach for E-Governance Systems", *International Journal of Information Technology Convergence and Services (IJITCS)*, Vol.3, No.3, June 2013.
- [9] M. Gudavalli, D.S. Kumar and S.V. Raju, "Securing E-Governance Services through Biometrics", *International Journal of Security and Its Applications*, Vol.8, No.1 (2014), pp.103-112.

- [10] 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 (IJARET)*, Vol. 5, Issue. 6, pp. 23-26, 2014.
- [11] A. Elia, D. Vellutino, F. Marano, A.M. Langella, and A. Napoli, "Semantic Web and language resources for e-Government: linguistically motivated Data Mining", In Proceedings of the International Conference on Web Intelligence, Mining and Semantics, pp. 31, 2011.
- [12] G.K. Rao and S. Dey, "Decision Support for E-Governance: A Text Mining Approach", *International Journal of Managing Information Technology (IJMIT)*, Vol.3, No.3, August 2011.
- [13] P. Bhandi, U. Kaushal, A. Pandey, "E-Governance in Higher Education: Concept and Role of Data Warehousing Techniques", *International Journal of Computer Applications, (0975 – 8887)*, Vol. 18, No.1, March 2011.
- [14] H. Kaur and S.K. Wasan, "Empirical Study on Applications of Data Mining Techniques in Healthcare", *Journal of Computer Science*, Vol. 2 (2), pp. 194-200, 2006.
- [15] C.C. Aggarwal and P.S. Yu, "On Indexing High Dimensional Data with Uncertainty" In *SDM*, pp. 621-631. 2008.
- [16] K.Srinivas, B.K. Rani, A.Govrdhan, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks", *International Journal on Computer Science and Engineering*, Vol. 02, No. 02, pp.250-255, 2010.
- [17] K. Radha, B.T. Rao, S.M. babu, "Service Level Agreements in Cloud Computing and Big data", *International Journal of Electrical and Computer Engineering*, Vol.5, No.1, pp.158-165, 2015.
- [18] L. Elaachak, A. Belahbibe, M.Bouhorma, "Towards a System of Guidance, Assistance, and Learning Analytics Based on Multi Agent System Applied on Serious Games", *International Journal of Electrical and Computer Engineering*, Vol.5, No.2, pp.344-354, 2015.

BIOGRAPHIES OF AUTHORS



Sangeetha Govinda is working as Assistant Professor, Department of MCA, Adarsh Institute of Management and Information Technology, Bangalore. She has got 10 years of teaching experience. She has obtained Bachelor of Science from Bangalore University in the year 2000. She studied Masters of Computer Application from IGNOU and was awarded in the year 2006. She obtained Bachelor of Education in 2008. In 2014 did Master of Philosophy from Bharathiar University, Coimbatore. Now she is a Ph.D. student 3rd year of CSE at Bharathiar University, Coimbatore, India. She has authored one (1) Textbook. She has published papers in both national and international conferences and 2 research papers in International Journals. Now currently working in Bangalore, India.



Dr. L. Manjunatha Rao is working as Professor and Head, Department of MCA, Dr. AIT, Bangalore. He has got 25 years of teaching experience. He did his Bachelor of Science from Bangalore University in the year 1990. He Studied Masters of Computer Application from Madhurai Kamaraj University and was awarded in the year 1999. In 2002 did Master of Philosophy from Mononmanium Sundaranar University. He has awarded Ph.D from Vinayaka Mission University, Tamil Nadu. He has published research papers in both national and international Journals and has authored 2 textbooks.