Role of Data Warehousing & Data Mining in E-Goverance

*Vaibhav Panwar**

Abstract

E-governance is the application of information & communication technologies to transform the efficiency, effectiveness, transparency and accountability of informational & transactional exchanges with in government, between Government & Government agencies of National, State, Municipal & Local levels, Citizens & Businesses, and to empower citizens through access & use of information. Governments deal with large amount of data. To ensure that such data is put to an effective use in facilitating decision-making, a data warehouse is constructed over the historical data. It permits several types of queries requiring complex analysis on data to be addressed by decision-makers.

This Paper deals with scope and use of data warehousing & Data mining in all the dimensions of e-governance like ***Government to Citizen (G2C) Citizen to Government (C2G), Government to Government (G2G) Government to Business ,Government to NGO (G2N).*** *Their are many methodologies used to increase the efficiency of E-governance .Three complimentary trends are* ***Data warehousing, OLAP ,Data mining.*** *By using these techniques we find that data warehousing is very helpful in analyzing Current & Historical data, finding useful pattern & support decision strategies .OLAP is useful in solving complex queries & views , interactive online analysis of data .Using Data mining technique & algorithm, automatic discovery of pattern & other interesting trends are found out*

Keywords: *E-governance, OLAP ,Data warehousing, Data Mining*

1. Introduction

E-governance involves the application of Information and Communication Technologies by government agencies for information and service delivery to citizens, business and government employees. It is an emerging field, faced with various implementation problems related to technology, employees, flexibility and change related issues, to mention a few. Global shifts towards increased deployment of IT infrastructure by governments emerged with the advent of the World Wide Web. With the increase in Internet and mobile connections, the citizens are learning to exploit their new mode of access in wide ranging ways. They have started expecting more and more information and services online from governments and corporate organizations to enhance their civic, professional and personal lives .The concept of e-governance came into existence in India during the seventies with a focus on development of government applications in the areas of defense, economic monitoring, planning and the inclusion of Information Technology to manage data intensive functions related to elections, census, tax administration etc. The role & efforts made by National

* *Vaibhav Panwar*
Asst Prof & Head MCA Deptt
Institute of Management Studies Dehradun
Email:panwar.vaibhav@gmail.com

Informatics Center (NIC) to connect all the district headquarters during the eighties was a very new innovative approach.

From the early nineties, IT technologies were supplemented by ICT technologies to extend its use for wider applications with policy implementation & emphasis on reaching out to rural areas and taking in greater inputs from NGOs and private sector. There has been an increased involvement of international agencies under the framework of e-governance for development to catalyze the development of e-governance laws and technologies in developing countries. For governments, the motivation to shift from manual processes to IT-enabled processes may be increased efficiency in administration and service delivery, but this shift can be conceived as a worthwhile investment with potential for returns.

E-governance is the process of service delivery and information dissemination to citizens using electronic means providing the following benefits over the conventional system

- Increased efficiency in various Governmental processes
- Transparency and anticorruption in all transactions
- Empowerment of citizens and encouragement of their participation in governance.

The main objective of E-Governance is to change organization into e-organization. An e-organization needs to focus on the following things:-

- ✓ develop customer orientation
- ✓ manage customer relationships
- ✓ streamline business processes
- ✓ communicate better
- ✓ organize information
- ✓ work more flexibly
- ✓ make better decisions.
- ✓ coordinate activities better

2. NEED FOR E-GOVERNANCE

2.1 Pre conditions of E-Governance

Some of the pre conditions for an effective e-governance that could be listed as

- ➢ Formulation of new set of cyber laws to replace traditional set of rules & regulations for effective replacement of e-governance
- ➢ Simplification of procedures, rationalization of various administrative processes restructuring of government and mindset of bureaucrats to adapt according to e-governance.
- ➢ De-layering or re-layering of decision-making of levels
- ➢ Security and privacy are the two major concerns

2.2 Factors necessary for successful e-governance

- ➢ Political commitment
- ➢ Effective administrative leadership
- ➢ Effective handling of HR issues
- ➢ Involvement of staff at design stage
- ➢ Innovative funding strategy and revenue model
- ➢ Appropriate administrative structure
- ➢ Common infrastructure and database creation
- ➢ Training & Motivation

3. IMPLEMENTATION ISSUE IN E-GOVERNANCE

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 issues which need to be discussed apart from above mentioned issues. The Above mentioned issues can be resolved by the government but as far as technical issues are concerned they need more focus to resolve.

Some of technical issues related to g-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 departments of

a Government organization

- ✓ Retrieval of meaningful Data
- ✓ Presentation of meaningful data to enable fast decisions.

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.

Increasingly, government organizations, 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 parts of the organization and that data is fairly static Three complementary trends are there:

1) Data warehouse
2) OLAP
3) Data Mining

4. ROLE OF DATA WARE HOUSE IN E-GOVERNANCE

4.1 Need for data warehouse

Governments deal with enormous amount of data. In order that such data is put to an effective use in facilitating decision-making, a data warehouse is constructed over the historical data. It permits several types of queries requiring complex analysis on data to be addressed by decision-makers.

When used properly, it can help planners and decision makers in making informed decisions leading to positive impact on targeted group of citizens. However to use information to it's fullest potential, the planners and decision makers need instant access to relevant data in a properly summarized form. In spite of taking lots of initiative for computerization, the Government decision makers are currently having difficulty in obtaining meaningful information in a timely manner because they have to request and depend on IT staff for making special reports which often takes long time to generate. An Information Warehouse can deliver strategic intelligence to the decision makers and provide an insight into the overall situation. This greatly facilitates decision-makers in taking micro level decisions in a timely manner without the need to depend on their IT staff. By organizing person and land-related data into a meaningful Information Warehouse, the Government decision makers can be empowered with a flexible tool that enables them to make informed policy decisions for citizen facilitation and accessing their impact over the intended sections of the population.

4.2 Benefit of a data warehouse for e-governance

Citizen facilitation is the core objective of any Government body. For facilitating the citizens of a state or a country, it is important to have the right information about the people and the places of the concerned territory. Hence a data warehouse built for eGovernance can typically have data related to persons and land. Such a data warehouse can be beneficial to both the Government decision makers and citizens as well in the following manner:

4.2.1 Benefit for the decision makers

- ➢ 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

4.2.2 Benefit for the citizens

- ➢ 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.

The data warehouse has enough potential to access the impact of various welfare schemes across the population of the state. The planners can design schemes focused on specific target groups and achieve high impact. The decision-makers can carry out analysis of population profile across the state in areas of economy, education, family units, shelter, etc. The warehouse can also be used for rural and urban development planning, agricultural yield and cropping pattern analysis and much more. These analyses will help in making decisions that are focused and the benefit of the government policies can reach the intended group. The various types and number of queries that can be handled by the data warehouse are limited only by the intelligence of the person using the data warehouse and the data fed to it.

5. ROLE OF DATA MINING IN E-GOVERNANCE

It is well known that in Information Technology (IT) driven society, knowledge is one of the most significant assets of any organization. The role of IT in E-governance is well established. Knowledge Pragmatic use of Database systems, Data Warehousing and Knowledge Management technologies can contribute a lot to decision support systems in E-governance. Knowledge discovery in databases is well-defined process consisting of several distinct steps. Data mining is the core step, which results in the discovery of hidden but useful knowledge from massive databases. A formal definition of Knowledge discovery in databases is given as :"Data mining is the non trivial extraction of implicit previously unknown and potentially useful information about data".Data mining technology provides a user- oriented approach to discover novel and hidden patterns in the data. The discovered knowledge can be used by the E-governance administrators to improve the quality of service. Traditionally, decision making in E-governance is based on the ground information, lessons learnt in the past resources and funds constraints. However, data mining techniques and knowledge management technology can be applied to create knowledge rich environment. An organization may implement Knowledge Discovery in databases (KDD) with the help of a skilled employee who has good understanding of organisation. KDD can be effective at working with large volume of data to determine meaningful pattern and to develop strategic solutions. Analyst and policy makers can learn lessons from the use of KDD in other industries. E-governance data is massive. It includes centric data, resource management data and transformed data. E-governance organizations must have ability 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.

5.2 Knowledge Discovery in E-governance

Data mining is an essential step of knowledge discovery. In recent years it has attracted great deal of interest in Information industry. Knowledge discovery process consists of an iterative sequence of data cleaning, data integration, data selection, data mining pattern recognition and knowledge presentation. In particulars, data mining may accomplish class description, association, classification, clustering, prediction and time series analysis. Data mining in contrast to traditional data analysis is discovery driven. Data mining is a young interdisciplinary field closely connected to data warehousing, statistics, machine learning, neural networks and inductive logic programming. Data mining provides automatic pattern recognition and attempts to uncover patterns in data that are difficult to detect with traditional statistical methods. 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.

Data Mining Cycle

For Goverment organization to succeed they must have the ability to capture, store and analyze data. Online analytical processing (OLAP) provides one way for data to be analyzed in a multi-dimensional capacity. With the adoption of data warehousing and data analysis/OLAP tools, an organization can make strides in leveraging data for better decision making. Many organizations struggle with the utilization of data collected through an organization online transaction processing (OLTP) system that is not integrated for decision making and pattern analysis. 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. 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 format 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. A data warehouse is also often viewed as architecture constructed by integrating data from multiple heterogeneous sources to support structured and/or ad-hoc queries, analytical reporting and decision making

5.3 Data Mining technique in E-governance

There are various data mining techniques available with their suitability dependent on the domain application. Statistics provide a strong fundamental background for quantification and evaluation of results. However, algorithms based on statistics need to be modified and measured before they are applied to data mining.

5.4 CLASSIFICATION OF DATA MINING TECHNIQUES

5.4.1 Rule Induction

Rule induction is the process of extracting useful 'if then' rules from data based on statistical significance. A Rule based system constructs a set of if-then-rules. Knowledge represents has the form IF conditions THEN conclusion. This kind of rule consists of two parts. The rule antecedent (the IF part) contains one or more conditions about value of predictor attributes where as the rule consequent (THEN part) contains a prediction about the value of a goal attribute. An accurate prediction of the value of a goal attribute will improve decision-making process. IF-THEN prediction rules are very popular in data mining; they represent discovered knowledge at a high level of abstraction. Rule Induction Method has the potential to use retrieved cases for predictions.

5.4.2 Decision tree

It is a knowledge representation structure consisting of nodes and branches organized in the form of a tree such that, every internal non-leaf node is labeled with values of the attributes. The branches coming out from an internal node are labeled with values of the attributes in that node. Every node is labeled with a class (a value of the goal attribute). Tree based models which include classification and regression trees, are the common implementation of induction modeling. Decision tree models are best suited for data mining. They are inexpensive to construct, easy to interpret, easy to integrate with database system and they have comparable or better accuracy in many applications. There are many Decision tree algorithms such as HUNTS algorithm (this is one of the earliest algorithm), CART, ID3, C4.5.

CONCLUSION

A large number of e-Governance applications are already in operation in most of the states and at the centre. The necessary DWM infrastructure has been created at the head-quarter and sufficient number of officials have been trained on DWM. This is the right time for introducing DWM in the e-Governance arena and to further strengthen the e-Governance system. Once the desired results are achieved, the same can be replicated in other sectors of the government.

REFERENCES

1. Ester M., Wittmann R.: "Incremental Generalization for Mining in a Data Warehousing Environment", Proc. 6th Int. Conf. on Extending Database Technology, Valencia, Spain, 1998, in: Lecture Notes in Computer Science, Vol. 1377, Springer, 1998, pp. 135-152.

2. Fayyad U., Piatetsky-Shapiro G., and Smyth P.: "Knowledge Discovery and Data Mining: Towards a Unifying Framework", Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining, Portland, OR, 1996, pp. 82-88.

3. Gueting R. H.: "An Introduction to Spatial Database Systems", The VLDB Journal, Vol. 3, No. 4, October 1994, pp. 357-399.

4. Han J., Cai Y., Cercone N.: "Data-driven Discovery of Quantitative Rules in Relational Databases", IEEE Transactions on Knowledge and Data Engineering, Vol.5, No. 1, 1993, pp. 29-40.

5. Kaufman L., Rousseeuw P. J.: "Finding Groups in Data: An Introduction to Cluster Analysis", John Wiley & Sons, 1990.

6. Mumick I. S., Quass D., Mumick B. S.: "Maintenance of Data Cubes and Summary Tables in a Warehouse", Proc. ACM SIGMOD Int. Conf. on Management of Data, 1997, pp. 100-111.

7. Ng R. T., Han J.: "Efficient and Effective Clustering Methods for Data Mining", Proc. 20th Int. Conf. on Very Large Data Bases, Santiago, Chile, 1994, pp. 144-155.