



A Survey of Data Mining Research and Implementation for E-Governance Systems

KEYWORDS

Pushpal Desai

ABSTRACT *In this paper an overview of the current research being carried out using Data Mining algorithms for knowledge discovery from e-governance systems is discussed. The objective of this study is to identify various advantages of implementing Data Mining algorithms on e-governance systems. The review of literature indicates that Data Mining can be effectively utilized in government organizations. The review indicates that knowledge discovered from historical e-governance data can be utilized for better planning and decision making.*

Introduction

The Data Mining techniques are extensively used by private organizations and research communities to uncover hidden trends and knowledge from historical data. The Data Mining concepts are successfully implemented in several areas like "Banking", "Credit Card Business", "Insurance", "Customer Relationship Management", "Super Store Sales data analysis", "Stock Market", "Gaming", "Network and Security", "Financial Market", "Telecommunication", "Oil and Gas exploration", "Weather Forecasting" etc...In recent past, the government organizations have also realized the potential use of Data Mining on e-governance data. The Data Mining algorithms can be applied on e-governance data to find hidden trends and knowledge from historical data. In this paper, literature related to Data Mining implementation on e-governance data is discussed.

Survey of Data Mining for e-governance data

Matjaž Gams et al. applied data mining techniques to understand impact of various attributes on fertility rates. In their research paper, demographic analysis of 147 countries with 95 basic attributes is considered. They applied decision trees for uncovering knowledge from the data. They discussed about several decision trees based on various attributes. They discussed about various decision trees such as decision trees with all attributes, decision trees with selected attributes, decision trees with direct attributes, decision trees with social attributes, decision trees with economical attributes, decision trees with education attributes and decision trees for developed countries etc... Their study showed that decision trees can be effectively used to study fertility rate considering various attributes [1].

Neera Singh et al. in their research work analyzed the data for the state of Uttar Pradesh and study the age, place, gender, number of doses of polio and various other attributes. They collected and used data of recorded cases in Greater Bombay and districts of Uttar Pradesh. They used WEKA data mining software for the practical implementation. They used data mining methods such as association rules, clustering, decision trees etc...to predict which age group, communities, and districts were prone to polio. Their research paper highlighted that seasons like July-September and certain regions are prone to having polio cases. Their study concluded that the Muslim community and rural population are more prone to the disease. Their research work confirmed that data mining can be effectively used in government sector for improvement in planning and decision making [2].

Kishori Lal Bansal and Satish Sood proposed the use of data warehousing and data mining in all spheres of e-governance like government to citizen, government to government and government to business. Their research paper suggested

that efficiency of Government can be increased by using the data warehouses and data mining [3].

G. Koteswara Rao et al. discussed about using text mining in e-governance. They proposed text mining techniques in web forums and blogs to discover interesting association between government policies and citizen opinions. Their approach suggested that citizen services can be improved using text mining. Furthermore, they discussed about important points like multilingual text mining, multilingual and cross lingual project, machine translation, technical architecture of bi-lingual text mining, etc...Their research work concluded that India can initiate text mining project at national level and later on it can be implemented at state level [4].

Behrouz Minaei-Bidgoli et al. proposed data mining technique for customer complaint management. Behrouz Minaei-Bidgoli et al. used citizen complain data of Tehran Municipal Corporation for their research work. They used various attributes to find association among them. They used various attributes like subject, message time, gender, education, region etc...and found interesting relationship. Their work demonstrated association rules between region, season and subject attributes. These association rules make it possible for the municipality administrator to find out the causes of complaints. Their results concluded that association mining helps the municipality to optimize its services [5].

Sushil Kumar et al. discussed about conceptual model for electronic voting and various advantages of using data mining on data collected via e-voting. Their conceptual study shows that data mining can be very useful in discovering knowledge from e-voting database [6].

Ebrahim Sahafzadeh et al. used data mining technique such as clustering and decision trees to predict dusty days of Boushehr city. They used data of past 53 year for research work. They derived various rules from data by using many attributes such as air pressure, air humidity and dusty days. They used Clementine software for practical implementation and they were able to classify dusty days into five different classes [7].

Wei Cheng et al. applied data mining on traffic violation data. They analyzed different type of information like traffic violate type, traffic violated people, traffic accidents, weather, time, road type etc...Their research work concluded that data mining could find interesting relationships and hidden knowledge from traffic violation data [8].

Malathi et al. discussed about enhanced data mining algorithm for crime prevention. They proposed new approach to predict crime patterns. They focused on MV algorithm and

Apriori algorithm for missing value and crime patten generation. Malathi et al. also implemented their algorithm practically on real crime data. Their research work demonstrated that they were able to derive knowledge from crime data. They used crime dataset from the Integrated Network for Societal Conflict Research (INSCR) website (Marshall and Marshall, 2008). Their practical implementation results proved that their tool is effective in terms of analysis speed, identifying common crime patterns and future prediction [9].

Sujatha et al. in their research work considered various socio and economic factors data like Agriculture, Rainfall, Sex-ratio, Unemployment rate, GDP, Literacy rate, Population, Area along with the crime data. They proposed data mining model based on these factors. The proposed to used classification for crime count prediction, clustering for generating crime patterns and association rule for finding relationship among these factors. They practically implemented to their proposed model using Weka data mining software. They used classification of crime data using J48 classifier for different crimes against women in India. These crimes are rape, cruelty by husband & relatives, kidnapping & abduction of women & girls, dowry death, molestation, sexual harassment and importation of girls. Their results demonstrate different types of crime against women in India in the year 2011. Their finding suggested that cruelty by husband & relatives is the largest crime compared to all other crimes considered in their study. Their finding also suggested that cruelty by husband & relatives is too high in West Bengal followed by Andhra Pradesh and Rajasthan. This cruelty is the least in Nagaland as per their study [10].

Anjum Mujawar et al. proposed fuzzy based data mining system for e-government. Their research paper proposed an algorithm of the fuzzy data mining. They discussed regarding use of the fuzzy data mining to find the useful information in the electronic government affairs system of the state of Maharashtra. They considered data of Maharashtra tax bureau for their practical implementation. They concluded that fuzzy data mining on e-government data will be great help for policy maker [11].

Hanmant N. Renushe et al. proposed short term crime forecasting using data mining. They considered crime data of Satara district for practical implementation. They considered the crime incidences of different categories of crimes from January 2010 to November 2010 of Satara district to forecast crime incidence for the December 2010. They study concluded that actual no of crime reported during the December 2010 are approximately same as the forecasted crimes using short term forecasting. They also concluded that the short term forecasting method is very useful in crime prevention and investigation process [12].

Conclusion

The survey suggest that data warehouse and data mining concepts are adopted in many government sectors like healthcare, agriculture, education, social security fund, pollution control, electronic voting, rainfall prediction, customer complain, road traffic violation, crime control, crime forecasting, tax department etc... The Table 1 provides summary of the literature review.

Table 1 Summary of literature review on data mining implementation on e-governance data

Authors	Year	Research Area	Country	Practical Implementation?	Remarks
Matjaž Gams et al. [1]	2008	demography, fertility	Multinational	YES	Analysis performed using Decision Trees
Neera Singh et al. [2]	2011	Healthcare	India	YES	Knowledge discovered using Association rules, Clustering, Decision Trees
Kishori Lal Bansal et al. [3]	2011	E-governance	India	NO	Conceptual discussion about use of data warehouse and data mining in e-governance
G. Koteswara Rao et al. [4]	2011	DSS	India	NO	Conceptual discussion about Text mining
Behrouz Minaei-Bidgoli et al. [5]	2010	Customer complain system	Iran	YES	Knowledge discovery using Association rules
Sushil Kumar et al. [6]	2009	E-voting	India	NO	Conceptual discussion
Ebrahim Sahafizadeh et al. [7]	2009	Air Pollution	Iran	YES	Knowledge discovery using Clustering and decision trees
Wei Cheng et al. [8]	2010	Road Traffic	China	YES	Knowledge discovery using Association rules
Malathi. A et al. [9]	2011	Crime Detection	India	YES	Enhance Data Mining algorithm for Crime Detection
R Sujatha et al. [10]	2013	Crime Detection	India	YES	Crime Detection using Classification
Anjum Mujawar [11]	2012	Tax	India	YES	Better decision making in Tax department using Fuzzy Data Mining
Hanmant N. Renushe et al. [12]	2012	Crime Forecasting	India	YES	Crime forecasting and prevention using data mining

The survey confirmed that many researchers and government bodies have already implemented data mining.

REFERENCE

Matjaž Gams and Jana Krivec, Demographic Analysis of Fertility Using Data Mining Tools, *Informatica* 32, 2008, pp. 147–156, 2008 | Neera Singh, Sonali Agarwal, Ramesh C. Tripathi, A Data Mining Perspective on the Prevalence of Polio in India, *International Journal on Computer Science and Engineering (IJCSE)*, ISSN : 0975-3397 Vol. 3 No. 2 Feb 2011, pp. 580-585., 2011 | Kishori Lal Bansal and Satish Sood, ROLE OF DATA WAREHOUSING & DATA MINING IN E-GOVERNANCE, *International Journal of Artificial Intelligence and Knowledge Discovery* Vol.1, Issue 1, Jan, 2011, pp. 38-41, 2011 | G. Koteswara Rao and Shubhamoy Dey, DECISION SUPPORT FOR E-GOVERNANCE: A TEXT MINING APPROACH, *International Journal of Managing Information Technology (IJMIT)* Vol.3, No.3, August 2011 DOI : 10.5121/ijmit.2011.3307 pp. 73-91., 2011 | Behrouz Minaei-Bidgoli, Elham Akhondzadeh, A New Approach of Using Association Rule Mining in Customer Complaint Management, *IJCSI International Journal of Computer Science Issues*, Vol. 7, Issue 5, September 2010, pp. 142-147, 2010 | Sushil Kumar, Himanshu Aggarwal, Upinder Kaur & Sumati, E-GOVERNANCE AND DATA MINING: A METHODOLOGY FOR E-VOTING, *International Journal of Information Technology and Knowledge Management* January June 2009, Volume 2, No. 1, pp. 139-144, 2009 | Ebrahim Sahafizadeh, Esmail Ahmadi, Prediction of Air Pollution of Boushehr City Using Data Mining, 2009 Second International Conference on Environmental and Computer Science, DOI 10.1109/ICECS.2009.18, pp. 33-36, 2009 | Wei Cheng, Xiaofeng Ji, Chunhua Han, Jianfeng Xi, The Mining Method of the Road Traffic Illegal Data Based on Rough Sets and Association Rules, 2010 International Conference on Intelligent Computation Technology and Automation, DOI 10.1109/ICICTA.2010.803, pp. 856-859., 2010 | Malathi. A and Dr. S. Santhosh Baboo, An Enhanced Algorithm to Predict a Future Crime | using Data Mining, *International Journal of Computer Applications* (0975 – 8887), Volume 21– No.1, May 2011, | R.Sujatha and D.Ezhilmaran, A Proposal for Analysis of Crime Based on Socio – | Economic Impact using Data Mining Techniques, *International Journal of Societal Applications of Computer Science*, Vol 2 Issue 2 February 2013, ISSN 2319 – 8443 | Anjum Mujawar and Vijay Patil , Fuzzy based data mining system for E-government, *Elixir Comp. Sci. & Engg.* 51A, 11066-11068, 2012. | Hanmant N. Renushe et al., Short term crime forecasting for prevention of crimes: A | study of Satara district, *Int. J. Comp. Tech. Appl.*, Vol 2 (3), pp. 608-611, ISSN:2229-6093, 2012. |