

# Web Mining- Concepts and Application

S.Vidya, K.Banumathy

Department of computer science, Head of the Department  
KSG College of arts and science, Coimbatore, India

**Abstract-**Web mining is the use of data mining techniques to automatically discover and extract information from web. web mining helps to solve the problem of discovering how users are using a websites. It involves mining logs or log analysis and get meaning full data from web logs. web mining is the application of data mining techniques to web data where at least one of structure or usage data is used in the mining process.The present paper deal with primary discussion of web mining.And also help to future research and extend from this paper.

## I.INTRODUCTION

Web mining is the application of data mining techniques to extract knowledge from web data, including web documents,hyperlinks between documents usage of web sites.The web is the interesting area of research.Its help to extract knowledge from web data.In which atleast one of structure or usage data is used in the mining process. according to analysis target, web mining can dividved into three different types.

## II.REASONS FOR WEB MINING:

In web area world wide web is act as a two side one is a user side and another one is an information provider.Both a sides are face problems while dealing with the web data.So Web Usage mining retrieve useful data.but there will be many copies of the same useful data available.So Web usage mining makes use of SOM model cluster only the similar data and eliminate redundancy.Self Organizing Map(SOM) is one of the unsupervised learning method in the family of artificial neural network(ANN) and it's also used in web usage mining for getting similar data and avoid redundancy.

### A. A User Problem:

**Finding Relevant information**→ Users browse or use the search service to find a relevant information from the web.we face a two problems here.

- **Low Precision:**We get an irrelevant information from various search result.
- **Low Recall:**We get an relavant information from the web,but the pages are not properly indexed.

### B. The Information Providers Problem:

- What do customers do?
- what do the customers want?
- How effectively use the web data to market products and service to the customers?

## III. WEB MINING:

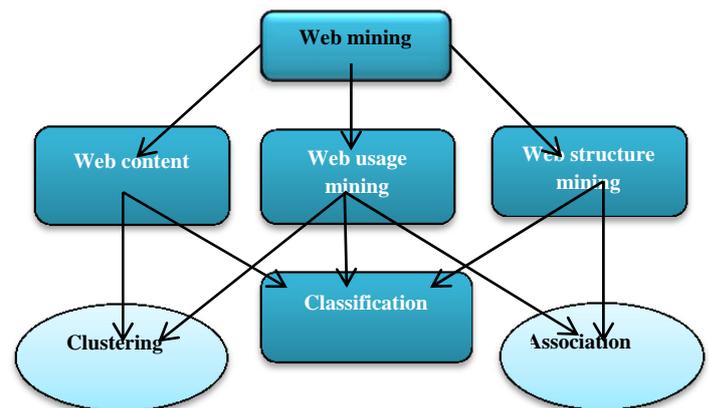


Fig.1 Web mining classification

## IV. WEB MINING

Web mining can be broadly divided into three distinct categories,according to that data to be mined.

### A. Web Content Mining:

Web content mining is the process of extracting useful information from the content of the web documents. Content data is the collection of facts a web page is designed to contain. It can provide useful and interesting patterns about user needs and contribution behaviour.Web content mining also known as text mining ,is generally the second step in web data mining .content mining is the scanning and minining it consists of images, text, audio, video or structured records such as lists and tables, application of text mining to web content has been the most widely researched. Issues addressed in the text mining include topic discovery and tracking, extracting associated patterns, clustering of web documents and classification of web pages.web content mining also distinguishes personel home pages with other web pages.Research in web content mining encompasses resource discovery from the web, document categorization and clustering, and information extraction from web pages.

### B. Web Structuring Mining:

The structure of typical web graph consists of web pages as nodes, and hyperlinks as edges connecting related pages.Technically Web content mining mainly focus on the structure of inner document but web structure mining tries to discover the link structure of the hyperlinks at the inter

document level based on the topology of the hyperlinks and generate the information, such as the similarity and relationship between different web sites.

Web structure mining is the process of discovering structure information from the web. This can be further divided into two kinds based on the kind of structure information used.

#### **Hyperlinks:**

A hyperlink is a structural unit that connects a location in the web page, either within the same web page or on a different web page.

A hyperlink that connects to a different part of the same page is called as intra-document hyperlink and the hyperlink that connects two different pages is called as inter document hyperlink.

Document structure In addition, within a web page can also be organised in tree structured format, based on various HTML and XML tags within the page.

Mining efforts here have focused on automatically extracting document object model (DOM) structures out of the document.

#### **C. Web Usage Mining:**

**Web usage mining** is the application of data mining techniques to discover interesting usage patterns from the web usage data in order to understand and better serve the needs of web-based application origin of web users along with their browsing behaviour at a web site.

Web usage mining itself can be classified further depending on the kind of usage data considered: web server data user log are collected by the web server and typically include IP address, page reference and access time.

### **V. WEB MINING BENEFICIAL AREAS:**

Application of web mining is connected with the rapid growth of world wide web, web mining becomes a very hot and popular topic in web research area. web mining it also plays an important role for E-Commerce and E-Service web site to understand their web sites and services are used and provide better service for both customers and users.

Few applications are:

#### **A. E-Learning:**

Web mining can be used for improving and enhancing the process of E-learning environments. Applications of web mining to e-learning are usually web usage based. Machine learning techniques and web usage mining enhance web-based learning environments.

#### **B. Digital libraries:**

Digital libraries services provide precious information distributed all around world, eliminating the necessity to be physically present at different libraries in different parts of world.

#### **C. E-Government:**

Organisations that interact with the citizen of the country lead to better social services. The main characteristics of the e-government systems is related to the use of technology to deliver services electronically, focusing on the citizen needs

by providing better information and enhanced service in the support of government. E-government system may provide customized services to citizen resulting in user satisfaction and quality of services and support in citizen decision making, which leads to social benefits.

#### **D. Electronic commerce:**

A major challenge e-commerce is to understand visitors or customers' needs and to value orientations as such as possible. It can improve capacity of service for consumer and competitive advantages.

#### **E. E-Politics and E-Democracy:**

E-politics provides political information and politics on demands to the citizen improving political transparency and democracy. Election information, parties, members of parliament, members of local government on the web are part of e-politic services. Despite the importance of e-politics in democracy there is limited web mining methods to meet citizen needs.

#### **F. Security and Crime Investigation:**

Web mining techniques are also used for protection of user system or logging information against such cybercrimes as hacking, internet fraud, fraudulent websites, illegal online gambling, virus spreading, child pornography distribution and cyber terrorism. Clustering and classification techniques of web mining can reveal identities of cyber criminals whereas neural network, decision trees, genetic algorithm and support vector machines can be used to trace criminal patterns and network visualization on websites.

#### **G. Electronic Business:**

Web mining techniques can support a web enabled electronic business to improve on marketing, customer support and sales operations.

### **V. CONCLUSION:**

Web and web usage continues to grow, so too grows the opportunity to analyse web data and extract all manner of useful knowledge from it. In this paper briefly described the web mining and applications of web mining some areas of future research. It can also be used to provide fast and efficient services for business. It is expected that more applications of web mining will be developed.

### **FUTURE SCOPE**

Research activities on this topic have drawn heavily on techniques developed in other disciplines such as Research activities on this topic have drawn heavily on techniques developed in other disciplines such as Information Retrieval (IR) and Natural Language Processing (NLP). While SRIVASTAVA, DESIKAN and KUMAR [40] there exists significant body of work in extracting knowledge from images in the field of Image Processing and computer vision, the application of these techniques to web content mining has been limited.

## REFERENCE

- [1] Zdravko Markov ,Daniel T.Larose “*Webcontent structure and Usage*”,Wiley,2007.
- [2] V.Bharanipriya & V.Kakakshi Prasad “*WEB CONTENT MINING TOOLS:A COMPARITIVE STUDY*”.
- [3] Rekha Jain and DR G.N. “Purohit page ranking algorithm for web mining.*International journal of computer applications*”,(0975 .8887 volume13.No.5,January 2011.
- [4] Wang and Liu 1998;Moh, Lim and Ng 2000
- [5] Srivastava, Cooley, Deshpande, and Tan 2000
- [6] JaideepSrivastava\_y,Robert cooleyz,Mukund deshpande and pang-ning tan,web usage mining:discovery and applications of usage patterns from web data”published in ACM SIGKDD Explorations.copyrightc 2000 ACM SIG KDD,Volume I,Issue 2,jan 2000,pp. 12-23.
- [7] Dr.A.C.Mondal and Sourav Maitra,”A Study of web mining Research- last few years and the road Ahead”publish in ICCS,Burdwan University 2010.
- [8] Wangbin Hu.Junpeng Yuan and Yuantao Song,”The Research of a web mining method in Research Areas”published in Sixth Wahon international center on E-Business,e-Business Track.
- [9] Kosala,Raymond;Hendrik Blockeel,”Wen Mining Research :A Survey” SIGKDD Explorations.
- [10] Mustapasa,A.Karahoca,D.Krahoca and H.Uzunboylu, “Hello World,Webmining for E-Learning”,Procedia. Computer science vol3,2011.