WEB USAGE MINING AND DATA MINING TECHNIQUES

Pankaj Dhumane[1] and Vijaya Kamble[2]

[1]Assistant Professor, Department Of Computer Science, Sardar Patel Mahavidyalaya, Chandrapur
Email: pdhumane@rediffmail.com
[2]Lecturer, Department of Computer Application, Sardar Patel Mahavidyalaya, Chandrapur
Email: sairamvijaya@gmail.com

ABSTRACT

Since in our day to day life is lot of information is available online. The web mining is a fertile area of research where data mining can be applied. It has large computational and research work. It requires lots of information retrieval techniques. The web mining can be categorized into three area, web content mining, web structure mining and web usage mining. This paper presents some techniques of web usage mining.

***Keywords*:** Data mining, web mining, web usage mining, web structure mining

INTRODUCTION

In the world of information on web some tools are must to extract title and get the needed information .Since many organizations have invested in web information, it is very much necessary to find the interest of user. So it is very much necessary to find the track the detail on client terminal to server side. It needs a system to keep track all details and keep knowledge and find users access pattern .User retrieve information from the web. The services that are provided by the organization to customer are online shopping, resource material, distance education, online collaboration, news broadcasts. The www is becoming everyday activity of every people from child sharing music files with friends to senior receiving photograph and message from grandchild. We can also see web pages for courses for all fields in university and college. There are several issues to be considered regarding the web log and the processing should be done on server side data collection. Then the information can be gathered for user registration details, session details, transaction detail, site topologies and study of user behaviors and structure. Information retrieval is very important process of data mining and knowledge discovery process. The quality of knowledge depends on the data mined quality. The data mining techniques and algorithms are applied to mine the data. The success depends on data mining algorithms applied algorithms should be applied. There for appropriate selection algorithms should be applied .The main aim is to select necessary features from the data to be mined. Feature selection can be visualized as a selection of subset of features from given phases. A set of labeled is

determined from given set of samples for feature selection,. But it is very difficulties areas when the sample is not properly labeled. The feature selection is very important in the phase of knowledge discovery. There for how to select the best features is regarded as key features in machine learning. Further more among the different categories of feature selection the very specialized categories are feature selection in genetic algorithms. The genetic based features selection has become very important .The reason behind it is suppose there are m numbers of features in the data being mined. Then the total numbers of candidate features subset is 2m that is the size of search space of feature selection grows exponentially with the the number of features. The genetic algorithm is biologically inspired and has many mechanisms mimicking natural evaluation. It has a great deal of potentiality in scientific and engineering optimization on search problems. The research done by siedleck has find out the work done in the field of genetic algorithms and compared to the representative level of classical algorithms. How ever genetic algorithms does not guarantee the optimal solution and they are strongly found out by the the value of size n in the population. For the good results genetic algorithms has to pay high computational cost. In contrast parallel genetic algorithm can keep the quality of results high and find them fast result because of processing mechanism in parallel and less time. This keeps the confidence factor high and the response time low , opening opportunities to apply genetic algorithms in time constrained application Additionally parallel genetic algorithms may find difficult solution which can be independently compared.

Useful Tasks in Data Mining

Due to flourishing in web user and e-commerce, it has become very important to understand the online customer behavior. There is huge amount of work and effort is needed to do this. Data mining algorithms and the web usage mining can do this. Statistical method for analysis in web log to provide and help in calculating the frequency average and session time. These has been used some data mining techniques adapted for web usage mining. The techniques are association rules for mining, clustering and classification, sequential pattern analysis and dependency modeling and prediction. These techniques are used for various purposes. The purpose can be web improvement, web tracking ,networking, site update, marketing and intelligent performance. Association rules is used for discovery of relationship between items intersection. It is typically used for market basket analysis ie to find out which items being buyer by the customer. Clustering is an unsupervised grouping of object. While classification is supervised grouping of object. Sequentially pattern analysis is similar to association rules but takes in to account the sequence of events. In the other words that if a page d is required before page f and the pattern is discovered. The various data mining techniques has been adapted for knowledge discovery and also for web mining.

Web Data Mining

The www serves a huge widely distributed global information service center for news advertisement, consumer, information, financial management, education, ecommerce etc. The web also contains a rich and dynamic collection of hyperlink, providing rich source for data mining and web mining. Web mining can be categories into following sub tasks.

1. Resource finding; the tasks of retrieving intended wed document

2. Information selection and preprocessing: automatically selecting and pre-processing specific information from retrieved web resource
3. Generalization: automatically discover general pattern at individual web site as well as multiple web site
4. Analysis: Validation and Interpretation of the mined pattern

Knowledge Discovery is an very important step from web mining. This needs validation and interpretation as on step 4. So an interactive query is being triggered and data is being processed to give final output. The Web Mining has to face the challenges such as the complicity of web page is far greater. The Web servers a broad diversity of information on the web is truly relevant and useful. However web mining refers to overall process of discovering useful information or knowledge from web data. IT implicitly covers the standard process of knowledge discovery .Web mining tasks can be classified into three categories, web content mining, web structure mining and web usage mining.

Web Content Mining

Web content mining is the process of automatic search of information resource available online and involves mining wed content data. In web mining domains, web content mining essentially is an analog of data mining techniques from relational database, since it is possible to find similar type of knowledge from the unstructured data. Similar type of knowledge from unstructured data residing in web documents. The web documents usually contains server type data such as text , images, audio, video, metadata and hyperlink. Some are semi structured and html documents, more structured data in table or database generated in html page.. The unstructured characteristic of web data forces web content mining towards more complicated approach.

Web Structure Mining

Most of the Web Information retrieval tools uses the textual information, while ignore the link information that could be valuable. The goal of web structure mining is to generate structural summery about the web site and web pages. Technically web content mining focuses on the structure of inner documents while web structure mining tries to discover the link structure of hyperlink at the inner document level.

Web structure mining will be categories the web pages and generate the information such as the similarity and the relationship between the web sites. Web structure mining can also have direction and discovering the structure of web documents itself.

This type of structure mining can be used to reveal the structure of web pages. This would be good for navigation purpose and make it possible to compare integrate web page schemes. The structure information generated from web structure mining includes the following: The information measuring the frequency of the local in the web tuples in a web table containing links that are interior and the links that are containing links that are interior and link that are written within same document.

Web Usage mining

Web usage mining tries to discover the useful information from the secondary data derived from the interaction of users while surfing the web. It focuses on the techniques that could predict user's behaviors while the user interacts with web. M Spilious abstract the potential strategies aims in each users behaviors within the site , comparison between expected and actual web site usage adjustment of web site to the interest of its user. There are no definite destinations between the web usage mining and other two categories. In the process if data presentation of web usage mining, the web site topology will as inform sources, which interact web usage mining with content mining and web structure mining moreover, the clustering process of pattern discovery is a bridge to web content and structure mining from usage mining. There is lot of work have been done in the IR, database, Intelligent Agents, web contents. Web structure mining web usage mining is a relatively new research area and gains more and more attention in recent years.

CONCLUSION

Web mining is the fastest growing research area. We surfaced the research in the area of web data mining .Three recognized types of web content mining is related but different from data mining and text mining. Web data are mainly semi-structured and unstructured. Web content mining requires creative application of data mining and text mining techniques and also unique approach. Due to the heterogeneity and lack of structure of web data automatically discovered targeted and unexpected knowledge information still present many challenging research problem.

REFERENCES

1. G. Salton and M.Mc.Gill "Introduction to Mordern Information Retrival". Mc Graw Hill
2. U.Fayyad . G. P. Dantesky-shappira. P. Smith"Data Mining To Knowledge discovery"
3. R. Kosala .H.Blocked" Web Mining Reasearch: A Survey"
4. J. Srivastava .R.Cooley . M.Deaspandey"P.Tan Web Usage mining"
5. J.Hann and M, Kamber"Data Mining Concept and Tecaniques