Journal of Advances and Scholarly Researches in Allied Education
Vol. 15, Issue No. 9, October-2018, ISSN 2230-7540

Web Mining Opportunities and Privacy Challenges

Vipula Vinaykumar Mahindrakar[1]* Dr. Bechoo Lal[2]

[1] Teaching Assistant, Karnataka College

[2] Pebble Hills University

Abstract – This paper is a work on survey on the existing techniques of web mining and the issues related to it. The World Wide Web acts as an interactive and popular way to transfer information. Due to the enormous and diverse information on the web, the users cannot make use of the information very effectively and easily. Data mining concentrates on non-trivial extraction of implicit previously unknown and potential useful information from the very large amount of data. Web mining is an application of data mining which has become an important area of research due to vast amount of World Wide Web services in recent years. The aim of this paper is to provide the past and current techniques in Web Mining. The research work done by different users depicting the pros and cons are discussed. It also gives the overview of development in research of web mining and some important research issues related to it.

The web mining introduces unique computational and statistical challenges, including scalability and storage bottleneck, noise accumulation, spurious correlation and measurement errors. These challenges are distinguished and require new computational and statistical paradigm. This paper presents the literature review about the web Mining and the Opportunities and Privacy challenges with emphasis on the distinguished features of web mining. It also discusses some methods to deal with web mining.

Keywords: Web Mining, Opportunities, Privacy Challenges, techniques, World Wide Web, information, effectively, application, important, services, etc.

----------X----------

INTRODUCTION

Web is a collection of billions of documents. The web is very enormous, diverse, flexible, and dynamic. The World Wide Web continues to grow both in the huge volume of traffic and the size and complexity of Web sites. It is difficult to identify the relevant information present in the web. Most of the contents in the web are unstructured in nature, but very little work deals with unstructured and heterogeneous information on the Web. The emerging field of web mining aims at finding and extracting relevant information that is hidden in Web related data, in particular in text documents published on the Web. Data Mining involves the concept of extraction meaningful and valuable information from large volume of data. Web mining is an important area in data mining where we extract the interesting patterns from the contents. Based on these kinds of information the Web Mining consists of 3 processes namely Web Content Mining, Web structure Mining and Web Usage Mining as shown in fig1. Web content mining deals with the raw data that is available on the web. The web structure mining mainly deals with the structure of the web sites (Chidansh & Kankanhalli, 2010). Web Usage mining involves mining the usage characteristics of the users of Web applications. It is in a semi structured format so that it needs lots of pre-processing and parsing before the actual extraction of the required information. This paper gives the survey of web mining techniques. Data mining process consist of several stages namely (Wenguo, 2008) Domain Understanding, Data selection, Data pre-processing and cleaning, Pattern discovery, Interpretation and Reporting.

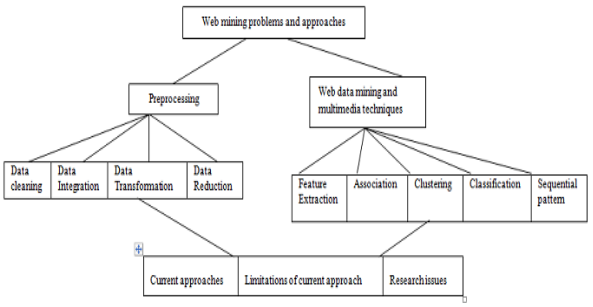

Fig 1- Web Mining Techniques

REVIEW OF LITERATURE:

Web mining is a technique in data mining that automatically retrieves extracts and analyzes the information from web. Yang et al, (2012) [3] discuss about the various issues to be addressed in data mining. The major issues include Automated Data Cleaning, Over Fitting, Under Fitting and Oversampling of data, Scaling up for high dimensional data, Mining sequence and time series data. A poll was conducted and given by k d nuggets and many of the researchers suggested the important work for research as Scaling up Data Mining algorithms for huge data, mining text and automated data cleansing as the major issues discussed with highest priorities (Vellingiri & Pandian, 2011). Other issues include dealing with unbalanced data, mining data streams, link and networks. Security in mining and distributed data mining also caught the significance but not to as greater extent. A hotly debated technical issue is whether it is better to set up a relational database structure or a multidimensional one. Finally, there is the issue of price.

Challenges to Security and Privacy in web mining: The monstrous maintenance of financial, demographic, behavioral, monetary, and other value-based information for expository purposes may prompt the disintegration of common freedoms because of lost security and individual self-rule. From a protection and security point of view, the test is to guarantee that information subjects (i.e., people) have supportable control over their information, to anticipate abuse and mishandle by information controllers (i.e., enormous information holders and other outsiders), while saving information utility, i.e., the estimation of huge information for learning/designs revelation, advancement and financial development. The accompanying areas depict some applicable difficulties to security and protection with regards to enormous information.

Big information Challenges: While the ascent of enormous information yields colossal open doors for people, associations and the general public everywhere, it additionally raises vital protection and moral issues (Oard, et. al., 2008). These issues are elements that may prompt circumstances in which the basic scientific models and frameworks are liable to effect protection adversely from both a lawful and a moral point of view, and consequently speak to conceivable hindrances for the huge information's capability to be completely figured it out.

Web mining is Reshaping Medicine and Health Care: As all parts of social insurance (counting general wellbeing checking, human services conveyance and examination) turn out to be increasingly reliant on data innovation, partners in the medicinal services industry and wellbeing financial aspects are progressively ready to gather, process and share different sorts of information efficiently, including individual related organic specimens, therapeutic imaging information, understanding cases, solutions, clinical notes, and other restorative insights. In actuality, gathering, preparing, and sharing this sort of information is just about as old as clinical drug. What makes this a more prominent theme of enthusiasm for the huge information age, in any case, is that social insurance investigators and experts can now:

(i) Consolidate conventional wellbeing information with outside information - demographic, behavioral and therapeutic/wellness related sensor information - so as to gather bits of knowledge into human exercises and connections, and afterward

(ii) Influence these experiences to enhance restorative research, find and screen generally imperceptible wellbeing pat-terns in a vast part of the populace, or to give new inventive customized medicinal items and administrations.

Web mining and Financial Services: Money related foundations are progressively benefiting from late advancements in the field of IT and enormous information related apparatuses. They utilize these advances to gather and dissect gigantic measures of individual, monetary, and money related information, some of which are ongoing streams (e.g. those from stock and money related markets), keeping in mind the end goal to better comprehend and control the perplexing consistence challenges and budgetary dangers connected with conceivable new speculations (Schafer, et. al., 2001). As of late, credit offices and safety net providers have gotten to be energetic to gain by re-penny advancements in the field of IT and the simple access to person to person communication information to break down years of value-based information reflectively as they look to distinguish exceptionally complex pat-terns, which they can use for extortion location Brokerage firms' developing capacity to evaluate vast quantities of conceivable business sector situations, examine new sorts and wellsprings of information (e.g. breaking news and climate data, continuous sub-prime business sector information, online networking) may permit them to coax out possibly significant examples that would some way or another remain shrouded lair. They can then utilize those bits of knowledge to foresee securities exchange exhibitions and enhance exchanging choices. One illustrative case of this computational way to deal with securities exchange is high-recurrence stock exchanging: a developing type of exchanging that depends completely on fast PCs and smart calculations to settle on exact exchanging choices at rates measured in the request of milliseconds. Additionally, whole business sections are progressively depend loaning choices or overseeing dangers connected with client installments on the web (Michael & Motilal, 2008).

Non-bank moneylenders, specifically, are relied upon to apply progressed investigation progressively on constant aggregations of cross-space information keeping in mind the end goal to pick up knowledge into customer conduct, recognize conceivably suspicious clients exercises, and subsequently settle on exact loaning choices with an exactness to a great extent thought inconceivable only a couple of years prior.. Summing up, enormous information driven monetary administrations can possibly add to more prominent money related incorporation which thusly is imperative for documenting comprehensive financial development.

CONCLUSION:

In this paper we have discussed about the research issues and the drawbacks of the existing techniques. More research work need to be done on the web mining domain as it will rule the web in the near future. Web mining along with semantic web known as semantic web mining is to be concentrated that is evolving which helps us to overcome the cons of web mining. Though various algorithms and techniques have been proposed still work has to be done in discovering new tools to mine the web. Web mining is suitable an active concerning field of research because of its potential marketable benefits. It is further possible to analyze the visitor's performance by linking the Web logs with cookies and forms, and which could help e-services site to address several business questions. Its awareness in analyzing user's actions on the web after discovering access logs made its fame very rapidly specially in E-services areas. Details like user log files, request for resources etc. are uphold in web servers, which is the core mining area of web usage. The study of these gives the user browsing case and that can be utilized for target advertisement, improvement of web design, satisfaction of clientele and making market analysis. Most of the e-service suppliers understand the fact that they can apply this tool to keep their clientele as the web and its usage continues to grow, so too grows the chance to analyze web data and remove all manner of useful knowledge from it.

REFERENCES:

[1] Chidansh Amitkumar Bhatt, Mohan S. Kankanhalli (2010). "Multimedia Data Mining: State Of The Art And Challenges" Published Online: 16 November 2010© Springer Science+Business Media, LLC 2010.

[2] Wenguo Wu (2008). "Study On Web Mining Algorithm Based On Usage Mining", Computer- Aided Industrial Design And Conceptual Design, 2008. CAID/CD 2008. 9th International Conference on 22-25 Nov.2008.

[3] J. Shao, X. He, C. Bohm, Q. Yang, C. Plant (2012). "Synchronization-Inspired Partitioning and Hierarchical Clustering," IEEE Transactions on Knowledge and Data Engineering.

[4] J. Vellingiri, S. Chenthur Pandian (2011). "A Survey on Web Usage Mining", Global Journal of Computer Science and Technology .Volume 11 Issue 4 Version 1.0 March 2011.

[5] Oard D., He D., and Wang J. (2008). "User-assisted query translation for interactive cross- language Information retrieval" International Journal of Information Processing and Management, vol.44, no. 1, pp. 181-211.

[6] J. B. Schafer, J. A. Konstan, J. Riedl (2001). E-commerce recommendation applications. Data Mining and Knowledge Discovery, 2001(5): pp. 115-153.

[7] Michael C. Shapiro, Motilal Banarsidass (2008). "A Primer of Modern Standard Hindi", Publishers Private Limited, 2008 Reprint

Corresponding Author

Vipula Vinaykumar Mahindrakar*

Teaching Assistant, Karnataka College