

Web Mining: It's Application to Web Search & Tracking Changes in the Web

Vipula Vinaykumar Mahindrakar^{1*} Prof. Dr. Bechoo Lal²

¹Teaching Assistant, Karnataka College

²Assistant Prof at Western College, University of Mumbai

Abstract – Web mining is the application of data mining. It is a process to remove knowledge from web data, including web documents, hyperlinks between credentials, and practice logs of web sites. A panel prearranged at ICTAI 1997 asked the question “Is there something different about web mining?” As no eventual conclusions were reached then the marvelous notice on web mining in the past five years, and a number of noteworthy ideas that have been developed. It has certainly answered this study in the assenting in a big way. In accumulation a quite steady community of researchers paying notice in the area has been formed, mostly through the successful series of Web KDD workshops, which have been held yearly in conjunction with the ACM SIGKDD Conference since 1999 and the web analytics workshops, which have been held in conjunction with the SIAM data mining conference. The World Wide Web has multiplicity of knowledge service centers, like news sites, encyclopedias, education sites, ecommerce etc. So the information in WWW is extended widely in theses information hub worldwide. To recover from these disseminated storage areas is a quite hard process and it necessary requires a competent tool to find the preferred information. Only a clever system which effectually mine for knowledge can decide these troubles. The subsequent factors made it hard for an effectual data warehousing and data mining.

Keywords: Web Mining, Application, Web Search, Tracking, Changes, Web, process, knowledge, workshops, data mining, World Wide Web, information, data warehousing, etc.

----- X -----

INTRODUCTION

Recover information from World Wide Web is a boring assignment since the expansion in the ease of use of knowledge backup supply on it. So this raises the need to utilize a clever system to recover the information from World Wide Web. The way in which Web information of getting back and Web base data warehousing are boosted with the removal of facts from the Web using web mining tools (Ning and Yuan, 2011. Palmer, 2012. Ventura *et. al.*, 2010. Ebrahim and Irani, 2005). Web usage mining is one of the best developing areas of web mining. Its notice in analyze users recital on the web after exploring right to use logs made its fame very quickly in Eservices areas. Most of the e-service providers realized the fact that they can relate this tool to keep hold of their clientele. This study tries to provide an insight into web mining and the different areas of web mining. Then it spotlight on Web usage mining, its application and collision in E-services.

As the web and its usage continues to grow, so too grows the opportunity to analyze web data and extract all manner of useful knowledge from it. The past five

years have seen the emergence of web mining as a rapidly growing area, due to the efforts of the research community as well as various organizations that are practicing it.

REVIEW OF LITERATURE:

In this chapter we follow the data-centric view of web mining which is defined as follows, Web mining is the application of data mining techniques to extract knowledge from web data, i.e. web content, web structure, and web usage data.

- The massive size of the web
- No proper arrangement for the web documents.
- The active environment of the information source
- The variety in usage and consumer society

However these are confront that stand as the powerful strength for the research into competent and effectual discovery and use of property on the Internet. The basic distinctiveness of web make us have to think about to form and get longer the unfashionable research methodologies. Two dissimilar approaches were taken in originally important web mining. First was a "process-centric view," which define web mining as a series of task (Chakraverty *et. al.*, 2012). Second was a "data-centric view," defined web mining as a stipulations of the types of web data that was being worn in the mining process. The second meaning has become more satisfactory, as is apparent from the approach adopted in most studys that have physically location the issue. Web Mining is the function of data mining method to remove information from web data, i.e. web content, web structure, and web usage data (Rani and Chakraverty, 2012).

Process Mining: Mining of market storage bin data, scatterbrained at the point-of-sale in any shop, has been one of the noticeable successes of data mining. Yet, this data offers only the end result of the process, and that too judgments that ended up in product purchase. Clickstream data make available the opportunity for a complete look at the decision making process itself, and information extracted from it can be used for optimizing, manipulate the process, etc. (Nasraoui *et. al.*, 2008) has decisively established the value of process information in recognizing users' performance in conventional shops. Research requires to be carried out in

- (1) Remove process models from usage data,
- (2) Accepting how different parts of the process model collision various web metrics of interest,
- (3) How the process models change in reply to various changes that are made, i.e. changing stimuli to the user. Figure 1 shows an approach of modeling online shopping as a state evolution figure.

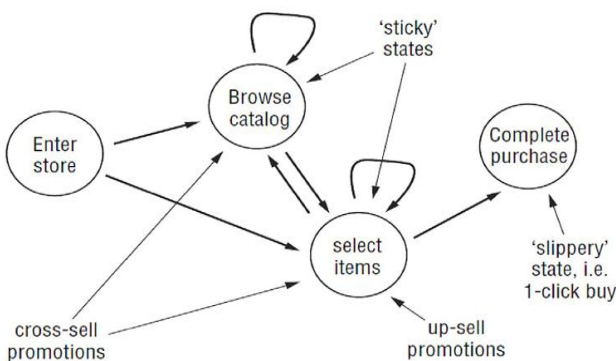


Figure 1- Shopping pipeline model as state evolution figure

In general goal:

- Make the most of probability of achieving final state
- Maximize predictable sales from every visit

Chronological Transition of the Web: People's interface with the web is altering the web as well as the way public cooperate with each other. As storing the history all of this communication in one place is obviously too amazing a task, at least the altering to the web is being evidence by the pioneering internet archive project. Research requires to be carried out in remove temporal models of how web contented, web structures, web communities, establishment, hubs, etc. (Fengrong, 2004). evolve over time. Large organizations usually archive usage data from their web sites. With these sources of data accessible; there is a large scope of research to expand techniques for scrutinizing of how the web changes over time

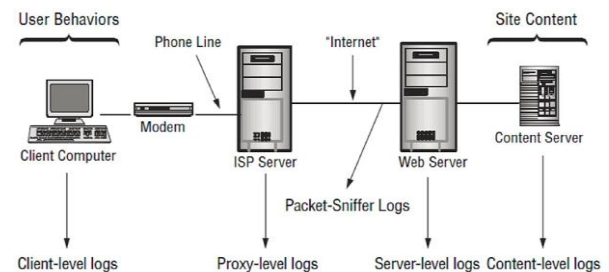


Figure 2- Elevated level architecture of dissimilar web logs

Privacy and Web Mining: Although there are so many benefits to be expanded from web mining. A clear disadvantage is the possible for severe infringement of privacy. Public approach towards privacy give the feeling to be nearly schizophrenic, i.e. people say one thing and do quite the differing have confirmed that people were eager to provide fairly personal information about themselves, which was completely inappropriate to the task at hands if provided the right motivation to do so. Also, clearly bringing notice to information privacy policies had virtually no effect (Lee *et. al.*, 2011. Webber, 2010. Nagarathinam and Saraswathi 2011). One explanation of this apparently contradictory attitude towards privacy may be that we have a bi-modal view of privacy, namely that "I'd be willing to share information about myself as long as I get some (tangible or intangible) benefits from it, and as long as there is an implicit guarantee that the information will not be abused." The examining issue generated by this attitude is the need to develop move toward, methodologies and tools that can be used to verify and authenticate that a web service is certainly using

user's information in a manner reliable with its stated policies.

CONCLUSION:

In this study, a study on Web mining has given with research point of view. Misperceptions regarding the usage of the term Web mining is elucidated and discussed briefly about web mining categories and various approaches. In this survey, we focus on representation issues, various techniques of web usage mining and web structure mining and information retrieval and extraction issues in web content mining, and connection between the web content mining and web structure mining.

The hasty growth of the web is causing the stable growth of information, leading to several problems such as an increased difficulty of extracting potentially useful knowledge. The huge amount of information available online, the World Wide Web is a fertile area for web mining research. The research in web mining aims to develop new techniques to effectively extract and mine useful knowledge or information from these web pages. Due to the heterogeneity and lack of structure of Web data, automated discovery of targeted or unexpected knowledge/information is a challenging task. In this study, we survey the research in the area of Web mining, point out the categories of Web mining and variety of techniques used in those categories. In this study we elicit research scope in the areas of web usage mining, web content mining, web structure mining and concluded this study with a brief discussion on data managing, querying, representation issues.

REFERENCES:

- A. Nagarathinam and Dr. S. Saraswathi (2011). "State of Art: Cross Lingual Information Retrieval System for Indian Languages" *International Journal of Computer Applications* (0975 – 8887) Volume 35– No.13.
- A. S. Chakraverty, B. G. Rani, C. B. Singla and D. D. Anand (2012). Experience based recommendations system for E-governance.
- Bart C. Palmer (2012). *Web Usage Mining: Application to an online educational digital library service*; Digital Commons@USU.
- Chu Hue Lee, Yo Lung Lo, Yu Hsiang Fu (2011). A novel prediction model based on hierarchical characteristic of web site; *Elsevier*; Volume 38 Issue 4, Pages 3422 – 3430.
- G. Rani and S. Chakraverty (2012). Boosting Interactivity of EGovernancell, *International*

Conference on Communication Languages and Signal Processing- with Preference to 4 G TechnologiesII, ICCLSP 4G.

- Jin Fengrong (2004). *Study of Web Usage Mining and Discovery of Browse Interest*. Master's Degree Thesis of Beijing Science and Technology University.
- Nasraoui O., Soliman M., Saka E., Badia A., Germain R. (2008). "A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites," *Knowledge and Data Engineering, IEEE Transactions on*, vol.20, no.2, pp. 202-215.
- Ning Bin, Lei Yuan (2011). *Research on Application of Web Mining in E-commerce*; *Advanced Materials Research - Scientific. Net*; Volume 403 – 408; Pages 1830 – 1833.
- Romero C. Ventura S., Pechenizky M., Baker R. S. (2010). *Handbook of educational data mining*; CRC Press.
- William Edward Webber (2010). "Measurement in Information Retrieval Evaluation" Department of Computer Science and Software Engineering The University of Melbourne Doctor of Philosophy, Web Source ww2.cs.mu.oz.au/~wew/wew-thesis-PhD.pdf
- Zakareya Ebrahim and Zahir Irani (2005). E-government adoption: architecture and barriersII *Emerald Business Process Management journal*, vol.II, No.5 2005, pp. 589-611.

Corresponding Author

Vipula Vinaykumar Mahindrakar*

Teaching Assistant, Karnataka College

E-Mail – operations@ima.edu.in