Abstract

Web usage mining (WUM) is one of the types of data mining method which is used for analysing web usage patterns with the help of users’ session and behavior. It is the technique to classify the web pages and internet users by taking into consideration the contents of the page and behavior of internet user in the past. Web mining extracts data accumulated in server access logs, referrer logs, agent logs, client-side cookies, user profile and Meta data. Usually the WUM techniques study the visitors’ browsing behavior to obtain interesting knowledge. There are existing different techniques for web usage mining. Those existing techniques have their own advantages and disadvantages. Here we present a survey on some of the existing web usage mining techniques. First technique that we discuss is based on Sequential Pattern Algorithm. Sequential mining is the process of applying data mining techniques to a sequential database for the purposes of discovering the correlation relationships that exist among anordered list of events. Sequential mining techniques is web usage mining technique, where the sequences of web page accesses made by different web users over a period of time, through a server, are recorded. Ford Lumban Gaol has proposed a web log sequential pattern mining using Apriori-all algorithm. Second technique that we discuss in this paper is based on distance between 2 sequences using no-Euclidean distance formula. Peiqian Liu and Wei Li have proposed an improved Ward’s method for web user clustering. They
have given a formula, to calculate distance between elements which is a no-Euclidean distance measure. Since it is not a Euclidean distance the output preserves the ordering of events. In the last section of this document we have proposed a new web mining algorithm for analysing usage pattern of users. Proposed algorithm is based on WebLog Sequential Pattern Algorithm[Ford Lumban Gaol] and DBS[Peiqian Liu and Wei Li 2011].

References

- Anuradha Yadav, Satbir Jain; Analyses of Web Usage Mining Techniques To Enhance the Capabilities of E-Learning Environment; 2011 IEEE
- Jinguang Liu & Roopa Datla, "Web Usage Mining -- Pattern Discovery and its applications"; 2013
- P. Nithya, Dr. P. Sumathi; Novel Pre-Processing Technique for Web Log Mining by Removing Global Noise and Web Robots; 2012 National Conference on Computing and Communication Systems (NCCCS)
- P. Nithya, Dr. P. Sumathi; A Survey on Web Usage Mining: Theory and Applications; IJCTA | July-August 2012
- Peiqian Liu, Wei Li; Navigation Pattern Discovery on Web Site Based on the Distance Between Sequences; 2011 IEEE
- Sisodia, D. S.; Verma, S; Webusage pattern analysis through web logs: A review; Computer Science and Software Engineering (JCSSE), 2012 International Joint Conference
- Varnagar, C. R.; Madhak, N. N.; Kodinariya, T. M.; Rathod, J. N. &apos; &apos; Webusagemining: A review on process, methods and techniques; Information Communication and Embedded Systems (ICICES), 2013 International Conference

Index Terms

Computer Science

Web Mining
Keywords
Web Usage Mining  Web Log sequential  no Euclidean distance formula