

Survey on Techniques for Improving user Navigation by Reorganizing Web Structure

Priyanka Dhas

Department of Computer Science and Engineering
Deogiri Institute of Engineering and Management Studies
Aurangabad, India

Sarika Solanke

Department of Computer Science and Engineering
Deogiri Institute of Engineering and Management Studies
Aurangabad, India

Abstract—The growing availability of information on the web has raised a challenging problem such as web based information system can satisfy itself to different user requirement with ultimate of personalization or web transformation and improving user navigation in accessing content of a website. This paper reviews for basics of web mining, various techniques and algorithms for improving user navigation by reorganizing website structure as per user's requirement

Keywords— User navigation, Website restructures, k-means, Weblogs.

I. INTRODUCTION

The prolonged flow of interaction between users and website is a beneficial source of information about users browsing pattern. On other side information related to user's topic of interest is usually scattered on web environment or on website [6]. As website is huge source of information it consists of user's access website at a time and may require different pages at same time or same user may access different pages at different time. To satisfy user we require making website or webbing environment intelligent. Hence modern web based information system assures improvement in navigation pattern of user in accessing contents available from website. Data mining is analytic process for extraction of hidden predictive information from large databases [11].

A. Web mining overview

Web mining is application of data mining technique to extract knowledge from web data. Web is collection of inter related files on one or more web server and web data can be web content, web structures and web usage data. According to usage of web data as a input in data mining process it is mainly divided into three domains namely, Web Content Mining (WCM), Web Usage Mining(WUM) and Web Structure Mining (WSM).

Web content mining is process of extracting useful information from the content of web documents. Web documents may consist of text, image, audio, video or structured record. Two types approach in content mining agent based approach and database based approach.

Web usage mining is extracting information according to user navigation and behavior patterns like time spent on pages, traversing path, client side cookies, metadata and number of clicks on pages. User access patterns called as web logs or profile. Through web usage mining we can

predict pages required to add, number of pages are useless and user's interest.

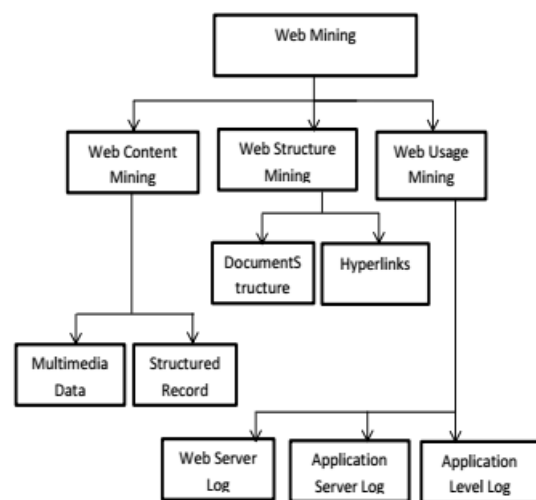


Fig.1 Web Mining Classification

Web structure mining generates structural summary about web sites and web pages. It discovers useful knowledge from link structure of the hyperlink which helps user to access website in the form of URL and navigate users. Web structure mining categorizes web pages and generates information.

Incentive for selecting web structure mining is declining interest of user due to website's inefficient navigation to page which are most required by users. Key reason for inefficient user navigation is poor website design as it's designed only with perception of web developer which can be different than user. Website's effectiveness is measured by user's satisfactions rather than developers. Hence web pages are linked in such a way that satisfies user's navigation pattern.

Mining of a web server log or improvement of user navigation can be done in two ways, 1) web personalization 2) web transformation. Web personalization is concerned with web logs that are user behavior, user profile sessions and history of data which is created by user's activity on web site. On other hand transformation focuses on developing methods to completely reorganize the link structure of website [15]. Studying link mining is preferred in this paper. We have focused on web transformation technique where website can be considered as a graph. In fig.2.1 website is shown in the

form of graph and pages are shown as nodes A, B, C, D and links between pages are edges of the graph through which user can navigate.

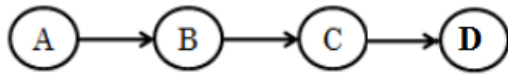


Fig. 2.1 Normal website structure

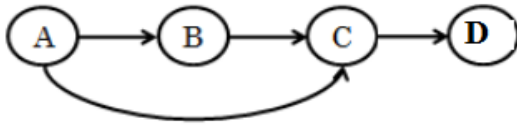


Fig.2.2 Reorganize website structure

If in weblogs we found user is accessing page 'C' frequently then website is reorganized by creating new link between page 'A' and page 'C' so user can access page 'C' in less time and clicks. Pages can be reorganized with two parameters first, in-links and out-links of web pages and second is users access pattern and traversing path of users. Our goal for reorganization is to provide required information to users within less time and clicks. Succeeding section we provide details of various techniques and clustering algorithms which can be used for improving user navigation [1].

II. LITERATURE REVIEW

Improving website structure can be done with various techniques. Links reorganization works on website structure. Data can be mined using two learning approaches supervised learning or unsupervised learning. Clustering plays vital role in data analysis and data mining application. Clustering can be done with number of different algorithms such as hierarchical, partitioning, grid and density based algorithms.

A Hierarchical Clustering Algorithm

Hierarchical clustering is connectivity based clustering algorithms. It is a method of cluster analysis which seeks to build a hierarchy of clusters. It does not partitions data in a single step rather it may start with single cluster including all object and may tend to number of clusters with few object in each.

This algorithm uses agglomerative and divisive method for partitioning [2].

B Density Based Algorithms

In density based algorithms clusters are defined as areas of higher density than the remaining of the data set. Major approaches of density based algorithms are 1) it fix density to training data points and measures density and connectivity in terms of local distribution of nearest neighbor objects. 2) it uses point in attribute space to fix density. It measures density with all objects density function. DBSCAN (Density Based Spatial Clustering of Applications with Noise), DENCLUE are some representative algorithms [2].

C Farthest First Traversal Clustering Algorithm

The strategy is farthest first traversal clustering algorithm. Before link mining they performed preprocessing using session-threshold and click-threshold where time spent on

page and number of clicks on links respectively and perform reorganization with minor changes in website[3]. Reconciling Website System which makes hit pages more accessible, interested links and connected links hence improves web navigation efficiency and offers reorganization of website [4].

D Latent Linkage and Cocitation Algorithm

To find relevant pages for given URL two algorithms are explained based on hyperlink analysis. First algorithm is on extended cocitation algorithm which were developed for scientific literature index and clustering and then extended to web page analysis. Second is latent talent linkage information (LLI) algorithm reveals deeper relationship among the pages and find out relevant pages more precisely and efficiently [7].

E K-means Clustering Algorithm

In clustering algorithm widely used partitioning algorithm is K-means clustering where objects are classified as they belong to one of K group. K-means is data mining algorithm which performs clustering of data samples uses iterative approach follows four steps: 1) Initialization 2) classification 3) centroid recalculation 4) convergence condition. Weighted page rank algorithm works on in links and out links and accordingly gives ranks to the website. It takes less execution time than K means algorithm [8].

F Farthest Point Heuristic Based Algorithm

K means algorithm exemplar is extended in k modes algorithm by using 1) modes instead of means 2) frequency based methods to minimize cost function of clustering 3) simple matching dissimilarity measure for categorical object. Farthest point heuristic based algorithm is initialization methods for k modes to improve efficiency of it. This method starts by arbitrary points. It is suitable and fast for large scale data mining application also overcome with minmax radius clustering problem of k centre clustering [9].

G Weighted Page Content Rank Algorithm

Weighted page content rank algorithm is based on structure mining and content mining which shows relevancy between pages so that query can be determined better than page rank and weighted page rank algorithm. It is used to give sorted order web pages returned as response to users query by search engine. It uses web structure mining to calculate importance of webpage whereas content mining works to get relevancy of returned pages with users query. [12].

H Clustering Technique for Improving Navigational Behaviors

Mathematical model suggested for improving user navigation is applied on static website which were having informative structure [10]. Navigational behavior of customer can be studied by analyzing the web logs. This can be done with web utilization manner which gives generalized sequence and aggregate tree as an output. Hence improvement is required or not decided by structure produce by WUM [13]. Concept based clustering approach is used by number of data mining techniques. Cluster formation uses some attributes which are differentiated according to types, attribute scale and proximity measures. Binary, discrete, continuous are attribute types. Quantitative-nominal, ordinal, ratios are different attribute scale and Min kowski metric

which abstraction of distance between points in Euclidean space is.

$$P_{ij} = \left(\sum_{k=1}^d |x_{ik} - x_{jk}|^r \right)^{1/r} \quad (1)$$

where, r is a parameter, d is the dimensionality of the data object, and x_{ik} and x_{jk} are, respectively, the k^{th} components of the i^{th} and j^{th} objects, x_i and x_j [14].

III. CONCLUSION

Website reorganization can be beneficial to user as well as owner as it can be done with various techniques which mainly focus on user's requirement. This paper surveys number of techniques to improve user navigation mainly focused on clustering algorithm also included minute details of web mining domain. We specifically studied and reviewed for web transformation. Future scope for the paper is to get detailed study for web personalization techniques which focuses on peculiar users log details or data produced through uses activity on web. This review on website reorganization can be helpful for web developers as well as researchers to understand features of website, process for improving user navigation.

REFERENCES

- [1] Deepshree A.vadeyar, "Study on Improving User Navigation by reorganizing web structure based on Link Mining "International Journal of Computer Science and Information Technologies, Vol. 5, 2014.
- [2] Amandeep Kaur Mann, Navneet Kaur "Survey Paper on Clustering Techniques" International Journal of Science, Engineering and Technology Research (IJSETR) ISSN: 2278 – 7798Vol. 2, Issue 4, April 2013.
- [3] Deepshree A.vadeyar, "Reorganization of Links to Improve User Navigation", cited on July 2014.
- [4] Deepshree A. Vadeyar, Yogish H.K, "Farthest First Clustering in Links Reorganization" International Journal of Web & Semantic Technology (IJWesT) Vol.5, No.3, July 2014.
- [5] Joy Shalom Sona, Asha Ambhaikar "Reconciling the Website Structure to Improve the Web Navigation Efficiency" International Journal of Advanced Research in Computer Engineering & Technology ISSN: 2278 – 1323 Vol. 1, Issue 4, June 2012.
- [6] Federico Michele Facca, Pier Luca Lanzi, "Mining interesting knowledge from weblogs: a survey" 0169-023 2004 Elsevier B.V.
- [7] Jingyu Hou and Yanchun Zhang, "Effectively Finding Relevant Web Pages from Linkage Information" IEEE transactions on knowledge and data engineering, Vol. 15, NO. 4, August 2003
- [8] Amar Singh, Navjot Kaur "To Improve the Convergence Rate of K-Means Clustering Over K-Means with Weighted Page Rank Algorithm" Vol. 3, Issue 8, August 2013 ISSN: 2277 128X.
- [9] Zengyou He, "Farthest-Point Heuristic based Initialization Methods for K-Modes Clustering" CoRR, abs/cs/0610043, 2006.
- [10] Min Chen, Young U. Ryu "Facilitating Effective User Navigation through Website Structure Improvement" IEEE transactions on knowledge and data engineering, Vol. 25, No. 3, MARCH 2013.
- [11] <http://www.thearling.com/text/dmwhite/dmwhite.html> ,cited on 12th march 2015.
- [12] Gurpreet Kaur, Shruti Aggarwal, "A Survey- Link Algorithm for Web Mining" International Journal of Computer Science & Communication Networks, Vol.3 (2), 2013.
- [13] Shweta Mohod, Prof. Vishal gangawane, "Data Mining to Facilitate Effective User Navigation and Improve Structure of a Website" ISSN: 2248-9622, Vol. 4, Issue 8(Version 5), August 2014.
- [14] Er. Arpit Gupta, Er.Ankit Gupta, Er. Amit Mishra, "Research Paper on Cluster Techniques of Data Variations" International Journal of Advance Technology & Engineering Research (IJATER) ISSN NO: 2250-3536 Vol. 1, Issue 1, November 2011.
- [15] Sergio Flesca, Sergio Greco, Andrea Tagarelli, Ester Zumpano, "Mining User Preferences, Page Content and Usage to Personalize Website Navigation" Springer Science. Published on 1 August 2005.