Developing a Web Recommendation System Based on Closed Sequential Patterns

Utpala Niranjan[1], R.B.V. Subramanyam[2], and V.Khanaa[3]

[1] Department of Information Technology, Rao & Naidu Engineering College, Ongole,
Andhra Pradesh, 523 001, India
`niranjan_utpala@yahoo.com`
[2] Department of Computer Science and Engineering, National Institute of Technology,
Warangal, Andhra Pradesh 506 004, India
`rbvs66@gmail.com`
[3] Department of Information Technology, Bharath University, Chennai,
Tamilnadu 600 020, India
`drvkannan62@yahoo.com`

Abstract. The proposed system is mainly based on mining closed sequential web access patterns. Initially, the PrefixSpan algorithm is employed on the preprocessed web server log data for mining sequential web access patterns. Subsequently, with the aid of post-pruning strategy, the closed sequential web access patterns are discovered from the complete set of sequential web access patterns. Then, a pattern tree, a compact representation of closed sequential patterns, is constructed from the discovered closed sequential web access patterns. The Patricia trie based data structure is used in the construction of the pattern tree. For a given user's web access sequence, the proposed system provides recommendations on the basis of the constructed pattern tree. The experimentation of the proposed system is performed using synthetic dataset and the performance of the proposed recommendation system is evaluated with precision, applicability and hit ratio

Keywords: Sequential pattern mining, Web Personalization, Web recommendation, Web server log data, Prefixspan, Sequential database, Web access pattern, Pattern tree.

1 Introduction

The development of data mining techniques has been centralized on discovering hidden data in an efficient way that is beneficial for corporate decision-makers [1, 2]. Sequential pattern mining is an important subject of data mining which is extensively applied in several areas [3]. In general, sequential pattern mining is defined as determining the complete set of frequent subsequences in a set of sequences [4, 8]. Each event is considered as a collection of items/itemset occurring at the same time [5]. Generally, all the transactions of a customer are collectively considered as a sequence, known as customer-sequence, where each transaction is denoted as an item set in that sequence and all the transactions are listed in a particular order in connection with the transaction-time [6].

The rest of the paper is organized as follows: Section 2 presents a brief review of the researches related to the proposed system. Section 3 details the proposed web recommendation system. The experimental results and performance evaluation of the proposed system are given in Section 4 and finally the conclusions are summed up in Section 5.

2   Review of Related Research

Numerous researches are available in the literature for web recommendation system using sequential pattern mining. Here, we present some of the researches related with closed sequential pattern mining along with web recommendation system based on sequential pattern mining.

Zhou. B *et al*. [9] have proposed an intelligent Web recommender system identified as SWARS (sequential Web access-based recommender system) that employs sequential access pattern mining. In the proposed system, CS-mine, an efficient sequential pattern mining algorithm was made use of to recognize frequent sequential Web access patterns. The access patterns were then stored in a compact tree structure (Pattern-tree) which was then employed for matching and generating Web links for recommendations. The performance of the proposed system was assessed on the basis of precision, satisfaction and applicability. An efficient sequential access pattern mining algorithm, called CSB-mine (Conditional Sequence Base mining algorithm) was presented by Baoyao Zhou *et al.*.[10] The presented CSB-mine algorithm was on the basis of conditional sequence bases of each frequent event which removes the need for constructing WAP-trees. This enhanced the efficiency of the mining process considerably in comparison to WAP-tree based mining algorithms, particularly when the value of support threshold becomes smaller and the database size gets larger. Cui Wei *et al.* [11] have presented a hybrid web personalization system which was on the basis of clustering and contiguous sequential patterns. Their system clustered log files to find out the basic architecture of websites, and for each cluster, they employed contiguous sequential pattern mining to optimize the topologies of websites further. They have presented two evaluating parameters to test the performance of our system. Zhenglu Yang *et al.* [8] have presented an efficient sequential mining algorithm (LAPIN_WEB: LAst Position INduction for WEB log), which is an extension of previous LAPIN algorithm to extract user access patterns from traversal path in Web logs. Web log mining system comprises of data preprocessing, sequential pattern mining and visualization. The experimental results and performance studies established that LAPIN_WEB was efficient and outplayed familiar PrefixSpan by up to an order of magnitude on real Web log datasets.

3   A Proposed Web Recommendation System Based on Sequential Patterns

Web Personalization is an application of data mining and machine learning techniques to build models of user behavior. It can be useful to the task of predicting user needs and adapting future interactions with the main aim of improved user satisfaction. A