Chapter 11

A Robust Biclustering Approach for Effective Web Personalization

H. Hannah Inbarani
Periyar University, India

K. Thangavel
Periyar University, India

ABSTRACT

Web recommendation or personalization could be viewed as a process that recommends the customized web presentations or predicts the tailored web contents to web users according to their specific need. The first step in intelligent web personalization is segmenting web log data into web user sessions for constructing user model. These segments are later used to recommend relevant URLs to old and new anonymous users of a web site. The knowledge discovery part can be executed offline by periodically mining new contents of the user access log files. The recommendation part is the online component of a usage-based personalization system. In this study, we propose a robust Biclustering algorithm to disclose the correlation that exists between users and pages. This chapter proposes a Robust Biclustering (RB) method based on constant values for integrating user clustering and page clustering techniques which is followed by a recommendation system that can respond to the users’ individual interests. To evaluate the effectiveness and efficiency of the recommendation, experiments are conducted in terms of the recommendation accuracy metric. The experimental results have demonstrated that the proposed Biclustering method is very simple and is able to efficiently extract needed usage knowledge accurately for web page recommendation.

DOI: 10.4018/978-1-60960-102-7.ch011
1. INTRODUCTION

Web Mining specifies three domains: web content mining, web structure mining and web usage mining. The last domain is the most popular one in the area of recommendation (Claypool et al., 2001). Web usage mining, also known as web log mining, aims to discover user patterns from the data stored in server logs or browser logs while surfing the web system. Most of these efforts have been proposed to discover web usage patterns using various data mining or machine learning techniques to model and understand web user activity (Haibin Liu et al., 2000). The mined knowledge can improve the design of web pages, and develop adaptive usage scenarios more efficiently and effectively.

Web usage mining consists of three steps, i.e. data collection and preprocessing, pattern mining as well as knowledge application. Feature selection is a preprocessing step in data mining, and it is very effective in reducing dimensions, reducing the irrelevant data, increasing the learning accuracy and improving comprehensiveness. Log files usually contain nonessential information from the analytical point of view. Thus the first data pre-processing step is the selection of features. The output of the pre-processing phase must be divided into sessions (Hannah Inbarani et al., 2007). These feature selection algorithms are used for selecting significant attributes for describing a session which is suitable for pattern discovery phase.

Pattern mining draws upon the methods and algorithms developed from several fields such as statistics, data mining, machine learning and pattern recognition. The usage mining tasks can involve the discovery of association rules, sequential patterns (Mamata Jenamani et al., 2002) page view clusters (Mobasher, 2002) transaction clusters (Hannah Inbarani et al., 2009) or any other pattern discovery method from user transactions.

The aim of web recommendation is to find the most matched user access pattern to the active user session, which is derived from web usage mining, and to recommend a list of pages that the users might be interested in, via referring to the visiting preferences of the chosen usage pattern. To perform recommendations efficiently and effectively, there are a variety of machine learning algorithms that have been well studied and developed, and can be used in web recommendation (Guandong Xu, 2008).

Basically, there are two kinds of technologies to carry out personalization recommendation: association rules and cluster analysis (TAN Xiaoqiu et al., 2006).

In this chapter, we present and experimentally evaluate a new technique, based on integrated clustering of both user transactions and of page views, in order to discover user and page biclusters of users and pages that can be effectively used by recommender systems for real-time web personalization.

The main contribution of this chapter is to propose Robust Biclustering (RB) algorithm for page recommendation. The proposed algorithm performs simultaneous clustering of users and pages to provide more accurate recommendations.

The rest of this chapter is organized as follows. Section 2 describes the Background Section 3 describes the Methodology for Recommendation process. Section 4 summarizes the related work and Experimental results are given in Section 5. Finally, Section 6 concludes this chapter with observations from the experimentation.

2. BACKGROUND

The stages of page recommender systems include preprocessing, segmenting web log data into web user sessions, and learning a usage model from this data. The usage model can come in many forms: from the modeling used in collaborative filtering, that simply stores all other users’ information and then relies on K Nearest Neighbors to provide recommendations from previous history of neighbors or similar users, to a set of frequent