Chapter 15

Extraction of Target User Group from Web Usage Data Using Evolutionary Biclustering Approach

R. Rathipriya  
Periyar University, India

K. Thangavel  
Periyar University, India

J. Bagyamani  
Government Arts College, India

ABSTRACT

Data mining extracts hidden information from a database that the user did not know existed. Biclustering is one of the data mining techniques which helps marketing user to target marketing campaigns more accurately and to align campaigns more closely with the needs, wants, and attitudes of customers and prospects. The biclustering results can be tuned to find users’ browsing patterns relevant to current business problems. This paper presents a new application of biclustering to web usage data using a combination of heuristics and meta-heuristics algorithms. Two-way K-means clustering is used to generate the seeds from preprocessed web usage data. Greedy Heuristic is used iteratively to refine a set of seeds, which is fast but often yield local optimal solutions. In this paper, Genetic Algorithm is used as a global optimizer that can be coupled with greedy method to identify the global optimal target user groups based on their coherent browsing pattern. The performance of the proposed work is evaluated by conducting experiment on the msnbc, a clickstream dataset from UCI repository. Results show that the proposed work performs well in extracting optimal target users groups from the web usage data which can be used for focalized marketing campaigns.

DOI: 10.4018/978-1-4666-2145-9.ch015
INTRODUCTION

Customer/User segmentation has been a critical element of the marketing and is one of the most important strategic concepts contributed by the marketing discipline to business firms (Shina et al., 2004). User Segmentation is the process of developing effective schemes for categorizing and organizing meaningful groups of customers. A user segment is a group of prospects or customers who are selected from a database based on characteristics they possess or exhibit. It also allows company to differentially treat consumers in different segments. User Profiling is the process of analyzing the customers of each segment in order to generalize, describe or name this set of customers based on common characteristics. It is the process of understanding and labeling a set of users. It provides valuable information about users/customers so marketers can furnish stronger, more targeted offers and each user segment is the target group of users.

One-to-one marketing is the ideal marketing strategy, in which every marketing campaign or product is optimally targeted for each individual customer; but this is not always possible. Therefore, segmentation is required to distinguish similar users and put them together in a segment/group. Doubtlessly using segmentation to understand user’s needs is much easier, faster and more economical than uniquely investing to understand them particularly (Jonker et al., 2004).

With proper market segmentation, enterprises can arrange the right web pages, services and resources to each target user group and build a close relationship with them. Market segmentation has consequently been regarded as one of the most critical elements in achieving successful modern marketing and customer relationship management.

Click stream data is a sequence of Uniform Resource Locators (URLs) browsed by the user within a particular period of time. To discover group of users with similar behavior and motivation for visiting the particular website can be found by clustering. Traditional clustering (Lee et al., 2008) is used to segment the web users or web pages in to groups based on the existing similarities. When a clustering method is used for grouping users, it typically partitions users according to their similarity of browsing behavior under all pages. But, in the most cases web users behave similarly only on a subset of pages and their behavior is not similar over the rest of the pages. Therefore, traditional clustering methods fail to identify such user groups.

To overcome this problem, concept of Biclustering was introduced. Biclustering (Bleuler et al., 2004; Chakraborty et al., 2005; Madeira et al., 2004) was first introduced by Hartigan (1972). Biclustering is the simultaneous clustering of rows and columns of the data matrix. In literature, biclustering algorithms are widely applied to the gene expression data. In this paper, it is used to mine clickstream data in order to extract target usage groups. These groups are analyzed to determine user’s behavior which is an important element in the E-Commerce applications.

The rest of the paper is organized as follows. In Section 2, some of the existing work related to the biclustering approaches and user segmentations are discussed. Methods and materials required for biclustering approach are described in Section 3. Section 4 focuses on the proposed Biclustering approach using Genetic Algorithm. Analysis of experimental results is discussed in Section 5. Section 6 concludes the paper.

RELATED WORK

Tsai et al. (2004) developed a market segmentation methodology based on product specific variables such as purchased items and the associative monetary expenses from the transactional history of customers to address the unreliable results of segmentation based on general variables like customer demographics. Shina et al. (2004) used three clustering methods such as K-Means,
Related Content

A Rigorous Analysis of the Harmony Search Algorithm: How the Research Community can be Misled by a “Novel” Methodology
www.igi-global.com/chapter/rigorous-analysis-harmony-search-algorithm/63805?camid=4v1a

GA_SVM: A Classification System for Diagnosis of Diabetes
www.igi-global.com/chapter/gasvm/179399?camid=4v1a

Pseudorandom Number Generators Based on Cellular Automata With the Hexagonal Coverage
(2018). *Formation Methods, Models, and Hardware Implementation of Pseudorandom Number Generators: Emerging Research and Opportunities* (pp. 139-156).
www.igi-global.com/chapter/pseudorandom-number-generators-based-on-cellular-automata-with-the-hexagonal-coverage/190217?camid=4v1a

Optimization of Drilling Process via Weightless Swarm Algorithm
www.igi-global.com/chapter/optimization-of-drilling-process-via-weightless-swarm-algorithm/115304?camid=4v1a