

Chapter 7

Hybrid Swarm Intelligence–Based Biclustering Approach for Recommendation of Web Pages

R. Rathipriya

Periyar University, India

K. Thangavel

Periyar University, India

ABSTRACT

This chapter focuses on recommender systems based on the coherent user's browsing patterns. Biclustering approach is used to discover the aggregate usage profiles from the preprocessed Web data. A combination of Discrete Artificial Bees Colony Optimization and Simulated Annealing technique is used for optimizing the aggregate usage profiles from the preprocessed clickstream data. Web page recommendation process is structured in to two components performed online and offline with respect to Web server activity. Offline component builds the usage profiles or usage models by analyzing historical data, such as server access log file or Web logs from the server using hybrid biclustering approach. Recommendation process is the online component. Current user's session is used in the online component for capturing the user's interest so as to recommend pages to the user for next navigation. The experiment was conducted on the benchmark clickstream data (i.e. MSNBC dataset and MSWEB dataset from UCI repository). The results signify the improved prediction accuracy of recommendations using biclustering approach.

INTRODUCTION

Web usage mining (WUM) systems use data mining algorithms on both usage and clickstream data from one or more websites to discover usage patterns/user profiles. These patterns are analyzed to determine user's behavior and these patterns are used by the recommendation systems to provide

the recommendation of appropriate pages to the web users. Web personalization is the process of personalizing web sites or web services according to the specific user's profile to achieve more efficient web browsing. User's browsing efficiency can be increased by altering the web sites' structures, and by employing recommender systems to produce user-tailored recommendations.

DOI: 10.4018/978-1-4666-6328-2.ch007

User profiles or Usage models are the generalization of the collected data about the user behavior from web usage data (such as clickstream data). The main goal of using user profiling is to increase the efficiency of user activities by delivering more personalized information when users interacting the web site. Usage-based Personalized Recommendation systems (Cooley *et al.*, 1997; Srivastava *et al.*, 2000) analyzed the user's navigation pattern to provide personalized recommendations of web pages according to the current interests of the user.

Recently, web usage mining has gained much attention as it is found to fulfill the needs of web personalization. In the literature, various clustering algorithms can be applied to detect the user profiles as well as other web mining techniques such as association rule, have been explored by several research groups. Clustering is the process of grouping the users into clusters such that users within a cluster have high similarity compared to each other but dissimilar to users in other clusters.

The discovery of patterns from usage data by clustering the web transaction into clusters of user sessions or pages, by itself is not sufficient for performing the personalization tasks (Mobasher *et al.*, 2002). The critical step in the recommendation system is the effective derivation of good quality and useful usage profiles from the web usage data. Clustering, in general tries to partition the set of web users according to their browsing interest of all pages of a web site. Assuming that users may show interest only for a particular subset of pages, conventional clustering approaches may not be sensitive and/or specific enough to find and present correlation between users in an appropriate and comprehensive manner.

To address the limitation of traditional clustering, the biclustering method was introduced. In contrast to traditional clustering, a biclustering method produces biclusters, which consists of subset of users and a subset of pages under which these users behave similarly.

Greedy search algorithms are used as the promising approach in the biclustering algorithms. Greedy search algorithms start with an initial solution and find a locally optimal solution by successive transformations that improve some fitness function. Most of the times, the results of biclustering suffer from local optima problem. Meta-heuristics optimization algorithms such as Particle Swarm Optimization (PSO) (Kennedy & Eberhart, 1995) Genetic Algorithm (GA) (Goldberg, 1989) and Simulated Annealing (SA) are used along with greedy biclustering to improve the results because it has potential to escape local optima.

The term Swarm Intelligence (SI) was coined by Beny and Wang in late 1980s in the context of cellular robotics. Swarm Intelligence is a collection of nature based algorithms which has attracted several researchers from the field of pattern recognition, information retrieval and clustering. SI systems are typically made up of a population of simple agents interacting locally with one another and with their environment. The efforts to mimic such behaviors through computer simulation finally resulted into the fascinating field of SI. An example of swarm intelligence is Particle Swarm Optimization (PSO) and it is a very popular SI algorithm for global optimization over continuous search spaces and also for discrete optimization problems.

The search algorithms achieve the two goals that are exploration (diversification) and exploitation (intensification) by using local search methods or global search approaches, or an integration of both global and local strategies, these algorithms are commonly known as hybrid methods. The present study focuses on the hybrid swarm algorithm in which one of the main algorithms is a well known search strategy called Discrete Artificial Bees Colony Optimization (DABC) is combined with Simulated Annealing (SA) technique. SA is a powerful optimization procedure that has been successfully applied to a number of combinatorial optimization problems.

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/hybrid-swarm-intelligence-based-biclustering-approach-for-recommendation-of-web-pages/115303

Related Content

The Information Paradox: Researching Health Service Information Systems Development

Said Shahtahmasebi (2016). *International Journal of Computers in Clinical Practice* (pp. 28-41).

www.irma-international.org/article/the-information-paradox/152591

Automatic Circuit Design of CMOS Miller OTA Using Cuckoo Search Algorithm

Pankaj P. Prajapati and Mihir V. Shah (2020). *International Journal of Applied Metaheuristic Computing* (pp. 36-44).

www.irma-international.org/article/automatic-circuit-design-of-cmos-miller-ota-using-cuckoo-search-algorithm/239556

New Evolutionary Algorithm Based on 2-Opt Local Search to Solve the Vehicle Routing Problem with Private Fleet and Common Carrier

Jalel Euchi, Habib Chabchoub and Adnan Yassine (2013). *Trends in Developing Metaheuristics, Algorithms, and Optimization Approaches* (pp. 125-148).

www.irma-international.org/chapter/new-evolutionary-algorithm-based-opt/69722

A Survey of the Cuckoo Search and Its Applications in Real-World Optimization Problems

Christopher Expósito-Izquierdo and Airam Expósito-Márquez (2017). *Handbook of Research on Soft Computing and Nature-Inspired Algorithms* (pp. 541-555).

www.irma-international.org/chapter/a-survey-of-the-cuckoo-search-and-its-applications-in-real-world-optimization-problems/179406

An Integrated Model of Data Envelopment Analysis and Artificial Neural Networks for Improving Efficiency in the Municipal Solid Waste Management

Antonella Cavallin, Mariano Frutos, Hernán Pedro Vigier and Diego Gabriel Rossit (2018). *Handbook of Research on Emergent Applications of Optimization Algorithms* (pp. 206-231).

www.irma-international.org/chapter/an-integrated-model-of-data-envelopment-analysis-and-artificial-neural-networks-for-improving-efficiency-in-the-municipal-solid-waste-management/190161