International Journal of Innovative Computing, Information and Control
Volume **4**, Number **4**, April **2008**

ICIC International ©2008 ISSN 1349-4198
pp. **897–904**

KNOWLEDGE DISCOVERY FROM WEB USAGE DATA: A NOVEL APPROACH FOR PREFETCHING OF WEB PAGES BASED ON ART NEURAL NETWORK CLUSTERING ALGORITHM

G. T. Raju[1] and P. S. Satyanarayana[2]

[1]Department of Computer Science and Engineering
[2]Department of Electronics and Communication
B. M. S. College of Engineering
Visvesvaraya Technological University, Karnataka India
gtraju1990@yahoo.com; pssvittala@gmail.com

Received January 2007; Revised June 2007

Abstract. *The exponential growth of the web in terms of web sites and their users during the last decade has generated huge amounts of data related to the user's interactions with the web sites. This data is recorded in the web log files and usually referred to as Web Usage Data (WUD). Knowledge Discovery from Web Usage Data (KDWUD) has become very critical for efficient and effective management of the activities related to: e-business, e-education, personalization, website management, network traffic analysis, the cache, the proxies, great diversity of web pages in a site, search engine's complexity, and the ability to predict user's actions. In this paper, we propose a novel approach called Cluster and PreFetch (CPF) for prefetching of web pages based on the Adaptive Resonance Theory (ART) neural network clustering algorithm. First, we cluster the users and then we prefetch the web pages for each cluster before the users request them. Experiments have been conducted and results show that prediction accuracy of our CPF approach is as high as 98.38 percent. Our CPF approach effectively reduces the user perceived latency without wasting the network resources.*
Keywords: Prefetching, Clustering, Knowledge discovery, ART neural network

1. **Introduction.** The WWW has had an impressive development in the past decade, today reaching more than 10 billion documents online and having 25 million new web pages published each day. This phenomenal growth of the web triggered the development of new domains of applications. KDWUD is the application of data mining techniques to large web usage data repositories in order to extract potentially useful and previously unknown patterns. KDWUD comprises of three main stages: Preprocessing of raw web log data, Discovery of patterns and the Pattern Analysis. Once the raw web usage data is preprocessed into server sessions or episodes, a number of techniques can then be used to discover clustering and sequential patterns. Finally, the OLAP/Visualization techniques may be used to extract interesting patterns to help an analyst. Since the amount of network traffic has rapidly increased with the WWW expansion, Users have experienced a long latency when retrieving the web pages. To solve the latency problem, a prefetching scheme that predicts the destination pages for users; grouping users into community; extracting patterns for each community has become critical to save the communication overhead. Prefetching means fetching of web pages before the users request them. Prefetching technique is like a web caching technique to reduce the user perceived latency. Prefetching technique is motivated by the fact that, in general, once a user goes to a Web site; he/she generally browses around for several pages before leaving for another site. Since the user follows hyperlinks upon his/her interests, it is likely that links are not followed uniformly.