14th Int'l CSI Computer Conference
(CSICC'09)
Amirkabir University of Technology
Tehran, IRAN
October 20-21, 2009

Identification of Web Communities using Cellular Learning Automata

S. Motiee, M. R. Meybodi
Soft Computing Laboratory, Computer Engineering and Information Technology Department
Amirkabir University of Technology, Tehran, Iran
motiee@aut.ac.ir, mmeybodi@aut.ac.ir

Abstract

A collection of web pages which are about a common topic and are created by individuals or any kind of associations that have a common interest on that specific topic is called a web community. Since at present, the size of the web is over 3 billion pages and it is still growing very fast, identification of web communities has become an increasingly hard task. In this paper, a method based on asynchronous cellular learning automata (ACLA) for identification of web communities is proposed. In the proposed method first an asynchronous cellular learning automaton is used to determine the related pages and their relevance degree (the relationship structure of web pages). For determination of relationship structure of web pages information about hyperlinks and the users' behaviour in visiting the web pages are used. Then, an algorithm similar to the HITS algorithm is applied on the obtained structure to identify the web communities. One of the advantages of the proposed method is that the web community obtained using this method is not dependent on a specific web graph structure. To evaluate the proposed approach, it is implemented and the results are compared with the results obtained for two existing methods, HITS and a complete bipartite graph based method. Experimental results show the superiority of the proposed method.