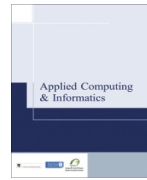




Saudi Computer Society, King Saud University

Applied Computing and Informatics

(<http://computer.org.sa>)
www.ksu.edu.sa
www.sciencedirect.com



ORIGINAL ARTICLE

Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method



D.A. Adeniyi, Z. Wei, Y. Yongquan *

Department of Computer Sc. & Technology, College of Information Sc. & Engineering, Ocean University of China, Qingdao, Shandong, China

Received 3 August 2014; revised 29 September 2014; accepted 17 October 2014

Available online 28 October 2014

KEYWORDS

Automated;
Data mining;
K-Nearest Neighbor;
On-line;
Real-Time

Abstract The major problem of many on-line web sites is the presentation of many choices to the client at a time; this usually results to strenuous and time consuming task in finding the right product or information on the site. In this work, we present a study of automatic web usage data mining and recommendation system based on current user behavior through his/her click stream data on the newly developed Really Simple Syndication (RSS) reader website, in order to provide relevant information to the individual without explicitly asking for it. The K-Nearest-Neighbor (KNN) classification method has been trained to be used on-line and in Real-Time to identify clients/visitors click stream data, matching it to a particular user group and recommend a tailored browsing option that meet the need of the specific user at a particular time. To achieve this, web users RSS address file was extracted, cleansed, formatted and grouped into meaningful session and data mart was developed. Our result shows that the K-Nearest Neighbor classifier is transparent, consistent, straightforward, simple

* Corresponding author.

E-mail address: i@yangyongquan.com (Y. Yongquan).

Peer review under responsibility of King Saud University.



Production and hosting by Elsevier

<http://dx.doi.org/10.1016/j.aci.2014.10.001>

2210-8327 © 2015 Production and hosting by Elsevier B.V. on behalf of King Saud University.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

to understand, high tendency to possess desirable qualities and easy to implement than most other machine learning techniques specifically when there is little or no prior knowledge about data distribution.

© 2015 Production and hosting by Elsevier B.V. on behalf of King Saud University.
This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

1. Introduction

Data mining is the extraction of knowledge from large amount of observational data sets, to discover unsuspected relationship and pattern hidden in data, summarize the data in novel ways to make it understandable and useful to the data users [13,31,2]. Web usage mining is the application of data mining technique to automatically discover and extract useful information from a particular web site [2,22,30].

The term web mining was believed to have first came to be in 1996 by Etzioni in his paper titled “The World Wide Web: Quagmire or Gold mine” and since then attention of researchers world over has been shifted to this important research area [26]. In recent years, there has been an explosive growth in the number of researches in the area of web mining, specifically of web usage mining. According to Federico and Pier [9], over 400 papers have been published on web mining since the early paper published in 1990s.

The Really Simple Syndication (RSS) reader website was developed for the purpose of reading dailies news on-line across the Globe, but lack ways of identifying client navigation pattern and cannot provide satisfactory Real-Time response to the client needs, so, finding the appropriate news becomes time consuming which makes the benefit of on-line services to become limited. The study aimed at designing and developing an automatic, online, Real-Time web usage data mining and recommendation system based on data mart technology. The system is able to observe users/clients navigation behavior by acting upon the user’s click stream data on the RSS reader web site, so as to recommend a unique set of objects that satisfies the need of an active user in a Real-Time, online basis. The user access and navigation pattern model are extracted from the historical access data recorded in the user’s RSS address URL file, using appropriate data mining techniques.

The K-Nearest Neighbor classification method was used online and in Real-Time to exploit web usage data mining technique to identify clients/visitors click stream data matching it to a particular user group and recommend a tailored browsing option that meet the need of the specific user at a given time [24]. For instance, if a user seems to be searching for politics news on china daily on his/her visit to the RSS reader site, more politics news headlines from other dailies such as CNN politics news will be recommended to the user with the required feed needed to be added to his/her profile in order to access such news headlines asides his/her originally requested news. This is aimed at assisting the user to get relevant

Download English Version:

<https://daneshyari.com/en/article/467400>

Download Persian Version:

<https://daneshyari.com/article/467400>

[Daneshyari.com](https://daneshyari.com)