UNIVERSITI PUTRA MALAYSIA

ADAPTIVE METHOD TO IMPROVE WEB RECOMMENDATION SYSTEM FOR ANONYMOUS USERS

YAHYA MOHAMMED ALMURTADHA

FSKTM 2011 27

ADAPTIVE METHOD TO IMPROVE WEB RECOMMENDATION SYSTEM FOR ANONYMOUS USERS

By

YAHYA MOHAMMED ALMURTADHA

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfillment of the Requirement for the Degree of Doctor of Philosophy

March 2011

Dedicated to

my parents, my wife,

my brothers, and my sisters

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the degree of Doctor of Philosophy

ADAPTIVE METHOD TO IMPROVE WEB RECOMMENDATION SYSTEM FOR ANONYMOUS USERS

By

YAHYA MOHAMMED ALMURTADHA

March 2011

Chairman : Associate Professor Hj. Md. Nasir Sulaiman, PhD

Faculty : Computer Science and Information Technology

Today the major concerns are not the availability of information but rather obtaining the right information. Web Recommendation system is a specific type of Web personalization system technique that attempts to predict the user next browsing activity then recommend the web pages items that are likely to be of interest to the user. The ability of predicting the next visited pages and recommending it to the short term navigation user (anonymous user) is highly needed. Presently, there are many recommendation systems (e.g. Analog, Web Miner, WebPersonalizer, PACT, SWARS, EntreeC, SUGGEST, one-and-only items, Hybrid and NEWER) that can be used to make recommendation to the current online user, but, recommendation to anonymous users needs to be adaptive (up to date) to the changes in users' interests' over time.

This research focuses on improving the prediction of the next visited web pages and introduces them to current anonymous user. An enhanced classification algorithm is

used to assign the current anonymous user to the best web navigation profile. As the users' interests change over time, the recommender system has the ability to modify the current web navigation profiles and keep them updated. These adaptive profiles help the prediction engine to predict and then recommend the next visited pages to the current user in an accurate manner.

This research proposed two web page recommendation systems. The first is iPACT, an improved recommendation system based on PACT methodology to demonstrate the prediction accuracy of the proposed enhanced classification algorithm in this research. The prediction accuracy was evaluated against two previous recommendation systems PACT and HyperGraph. The second is Adaptive Web page Recommendation System (AWRS) which combines the classification algorithm of iPACT in addition to the ability of adaptive recommending due to the changes of the users' interests and weighting methods to deal with unvisited or new added pages. For the evaluating purpose, the experiments were carried out on the public CTI logs file dataset which contains the preprocessed and filtered sessionized data for the main DePaul CTI Web server. AWRS was evaluated and shows better performance as compared to several recommendation systems namely, iPACT, Association Rules and Hybrid systems.

Based on the experimental results, the outcome of a considerably good accuracy is mainly due to the right classification of the current user to the best web navigation profile with similar browsing activities. Also, the adaptive phase is able to update the web navigation profile(s) based on the interest's changes and predict the next visited pages in accurate manner to the anonymous users based on their early stage navigation. This research opens a wide range of future works to be considered,

including the investigation of the dependency between the recommended web pages for each web navigation profile, investigating the quality of the method on different datasets, and finally, the possibility to apply the proposed method in other area like the misuse detection systems.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

KAEDAH ADAPTIF UNTUK MENAMBAHBAIK SISTEM CADANGAN PENGGUNAAN LAMAN SESAWANG BAGI PENGGUNA TANPA NAMA

Oleh

YAHYA MOHAMMED ALMURTADHA

Mac 2011

Pengerusi : Profesor Madya Hj. Md. Nasir Sulaiman, PhD

Fakulti : Sains Komputer dan Teknologi Maklumat

Hari ini, fokus utama bukan terhadap wujudnya maklumat tetapi lebih kepada mendapatkan maklumat yang betul. Sistem Cadangan Sesawang ialah satu contoh spesifik teknik sistem sesawang Peribadi yang cuba meramal aktiviti pengguna berikutnya dan seterusnya mencadangkan beberapa laman sesawang yang mungkin menarik bagi pengguna tersebut. Keupayaan meramal laman kunjungan seterusnya dan mencadangkan kepada pengguna untuk navigasi tempoh jangka pendek (pengguna tanpa nama) amat diperlukan. Pada masa ini, terdapat banyak sistem cadangan (contohnya Analog, Web Miner, WebPersonalizer, PACT, SWARS, EntreeC, SUGGEST, one-and-only items, Hybrid and NEWER) yang boleh digunakan untuk membuat cadangan bagi pengguna dalam talian semasa, tetapi,

cadangan untuk pengguna-pengguna tanpa nama perlu adaptif (terkini) terhadap perubahan minat pengguna dari masa ke semasa.

Penyelidikan ini fokus kepada meningkatkan ramalan laman sesawang yang akan dikunjungi seterusnya dan memperkenalkan laman-laman sesawang tersebut kepada pengguna tanpa nama semasa. Algoritma klasifikasi yang telah ditambahbaik digunakan untuk menentukan pengguna tanpa nama semasa kepada profil navigasi sesawang yang terbaik. Oleh kerana minat penguna yang berubah mengikut masa, sistem cadangan harus mempunyai keupayaan mengubahsuai profil navigasi sesawang semasa dan juga boleh melakukan pengemaskinian. Profil adaptif ini dapat membantu enjin ramalan untuk meramal dan seterusnya mencadangkan laman sesawang yang akan dikunjungi seterusnya kepada pengguna semasa dalam dalam kaedah yang betul.

Penyelidikan ini mencadangkan dua sistem cadangan laman sesawang. Pertama ialah iPACT, merupakan sistem cadangan yang telah ditambahbaik berdasarkan kaedah PACT untuk menunjukan ketepatan ramalan algoritma klasifikasi yang telah ditambahbaik seperti yang dicadangkan dalam penyelidikan ini. Ketepatan ramalan ini dinilai dengan membuat perbandingan menggunakan dua sistem cadangan yang sedia ada iaitu PACT AND HyperGraph. Kedua ialah Adaptive Webpage Recommendation System (AWRS) yang menggabungkan algoritma klasifikasi iPACT dengan penambahan terhadap keupayaan cadangan yang adaptif selari dengan perubahan minat pengguna dan kaedah pemberat yang digunakan untuk menguruskan laman yang tidak dilawati atau yang baru ditambah. Untuk tujuan penilaian, beberapa siri eksperimen telah dilakukan dengan menggunakan set data awam CTI *logs* yang mengandungi data pra-proses dan ditapis secara *sessionized* yang diperolehi dari

pelayan sesawang utama DePaul CTI. Penilaian terhadap AWRS menunjukkan prestasi lebih baik berbanding dengan beberapa sistem cadangan yang lain iaitu, iPACT, Association Rules System dan Hybrid System.

Berdasarkan dapatan daripada eksperimen, keputusan dapatan menujukkan hasil ketepatan yang baik selaras dengan hasil klasifikasi yang tepat oleh pengguna semasa berdasarkan profil navigasi sesawang yang terbaik dan aktiviti browsing yang sepadan. Disamping itu, fasa adaptif juga mampu mengemaskini profil navigasi sesawang berdasarkan perubahan minat dan ramalan laman sesawang yang akan dikunjungi berikutnya dalam cara yang lebih tepat bagi pengguna tanpa nama berdasarkan navigasi pada peringkat awal. Penyelidikan ini juga membuka ruang yang luas kepada kajian penyelidikan di masa akan datang termasuk kajian terhadap kebergantungan di antara cadangan laman sesawang untuk setiap profil navigasi, kajian mengenai kualiti kaedah untuk set-set data yang berbeza , dan juga kajian terhadap kebarangkalian kaedah yang disarankan dalam bidang lain seperti penyalagunaan sistem pengesanan.

ACKNOWLEDGEMENTS

In the name of Allah, the most Merciful, Gracious and most Compassionate, Creator of all knowledge for eternity. We beg for peace and blessing upon our Master the beloved Prophet Mohammed (Peace and Blessing be Upon Him) and his progeny, companions and followers.

I wish to extend my deepest appreciation and gratitude to the supervisory committee led by Assoc. Prof. Dr. Hj. Md. Nasir bin Sulaiman and committee members Dr. Norwati Mustapha snd Dr. Nur Izura Udzir for the various guidance, sharing of intellectual experiences and in giving me the vital to undertake the numerous of this study.

Also special thanks to my parents and my wife for making the best of my situation. May thanks are also extended to the Quraan teachers, my friends and the colleagues for sharing experiences throughout the years.

I certify that an Examination Committee has met on 15th March 2011 to conduct the final examination of Yahya Mohammed Al-Murtadha on his Doctor of Philosophy thesis entitles "Adaptive Method to Improve Web Recommendation System for Anonymous Users" in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulation 1981. The Committee recommends that the candidate be awarded the relevant degree. Members of the Examination Committee are as follows:

Mohamed Othman, PhD
Professor/ Deputy Dean
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Abu Bakar Md. Sultan, PhD
Associate Professor/ Head, Information System Department
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Lilly Suriani Affendey, PhD
Head, Computer Science Department
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

R. Ajith Abraham, PhD
Professor
Machine Intelligence Research Lab (MIR Labs)
Scientific Network for Innovation and Research Excellence (SNIRE)
USA
(External Examiner)

SHAMSUDDIN SULAIMAN , PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Md Nasir Sulaiman, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Norwati Mustapha, PhD
Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Nur Izura Udzir, PhD
Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

HASANAH MOHD GHAZALI, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

DECLARATION

I hereby declare that the thesis is based on my original work except for quotation and citation which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any degree at UPM or other institution.

_____

YAHYA MOHAMMED AL MURTADHA

Date: 15 March 2011

TABLE OF CONTENTS