

Research and Design of Web Data Mining in Personalized E-Business

Yun Xue¹, Jianbin Chen²

¹ Business College of Beijing Union University, Beijing, China

Email: yun.xue@bcbuu.edu.cn

² Business College of Beijing Union University, Beijing, China

Email: jianbin.chen@bcbuu.edu.cn

Abstract—As e-business is widely applied, web data mining technology is used for e-business to provide personalized e-business and better meet the requirements of users. Beginning from the concept of personalized information services, this paper focuses on detailing six personalized services available in an e-business environment, proposes the personalized e-business based on web data mining, and makes research and discussion on data resources, key technology and basic flow in relation to this model.

Index Terms—Web Data Mining, E-Business, Personalized Service, Agent

I. PERSONALIZED INFORMATION SERVICE

A. Concept

Research on personalized information services has its roots in the personalized navigation system [1] raised by Robert Armstrong and others of Carnegie Mellon University at the National Conference on Artificial Intelligence (AAAI) in 1995. Personalized information services are now a hot research topic in global information services. Traditional universal information services are giving gradually away to personalized information services that, according to the interest, status and specific requirements of a user, serve the user on a targeted basis. [2]

B. Personalized E-Business

As an extension of personalized information services in e-business, personalized services in an e-business environment are also a field that is prepared for key application, development and research in personalized information services. Personalized services in e-business mean that in the mode of e-business a businessman obtains his user's data and access information by depending on his users' access to his website, and analyzes and processes such information by using web data mining technology to guide his business decision so that based on the requirements of his users the businessman could embark on e-business activities, offer personalized information services, improve the awareness and satisfaction of the users and gain win-win for the businessman and his users. Personalized services in e-business, in nature, are web services centered on user requirements. See Figure 1, where "User Analysis Module" functions to learn the user's features, create user access model and offer personalized services to the user by using technology processing and web resources.

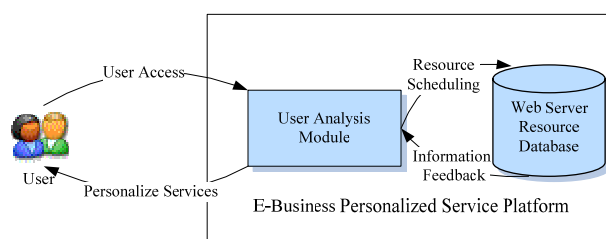


Figure 1. Meaning of Personalized Services in E-Business

C. Personalized Services Available under E-Business Environment

Personalized services in e-business are divided into several types. For more details, see Figure 2.

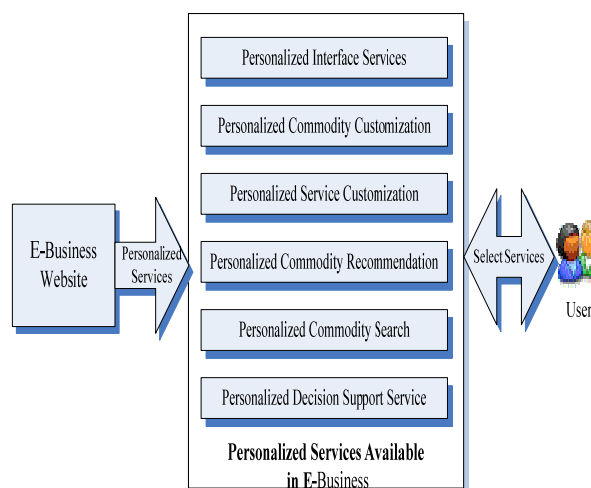


Figure 2. Specific Personalized Services Available in E-Business

1) Personalized Interface Customization

Interface customization means that a user, according to his requirements and interest, designs the personal user interface of an e-business website on a personalized basis, including module layout, display and hiding of function module, button style, picture choice, interface color design and commodity display mode. A survey shows that a personalized interface for users helps to get more attention paid to an e-business website. Now, some websites have launched such similar services.

2) Personalized Commodity Customization

Since personalization is highlighted in our society, an e-business website offers the personalized product design to its users on the basis of its product features. For instance, to meet the personalized requirements of consumers, Nike Company launches the personalized product customization whereby a user can choose the styles and colors of sports shoes freely and place his name and favorite number on the face of a pair of shoes to show his personality and avoid any conformity. A user first submits a unique shoe sample he designs himself via the website, and then makes the payment online; they will be sent to the user as soon as the shoes are made. Personalized commodities serve as the choice of a new generation trend while free customization or DIY commodities will be the one of the greatest potential and development room.

3) *Personalized Service Customization*

An e-business website can, according to the personalities and requirements of users, provide more services, including online consulting service, e-mail, SMS, telephone service, expert consultation and service software (weather software and news software). A user can learn about update, prices and types of commodities for sale by such personalized services. Furthermore, this website can offers some special services on the grounds of users' information feedback and requirements, for instance, provide interactive multimedia information services such as 3D animation model, animation video data, or sounds and words.

4) *Personalized Commodity Recommendation*

When providing more and more choices, an e-business website has seen its increasingly complicated site structure so that a user is often troubled with "overloading data" and "getting lost in information", and finds it very difficult to search smoothly for his desirable commodities. The personalized recommendation system at an e-business website learns the features of users unceasingly to create the corresponding visit models. Before recommendation, based on the access model of a user, this website interacts directly with the user, imitates the salesman in a store to recommend commodities to the user and assist the user in finding his desirable products. The personalized recommendation in the increasingly competitive e-business environment attracts the users effectively, improves the customer loyalty and commodity sales.

5) *Personalized Commodity Search*

It is not easy to find a cheap and desirable website to buy products among thousands of e-business websites across Internet. The personalized commodity search provides a good solution to this problem. It helps a user to enjoy more convenience and tangible benefits brought by shopping online. This service, based on the interest and features of users, provides two ways: search the whole network, and search intra-website. Depending on the search, a user can find quickly relevant commodities on numerous e-business websites and in this website across Internet, look through product pictures, market prices, member prices, brief introduction and source websites via browser. He can also click the link of a source website to visit the final page of the website where

commodities are available for shopping online. For instance, the search engine of 8848 shopping online provides search service for commodities in China.

6) *Personalized Decision Support Service*

Providing users with some rules and modes in relation to decision support, intelligent query, scientific research and problem solving and so on, this service still in research can be used to improve the competitiveness of e-business websites and make them more vital.

II. WEB DATA MINING

As a key technology to provide personalized e-business and help to collect user information, web data mining can be used to analyze user data, create access model, requirement model and interest model that accord with user features, making personalized e-business possible.

A. *Concept*

Oren EtZioni put forward the concept of web data mining in his thesis in 1996 for the first time: applying data mining to web helps to automatically discover potential and useful models or information [3] among a great number of web documents and services.

B. *Classification*

According to data mining behaviors, web data mining is classified as web content mining, web structure mining and web usage mining. For more information, see Table 1. Web usage mining means that by mining the log files and data at the corresponding site, you discover the behaviors of visitors and users having access to this site. Data mining methods include path analysis, association rules, classification rule, sequential patterns, statistical analysis, dependent relationships modeling and cluster analysis [4].

TABLE I. CLASSIFICATION OF WEB DATA MINING

Classification	Secondary Classification
Web Content Mining	Text Mining
	Multimedia Mining
Web Structure Mining	Organizational Structure Mining
	Page Structure Mining
Web Usage Mining	User Record Data Mining
	Customization Mode Mining

C. *Processing Flow*

Data on web is unstructured, semi-structured and dynamic, so web data mining has to go through the corresponding processing flow that is composed of data positioning, data preprocessing, pattern recognition and pattern analysis. In this process, you should first determine the source of data, including web document, e-mail, website log data and transaction data. Next, you should preprocess data, i.e., delete some redundant information and unify information recognition, session recognition and transaction recognition, and then carry out pattern recognition of the preprocessed data, i.e., use the data mining method to mine useful, potential and understandable information. Finally, through pattern analysis, you can convert the filtered data into useful

rules and patterns to guide the practical e-business activities.

III. DESIGN OF WEB USAGE MINING BASED PERSONALIZED E-BUSINESS MODEL

A. Use of Agent Technology

Agent is a computer system in a certain environment that imposes flexible autonomous actions on its environment to reach its expected target. Agent has the following essential features [5] as: autonomy, communication capacity, interaction capacity, proactive capacity, viability, perception capacity, initiative and sustainability.

In this model, Agent technology is used to help support decisions and collect information, i.e., screen out the qualified information from data in quantity according to relevant user information, update information resource database dynamically, reduce the working stress of a server and improve the efficiency.

This model has the function module of Agent technology as user Agent. The module, consisting of input interface, history database, reasoning machine and output interface, interacts with users. User Agent records user usage data via the input interface, and saves it in a history database. The reasoning machine, based on the user data in the history database, analyzes the current user intent by using knowledge base in collaboration with user model, and assists the user in using both actively or semi-actively. Meanwhile, the reasoning machine always updates or optimizes user models on the basis of users' new usage conditions, and eliminates some outdated applicable records. Additionally, this machine makes query more detailed. The output interface is used to show results. As user Agent technology is used, accurate user interest model and certain applicable experience are a great help to personalized e-business.

B. Data Source of Web Usage Mining

1) Server Data

The access log file of web server records access and interaction information of visitors. This web log file, containing many records, is used to record users' access to this website, including Server log (user IP, server name, URL, time to browse this site, Cookie identification number), Error log (lost connection, authorization failure, timeout, etc.), Cookie (user state and access path).

2) User Registration Information

Having access to an e-business website, a user inputs and delivers some information to the server via web page, including registration information and exchange information. Analysis of user registration information helps you to analyze user behavior pattern and formulate the corresponding e-business policies aiming at specific users.

3) Transaction Data

The background database of an e-business website saves user information, goods information table and order information table. Each time a user completes a commodity transaction, the order information table will have a new record to record the user purchase

information. This information mining is of great significance to analyze users' interest.

C. System Model Architecture

Use of web data mining technology in combination with Agent technology allows you to model a web-based mining personalized e-business system. For more information, see Figure 3 below.

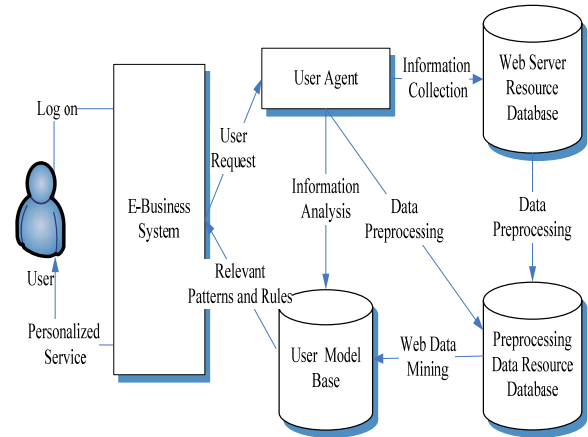


Figure 3 The Model Architecture of Web Data Mining Personalized E-Business

D. Work Flow

- A user logs on to an e-business website platform. Then, user Agent enables this function module on the basis of user usage information.
- According to the personalized requirements of users, organize data resource and find the original use data of this user.
- Preprocess the user data, including data cleansing, conversion, integration and formatting, and load the results to the preprocessing data resource bank.
- Select a suitable data mining method in collaboration with user Agent to build user model and model base.
- Based on data mining results, integrate with expertise and area rules, and offer users personalized e-business services via an e-business system.

IV. CONCLUSION

As Internet-based e-business grows rapidly, personalized e-business should be worth paying more attention and developing from the theoretical and practical standpoints. Obviously the personalized e-business have some room for further improvement and research, for instance, collection of user registration information cannot violate users' privacy during web data mining while optimization of web data mining algorithm and user modeling, etc, and that will be a research trend in the future personalized e-business.

REFERENCES

- [1] Yang Lin and Mao Yurong, "Personalization: Customize Your Network Service", *Software Engineer*, v.7, 2003.
- [2] James C and Charles L., "Personalized Information Environment: An Architecture for Customizable Access to Distributed Digital Libraries", *Library Magazine*, v.6, 1999.
- [3] Scafer J.B., Konstan J.A., Riedll J.: "E-Business Recommendation Applications" [J]. *Data Mining and Knowledge Discovery*, vol. 5, no. 1/2, 115-153, 2001
- [4] Zhao Hongxia and Zhao Yumei, "Web Data Mining Applications in E-Business", *Financial Computer of Huanan*, v.11, 2004.
- [5] Tang Chunsheng and Jin Yihui, "Proxy Mechanism Based Internet Information Automatic Extraction", *Computer Engineering and Applications*, v.10, 2001: 38-41.
- [6] Servastava J., Cooley R., Deshpande M., Tan P.-N.: "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data [C]., *SIGKDD Explorations*, 2000.
- [7] Rosario Sotomayor, Joe Carthy and John Dunnion, The Design and Implementation of an Intelligent Online Recommender System [C], Mixed-Initiative Problem – Solving Assistants, 2005 Fall Symposium Series, AAAI, Virginia, USA, 3rd – 6th November 2005, AAAI Press.
- [8] Beale Russell Supporting serendipity. Using ambient intelligence to augment user exploration for data mining and web browsing [J]. *International Journal of Human Computer Studies*, 2007, 65(5):421-433.
- [9] Kwork KurLan, Grunfeld Laszlo, Deng Peter ,Employing web minging and data fusion to improve weak and hoc retrieval[J].*Information Processing and Management*, 2007, 43(2):406-419
- [10] Brian Pinkerton, Web Crawler Finding What People Want [D].*University of Washington*, 2000.
- [11] Eirinaki M, Vazirgiannis M. Web Mining for Web Personalization [J]. *ACM Transactions on Internet Technology*, 2003, 3(1):1-27.
- [12] Baglioni M, Ferrara U, Romei A. Preprocessing and Mining Web Log Data for Web Personalization [J]. *Advances in Artificial Intelligence*, 2003(2):237-249.