

The Research of Personalization E-commerce Model Based on Data Mining

Cheng Deli
Law school of Tongji university
Fudan university
Shanghai ,200092,china

Abstract: with the development of supply chain and logistic, the commodity classification and amount provided by E-commerce system rise increasingly. The traditional web-centered searching commerce model has become the bottleneck hindering value-adding of e-commerce enterprises, while personalized e-commerce system is one effective way to solve this problem. Incorporating the demand for personalized e-commerce system and data mining technology, this paper explores personalized e-commerce model based on data mining. The practice proves that this model can realize one-to-one service for the customer provided by e-commerce system, shorten the shopping time of customers, maximize the attractions for customers and increase the loyalty of clients.

Key Words: Data Mining Personalization E-commerce Model Design

In the traditional commerce era, the personal need of people can not be met, so the occurrence, development and popularization of internet liberalize the individuals in the economic organization with traditional order, which provides a more effective platform for publicizing individuals and giving play to creativity, and also provide stronger technical support for the realization of customer rights. When certain commodities, service or internet are considered personal, it means they can attract more concern and attention which can be transformed into rich economic resource. As a certain stage of developing entrepreneurial e-commerce, personal e-commerce conducts deeper development and management of information dissemination, transaction means and other things by maximizing the attraction for customers, maintaining and increasing customer loyalty to enable customers to enjoy customized one-to-one service and information. Each kind of commodity has its own certain user group. Different groups have different consumption habits, preference and consumption capacity. In the increasingly fierce market competition, personalization, specialization and individualization

has taken up an increasingly important role in the service for clients, therefore, enterprises and trade companies should fully subdivide the market, find exact commodity positioning, and implement personal market activities for target customers.

The core for realizing personal e-commerce system is to analyze accurately and dynamically customers' preference. This paper introduces data mining theory into the design of e-commerce system, realizes the personal service provision and increases the attraction of commodities for customers, by using data-mining to analyze dynamically the preference and demand of customers, which is conducive for the smooth implementation of e-commerce.

1. Systematic Structure

Personal e-commerce system is a three-layer structure, as shown in diagram 1. System is composed by three layers, customer application service and data.

a. Customer layer. Provide interface for users and system, take responsibility for receiving the request proposed

by users for system, and display feedback of system to the users in text, graph, pictures, audios and videos. The main component of customer layer is IE or web-centered desktop system.

b. Application service layer. Application service layer integrates the processing of service logic of e-commerce system, which is the core part of application system. The core of application service layer is the Web server which boasts traditional functions and data-mining module, a core one to realize personalized e-commerce.

c. Data layer. Store mainly all kinds of data and resource needed in the operation of application service. The database in data layer adopts popular relational database system to realize system easily.

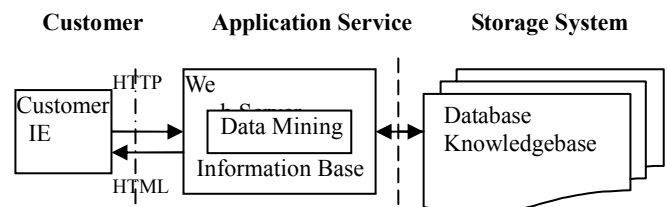


Diagram 1 Systematic Structure

2. Model realization

Most e-commerce websites can not meet the need of personalized service for customers currently. Given this, we conducted the research on some domestic large-sized e-commerce internets and some of their relatively stable customers, and initially met the basic requirements of commerce websites and customers for personalized service, based on which we defines the research on implementation technique of personalized e-commerce system model based on data-mining from proceeding design, data mining and system interface.

2.1. Process design

The process design generally includes:

a. customer send out HTTP request from IE to Web server.

b. Web server conduct different operations based on two different categories. For new customers, the relative module of server records access track and detention time on different pages of users; for existing customers, server extracts historical record of customers from storage system as well as records update access track and detention time of customers.

c. server send out the collected information to data-mining module which will complete the related operation.

d. explanation and evaluation module seizes the processing result of data-mining module and feedback of customer IE, adopts certain algorithm for integrated processing and forms two part information. One is to send back IE information, and the other one is to correct the data updating information in the storage system.

e. the data updating information of explanation and evaluation module is sent back to updating data module, which completes the updating of storage system.

f. the IE information of explanation and evaluation module is sent back to IE in HTML form.

The difference between personalized e-commerce system based on data-mining and traditional e-commerce system is that it introduces data –mining and explanation and evaluation module. The key of systematic implementation is how to complete the design of these two modules.

2.2 Data-mining Module

Some research has been conducted on the algorithm of data-mining in current e-commerce environment, but few are mature. The main reason for it is that structural data (database) and non-structural data (web log or pages) should be operated simultaneously during data-mining process. To solve this problem, data-mining framework is designed which is shown in diagram 2.

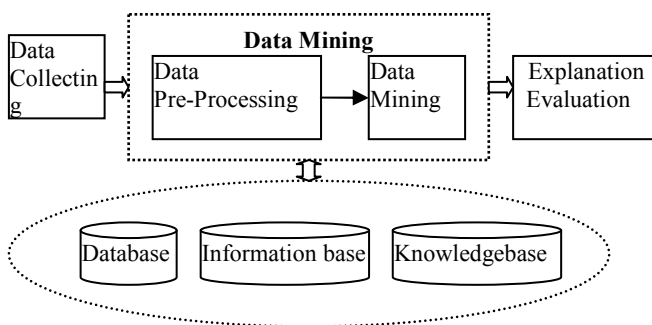


Diagram 2 Data Mining Framework

Data-mining framework is composed by data-mining module and storage system which includes database, information base and knowledgebase. Operated on database server, database stores detailed information of commodity, registration information of users and other system data. Operated on Web server, information base stores mainly such dynamic information as Web log. Operated on database server, knowledgebase stores data-mining algorithm and information of general knowledge to judge customers' preference. Data-mining is the core of framework model, including mainly data pre-processing and data-mining.

2.2.1 Data Pre-processing

Clean, accurate and clear data should be provided to make data-mining module work effectively. However, the original information flow formed by data collecting has the features of scrambling, repetition and imperfection. Data pre-processing is extracting the part of attribute which can exert important impact on target output from numerous data attributes, that is, to lower dimension of original data, so that to improve the data quality of original information flow and increase data-mining speed. Data pre-processing module generation result is regarded as data-mining information flow. Data pre-processing has such basic functions as data integration, data cleaning, data transformation and data reduction.

The information formed by data collecting includes users' access record and information extracted from storage system which comes from database and information base and knowledgebase. The features of storage system decide

that it boasts features like small data and high integration, which, generally speaking, doesn't need pre-processing. Therefore, data pre-processing works on the information from users' access record. Generally, the information from users' access record is Web data, which boasts features like big data, complex pages and skimble-scramble structure. Data pre-processing is the basis of realizing data-mining. The intention of customer can not be correctly reflected unless the accuracy of the data can be ensured, so that the following analysis can be conducted in the right direction. From the existing bibliographies, Web data pre-processing includes the pre-processing of service data and content and data structure.¹

a. The Pre-processing of Service Data

The specific steps include data cleaning, user recognition, user dialogue recognition, path supplement, and business recognition. Data cleaning refers to deleting data unrelated to data mining in Web data, such as pictures and frameworks. User recognition refers to correctly distinguishing each customer accessing the web. User dialogue recognition refers to cutting customers' record into each dialogue generally with a fixed time period as the time limit, e.g. 30 minutes, every 30 minutes access sequences of a user can be regarded as a dialogue of this user. The selection of fixed time period can be defined by statistically analyzing the blog. Path supplement refers to supplementing the access track of user dialogue so as to better reflect the access process of the users, which mainly functions as solving the problem of failing to record completely the access track of users for the reason of slow storage. Business recognition refers to distinguishing the meaningful page access sequences from user access dialogue.

b. Pre-processing of content and data structure

The pre-processing of content and data structure is to transform context, graph, Script of Web pages and super links between Web pages into the form of Web service data mining according to specific application. For example, the concept and topic concerning the page can be described for the clustering of Web pages according to the text content. Topological Structure chart in Websites should be built based on the super link information between web pages to recognize users.

c. Result of data pre-processing

The result of data pre-processing is pages collecting $P = \{p_1, p_2, \dots, p_n\}$ and user business collecting $T = \{t_1, t_2, \dots, t_m\}$. Set P is the subset of websites topological structure; Each component of set T is the subset of websites topological structure P as well as subset of P , which is composed by websites with relatively high correlation. Any component $t_i(p_j, w_j \mid j=1 \dots k)$, while p_j is one page of set p , w_j is the weights in t_i of this page. In the design of system, weights of w_j are based on the detention time of customers on p_j page.

2.2.2 Data Mining

Data mining refers to the process of extracting potentially useful information and knowledge inferred which is unknown to the people from numerous, incomplete, noisy, blurry and random data. Data mining take advantage of the research fruit of Mathematical statistical technique and humane intelligence as well as knowledge engineering to construct its own theoretical system, an cross-course technique which can integrate database, humane intelligence,

mathematical statistic, visional and parallel algorithm.²⁻⁷ Data Mining is not simple indexing, searching, transferring aimed at specific data, but also conducts micro, medium and macro statistic, analysis, combining and reasoning of these data to guild the solving of these practical problems. Effort is made to discover the correlation between events, and even to predict the future activities with existing data. In the design of personalized e-commerce system, enormous business opportunities can be produced by using data-mining technique. The flow chart of data-mining in personalized e-commerce system based on data-mining is shown on diagram 3.

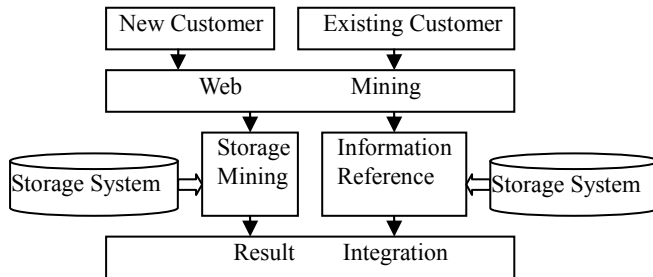


Diagram 3 Data Mining Process

There are four core modules in the data mining: web mining, storage system mining, information reference, and result integration. Web mining refers to mining customers' information from Web data from data collected. Information reference refers to comparing the present customers' record with existing information of storage system, from which customers' reference can be derived. Storage mining actually refers to extracting customer-related information and data mining algorithm stored in knowledgebase from storage system. Result integration functions as realizing format conversion of data. The first three realize data mining to shorten system design period. Some existing algorithm is used in realizing data mining module, while Web mining adopts algorithm of reference 8. Information reference adopts data mining method based on neural network, and storage mining is realized by information indexing. Result integration converts result of data mining into standard XML document for follow-up module to use. It is to increase expandability of system by adding data integration module.

2.2.3 Related Interface

a. Data collecting

The input information of data mining is the result of data collecting. From system operation process shown in diagram2, the result of data collecting includes user access record and information extracted from storage system. The latter one is basically composed by: information from database which is a two- dimension table composed by attribute and field, information from information base, which is previous access information, including content information and structure information, and which is in tree structure where the nodes store the abstract of page information and access track (detention time)of customers; preference algorithm of data mining from knowledgebase (adopting higher grade preference algorithm aimed at customers with different preference, different preference grade of mining algorithm)and some common knowledge defining customers' preference (e.g. residence and age of customers can be regarded as key words in the cosmetic web marketing to

abstract common information)

b. Explanation and evaluation

Explanation and evaluation module refers to evaluating accuracy and feasibility of data mining results. The input information of this module is the operation result of data mining module and feedback of customer IE. It is an interactive module which sends result formed by data mining to Web database interface and forms HTML text indicating that users are on IE and meanwhile Web server feed back the operation of users on pages to explanation and evaluation module to realize the evaluation of data mining results. The testing system in this paper adopts traditional weighted average to evaluate results.

2.3 Systematic Features

The main functions of personalized e-commerce system based on data-mining include:

a. Automatic tendency prediction. E-commerce system based on data mining technique can find potential predicting information from numerous data and gain the prospect development tendency of commodity sale to deal with the fast-changing market.

b. Acquisition of users' personalized need. The system can adopt relative algorithm to analyze the personalized need of users and provide more humane service for users based on the result analyzed by analyzing users' such historical record as IE pages and purchased commodities.

c. Dynamical probing of unknown mode. Adopting data mining technique, e-commerce system can scan all the data and recognize the hidden modes. E.g. recognizing seemingly non-related commodities which are sold together under most circumstances by analyzing zero sale data.

d. Data mining technique can make existing software and hardware more automatic, and enable them to be operated on the updated or newly-developed platform. When operated on parallel processing system with high performance, data mining tools can analyze a super large-scale database within several minutes. This faster processing means that users can have more opportunities to analyze data and make the result analyzed more accurate, reliable and easier to understand.

3. Conclusion

As a website to sell books and audio-video commodities, testing system adopts ASP NET compilation in a logically tree structure where commodities are stored in category. Testing group are 60 people in total, each of who will visit 30 times according to their own needs. The statistical result of different kinds of customers and different access are shown in diagram 4 and 5.

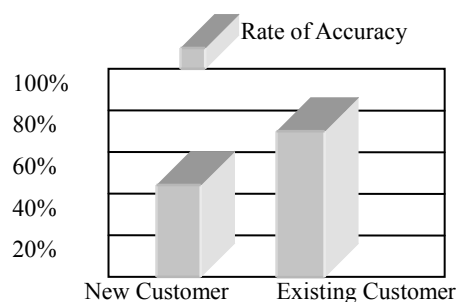


Diagram4 Customer Classification Statistics

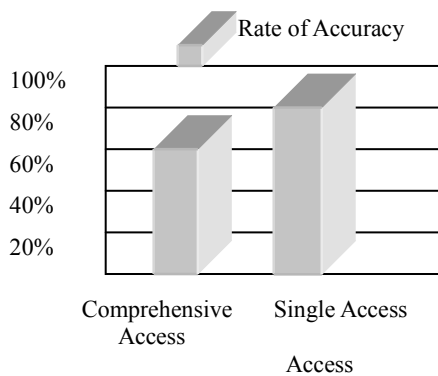


Diagram 5 Access Method Statistics

We can come to conclusion from diagram 4 and 5:

- the rate of accuracy of new customers is much lower than that of existing customers. The main reason for this is that the information in the testing from storage system is limited, which leads to the deviation of inference result in the information reference module. This situation can be improved by expanding the information volume in storage system. Currently, the system will not provide personal recommendation toward new users to avoid this problem, but return to HTML page directly according to customers' HTTP request.
- comprehensive access in diagram 6 refers to multi categories of customers' access websites. Single access refers to a single category of customers' access websites. The rate of accuracy of comprehensive access is lower than that of single access, because the hit of customers is too low in system testing. The result states that rate of accuracy will increase with the rise of hit of effective customers.
- practice shows that the rate of accuracy of the other three kinds of customers will increase too with the rise of effective hit, besides that of new customers.
- the rate of accuracy will decrease greatly when customers change their preference which is discovered during system operation. This phenomenon is related to non-intelligence of storage mining, which needs further improvement.
- the rate of accuracy of system is related greatly to the rationality of explanation and evaluation module. The weighted average used in testing system is too generalized, which needs improving in the follow-up research.

Reference

- Guo Yan, Bai Shuo, Yu Manquan, Survey of Web Usage Mining. Computer Science, 2005;(1)
- Ning Zhong, Jiming Liu, Ron Sun. Intelligent Agents and Data Mining for Cognitive systems. Cognitive System Research, 2004;(5)
- Guo Xiujuan, Summery of Data Mining Methods, Journal of Jilin Architectural and Civil Engineering Institute, 2004;(1)
- Tsumoto S. Automatic Discovery of Positive and Negative Knowledge in Clinical Database Based on Rough set Model. IEEE

EMB Magazine, 2000;(4)

- Ioannis Kopanakis, Babis Theodoulidis. Visual Data Mining Model Techniques for the Visualization of Mining Outcomes. Journal of Visual Languages and Computing, 2003;(14)