

Research on Information Recommendation System of Internet Bookstore

Hualin Ma, Yinchu Ma, Liyan Zhang

Zhejiang Business Technology Institute, Ningbo 315012, China

email: 39679927@qq.com

Keywords: Internet bookstore; personalized service; data mining; recommendation system

Abstract. This paper proposes a strategic data-mining and recommendation method. It helps us to get the weight attribute of type of book by using AHP, and we add it in association rules. Then the method clusters the customer and type of book, and gives some strategies of personalized recommendation. Internet bookstore recommendation system is implemented with ASP.NET in this article. The experimental results indicate that the Internet bookstore recommendation system is feasible.

Introduction

This paper implemented an Internet bookstore prototype system with asp.net (C#) and Sql Server 2008. It consists of five modules, namely, membership management, customer service, background management, book management and personalized recommendations. Membership module is in charge of registering of new members, log-on of members and modification of member information. The book management is in charge of search and query, publishing of books, online query. The personalized recommendations module consists of recommendation, mining and configuration sub-modules. The whole system is shown in Fig. 1.

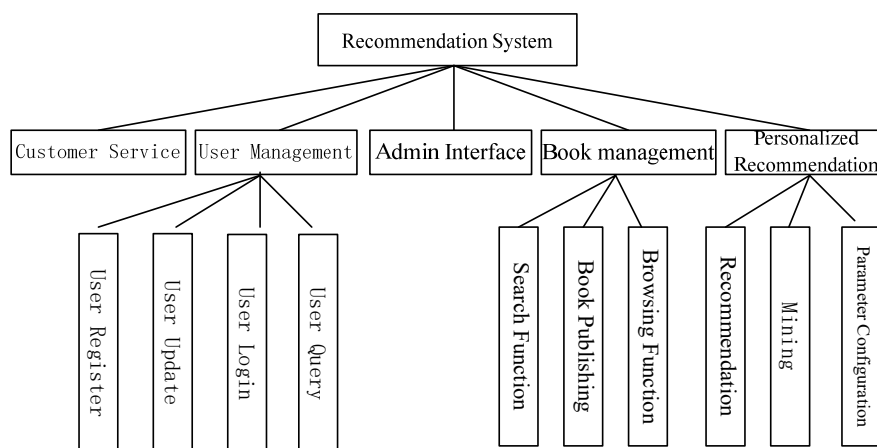


Fig.1. Functional structure

When customers first log on to the system, they will be asked to register some information and fill in the interest survey table. Then they begin to browse the web-pages and every activity will be recorded in the database. If it is an existing customer who logs on to the system, recommendations will be presented by the recommendation module according to his/her characteristic information in the database. The analysis of customers' browse activities, the mining of association rules, the clustering of book types and customers are all dealt with offline. System administrator can configure the system and manage the customers with the customer interaction module.

The algorithm of the Internet bookstore recommendation system

This paper proposes a strategic data-mining and recommendation algorithm. The former involves the calculation of the weight attribute of type of and interest for a book, the clustering of customers,

the clustering of book types, the mining of association rules. The latter involves the recommendation of the strong items within groups, the recommendation of association differentiation and the recommendation of domain knowledge. It helps us to get the weight attribute of a book type with AHP, the weight attributes spoken on behalf of its owner, and we add it in association rules. Then the method clusters the customers and book types, and gives some strategies of personalized recommendation.

The calculation of the interest weight attribute of a book type

This paper proposes two approaches to calculate the interest weight. The first is to calculate the weight with AHP according to the customers’ interest survey table. It serves as the default algorithm for the system. The second argues that customers tend to have more interest on the recently browsed books, and the weight for such books, therefore, shall be higher [1]. However it is inappropriate to consider just the browse sequence. Suppose a customer hasn’t browsed anything for a long time but browsed a lot of times within short intervals, it is meaningless if such approach is applied in such cases. Therefore in this paper the browse sequence is broken into time intervals and every interval is given its corresponding weight.

Analytical Hierarchy Process (APH) in [2] is deemed as an easy approach to make decisions for some complicated, fuzzy problems. Three steps are involved to calculate the weight among elements by means of APH, 1) yield a paired matrix; 2) get the weight among the elements from the paired matrix; 3) check the consistency.

The refined Apriori mining association rule

The common Apriori algorithm in [3] is an effective way for association rule mining. However as it doesn’t take customers’ interest into consideration when conducting mining from the frequent item sets, it is unable to mine the books which are actually to the customers’ interest in case degree of support of such books are lower than the minimum degree of support as a result of low frequency of browsing. This paper proposes a way to refine the Apriori. The weight is added during the association rule mining, in the hope of enhancing the Apriori by means of close integration of association rules and customers’ preferences.

The clustering of customers and types of books

Customers with similar interests and hobbies can be allocated to the same cluster (group) with the clustering techniques [4]. The subjects in the same cluster (group) enjoy a high level of similarity while they are different in different clusters (groups). This paper intends to cluster the customers and types of books using characteristic comparison formula shown as follows,

$$V = \frac{\sum_{i=1}^n (1 - |X_i - Y_i|)}{n} \tag{1}$$

Where n is the number of dimensions, V is the similarity between vector X and vector Y . Provided that vector X is [0 1 1 0] and Vector Y is [1 1 0 0], V is 0.5.

(1)The clustering of customers

The objective of the clustering of customers is to find the customers’ group which enjoys the similar browse activities so as to share the information in the same group. For example, we can get the browse records for Zhangqiang, Wangming and MaYu from the database. First Zhangqiang browses books involving C++, and then he browses books involving Java. According to the predefined codes, this browse activity can be recorded as [112→115]. The rest can be recorded in the same way. The browse activities of Zhangqiang and Wangming can be transformed into Table 1. We can observe their similarities and check whether they belong to the same group.

Table 1. Vector transformation table

Zhangqiang	Wangming
[112→115]	[111→113]
[112→113]	[112→113]
[115→113]	[113→115]

The vector for Zhangqiang is [000001010010000] as shown in Table 2.

Table 2. Vector transformation table for Zhangqiang

	111	112	113	114	115	116
111	0	0	0	0	0	0
112		1	0	1	0	
113			0	1	0	
114				0	0	
115					0	
116						

The vector for Wangming is [010001000010000] as shown in Table 3.

Table 3. Vector transformation table for Wangming

	111	112	113	114	115	116
111	0	1	0	0	0	0
112		1	0	0	0	
113			0	1	0	
114				0	0	
115					0	
116						

According to (1), the similarity between Zhangqiang and Wangming is 0.866. Suppose the vector for Mayu is [000010000100010], in the same way, the similarity between Zhangqiang and Mayu can be calculated as 0.8. If the threshold for the similarity is 0.8, then Zhangqiang and Wangming belong to the same group while Zhangqiang and Mayu don't.

When clustering, the system picks up one customer at first randomly, calculates the similarity with other customers, clusters the customers with their similarities above a given threshold as one group and the rest as the other group. And then the same process goes to the latter group again until no further groups can be clustered.

(2)The clustering of books

The clustering of books are roughly the same as that of customers. What is different is that books are classified according to book types which are usually defined by experts. For convenience, take computer books as an example, the types and codes are denoted as in Table 4 and Table5.

Table 4. Concept and code of computer books

Concept	Code
Software	A
Hardware	B
Method	C
System	D
Management	E
Network	F
Database	G

Table 5. Concept and Code of Category of Computer

Category of Book	Code
programming language	{ a, d }
database	{ a, d, g }
data structure	{ a, d, g }
web-page design	{ a, c, f }
introduction to computer science	{ a, b, c, d, e, f, g }
it	{ a, b, c, d, f, g }
computer hardware	{ b, c }
system maintenance	{ a, b, c }
image processing	{ a, c }
e-commerce	{ a, b, c, d, e, f, g }

Before clustering, it is necessary to transform a book type into a book type vector. for example,

the type for programming languages is {software, system}. as a which stands for software is defined as 1 and d which stands for system is also defined as 1, we can get a vector [1001000]. the transformed book type vectors are shown in Table 6.

Table 6. Vector similarity

category of book	vector	value
programming language	[1001000]	-----
database	[1001001]	$v=6/7=0.8571$
data structure	[1001001]	$v=6/7=0.8571$
web-page design	[1010010]	$v=5/7=0.7143$
introduction to computer science	[1111111]	$v=2/7=0.2857$
it	[1111011]	$v=3/7=0.4286$
computer hardware	[0110000]	$v=3/7=0.4286$
system maintenance	[1110000]	$v=4/7=0.5714$
image processing	[1010000]	$v=5/7=0.7143$
e-commerce	[1111111]	$v=2/7=0.2857$

The clustering process is the same as that in customer clustering. First pick up one type vector at random (programming languages here), calculate the similarity with other book types, cluster books with a high level of similarity as one group and the rest as the other group and repeat the same process until no further groups can be clustered. The final results are shown in Table 7.

Table 7. Clustering results of computer books

code	types included
1	(programming languages, database, data structure)
2	(web-page design, image processing)
3	(introduction to computer, information technology, E-commerce)
4	(computer hardware, system maintenance)

The strategic data-mining recommendation method

The personalized recommendations method can be classified into content filter and collaborative filter [5]. This paper proposes three recommendation strategies, namely, the recommendation of the strong items within groups, the recommendation of association differentiation and the recommendation of domain knowledge. The last one belongs to the personalized recommendation based on content filter, while the first and second one belong to the personalized recommendation based on collaborative filter and can't do without the clustering of customers.

We will elaborate the three recommendation methods with customer w1 as an example.

(1)The recommendation of the strong items within groups

The most frequently browsed items will be located and then introduced to those who haven't read this book yet. The prior favorite book ranking list will be retrieved and naturally it will change over time. For instance, the favorite book for customers within the group is .net Programming in Practice, but w1 hasn't browsed this book. In such case, the system will recommend this book to him.

(2)The recommendation of association differentiation

After the clustering of customers, w1 and a are clustered into the same group. Both a and w1 have read Advanced C# Programming and at the same time a has read The Construction and Management of Web-sites but w1 has not. In such case, the system will recommend The Construction and Management of Web-sites to w1.

(3)The recommendation of domain knowledge

After the clustering of book types, type a and b belong to the same domain. If customers have read books in type a, then books in type b which they haven't read will be recommended to them. For instance, w1 has read books in type C and C++ and the system will recommend the books such as type VB, Delphi and Java which belong to the same group programming languages to him.

As we can see, customers' interest can be widened with the help of the three methods aforesaid. Sometimes customers are not sure about their interest and will miss some books. With the help of

the personalized recommendation methods in this paper, customers will have more chance to choose books which are probably to their interest.

The system's effect

In order to check the effect of the recommendation module, tests have been conducted. The average degree of satisfaction is thus applied to observe whether the recommendations effect is influenced given that Top-N is on the increase.

The collection of experiment data

96 students in 3 classes in our university were invited to participate in the experiment for the Top-10, Top-20 and Top-30 of the three recommendation strategies. The detailed information for every book type is shown in the recommendations list web-page. After they browsed, they were asked to give the corresponding degree of satisfaction for every type. We set 5 options ranging from very satisfied, satisfied, so-so, not satisfied and very unsatisfied with the corresponding rankings ranging from 5, 4, 3, 2, 1 in the system.

The standard for the appraisal of experiment

The average degree of satisfaction is applied as the standard. First get the average degree of satisfaction for every individual as the individual average degree of satisfaction and then get the average degree of satisfaction for every individual who has received recommendation. Thus we can get the entire average degree of satisfaction. Formula is denoted as follows,

$$avg = \frac{\sum_{u=1}^m \frac{1}{n} \sum_{i=1}^n e_{u,i}}{m} \quad (2)$$

Where m stands for the number of customers, n stands for the times a customer receives recommendations.

The analysis of the experiment results

The results are shown as in Fig.2. As we can see, the average degree of satisfaction of the three recommendation algorithms ranges from 1.8 to 3.3. With the increase of the number of recommendations, the effect of recommendations decreases. Top-10 recommendation is preferred if a satisfactory recommendation effect wants to be achieved. As far as the recommendation strategy is concerned, the recommendation of the strong items within groups and the recommendation of association differentiation do a better job than the recommendation of domain knowledge.

In sum, the experiment shows that customers' interest can be mined with the recommendation system of the Internet bookstore implemented in this paper.

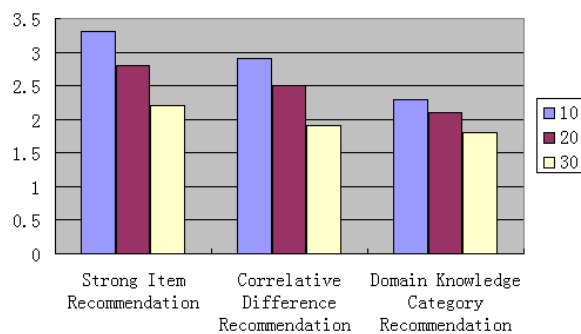


Fig. 2. Average degree of satisfaction

Acknowledgements

This project was supported by Education Department of Zhejiang province(Y201534745), by students's technology innovation practice projects of Zhejiang Business Technology Institute.

References

- [1] LU Hai-Ming, LU Zeng-Xiang, Li Yan-Da. Personal Informational Recommendation Service Based on Multi-Agent Hybrid Intelligence[J]. Communications of Advance Technical, 2001:31. (Chinese)
- [2] Introduce of AHR[EB/OL]. <http://www.doc88.com/p-61484822590.html>. (Chinese)
- [3] LIU Ji. Design of Association Rules-based Recommended Agent for business Register Access[J]. Computer Engineering, 2006,32(18): 50-51. (Chinese)
- [4] DENG Ai-Lin, ZHU Yang-Yong, SHI Bai-Le. A Collaborative Filtering Recommendation Algorithm Based on Item Rating Prediction[J]. Journal of Software, 2003,14(9):1622. (Chinese)
- [5] TAO Jian-Wen, PAN Hong-Yan. Collaborative Filtering Recommendation Algorithm Based on Similar Items and Users Rating[J]. Journal of China Society for Scientific and Technical Information, 2008, 27(2):199. (Chinese)