An Approach of Personalization for Electronic Commerce Websites Based on Ontology

Weihua Dei [1],,and Ming Yi [2]

1 Dei Weihua, School of Economy & Management, Huazhong Agriculture University, Wuhan, 430070, China
dengwhyi@yahoo.com.cn

2 Department of Information Management, Huazhong Normal University, Wuhan, 430079, China
beiyong0415@yahoo.com.cn

Abstract Aiming at the limitations of traditional personalization approaches, this article analyzes the approach based on ontology, and proposes its practical method. This approach retains the relationships both between attributes of concepts and between concepts, providing more flexibility in matching usage profiles with current user session, which can improve the precision and coverage of the recommendation sets for personalization.

1 Introduction

More recently, Web usage mining has been proposed as an underlying approach of personalization for the e-commerce website. The goal of personalization based on Web usage mining is to recommend a set of objects to the current user, possibly consisting of links, ads, text, products, etc., tailored to the user's perceived preferences as determined by the matching usage patterns. This task is accomplished by matching the current user session against the usage patterns discovered through Web usage mining. However, related approaches mainly utilize the formal characteristics of user click behaviors (syntactic information), does not utilize the internal semantics of users' click behavior (semantic information), which can improve the accuracy and coverage of the final personalized recommendation sets.

The research on utilizing semantic information of user click behaviors focus on two sides: features and ontology. B. Mobosher, Honghua Dai, Tao Luo et al. present an approach of personalization recommendation based on integrating web usage and content mining [1]; Xin Jin, Yanzan Zhou and B. Mobasher propose a maximum entropy web recommendation system: combining collaborative and content features [2]; Mooney and Roy also propose an approach of personalization recommendation based on context Categorization [3]; Raymond J. Mooney and Raymond J. Mooney present an approach of personalization recommendation based on integrating evaluating information and context features [4]. Obviously, these approaches are all

Please use the following format when citing this chapter:

Dei, W., Yi, M., 2007, in IFIP International Federation for Information Processing, Volume 251, Integration and Innovation Orient to E-Society Volume1, Wang, W. (Eds), (Boston: Springer), pp. 491-498.

based on the content features of the e-commerce website. However, such approaches cannot capture the underlying attributes of these objects and their complex relations, for example, potentially valuable relational structures among objects such as relations between books, authors and publishers of online bookshops may be missed, if one can only rely on the description of these entities using features, and the final recommendation sets may have many limitations. Hence, a few scholars propose a new approach of personalization for the e-commerce website based on ontology, which may remedy the shortcomings of the approaches based on features [5-7]. However, they have just presented a framework about this approach, and there are many problems needed to be further researched.

2 The Process of the E-commerce Website Personalization Based on Ontology

The overall process is divided into two components: the offline component and the online component.

The task of the offline component is generating semantic usage profiles based on the e-commerce website ontology. The first step is Web usage preprocessing, which includes data cleaning, user identification, session identification and path completion. The data preprocessing ultimately results in a set of syntactic transactions. The second step is clustering the syntactic transactions to discover syntactic usage profiles. Finally, according to the e-commerce website ontology, the syntactic usage profiles will be transformed into semantic usage profiles.

The task of the online component is online recommendation. Firstly, transform the current user session into semantic one, according to the e-commerce website ontology. Secondly, match the semantic current user session against the semantic usage profiles, which may result in an extended user profiles. Finally, the extended user profiles may be instantiated to real Web objects, which may be recommended to the user. For the sake of our research, this paper assumes that the e-commerce website ontology has been constructed.

3 Discovery of Semantic Usage Profile

3.1 Discovery of Syntactic Usage Profile

Web usage preprocessing ultimately results in a set of m pageview records and a set of n user transactions.

Definition 1: $T=\{t_1,t_2,\ldots,t_n\}$, T is the set of syntactic transactions as a result of Web usage preprocessing.

Definition 2: $P=\{p_1,p_2,\ldots p_m\}$, P is the set of pageview records as a result of Web usage preprocessing.

Definition 3: Each transaction $t_i = <w(p_1,t_i), w(p_2,t_i),..,w(p_j,t_i),..,w(p_m,t_i)>$ is an m-dimensional vector, $w(p_j,t_i)$ is the weight of the pageview p_j in the transaction $t_i, i \in \{1,2,\ldots,n\}$, $j \in \{1,2,\ldots,m\}$.

We may use standard clustering algorithms to partition T into groups of transactions that are close to each other based on a measure of distance or similarity. Such a clustering will result in a set $U = \{u_1, u_2, \ldots, u_k\}$ of clusters, where each u is a subset of the set of T. However, each transaction cluster contains thousands of user transactions, which are composed of millions of pageviews, so these transaction clusters cannot capture an aggregated view of common user profiles. Therefore, for each transaction cluster $u \in U$, we compute the mean vector to discover syntactic usage profiles.

Definition 4: $Tpr = \{pr_1, pr_2, \ldots, pr_k\}$, Tpr is the set of syntactic usage profiles , pr is a syntactic usage profile, which is defined as a set of pageview-weight pairs[8]:

$$pr = \{< p, weight(p, pr) > | p \in P\}.$$

Where, the weight of the pageview p in the usage profile pr, $weight(p, pr)$, is given by:

$$weight(p, pr) = \frac{1}{|u|} \times \sum_{t \in u} w(p,t),$$

and $w(p,t)$ is the weigh of the pageview p in the transaction profile $t \in u$.

3.2 discovery of semantic usage profile

Given the syntactic usage profile pr, we can transform pr into semantic web usage profile $spr = \{< o_1, ow_1 >, < o_2, ow_2 >, \ldots, < o_x, ow_x >\}$ by extracting instance from each pageview based on the e-commerce website ontology. Where,

o is a conceptual instance in the ontology, and ow is its weight. However, spr may potentially contains thousands of conceptual instances, we should combine the conceptual instances belonging to the same concept to reduce spr.

Definition 5: $C=\{c_1,c_2,\ldots,c_f\}$, $A_c=\{a_1^c,a_2^c,\ldots,a_l^c\}$, C is the set of concepts in the ontology, f is the total number of the concept, A_c is the set of attributes of the concept c, and l is the total number of the attribute of the concept c.

Definition 6: $spr=\{g_1,g_2,\ldots,g_f\}$, $g_i=\{<o_1^c,w_{o_1}^c>,<o_2^c,w_{o_2}^c>,\cdots,<o_y^c,w_{o_y}^c>\}$, $i\in\{1,2,\ldots,f\}$. Where, g_i is the set of instances in the concept c, o^c is an instance, $w_{o_y}^c$ is its weight, and y is the total number of instance in the concept c.

As for the concept c, we should provide a combination function φ_a for its each attribute a^c. The combination function φ_a can be represented by:

$$\varphi_a(<a_{o_1}^c,wa_{o_1}^c>,<a_{o_2}^c,wa_{o_2}^c>,\cdots,<a_{o_l}^c,wa_{o_l}^c>)=<o_{agg},w_{agg}>.$$

Where, a_o^c is an instance of the attribute a^c of the concept c, and wa_o^c is its weight. Further more o_{agg} is a pseudo instance of a meaning that it is an instance of the attribute a^c which does not belong to a real object in the underlying ontology, and w_{agg} is its weight.

Given a set of instances, $\{o_1^c,o_2^c,\ldots,o_y^c\}$, of the concept c, the aggregated g_i and spr may be obtained by applying the combination function for each attribute in the concept c to all of the corresponding attribute instances across all instances $o_1^c,o_2^c,\ldots,o_y^c$. Table 1 is an example of the concept "Book", and the following analysis is based on it.

Table 1. example 1 of the concept of "Book"

o^{Book}	$ow_o^{c^{Bo}}$	name	author	publisher	year	genre	…
Book1	1	{name1}	{A:1}	{publisher 1}	{2001}	book→science and technology→ computer/network→ network and communication →e-commerce	…
Book2	0.8	{name2}	{B:0.6; A:0.4}	{publisher2}	{2002}	book→ science and technology→ computer/network→ network and communication→ network management	…
Book3	0.6	{name3}	{C:0.5; B:0.3; D:0.2}	{publisher 33}	{2002}	book→ science and technology→ computer/network→ database→data warehouse and data mining	…
Book4	0.3	{name4}	{D:0.6; A:0.4}	{publisher 22}	{2003}	book→ science and technology → computer/network → network and communication →network protocol	…

The combination function φ_{name} of attribute "name" is a union operation. For example, applying φ_{name} to {< {name1}, 1>, < {name2}, 0.8>, < {name3}, 0.6>, < {name4}, 0.3>} will generate an aggregate instance about attribute "name" {<name1, 1>, <name2, 0.8>, <name3, 0.6>, <name4, 0.3>}.

As for the attribute "author", its value contains a weighted object set. In such cases we can use a vector-based weighted mean operation. The computation method of each object's weight is given as:

$$ww'_o = \frac{\sum_l ow_{o_l}^c \times ww_o}{\sum_l ow_{o_l}^c}$$

Where, $l \in \{1,2,\ldots,y\}$, y is the total number of the instance in the concept c, $ow_{o_l}^c$ is the weight of the conceptual instance, ww_o is the weight of each object (author) in the original attribute instance, and ww'_o is the weight of the aggregate object. For example, applying φ_{author} to {< {A, 1}, 1>, < {B, 0.6; A, 0.4}, 0.8>, <{C, 0.5; B, 0.3; D, 0.2}, 0.6>, <{D, 0.6; A, 0.4}, 0.3>} will generate an aggregate instance about attribute "author" {<A, 0.53>, <B, 0.24>, <C, 0.11>, <D, 0.11>}.

$$ww'_A = \frac{1\times 1 + 0.4\times 0.8 + 0.4\times 0.3}{1+0.8+0.6+0.3} = 0.53$$

$$ww'_B = \frac{0.6\times 0.8 + 0.3\times 0.6}{1+0.8+0.6+0.3} = 0.24$$

$$ww'_C = \frac{0.5\times 0.6}{1+0.8+0.6+0.3} = 0.11$$

$$ww'_D = \frac{0.2\times 0.6 + 0.6\times 0.3}{1+0.8+0.6+0.3} = 0.11$$

As for attribute "publisher", its combination function means a union operation. For example, applying $\varphi_{publisher}$ to {<{publisher1}, 1>, <{publisher2}, 0.8>, <{publisher3}, 0.6>, <{publisher2}, 0.3>} will generate an aggregate instance about attribute "publisher" {<publisher1, 1>, <publisher2, 1.1>, <publisher3, 0.6>}.

As for attribute "year", its combination function also means a union operation. For example, applying φ_{year} to {<{2001}, 1>, <{2002}, 0.8>, <{2002}, 0.6>, <{2003}, 0.3>} will generate an aggregate instance about attribute "year" {<2001, 1>, <2002, 1.4>, <2003, 0.3>}.

As for attribute "genre", it contains a partial order representing a concept hierarchy among different genre values. The combination function, in this case, can perform tree (or graph) matching to extract the common parts of the conceptual hierarchies among all instances. For example, applying φ_{genre} to the example of table 1 will generate an aggregate instance about attribute "genre" {book→ science and technology → computer/network}.

Hence, aggregate semantic web usage profile $nspr$ will be formed by implying the combination functions to all attributes, which may be defined as:

$$nspr = \left\{ <o_1, nw_1>, <o_2, nw_2>, \ldots, <o_f, nw_f> \right\}.$$

Where, o is the aggregate instance of each concept formed by performing combination function, nw is its weight, which can be determined by the significance of the concept in the e-commerce website domain ontology. Therefore, table 1 can be transformed into table 2.

Table 2. example 2 of the concept "Book"

o^{Book}	name	author	publisher	year	genre	…
Book	{<name1, 1>, <name2, 0.8>, <name3, 0.6>,	{<A, 0.53>, <B, 0.24>, <C, 0.11>,	{<publisher1, 1>, <publisher2, 1.1>, <publisher3,	{<2001, 1>, <2002, 1.4>, <2003, 0.3>}	book→ science and technology → computer/ network	…

	<name4, 0.3>}	<D, 0.11>}	0.6>}			

4 Online Recommendation

In online recommendation phase, syntactic current user session should be transformed into semantic current user session $S=\{<o_{s_1},ws_1>,<o_{s_2},ws_2>,\cdots,<o_{s_f},ws_f>\}$ firstly. Where, o_s is the aggregate instance of each concept formed by performing combination functions, ws is its weight representing the significance of the concept in the ontology. Secondly, match semantic current user session against the semantic web usage profiles by means of the semantic similarity measures [9].

As for each $nspr=\{<o_1,nw_1>,<o_2,nw_2>,\ldots,<o_f,nw_f>\}$, the similarity $Sim(S,nspr)$ of semantic vector s and $nspr$ lies on the semantic similarity $SemSim(s_i,nspr_i)$ of each aggregate conceptual instance:

$$Sim(S,nspr)=\sum_{i=1}^{f}\alpha_i\times SemSim(o_{s_i},o_i)$$

$$SemSim(o_{s_i},o_i)=\sum_{j}\beta_j\times Simlarity(o_{s_i}.a_j,o_i.a_j)$$

Where, α_i is the significance of concept c in the e-commerce website ontology, and β_i is the significance of the attribute a_j in the concept c.

According to the above analyses, the computation of $SemSim(o_{s_i},o_i)$ and $Sim(S,nspr)$ can be accomplished. Therefore, as for semantic current user session S and each semantic web usage profile $nspr$, the recommendation value $\mathrm{Re}c^o(o^{c_i},S)$ of each aggregate conceptual instance o^{c_i} is denoted as:

$$\mathrm{Re}c^o(o^c,S)=\begin{cases}0, & if\ \ o^c\in S\\ nw\times Sim(S,nspr), & otherwise\end{cases}$$

Where, nw is the weight of the aggregate conceptual instance o^c in $nspr$, $Sim(S,nspr)$ is the similarity of the semantic current user session S and the semantic usage profile $nspr$ which has the conceptual instance o^c. And the extended user profile $URec(S)$ can be represented by:

$$URec(S)=\{<o^c,\mathrm{Re}c^o(o^c,S)>\mid \mathrm{Re}c^o(o^c,S)\geq\varepsilon\}.$$

Those aggregate conceptual instances whose recommendation value is less than a certain threshold ε will be filtered out. Then, the extended user profiles may be instantiated to real pageviews, so that appropriate pageviews may be recommended to the user for the purpose of personalization.

5 Conclusions

In this paper we have proposed an approach of the e-commerce website personalization Based on Ontology. This approach retains the relationships both between attributes of concepts and between concepts, providing more flexibility in matching usage profiles with current user session, which can improve the precision and coverage of the recommendation sets for personalization. The examples provided throughout this paper reveal how such a framework can provide insightful patterns and smarter personalization services. However, We have only provided an overview of the relevant issues and suggested a road map for further research and development in this area. One area of future work involves the study of machine learning techniques in order to discover the best way to summarize the attribute automatically. Another area of future work will be to explore use of discovered domain-level aggregates from Web usage mining to enrich the existing domain ontology for a Web site.

References

1. B. Mobasher, Honghua Dai, Tao Luo et al. Integrating Web Usage and Content Mining for More Effective Personalization. In E-Commerce and Web Technologies: Proceedings of the EC-WEB 2000 Conference, Springer, 2000.165～176

2. Xin Jin, Yanzan Zhou and B. Mobasher. A Maximum Entropy Web Recommendation System: Combining Collaborative and Content Features. In KDD'05, Chicago, 2005.21～24

3. Raymond J. Mooney and Raymond J. Mooney. Content-Based Book Recommending Using Learning for Text Categorization. In Proceedings of the Fifth ACM Conference on Digital Libraries,San Antonio, 2000.195～240

4. C. Basu, H. Hirsh and W. Cohen. Recommendation as classification: Using social and content-based information in recommendation. In Proceedings of the AAAI-98, 1998.714～720

5. H. Dai and B. Mobasher. Integrating Semantic Knowledge with Web Usage Mining for Personalization. In Web Mining: Applications and Techniques.Anthony Scime (ed.), Idea Group Publishing, 2005

6. Bangyong Liang, Juanzi Li and Kehong Wang. Web page recommendation model for the semantic web. Journal of Tsinghua Univ (Sci&Tech), 2004(9):1271～1277

7. Hongyan Pan, Hongfei Lin and Jin Zhao. Ontology-Based Personalized Recommendation system. Computer Engineering and Applications, 2005(20):176～180

8. B. Mobasher, Honghua Dai, Tao Luo, et al.Discovery of Aggregate Usage Profiles for Web Personalization. In Proc. of the Web Mining for E-Commerce Workshop (WEBKDD'00), Boston, 2000

9. B. Mobasher and Honghua Dai. A road map to more effective web personalization: Integrating domain knowledge with web usage mining. In International Conference on Internet Computing, 2003