Praise Worthy Prize

International Review on Computers and Software (I.RE.CO.S.), Vol. 8, N. 6
ISSN 1828-6003 June 2013

Building Automatic Web Customer Profiling Service

B. Alkhatib, A. Alnahhas, H. Ezaldeen

Abstract – *As the web has spread widely in the past years, it is now very important to personalize and customize websites so that we can improve user experience. Customer profiling is one of the most important methods to add personalization and customization to websites as it can capture the properties and interests of the customer in order to use them later to service him properly. In this paper, we present a generalized automatic customer profiling model that can suite most websites. Our model is based on probability theory; it captures user behavior slots in order to predict his profile. These probabilistic slots values are used to generate the customer adequate services. We show how to incorporate this model in a web service so that any website developer can use this service to add automatic user profiling to his website.* ***Copyright © 2013 Praise Worthy Prize S.r.l. - All rights reserved.***

***Keywords**: Customer Profiling, Automatic Web Site Personalization, Web Service*

Nomenclature

F	Features (Slots) set of customer profile
f_i	Feature i
C	Customer
P_c	Profile of customer C
s_i	Certainty (Value) of C having feature F_i
A	Actions set that can be performed by customer
a_i	Action i
H	Change of s_i for some feature f_i
V	Services set that can served to customer
v_i	Service i
o_i	Score of service v_i
$l_{a,f}$	Certainty of a user who does action a has the feature f
$t_{f,v}$	Likelihood of a customer has the feature f should be served with the service v

I. Introduction

Customer profiling [1] is the process of identifying your customers to reach them and know why they are in your website. User profile is a set of information about a specific customer. Many methods have been presented earlier in the field of automatic customer profiling. Generally, these methods can be classified into three major classes: Content-based profiling service [2], [3], [4], that is based on customer evaluations, these evaluations are used to predict recommendations for other customers.

The second class is collaborative filtering systems [14], [15], [16] that depend on the similarity among customers, where each customer get a recommendation that is chosen by similar customers.

Some systems [5] combine both and are called hybrid systems.

In this paper, we present a new model for customer profiling, which is described in section 2, section 3 shows how we can use this model as a web service, we show some implementation details in section 4, and results in section 5, we finally conclude in section 6.

II. The Proposed Customer Profiling Model

One can easily observe that customers' behavior on the web can be very expressive, and reflects customers' properties. For example, people surfing books section of a website are mostly educated.

Our model defines the customer profile as a set of numbers; each number represents the certainty that the user has a specific feature, so if we denote F as a set of possible customer profile slots:

$$F = \{f_1, f_2, \dots, f_n\}$$

For example f_1 can be "is married", f_2 is "is Student", etc…

The customer profile P_c for a customer C is the set of numbers $s_1, s_2, \dots, s_n$ where s_i is the certainty of customer C being f_i, for former example if:

$$P_c = \{0.2, -0.7, \dots\}$$

It means that the certainty of (C is married) is 0.2 and the certainty of (C is student) is -0.7.

In our model certainty values are within the range [-1,1], the value of 0 indicates complete uncertainty, the value of +1 indicates complete certainty of customer having the feature s_i, whereas -1 indicates complete certainty of customer not having this feature.

Manuscript received and revised May 2013, accepted June 2013 *Copyright © 2013 Praise Worthy Prize S.r.l. - All rights reserved*

As we mentioned above, the customer profile is related to customer actions on the web. This fact can be modeled as: if Customer C does an action a then it is possible that he has the feature f with certainty value s.

Formally, let A be the set of actions users can do:

$$A = \{a_1, a_2, \ldots, a_m\}$$

For example a_1 can be "visited geography books section", a_1 can be "bought cooking book".

Let's define a function L:

$$L: A \times F \rightarrow [-1{,}1]$$

where $l_{a,f}$ is the certainty of a user who does action a has the feature f, the value of $l_{a,f}$ is within the range $[-1{,}1]$ as we mentioned above. These values are provided by a domain expert and are considered as inputs to the proposed model.

If Customer C has a value s_i for some feature f_i in his profile, doing an action a_j will affect s_i as follow:

$$s_i = \begin{cases} l_{a,f}, & s_i = 0 \\ s_i + H, & \text{otherwise} \end{cases} \tag{1}$$

where H is:

$$H = |s_i \times (1 - s_i)| \times l_{a,f} \tag{2}$$

Note that H corresponds to the graph in Fig. 1 and reflects the change that is applied to s_i.

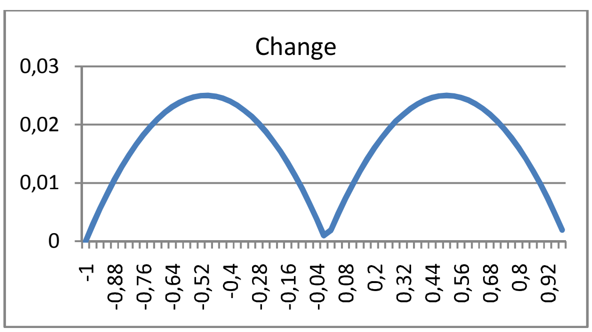

Fig. 1. The change of the slot value

The change H is chosen carefully to meet the following conditions:

- The value of s_i shouldn't go outside the interval $[-1, 1]$.
- The value of H should not be inversely proportional to the value of s_i.

The first condition is achieved by the $(|1 - s_i|)$ component of H, the second condition is achieved by the (s_i) component.

For example: if a_i is "buying a math book" and f_i is "being a student", $l_{a,f}$ can be 0.4, note that the value of L is set by an expert and is an input for the model.

If $s_i = 0.6$, then applying Eq. (2) would result:

$$H = |0.6 \times (1 - 0.6)| \times 0.4 = 0.096$$

The new value of s_i is:

$$s_i = 0.696$$

Note that the value of H is getting smaller as s_i is getting larger than 0.5, in this way s_i will reach certainty after a significant number of actions, normally.

Besides, the sign of H value is related to the sign of L, so the characteristics of H are identical in both the positive and negative fields.

The aim of customer profiling is to provide the best service for each customer. To achieve this goal, we should link the customer profile to the available service set that can be denoted as V, this set represents all available services that can be served:

$$V = \{v_1, v_2, \ldots, v_m\}$$

v_1 can be for instance "Show books ads".

It is clear that the service is related to the profile slots of the customer in some way. For example: if customer "is a student" then it is good to "Show books ads"

Let's define the function T:

$$T: F \times V \rightarrow [-1{,}1]$$

where $t_{f,v}$ is the likelihood of a customer has the feature f should be served with the service v. Values of this function are determined by a domain expert and are considered as inputs to the model along with the values of function L described above.

We can find that each service is linked to all profile slots using the function T, Let's define the score function o_j for the service j as follows:

$$o_j = \sum_i s_i \times t_{i,j} \quad for\ all\ profile\ slots\ i \tag{3}$$

Now each service has a score according to the function above (3), the best service is the one with the highest score, o.

For example: if we have two services: v_1 which is "Show book ads" and v_2 which is "Show stationary ads", suppose we have 3 profile slots:

f_1	Is student
f_2	Like books
f_3	Is married

Let the values of the function T are in the following table:

	f_1	f_2	f_3
v_1	0.6	0.9	0
v_2	0.7	0	0.4

The values of the above table are provided by expert in the business domain.

Copyright © 2013 Praise Worthy Prize S.r.l. - All rights reserved

If a customer C has the current profile as follows:

f_1	0.4
f_2	-0.1
f_3	0.3

We can calculate the score function value of both services by applying Eq. (3) as follows:

$$o_1 = 0.4 \times 0.6 - 0.1 \times 0.9 + 0.3 \times 0 = 0.15$$

$$o_2 = 0.4 \times 0.7 - 0.1 \times 0 + 0.3 \times 0.4 = 0.4$$

The score of the second service is larger than the score of the first, so the user will receive the second service as it can fit his interests more likely.

III. Customer Profiling as a Web Service

One can find that the model described in the previous section has many variables that control the process of finding the best service, namely those variables are F, A, L, V, and T, the existence of those variables makes the model general and suitable for many cases where each case can be represented by applying a change to them.

Actually these variables are much related to business specifications where the customer profiling is applied, and can be set by experts in that business domain.

Thus, we can provide the customer profiling as a service configured using the variables above. This service has two basic interfaces, one to provide customer actions and another to get the score of the services, the client then chooses to view the highest score service.

The client here is a website that wishes to add customer profiling feature in order to provide user-specific services. Website developers can use an interface provided by the system based on the model proposed to include customer profiling taking into consideration the following two points:

1. When a customer does an important action, the developer should call the web service interface to tell about that action.
2. When the service should be provided, the developer should call the web service to get the best service to provide.

An expert in a specific business domain is provided by an interface used to set the domain-related variables mentioned at the top of this section, Fig. 2 illustrates the process of the system.

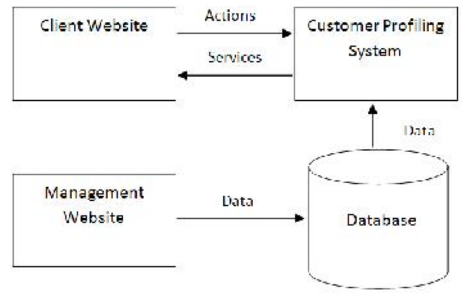

Fig. 2. The proposed service model

IV. Implementation

As mentioned earlier, the service we propose in this paper is provided for websites that need to add customer profiling for personalization, so it should be provided as a web service. Practically, we use WCF [17] to implement the service, so we support ASP.net website along with other types of website development languages as WCF has back compatibility with SOAP.

Our implementation includes the following messages that can be sent to the service:

Message	Description
AddClient	Called when a new customer is registered on the client website
AddAction	Called from the interface website to add a new action type
AddProfileSlot	Called from the interface website to add a profile slot
AddService	Called from the interface website to add a service type
SetActionSlotRatio	Called from the interface website to change the relation between some action and some slot
SetSlotGoalRatio	Called from the interface website to change the relation between some slot and some service
ClientDidAction	Called from client website when the customer did an action
GetbestGoalForClient	Called from client website to get the best service for one customer
GetProfileOfClient	Called from the interface website to view some customer current profile
GetActionsOfCustomer	Called from the interface website to get the actions of one website
GetProfileSlotsOfCustomer	Called from the interface website to get available profile slots for one website
GetServicesOfCustomer	Called from the interface website to get the services of one website
GetActionSlotRation	Called from the interface website to get the relation between some action and some slot
GetSlotGoalRatio	Called from the interface website to get the relation between some slot and some service
CheckUserValidity	Used to check the client credentials.

V. Online Book Search Website as a Testing Example and Results

In order to test our service we have built a standalone website that is used to search for books online.

This website uses customer profiling to enhance search results, where the books in search results are sorted according to the customer profile. To use the proposed model in this website we should define actions, profile slots, services and relations between actions and profile slots and between profile slots and services. An action of opening book details of some category T is considered to be T; hence we have one action for each book category. In our example, we have 64 book categories; some of them are in the following table:

Copyright © 2013 Praise Worthy Prize S.r.l. - All rights reserved

Accounting	Diagnosis & Skills	Geology
Allergy	Cooking	Haematology
Alternative Med	Dictionaries	Histology

The profile slots are chosen to be related to books domain, and they are as follow:

scientific educated
literally educated
Student
general educated
Gender
Underage
Elder
sport fans
art lovers
not academic

Each book category corresponds to service type, so customer profiling in our case study is used to sort search results according to customer interests. Books with categories that correspond to services with higher scores should be boosted.

If the book score without customer profiling service is $Score_1$ and the score of the book category is $Score_2$ then the books are sorted according to function Score where:

$$Score = Score_1 + Score_2$$

The values of the function that relates actions with slots are considered the same as the function that relates slots to services. Those values are chosen with a help from a book store supervisor. The following table shows an example of these values:

value	scientific educated	Underage	Student
Biology	0.5	0.1	0.3
Bioscience	0.7	-0.2	0.2
Cardiology	0.5	-0.2	0.3
Chemistry	0.8	-0.2	0.4
Civil Engineering	0.8	-0.2	0.4
Computing	0.7	-0.2	0.6
Critical Care	0.9	-0.2	0.5

We have published our website online, and about two hundred customers used it.

As the aim of the customer profiling is to view books that interests user at the top of his search results, we take the measure of the first book clicked from the search results to test our model and we calculated the average of the results for all customers.

Fig. 3 shows that choosing the first book is getting better as the customer keeps trying; that is, book that interests a customer is getting up each time he searches for books.

VI. Conclusion and Future Work

We showed in this paper how a generalized customer profiling model that can suite many cases on the web can be built.

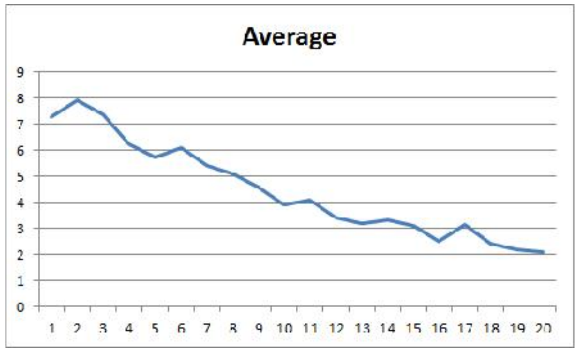

Fig. 3. Shows how the first chosen result of the user is getting better

We also showed how we can apply this model using web service model and demonstrated a brief implementation issues, and finally presented a practical example with results.

The model proposed can be revised to include more issues such as the relation among profile slots themselves. We also can improve the logical model to include fuzzy logic as it can reflect real world more precisely.

References

[1] Palgrave Macmillan: Customer profiling in e-commerce: Methodological aspects and challenges. The Journal of Database Marketing, Volume 9, Number 2, 1 January 2002 , pp. 170-184(15).

[2] Mulvenna, M., Anand, S.S., Buchner, A.G.: Personalization on the net using web mining. Communication of ACM 43

[3] Intelligent Techniques for Web Personalization. B. Mobasher and S.S. Anand. Springer-Verlag Berlin Heidelberg 2005.

[4] Automatic Personalization Based on Web Usage Mining. B. Mobasher, R. Cooly, and J. Srivastava. COMMUNICATIONS OF THE ACM.

[5] Analysis of Recommendation Algorithms for E-Commerce. B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. ACM.

[6] Knowledge Discovery from Users Web-Page Navigation. C. Shahabi, et al. IEEE.

[7] Category-Based Web Personalization System. Ching-Cheng Lee, Wei Xu. IEEE.

[8] Improving New User Recommendations with Rule-based Induction on Cold User Data. An-Te Nguyen.

[9] Yoda: An Accurate and Scalable Web-based Recommendation System. C. Shahabi, F. Banaei-Kashani, Y. Chen, D. McLeod. In the Proceedings of The Sixth International Conference on Cooperative Information Systems. Trento, Italy. (September 2001).

[10] Improving User Profiles for E-Commerce by Genetic Algorithms. Y. Chen, C. Shahabi. Integrated Media Systems and Computer Science Department. USC, LA.

[11] Mladenic, D.: Personal web watcher: Implementation and design. Technical Report IJSDP- 7472, Department of Intelligent Systems, J. Stefan Institute, Slovenia.

[12] Krulwich, B., Burkey, C.: Learning user information interests through extraction of semantically significant phrases. In: Proceedings of the AAAI Spring Symposium on Machine Learning in Information Access.

[13] Goldberg, D., Nichols, D., Oki, B.M., Terry, D.: Using collaborative filtering to weave an information tapestry. Communications of the ACM 35.

[14] Herlocker, J. L.; Konstan, J. A.; Terveen, L. G. &Riedl, J. T. “Evaluating collaborative filtering recommender systems”, ACM Trans. Inf. Syst. 22 (1): 5–53 (January 2004).

[15] Amazon.com Recommendations, Item-to-Item Collaborative

Copyright © 2013 Praise Worthy Prize S.r.l. - All rights reserved

Filtering. G. Linden, B. Smith, and J. York. IEEE INTERNET COMPUTING (February 2003).
[16] Ungar, L., Foster, D.P.: Clustering methods for collaborative filtering. In: Proceedings of the Workshop on Recommendation Systems.
[17] David Chappell, Chappell & Associates: Introducing Windows CommunicationFoundation, Microsoft publication, July 2013.

Authors' information

Syrian Virtual University, Syrian Arab Republic.

Bassel Alkhatib is the web science master director at the Syrian Virtual University and the head of Artificial Intelligence department at Information Technology Faculty at Damascus University. He holds PhD degree in computer science from the University of Bordeaux-France, 1993. Dr. Alkhatib supervises many PhD students in web mining, and knowledge management. He also leads and teaches modules at both BSc and MSc levels in computer science and web engineering in both Syrian Virtual University and Damascus University

Ammar Alnahhas is a teacher assistant at the Syrian Virtual University and Damascus University, he has got the Master degree in Artificial Intelligence from Damascus University, He has achieved a lot of interesting projects in the field of AI and web engineering, and has participated in many conferences.

Hadi Ezaldeen is a Web Science master student. Bachelor of information technology degree, SVU 2009.

Copyright © 2013 Praise Worthy Prize S.r.l. - All rights reserved