Information Systems 49 (2015) 182–199

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/infosys

Contextual preference mining for user profile construction ☆

Sandra de Amo [a], Mouhamadou Saliou Diallo [b,c], Cheikh Talibouya Diop [c], Arnaud Giacometti [b], Dominique Li [b], Arnaud Soulet [b]

[a] Universidade Federal de Uberlândia, Av. João Naves de Avila 2121 38400-902 Uberlândia, MG, Brazil
[b] Université François Rabelais de Tours, 3 Place Jean Jaurès 41029 Blois, France
[c] Université Gaston Berger, BP 234, Saint-Louis, Senegal

ARTICLE INFO

Article history:
Received 5 February 2014
Received in revised form
13 November 2014
Accepted 22 November 2014
Recommended by: F. Korn
Available online 3 December 2014

Keywords:
Pattern mining
Preference elicitation
Contextual preference rule
User profile mining

ABSTRACT

The emerging of ubiquitous computing technologies in recent years has given rise to a new field of research consisting in incorporating context-aware preference querying facilities in database systems. One important step in this setting is the *Preference Elicitation* task which consists in providing the user ways to inform his/her choice on pairs of objects with a minimal effort. In this paper we propose an automatic preference elicitation method based on mining techniques. The method consists in extracting a *user profile* from a set of user preference samples. In our setting, a profile is specified by a set of contextual preference rules verifying properties of soundness and conciseness. After proving that the problem is NP-complete, we propose a resolution in 2 phases. The first phase extracts all individual user preferences by means of contextual preference rules. The second phase builds the user profile starting from this collection of rules using a greedy method. To assess the quality of user profiles, we propose three ranking techniques benefiting from these profiles that enable us to rank objects according to user preferences. We evaluate the efficacy of our three ranking strategies and compare them with a well-known ranking method (SVMRank). The evaluation is carried out through an extensive set of experiments executed on a real-world database of user preferences about movies.

1. Introduction

The topic of enhancing database systems with context-aware preference querying functionalities has been attracting a lot of interest in the database community in recent years, due to the variety of applications ranging from e-commerce to personalized search engines where user's preferences are essential in the system model design. Several important research work has been dedicated to this topic, including research on the development of (1) powerful frameworks for preference modeling and reasoning [1,2] and (2) preference query languages with high declarative and expressive power for personalized database applications [3–6]. However, little work has been dedicated so far to the topic of *preference elicitation*, that is, in the way user preferences are obtained. This paper is focused on this topic.

Elicitation of preferences consists basically in providing the user a way to inform his/her preferences on objects belonging to a dataset, with a minimal effort for him/her. It can be achieved by following different strategies: (a) by using a query interface where users are asked to express their preferences [1], or (b) by capturing implicit user's choices and applying preference mining algorithms [7]. The first alternative is not

☆ This work has been financially supported by the French Embassy of Dakar, the Brazilian Research Agencies CNPq, CAPES (SticAmSud Project 016/09) and FAPEMIG.

E-mail addresses: deamo@ufu.br (S.d. Amo), ms.diallo@hotmail.fr (M.S. Diallo), cheikh-talibouya.diop@ugb.edu.sn (C.T. Diop), arnaud.giacometti@univ-tours.fr (A. Giacometti), dominique.li@univ-tours.fr (D. Li), arnaud.soulet@univ-tours.fr (A. Soulet).

http://dx.doi.org/10.1016/j.is.2014.11.009
0306-4379/

efficient since the users in general are not able to express their preferences in an exact and consistent way. This paper is focused on the second alternative for preference elicitation. We assume our data is constituted by pairwise comparisons. We do not discuss in this paper the way the user informed his/her choices, knowing that different strategies can be applied [8]. Our method simply assume that *pairs of objects* expressing the user preferences have been collected somehow. The running example below illustrates the preference mining problem we tackle in this paper. In this example we assume that the user preferences are informed by means of the number of clicks on certain tags.

Motivating example. A web service regularly provides recommendation about movies to its subscribers. In order to capture their preferences on films without being too annoying and intrusive, the service offers its subscribers a trial period during which they can freely access information about films. They indicate the films they are interested in by clicking on different tags. For instance, a subscriber can click on tags *Action*, *Spielberg* and *War* to indicate that he/she is interested in obtaining information on films directed by Steve Spielberg, with a script based on a war story, and having a lot of action. His/her clicks are automatically collected during the trial period. The relation $\mathcal{D}$ depicted on Table 1 presents some of the access of a subscriber during a trial period. Tags A, B, C, D and E stands for *Spielberg*, *Tom Hanks*, *Action*, *Leonardo di Caprio* and *War* respectively. Each $t_i\,(i=1,\ldots,5)$ represents the set of tags I selected each time the subscriber accessed the service. They are called *transactions*. Let us suppose that during the trial period the subscriber accessed the service ten times by clicking on the set of tags t_1 and only five times by clicking on the set of tags t_3. Thus, he/she implicitly indicated that he/she is more interested on films associated to tags t_1 than to tags t_3 as indicated by the first pair $\langle t_1, t_3\rangle$ in relation $\mathcal{P}$ depicted in Table 1. Notice that both t_1 and t_3 contain the tags A(Spielberg) and C(Action). Between them he/she prefers the one containing the tag D(Leonardo di Caprio) than the one containing the tag B(Tom Hanks). So, the following *contextual preference rule* can be inferred: Between two *action* movies directed by *Spielberg* he/she prefers the one played by Leonardo di Caprio than the one played by Tom Hanks. Tags *Action* and *Spielberg* constitute the *context* of the rule. Notice that some pairs of transactions (for instance, $\langle t_1, t_2\rangle$) do not appear in relation $\mathcal{P}$, indicating that the number of clicks on each of these sets of tags is the same or differs by a *negligible* amount of clicks (below a given threshold).

In this paper we propose the method *ProfMiner* for building the profile of a user from a sample of his/her preferences previously captured by the system. A user's profile is specified by a set of *contextual preference rules* [6] satisfying some interestingness criteria, namely *soundness* and *conciseness*. The *soundness* property guarantees that the preference rules specifying the profiles are in agreement with a large set of the user preferences, and contradicts a small number of them. On the other hand, *conciseness* implies that profiles are small sets of preference rules. Formally, our problem is to find a set of contextual preference rules that minimizes a cost function that takes into account the soundness criterion. As this problem is NP-complete, our method consists of two separated phases: first, a set of contextual preference rules S is mined; and second a user profile is extracted from S. We argue that our approach based on preference rules has many advantages if compared to other preference models found in the literature. The model is easy to understand and manage due to its conciseness and its *qualitative* aspect (it is constituted by a set of preference rules and it does not employ score functions explicitly assigning grades to each transaction [9–12]). Moreover, the soundness property guarantees that our method builds user profiles with good predictive properties.

Besides introducing the method *ProfMiner* for building the user profile, we also propose three ranking strategies which make use of the mined profile in order to rank objects according to the user preferences. The purpose of these ranking strategies is not only to show the usefulness of built user profiles but also to assess their quality. We obtain in this way three different ranking methods, all of them using the mined profile coupled with a particular ranking strategy. We compare the accuracy of these ranking methods with each other as well as with a well-known ranking technique (the SVMRank introduced in [11]) throughout a series of experiments executed on a real-world database of user preferences about movies. Our experiments show that some of our

Table 1
A transactional database and a preference database.

$\mathcal{D}$

Tid	Transactions				
t_1	A		C	D	
t_2	A	B		D	
t_3	A	B	C		E
t_4			C	D	
t_5	A	B			

$\mathcal{P}$

Pid	User preferences
p_1	$\langle t_1, t_3\rangle$
p_2	$\langle t_2, t_3\rangle$
p_3	$\langle t_2, t_4\rangle$
p_4	$\langle t_3, t_4\rangle$
p_5	$\langle t_4, t_5\rangle$