



A data-mining approach to preference-based data ranking founded on contextual information



Antonio Miele, Elisa Quintarelli, Emanuele Rabosio*, Letizia Tanca

Politecnico di Milano, Dipartimento di Elettronica e Informazione, via Ponzio 34/5, 20133 Milano, Italy

ARTICLE INFO

Article history:

Received 27 September 2012

Received in revised form

3 December 2012

Accepted 4 December 2012

Recommended by: D. Shasha

Available online 20 December 2012

Keywords:

Data personalization

Context-awareness

Pervasive computing

Context-aware preference systems

ABSTRACT

The term *information overload* was already used back in the 1970s by Alvin Toffler in his book *Future Shock*, and refers to the difficulty to understand and make decisions when too much information is available. In the era of Big Data, this problem becomes much more dramatic, since users may be literally overwhelmed by the cataract of data accessible in the most varied forms. With *context-aware data tailoring*, given a target application, in each specific context the system allows the user to access only the view which is relevant for that application in that context. Moreover, the relative importance of information to the same user in a different context or, reciprocally, to a different user in the same context, may vary enormously; for this reason, *contextual preferences* can be used to further refine the views associated with contexts, by *imposing a ranking on the data of each context-aware view*. In this paper, we propose a methodology and a system, PREMINE (PREference MINEr), where data mining is adopted to infer contextual preferences from the past interaction of the user with contextual views over a relational database, gathering knowledge in terms of *association rules* between each context and the relevant data.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

The current ecosystem of available digital information represents an unprecedented opportunity for the users, but at the same time risks to overwhelm them during decision-making [1]. The effect of this problem is amplified for users who access data by means of mobile devices, which are equipped with limited resources and connectivity and thus impose that only the most valuable information should be kept on board. Imagine you want to keep on your smartphone some information for on-line trading but also to support your shopping activity and your travels: some of the personal data you need for these operations resides on your device, but keeping what is necessary for *all three* operations on the smartphone all

the time is not really sensible. Instead, eliminating, at any time, the redundant information will speed-up your work both in terms of device efficiency and of the effectiveness that you can achieve by working in the absence of *information noise*.

However, distinguishing useful data from all the information which is irrelevant to the specific application or user is not a trivial task, since the same piece of information can be considered differently, even by the same user, in different situation or places—in a single word, in a different *context*.

This emergent problem has been tackled in the literature by introducing context models (see [2–4] for surveys) allowing the personalization of data repositories on the basis of a set of perspectives, or *dimensions*, such as the user's role and location, the time, his or her interests and the situations he or she is involved in [5]. However, data personalization based on context may be only a partial solution, since the tailoring of the available dataset may still be too coarse-grained. For example, if we consider a movie dataset and Bob – a young teenager who is interested in

* Corresponding author. Tel.: +39 02 2399 3482.

E-mail addresses: miele@elet.polimi.it (A. Miele), quintare@elet.polimi.it (E. Quintarelli), rabosio@elet.polimi.it (E. Rabosio), tanca@elet.polimi.it (L. Tanca).

movies – a contextual system will suggest the movies played in cinemas close to Bob's location and appropriate for people of his age, but will not be able to propose any ranking or further filtering of this contextual data according to Bob's personal tastes: for example, Bob might like watching comedies when alone and thrillers when with his friends.

Therefore, to attain more effective personalization, this work couples the notion of context with the user personal preferences: this allows to rank the information delivered to Bob differently in each different context (alone or with friends).

The approaches already proposed for personalizing relational data (tuples or attributes) on the basis of contextual preferences [6–8] rely on the collaboration of the users for preference indication. However, with a large variety of data and a considerable number of possible contexts, the manual specification of an extensive list of preferences may be a trying experience which discourages the user. A way around this problem is exploiting other information, implicitly provided by the past querying activity of the user. This activity can be of various kinds, e.g. Bob might formulate queries to visualize the titles of the available comedies, then select "The Muppets" to see further details and subsequently repeat the same operation for other Disney movies in the list, and finally decide to watch one of them. A system analyzing Bob's activity may discover that he is often attracted by Disney comedies.

Given this rationale, this paper's contribution is the PREMINE (PREference MINER) methodology and the related system, which use data mining algorithms to learn the contextual preferences of the users on both tuples and attributes of relational databases. Our interest towards the relational technology is motivated by the fact that most commercial databases, and also a significant part of the deep web rely on it, therefore handling relational preferences, have long been recognized as an important issue [9].

Contextual preferences are thus used to further personalize the set of data associated with each context (called *contextual view*) and can be applied with two goals: (1) to minimize the information noise, presenting a list of the data ordered by their relevance for the user with the effect of "recommending" the highest-ranked data, (2) to fulfill the memory requirements imposed by small devices, by loading only the data which have been ranked high according to the user preferences. Our approach starts from the contextual preference model introduced in [7] and adds a sophisticated technique to mine *contextual association rules* (that is, co-occurrences between each context and the browsed data) from the past interaction of the user with the contextual views over a given relational dataset.

Although there are several degrees of freedom for the personalization, leading to a large set of possible approaches, in this paper we focus on the preference mining part and give a quick account of how the mined preferences are used to produce the personalized contextual view.¹ Also, we

```

CINEMA(cinema_id, name, address, zipcode, city, cinema_type)
MOVIE(movie_id, title, genre, year, director)
SHOWING(movie_id, cinema_id, day, time)
ACTOR(actor_id, name, lastname, citizenship)
MOVIE_ACTOR(movie_id, actor_id, mainrole)

```

Fig. 1. Database schema of the running example.

remark that our proposal, differently from the majority of recommendation systems, *does not require any explicit input from the users* about their preferences.

The procedure goes as follows: on the user's device runs a client application accessing a contextual view of the global database. This portion of data is initially selected only on the basis of the user current context; the user's querying activity and subsequent browsing in the list of the returned tuples allow the PREMINE server-side application to gain knowledge about the correlations between a context and the properties of the data preferred in that context. Afterwards, when the device connects to the application server, this knowledge is used to further filter and personalize the contextual view.

Note that the proposed approach does not completely exclude the manual specification of preferences; in fact, the two approaches can be used in conjunction: the user can manually add preferences or adjust the mined ones, when they do not reflect any more his or her actual needs. Some encouraging experiments performed with real users interacting with the dataset of a European company of video on demand show the practical impact of our proposal.

Running example: Fig. 1 shows the relational schema of the running example we use throughout the paper (a simplification of the mentioned case study), namely the information system of a company offering services of video on demand and reservation of movie tickets. All the applications composing the information system rely on a central database storing all the managed information. This database is also used for the experimental session at the end of the paper.

Paper structure: The structure of the paper is as follows. Section 2 presents the state of the art, Section 3 introduces some preliminary notions and Section 4 presents the mining framework. Sections 5 and 6 describe our strategies for mining preferences, respectively, on tuples and attributes. Section 7 shows the effectiveness of the approach illustrating the experiments we have performed and, finally, Section 8 draws the conclusions.

2. State of the Art

The technique of mining contextual association rules has been proposed in [10] with the purpose of analyzing frequent user accesses to available services; however, in that case the authors focus on the mining process without using the discovered association rules for data personalization.

The problem of learning user preferences has been recognized as important in many applications. For example, problems related to ordering data on the basis of preference information explicitly provided by the users are discussed in [11,12]. Other notable approaches, instead, extract preferences from logs recording past user

¹ For further details, refer to [7].

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات