



Automatic preference learning on numeric and multi-valued categorical attributes



Lucas Marin*, Antonio Moreno, David Isern

Intelligent Technologies for Advanced Knowledge Acquisition (ITAKA), Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Av. Països Catalans, 26, 43007 Tarragona, Catalonia, Spain

ARTICLE INFO

Article history:

Received 11 December 2012
Received in revised form 11 November 2013
Accepted 13 November 2013
Available online 22 November 2013

Keywords:

Recommender systems
Preference learning
Aggregation operators
Fuzzy sets
Ranking

ABSTRACT

One of the most challenging tasks in the development of recommender systems is the design of techniques that can infer the preferences of users through the observation of their actions. Those preferences are essential to obtain a satisfactory accuracy in the recommendations. Preference learning is especially difficult when attributes of different kinds (numeric or linguistic) intervene in the problem, and even more when they take multiple possible values. This paper presents an approach to learn user preferences over numeric and multi-valued linguistic attributes through the analysis of the user selections. The learning algorithm has been tested with real data on restaurants, showing a very good performance.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Nowadays it is practically unconceivable to select our summer holiday destination or to choose which film to see in the cinema this weekend without consulting specialized sources of information in which, in some way or another, our preferences can be specified to aid the system to recommend us the best choices. That is because we live in an era where there are so many data easily available that it is impossible to manually filter every piece of information and evaluate it accurately. *Recommender Systems (RS)* have been designed to do this time-consuming task for us and, by feeding them with information about our interests, they are capable enough to tell us the best alternatives for us in a personalized way.

The preferences of the user are stored in a structure called *user profile*. In this work, as usual in the literature, it will be considered that each decision alternative is represented through a set of values assigned to a certain set of predetermined attributes or criteria. In these situations, the user profile must somehow represent the preference of the user with respect to each of the possible values of the attributes. With this information, the RS may rate and rank the corpus of available decision alternatives and show it to the user to help him/her to make the final choice. The representation of the preferences, the recommendation process and the automatic

management of the dynamic evolution of the preferences are three of the most challenging issues in the development of this type of systems [24].

Concerning the latter problem (how to learn automatically the preferences of the users), the RS requires some kind of information from the users to guide the learning process. This feedback may be obtained implicitly, explicitly or combining both approaches. *Explicit feedback* forces users to evaluate items, indicating how relevant or interesting they are to them using some numeric or linguistic scale. These systems offer high performance and simplicity [25,26,29,33]. However, explicit feedback has some serious limitations: the user must spend some time and effort, the rating action distracts the attention of the user from his/her standard workflow, and a subjective numerical or linguistic scale is needed to rate each item [12]. Moreover, users are usually reluctant to spend time giving explicit feedback and some studies argue that only 15% of the users would supply it even if they were encouraged to do so [30].

On the other hand, *implicit feedback* is obtained by monitoring the actions of the users and automatically inferring their preferences. The amount of collected data is consequently very large, the computation needed to derive the profile adaptations is extensive, and the confidence in their suitability is likely to be relatively low. This approach has been less explored, although some existing methods have shown promising results (e.g. [3,9,24,32]). This paper discusses an unsupervised way to infer the user interests over numerical and multi-valued categorical attributes, which observes the user interaction and does not require any explicit information from him/her [17].

* Corresponding author.

E-mail addresses: lucas.marin@urv.cat (L. Marin), antonio.moreno@urv.cat (A. Moreno), david.isern@urv.cat (D. Isern).

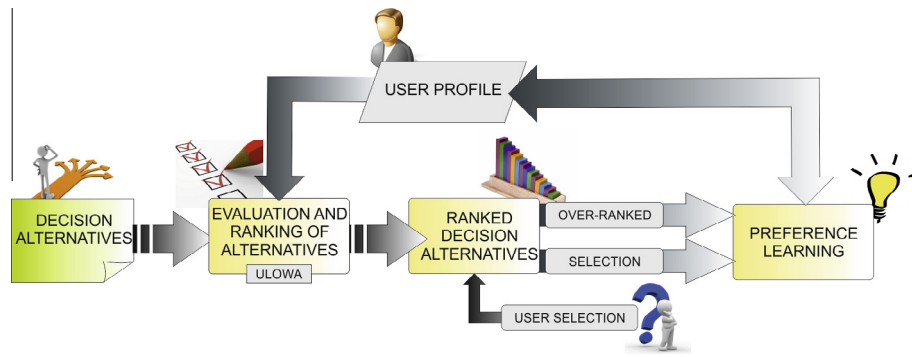


Fig. 1. System architecture.

The basic idea of the preference learning framework proposed in this paper is shown in Fig. 1. There is a set of alternatives that can be recommended to the user as solutions of a decision problem, represented with a set of predetermined criteria. These options are evaluated and ranked according to the current preferences of the user, stored in his/her profile. The user is shown the ranked list of alternatives, and his/her preferred option is selected. The preference learning algorithms analyze the option chosen by the user and the alternatives that were ranked above it in order to decide which changes must be made to the user profile so that it captures better the user preferences and the next recommendation is more accurate. It can be seen that the adaptation process is iterative and, the more choices are made by the user, the more information will be especially suitable in those cases in which the user is confronted with frequent decisional problems (e.g. which news to read every morning, which messages to read from a social network every day). Due to its progressive, continuous and dynamic nature, it will also be applicable in realistic settings in which the preferences of the user are not static but change dynamically over time [22,23].

The rest of the paper is organized as follows. Section 2 includes a brief description of the previous work of the authors in the area of preference learning over single-valued linguistic and numeric attributes, explaining how the interests of users over the values of these kinds of criteria are learned. Section 3 presents a new approach to learn preferences over the values of categorical attributes when the decision alternatives can take multiple linguistic values in a single attribute. Section 4 describes a novel expressive preference function on numeric attributes and how the parameters that define this function can be automatically learned. In Section 5 the case study where our approach has been tested (restaurant recommendation) is explained, describing the data set used and the results obtained. Section 6 gives an overview of related work on multi-criteria preference learning and analyses previous proposals. Finally, Section 7 outlines the main conclusions of the paper and identifies some lines of future research.

2. Background

This section describes the previous work of the authors in the area of preference learning for single-valued numerical and categorical attributes, in order to set the background of the new results introduced in this paper. The first subsection explains how the information about the user preferences on the values of numerical and categorical attributes is represented in the user profile. After that, it is explained how this preference information is used to evaluate each of the alternatives of a decisional problem, so that the user may be given a ranked set of options before making a decision. Finally, the last subsection presents an overview of the basic

techniques used to analyze the choices of the user and adapt dynamically the content of his/her profile. Fig. 1 represents the whole system architecture, where all the steps in the recommendation and learning processes are depicted.

2.1. Representation of preferences

The user profile contains information about the preferences of the user regarding the values that can be taken by numerical and categorical attributes. The preferences over these two kinds of criteria are represented in different ways. In a recent work [23] we proposed to represent the level of interest over each value in the domain of the *categorical* attributes by using a linguistic scale in which the semantics of preference labels is defined using fuzzy sets (see Fig. 2a with an example with the labels “Very Low”, “Low”, “Medium”, “High” and “Very High”). This example shows a set of linguistic labels that are symmetrical and uniformly distributed. However, in some situations it can be more appropriate to represent the preferences using fuzzy sets that are not symmetrical or are not distributed uniformly through the domain, as shown in Fig. 2b with the label set (Very Low, Medium, Almost High, High, Perfect). As it will be explained later, the ULOWA aggregation operator [15] permits to aggregate preference information in both scenarios.

For the case of each numeric attribute, the profile contains a value, v_{pref} , that represents the preferred value of the user in the domain of the attribute. In order to evaluate the degree of preference

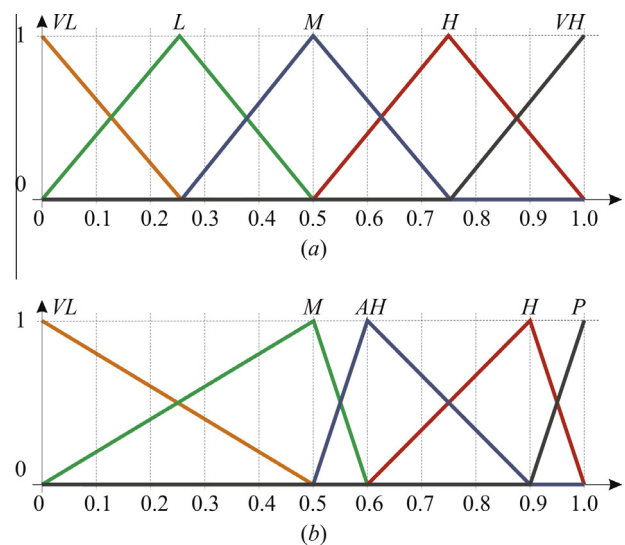


Fig. 2. Examples of (a) balanced, and (b) unbalanced linguistic preference sets.

Download English Version:

<https://daneshyari.com/en/article/6862649>

Download Persian Version:

<https://daneshyari.com/article/6862649>

[Daneshyari.com](https://daneshyari.com)