Knowledge-Based Systems 23 (2010) 32–39

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries

C. Porcel [a,*], E. Herrera-Viedma [b]

[a] University of Jaén, Department of Computer Science, Jaén, Spain
[b] University of Granada, Department of Computer Science and Artificial Intelligence, Granada, Spain

ARTICLE INFO

Article history:
Available online 5 August 2009

Keywords:
Recommender systems
Fuzzy linguistic modeling
University digital libraries
Incomplete fuzzy linguistic preference relation

ABSTRACT

As in the Web, the growing of information is the main problem of the academic digital libraries. Thus, similar tools could be applied in university digital libraries to facilitate the information access by the students and teachers. In [46] we presented a fuzzy linguistic recommender system to advice research resources in university digital libraries. The problem of this system is that the user profiles are provided directly by the own users and the process for acquiring user preferences is quite difficult because it requires too much user effort. In this paper we present a new fuzzy linguistic recommender system that facilitates the acquisition of the user preferences to characterize the user profiles. We allow users to provide their preferences by means of incomplete fuzzy linguistic preference relation. We include tools to manage incomplete information when the users express their preferences, and, in such a way, we show that the acquisition of the user profiles is improved.

1. Introduction

Digital libraries are information collections that have associated services delivered to user communities using a variety of technologies [8,15,48]. Therefore, digital libraries are the logical extensions of physical libraries in the electronic information society. These extensions amplify existing resources and services. As such, digital libraries offer new levels of access to broader audiences of users and new opportunities for the library. In practice, a digital library makes its contents and services remotely accessible through networks such as the Web or limited-access intranets [39,50].

As digital libraries become commonplace and as their contents become more varied, the users expect more sophisticated services from them [8,15,48,50]. A service that is particularly important is the selective dissemination of information or filtering, to help the users to access interesting information for them. Users develop interest profiles and as new materials (books, papers, reports, and so on) are added to the collection, they are compared to the profiles and relevant items are sent to the users [39].

Moreover, digital libraries have been applied in a lot of contexts but in this paper we focus on an academic environment. University Digital Libraries (UDL) provide information resources and services to students, faculty and staff in an environment that supports learning, teaching and research [11].

Recommender systems are becoming popular tools for reducing information overload and to improve the sales in e-commerce web sites [7,9,35,40,49]. The use of this kind of systems allows to recommend resources interesting for the users, at the same time that these resources are inserted into the system. In the UDL framework, recommender systems [7,49] can be used to help users (teachers, students and library staff) to find out and select their information and knowledge sources [43].

Generally, in a recommender system the users' information preferences can be used to define user profiles that are applied as filters to streams of documents [7,47,49]. In [45,46] we developed some recommender systems in an academic context. For instance, in [45] we proposed a fuzzy linguistic recommender system for a technology transfer office which helps researchers and environment companies allowing them to obtain information automatically about research resources (calls or projects) in their interest areas; in [46] we proposed a fuzzy linguistic recommender system to achieve major advances in the activities of UDL, which recommends researchers specialized resources and complementary resources related with their respective research areas. The problem of both recommender systems is that users must directly specify their user profiles by providing their preferences on all topics of interest and it requires too much user effort.

In this paper, we focus on the idea of that a recommender system could be seen as a decision support system (DSS) [37,38,44], where the solution alternatives are the digital resources inserted into the library, and the criteria to satisfy are the user profiles. The proper use of these recommendation systems is essential to

* Corresponding author.
E-mail addresses: cporcel@ujaen.es (C. Porcel), viedma@decsai.ugr.es (E. Herrera-Viedma).

0950-7051/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2009.07.007

provide real personalized services, and it can substantially reduce information overload and increase user satisfaction. Therefore, it has become an important area in information systems and decision support research [37,38,44]. So, the activity of a recommender system can be seen as a group decision making (GDM) problem, so we can adopt the typical representation formats used in GDM, as for example, fuzzy preference relations [19,20,28,32,41]. This representation format presents a high expressivity and some interesting properties that allow us to work easily. However, in real world problems it is common to find situations in which users are not able to provide all the preference values that are required, and then, we have to deal with *incomplete fuzzy preference relations* [1–3,25,26,41].

The aim of this paper is to present a new fuzzy linguistic recommender defined in a UDL framework which overcomes the problem of user profile characterization observed in the recommender systems defined in [45,46]. In order to improve the system performance, we propose an alternative way to obtain accurate and useful knowledge about the user preferences. This new recommender system allows users to provide their preferences by means of incomplete fuzzy linguistic preference relations [1], and in such a way, we facilitate users the expression of their preferences and, consequently, the determination of user profiles process. The recommender system is able to complete the incomplete preference relations using the tools proposed in [1,2,26]. Each user profile is composed of both user preferences on topics of interest and user preferences on collaboration possibilities with other users. Then, the recommender system is able to recommend both research resources and collaboration possibilities to the users of a UDL. As in [45,46] we define this recommender system in a multi-granular fuzzy linguistic context [10,12,22,27,32,42]. In such a way, we incorporate in the recommender system flexible tools to handle the information by allowing to represent the different concepts of the system with different linguistic label sets.

The rest of the paper is set out as follows. Section 2 presents the preliminaries necessary to develop the proposed model. Section 3 presents the new recommender system to the dissemination of knowledge in a UDL. Section 4 reports the system evaluation and the experimental results. Finally, our conclusions are pointed out in Section 5.

2. Preliminaries

2.1. Recommender systems

Recommender systems could be defined as systems that produce individualized recommendations as output, or have the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options [6].

It is a research area that offers tools for discriminating between relevant and irrelevant information by providing personalized assistance for continuous information accesses [43,49]. Automatic filtering services differ from retrieval services [23,24,29–31] in that in filtering the corpus changes continuously, the users have long time information needs (described by means of user profiles) instead of introducing a query into the system, and their objective is to remove irrelevant data from incoming streams of data items [17,39,49]. A result from a recommender system is understood as a recommendation, an option worthy of consideration, while a result from an information retrieval system is interpreted as a match to the user's query [7]. However both systems present some analogies, and in this sense they could be considered a DSS [44]. In both cases, the solution alternatives would be the documents to recommend or retrieve and the criteria to satisfy would be the user profiles and user queries, respectively.

A variety of techniques have been proposed as the basis for recommender systems [7,17,40,49]; all of these techniques have benefits and disadvantages. The use of an hybrid approach is proposed to smooth out the disadvantages of each one of them and to exploit their benefits [5,13,16]. In these kind of systems, the users' information preferences can be used to define user profiles that are applied as filters to streams of documents. The construction of accurate profiles is a key task and the system's success will depend on a large extent on the ability of the learned profiles to represent the user preferences [47].

The recommendation activity is followed by a relevance feedback phase. *Relevance feedback* is a cyclic process whereby the users feed back into the system decisions on the relevance of retrieved documents and the system uses these evaluations to automatically update the user profiles [17,49].

2.2. The 2-tuple fuzzy linguistic approach

The fuzzy linguistic modeling (FLM) is a tool based on the concept of *linguistic variable*[52] which has given very good results for modeling qualitative information in many problems, e.g., in decision making [20], quality evaluation [33,34], models of information retrieval [23,24,29–31], political analysis [4], etc.

The 2-tuple FLM [21] is a continuous model of representation of information that allows to reduce the loss of information typical of other fuzzy linguistic approaches (classical and ordinal [18,52]).

Let $S = \{s_0, \ldots, s_g\}$ be a linguistic term set with odd cardinality, where the mid term represents a indifference value and the rest of the terms are symmetrically related to it. We assume that the semantics of labels is given by means of triangular membership functions and consider all terms distributed on a scale on which a total order is defined, $s_i \leqslant s_j \iff i \leqslant j$. In this fuzzy linguistic context, if a symbolic method [18,20] aggregating linguistic information obtains a value $\beta \in [0, g]$, and $\beta \notin \{0, \ldots, g\}$, then an approximation function is used to express the result in *S*. β is represented by means of 2-tuples (s_i, α_i), $s_i \in S$ and $\alpha_i \in [-.5, .5)$ where s_i represents the linguistic label of the information, and α_i is a numerical value expressing the value of the translation from the original result β to the closest index label, *i*, in the linguistic term set $(s_i \in S)$. This 2-tuple representation model defines a set of transformation functions between numeric values and 2-tuples $\Delta(\beta) = (s_i, \alpha)$ and $\Delta^{-1}(s_i, \alpha) = \beta \in [0, g]$ [21].

The computational model is defined by presenting a negation operator, comparison of 2-tuples and aggregation operators [21]. Using functions Δ and Δ^{-1} that transform without loss of information numerical values into linguistic 2-tuples and viceversa, any of the existing aggregation operators can be easily extended for dealing with linguistic 2-tuples. Some examples are

Definition 1 (*Arithmetic mean*). Let $x = \{(r_1, \alpha_1), \ldots, (r_n, \alpha_n)\}$ be a set of linguistic 2-tuples, the 2-tuple arithmetic mean $\bar{x}^e$ is computed as,

$$\bar{x}^e[(r_1, \alpha_1), \ldots, (r_n, \alpha_n)] = \Delta\left(\sum_{i=1}^{n} \frac{1}{n} \Delta^{-1}(r_i, \alpha_i)\right) = \Delta\left(\frac{1}{n} \sum_{i=1}^{n} \beta_i\right). \quad (1)$$

Definition 2 (*Weighted average operator*). Let $x = \{(r_1, \alpha_1), \ldots, (r_n, \alpha_n)\}$ be a set of linguistic 2-tuples and $W = \{w_1, \ldots, w_n\}$ be their associated weights. The 2-tuple weighted average $\bar{x}^w$ is

$$\bar{x}^w[(r_1, \alpha_1), \ldots, (r_n, \alpha_n)] = \Delta\left(\frac{\sum_{i=1}^{n} \Delta^{-1}(r_i, \alpha_i) \cdot w_i}{\sum_{i=1}^{n} w_i}\right) = \Delta\left(\frac{\sum_{i=1}^{n} \beta_i \cdot w_i}{\sum_{i=1}^{n} w_i}\right). \quad (2)$$