Using linguistic incomplete preference relations to cold start recommendations

Rosa M. Rodríguez, Macarena Espinilla, Pedro J. Sánchez and Luis Martínez-López
Computer Science Department, University of Jaén, Jaén, Spain

Abstract

Purpose – Analyzing current recommender systems, it is observed that the cold start problem is still too far away to be satisfactorily solved. This paper aims to present a hybrid recommender system which uses a knowledge-based recommendation model to provide good cold start recommendations.

Design/methodology/approach – Hybridizing a collaborative system and a knowledge-based system, which uses incomplete preference relations means that the cold start problem is solved. The management of customers' preferences, necessities and perceptions implies uncertainty. To manage such an uncertainty, this information has been modeled by means of the fuzzy linguistic approach.

Findings – The use of linguistic information provides flexibility, usability and facilitates the management of uncertainty in the computation of recommendations, and the use of incomplete preference relations in knowledge-based recommender systems improves the performance in those situations when collaborative models do not work properly.

Research limitations/implications – Collaborative recommender systems have been successfully applied in many situations, but when the information is scarce such systems do not provide good recommendations.

Practical implications – A linguistic hybrid recommendation model to solve the cold start problem and provide good recommendations in any situation is presented and then applied to a recommender system for restaurants.

Originality/value – Current recommender systems have limitations in providing successful recommendations mainly related to information scarcity, such as the cold start. The use of incomplete preference relations can improve these limitations, providing successful results in such situations.

Keywords Uncertainty management, Catering industry, Internet

Paper type Research paper

1. Introduction

The viral growth of Internet has produced huge amounts of information that the users cannot manage directly unless they use different tools such as searchers, meta-searchers, etc. Similarly, in the e-commerce arena customers face to huge amounts of information about items that are hard to check in an affordable time in order to buy the most suitable item/s. Therefore, in order to support customers in their buying processes different tools have arisen, the most successful one in such a duty it has been the recommender systems (Martínez *et al.*, 2008b; Schafer *et al.*, 2001; Uchyigit and Ma, 2008); these try to lead customers to the most suitable items according to their requirements, needs, tastes, preferences and so on.

This paper has been partially supported by the research projects TIN2009-08285, P08-TIC-3548 and Feder Funds.

Internet Research
Vol. 20 No. 3, 2010
pp. 296-315

1066-2243
DOI 10.1108/10662241011050722

The general scheme of a recommender system (see Figure 1) consists of generating an item database and a customers database that store the items and customers' profiles respectively and perform a filtering process by matching a customer's profile and the items profiles. This allows finding out those items, which better meet customers' needs expressed by their profiles.

The scheme showed in Figure 1 has been implemented with different techniques producing different types of recommender systems:

(1) *Content-based recommender systems* (Horvath, 2009; Martínez *et al.*, 2007): They learn a customer profile based on features of items experienced by the customer and it uses this profile to find out similar items.

(2) *Collaborative filtering recommender systems* (Goldberg *et al.*, 1992; Takacs *et al.*, 2009): They use customers' ratings to filter and recommend items to a specific user based on the similarity of the target customer and the others.

(3) *Knowledge-based recommender systems* (Burke, 2000; Zhen *et al.*, 2010): These systems use the knowledge about customers' necessities and how an item matches these necessities to infer recommendations that fulfil user's expectations.

(4) *Demographic recommender systems* (Krulwich, 1997): They categorize customers into demographic groups and make recommendations based such groups.

(5) *Utility-based recommender systems* (Guttman, 1998): They make recommendations by computing the utility of each object for the user.

(6) *Hybrid recommender systems* (Albadvi and Shahbazi, 2009; Burke, 2002): Their aim is to sort out drawbacks presented in the aforementioned recommender systems. To accomplish this aim, these systems combine different techniques to improve the accuracy of the recommendations.

The classical and most spread recommender systems are collaborative and content based systems both have provided good results in different areas as tourism (Sebastia *et al.*, 2009), e-learning (Romero *et al.*, 2009), academic orientation (Castellano and Martínez, 2009), etc. Notwithstanding, they require a significant amount of information about customers to infer accurate recommendations (Albadvi and Shahbazi, 2009; Goldberg *et al.*, 1992; Takacs *et al.*, 2009). To obtain such information is usually carried

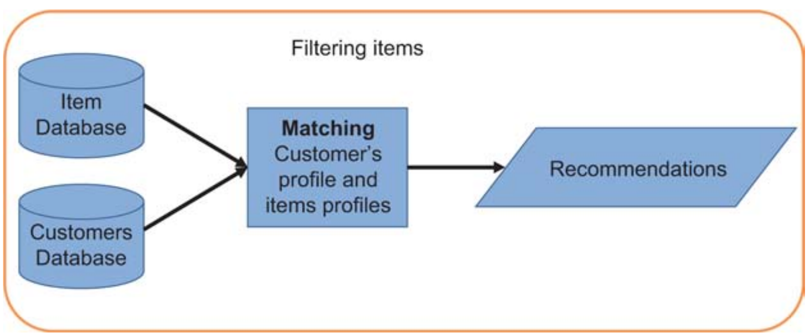

Figure 1.
Recommendation scheme

out a gathering process in which is expected to obtain much information, the more the better, about customers' preferences, tastes, necessities, etc. This process is time consuming, tedious and somehow problematic; because customers do not feel comfortable providing this information because of privacy issues either because they cannot afford to spend so much time in this process. But sometimes it is due to the difficulty to express appropriately such information, because customer's profile is mainly based on preferences, tastes and perceptions. This type of information is qualitative rather than quantitative and implies uncertainty. However, this information, is usually required, by precise numbers. In the literature, the use of the Fuzzy Linguistic Approach (Zadeh, 1975) has provided successful results to manage qualitative information. Due to this fact, we shall consider the use of the Fuzzy Linguistic Approach to model information about customers.

Consequently, in those situations with lack of customers' information, classical recommender systems cannot provide successful recommendations provoking that customers distrust about the reliability of the recommendations.

The previous problem is so-called cold-start (Ahn, 2008; Victor *et al.*, 2008) and it is due to the lack of information about customers in order to infer accurate recommendations. Different proposals have been developed to solve it (Cane *et al.*, 2007; Victor *et al.*, 2008). In this paper we focus on how to solve the cold start problem in collaborative recommender systems. To do so, we propose hybridizing by commutation a collaborative model and a knowledge-based one. The latter will provide the cold start recommendations by using incomplete linguistic preference relations. This proposal will facilitate the gathering process because the customer has to provide just a few data, shortening the time of this process and the system builds a whole customer's profile from such few data that will be big enough to compute accurate recommendations.

This paper is structures as: section 2 reviews in short some general concepts about collaborative recommender systems paying special attention to the cold start problem. Section 3 introduces a necessary linguistic background about linguistic and preference relation concepts that will be used in our proposal. Section 4 presents the knowledge-based model that uses incomplete linguistic preference relations for cold start recommendations and that will be hybridized with a collaborative one. In section 5, an application of this model to a restaurant recommender system is introduced. Finally section 6 shows some implications and conclusions of our proposal.

2. Collaborative recommender systems

Collaborative recommender systems (CRS) gather human judgments (known as ratings) for items in a given domain and group customers with similar needs, preferences, tastes, etc. (Herlocker *et al.*, 1999). In a CRS, customers share each other their judgments and opinions about items that they have already experienced, such that, the system can support them in order to make right and better decisions about the items involved in the system. The CRS provide useful customized recommendations of interesting items by using collaborative filtering algorithms. These algorithms try to predict user's satisfaction regarding an unrated item based on similar users to the target user.

Judgments and opinions used by the CRS are classified into two main categories:

(1) *Explicit data:* which is directly provided by the users, according to their own experience, and knowledge.

(2) *Implicit data:* they are inferred through knowledge discovery processes like data-mining, navigation monitoring, etc. (Herlocker *et al.*, 2004).

Most of CRS use explicit data related to customers' perceptions and preferences that implies uncertainty, though it has been fairly usual the use of precise scales to gather such information, the use of linguistic information to model such an information seems more suitable and several proposals have been developed (Martínez *et al.*, 2007; Porcel *et al.*, 2009).

There exist different collaborative approaches (Adomavicius and Tuzhilin, 2005):

(1) Memory-based.

(2) Model-based.

According to Figure 2, all of them fulfil three general tasks to elaborate the recommendations demanded by users:

(1) *Analyzing and selecting data sets*: A dataset must be collected and optimized for the system (Herlocker *et al.*, 2004).

(2) *Grouping users*: In order to elaborate recommendations, collaborative algorithms select a group of users with similar tastes and preferences.

(3) *Generating predictions*: Once users have been grouped by interest (similarity), the system uses them to compute predictions for the target customer by using different aggregation methods (Herlocker *et al.*, 1999; Adomavicius and Tuzhilin, 2005).

2.1 Advantages and weaknesses

Collaborative filtering methods provide several advantages regarding other techniques used in recommender systems (Herlocker *et al.*, 1999):

- Support for filtering items whose content is not easily analyzed automatically.
- Ability to filter items based on quality and taste, not only on its features. Do not need knowledge domain, i.e. no information or knowledge about the products is needed.
- Ability to provide serendipitous recommendations. Other systems never recommend products which are outside the box, i.e. recommended products are not very different to the ones positively rated by the customer.

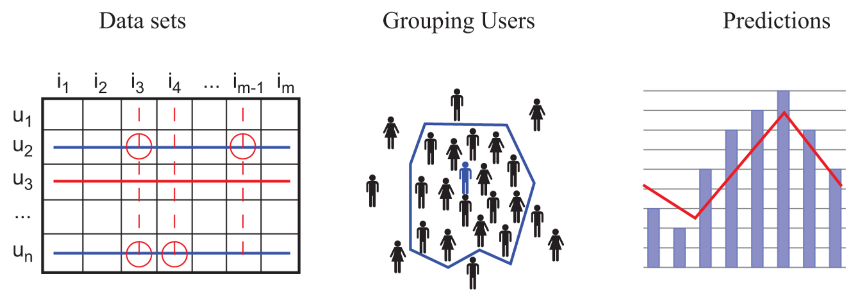

Figure 2. Collaborative recommendation scheme

- Adaptivity, its quality is improved along the time. When the number of customers and rates increases these systems work better, the more density of users and rates the better performance.

Despite the general good performance of these systems, they present some weaknesses and limitations:

- *The "grey sheep" problem:* This system does not work properly with "grey sheep" users, which are in the frontier of two groups of users.
- *Historical data set:* Being an adaptive system can be an advantage but can be also a disadvantage when the system is starting and the historical data set is small.
- *The cold-start problem*: This problem is presented with both users and products. When a new user access to the system, it has not any information about him/her. Therefore, the system cannot compare him/her with the users of the database and cannot provide recommendations. When a new item is added and it has not been assessed by any user yet, it cannot be recommended. This is the problem we focus on this paper, so it will be further detailed in the following.

2.2 Cold start problem

Probably, the main disadvantage of collaborative recommender systems is the cold-start problem, that is related to the situation when a new or recent customer logins to the system and has not provided enough ratings yet. Hence the collaborative system cannot compute a recommendation under such an situation (Massa and Bhattacharje, 2004). This problem is also applied to new and obscure items and to customers with eclectic tastes (Trujillo *et al.*, 2007). Due to the fact that, a CRS works computing similarities and correlations between the target customer and other users, when there not exist any initial information for the target one, or it is scarce, the similarity between the concepts involved is low or null, hence no recommendations are produced. This problem decreases the overall efficiency of the CRS and, another issue much more important is that the user confidence on the recommender system decreases too.

In the literature different proposals have been proposed for solving this problem. Trujillo *et al.* (2007), proposed a two-phased model:

(1) *An off-line clustering phase*: It deals with demographics features, such as age, university relation, higher academic degree, etc. and psychographic features, such as interest areas. The system calculates the similarity between users and classifies them into clusters.

(2) *An online probabilistic phase*: It calculates the probability of a user, u, is interested in a product, p, according to the ratings provided by u and the ratings received by p.

The second phase is purely collaborative, but the first one works rightly in cold-start situations:

- In (Wang and Kong, 2007) was proposed a semantic-enhanced collaborative filter recommendation method, in which the recommendation is produced by using semantic information of the category features of items as well as the user's demographical data.

- Ahn (2008), proposed a new similarity measure to alleviate the new user cold-starting problem by considering three factors: proximity, impact and popularity.
- Cane *et al.* (2007), proposed a model with association rules for cold-start recommendations. It uses associations between a given item's attributes and other domain items.
- Li *et al.* (2007), describe a collaborative music recommender system (CMRS) based on an item-based probabilistic model. It has been extended for improving recommendation performance by utilizing audio features that help alleviate the cold start problem for new items.
- In (Diez and Villegas, 2007) the authors worked in a mixed system that takes advantage of the knowledge defined in an ontology used in the items' data acquiring process.
- Kim *et al.* (2007), proposed a new method of building a model, namely a user-item error matrix, for CRS. The main advantage of such an approach is that it supports incremental updating of the model by using explicit user feedback.
- Victor *et al.* (2008), proposed a method that connects new users to an underlying trust network among the users of the recommender system alleviating the cold start problem.

Obviously, no recommender system can work without some initial information but the quality and efficiency of the system depends on the ability predicting successful recommendations with the minimum amount of information about users and items. From the previous proposals, we can observe that models based on demographic and psychographic features need personal information about users like academic degree, age, interest areas, etc. But users may be reluctant to provide this type of information and, so, they will reject this kind of systems. Some solutions require knowledge about the items, or content-based-information, for example, a movie recommender system needs to know attributes like actors, the genre, etc. This kind of knowledge is not always available or is scarce. Other proposed solutions are only partial solutions because improve the recommendations when the data about the user is small but do not work when this set is empty (new user).

Another promising method to solve the cold start problem is the hybridization with a knowledge-based recommender system (Burke, 2000; Burke *et al.*, 1996; Martínez *et al.*, 2008a). This is the alternative that we have chosen for our proposal and it will be detailed in section 4.

3. Linguistic background

Owing to the fact that our proposal models customers' data by using linguistic information and deals with incomplete linguistic preference relations, this section reviews some concepts about linguistic information and preference relations that we shall use in our proposal.

3.1 Fuzzy linguistic approach

Information in a quantitative setting is usually expressed by means of numerical values. However, many aspects in the real world (perceptions, preferences, tastes, etc.)

cannot be assessed in a quantitative form, but rather in a qualitative one, i.e. with vague or imprecise knowledge. In such a case, a better approach may be the use of linguistic assessments instead of numerical ones. The fuzzy linguistic approach represents qualitative aspects as linguistic values by means of linguistic variables (Zadeh, 1975). This approach has been successfully applied in different areas such as information retrieval (Herrera-Viedma *et al.*, 2009), marketing (Yager *et al.*, 1994), education (Lalla *et al.*, 2004), decision-making (Chen and Ben-Arieh, 2006; Herrera *et al.*, 2009), sensory evaluation (Martínez, 2007), consensus (Mata *et al.*, 2009), recommender systems (Martínez *et al.*, 2008a; Yager, 2003), etc.

The use of the fuzzy linguistic approach implies to choose the appropriate linguistic descriptors for the term set and their semantics. The universe of the discourse defined for the term set is problem specific, and linguistic term sets are usually defined in the interval [0, 1].

One possibility of generating the linguistic term set is to directly supply the term set by considering all terms distributed on a scale on which a total order is defined (Herrera *et al.*, 2009), e.g. a set of five terms S, could be given as:

$$S = \{s_0 : \text{Poor}; s_1 : \text{Low}; s_2 : \text{Average}; s_3 : \text{High}; s_4 : \text{Good}\}$$

In these cases, it is usually required that there exist:

- A negation operator Neg $(s_i) = s_j$ such that $j = g$-i ($g + 1$ is the cardinality of S).
- A min and a max operator in the linguistic term set: $s_i < = s_j \leftrightarrow i < = j$.

The semantics of the terms are given by fuzzy numbers defined in the interval of [0,1], which are described by membership functions. One way to characterize a fuzzy number is to use a representation based on the parameters of its membership function. This parametric representation is achieved by the four-tuple (a, b, d, c), where b and d indicate the interval in which the membership value is 1, with a and c indicating the left and right limits of the definition domain of the trapezoidal membership function. For example, we may assign the following semantics to the set of five terms, see Figure 3:

3.2 The two-tuple fuzzy linguistic representation model

The two-tuple fuzzy linguistic representation model (Herrera and Martínez, 2000) is based on the symbolic method and takes as the base of its representation the concept of Symbolic Translation. This model has initially overcome the drawback of the loss of

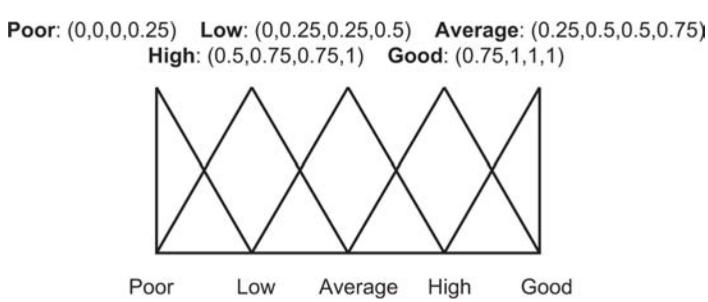

Figure 3.
A linguistic term set of five terms and its semantics

information presented by the classical linguistic computational models (Degani and Bortolan, 1988; Herrera *et al.*, 1995).

Definition 1. The Symbolic Translation of a linguistic term $s_i \in S = \{s_0, \ldots, s_g\}$ is a numerical value assessed in $[-0.5, 0.5)$ that supports the "difference of information" between an amount of information $\beta \in [0, g]$ and the closest value in $\{0, \ldots, g\}$ that indicates the index of the closest linguistic term in S (s_i), being $[0,g]$ the interval of granularity of S.

This model represents the linguistic information by means of two-tuples $(s_i, \alpha_i), s_i \in S$ and $\alpha_i \in [-0.5, 0.5)$. Furthermore, it defines a set of functions between the linguistic 2-tuple and numerical values.

Definition 2. Let $S = \{s_0, \ldots, s_g\}$ be a linguistic term set and $\beta \in [0, g]$ a value supporting the result of a symbolic aggregation operation. Then the 2-tuple that expresses the equivalent information to β is obtained using the following function:

$$\Delta : [0, g] \rightarrow S \times [-0.5, 0.5)$$

$$\Delta(\beta) = (s_i, \alpha), \text{with} \begin{cases} s_i & i = \text{round}(\beta) \\ \alpha = \beta - i & \alpha \in [-0.5, 0, 5) \end{cases}$$

where s_i has the closest index label to "β" and "α" is the symbolic translation.

We note that Δ is a one to one mapping and $\Delta^{-1} : S \times [-0.5, 0.5) \rightarrow [0, g]$ is defined by $\Delta^{-1}(s_i, \alpha) = i + \alpha$.In this way, the 2-tuples of S will be identified with the numerical values in the interval $[0, g]$.

This model has a computational technique with aggregation, comparison and negation operators (Herrera and Martínez, 2000).

3.3 Linguistic preference relations

We have aforementioned in section 2.2 the importance of the structure in which the experts express their preferences. There exist different structures to represent preferences about a set of items, $X = \{x_1, \ldots, x_n\}$, such as preference orderings (Chiclana *et al.*, 1998), utility vectors (Tanino, 1990) and preference relations (Tanino, 1984). In our proposal, we shall use preference relations that represent the information by means of a preference matrix $P \subseteq X \times X, P = (p_{ij})$,where p_{ij} is the intensity of preference of item x_i regarding item x_j:

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix}$$

Definition 3. A linguistic preference relation, P, on a set of alternatives, X, is characterized by a function $\mu_p : X \times X \rightarrow S$,where S is the linguistic term set to express the preference degrees.

As we have aforementioned our proposal deals with linguistic preference relations in order to facilitate customers the provision of more detailed information about his/her preferences. Initially this representation could be time-consuming and lead to inconsistencies. But to overcome these drawbacks we use incomplete preference *relations* (Alonso *et al.*, 2008; Herrera-Viedma *et al.*, 2004; Xu, 2006) that are suitable in situations where the users are under time pressure, there is a lack of information or some alternatives could be unknown. In these cases some preferences, p_{ij}, could be missed (Alonso *et al.*, 2008; Martínez *et al.*, 2008b). These preference relations improve the time cost and the consistency of the information gathering process. And they will be completed keeping consistency property (Martínez *et al.*, 2008b).

Definition 4. A function $f: X \rightarrow Y$,is partial when not every element in X necessarily maps to an element in Y. Otherwise is a total function.

Definition 5. A preference relation P on a set of alternatives X with a partial function is an incomplete preference relation.

Definition 6. (Alonso *et al.*, 2008). A linguistic preference relation will be considered additive consistent if for every three alternatives $x_i, x_j$and x_k it holds the following condition:

$$p_{ik} = \Delta(\Delta^{-1}(p_{ij}) + \Delta^{-1}(p_{jk}) - g/2) \forall i,j,k \in \{1, \ldots n\}$$

4. A knowledge-based recommender system based on incomplete linguistic preference relations for cold start recommendations

Our proposal to solve the cold start problem consists of hybridizing a CRS with a knowledge-based system. To achieve our aim, first we review a general scheme of the working of this type of system. Second, we will introduce an enhanced knowledge based system dealing with incomplete linguistic preferences that uses a filling algorithm based on the additive transitivity (definition 6). And finally, we will show the whole hybrid scheme proposed to solve cold start in CRS.

4.1 Knowledge-based recommender systems

Knowledge based recommender systems (Burke, 2000) use case based reasoning (Kolodner, 1993) to make recommendations, i.e. they starts with an example that the customer points out, according to his/her tastes, preferences or necessities. Using this information the system infers a customer's profile that will be utilized to find the items that best match it in the search space. For reaching its purpose, this type of systems matches the customer's profile and the possible recommended items. The user knowledge can be expressed in any knowledge structure that allows building a user profile. The simplest case could be that the user chooses among all the available products, one of them that acts as an example of his/her necessities or tastes. These systems manage three types of knowledge:

(1) *Catalog knowledge*: knowledge about the products being recommended.

(2) *Functional knowledge*: how the features of the products meet the user's necessities.

(3) *User knowledge*: it is the knowledge that the system has gathered about the user. It could be the necessities that the user has stated as well as all the knowledge that can be obtained by other means (for example, using demographic information).

Knowledge based recommender systems are especially suitable for casual searching when customer's information does not exist or is scarce. We have seen that other systems (collaborative, content based, . . .) need a start-up period until the system gathers enough information about customers. The quality of the recommendations during that period is quite low and sometimes no recommendations can be issued. Knowledge based systems do not suffer this drawback because they do not need such kind of historical information. They work very well with just a small amount of knowledge about the user. So, we can enumerate the following advantages of this kind of system:

- They do not suffer cold start problem.
- The grey sheep problem does not appear in these systems.
- They do not depend on large historical data set.

On the other hand, these systems present two disadvantages related to the gathering of user knowledge:

(1) When the amount of products is very large, the process of providing an example to express the user necessities may be a hard task.

(2) It is possible that the user does not find an example that fits exactly his/her necessities. So, the system recommends him/her products that perhaps do not satisfy the user.

The main disadvantage of knowledge based recommender system is that they still require explicit knowledge acquisition. Many times this knowledge is not easy to obtain or cannot be gathered with automatic tools.

A way to improve the information gathered and overcome the previous drawback is to infer the recommendation from several examples instead of a unique one. For example, in (Martínez *et al.*, 2008a; Porcel *et al.*, 2009) the recommender system makes the recommendations from a small set of items related to one another by means of a preference relation. In our proposal we shall use a similar gathering process based on incomplete preference relations.

4.2 Knowledge-based recommender systems with incomplete linguistic preference relations

The general scheme of the knowledge-based recommendation model that deals with incomplete linguistic preference relations is composed of three different phases (see Figure 4):

(1) *Acquiring customer's preference information:*
   - setting the favorite examples; and
   - filling the preference relation up.

(2) *Building the user profile:*
   - building partial customer profiles; and
   - obtaining the customer's profile.

(3) *Recommendation.*

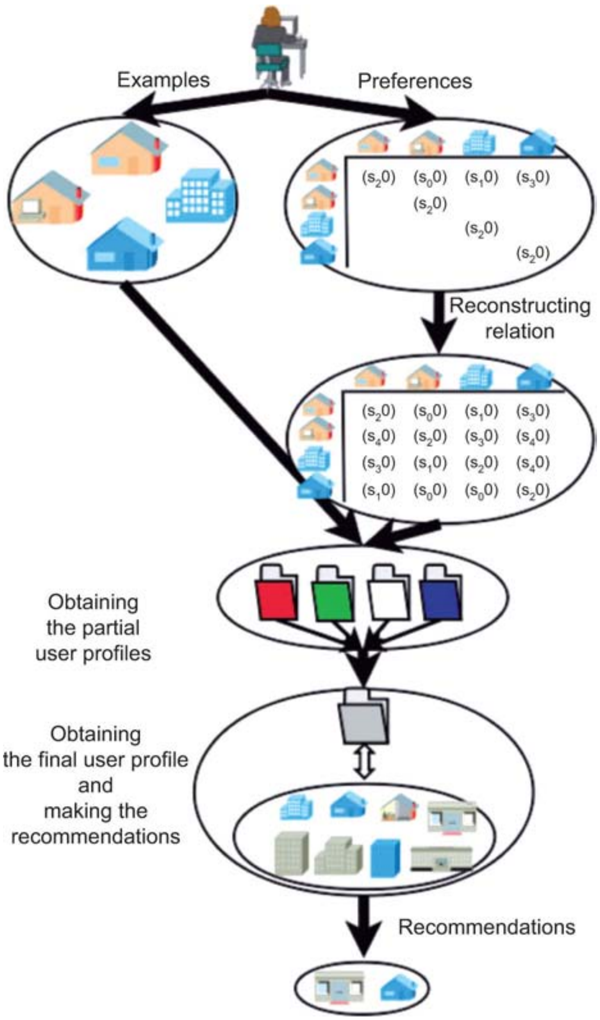

Figure 4.
A knowledge-based model with incomplete linguistic preference relations

Following, we shall explain in further detail each phase of the recommender model, paying more attention to phase 1b) that makes possible the improvement of the cold start problem.

4.2.1 Acquiring customer's preference information. This phase aims to obtain information about the customer's preferences. First, the user must choose a small set of items (from three to five) as examples of his/her preferences, tastes or necessities. Due to the fact that the item database, $X = \{x_1, \ldots, x_m\}$, can be huge in order to make easier this choice task, the system offers a subset of items, $X_r = \{x_1^r, \ldots, x_{m'}^r\} m' < m$, that should be big enough to include items that represent any kind of user's necessities, but not too big to avoid that the task of choosing the examples of his/her necessities would be too tedious and time-consuming. Moreover, these items ought to be well known.

Remark 1. There is not any correlation between well known and preferred. For instance, in recommendation of hotels the Sheraton Hotels are well known but it might be that the user does not like.

This process consists of two phases:

A. Setting the favorite examples

The set, X_r, is shown to the user, so he/she can choose a subset $X_u = \{x_1^u, ..., x_n^u\}$, with four or five elements, according to his/her needs.

Afterwards, the system inquires the users to express their preferences among the elements of X_u, by means of an incomplete linguistic preference relation, *P*, assessed in a linguistic term set$S = \{s_0, ..., s_g\}$.

Owing to the fact that one goal of our proposal is to decrease the time of the information gathering process. The system will require to the user only a row of the preference relation.

B. Completing the preference relation

The incomplete linguistic preference relation, *P,* should be then completed, *P*'. There exist different algorithms based on the additive transitivity to carry out this task (Alonso *et al.*, 2008; Herrera-Viedma *et al.*, 2004, Xu, 2006), but we propose an extension of the algorithm presented in (Alonso *et al.*, 2008).

In the algorithm presented in (Alonso *et al.*, 2008) can occur that an unknown, p_{ij},can be estimated with different values. In such cases Alonso *et al.* apply an average to obtain a single value. However, with the aim of giving more relevance to the preferences directly provided by the customer, we modify the previous algorithm such that we find a situation in which an unknown, p_{ij},that can be estimated with several values, choosing that one closest to the indifference. To do so, we define a function that computes between two values, the closest to the indifference, taking into account that giving a linguistic term set $S = \{s_0, ..., s_g\}$the term that means indifference is, $s_{g/2}$.

Definition 7. Let $p, q \in [0, g]$ be two values obtained by using Δ^{-1} with two linguistic terms in the incomplete preference relation. The function that computes the closest value to the indifference is defined as:

$$cti(p,q) = \begin{cases} p, & if \quad 1 - \frac{|p-g/2|}{g} > 1 - \frac{|q-g/2|}{g} \\ \\ q, & \text{otherwise} \end{cases}$$

Therefore, the proposed additive transitivity algorithm to fill up the incomplete preference relation obtaining a consistent preference relation is the following:

(1) Initialization
 $P' = \Delta^{-1}(P)$
 $EMV_0 = ?$
 $h - 1$

(2) While $EMV_h \neq ?\{$
(3) For every $(i,k) \in EMV_h\{$
(4) $K = ?$
(5) $H^1_{ik} = \{j \neq i,k|(i,j),(j,k) \in KV_h\}$; if $(H^1_{ik} \neq\ )$ then $K = K \cup \{1\}$
(6) $H^2_{ik} = \{j \neq i,k|(i,k),(j,i) \in KV_h\}$; if $(H^2_{ik} \neq\ )$ then $K = K \cup \{2\}$
(7) $H^3_{ik} = \{j \neq i,k|(i,j),(k,j) \in KV_h\}$; if $(H^3_{ik} \neq\ )$ then $K = K \cup \{3\}$
(8) Calculate $p'_{ik} = cti(cp^{jl}_{ik}, \forall l \in K, \forall j \in H^l_{ik})$
(9) $h++$
(10) $\}\}$
(11) $P'' = \Delta(P')$

where

KV_h = the known values for the iteration h
UV_h = the unknown values for the iteration h
EMV_h = the subset of unknown values that can be computed in the iteration h
$EMV_h = \{(i,k) \in UV_h\} | \exists j \in H^1_{ik} \cup H^2_{ik} \cup H^3_{ik}$
$cp^{j1}_{ik} = \min(\max((p'_{ij} + p'_{jk} - g/2), 0), g)$
$cp^{j2}_{ik} = \min(\max((p'_{jk} + p'_{ji} - g/2), 0), g)$
$cp^{j3}_{ik} = \min(\max((p'_{ij} + p'_{kj} - g/2), 0), g)$

The proposed process of acquiring customer's preferences provides three main benefits in order to gather the user's information:

(1) The task is easier and quicker for the user: he/she provides the minimum information necessary.
(2) The proposed algorithm completes an incomplete preference relation with an only row of known values and avoids inconsistencies.
(3) Since the system uses several examples, the recommendations are less dependent on the adequacy of the examples than in Classical Knowledge Based Recommender System. Classical Knowledge Based Recommender Systems compute the recommendations by using just one example. If such an example is not completely adequate, the recommendations will be fairly inaccurate. When recommendations are led by several examples, it will be more likely to obtain better recommendations whenever some of the examples are adequate.

4.2.2 Building the user profile. Now the system has a complete linguistic preference relation, *P''*, that contains much more information about the customer than the initially provided by him/her. The following phase in the knowledge based recommender system will be to compute a customer's profile in order to compare his/her necessities with the items stored in the database. The system computes the customer's profile by using the complete preference relation, *P''*, and the descriptions of the items in X_u, considered in such preference relation (Martínez *et al.*, 2008b). The user profile is computed in two steps:

(1) *Building partial customer's profiles*: The system exploits the preference relation to obtain partial user profiles. For each column, *j*, of the preference relation, *P''*, the system obtains a partial user profile that represents the user's preferences related to the example *j*.
(2) *Obtaining the customer's profile*: from the previous profiles, the final one is computed by aggregating all the partial profiles.

The customer's profile obtained will be used to compute the recommendations for him/her.

4.2.3 Recommendation. This is the most important phase of the Recommender System. Once the customer profile has been computed, the system should recommend the closest items to the customer's necessities.

The process of computing the items that better matches the customer's profile consists of comparing each item description with the user profile. To do so, the system will compute the similarity between the items and the user's profile different measures and proposals have been proposed to carry out this process (Martínez *et al.*, 2008a, b).

Finally, the system will recommend to the customer the set of more similar items to his/her necessities.

4.3 Hybridizing collaborative and knowledge-based recommender systems for cold start recommendations

Once we have reviewed CRS and presented a knowledge-based recommender model dealing with incomplete linguistic preference relations. Here we are going to present a recommender model that hybridizes both of them in order to provide cold start recommendations.

Our proposal consists of hybridizing by commutation (Burke, 2002) a collaborative and the knowledge-based model dealing with linguistic incomplete preference relations proposed previously. Such that, the customers obtain their recommendations through the collaborative model whenever would be possible, but in those cases where the collaborative model does not obtain good recommendations because of the information scarcity the customer will obtain the recommendations by using the knowledge-based model.

The scheme of this hybrid system can be seen graphically in Figure 5. In the following section we show the implementation of a restaurant recommender system based on such a scheme.

5. REJA: a restaurant recommender system

This section introduces the recommender system for restaurants of Jaén (a small city in the south of Spain), called REJA, whose URL is: http://sinbad2.ujaen.es/~reja/joomla/index.php (see Figure 6). The main aim of REJA is to provide successful recommendations to users about the existing restaurants in the city of Jaén.

When we started this system, we noticed that many potential users would be tourists that come over Jaén for one or two days. So, it would not be useful or possible to obtain information about their preferences regarding restaurants of Jaén. Therefore, the necessity to provide good recommendations in those situations was our challenge.

Consequently, once we have developed a method to provide cold start recommendations, we implemented in REJA the hybrid scheme presented in the

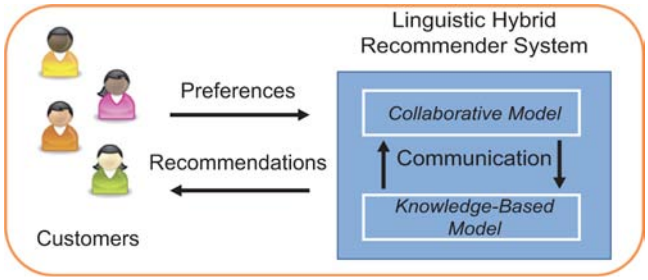

Figure 5.
Scheme for a linguistic hybrid recommender system

Figure 6.
Restaurant recommender system (REJA)

Figure 5 such that, hybridizes by commutation a CRS and a knowledge based system as we proposed in the previous section. This hybridizing scheme makes of REJA a system that takes the advantages of each model in order to overcome their own drawbacks.

Following, we are going to present the working of each one of the different recommendation models provided by REJA.

5.1 Collaborative recommendations in REJA

The collaborative system used by REJA has been implemented by using the collaborative filtering engine CoFE (http://eecs.oregonstate.edu/iis/CoFE/) that uses a database with all the restaurants in the province of Jaén and the users registered in the system.

In order to obtain a recommendation of the collaborative system, the user must be registered, then he/she should login and provide enough ratings about the restaurants (this system requires at least 20 ratings). The system obtains explicit information from the user who may rate the restaurants in which he/she has already been. This information is used to build and modify the customer's profile and to compute suitable recommendations for him/her. The system utilizes the profile of the target customer to recommend him/her ten restaurants (see Figure 7) that have not been rated yet by him/her, but have got good ratings by other users of his/her neighbourhood.

Such as it was aforementioned the collaborative model suffers the cold start problem that affects new users. In REJA we have solved this problem by hybridizing as showed in Figure 5 the collaborative model with the knowledge-based one dealing with incomplete preference relations.

5.2 Cold start recommendations in REJA. Knowledge-based model

The aim of the knowledge-based model is to provide cold start recommendations when the collaborative system does not have enough information about customers. To do so, REJA follows the scheme presented in section 4 and showed in Figure 4:

(1) *User's profile:*
  - *to gather user's preference information*: the system gathers the user's preferences and needs by means of an incomplete preference relation to minimize this process, as follows (see Figure 8).
  - *to build a user's profile:* with these three pieces of data, the system fills up the incomplete preference relation with the algorithm presented in section 4 and computes the customer's profile that will be used to obtain the recommendations.

(2) *Product filtering*: the system filters the restaurants according to the customer's profile in order to take into account only those restaurants that fulfil user's needs. In our case, the filter uses as filtering attributes the type of food, the price, and so on.

(3) *Recommendations*: from those restaurants that fulfil pretty much the user's needs, *REJA* recommends those that better satisfy these requirements (see Figure 7).

6. Concluding remarks

This paper has focused on the cold start problem of collaborative recommender systems that occurs when the system has not enough information about customers and

Figure 7.
REJA: collaborative recommendations

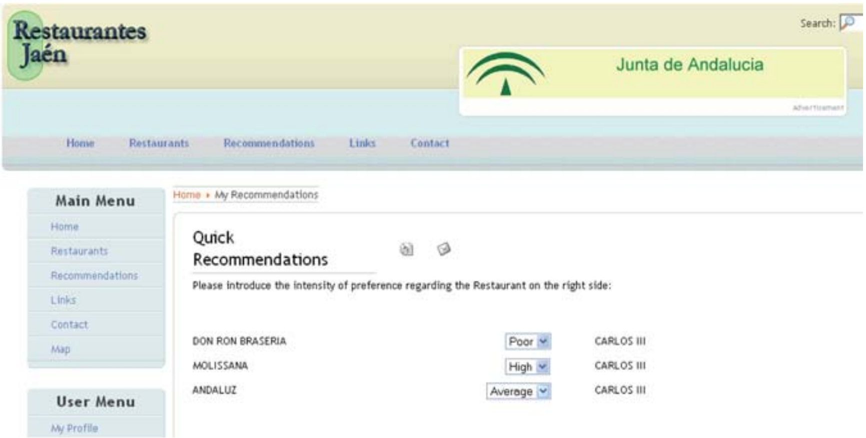

Figure 8.
REJA: gathering an incomplete linguistic preference relation

items as to compute a recommendation. To solve this problem we have proposed a hybrid system composed by a collaborative and a knowledge-based model, such that the knowledge-based model uses incomplete linguistic preference relations in order to facilitate the gathering process of the recommender system when the information is scarce. We have developed an algorithm to fill up incomplete preference relations by using the additive transitivity that keeps more relevance to the data directly provided by the customers than to the estimated data.

Such a system can be applied any recommender system that have the cold start problem in order to improve its recommendations. This involves an increase of the users' trust on the system, a greater use of it and users do not need to provide too rating.

Finally, we have presented a restaurant recommender system, so called REJA, which implements and applies the models proposed to a real world application.

References

Adomavicius, G. and Tuzhilin, A. (2005), "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17 No. 6, pp. 734-49.

Ahn, H.J. (2008), "A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem", *Information Sciences*, Vol. 178 No. 1, pp. 37-51.

Albadvi, A. and Shahbazi, M. (2009), "A hybrid recommendation technique based on product category attributes", *Expert Systems with Applications*, Vol. 36 No. 9, pp. 11480-8.

Alonso, S., Chiclana, F., Herrera, F., Herrera-Viedma, E., Alcalá, J. and Porcel, C. (2008), "A consistency-based procedure to estimate missing pairwise preference values", *International Journal of Intelligent Systems*, Vol. 23 No. 2, pp. 155-75.

Burke, R. (2000), "Knowledge-based recommender systems", *Encyclopedia of Library and Information Systems*, Vol. 69.

Burke, R. (2002), "Hybrid recommender systems: survey and experiments", *User Modeling and User-Adapted Interaction*, Vol. 12 No. 4, pp. 331-70.

Burke, R.D., Hammond, K.J. and Young, B.C. (1996), "Knowledge-based navigation of complex information spaces", *AAAI/IAAI*, Vol. 1, pp. 462-8.

Cane, W.-K. L., Chan, S.C.-F. and Chung, F.-L. (2007), "Applying cross-level association rule mining to cold-start recommendations", *IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, IEEE, Hong Kong, pp. 133-8.

Castellano, E.J. and Martínez, L. (2009), "A web-decision support system based on collaborative filtering for academic orientation. Case study of the Spanish secondary school", *Journal of Universal Computer Science*, Vol. 15 No. 14, pp. 2786-807.

Chen, Z. and Ben-Arieh, D. (2006), "On the fusion of multi-granularity linguistic label sets in group decision making", *Computers and Industrial Engineering*, Vol. 51 No. 3, pp. 526-41.

Chiclana, F., Herrera, F. and Herrera-Viedma, E. (1998), "Integrating three representation models in fuzzy multipurpose decision making based on fuzzy preference relations", *Fuzzy Sets and Systems*, Vol. 97 No. 1, pp. 33-48.

Degani, R. and Bortolan, G. (1988), "The problem of linguistic approximation in clinical decision making", *International Journal of Approximate Reasoning*, Vol. 2 No. 2, pp. 143-62.

Diez, M. and Villegas, P. (2007), "Automatic recommendations for machine-assisted multimedia annotation: a knowledge-mining approach", *Semantic Multimedia*, Springer, Berlin, pp. 95-8.

Goldberg, D., Nichols, D., Oki, B.M. and Terry, D. (1992), "Using collaborative filtering to weave an information tapestry", *ACM. Communications of the ACM*, Vol. 35 No. 12, pp. 61-70.

Guttman, H.R. (1998), *Merchant Differentiation through Integrative Negotiation in Agent-mediated Electronic Commerce*, School of Architecture and Planning, Massachusetts Institute of Technology, Cambridge, MA.

Herlocker, J.L., Konstan, J.A., Borchers, A. and Riedl, J. (1999), "An algorithmic framework for performing collaborative filtering", *Proceedings of the 22nd International Conference on Research and Development in Information Retrieval, ACM, Berkeley, CA*, pp. 230-7.

Herlocker, J.L., Konstan, J.A., Terveen, K. and Riedl, J.T. (2004), "Evaluating collaborative filtering recommender systems", *ACM Transactions on Information Systems*, Vol. 22 No. 1, pp. 5-53.

Herrera, F. and Martínez, L. (2000), "A two-tuple fuzzy linguistic representation model for computing with words", *IEEE Transactions on Fuzzy Systems*, Vol. 8 No. 6, pp. 746-52.

Herrera, F., Herrera-Viedma, E. and Verdegay, J.L. (1995), "A sequential selection process in group decision making with linguistic assessment", *Information Sciences-Intelligent Systems*, Vol. 85 No. 4, pp. 223-39.

Herrera, F., Alonso, S., Chiclana, F. and Herrera-Viedma, E. (2009), "Computing with words in decision making: foundations, trends and prospects", *Fuzzy Optimization and Decision Making*, Vol. 8 No. 4, pp. 337-64.

Herrera-Viedma, E., Herrera, F., Chiclana, F. and Luque, M. (2004), "Some issues on consistency of fuzzy preference relations", *European Journal of Operational Research*, Vol. 154 No. 1, pp. 98-109.

Herrera-Viedma, E., López-Herrera, A.G., Alonso, S., Moreno, J.M., Cabrerizo, F.J. and Porcel, C. (2009), "A computer-supported learning system to help teachers to teach fuzzy information retrieval systems", *Information Retrieval*, Vol. 12 No. 2, pp. 179-200.

Horvath, T. (2009), "A model of user preference learning for content-based recommender systems", *Computing and Informatics*, Vol. 28 No. 4, pp. 453-81.

Kim, H.N., Ji, A.T., Kim, H.J. and Jo, G.S. (2007), "Error-based collaborative filtering algorithm for top-*n* recommendation", *Advances in Data and Web Management*, Springer, Berlin, pp. 594-605.

Kolodner, J. (1993), in Kolodner, J. and Kaufmann, M. (Eds), *Case Based Reasoning*, Morgan Kaufmann, San Mateo, CA.

Krulwich, B. (1997), "Lifestyle finder: intelligent user profiling using large-scale demographic data", *AI Magazine*, Vol. 18 No. 2, pp. 37-45.

Lalla, M., Facchinetti, G. and Mastroleo, G. (2004), "Ordinal scales and fuzzy set systems to measure agreement: an application to the evaluation of teaching activity", *Quality and Quantity*, Vol. 38 No. 5, pp. 577-601.

Li, Q., Myaeng, S.H., Guan, D.H. and Kim, B.M. (2007), "A probabilistic music recommender considering user opinions and audio features", *Information Processing and Management*, Vol. 43 No. 2, pp. 473-87.

Martínez, L. (2007), "Sensory evaluation based on linguistic decision analysis", *International Journal of Approximated Reasoning*, Vol. 44 No. 2, pp. 148-64.

Martínez, L., Pérez, L.G. and Barranco, M. (2007), "A multi-granular linguistic content-based recommendation model", *International Journal of Intelligent Systems*, Vol. 22 No. 5, pp. 419-34.

Martínez, L., Barranco, M.J., Pérez, L.G. and Espinilla, M. (2008a), "A knowledge-based recommender system with multigranular linguistic information", *International Journal of Computational Intelligence Systems*, Vol. 1 No. 3, pp. 225-36.

Martínez, L., Pérez, L.G., Barranco, M.J. and Espinilla, M. (2008b), "Improving the effectiveness of knowledge-based recommender systems using incomplete linguistic preference relations", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 16 No. 2, pp. 33-56.

Massa, P. and Bhattacharje, B. (2004), "Using trust in recommender systems: an experimental analysis", *2nd International Conference on Trust Management, Oxford*, Springer, Berlin.

Mata, F., Martínez, L. and Herrera-Viedma, E. (2009), "An adaptive consensus support model for group decision-making problems in a multi-granular fuzzy linguistic context", *IEEE Transactions on Fuzzy Systems*, Vol. 17 No. 2, pp. 279-90.

Porcel, C., López-Herrera, A.G. and Herrera-Viedma, E. (2009), "A recommender system for research resources based on fuzzy linguistic modeling", *Expert Systems with Applications*, Vol. 36 No. 3, pp. 5173-83.

Romero, C., Ventura, S., Zafra, A. and De Bra, P. (2009), "Applying web usage mining for personalizing hyperlinks in web-based adaptive educational systems", *Computers and Education*, Vol. 53 No. 3, pp. 828-40.

Schafer, J.B., Konstan, J.A. and Riedl, J. (2001), "E-commerce recommender applications", *Data Mining and Knowledge Discovery*, Vol. 5 Nos 1-2, pp. 115-53.

Sebastia, L., Garcia, I., Onaindia, E. and Guzman, C. (2009), "E-tourism: a tourist recommendation and planning application", *International Journal on Artificial Intelligence*, Vol. 18 No. 5, pp. 717-38.

Takacs, G., Pilaszy, I., Nemeth, B. and Tikk, D. (2009), "Scalable collaborative filtering approaches for large recommender systems", *Journal of Machine Learning Research*, Vol. 10, pp. 623-56.

Tanino, T. (1984), "Fuzzy preference orderings in group decision making", *Fuzzy Sets and Systems*, Vol. 12 No. 2, pp. 117-31.

Tanino, T. (1990), "Multiperson decision making using fuzzy sets and possibility theory", in Kacprzyk, J. and Fedrizzi, M. (Eds), *Group Decision Making under Fuzzy Preferences*, Kluwer Academic, Norwall, MA, pp. 172-85.

Trujillo, M., Millan, M. and Ortiz, E. (2007), "A recommender system based on multi-features", in Heidelberg, S.B. (Ed.), *Computational Science and Its Applications – ICCSA 2007*, Springer-Verlag, Berlin, pp. 370-82.

Uchyigit, G. and Ma, M.Y. (2008), *Personalization Techniques and Recommender Systems*, World Scientific, Singapore, p. 323.

Victor, P., De Cock, M. and Cornelis, C. (2008), "Getting cold start users connected in a recommender system's trust network", *8th International Conference on Fuzzy Logic and Intelligent Technologies in Nuclear Science, Spain*, World Scientific, Singapore, pp. 877-82.

Wang, R.Q. and Kong, F.S. (2007), "Semantic-enhanced personalized recommender system", *6th International Conference on Machine Learning and Cybernetics*, IEEE, Hong Kong, pp. 4069-74.

Xu, Z.S. (2006), "Incomplete linguistic preference relations and their fusion", *Information Fusion*, Vol. 7 No. 3, pp. 331-7.

Yager, R.R. (2003), "Fuzzy logic methods in recommender systems", *Fuzzy Sets and Systems*, Vol. 136 No. 2, pp. 133-49.

Yager, R.R., Goldstein, L.S. and Mendels, E. (1994), "FUZMAR: an approach to aggregating market research data based on fuzzy reasoning", *Fuzzy Sets and Systems*, Vol. 68 No. 1, pp. 1-11.

Zadeh, L.A. (1975), "The concept of a linguistic variable and its application to approximate reasoning – I, II, III", *Information Sciences*, Vol. 8, pp. 199-249; 301-357; Vol. 9, pp. 43-80.

Zhen, L., Huang, G.Q. and Jiang, Z.H. (2010), "An inner-enterprise knowledge recommender system", *Expert Systems with Applications*, Vol. 32 No. 2, pp. 1703-12.

Corresponding author
Rosa M. Rodríguez can be contacted at: rmrodrig@ujaen.es