MODELS FOR MANAGING INCOMPLETE INFORMATION IN RECOMMENDER SYSTEMS

L. Martínez, L.G. Pérez, M. Barrranco
Department of Computer Sciences, University of Jaén
Campus Las Lagunillas s/n. 23071. Jaén. Spain.
martin,lgonzaga,barranco@ujaen.es

ABSTRACT

Recommender Systems have recently emerged to assist network users in their search processes, due to the fact, these users must deal with a vast quantity of information that is stored in huge data bases of different e-shops, e-libraries, etc. The Recommender Systems help them in their search by means of recommendations that arise from information provided by different sources as the proper user, experts, other users, etc. Most of the current Recommender Systems force their users to provide the information in an only way, usually a numerical scale. Nevertheless, the information provided by the different sources get to use incomplete, vague and imprecise because it is related to their own perceptions, tastes and preferences. In other research areas as decision analysis, planning, scheduling, etc., this type of information has been successfully managed allowing the sources of information to express their preferences by means of different representation models and preference structures. In this contribution we shall review these models and structures together several resolution processes dealing with them in order to propose their use in the Recommender Systems to improve their accuracy and success in the recommendations.

KEYWORDS

Recommender systems, preference modeling, uncertainty, accuracy,

1. INTRODUCTION

One of the main problems users face navigating in Internet is the vast quantity of information they find. It is such amount of information that companies such as google (www.google.com) or amazon (www.amazon.com) use Recommender Systems to assist users in their search. The Recommender Systems are a class of software (P. Resnick and H.R.Varian, 1997) that has emerged in the last years within the domain of the E-Commerce (J.B. Schafer et al., 2001). Traditionally, these systems have fallen in three main categories: (i) *Collaborative filtering systems* (ii) *Content-based filtering systems* and (iii) *Hybrid content-based and collaborative recommender systems*.

These systems gather preference information from users, experts, etc., related to their preferences, tastes, and opinions about a given type of items (books, music, etc). Such a way that using this information the system ranks the items, and makes a recommendation about what items are the most attractive for each user. These systems can be seen as a kind of decision making process that tries to obtain a solution set of alternative/s (recommendations) from the information provided by the sources of information. And this information get to use subjective, vague and incomplete because it is related to the own perceptions of the sources. In spite of this, the most common way that the Recommendation Systems require the information to their sources is by means of a precise numerical scale (C. Hayes and P. Cunningham, 2001). This obligation implies a lack of expressiveness and hence a lack of precision in the recommendations. In the decision analysis literature we can find different ways to express this type of information by means of different representation models and preference structures together their respective resolution schemes have provided successful results (F. Chiclana et al., 1998; F. Herrera et al., 2004, Q. Zhang et al., K. Jae Kyeong and C. Sang Hyun, 2001). The application of these tools to Recommendation Systems can improve the effectiveness of the recommendations given by the Recommender Systems if they are able to deal with different representation models and preference structures according to the nature and uncertainty of the features and aspects assessed by the sources of information. Here, we shall review different representation models, preference structures and resolution models that have been used in the literature to deal with incomplete, vague and uncertain information in order to propose that Recommendation Systems can handle and offer these tools to support their users in their searching process improving their recommendations.

This contribution is set out as follows: Section 2 we present a general view of Recommender Systems. Section 3 we show different representation models and preference structures used to deal with vague and imprecise information. In the section 4, we review different resolution methods dealing with the above models and structures and finally we shall outline our future work in recommender systems.

2. A GENERAL VIEW OF RECOMMENDER SYSTEMS

Recommender Systems are a class of software to assist people in their search process in Internet or other networking processes. Their purpose is to recommend the most suitable items, from a set of items according to the customer, user's desires. This recommendation is made using different preferences and opinions gathered by the Recommender Systems from different types of sources. Most of Recommender Systems use one or two types of information sources. However, there exist at least five different types of information sources that can be used to obtain better recommendations (A. Ansari et al., 2000):

1. A person's expressing preferences or choices among alternative products.
2. Preferences for product attributes.
3. Other people's preferences or choices.
4. Expert Judgments.
5. Individual characteristics that may predict preferences.

A good recommendation system should be able to use any or all of these five types of information, potentially making better recommendations as more information become available.

Current recommender systems fall into three classes depending on the type of information and sources they use to make recommendations:

1. ***Collaborative filtering systems*** (D. Goldberg et al., 1992): they use explicit and implicit preferences from many users to filter and recommend objects to a given user, ignoring the representation of the objects. In the simplest case, these systems predict a person's preference as a weighted sum of other people's preferences, in which the weights are proportional to correlations over a common set of items evaluated by two people. Collaborative filtering algorithms were first introduced by Golberg and colleagues (D. Goldberg et al., 1992). They are used by Los Angeles Times, London Times, CRAYON, and Tango to customize online newspaper; by Movie Critic, Moviefinder and Morse to recommend movies; by barnesandnoble.com to recommend books, etc.
2. ***Content-based filtering systems:*** filter and recommend the items by matching user query terms with the index term used in the representation of the items, ignoring data from other users (M. Pazzani et al., 1996). There are some commercial systems has been offered by PersonalLogic, Frictionless Commerce, and Active Research that use self-explicated importance ratings and/or attribute trade-offs to make their recommendations.
3. ***Hybrid content-based and collaborative recommender systems***: This new class has emerged between the content-based and collaborative recommender systems and its aim is to smooth out the disadvantages of each one of them. A usual way to hybrid both classes is to make a two level filter algorithm, where we use first one of the algorithm (the content-based filtering algorithm) to obtain the first set of items and afterwards, we use the second algorithm (the collaborative filtering algorithm) to filter and recommend items from this set (C. Basu et al., 1998).

So far, we have seen what type of information is used by the Recommender Systems and we have aforementioned that most of the Recommender Systems force their sources to provide their information using a numerical scale despite, this information is vague, imprecise and uncertain. Our aim is to propose that these systems manage this information as in other areas in the literature such as Decision Making (Q. Zhang et al.); System for management support for restoring aquatic ecosystems and drainage areas (A. Jimenez et al., 2003), Information Systems evaluation problem (K. Jae Kyeong and C. Sang Hyun, 2001.) in order to improve the effectiveness of their recommendations.

3. INFORMATION REPRESENTATION MODELS AND STRUCTURES

In this section, we show different information representation models and preference structures that have been used in the literature to express and handle incomplete and vague information, where the information structures are referred to how the information is organized and the representation model is referred to the nature and domain in which the information is assessed by the source of information.

3.1 Information representation models

When we deal with incomplete, vague and imprecise information, the aspects assessed could have different nature (quantitative or qualitative) or the knowledge that the sources have on the aspects could be different as well. So according to the nature of the aspects qualified and the knowledge that the sources have on them this information could be expressed in different ways:

A) QUANTITATIVE ASPECTS

- Numerical values: if they are easy measurable the sources of information assess them by means of precise numerical values, so, we could work with a value scale, where the maximum value is the best rating and minimum value is the worst rating. However, when the items are related to qualitative aspects or when the knowledge about the items presents uncertainty it may be difficult to qualify them using precise values.
- Interval values: sometimes uncertainty is quite difficult to assess using a crisp and precise number although the aspect is quantitative. The first approach used in the literature to add some flexibility to the uncertainty representation problem was the use of intervals. In this case, the sources are not able to give a numerical value of a quantitative attribute but can provide their preference information using interval values assessed in a range (J.F. Le Téno and B. Mareschal, 1998).

B) QUALITATIVE ASPECTS

- Linguistic values: these aspects are realted to human perception. To model them, a better approach may be to use linguistic assessments instead of numerical values. The fuzzy linguistic approach represents qualitative aspects as linguistic values by means of linguistic variables (L.A. Zadeh, 1975).

  When we deal with linguistic information, we have to choose the appropriate linguistic descriptors for the term set and their semantics. In order to accomplish this objective, an important aspect to analyze is the "granularity of uncertainty", i.e., the level of discrimination among different counts of uncertainty. The universe of the discourse over which the term set is defined can be arbitrary, usually linguistic term sets are defined in the interval [0, 1].

  One possibility of generating the linguistic term set consists of directly supplying the term set by considering all terms distributed on a scale on which a total order is defined (R.R. Yager, 1995). For example, a set of seven terms S, could be given as follows:

  S ={s_0 = None; s_1 = Very Low; s_2 = Low; s_3 = Medium; s_4 = High; s_5 = Very High; s_6 = Perfect}

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

  (a) A negation operator Neg(s_i) = s_j such that $j = g\text{-}i$ (g+1 is the cardinality).

  (b) A minimization and a maximization operator in the linguistic term set.

  The semantics of the terms is given by fuzzy numbers defined in the [0,1] interval, which are described by membership functions. A way to characterize a fuzzy number is to use a representation based on parameters of its membership function (Bonissone and Decker (1986)). For example, we may use triangular membership functions and assign the following semantics to the set of seven terms, S, which is graphically shown in Figure 1.

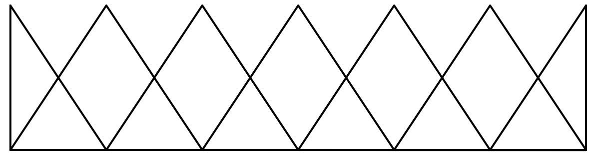

Figure 1: A set of seven terms and their semantics

3.2 Information Structures

When the sources of information provide their preferences so important is the representation model as the way to organize the information. Here we are going to show the different ways the sources can organize their information over a set of items $X = \{x_1, \ldots, x_n\}$ using different information structures:

1. *Preference ordering of the items*: In this case, the source, s_k , provides his preferences on X as an individual preference ordering, $O^k = \{o^k(1),..., o^k(n)\}$, where $o^k(\cdot)$ is a permutation function over the index set, $\{1,...n\}$ (F. Chiclana et al., 2001) . Therefore, according to this viewpoint, an ordered vector of items, from the best one to the worst one, is given: $O^k = \{i_1, i_2, i_3, i_5, i_4\}$.
2. *A preference relation*: With this representation, a source's preference, s_k , on X is describe by a preference matrix $A^k \subset X \times X$, $A^k = [a_{ij}^k]$, where a_{ij}^k indicates the preference intensity for item x_i to that of x_j (T. Tanino, 1984), i.e. $R : X \times X \rightarrow A$. The value, $R(x_i, x_j) = p_{ij}$, denotes the degree to which an item x_i is preferred to an item x_j where $p_{ij} \in A$:

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

   Where A can be any of the information representation models that has been seen in the section 3.1.
3. *An utility vector*: In this case, the source, s_k , provides its preferences on X as a set of n utility values, $U^k = \{u_i^k, i = 1,...n\}$, $u_i^k \in A$, where u_i^k represents the utility evaluation given by the source s_k to the item x_i (T. Tanino, 1990), i.e.: $U^k = \{u_1^k, u_2^k, u_3^k, u_4^k\}$ is the evaluation of the set of items $I = \{i_1, i_2, i_3, i_4\}$.Where $u_i^k \in A$ and A can be any of the information representation models that has been seen in the section 3.1. u_i^k represents the evaluation given to the item i_i and the most preferred item is the item that has the greatest evaluation, the worst item is the item that has the lowest evaluation.

4. DEALING WITH INCOMPLETE INFORMATION.

In the before section we have shown how can be represented and organised the vague and uncertain information by the sources of information. Due to the similarity between the processes carry out by the Recommendation Systems and the Decision Making processes in this section we shall review different literature about decision processes dealing with incomplete information modelled wit different representation models and preference structures (See Table 1).

Table 1. Decision Resolution Methods Dealing With Incomplete Information.

INFORMATION CONTEXT	MODEL
- Numerical and interval values. - Preference ordering of the items and utility vector.	P. Kyung Sam et al., 1996; K. Soung Hie et al., 1999; K. Jae Kyeong et al., 1998; K. Jae Kyeong and C. Sang Hyun, 2001; K. Soung Hie and A. Byeong Seok, 1997; K. Soung Hie and A. Byeong Seok, 1999
- Numerical and interval values. - Preference relations, preference ordering of the items, utility vector.	Zhang et al.
- Linguistic values, numerical values. - Preference ordering of the items, preference relations, utility vector.	F. Herrera and L. Martinez. 2001, F. Chiclana et al., 1998
- Linguistic values, numerical values. - Preference relations.	M. Delgado et al., 1998
- Linguistic values, numerical values. - Utility vector.	F. Herrera and L. Martinez, 1999
- Multi-granularity linguistic values. - Utility vector.	F. Herrera et al., 2000;F. Herrera and L. Martinez. 2001;F. Herrera and L. Martinez, 2001 (b)

5. FUTURE WORKS

In the future we want to use several of the methods presented in the Table 1 in the Recommendation Systems processes to improve the results of these systems. We shall mainly focus on the methods and models presented in (F. Chiclana et al., 1998; F. Herrera et al., 2004) such that the Recommendation Systems can deal with:

1. Numerical information.
2. Interval information.
3. Linguistic information.
4. Preference ordering vectors.
5. Preference relations.
6. Utility vectors.

Several of these methods and models have been used successfully in Decision Making and evaluation processes in (L. Martínez et al, 2004) and in (L.G. Pérez et al., 2004) we can find a first approach, in recommender systems, of a model that try to manage this kind of problems.

ACKNOWLEDGEMENT

This work is partially supported by the Research Projects TIC2002-03348, UJA23(UCUA), PromoJaén Project belongs to EQUAL Initiative and FEDER Funds.

REFERENCES

1. A. Ansari et al., 2000 . *Internet Recommendation Systems*. Journal of Marketing Research XXXVII:363- 375.
2. C. Basu et al., 1998. *Recommendation as CIassification: Using Social and Content-Based Information in Recommendation*. Proceedings of the Fifteenth National Conference on Artificial Intelligence, pp. 714 720.
3. P. Bonissone and K. Decker, 1986. *Selecting Uncertainty Calculi and Granularity: An Experiment in Trading-Off Precision and Complexity*. In L.N. Kanal and J.F. Lemmer, editors, Uncertainty in Artificial Intelligence. Norh-Holland,.
4. F. Chiclana et al., 2001. *Integrating Multiplicative Preference Relations in a Multipurpose Decision-Making Model Based on Fuzzy Preference Relations*. Fuzzy Sets and Systems 122 (2): 277-291.
5. F. Chiclana et al., 1998. *Integrating Three Representation Models in Fuzzy Multipurpose Decision Making Based on Fuzzy Preference Relations*. Fuzzy Sets and Systems 97 (1998), pp. 33-48

6. M. Delgado et al., 1998. *Combining Numerical and Linguistic Information in Group Decision Making*. Information Science, no. 107:177-194.
7. D. Goldberg et al., 1992. *Using Collaborative Filtering to Weave an Information Tapestry*. Communications of the ACM 35 (12): 61-70.
8. C. Hayes and P. Cunningham. 2001. *Smart Radio - Community Based Music Radio*. Knowledge-Based Systems 14 (3-4): 197-201.
9. F. Herrera et al., 2000. *A Fusion Approach for Managing Multi-Granularity Linguistic Term Sets in Decision Making*. Fuzzy Sets and Systems 114 (1): 43-58.
10. F. Herrera and L. Martinez, 1999. *An Approach For Combining Linguistic And Numerical Information Based On The 2-Tuple Fuzzy Linguistic Representation Model In Decision-Making*. International Journal of Uncertainty , Fuzziness and Knowledge -Based Systems. 8.5, pp. 539-562.
11. F. Herrera and L. Martinez. 2001. *The 2-Tuple Linguistic Computational Model. Advantages of its Linguistic Description, Accuracy and Consistency*. International Journal of Uncertainty , Fuzziness and Knowledge-Based Systems 9:33-48.
12. F. Herrera and L. Martinez, 2001 (b). *A Model Based on Linguistic 2-Tuples for Dealing with Multigranularity Hierarchical Linguistic Contexts in Multiexpert Decision-Making*. IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics 31 (2): 227-234.
13. F. Herrera, et al., 2004. *Managing non-homogeneous information in group decision making*. European Journal of Operational Research. In Press.
14. K. Jae Kyeong and C. Sang Hyun, 2001. *A Utility Range-Based Interactive Group Support System for Multiattribute Decision Making*. Computer & Operations Research, no. 28:485-503.
15. K. Jae Kyeong et al., 1998. *An Interactive Procedure for Multiple Criteria Group Decision Making with Incomplete Information*. Computers and Industrial Engineering 35 (1-2): 295-298.
16. A. Jimenez et al., 2003. *A Decision Support System for Multiattribute Utility Evaluation Based on Imprecise Assignments*. Decision Support Systems, no. 36:65-79.
17. P. Kyung Sam et al., 1996. Establishing Strict Dominance Between alternatives with Special Type of Incomplete Information. European Journal of Operational Research, no. 96:398-406.
18. L. Martínez et al, 2004. *A Linguistic Evaluation Model For Engineering Systems.* Proceedings of the XII Spanish Conference on Fuzzy Logic and Technology, pp. 153-158.
19. M. Pazzani et al., 1996. *Syskill & Webert: Identifying Interesting Web Sites*. In Proceedings of the Thirteenth National Conference on Artificial Intelligence AA.AI, no. 96:54-61.
20. L.G. Pérez et al., 2004. *Un modelo de recomendación con perfiles de usuario lingüísticos.* Proceedings of the XII Spanish Conference on Fuzzy Logic and Technology, pp. 521-526.
21. P. Resnick and H.R. Varian, 1997. *Recommender Systems*. Communications of the ACM 40:56-58.
22. J.B. Schafer et al., 2001. *E-Commerce Recommendation Applications*. Data Mining and Knowledge Discovery, no. 5:115-153.
23. K. Soung Hie et al., 1999. *An Interactive Procedure for Multiple Attribute Group Decision Making with Incomplete Information: Range-Based Approach.* European Journal of Operational Research, no. 118:139-152.
24. K. Soung Hie and A. Byeong Seok, 1997. *Group Decision Making Procedure Considering Preference. Strength under Incomplete Information*. Computer & Operations Research 24 (12): 1101—1112.
25. K. Soung Hie and A. Byeong Seok, 1999. *Interactive Group Decision Making Procedure under Incomplete Information*. European Journal of Operational Research, no. 116:498—507.
26. T. Tanino, 1984. *Fuzzy Preference Ordering in Group Decision Making*. Fuzzy Sets and Systems, no. 12:117-131.
27. T. Tanino, 1990. *On Group Decision Making under Fuzzy Preferences*. In: J. Kacprzyk, M. Fedrizzi (Eds.), Multiperson Decision Making Using Fuzzy y Sets and Possibility Theory, Kluwer Academic Publisher, Dordrecht, pp. 172- 185.
28. J.F. Le Téno and B. Mareschal, 1998. *An Interval Version of PROMETHEE for the Comparison of Building Products Design with Ill-Defined Data on Environmental Quality*. European Journal of Operational Research 109 (2): 522 -529.
29. R.R. Yager, 1995. *An Approach to Ordinal Decision Making*. International Journal of' Approximate Reasoning, no. 12:237 261.
30. L.A. Zadeh, 1975. *The Concept of a Linguistic Variable and its Applications to Approximate Reasoning*. Information Sciences, Part I, II, III, pp. 8,8,9:199-249,301-357.43-80.
31. Q. Zhang et al., In Press, Corrected Proof. *Decision Consolidation: Criteria Weight Determination Using Multiple Preference Formats*. Decision Support Systems.