A Multi-granular Linguistic Based-Content Recommender System Model

L. Martinez L.G. Pérez M. Barranco F. Mata

Dept. Of Computer Sciences, University of Jaén, 23071, Jaén, Spain

Abstract

Recommender systems have recently emerged in the area of e-commerce to assist consumers in their search processes by means of recommendations that arise from the information provided by different sources, due to the fact, they must deal with a vast quantity of information that is stored in huge data bases of different e-shops. We focus in based-content recommender systems that deal with information provided by customers regarding the features they wish, in order to reach the most suitable item according to their preferences. Most of the current recommender systems force their customers to provide the information in a numerical scale in spite of it is usually incomplete, vague and imprecise. We propose a multi-granular linguistic based-content recommender system model that manage the uncertainty of the information using linguistic terms and allow each source to use their own linguistic term set to express their preferences according to their knowledge.

Keywords: e-commerce, e-services, recommender systems, fuzzy linguistic approach, decision-making.

1. Introduction

One of the main problems users face navigating in Internet is the quantity of information they find, being most of it useless for their aims. Due to this fact, different e-services have raised to help them to reach easily and quickly their necessities. In this paper, we focus in Recommender Systems [11], a class of software that has emerged in the last years within E-Commerce area [13], that helps the customers to find out the most suitable items according to their preferences, necessities or taste, hiding or removing the useless information. Companies such as google or amazon use them to assist people in their searches.

The purpose of these systems is to recommend the most suitable items, from a set of them (item database) according to the customer's desires. Traditionally, these systems have fallen in three main categories: (i) Collaborative filtering systems [5]. (ii) Content-based filtering systems [9] and Hybrid systems [8]. They gather preference information from the customers, experts, etc., rank the items belonging to the item database according to their preferences, and make a decision about which items are the most attractive to the customers. This decision is made taking into account the information gathered by these systems from different types of information sources [1]. This information gets to use incomplete, vague and imprecise because it is related to perceptions, tastes and preferences. However, most of recommender systems force their sources to express their preferences using only one numerical scale [6]. This fact implies a lack of expressiveness for the sources and maybe a lack of precision in the recommendations made by the systems.

In this paper, we propose a new model to improve the effectiveness of the recommendations given by the content-based recommender system models. It consist of offering the customers the possibility to express their preference information using linguistic assessments instead of numerical ones, due to the fact that linguistic information is more suitable to assess qualitative information (human perceptions, taste, necessities). In addition, our model will let the customers choose their own linguistic term set to provide their preference information according to their knowledge about the items, therefore the context in which the recommendations are computed is a multi-granular linguistic context [7]. To deal with the multi-granular linguistic information in our recommendation model we shall use the fuzzy linguistic approach [14] to model the customer profiles and fuzzy tools, such as, fuzzy measures of comparison [4] to make comparison among the different elements (items and consumer profiles) and the non-dominance choice degree [10] to rank the items.

We focus on content-based recommender models that filter and recommend items according to a matching process among the customer profile and the description of the items in order to choose those one(s) pretty similar to the customer's tastes.

Our proposal for a multi-granular linguistic content-based recommendation model will act according to the followings steps (figure 1):

1. *Acquisition of the customer profile:* The customer profile is an information structure, in our case an utility vector, used by our model to gather the information provided by the customer about his/her opinions, necessities... In this model the customers will provide their profiles by means of linguistic information and depending on the aspect they are assessing they can use linguistic assessments belonging to different linguistic term sets.
2. *Matching items for the customer:* to find out the most suitable items for the customers, the model compares the features of every item of the item database with the customer necessities by means of fuzzy measurements of comparison. Therefore, after this step, the model obtains a fuzzy set that measures the similarity between the customer profile and each item of the item database.
3. *Making a Recommendation:* to recommend the most suitable items for a customer, the similarity values obtained in the step before are ranked by means of a non-dominance degree, such that, the top ranked items will be recommended to the customers.

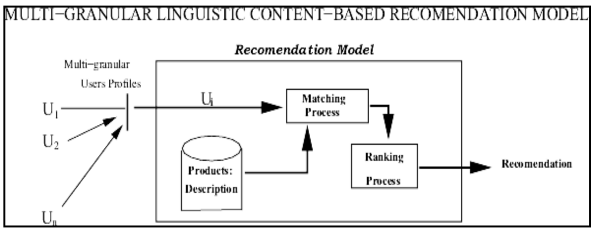

Fig. 1: The Content-Based Recommendation Model

This paper is structured as follows. In the Section 2 we shall make brief review of the fuzzy linguistic approach. In the Section 3 we present our multi-granular based content recommendation model, and finally, some conclusions are point out.

2. Fuzzy Linguistic Approach

Usually, we work in a quantitative setting, where the information is expressed by means of numerical values. However, many aspects of different activities in the real world cannot be assessed in a quantitative form, but rather in a qualitative one, i.e., with vague or imprecise knowledge. In that case, 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 [14]. This approach is adequate in some situations where the information may be unquantifiable due to its nature, and this, it may be stated only in linguistic terms.

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. Therefore, according to the source of information knowledge it can choose different counts of uncertainty. Typical values of cardinality used in the linguistic models are odd ones, such as 7 or 9, where the mid term represents an assessment of "approximately 0.5", and with the rest of the terms being placed symmetrically around it. In this paper, we shall deal with sources of information with different degrees of knowledge, so each one could use different linguistic term sets with different granularity. We call this context as multi-granular linguistic context [7].

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. For example, a set of seven terms S, could be given as follows:

$\{s_0 : N; s_1 : VL; s_2 : L; s_3 : M; s_4 : H; s_5 : VH; s_6 = P\}$

In these cases, it is required that there exist:

- A negation operator $\mathrm{Neg}(s_i) = s_j$ such that $j = g-i$ (g+1 is the cardinality).
- A minimization and a maximization 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 [0,1] interval. A way to characterize a fuzzy number is to use a representation based on parameters of its membership function [2]. The linguistic assessments given by the users are just approximate ones, some authors consider that linear trapezoidal membership functions are good enough to capture the vagueness of those linguistic assessments.

This parametric representation is achieved by the 4-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 [2]. A particular case of this type of representation are the linguistic assessments whose membership functions are triangular, i.e., b = d, so we represent this type of membership function by a 3-tuple (a; b; c). An example may be the following figure:

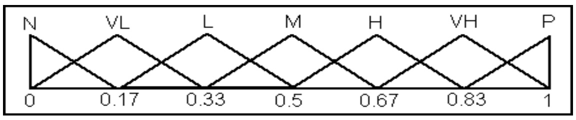

Fig. 2: A linguistic term set of seven terms and its semantics

Other authors use a non-trapezoidal representation, e.g., Gaussian functions [3].

3. Multi-granular Linguistic Content Based Recommendation Model

Here, we present our proposal for a multi-granular linguistic content-based recommendation model. The recommendation process will consist of a matching process between the customer profiles and the items features of the item database. In our proposal, the consumer profiles can be assessed by means of multi-granular linguistic information, it means, different customers can use different linguistic term sets to assess their profiles. And the item features are provided by experts in the area and as well could be assessed in a multi-granular linguistic context. Our model has the following three stages (figure 1).

1. Acquisition of the customer profile.
2. Matching items for the customer.
3. Making a recommendation.

In the following subsections we shall present in detail these stages.

3.1. Acquisition of customer profile and item features.

The aim of this stage is to gather information about the preferences, necessities or tastes of a customer, u_k, and build an utility vector, $P_k=\{p_1^k,\ldots,p_l^k\}$, by means of a set of criteria, $C=\{c_1,\ldots,c_l\}$, that meet this information, where $p_i^k \in S_{ki}$ is the assessment of the criterion c_i given by the customer u_k and where S_{ki} is a linguistic term set chose by u_k to assess c_i according to his/her knowledge.

The recommender system model has a set of items or products (the item database), $A=\{a_1,\ldots,a_n\}$, that can be recommended. Each item, a_i, is described in the database by an utility vector $F_i=\{v_1^i,\ldots,v_l^i\}$, by means of the same criteria set C used with the customer profile, being v_i^j the assessment of the criterion c_j of the item a_i. This assessment is a linguistic label that belongs to a linguistic term set.

In this stage we offer the possibility that each customer can assess their preferences or necessities in different linguistic term sets according to their knowledge. So, in our proposal we offer the customers a flexible multi-granular linguistic context instead of forcing all of them to provide their preference in the same scale.

3.2. Matching Items for each customer

Once we have got the consumer profiles the recommendation model will have:

a) A consumer profile $P_k=\{p_1^k,\ldots,p_l^k\}$ with the consumer preferences provided by the consumer, u_k,
b) A set of items $A=\{a_1,\ldots,a_n\}$, where each item is described in the database by means a vector of features $F_i=\{v_1^i,\ldots,v_l^i\}$ (Table 1).

	c_1	,...,	c_h	,...,	c_l
a_1	v_1^1	,...,	v_k^1	,...,	v_l^1
...	...		...		...
a_j	v_1^j	,...,	v_k^j	,...,	v_l^j
...	...		...		...
a_n	v_1^n	,...,	v_k^n	,...,	v_k^n

Table 1: Items of the item database

In order to find out which is the most suitable item/s for a customer, u_k, we shall compare customer profile, P_k, with the features of all the items of the recommender system by means of a matching process in order to obtain the closest items in the database according to consumer preferences or necessities.

The process of matching items involves the comparison between fuzzy sets. These comparisons are usually carried out by means of a measure of comparison. We focus in measures of comparison which evaluate the resemblance or likeness of two objects (fuzzy sets in our case) [12]. For simplicity, in this paper we shall choose a measure based on a possibility function, $D(A,B) = sup_x \min (f_A(x), f_B(x))$, where f_A and f_B are the membership functions of the fuzzy sets A and B respectively.

To accomplish the task of matching the items we shall measure the resemblance between the customer profile and the features of each item of the item database using the measure function D. For example, to compute the resemblance, R_i^k, between the customer profile P_k and the item a_i we shall calculate:

$$R_i^k = (r_1^i,\ldots,r_l^i) = (D(p_1^k, a_1),\ldots,D(p_l^k, a_n))=$$
$$=(sup_x \min (p_1^u, v_1^i),\ldots,sup_x \min(p_l^u, v_l^i))$$

Finally, we shall obtain the matching between the customer profile and the items as:

	c_1	,...,	c_h	,...,	c_l
R_1^k	r_1^1	,...,	r_k^1	,...,	r_l^1
...	...		...		...
R_j^k	r_1^j	,...,	r_k^j	,...,	r_l^j
...	...		...		...
R_n^k	r_1^n	,...,	r_k^n	,...,	r_k^n

Table 2: Similarity among the user profile and the items

Where $r_j^k \in [0,1]$.

3.3. Making a Recommendation

The aim of this stage is to rank the items by means of their similarity degree and recommend the best ones. This stage is performed in the following two phases:

1. *Computing a preference degree:* in this phase, we obtain the preference relation $B = \lfloor b_{ij} \rfloor$ where b_{ij} is a degree of possibility of dominance of a_i over a_j and is obtained according to the following formula:

$$b_{ij} = \inf_x \min \left(1 - f_{R_j^u}(x) + f_{R_i^u}(x), 1\right)$$

This formula is based on a measure of inclusion [4] and determines how much cover a_i over a_j or in our case, how much better is a_i than a_j.

2. *Applying the non-dominance choice degree:* for each item we calculate its non-dominance choice degree NDD_i [10] as:

$$NDD_i = \min_x \left\{1 - b_{ji}^s, j \neq i\right\}$$

where $b_{ji}^s = \max\{b_{ji} - b_{ij}, 0\}$ represents the degree to which a_i is strictly dominated by a_j.

Now, with the non-dominance choice degree we can rank the items. The best ones are those with a greater NDD, i.e., the items less dominated by the others. We must take into account that there can be several alternatives with the same NDD. These items will have the same position in the ranking. Finally, we shall make the recommendation selecting top item(s) ranked.

4. Concluding Remarks

In this paper, we have presented a multigranular linguistic content-based recommender system model that improve the recommendations because it offers customers the possibility to express their preference information using linguistic assessments instead of numerical ones, due to the fact that linguistic information is most suitable to assess qualitative criteria. Moreover, we let customer choose their own linguistic term set to provide their preference information according to their knowledge.

Acknowledgments

This work has been partially supported by the Research Projects TIC 2003-03348 and TIC 2002-11942-E

5. References

[1] A. Ansari, S. Essegaier, and R. Kohli. "Internet Recommendation Systems". *Journal of Marketing Research XXXVII* (2000) 363- 375.

[2] P.P. Bonissone and K.S. Decker, "Selecting uncertainty calculi and granularity: an experiment in trading-off precision and complexity", in: L.H. Kanal, J.F. Lemmer, (Eds.), *Uncertainty in Artificial Intelligence*, North-Holland, Amsterdam, (1986) 217-247.

[3] G. Bordogna and G. Passi, "A fuzzy linguistic approach generalizing boolean information retrieval: a model and its evaluation", *J. Amer. Soc. Inform. Sci.* 44 (1993) 70-82.

[4] B. Bouchon-Meunier, M. Rifqi and S. Bothorel, "Towards general measures of comparison of objects", *Fuzzy Sets and Systems*, vol. 84, n.2 (1996), p. 143-153.

[5] D. Goldberg, D. Nichols, B.M. Oki, and D. Terry. "Using Collaborative Filtering to Weave", *Communications of the ACM* 35 (12) (1992).

[6] C. Hayes and P. Cunningham, "Smart Radio - Community Based Music Radio". *Knowledge-Based Systems* 14 (3-4) (2001) : 197-201.

[7] F. Herrera, E. Herrera-Viedma and L. Martínez. "A fusion approach for managing multi-granularity linguistic term sets in decision making". *Fuzzy Sets and Systems*. **114** 43-58 (2000).

[8] E. Herrera-Viedma, F. Herrera, L. Martinez, J.C. Herrera, and A.G. Lopez, "Incorporating Filtering Techniques in a Fuzzy Linguistic Multi-Agent Model for Gathering of Information on the Web". *Fuzzy Sets and Systems*, (2004) no. To appear.

[9] H. Lieberman, "Letizia: An Agent That Assists Web Browsing". *In Proceedings of the 1995 International Joint Conference on Artificial Intelligence* (1995). Montreal, Canada.

[10] S.A. Orlovski, "Decision Making with a Fuzzy Preference Relation", *Fuzzy Sets and Systems* 1 (1978) 155-167.

[11] P. Resnick and H.R. Varian, "Recommender Systems". *Communications of the ACM* 40 (1997):56-58.

[12] M.Rifqi, V.Berger, B.Bouchon-Meunier, "Discrimination power of measures of comparison", *Fuzzy Sets and Systems* 110 (2000) 189-196.

[13] J.B. Schafer, J. A. Konstan, and J. Riedl, "E-Commerce Recommendation Applications". *Data Mining and Knowledge Discovery*, no. 5 (2001):115-153.

[14] L.A. Zadeh. "The concept of a linguistic variable and its application to approximate reasoning". *Information Science* (8 and 9): (Part I and II) 8, pp 199-249 and pp. 301-357, (Part III) 9, pp, 43-80, (1975).