2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology

A fuzzy linguistic recommender system to disseminate the own academic resources in universities

C.Porcel
Department of Computer Science
University of Jaén
23071 Jaén, Spain
Email: cporcel@ujaen.es

E.Herrera-Viedma
Department of Computer Science and A.I.
University of Granada
18071 Granada, Spain
Email: viedma@decsai.ugr.es

***Abstract*—We propose a recommender system for the dissemination of own academic resources in a University Digital Library (UDL). With this system the digital library offers to the university faculty the own indigenous knowledge generated in the university. The system recommends both specialized and complementary resources, and collaboration possibilities among university membership to form multidisciplinary working groups. So, this system increases the internal social collaboration possibilities in an academic environment and it contributes to improve the services provided by a UDL.**

***Keywords*-recommender systems; fuzzy linguistic modeling; university digital libraries.**

I. Introduction

Digital libraries are information collections that have associated services delivered to user communities using a variety of technologies [4]. As digital libraries become commonplace and as their contents and services become more varied, the users expect more sophisticated services from their digital libraries [3], [4], [14]. A service that is particularly important is the selective *dissemination of information*. In [17] the importance of the role of digital libraries in the preservation and dissemination of indigenous knowledge is emphasized. It will increase the visibility of the academic departments and research groups to the campus communities as well as to the society. Therefore, the dissemination of indigenous knowledge will allow the researchers to meet another researchers with the aim to discover collaboration possibilities, and so, to form multidisciplinary working groups.

In this paper we present a recommender system [6], [12], [15] designed using a hybrid approach and assuming a multi-granular Fuzzy Linguistic Modeling (FLM) [1], [7], [8], [19]. The system is oriented to researchers and it recommends them three types of resources:

1) Resources of the user research area to achieve the user specialization.
2) Other resources as complementary formation.
3) Partners or research collaborators of the proper institution, in order to include researchers of related areas that could be interesting to discover collaboration possibilities and to form multidisciplinary groups.

In section 2 we introduce the preliminaries, that is, the basis of recommender system and FLM. Section 3 presents the design of the system. Finally, in section 4 some concluding remarks are pointed out.

II. Preliminaries

A. Basis of recommender systems

Recommender systems help online users in the effective identification of items suiting their wishes, needs or preferences. They has the effect of guiding the user in a personalized way to relevant or useful objects in a large space of possible options [2]. These applications improve the information access processes for users not having a detailed product domain knowledge. They are becoming a popular tools to reduce the information overload and to improve the sales in e-commerce web sites [2], [15], [16].

In order to generate personalized recommendations that are tailored to the user's preferences or needs, recommender systems must collect personal preference information, such as user's history of purchase, items previously interesting for the user, click-stream data, demographic information, and so on. Two different ways to obtain information about user preferences are distinguished [15], implicit and explicit approaches, although many systems adopt a hybrid approach.

There are several approaches that have been proposed to the implementation of recommender applications [2], [5], [6], [15]. In this paper we propose the use of a hybrid approach to smooth out the disadvantages of each one of them and to exploit their benefits; using a hybrid strategy users are provided with recommendations more accurate than those offered by each strategy individually.

The recommendation activity is followed by a relevance feedback phase. It is a cyclic process whereby the users feed back into the system decisions on the relevance of retrieved documents and the system uses these evaluations to automatically update the user profiles [6], [15].

978-0-7695-3801-3/09 $26.00 © 2009 IEEE
DOI 10.1109/WI-IAT.2009.257

B. Fuzzy linguistic modeling

There are situations in which the information cannot be assessed precisely in a quantitative form but may be in a qualitative one. The use of Fuzzy Sets Theory has given very good results for modeling qualitative information [19] and it has proven to be useful in many problems. It is a tool based on the concept of *linguistic variable*.

1) The 2-tuple fuzzy linguistic approach: The 2-tuple FLM [8] is a continuous model of representation of information. To define it we have to establish a 2-tuple representation model and a 2-tuple computational model to represent and aggregate the linguistic information, respectively.

Let $S = \{s_0, ..., s_g\}$ be a linguistic term set with odd cardinality, where the semantics of the labels is given by means of triangular membership functions and we consider all terms distributed on a scale on which a total order is defined. If a symbolic method aggregating linguistic information [8] obtains a value $\beta \in [0, g]$, and $\beta \notin \{0, ..., g\}$, then β is represented by means of a 2-tuples (s_i, α_i), where s_i represents the linguistic label, and α_i is a numerical value expressing the value of the translation from the original result β to the closest index label, i, in the linguistic term set ($s_i \in S$). This model defines a set of functions for the transformation between numeric values and 2-tuples: $\Delta(\beta) = (s_i, \alpha)$ and $\Delta^{-1}(s_i, \alpha) = \beta \in [0, g]$ [8].

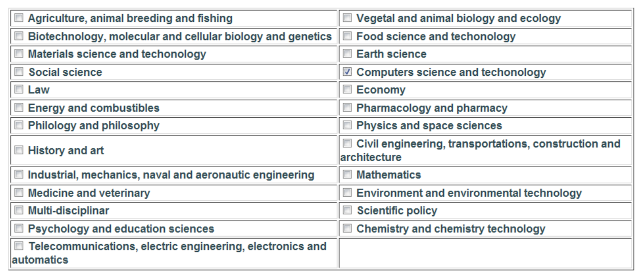
The computational model is defined by presenting the negation operator, comparison of 2-tuples and aggregation operators. Using functions Δ and Δ^{-1}, any of the existing aggregation operator (such as arithmetic mean) can be easily extended for dealing with linguistic 2-tuples [8].

2) The multi-granular fuzzy linguistic modeling: According to the uncertainty degree that an expert qualifying a phenomenon has on it, the linguistic term set chosen to provide his knowledge will have more or less terms. When different experts have different uncertainty degrees on the phenomenon or when an expert has to assess different concepts, then several linguistic term sets with a different granularity are necessary [9]. In such situations, we need tools to manage multi-granular linguistic information. In [9] a multi-granular 2-tuple FLM based on the concept of linguistic hierarchy is proposed. A *Linguistic Hierarchy (LH)*, is a set of levels *l(t,n(t))*, where each level t is a linguistic term set with different granularity, i.e., with different number of linguistic terms, $n(t)$ [9]. The levels are ordered according to their granularity. We can define a level from its predecessor level as: $l(t, n(t)) \rightarrow l(t+1, 2 \cdot n(t) - 1)$. In [9] a family of transformation functions between labels from different levels was defined. To define the computational model, we select a level to make the information uniform and then we can use the operators defined in the 2-tuple FLM.

III. A MULTIDISCIPLINARY RECOMMENDER SYSTEM TO DISSEMINATE INDIGENOUS KNOWLEDGE IN UDL

A. Information representation

To represent the resource scope we use the *vector space model* [10], where for each resource the system stores a vector VR. To build this vector we follow a generic classification with 25 disciplines (figure 1). In each position the vector stores a 2-tuple representing the importance degree for the resource scope of the discipline in that position. To represent the topics of interest of the researchers we also use the vector space model [10]. To build this vector, VU, we follow the same classification used in resource management. In each position the vector stores a 2-tuple representing the importance degree for the researchers of the discipline in that position. When a new researcher is inserted, the first action to confirm his/her register is to assess more than 15 resources.



☐ Agriculture, animal breeding and fishing	☐ Vegetal and animal biology and ecology
☐ Biotechnology, molecular and cellular biology and genetics	☐ Food science and techonology
☐ Materials science and techonology	☐ Earth science
☐ Social science	☑ Computers science and techonology
☐ Law	☐ Economy
☐ Energy and combustibles	☐ Pharmacology and pharmacy
☐ Philology and philosophy	☐ Physics and space sciences
☐ History and art	☐ Civil engineering, transportations, construction and architecture
☐ Industrial, mechanics, naval and aeronautic engineering	☐ Mathematics
☐ Medicine and veterinary	☐ Environment and environmental technology
☐ Multi-disciplinar	☐ Scientific policy
☐ Psychology and education sciences	☐ Chemistry and chemistry technology
☐ Telecommunications, electric engineering, electronics and automatics	

Figure 1. Disciplines used in the representation of the resources scope.

We distinguish three concepts: *Importance degree* of a discipline with respect to a resource scope or researcher preferences, *Relevance degree* of a resource for a researcher, and *Compatibility degree* between two researchers. So, we use different label sets S_1, S_2, and S_3 chosen from a LH, to represent the different concepts to be assessed. Specifically we use level 2 (5 labels) to assign importance degree ($S_1 = S^5$), and the level 3 (9 labels) to assign relevance degrees ($S_2 = S^9$) and complementary degrees ($S_3 = S^9$).

B. Recommendation strategy

The system uses a hybrid approach [2], [6], [11] with a posterior phase in which a *complementary recommendation* is carried out. The purpose of this complementary recommendation is to favor the interconnection between researchers of related areas, but not exactly of the same area [13]. To do this, it searches for a medium value of similarity, applying a Gaussian function to the similarity value obtained between the resource scope and the researcher topics of interest or between two researchers.

To calculate the similarity we use the following linguistic measure:

$$\sigma_l(V_1, V_2) = \Delta(g \times \frac{\sum_{k=1}^{n}(\hbar_1 \times \hbar_2)}{\sqrt{\sum_{k=1}^{n}(\hbar_1)^2} \times \sqrt{\sum_{k=1}^{n}(\hbar_2)^2}})$$

where g is the granularity of the used term set, n is the number of terms used to define the vectors (i.e. the number of disciplines), $\hbar_i = \Delta^{-1}(v_{ik}, \alpha_{vik})$ and (v_{ik}, α_{vik}) is the 2-tuple linguistic value of term k in the resource vector or in the researcher topics of interest vector (V_i).

When a new resource is inserted into the system, $\sigma_l(V_i, V_j)$ is computed between the new resource scope vector (V_i) against all the stored resources in the system (V_j). If $\sigma_l(V_i, V_j) \geq \alpha$ (linguistic threshold value), the resource j is chosen as similar to i. Next, the system searches for the users which were satisfied with these similar resources. To obtain the relevance of the resource i for a selected researcher x, the system aggregates (using the arithmetic mean defined in [8]) the $\sigma_l(V_i, V_j)$ with the recommendations previously provided by x about the similar resources. To aggregate the information we need to transform the value $\sigma_l(V_i, V_j)$ to a linguistic label in S_2, using the transformation function defined in [9].

If the calculated relevance degree is greater than a linguistic threshold μ, then, the system will send to the selected users the resource as specialized information. If not, the system estimates if the resource could be interesting as a complementary recommendation. The system calculates $\sigma_l(V_i, V_x)$ between the resource i and the researcher x (for all researchers). Then, it applies a so called multidisciplinary function to the value $\sigma_l(V_i, V_x)$. This function is called multidisciplinary because the idea is to give greatest weights to similarity middle values (near 0.5). Values of total similarity contribute with efficient recommendations but are probably known for the researchers. Similarly, null values of similarity show a null relationship between areas. To establish this function (figure 2) we use the centered OWA operators in which the OWA weights are generated from a Gaussian type function [18]. If the obtained multidisciplinary value is greater than a linguistic threshold γ, the system recommends the resource as complementary knowledge. To express the multidisciplinary value as a linguistic label in S_3, the transformation function defined in [9] is used.

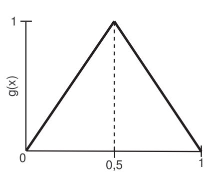


Figure 2. Triangular function.

Finally, the system extracts the authors (researchers) of the resources selected as recommendations. The system calculates $\sigma_l(V_x, V_a)$ between the researcher x and all the extracted authors a. Then, a multidisciplinary function is applied to the value $\sigma_l(V_x, V_a)$ to give greatest weights to middle similarity values. To establish this function the system also uses the centered OWA operators in which the OWA weights are generated from a Gaussian type function [18]. If the obtained value is greater than a threshold λ, the system considers that the two researchers could collaborate and it recommends this collaboration. The system expresses this multidisciplinary value as a complementary degree, that is, as a linguistic label in S_3.

Once completed this process, the system sends to the selected researchers the recommendations about the specialized resources and their estimated relevance degree, or the complementary resources and their estimated linguistic complementary degree. The system also sends the collaboration possibilities along with the estimated complementary degree between the researchers.

In the following, we describe the process when a new researcher is inserted. The first step is to identify the researchers most similar to the new researcher, using a similarity function. $\sigma_l(V_x, V_y)$ is calculated between the new researcher (V_x) against all researchers in the system (V_y). If $\sigma_l(V_x, V_y) \geq \delta$, the research y is chosen as near neighbor of x. Next, the system searches for the resources which were interesting for the neighbors of x to recommend them to x. To obtain the relevance of a resource i for the researcher x, the system aggregates $\sigma_l(V_x, V_y)$ with the assessments previously provided about i by the nearest neighbors of x. If the calculated relevance degree is greater than the linguistic threshold μ, then the system recommends to the new researcher the resource information and its calculated linguistic relevance degree (label of S_2).

The recommendation activity is followed by a relevance feedback phase in which the researchers feed back into the system decisions on the relevance of retrieved documents and the system uses these evaluations to automatically update the researchers profiles [6], [15].

C. System evaluation

At present we have implemented a trial version, in which the system works only with few researchers. This beta version has been used to prove the system functionality, but we are working to obtain a definitive version. The purpose of the experiments is to test the performance of the proposed system, so we compared the recommendations made by the system with the information provided by the library staff. When the users receive a recommendation, they provide a feedback to the system assessing the relevance of the recommended resource, i.e., they provide their opinions about the recommendation supplied by the system. If they are satisfied with the recommendation, they provide a higher value.

We have designed experiments in which the system is used to recommend research resources that best satisfy the preferences of 10 postgraduate students in Computer Science; all of them completed his/her registration process and evaluated 15 resources, that is, they expressed some of their preferences in the registration process. The resources and the provided evaluations constituted our training dataset. After this, we added 20 new resources that constituted the test dataset. The system filtered the 20 resources and recommends them to the suitable researchers. Then, we compared the recommendations provided by the systems with the recommendations provided by the library staff. With this information, we calculate the precision (ratio of the selected relevant items to the selected items), recall (ratio of the selected relevant items to the relevant items) and F1 (combination metric that gives equal weight to both precision and recall), which are measures widely used to evaluate the quality of the recommendations [16]. The average of precision, recall and F1 metrics are 62,35%, 69,25% and 64,65% respectively, improving the measures obtained with the previous proposal [13]. These values reveal a good performance of the proposed system and therefore a great satisfaction of the users.

IV. Conclusion

We have presented a recommender system to spread indigenous knowledge in a UDL. The system filters the incoming information stream to spread the own academic resources. To improve the services that a UDL provides, the system recommends both specialized and complementary resources, and collaboration possibilities among university membership to form multidisciplinary working groups. So, this system increases the internal social collaboration possibilities in an academic environment.

Acknowledgment

This paper has been developed with the financing of SAINFOWEB project (TIC00602) and FEDER funds in FUZZYLING project (TIN2007-61079), PETRI project (PET2007-0460), and project of Ministry of Public Works (90/07).

References

[1] Bordogna G. and Pasi G. (1993), A fuzzy linguistic approach generalizing boolean information retrieval: A model and its evaluation. Journal of the American Society for Information Science, 44, 70-82.

[2] Burke R. (2007), Hybrid Web Recommender Systems. P. Brusilovsky, A. Kobsa, and W. Nejdl (Eds.): The Adaptive Web, LNCS 4321, 377-408.

[3] Callan J., et. al. (2003), Personalisation and Recommender Systems in Digital Libraries. Joint NSF-EU DELOS Working Group Report.

[4] Chao H. (2002), Assessing the quality of academic libraries on the Web: The development and testing of criteria. Library & Information Science Research, 24, 169-194.

[5] Cornelis C., Lu J., Guo X. and Zhang G. (2007), One-and-only item recommendation with fuzzy logic techniques. Information Sciences, 177(22), 4906-4921.

[6] Hanani U., Shapira B. and Shoval P. (2001), Information Filtering: Overview of Issues, Research and Systems. User Modeling and User-Adapted Interaction, 11, 203-259.

[7] Herrera F. and Herrera-Viedma E. (1997), Aggregation operators for linguistic weighted information. IEEE Trans. on Systems, Man and Cybernetics, Part A: Systems, 27, 646-656.

[8] Herrera F. and Martínez L. (2000), A 2-tuple fuzzy linguistic representation model for computing with words. IEEE Transactions on Fuzzy Systems, 8(6), 746-752.

[9] Herrera F. and Martínez L. (2001), 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.

[10] Korfhage R.R. (1997), Information Storage and Retrieval. New York: Wiley Computer Publishing.

[11] Lekakos G. and Giaglis G.M. (2006), Improving the prediction accuracy of recommendation algorithms: Approaches anchores on human factors. Interacting with Computers, 18, 410-431.

[12] Marchionini G.. Research and Development in Digital Libraries. http://ils.unc.edu/~march/digital_library_R_and_D.html. Last access: April, 2009.

[13] Porcel C., López-Herrera A.G. and Herrera-Viedma E. (2008), A recommender system for research resources based on fuzzy linguistic modeling, Expert Systems with Applications, doi:10.1016/j.eswa.2008.06.038.

[14] Renda M.E. and Straccia U. (2005), A personalized collaborative Digital Library environment: a model and an application. Information Processing and Management, 41, 5-21.

[15] Resnick P. and Varian H.R. (1997), Recommender Systems. Special issue of Comm. of the ACM, 40(3), 56-59.

[16] Sarwar B., Karypis G., Konstan J., Riedl J. (2000), Analysis of recommendation algorithms for e-comerce. Proceedings of ACM E-Commerce 2000 conference, 158-167.

[17] Toong Tjiek L. (2006), Desa Informasi: The role of digital libraries in the preservation and dissemination of indigenous knowledge. The International Information & Library Review, 38, 123-131.

[18] Yager R.R. (2007), Centered OWA operators. Soft Computing, 11, 631-639.

[19] Zadeh L.A. (1975), The concept of a linguistic variable and its applications to approximate reasoning. Part I. Information Sciences, 8, 199-249. Part II, Information Sciences, 8, 301-357. Part III, Information Sciences, 9, 43-80.

Author Index