An Ontology-based Recommender System for Health Information Management

Francisco P. Romero[1], **Mateus Ferreira-Satler**[2], **Jose A. Olivas**[1], **and Jesus Serrano-Guerrero**[1]
[1]Dept. Information Technologies and Systems, University of Castilla La Mancha, Ciudad Real,Spain
[2]Dept. Computer Science / Unviersidade Federal de Juiz de Fora, Juiz de Fora, MG, Brazil

Abstract— *This paper shows how some fuzzy logic techniques applied to a recommender engine can be used in a Electronic Medical Records Repository. A Fuzzy Linguistic model based on three dimensions: intrinsic, contextual, personal is proposed. The contextual and personal dimensions are modeled using domain ontologies and a automatically built fuzzy ontology, respectively. The experiment results indicate that the presented approach is useful and warrants further research in recommending and retrieval information.*

Keywords: Recommender Systems, Fuzzy Ontology, healthcare Information Management, Fuzzy Logic

1. Introduction

Nowadays, new ways of managing and accessing to health care information are continuously appearing. Electronical Medical Records (EMRs) have the potential to make data about health care available to clinicians, researchers and students in different medical contexts and applications. Hundreds of Medical Records are stored and interchanged in Medical Records Repositories. One of the biggest challenges faced by healthcare systems is the growth of information accessible, i.e. the amount of information accessible has grown enormously and as a result health care professionals are currently burdened with more and more data, which unfortunately has not always the adequate levels of quality, making that their work cannot always be as successful as expected. A way of alleviating this situation consists in limiting somehow the number of Medical Records in a repository that are displayed for users. This can be done by means of filtering or recommendation techniques being capable to be adapted to different requirements for each one of the users. Therefore, the need of an efficient and reliable recommendation process is critical in order to provide a more personalized and tailored knowledge to clinicians, researchers and students..

In this paper, we propose the integration of the analysis of different dimensions in a recommendation system. This recommendation engine could be a useful tool to support Health Information Management in a Health Information and Management Systems, in order to improve information filtering and retrieval, as well as their classification. The analyzed dimensions in this proposal are the following: intrinsic (is the EMR complete or accurate?), contextual (is the EMR adequate according to the user context?) and personal (is the EMR adequate according to the user preferences?)

The intrinsic dimension is modeled using measures as completeness and timeliness; the contextual dimension is modeled using a domain ontology, for example, Medical Subject Headings (MeSH) [1] and the personal dimension is modeled by a fuzzy ontology automatically built from the EMR's provided or selected by the user [1]. Within this context, and taking advantage of Fuzzy Logic [2], we addressed the definition, implementation and validation of a process to construct the recommendation system. The main purpose of this work is to provide a method to describe the recommendation process as well as a linguistic model, i.e., we can only describe the whole recommendation system by using natural language.

The remainder of the work is structured as follows: Section 2 describes the background contents of this paper, i.e, recommendation systems and fuzzy ontologies. Section 3 describes the recommendation model and how the user profiles are built. In Section 4 the experiments that have been conducted to validate our proposal are explained and analyzed, and in Section 5 some conclusions and future works are pointed out.

2. Background

This section presents some concepts related to this research and proposed process. It starts describing recommendation systems, their general characteristics and their applications. Then, we discuss how the fuzzy ontologies can be used to represent user's preferences and some works about this topic.

2.1 Recommender Systems

According to [3], recommender systems are defined as systems that produce individualized recommendations as output, or have the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options. [4] enumerate the main characteristics of recommender systems: 1) Can be applied to unstructured data and semi-structured (for example, Web documents or e-mail messages); 2) Based on user profiles, rather than

[1]http://www.geneontology.org/

users expressing their needs through consultation; 3) Manage large amounts of information; 4) Works primarily with information in text mode; 5) Its goal is to eliminate irrelevant information from the input stream. Recommendation systems share similar tasks with information filters such as removing redundant or unwanted information and reducing overload.

In recent years, there are several studies about the applications of the recommender systems in healthcare environments, for example, in [5] a recommendation system is presented with the aim of making health events accessible in personalized way. The proposed system uses a set of "signal definitions", i.e., a predefined structured queries with parameters related to kinds of health threats the users' interest's. The systems provides ratings according to this "signal definitions" to give recommendations. The method to compare the queries and the documets is very straight forward (tf-idf representations and cosine similarity measure) but its results are satisfactory.

On the other hand, there are some proposal that include fuzzy logic in recommendation techniques. For example, Chao et al. [8], propose a recommendation mechanism focused on teachers in a content management system. The main components of these systems are: data pre-processing, association rule mining, associative classification, sequential pattern mining and fuzzy sets. In [6] propose a recommender system multi-granular fuzzy linguistic approach, where the solution alternatives are the digital resources stored into the library, and the criteria to satisfy in the user profiles. This recommender system, allows users to provide preferences on some research resources and from this information are calculate their respective preference vectors on topics of interest. The user profile is completed with user preferences on the collaboration possibilities with other users, with the objective of creating academic communities.

2.2 Ontologies and User Preferences

Ontologies have proved to be successful in handling a machine-processable information representation. They can take the simple form of a taxonomy (i.e., knowledge encoded in a minimal hierarchical structure) or as a vocabulary with standardized machine-interpretable terminology supplemented with natural language definitions. Furthermore, ontology-based user profiles are being widely applied in context representation and application customization so that they meet user requirements.

FCOU [7] is a fuzzy clustering method of ontology-based user profiles construction. The method employs fuzzy clustering techniques combined with optimization techniques and an augmented Lagrangian function to create a fuzzy clustering model for the construction of user profiles. The method allows some information to belong to several user profiles simultaneously with different degrees of accuracy, and makes it possible for a user profile to be represented by one or more ontologies.

In [8], the authors propose an approach that uses locally stored desktop documents to extract terms that will be used in query expansion for web search. Three possible techniques have been investigated. The first one proposes summarizing the entire desktop using term clustering methods. The second technique issues the original web user query on the desktop and extracts expansion keywords from the most significant sentences within the Top-30 documents selected by a scoring function. Similarly, the third technique suggests selecting query expansion keywords from the most dispersive lexical compounds within the Top-30 documents returned to the user's initial web query. Some experiments have also been performed to compare the proposed method with a regular Google web search.

Another approach to represent user preferences is by a domain ontology. This domain user preferences are called "user context" in this work. Lau et al [9] present a text mining methodology for the automatic discovery of fuzzy domain ontology from a collection of on line messages posted to blogs, emails, chat rooms, web pages, and so on. The collection of messages is treated as a textual corpus. The method consists of a document parsing (stop word removal, part-of-speech tagging, and entity tagging and stemming), concept extraction (pattern filtering, text windowing, and mutual information computation), dimensionality reduction (concept pruning and term space reduction), fuzzy relation extraction (computing fuzzy relation membership) and fuzzy taxonomy extraction (taxonomy generation and taxonomy pruning).

On the other hand, [10] show how a fuzzy ontology-based approach can improve semantic documents retrieval. The proposal is illustrated using an information retrieval algorithm based on an object-fuzzy concept network.

3. Fuzzy-Based Recommender System Approach

In this section we present a new fuzzy recommender system based on a matching process developed between user preferences and the EMR representation. For this purpose, we take into account the following parameters that can be assessed in the system: the intrinsic quality of the EMR, the compatibility between the EMR and the user context and the EMR and user preferences.

This system is applied to advise EMR repository users on the best EMR's that could satisfy their information needs. This recommender system also improves the services that a EMR repository provides to users, because it is easier to obtain the knowledge about users and it allows to decrease the time cost to establish the user preferences.

The model is based on a recommendation degree used to deliver the information resources to the fitting users (Eq. 1).

$$\Psi = \sqrt{\Upsilon\left(a,k,c,t\right)} \otimes \left(\chi \oplus \Phi\right) \qquad (1)$$

where Ψ represent the recommendation degree, Υ represents the intrinsic quality degree of the EMR based on the calculation of completeness c, reliability r and timeliness t, χ represents the contextual compatibility between the new EMR and the user context, and Φ represents the compatibility between the user preferences (represented by an ontology automatically built) and the EMR. The square root is used as the linguistic hedge "more or less" according to the explained in [11] about the decision criteria results. $\otimes$ denotes the fuzzy conjunction operator (t-norm). The product as t-norm allows us to obtain the best results on the empirical experiments carried out in this work. $\oplus$ denotes a fuzzy disjunction operator (t-conorm). The use of the algebraic sum as t-conorm allows us to obtain better results than applying classical functions on the empirical experiments carried out in this work.

Only those EMR's which have a recommendation degree, calculated according to the defined model, higher than a pre-established threshold are taken into account ($\Psi \geq \gamma$). This threshold (γ) is pre-established by the user as a part of the configuration process.

The model can be shown as a linguistic model, i.e., we can only describe the whole recommendation system by using natural language, for example, the model can be described as follows: "If the document has more or less quality AND is relevant to the user then it will be processed". The linguistic hedge, the fuzzy operators, the fuzzy rules and the considered intrinsic quality criterion dimensions can be changed to build an efficient recommendation system in any domain context. Each component of this approach is explained in the following subsection.

3.1 Intrinsic Quality

In order to determine if a EMR is useful enough for a user, data contained in it must be analyzed to check if it reaches an adequate level of quality according to certain parameters. Taking into account that the "fitness for use" depends on the task and role of the users who handle both EMR's and their sources, it is necessary to identify a set of dimensions that better represents quality requirements from user requirements specification. Intrinsic Dimensions denote that information has quality in its own right. These dimensions are independent of the userŠs context. They are capturing whether information correctly represents the real world and whether information is logically consistent in itself [12]

For making operative our proposal, three dimensions of a data quality (DQ) model for assessing data of new EMR's have been chosen in order to improve the performance of the proposed recommender system. These dimensions are the following:

- *Accuracy* (a): is the degree of correctness and precision with which information in an information system represents states of the real world [12].
- *Consistency* (k): implies that two or more values do not conflict with each other [13].
- *Completeness* (c): is the degree to which information is not missing [14], i.e. every item of a document is fulfilled and has information
- *Timeliness* (t): is the degree to which information is up-to-date, i. e. received information is adequate for the temporal context in which its topic is set [15].

Whole dimensions are calculated according to the specification shown in [16]. In order to get a summarized measure for a document we must take into account that the perception of quality is both subjective and inaccurate, and consequently it would be appropriate to use a fuzzy operator to measure/determine/assess DQ properly. In this work a Mamdani-style fuzzy system has been employed [17]. Linguistic labels and a set of rules were defined and optimized by a panel of experts. $\Upsilon\left(a,k,c,t\right)$ is a value obtained after a defuzzification process.

3.2 Contextual Compatibility

The Contextual Relevance (χ) of a EMR and a user is computed by using the compatibility between the EMR contents and the ontological definition of the user area of interests. This contextual representation based on ontologies is extracted from the definitions stored in domain ontologies, for example, MeSH [2], UMLS [3]

The degree of representativeness of each keyword is computed using different measures of similarity existing for this thesauri, for example, WordNetSimilarity [18] , UMLSSimilarity [19]. For example, some keywords that define the context "Pulmonary Medicine" in healthcare context using UMLS and UMLSSimilarity are shown in Table 1.

Table 1: Excerpt of "Pulmonary Medicine" Context Definition

Word	Degree
respiratory, lung	1.00
trachea, bronchial	0.50
bronchitis, pneumothorax, chest	0.25
thoracolumbar, abdominal	0.10

In this case the context and the EMR could be considered as fuzzy sets because they consist of words that have a membership degree. Therefore, the compatibility between context and document could be computed by using the generalized Jaccard coefficient as used in [20].

[2] http://www.nlm.nih.gov/pubs/factsheets/mesh.html
[3] http://www.nlm.nih.gov/research/umls/

3.3 Personal Compatibility

In this work, the user preferences will be represented as a fuzzy ontology automatically obtained from a set of EMR's previously selected by the user. A fuzzy ontology, in this context, may be considered as a set of directed graphs where each node represents an item and the edges denote that a term "is related with" other term. The proposal includes several stages of data processing, which were divided into five steps: linguistic pre-processing, term indexing (called pre-ontology), user relevant terms extraction, user ontology generation and user profile update as can be seen in [1]. In this fuzzy ontology, a relatedness degree (RD) is associated with each edge to represent the strength of the "is related with" association. In this way, the relatedness degree (RD) between two terms t_i and t_j is defined as (Eq. 2).

$$RD\left(t_i, t_j\right) = \frac{\sum_{o \in O} f - occur\left(t_i, o\right) \otimes f - occur\left(t_j, o\right)}{\sum_{o \in O} f - occur\left(t_i, o\right)} \tag{2}$$

where o is a EMR, O is the set of EMR's selected from the LOR by the user, $f - occur$ is the function of the relative frequency of a term in a EMR and $\otimes$ denotes a fuzzy conjunction operator.

The compatibility degree between the EMR and the user profile Φ assesses if the interests of the user are expressed in the EMR. The value 0 indicates that users preferences are totally different of the EMR content, i.e., the ontology extracted from the EMR and the user ontology represent different concepts, not necessarily contrary; whereas the value 1 indicates that the EMR contents are included in the users' preferences.

The EMR's ontology is built using a modification of the RD equation, where s is a section (or paragraph according to the kind of document) of the EMR δ (Eq. 3):

$$RD\left(t_i, t_j\right) = \frac{\sum_{s \in \delta} f - occur\left(t_i, s\right) \otimes f - occur\left(t_j, s\right)}{\sum_{s \in \delta} f \widehat{- occur}\left(t_i, s\right)} \tag{3}$$

The process to compute this value consists of the comparison between two ontologies. The comparison method is inspired by a set of ontology similarity measures proposed by [21]. In this case, the vector model is used to represent each concept in the ontology. Let v^i be the corresponding vector that represents a concept c_i, int v^i_j, the value in the position j, will be the RD degree between c_i and c_j when c_j belongs to the set of possible concepts that define a user profile, i.e., $v^i_j = RD(c_i, c_j)$. The detailed procedure is illustrated in Alg. 1, where $jaccard(v_1, v_2)$ represents the jaccard similarity function between two vectors.

Algorithm 1 Compatibility Degree Algorithm

```
O are the ontology corresponding to the new EMR
U are the ontology corresponding to the user
|O| and |U| are the size of each ontology
for c ∈ U do
  if c ∈ O then
    v_O = vector that represents the concept c in the ontology O
    v_U = vector that represents the concept c in the ontology U
    SC = jaccard( v_O, v_U)
  end if
  β = AVERAGE(SC) {β represents the similarity of common concepts}
end for
Φ = (|O ∩ C| ⊗ β) /min (|O|, |U|) {⊗ is a t-norm}
```

4. Experiment

An experiment was carried out to evaluate the performance of the proposed recommender system. In this section, the experiment, the performance measures used and the obtained results are described.

4.1 Experiment Description

The study group was shaped by 10 users from a healthcare organization. Most of them had experience in e-health technology as a researcher or as a physician. Each user has selected one of the following areas: traumatology (TRA), oftalmology (OFT), otolaryngology (OTO), surgery (SUR) and urology (URO). These areas and their ontological representation based on MeSH are the base for the user context definition. On the other hand, each user has selected 11 Medical Records as relevant in the Healthcare Information System. Once the selection process was finished, the user context and user preferences ontological definition were created for each participant, applying the methodology and algorithms described previously.

4.2 Experiment Results

In order to evaluate the feasibility of our approach, we compared the recommendations made by the system and the preferences for each new user of each new EMR. The contingency table (Table 2) used for this purpose is similar as the explained by [22] and by [6].

Table 2: Contingency Table

	Selected	Not Selected	Total
Relevant	Nrs	Nrn	Nr
Irrelevant	Nis	Nin	Ni
Total	Ns	Nn	N

Precision, recall and F-measure are measures useful to evaluate the quality of the recommendations [23] . Here, precision measures the probability of a selected item being relevant, recall represents the probability of a relevant items being selected and F-measure is the harmonic mean between precision and recall (Equations 4, 5, 6).

$$P = \frac{N_{rs}}{N_s} \quad (4)$$

$$R = \frac{N_{rs}}{N_r} \quad (5)$$

$$F = \frac{(2 * R * P)}{(R + P)} \quad (6)$$

We considered a test data set with 100 EMR's of different areas. The system filtered these EMR's and recommends them to the suitable users. Then, we compared the recommendations provided by the systems with the recommendations provided by the users. After this comparison the corresponding precision, recall and F-measure are obtained. The results of this process are shown in Table 3 (the values are in the range [0,1]).

Table 3: Results.

User	Precision	Recall	F-measure
TRA	0.98	0.81	0.89
OFT	0.95	0.91	0.93
OTO	0.94	0.89	0.91
SUR	0.86	0.91	0.88
URO	0.92	0.85	0.88

The average of precision, recall and F-measure are 0.93, 0.87% and 0.90%, respectively. These values reveal a good performance of the proposed system if compared with those obtained by using non-ontological approaches (0.70, 0.61 y 0.65, respectively). We have achieved a substantial improvement in precision and recall values, which means that the proposed system is flexible enough to provide good recommendations using our approach.

5. Conclusions and Future Work

In this paper, a recommendation method based on the fuzzy representation of user preferences was proposed. This approach has been applied to provide recommendations about new Medical Records that could be interesting for a user. This is an efficient solution to minimize the problem of access relevant information in Medical Records Repositories. The proposal combines the analysis of intrinsic and conceptual features for making decisions about recommendation. User preferences are represented by domain ontologies for user context and automatically built fuzzy ontologies. This combination allows us to make recommendations based on a richer description of the user preferences.

An experiment has been carried out in order to determine if the recommended EMR's are useful and interesting for the users. Experimental results show that the proposed system is reasonably effective in terms of precision and recall.

Further research is directed towards the task of improving the user profile quality considering the information provided by the user as feedback, and the application of some techniques of collaborative filtering. Moreover, more detailed evaluation experiments also will be necessary.

References

[1] M. Ferreira-Satler, F. P. Romero, V. H. MenÃl'ndez, A. Zapata, and M. E. Prieto, "A fuzzy ontology approach to represent user profiles in e-learning environments," in *FUZZ-IEEE 2010 IEEE International Conference on Fuzzy Systems - WCCI 2010 IEEE World Congress on Computational Intelligence*, Barcelona (Spain), 2010, pp. 161–168.

[2] L. A. Zadeh, "Fuzzy logic and approximate reasoning," *Synthese*, vol. 30, no. 3, pp. 407–428, 1975.

[3] R. Burke, "Hybrid recommender systems: Survey and experiments," *User Modeling and User-Adapted Interaction*, vol. 12, pp. 331–370, November 2002.

[4] U. Hanani, B. Shapira, and P. Shoval, "Information filtering: Overview of issues, research and systems," *User Modeling and User-Adapted Interaction*, vol. 11, pp. 203–259, August 2001.

[5] R. Lage, F. Durao, P. Dolog, and A. Stewart, "Applicability of Recommender Systems to Medical Surveillance Systems," in *Proceedings of the CIKM MedEX 2011*. ACM, 2001.

[6] C. Porcel and E. Herrera-Viedma, "Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries," *Knowledge Based Systems*, vol. 23, pp. 32Ű–39, 2010.

[7] L. Han and G. Chen, "A fuzzy clustering method of construction of ontology-based user profiles," *Advances in Engineering Software*, vol. 40, no. 7, pp. 535–540, 2009.

[8] P.-A. Chirita, C. S. Firan, and W. Nejdl, "Summarizing local context to personalize global web search," in *Proceedings of the 15th ACM international conference on Information and knowledge management*, ser. CIKM '06. New York, NY, USA: ACM, 2006, pp. 287–296. [Online]. Available: http://doi.acm.org/10.1145/1183614.1183658

[9] R. Y. Lau, D. Song, Y. Li, T. C. Cheung, and J.-X. Hao, "Toward a fuzzy domain ontology extraction method for adaptive e-learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 6, pp. 800–813, 2009.

[10] S. Calegari and E. Sanchez, "Object-fuzzy concept network: An enrichment of ontologies in semantic information retrieval," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 13, pp. 1532–2890, 2008.

[11] L. Zadeh, "A fuzzy set theoretic interpretation of linguistic hedges," *Journal of Cybernetics*, vol. 2, no. 3, pp. 4–34, 1972.

[12] Y. Wand and R. Wang, "Data quality in context," *Communications of the ACM*, vol. 39, no. 11, pp. 86–95, 1996.

[13] M. Mecella, M. Scannapieco, A. Virgillito, R. Baldoni, T. Catarci, and C. Batini, "Managing data quality in cooperative information systems," in *Proceedings of the Confederated International Conferences DOA, CoopIS and ODBASE*, 2002, pp. 486–502.

[14] L. Pipino, Y. Lee, and R. Wang, "Data quality assessment," *Communications of the ACM*, vol. 45, no. 4, pp. 211–218, 2002.

[15] B. Kahn, D. Strong, and R. Wang, "Quality benchmarks: Product and service performance," *Communications of the ACM*, vol. 45, no. 4, pp. 184–192, 2002.

[16] I. Caballero, E. Verbo, C. Calero, and M. Piattini, "A data quality measurement information model based on iso/iec 15939," in *Proceedings of the 12th International Conference on Information Quality*, 2007, pp. 393–408.

[17] E. Mamdani and S. Assilian, "An experiment in linguistic synthesis of fuzzy controllers," *International Journal Man-Machine Studies*, vol. 7, pp. 1–13, 1975.

[18] T. Pedersen, S. Patwardhan, and J. Michelizzi, “Wordnet::similarity: measuring the relatedness of concepts,” in *Demonstration Papers at HLT-NAACL 2004*, ser. HLT-NAACL–Demonstrations ’04. Stroudsburg, PA, USA: Association for Computational Linguistics, 2004, pp. 38–41.

[19] B. T. McInnes, T. Pedersen, and S. V. S. Pakhomov, “Umls-interface and umls-similarity : open source software for measuring paths and semantic similarity.” *AMIA Annu Symp Proc*, vol. 2009, pp. 431–5, 2009.

[20] V. V. Cross and T. A. Sudkamp, *Similarity and compatibility in fuzzy set theory: assessment and applications*. Heidelberg, Germany, Germany: Physica-Verlag GmbH, 2002.

[21] A. Maedche and S. Staab, “Measuring similarity between ontologies,” in *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*. Germany: Springer-Verlag, 2002, pp. 251–263.

[22] C. Porcel, J. Moreno, and E. Herrera-Viedma, “A multi-disciplinar recommender system to advice research resources in university digital libraries,” *Expert Systems with Applications*, vol. 36, no. 10, pp. 12 520–Ű12 528, 2009.

[23] Y. Cao and Y. Li, “An intelligent fuzzy-based recommendation system for consumer electronic products,” *Expert Systems with Applications*, vol. 33, pp. 230–240, 2007.