INTERNATIONAL JOURNAL OF SCIENTIFIC PROGRESS AND RESEARCH (IJSPR) ISSN: 2349-4689
Issue 104, Volume 36, Number 02, 2017

A Relational based Fuzzy clustering Ontology Model for Personalizing Multimedia Content

E. Maruthavani[1] , Dr. R. Lokesh Kumar[2]

[1]Department of Computer Science ,[2]Department of Information Technology

[1]Karpagam Institute of Technology,[2]Bannari Amman Institute of Technology

Coimbatore Tamil Nadu[1],Sathyamangalam, Erode District, Tamil[2] Nadu

Abstract - This paper provides a solution for the problem of adopting Ontologies in order to model the users and multimedia documents and to provide personalized search functionalities. Personalization is a difficult problem related to fields and applications ranging from information retrieval to multimedia content manipulation. Challenge is greater, when trying to combine traditional personalization techniques with novel knowledge representations like ontologies. In this paper we proposed a novel contextual knowledge modeling, based on Relational based Fuzzy clustering Ontology and exploits it in user profiling representation, extraction and use. The personalized results of the application of this methodology are then ranked accordingly. The performance of the proposed techniques is demonstrated through preliminary experimental results derived from a real-life data set.

Key words: Data mining, Multimedia Content analysis, Fuzzy clustering algorithm..

I. INTRODUCTION

Personalization forms an interesting asset used in the field of information retrieval (IR), suffering though from information overload, since IR usually tends to select documents, many of which are barely related to the user's wish [3]. Personalization uses information stored in user profiles, additionally to the user's current search or query, to estimate the users' wishes and select the set of relevant documents. In general no common distinction exists between different profiling algorithms. Handling of personalized information may be decomposed into three tasks tackled within this work: i) design of appropriate knowledge representation, ii) design, development and application of profiling algorithm and iii) presentation and ranking of results. Successful extraction of user profiles, using ontological knowledge [5] is still considered an open issue, because it is difficult to apply in multimedia environments. In order to interpret user queries, we consider contextual information available from prior sets of user actions. We refer to this information as contextual knowledge or just context. This work deals with exploiting ontology-based contextual information, specifically aimed towards its use in personalization tasks. The structure of the paper is as follows: in section 2, we present our knowledge infrastructure, introducing the notion of fuzzy relations in ontologies. In section 3 we explain our user profiling algorithm and we extract user preferences based on usage history, fuzzy hierarchical clustering and ontological knowledge. In section 4, we rank the retrieved results, while in section 5 we provide early experimental results and in section 6 we present our conclusions.

II. SYSTEM MODEL

It is very difficult to create generic personalization solutions, without having a large knowledge at hand. Enriching this knowledge with contextual information results in a useful and representative set of user preferences. We define this set as the contextualized set of user preferences. We restrict the notion of context in this work to the notion of ontological taxonomic context, defined on top of a "fuzzified" version of traditional ontologies. This context implements the necessary knowledge model and is strongly related to the notion of ontologies: ontology can be seen as an attempt for modelling real-world (i.e. fuzzy) concepts and context determines the intended meaning of each concept, i.e. a concept used in different context may have different meanings. In general, ontologies may be described as follows:

$$O = \{C,\{Rci,cj\}\},\ i, j= 1..n,$$

$$i \neq j,\ Rci,c\ C\rightarrow\{0.1\}.i = 1...n \qquad (1)$$

where O is an ontology, C the set of concepts it describes and ci,cj R the semantic relation amongst two concepts ci,cj Є C.

We define ontological context in the means of fuzzy taxonomic ontological relations. Although ontologies may contain any type of relations, only taxonomic relations are of our interest, since the use of such relations is necessary for the determination of the document's context [1]. Additionally, accurate representation of real-life information governed by uncertainty is only possible using fuzzy relations [6]. Consequently, we introduce a "fuzzified" definition of ontology:

F(O) = {C,{rci,cj}}, i, j = 1...n, i ≠j , F(Rci,cj) = rcjcj : C x C → [0,1] (2)

where F(O) forms a "fuzzified" ontology, C is the set of all possible concepts it describes and F(rci,cj) = ci,cj ci,cj F = denotes a fuzzy relation amongst two concepts. Unfortunately, current ontology languages (OWL, DL and plain RDF) are not powerful enough to model such ontology. Thus, we decided to enhance RDF, being a standardized, graph-modeled language, with novel characteristics like reification [7]. The proposed model is a graph, in which every node represents a concept and each edge between two nodes forms a contextual relation between the concepts. Additionally, each edge has an associated degree of confidence, implementing fuzziness. Describing the additional degree of confidence is carried out using "manual" reification, i.e. making a statement about the statement, which contains the degree information. In the next example concept holiday is related to concept sky with a fuzzy relation is RelatedTo and a degree of confidence equal to 0.75. Supposing an RDF namespace dom, we have: Following the above principles our knowledge model is able to utilize any type of real-life fuzzy relations between concepts. For personalization purposes, we utilize two of them, the specialization relation, Sp, and the part relation, P. Relation Sp is a fuzzy taxonomic relation on the set of concepts and Sp(x,y)>0 means that the meaning of x "includes" the meaning of y. Relation P is also a fuzzy taxonomic relation on the set of concepts and P(x,y)>0 means that y is a part of x. Combining the above relations, we construct a fuzzy taxonomic relation which is suitable for the handling of user preferences. T implies that if the user query contains x, then T(x,y) indicates that documents that contain y will also be of interest. The transitive closure Tr is necessary, since the union of transitive relations is not necessarily transitive [6].

III. PROPOSED METHODOLOGY

In this section Most personalized retrieval techniques (e.g. collaborative filtering) keep and process long records of accessed documents by each user, in order to infer potential preferences for new documents (e.g. by finding similarities between documents, or between users). The data handled by these techniques have been rather low-level and simple: document IDs, text keywords and topic categories at most the recent proposals and achievements towards the enrichment of multimedia content by formal, ontology-based, semantic descriptions open new opportunities for improvement in the personalisation field from a new, richer representational level. We see the introduction of ontology-based technology in the area of personalisation as a promising research direction Ontology's enable the formalization of user preferences in a common underlying, interoperable representation, whereby user interests can be matched to content meaning at a higher level, suitable for conceptual reasoning. An ontology-based representation is richer, more precise, and less ambiguous than a keyword-based model. It provides an adequate grounding for the representation of course to fine-grained user interests (e.g. interest for individual items such as a sports team, an actor, a stock value) in a hierarchical way, and can be a key enabler to deal with the subtleties of user preferences.

For instance, a personalisation framework may share domain ontology with a knowledge-based content analysis tool that extracts semantic metadata from audio/visual content, conforming to the ontology [10]. On this basis, it is easier to build algorithms that match preference to content, through the common domain ontology. In an ontology-based approach, semantic user preferences may be represented as a vector of weights (numbers from -1 to 1), representing the intensity of the user interest for each concept, being negative values indicative of a dislike for that concept Similarly, content is described by a set of weighted concepts (values from 0 to 1, indicating the intensity of relation between the content and the concept) in such a way that users can be related to the content units that make up the search space through the ontology layer (see Figure 1).

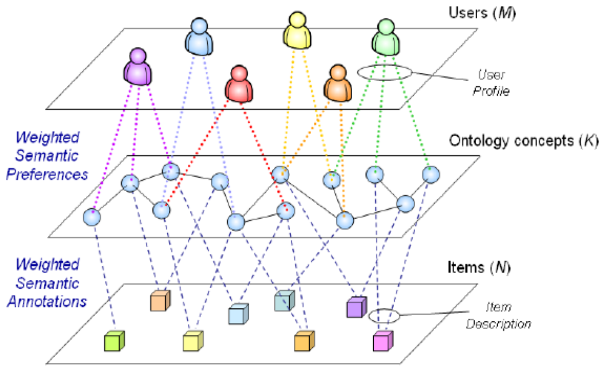


Figure 1. Association of users and content

If a content analysis tool identifies, for instance, a cat in a picture, and the user is known to like cats, the personalisation module can make predictions on the potential user interest for the picture by comparing the metadata of the picture, and the preferred concepts in the user profile. Furthermore, ontology standards backed by international consortiums (such as the W3C), and the corresponding available processing tools, support inference mechanisms that can be used to further enhance personalization, through the middle ontology layer, so that, for instance, a user interested in animals (superclass of cat) is also recommended pictures of cats. Inversely, a user interested in lizards, snakes, and chameleons can be inferred to be interested in reptiles with a certain confidence [11]. Also, a user keen of Sicily can be supposed to like Palermo, through the transitive located. In relation, assuming that this relation has been seen as relevant for inferring previous underlying user's interests.

We illustrated the modeling of contextual dependence between concepts and relations using an RDF-based representation and a fuzzy taxonomic relation T. We continue with the presentation, extraction and use of user preferences. In compliance with the fuzzy notation presented in [6], we adopt the following formal representation of user preferences

$$\mathrm{P} : P = \{U^+, U^-\} \qquad (3)$$

where U^+, U^- refer to the set of positive and negative preferences, respectively. Following the sum notation for fuzzy sets [6] U^+ and $U^-$are defined as follows:

$$U^+ = \{U_j^+\},$$
$$j \in N_k, U^- = \sum c_i / p_i^-, \; i \in N_n, n = |C| \qquad (4)$$

k is the count of distinct positive preferences contained in the user profile, p_{ij}^+ is the degree of participation of concept c_i in U_j^+, p_i^- is the degree of participation of concept c_i in U^- and $U_j^+ = \sum c_i / p_{ij}^+$, $i \in N_n, j \in N_k$, n=|C|.

This definition allows participation of a single concept in multiple preferences and to different degrees. As all relations existing in the ontology are defined on the set C of concepts, we define user preferences on the same set, i.e. user preferences are also concepts: $P \subseteq C$.

${}^{0+}d = \{c_1, ..., c_n\} \subseteq C$ and preferences are mined by applying clustering algorithms on it. Most clustering methods belong to either partitioning or hierarchical, however the former require the number of clusters as input and thus are inapplicable [6]. The proposed approach may be decomposed into the following steps:

- Perform a fuzzy clustering of concepts in order to determine the count of distinct preferences that a history document is related to, according to the following steps:

 1. Turn each available concept into a singleton, i.e. into a cluster k of its own.
 2. For each pair of clusters $k1$, $k2$ calculate their distance $d(k1,k2)$.
 3. Merge the pair of clusters that have the smallest distance $d(k1,k2)$.
 4. Continue at step 2, unless termination criteria are met; termination criterion most commonly used is a threshold for the value of $d(k1,k2)$.

- Find the user preferences that are related to each cluster.
- Aggregate the findings for each cluster to acquire an overall result for each d .

The key element of the above algorithm is the ability to define a unique distance among any pair of clusters, given the input space and the clustering features. We propose the following distance estimation:

$$d(k_1, k_2) = \sum_{i \in F} \sqrt[K]{\frac{\sum_{x \in k_1, y \in k_2} r_i(x_i, y_i)^\mu}{|k_1||k_2|}} \qquad (5)$$

where r_i, $i \in F$ is the metric that compares the i-th feature, F the overall count of features, k_1 the cardinality of cluster $k1$ and μ a constant. Obviously, μ=1 approaches the mean value and μ=2 yields the Euclidean distance. Still, this clustering method creates only crisp clusters and does not allow for overlapping among the detected clusters. In real life, a concept is related to a preference with a degree in [0,1] and is also related to more than one distinct preference, making "fuzzification" of the partitioning necessary. We construct a fuzzy classifier, in the means of a function C_k:$C \rightarrow$[0,1] that measures the degree of correlation of a concept c with cluster k. Then, we expand the detected crisp partitions to include more concepts. Partition k is replaced by cluster k^{fuzzy}, following again the sum notation for fuzzy clusters:

$$k^{fuzzy} = \sum_{c \in {}^{0+}d} c / C_K(c) \qquad (6)$$

Obviously $k^{fuzzy} \supseteq k$. The set of preferences that correspond to a history document is the set of preferences that belong to any of the detected fuzzy clusters of concepts.

Once user profiles are obtained by extracting user preferences from the semantically analyzed usage history, our approach to preference-based content retrieval [2][8] is based on the definition of a matching algorithm that provides a personal relevance measure prm(x,u) of a document x for a user u. The procedure for matching a content object to the user preferences is based on a cosine function for vector similarity computation. For this purpose, we build a vector based representation of user preferences from the fuzzy sets defined in the previous section. The user preference vector p is defined by $P_i = \sum_j P_{ij}^{+} - P_i^{-}$, for each concept c_i. Then the expected degree of preference of user *u* for a document *x* is computed by:

$$prm(x.u) = \cos(x,u) = \frac{x \cdot u}{|x||u|} \qquad (7)$$

where *x* stands for the vector of annotations of the document, so that x_i is the weight of the annotation of the document by each concept *ci* in the user profile. The measure above can be used as is to rank documents, based only on user preferences, as well as to personalize an explicit user query *q*, when combined with a query based score without personalization sim(*x*,*q*), to produce a combined ranking [4]. In our approach, we adopted the *combSUM* model, by which the two rankings are merged by a linear combination of the relevance scores:

$$\text{score}\ (x,q,u) = \lambda \cdot \text{prm}\ (x,u) + (1 - \lambda)\ \text{sim}\ (x,q),$$

$$\text{where } \lambda \in [0,1] \qquad (8)$$

The choice of the λ coefficient in (8) provides a way to gauge the degree of personalization, ranging from λ=0 producing no personalization at all, to λ=1, where the query is ignored and results are ranked only on the basis of global user interests.

IV. SIMULATION/EXPERIMENTAL RESULTS

In order to test the proposed techniques, we have conducted early experiments, which we describe next. The purpose of the experiments is to test the consistency of the preference learning by using them to personalize the output of a visual search engine on a corpus of images [9]. The test measures the overall effectiveness of the preference learning approach described in section 3, followed by the personalized ranking step described in section 4. The dataset set up for the experiments included:

- A sample "fuzzified" ontology, in RDF format, containing more than 1000 concepts. Relationships between concepts were defined by relation *T*.
- A set of 150 documents for usage tracking and preference learning, consisting of images with manual free-text annotations. A simple semantics extraction method was used to produce ontology-based metadata vectors from the textual annotations.
- A second set of 100 images for querying and retrieval with similar characteristics but separated from the first one, in order to show non-trivial results, i.e. the system being able to predict user preferences for images that were not available at the time the user's interest for specific documents was monitored.

Based on this corpus, the experiment consisted of the following steps:

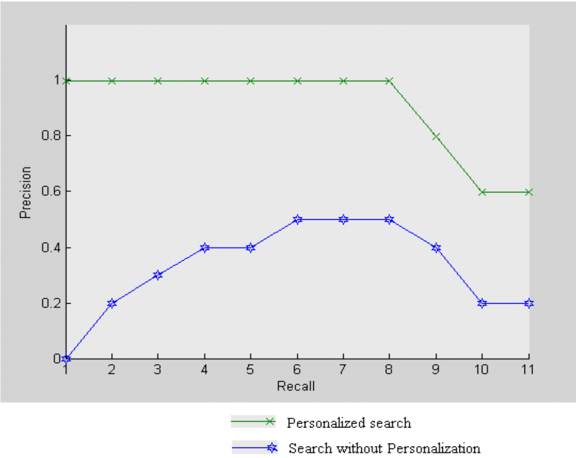
1. A subject selected 9 images from the first set of images displaying works of art, which had annotations by concepts such as *chapel*, *fresco*, *tower*, *fabric*, *Padua* and others. The concept vectors attached to the selected images are automatically stored by the system as history documents.

2. The preferences extraction algorithm is applied and for the sake of simplicity the fuzzy hierarchical clustering method identifies only positive user preferences, i.e. $U+$, yielding: 1 $U+ = health/\ 0.91 + leaders/\ 0.88 + art\ /\ 0.90$.

3. The subject is asked to provide preference-biased ground truth data for a "search for similar" query on the second document collection, the query consisting of a photo showing a horse. The user classifies each picture in the collection as relevant or non-relevant for the query, according to his own biased judgment.

4. The personalized search algorithm is run on the same query and collection, using an image-based search engine, the output of which is re-ranked by preference as described in section 4.

Figure 2 shows the performance of the ranked search results returned in step 4, compared to the results obtained without personalization. The poor precision of the search without personalization at the lowest recall levels is due to the fact that the image-based retrieval algorithm returns initially irrelevant results. Overall experiments show that the proposed ontology-based personalization is particularly helpful in difficult multimedia retrieval tasks.



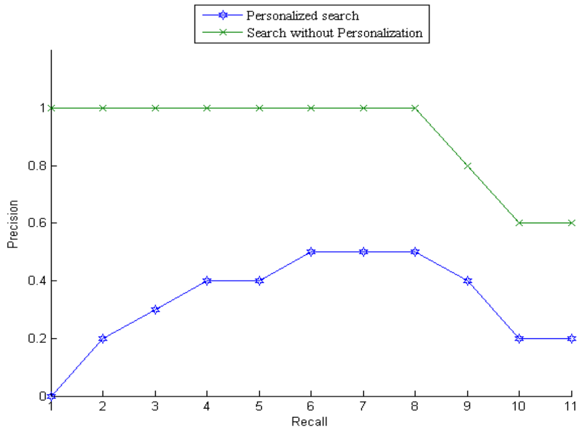


Figure 2. Recall/precision curve of the search with and without personalization

V. CONCLUSION

We have implemented and tested a personalized retrieval and ranking framework that can be exploited towards the development of more efficient personalization environments. Its core contribution has been the provision of personalized access to multimedia content. We based our efforts on a novel “fuzzified” ontological knowledge model, utilizing contextual information and fuzzy taxonomic relations, towards representing, extracting and using of user preferences. Early results on personalized content retrieval are very promising and form an interesting perspective.

REFERENCES

[1] Akrivas G., Stamou G., Kollias S., “Semantic Association of Multimedia Document Descriptions through Fuzzy Relational Algebra and Fuzzy Reasoning”, IEEE Trans. on Systems, Man, and Cybernetics, March 2009.

[2] Castells P., Fernández M., Vallet D., Mylonas P. and Avrithis Y., “Self-Tuning Personalized Information Retrieval in an Ontology-Based Framework”, LNCS vol. 3762, pp. 977-986, 2010.

[3] Chen P.M., Kuo F.C., “An information retrieval system based on a user profile”, Journal of Systems and Software 54, 3-8, 2014.

[4] Dwork C., Kumar R., Naor M. and Sivakumar D., “Rank Aggregation Methods for the Web”, Proc. of the 10th Intl. World Wide Web Conference (WWW10), Hong Kong, 2011.

[5] Xiaohui Tao, Yuefeng Li, and Ning Zhong, “A Personalized Ontology Model for Web Information Gathering”, IEEE transactions on knowledge and data engineering, vol. 23, no. 4, April 2014.

[6] O. Etzioni and D. Weld, “A softbot-based interface,” in Readings in Agents. San Mateo, CA: Morgan Kauffman, 2007, pp. 71–81.

[7] Dwi H. Widyantoro, John Yen," Using Fuzzy Ontology for Query Refinement in a Personalized Abstract Search Engine",IFSA World Congress and 20th NAFIPS International Conference, 2011.

[8] K. M. Sim, “Toward holistic web-based information retrieval: An agent based approach,” in Proc. 2003 IEEE/WIC Int. Conf. of Web Intelligence, pp. 39–46.

[9] M. Anandarajan, “Profiling web usage in the workplace: A behavior based artificial intelligence approach,” J. Manage. Inf. Syst., vol. 19, no. 1, pp. 243–266, 2002.

[10] Dwi H. Widyantoro, John Yen , A Fuzzy Ontology-based Abstract Search Engine and Its User Studies, The 10th IEEE International Conference on Fuzzy Systems December 2010.

[11] M.Vargas-Vera et al., “Knowledge Extraction Using an Ontology-Based Annotation Tool,” Workshop on Knowledge Markup & Semantic Annotation,ACM Press, New York, 2012, pp. 5–12 .

[12] R.Lokeshkumar, T.Gopalakrishnan, P.Sengottuvelan, “A Relational Based Fuzzy Clustering To Mine User Profiles for Web Directory Personalization”, International Journal of Engineering and Advanced Technology (IJEAT), 2015.