

# Fuzzy Induction in Dynamic User Profiling for Information Filtering

Rafal A. Angryk and Costin Barbu

Electrical Engineering and Computer Science Department  
Tulane University, New Orleans, LA, 70118  
{angryk, barbu}@eecs.tulane.edu

In this paper we investigate the role of the user profile in information filtering and we introduce a novel algorithm for learning the user profile based on user's initial profile and on a queries' interpretation using fuzzy generalization (Angryk and Petry 2003). Thousands of documents are usually retrieved by search engines for a given query during an information search on WWW. One way to prune irrelevant documents is to take advantage of the user's implicit interests to filter the documents returned by the search engine, or to reformulate the query based on these interests.

One of the common representations of the documents (and queries) in information retrieval is based on the vector hyperspace model (Salton and McGill 1993). We are using the expanded version of Salton's vector space model introduced by Barbu and Simina (2003) for its effectiveness in computing the dynamics of the user profile. In contrast with the classical vector space model, this recent model, time-words vector hyperspace, has an additional temporal dimension. The coordinates of the documents and queries vectors are calculated using the traditional TF-IDF technique. Only the queries have a temporal dimension (current interest weight) which is set to a preset positive initial value that decays in time, suggesting that some specific user interests could decrease as time goes on. The user's categories of interest are computed based on a novel approach using Fuzzy Concept Hierarchies (FCH) extracted from the WordNet<sup>1</sup> ontology. An FCH is built for each of the query's keywords, using their hypernym chains provided by WordNet. A subunitary membership degree (weight) is assigned for each of the edges of the hierarchy via a bottom-up approach, from the lowest level concept (initial keyword) to the more abstract one. The assignment of the weights is performed according to the following two conditions:

1) The sum of weights assigned to the links outgoing from the original keyword has to be equal to unity.

2) The sum of weights from all links incoming to each node in the FCH has to be equal to the sum of weights in all outgoing edges.

The user category of interest is found by intersecting keywords generalization models for each query. For each of the generalization paths ending at the same abstract concept we calculate the Average Generalization Path Value (AGPV) as the average of the products of the paths' weights, and choose the abstract with the largest value as the category of interest. An example of extracting the user category of interest given the  $Query = \{Shrimp, Chardonnay\}$  is illustrated in Figure 1. The Average Generalization Path Value was computed for both abstract concepts: *Food* and *Organism*.  $AGPV_{Food}$  was obtained by taking into consideration the path weights' product on the chains: *Shrimp - Food* and *Chardonnay - Food*.  $AGPV_{Organism}$  was calculated employing the path weights' product on the chains *Shrimp - Organism* and *Chardonnay - Organism*. The abstract Food was selected as the category of interest since  $AGPV_{Food} > AGPV_{Organism}$ .

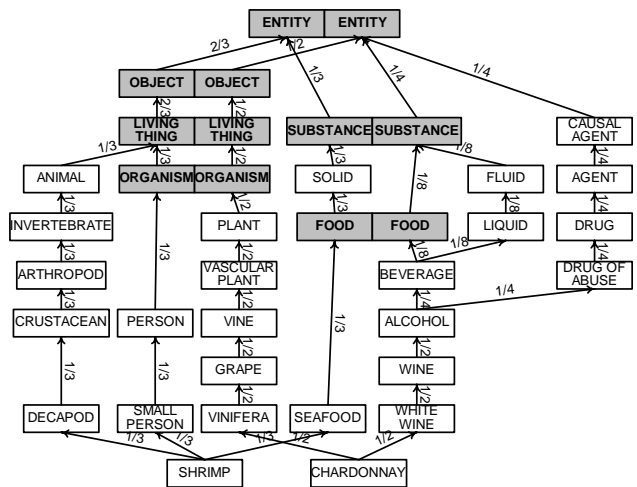


Figure 1. Intersection of Fuzzy Concept Hierarchies for  $Query = \{Shrimp, Chardonnay\}$

The rate of interest change  $\alpha_i$  computed as the cosine similarity between two sequential query feature vectors  $Q_i$

<sup>1</sup> WordNet is an on-line lexical reference system available at: <http://www.cogsci.princeton.edu/cgi-bin/webwn1.7.1>  
Copyright © 2004, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

and  $Q_{i-1}$  proved to be an effective measure of user's profile dynamics (Barbu and Simina 2003). We have modeled the user behavior by developing an adaptive algorithm for dynamic learning of the user profile (Figure 2). The scheme proposed in this work keeps track of both the user's Recent and Long-Term Profiles. The input of the algorithm is a query ( $Q_i$ ) and the output is one or more triplets (Category  $C_i$ , Current Interest Weight  $W_i$ , Rate of Interest Change  $\alpha_i$ ).

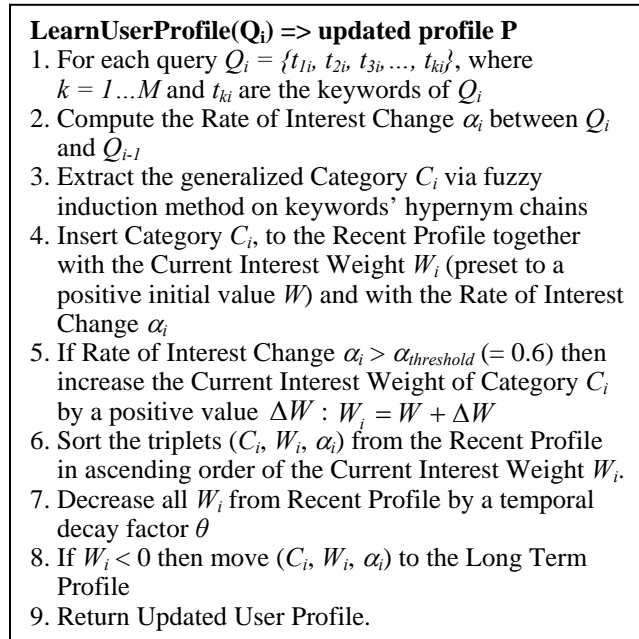


Figure 2. User Profile Learning algorithm

Experimental results are presented in Table 1. The filtering process is performed based on the Relevance Score determined as the cosine similarity between the User Profile feature vector and the feature vectors of the documents retrieved by a classical search engine (i.e. Yahoo).

|  | No Profile Filtering | Total Profile Filtering | Recent Profile Filtering |
|--|----------------------|-------------------------|--------------------------|
| Total Number Documents Retrieved       | 31,356,000           | 256                     | 11,756                   |
| Accuracy in Top 10 Documents Retrieved | 36.33 %              | 41.25 %                 | 70.66%                   |
| Accuracy in Top 20 Documents Retrieved | 28 %                 | 31 %                    | 65 %                     |

Table 1. Experimental results (on average) for various filtering methods on a set of queries

Our preliminary results demonstrate that the information filtering effectiveness is significantly improved by using the User Recent Profile as opposed to existing approaches that consider the total profile (Balabanovic 1997, Chen and Sycara 1998, Widyantoro et al. 1999). We plan to extend our work by integrating the background knowledge available in WordNet with a more specialized ontology in order to accommodate queries with brand names.

## References

- Angryk, R., and Petry, F. 2003. Data Mining Fuzzy Databases Using Attribute-Oriented Generalization. In Proceedings of the Third IEEE International Conference on Data Mining, 8-15. Melbourne, FL.: IEEE Computer Society.
- Angryk, R., and Petry, F. 2003. Consistent Fuzzy Concept Hierarchies for Attribute Generalization. In *Proceedings of the IASTED International Conference on Information and Knowledge Sharing*, 158-163. Scottsdale, AZ.: ACTA Press.
- Barbu, C., and Simina, M. 2003. Information Filtering Using the Dynamics of the User Profile. In *Proceedings of the 16th International Florida Artificial Intelligence Research Society Conference*, 245-249, St. Augustine, FL.: AAAI Press.
- Balabanovic, M. 1997. An Adaptive Web Page Recommendation Service. In *Proceedings of the First International Conference on Autonomous Agents*. 378 – 385, New York. N.Y.: ACM Press.
- Chen, L., and Sycara, K. 1998. WebMate: Personal Agent for Browsing and Searching. In *Proceedings of the Second International Conference on Autonomous Agents*, 132-139. New York. N.Y.: ACM Press.
- Cubero J. C., Medina J.M., Pons O. and Vila M.A. 1999. Data Summarization in Relational Databases Through Fuzzy Dependencies, *Information Sciences*, 121(3-4): 233-270.
- Lee D.H., and Kim M.H. 1997. Database Summarization Using Fuzzy ISA Hierarchies. *IEEE Transactions On Systems, Man, and Cybernetics* - part B, 27(1): 68-78.
- Lee M. 2001. Mining Generalized Fuzzy Quantitative Association Rules with Fuzzy Generalization Hierarchies. In *Proceedings of the Joint 9th IFSA World Congress and 20th NAFIPS International Conference*, 2977 - 2982. Vancouver, Canada: IEEE.
- Mock, K. J. 1996. Hybrid-Hill-Climbing and Knowledge-based Techniques for Intelligent News Filtering. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence and the Eighth Innovative Applications of Artificial Intelligence Conference*. 48-53. Menlo Park, California: AAAI Press
- Salton, G., and McGill, M. J. 1993. *Introduction to Modern Information Retrieval*. New York. N. Y.: McGraw-Hill.
- Voorhees, E.M. 1993. Using WordNet to Disambiguate Word Senses for Text Retrieval. In *Proceedings of the 16th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, 171-180. Pittsburgh, PA: ACM Press.
- Widyantoro, D. H., Yin J., El Nasr, M., S., Yang, L., Zacchi, A. and Yen J. 1999. Alipes: A Swift Messenger in Cyberspace. In *1999 AAAI Proceedings of the Spring Symposium on Intelligent Agents in Cyberspace*, 62-67. Palo Alto, CA: AAAI Press.