Fuzzy Logic in Web Intelligence

Masoud Nikravesh[2] Ali Jaoua[3] Chris Tseng[1,4]

[1]CS department, San Jose State University, One Washington Square, San Jose, CA 95192, USA
[2]University of California, Berkeley, CA 94720, USA
[3]University of Qatar, College of Engineering, Department of Engineering and Computer Science
[4]tseng@cs.sjsu.edu (corresponding author), Tel: 1-408-924-7255, Fax: 1-717-754-6112

Keywords: Semantic Web, Fuzzy Query, Fuzzy Search, PNL, NeuSearch, Z-Compact, BISC-DSS World Wide Web

1. Fuzzy Logic-Based Web Intelligence

Search engines have become the most heavily-used online services, with millions of searches performed each day. Their popularity is due, in part, to their ease of use. It is important to note that while the Semantic Web is dissimilar in many ways from the World Wide Web, the Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries through the World Wide Web. In this paper, we would like to go beyond the traditional semantic web which has been defined mostly as a mesh or distributed databases within the World Wide Web. For this reason, our view is that "Before one can use the power of semantic web, the relevant information has to be mined through the search mechanism and logical reasoning". The central tasks for the most
of the search engines can be summarize as

1) query or user information request- do what I mean and not what I say!,
2) model for the Internet, Web representation-web page collection, documents, text, images, music, etc, and
3) ranking or matching function-degree of relevance, recall, precision, similarity, etc.

Design of any new intelligent search engine should be at least based on two main motivations:
1) The web environment is, for the most part, unstructured and imprecise. To deal with information in the web environment what is needed is a logic that supports modes of reasoning which are approximate rather than exact.

While searches may retrieve thousands of hits, finding decision-relevant and query-relevant information in an imprecise environment is a challenging problem, which has to be addressed and
2) Another, and less obvious, is deduction in an unstructured and imprecise environment given the huge stream of complex information.

In this presentation, we will first present the state of the search engines and internet. Then we will focus on development of a framework for reasoning and deduction in the web. A web-based model to decision model for analysis of structured database will be presented. A framework to incorporate the information from web sites into the search engine will be presented as a model that will go beyond current semantic web idea. Another important and unique component of our system is compactification algorithm or Z-Compact. Z-Compact algorithm developed by L. A. Zadeh and it has been implemented for the first time as part of BISC-DSS for automatons multi-agents modeling as part of ONR project and has been extended to handle linguistic variables with deduction capability and currently is part of the BISC-DSS software and it has been applied in several applications.

2. Conceptual Fuzzy Web Reduction

When we browse by internet, search engines propose to us a quantity of web pages with different sizes depending on available data in the web. It is now well known that the main difficulty for the user is his incapacity to sort obtained references in decreasing importance level with respect to his request. The main reason is that this requires a big amount of human effort. Building computerized systems for data filtering and classification is now a high priority. Words used for searching in the web have a fuzzy meanings. Importance level

of a word in a document is itself fuzzy. For that reason, fuzzy clustering tools should be useful to decompose a documentary database into a minimal number of the most independent classes of web pages. Few selected representative documents should be enough to be given a highest importance level to the user.

In my intervention, I will present two major problems and resolve them by conceptual approaches. A first problem is about the different ways to reduce dynamically arriving web pages into a minimum most significant web pages to the user. The second problem is to find a database decomposition into a minimal number of classes.

Formal concept analysis emerged as a strong theoretical foundation for data structuring, and classifying or extracting information and/or knowledge [7,8,10,11,13]. As a matter of fact, it is proven that a database can be organized as a lattice of concepts. A concept is a maximal Cartesian product of two sets (a set of objects and associated properties). As for example the World Wide Web pages may be organized as a lattice of concepts, where a concept is a maximal set of web pages addresses all indexed by a maximal set of terms. Importance of concepts have been discovered from different ways, as for example from the definition of regular or difunctional relations [3,8,15]. Many algorithms for creating a lattice of concepts from a binary relation have been developed. The most efficient algorithms are incremental, because starting from an initial lattice [10], when new indexed web pages are integrated in the system, only few concepts from the initial lattice are updated, other new concepts are integrated.

Another alternative for a database structuring is to find a coverage of all data by a minimal number of concepts [13]. This coverage has the advantage to be economical with respect to the total lattice of concepts. But unfortunately both of these structuring methods have the inconvenient to occupy a lot of space even for only ordinary user request.

The idea of data reduction is to start from the conceptual coverage to find a priority between obtained addresses given by search engines. As a matter of facts different methods for concept ordering may be designed. Among these methods, we can define a measurement of concept importance, then assume that the most important concept as the one with the greatest measure. These methods may be used for database abstraction and text summarization.

Why do we need to extend these methods using fuzzy theory? The idea is natural because web documents are not indexed by all words with the same weight equal to one(1). But a degree is associated to a pair (document, term). By this way, we may represent the total web space by a fuzzy binary relation [12], that we can also generate automatically by an approximate method. Starting from this it is now possible to extent the lattice crisp structure to a lattice of fuzzy concepts. We can also find a coverage of all data space by a minimal number of fuzzy concepts. A fuzzy concept is a product of the set of web addresses by a fuzzy set of indexing terms. For the fuzzy conceptual coverage, we define an order between obtained concepts to give a priority to the best concept in terms of pertinence.

Another alternative is to extend direct crisp binary relation reduction to fuzzy binary relation, using different fuzzy logics. Experimentations realized on real database have given good results in terms of reduction rates. As a consequence, application of this method for approximate web reduction is possible.

In this talk, we will consider the problems related to fuzzy context conceptual decomposition by generalizing formal properties of binary relations to fuzzy binary relations. We will also discuss, the problem of semantic web structure and dynamic web database decomposition and reduction.

3. REFERENCES

[1] Beck H. W., Anwar T., Navathe S.B. (1994). “A Conceptual Clustering Algorithm for Database Schema Design.” *IEEE Transactions on Knowledge and Data Engineering*, **6**(3):396-411.

[2] Belkhiter N., Bourhfir C., Gammoudi M.M., Jaoua A., Le Thanh N., et Reguig M.

(1994). “Décomposition Rectangulaire Optimale d'une Relation Binaire: Application aux Bases de Données Documentaires.” *Information Systems and Operational Research Journal*., **32**(1): 33-54, Février.

[3] Jaoua A., Belkhiter N., Desharnais J., Moukam T. (1992). “Propriétés des Dépendances Difonctionnelles dans les Bases de Données Relationnelles.” *Information Systems and Operational Research Journal*, **30**(3): 297-316, Août.

[4] Norris E.M. (1978). “An Algorithm for Computing the Maximal Rectangles in a Binary Relation.” *Revue Roumaine de Mathématiques Pures et Appliquées*, **23**(2):243-250.

[5] Tarski A. “On the calculus of relations.", *Journal of Symbolic Logic* **6**, 1941. pp. 73-89.

[6] B.A. Davey and H.A. Priestley “Introduction to Lattices and Order,” *Cambridge Mathematical*, Textbooks, 1990.

[7] B. Ganter, and Wille, “Formal Concept Analysis,” *Springer-Verlag*,1999.

[8] G.Schmidt and S. Ströhlein, “Relations and Graphs,” *Springer-Verlag*, 1989.

[9] A. Jaoua and S. Elloumi, Galois Connection, “Formal Concepts and Galois Lattice in Real Relations: Application in a Real Classifier,” *The Journal of Systems and Software* 60 (2002) 149-163.

[10] G.W.Mineau and R. Godin, “Automatic Structuring of Knowledge Bases by Conceptual Clustering,” *IEEE Transactions on Knowledge and Data Engineering*, 7(5) :824-829, 1995.

[11] S. BenYahia, K.Arour, A. Slimani and A. Jaoua, “Discovery of Compact Rules in Relational Databases.” *Information Journal*, 3(4) :497-511, October 2000.

[12] H. Ounalli and A. Jaoua, “On Fuzzy Difunctional Relations,” *Journal of Information Sciences*, 95, 219-232 (1996).

[13] F. Hassan, H. Saleh, T. Mosaid, “Conceptual Text Summarization,” *Senior Project, University of Qatar*, January 2004.

[14] A. Jaoua, A. Mili, N. Boudriga, et “Regularity of relations: A measure of uniformity,” *J.L. Durieux Journal of Theoretical Computer Science*, Vol 79, February 1991, Pages 323-339.