FUZZY INFORMATION ENGINEERING: BACKGROUND AND APPLICATIONS

Didier Dubois – Henri Prade

Institut de Recherche en Informatique de Toulouse, Université Paul Sabatier-CNRS

118 route de Narbonne, 31062 Toulouse Cedex, France

Tel.: (+33) 61.55.63.31 / 65.79 – Fax: (+33) 61.55.62.39 – Email: {dubois, prade}@irit.fr

Abstract. We claim that information has become a matter of scientific investigation of its own. The aim of this invited talk is to show that fuzzy sets play an important role in information sciences and engineering for bridging the gap between man and computers, through a major contribution to data and knowledge representation.

I - A VIEW OF INFORMATION SCIENCES AND ENGINEERING

Beside its potentially enormous economical, industrial and social impact, *information* has become a matter of scientific investigation of its own, standing apart from other older established fields such as physics, mathematics, etc. This interest is in part strongly motivated by the confluence of computer and communication technologies which allows for the collection and storage of large amounts of information and its subsequent dissemination in a very efficient way. In several recent publications [7, 8, 11], we have tried to promote a unified view of information sciences and engineering that transcends boundaries existing between subfields that sometimes quarrel with one another, sometimes ignore one another, while having in common the purpose of processing some kind of information in some way. Moreover some newly emerged key-words such as Soft Computing, Computational Intelligence and the like are often misunderstood as to their actual potential and novelty [7, 8]. It is very important to organize research about information in order to achieve its recognition as a new scientific discipline with specific needs in terms of basic research, especially particular mathematical and logical tools, and a specific technology.

In [11] the following definition of information has been proposed: any collection of symbols or signs, produced either by the observation of natural or artificial phenomena, or by the cognitive activity of human beings, which can be used to understand the world around us, help in decision making or to communicate with one another. The concept of information engineering is an attempt to bring together under one umbrella many scientific investigations which have developed along with the emergence of the computer technology and have to do with information. Informally, we can tentatively see information science and engineering as concerned with three main tasks [11]: namely *clarifying, storing/retrieving,* and *exploiting information.*

1.1 Clarifying information

By *clarifying information* we mean the process of elaborating raw data into more useful and comprehensible forms. This task ranges from enhancing information via smoothing or eliminating redundant or erroneous data, completing information via interpolation, to higher cognitive activities such as abstracting and modeling. Some clarification tasks can be performed at the level of a sensor itself. A particularly extensive family of clarification methods is the arsenal of techniques used in computer vision, whose purposes are the reduction of uncertainty, the recognition of relevant patterns and the labelling of scenes. In addition many of the operations used in data fusion for synthesizing and combining pieces of data fall in the category of clarification methods, since more often than not, information is supplied by several, possibly conflicting sources and not in an homogeneous format. Clarification appears as a transformation process that turns primary information into a more structured information, that is more understandable and useful.

There is also an effort to go from information about particular cases (what is often called data) to more generic forms of information pertaining to a population of items (what is often called knowledge). In some cases, especially when abstracting is in order, it necessitates the use of linguistic labels to communicate the main features of the information. Alternatively, clarification may also come down to a kind of data compression by laying bare an underlying mathematical model. Clarified information then can be appropriately stored and be further exploited in reasoning and decision processes.

Typical of clarification are techniques such as clustering, filtering, other data analysis methods such as regression, identification, inductive learning and more generally any method aimed at abstracting and structuring information.

1.2 Storing/Retrieving information

Retrieving information means to extract, based on certain requirements and specifications, already existing, possibly hidden, pieces of information. This component of information engineering includes many of the methodologies used in databases, information retrieval and the new emerging field of multi-media. It is strongly dependent on the storage technology and the knowledge representation framework used.

In its most elementary form, which we may call querying, the retrieval of information does not involve a modification of existing information, as in the clarifying task. A large body of information is stored in some format and a query is entered. As a result of a query matching process a class of objects satisfying the query are provided. An important objective in the construction of these types of systems with respect to the querying task is the facilitation, for the user, of the expression of their goals regarding the information to be retrieved. This objective requires allowing for rather sophisticated forms of queries. It is also worth noting that a basic distinction between the database and information retrieval technologies is related to the use in information retrieval of intermediary descriptions of the information in terms of keywords, whereas in databases information is described directly in terms of attributes values. The handling of multimedia databases requires the joint use of the two approaches.

Some modes of storing information do not allow the retrieval process to be reduced to one of simply locating items in a file. For instance in deductive databases, information is encoded in a declarative way via some kind of logic-based language. The retrieval process requires that the information system be equipped with inferential power. In this environment retrieving information involves a reasoning task, whereby hidden information is made explicit for the user. Sometimes the information held for sure is too poor and one may wish to exhibit plausible information, possibly revised in a further step. It requires the use of some form of default inference. This type of concern is present in so-called intelligent information systems, which lie at the boundary between database research and artificial intelligence.

1.3 Exploiting information

In the *exploiting information* component we are encompassing a large range of activities including decision-making, optimization, design, and control. Generally in such tasks, information plays a supportive role to some sort of paradigm or model. In order to make the best use of the knowledge embedded in the model, the ability to represent and manipulate different types of information becomes very important here. The development of technologies that allow us to combine numerical, logical, linguistic and visual information is an important part of the agenda of this aspect of information engineering. All pieces of information are used in the problem-solving task, whose solution is presented to the users in order to guide their course of action. Several prototypical situations can be distinguished:

-) problems where the number of potential solutions is very small but the choice is made difficult due to many conflicting goals and/or the presence several of decision-makers. This is the realm of decision support systems for strategic decision-making;

-) problems where the computation of a solution is a highly combinatorial task but the local preference modeling is made very clear and simple. This is typically the case of optimization problems that occur in engineering design or large-scale systems where the difficulty lies in the intricacy of a large number of local constraints and decision variables;

-) problems where the solution is a complex entity that is built in several steps because time is involved in the picture. This is the case for planification problems in robotics or manufacturing;

-) lastly, problems where information must be exploited in a dynamic environment and reactivity is the key issue. Decision is then made in real time by exploiting some feedback. This is the realm of control problems.

In each of the above three components of information engineering that we have touched upon (clarifying, storing/retrieving and exploiting information), we want to emphasize the importance of two aspects: the representation of information and the communication of information between man and machine. Knowledge representation has become a central topic for basic research in Artificial Intelligence due to the necessity for a communication language between man and computers. This communication goes in both directions: information generated by the machine should be in a form easily comprehensible by human beings while information provided by human beings should be in a format exploitable by the computer. Logic has become the natural framework for the formalization of reasoning techniques. However classical logic is alone unable to face the major challenge of accounting for the various forms of human reasoning, necessary to clarify, retrieve and exploit information in an automated, yet human-friendly way. Moreover a great amount of information comes in the form of numerical data, as opposed to human knowledge. Data processing methods and symbolic knowledge representation and reasoning tools must thus be used conjointly, on the basis of their complementarity, and not be opposed as exclusive approaches to information engineering. This presupposes that the dogma of the monopoly of symbolic representations in AI be given up, and that data-driven schools of information processing open to logical approaches, in the scope of developing software systems that not only efficiently solve problems but are capable of explaining the obtained solutions to users.

II - FUZZY SETS: TECHNIQUES AND MEANINGS

Fuzzy set theory is composed of an organized body of mathematical tools particularly well-suited for handling the natural incompleteness of available information in a flexible way. It offers a unifying framework for modelling various types of information ranging from precise numerical, interval-valued data, to symbolic and linguistic knowledge, with a stress on semantics rather than syntax.

The main components of fuzzy set theory are as follows [9]:

Membership functions

The notion of membership function enables our notion of a class to be extended to categories that have no precisely defined boundaries, as is often encountered in linguistic information. Historically, fuzzy sets have been mostly introduced for this ability to model linguistic categories. The gradedness of concepts is pervasive both in the mental as well as the physical worlds. This idea has perhaps been most succinctly captured in Zadeh [4]'s view of fuzzy sets as "computing with words". This view highlights the importance of the interface between the data emanating from the physical world and the categories with which human beings are most comfortable in expressing information. The idea that a class is not an all-or-nothing matter sometimes results from the conflict between symbolic and continuous representations of information, and sometimes because the apparent incompleteness of information hides a notion of typicality or usuality that enables this incompleteness to be reduced in practical reasoning tasks.

Aggregation operators

A large body of information aggregation operations has been laid bare in the last twenty five years. Fuzzy set-theoretic operations unify connectives coming from logic and trade-off operations coming from decision theory and statistics. Generalizations of set theoretic unions and intersections, variants of mean aggregations have been studied in detail at the mathematical level on an axiomatic basis. This is the theory of triangular norms, and uninorms. The diversity of connectives concerns not only the type of aggregation (conjunctive, disjunctive, trade-off). It also pertains to the possibility of weighing arguments in the aggregation, and introducing quantifiers. More recently a unified view of quantified and weighted aggregations has been proposed whereby groups of arguments can be weighted. This is achieved by means of the Choquet integral on numerical universes and Sugeno integral in ordinal settings.

Fuzzy relations

Fuzzy set theory provides the opportunity of generalizing the theory of relations to many-valued relations in a general algebraic setting. Especially the max-min calculus of fuzzy relations offers a computationally nice and expressive setting for constraint propagation. Tools, both mathematical and algorithmic, for solving fuzzy relational equations have been developed. Fuzzy relations offer a sound theoretical basis for the study of fuzzy systems.

Similarity and fuzzy orderings

Important particular cases of relations are equivalence and ordering relations. Fuzzy relations have been used to generalize equivalence and give formal foundations to a theory of similarity. This theory is closely related to the theory of distances. However it also enable a generalization of partitions to be developed. It also leads to techniques for the comparison of fuzzy sets. Similarly, the concepts of partial ordering, preordering, linear ordering have been generalized to valued relations. More recently many results have been obtained on the decomposition of a fuzzy relation into the union of a similarity relation and a fuzzy ordering relation.

Possibility theory

A new theory of uncertainty has been developed on the basis of fuzzy sets: possibility theory. The main purpose of possibility theory is to model incomplete information in a flexible way. Possibility theory is very similar to probability theory, but possibility theory is maxitive, not additive: the possibility of a disjunction of events is the maximum of the possibilities of each event. Moreover the uncertainty of an event is captured by two degrees: possibility and necessity, instead of one. Possibility distributions are fuzzy sets of possible mutually exclusive outcomes. The possibility of an event evaluates its degree of unsurprizingness while its necessity evaluates its degree of acceptance. Possibility theory can be a purely ordinal construct. It is then the simplest ordinal uncertainty theory. Possibility degrees can also be numerical. Then they can be viewed as upper probability bounds. Numerical possibility theory is then the simplest theory of imprecise probability. All notions existing in probability theory (conditioning, independence, mean values, etc.) have counterparts in possibility theory [10].

New insights in information theory

Fuzzy set theory have given the opportunity to study evaluations of the quantity of information that differ from standard probabilistic notions based on entropy. Two points of view can be envisaged: to what extent a fuzzy set is fuzzy? to what extent a fuzzy set is specific (= focuses on a small set of values)? The first point of view has given rise to a literature on indices of fuzziness. The second is closely connected to possibility theory. Indices of specificity are obtained by fuzzy extensions of set cardinality. Fuzzy set inclusion also offers a tool for comparing fuzzy sets in terms of their relative specificity.

Fuzzy numbers and extensions of interval-valued analysis.

Fuzzy numbers are fuzzy sets of the real line that generalize intervals. A calculus of fuzzy numbers that parallels the calculus of random variables has been developed under various assumptions according to whether the underlying imprecise variables are interactive or not. Fuzzy counterparts of probabilistic limit theorems exist. On such a basis metric spaces of fuzzy quantities have been defined, that serve as a basis for fuzzy analysis. Many results exist on the integration and differentiation of fuzzy-valued functions. Also of interest is the available literature on fuzzy random variables which mixes statistics and fuzzy interval computations.

Approximate reasoning and fuzzy rules [4].

The logical side of fuzzy sets pertains to many-valued logics which have been recently investigated in a unified way. However fuzzy logic is a tool for reasoning with fuzzy restrictions on the universe of discourse, and not only a many-valued logic. Fundamental principles of approximate reasoning have been laid bare in terms of combination and projections of fuzzy restrictions. The many-valued extensions of implication provide a basis for modelling fuzzy rules which are basic building blocks of fuzzy logic. The main pattern of fuzzy reasoning is the generalized modus ponens.

What do membership functions mean?

Many basic notions used in the measurement and manipulation of information, such as *uncertainty, preference* and *similarity*, are naturally graded and can be captured by fuzzy set theory. Each of these concepts lead to a specific meaning of the membership function [7], that has been introduced by Zadeh in a specific paper.

A first meaning (and historically the oldest one [2]) is the expression of closeness, proximity, similarity, indiscernibility, indistinguishability and the like. Under this semantics elements with membership one are viewed as prototypical elements of the fuzzy set, while the other membership grades estimate the closeness of elements to the prototypical ones.

A second semantics associated with fuzzy sets is related to the representation of incomplete or vague states of information under the form of possibility distributions (Zadeh [16]). This view of fuzzy sets enables imperfect, imprecise or uncertain information to be stored and retrieved via appropriate reasoning tools.

A third semantics for a fuzzy set useful when modelling a flexible constraint, specification, or goal, expresses preferences between more or less acceptable solutions with respect to the constraint (Bellman and Zadeh [3]). The gradedness introduced by the use of fuzzy sets refines the simple binary distinction made by ordinary constraints, crisp specifications, or all-or-nothing goals between completely acceptable values and completely forbidden ones. Membership functions are then viewed as a kind of utility functions.

In this classification, the grades of membership of a fuzzy set such as *tall* may be understood in three different manners depending upon the context. First, 'tall' may describe a fuzzy class of sizes more or less close to prototypical values. Second, it may represent an incomplete state of information, e.g., we only know that 'John is tall' without knowing his height more precisely. Lastly, it may express a flexible constraint or a preference profile, e.g., we are looking for somebody who is tall.

III THE ROLE OF FUZZY SETS IN INFORMATION ENGINEERING

Let us look at the three components of information engineering and summarize the role fuzzy sets can play in each of them.

In clarification techniques such as clustering, recognition and classification, fuzzy set theory offers a tool for representing classes with soft boundaries, where objects which are judged to be sufficiently similar are gathered in the same (fuzzy) class (Bellman, Kalaba and Zadeh [2], Bezdek et al. [5]). We are also taking advantage of the similarity notion naturally embedded in the membership functions, and the proximity or possible overlap between classes. In fuzzy modeling and control, fuzzy rules are used for interpolation purposes, with a concern of linguistic interpretability [1, 15]. Other kinds of clarification techniques concern data analysis and data fusion. In data analysis, it has become possible to exploit fuzzy data in regression methods and other techniques such as least squares. Extensive use of possibility theory and fuzzy interval analysis is made. Statistical methods have been extended to linguistic data. In data fusion, be it sensor fusion, expert opinion pooling, or fuzzy database fusion, fuzzy set theoretic aggregations play a major role due to their flexibility. Lastly, several methods for extracting fuzzy rules from data have emerged, some based on decision trees, other using neural net methods and soft computing tools. While one aspect of clarification concerns the extraction of knowledge from data, another side of clarification concerns the estimation of a precise value from vague knowledge, what is often called defuzzification. The main asset of fuzzy set theory in this side of information engineering is the smooth interface between linguistic and numerical representations of information.

In retrieving information, it is often desirable to allow for flexible requests in order to more faithfully express the desires of the user of the information system. Flexible queries can be modelled by fuzzy sets. The aggregation of the satisfaction levels of individual objectives involved in the request can be achieved thanks to the vast array of aggregation operators available in fuzzy set theory. Another aspect of the use of fuzzy sets lies in the representation of stored information, and especially the representation of the uncertainty pervading the data or their descriptions. The ability of fuzzy sets to represent various types of information (precise, imprecise, partially conflicting, linguistic) is of great benefit in this task. Lastly, all systems whose purpose is to restore information via reasoning techniques can take advantage of approximate reasoning tools such as possibilistic logic, fuzzy rule-based inference. Qualitative possibility theory is a natural framework for exception-tolerant reasoning [9] Other more specific fuzzy reasoning framework have been proposed for temporal reasoning, order-of-magnitude reasoning, etc. Fuzzy arithmetics play a fundamental role in accomplishment of these more specialized tasks. Lastly, fuzzy relational methods for causal diagnosis have been devised. See [4].

When exploiting information, the use of fuzzy sets may appear in the representation of uncertainty (when making decision in uncertain environments), in the expression of preferences (in formulating objectives in optimization problems) or the capturing of similarities (as in case-based reasoning). An important asset of fuzzy sets in decision-making lies in the very general framework it provides that allows us to cope with the presence of multiple criteria, including the representation of complex relationships between criteria (using techniques such as OWA operators and fuzzy integrals for instance), the notion of flexible constraints in design and scheduling problems, the presence of uncertainty in the data (using possibility theory). It can bridge the gap between value functions that insist on numerical, totally ordered preference models, and relational approaches that handle non transitive notions of preference and indifference and capture notions of incomparability. See [14].

A detailed account of fuzzy sets theory and its applications in the scope of information engineering can be found in the Handbook of Fuzzy Set Series volumes, respectively devoted to the basic notions of fuzzy sets [10], the application of fuzzy sets to mathematics [12], the use of fuzzy sets in automated reasoning, learning, data fusion, information retrieval and databases [4], the use of fuzzy sets in pattern classification and image processing [5], the use of fuzzy sets in systems engineering [13], the use of fuzzy sets in optimization, preference modelling and statistics[14], and an overview of more specific applications where a proper representation of information requires fuzzy set techniques [18].

IV - REFERENCES

[1] Babushka R. (1998) *Fuzzy Modeling for Control*. Kluwer Academic, Dordrecht, The Netherlands.

[2] Bellman, R.E., Kalaba, R. and Zadeh, L.A. Abstraction and pattern classification. *J. of Mathematical Analysis and Applications 13*, 1-7, 1966.

[3].Bellman, R.E. and Zadeh, L.A. Decision-making in a fuzzy environment. *Management Sciences 17*, 141-164, 1970.

[4] Bezdek J., Dubois D., Prade H., Eds. (1999) *Fuzzy Sets in Approximate Reasoning and Information Systems*. The Handbooks of Fuzzy Sets Series, Kluwer Academic Publishers, Dordrecht, The Netherlands.

[5] Bezdek J., Keller J., Krishnapuram R., and Pal N. R. (1999) *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, The Handbooks of Fuzzy Sets Series, Kluwer Academic Publishers, Dordrecht, The Netherlands.

[6] Dubois D., Prade H. (1997) The three semantics of fuzzy sets. *Fuzzy Sets and Systems*., 90, 141-150

[7] Dubois D., Prade H. (1998) Soft Computing, Fuzzy Logic, and Artificial Intelligence *Soft Computing*, Vol 2,

[8] Dubois D., Prade H. (1998) Fuzzy information Engineering and Soft Computing. In *Computational Intelligence and Applications* (P.S. Szczepaniak, Ed.) Physica-verlag, Heidelberg, 18-33.

[9] D. Dubois, and H. Prade.(1998) Possibility theory: qualitative and quantitative aspects. in P. Smets, editor, *Handbook on Defeasible Reasoning and Uncertainty Management Systems — Volume 1: Quantified Representation of Uncertainty and Imprecision*. Kluwer Academic Publ., Dordrecht, The Netherlands, 169-226.

[10] Dubois D., Prade H., Eds. (1999) *Fundamentals of Fuzzy Sets*. The Handbooks of Fuzzy Sets Series, Kluwer Academic Publishers, Dordrecht, The Netherlands.

[11] Dubois D., Prade H., Yager R.R. (Eds.) (1997) *Fuzzy Information Engineering: A Guided Tour of Applications*. Wiley, New York.

[12] Hoehle U. and Rodabaugh S., Eds.(1999) *Mathematics of Fuzzy sets: Logic, Topology and Measure Theory* .The Handbooks of Fuzzy Sets Series, Kluwer Academic Publishers, Dordrecht, The Netherlands,

[13] Nguyen H.T. and Sugeno M. Eds. (1998) *Fuzzy Systems: Modeling and Control.* The Handbooks of Fuzzy Sets Series, Kluwer Academic Publishers, Dordrecht, The Netherlands.

[14] Slowinski R., Ed. (1998). *Fuzzy Sets in Decision Analysis, Operations Research and Statistics*, The Handbooks of Fuzzy Sets Series, Kluwer Academic Publishers, Dordrecht, The Netherlands.

[15] Verbruggen H.B., Zimmermann H.-J., Babushka R. (1999) *Fuzzy Algorithms for Control*. Kluwer Academic, Dordrecht, The Netherlands.

[16] Zadeh, L.A. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems 1*, 3-28, 1978.

[17] Zadeh, L.A. Fuzzy logic = Computing with words. *IEEE Trans. on Fuzzy Systems 4*, 103-111, 1995.

[18] Zimmermann H.-J.. (1999) *Practical Applications of Fuzzy Technologies*. The Handbooks of Fuzzy Sets Series, Kluwer Academic Publishers, Dordrecht, The Netherlands.