Fuzzy Databases in the New Era

B.P. Buckles and F.E. Petry,
Center for Intelligent and Knowledge-based Systems
Department of Computer Science
Tulane University, New Orleans, LA 70118

Abstract

The last five years have been witness to a revolution in the database research community. The dominant data models have changed and the consensus on what constitutes worthwhile research is in flux. Also, at this time, it is possible to gain a perspective on what has been accomplished in the area of fuzzy databases. Therefore, now is an opportune time to take stock of the past and establish a framework. A framework should assist in evaluating future research through a better understanding of the different aspects of imprecision that a database can model. Additionally, a framework can indicate which past efforts a specific instance of present research can be related.

1 Introduction

It is becoming evident on the commercial side that we are well within the era of the third generation of databases. (The first was graph-oriented models and the second relational systems.) Perusal of the proceedings of present-day conferences will strengthen this observation. Research on relational databases is now relegated to a few issues concerning efficiency, distributed systems, and transaction processing.

The trend is undeniably in the direction of object-oriented databases. Deductive databases are also important, but it appears that their contribution will be in the form of logic-enhanced object-oriented databases. Had the evolutionary approach from relations to objects been adopted [8], the situation would have been more favorable for incorporating the mechanisms and applying the knowledge gained in 15 years of fuzzy database research. But, that was not to be.

It would seem that the best we can do is separate the work of the past into categories that will more clearly distinguish what is comparable. Doing this will also assist in gaining a perspective on where effort in the future should be placed.

The premises on which this paper is based are illustrated in Figure 1. Enterprises are either *precise* or *vague*. The database extension that represents the enterprise is either *precise* or *imprecise*. Query languages are designed to express the user's retrieval requests in either a *crisp* manner or not.

2 Precise Enterprises

Databases are one form of model of aspects of the real world. The specific segment of the real world which a specific database models is called the enterprise. Nearly all present databases model enterprises that are crisp. (One is tempted to say that crisp enterprises are modelled without exception.) A crisp enterprise is one that is highly quantifiable – all relationships are fixed and all attributes have one value.

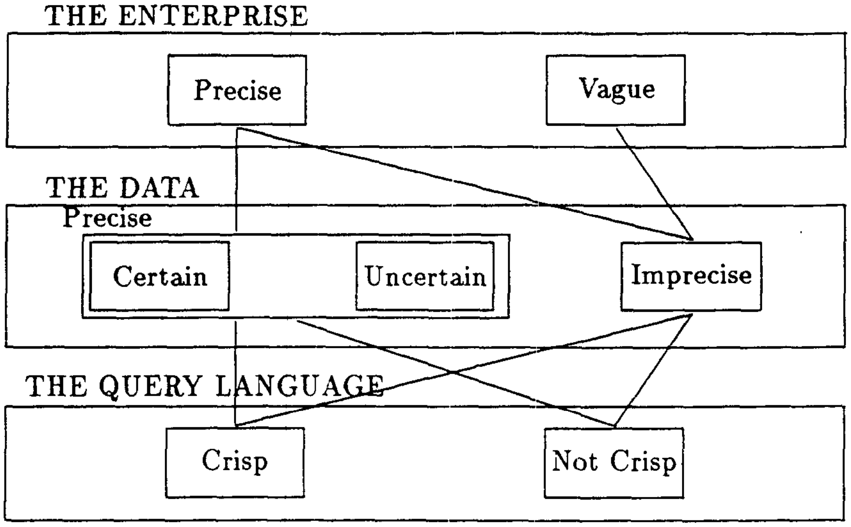

Figure 1: The Fuzzy Database Landscape

2.1 With Precise Data

The case of precise enterprise-precise data includes virtually all database systems in wide-spread use. The potential payoff of exploiting the subcase in which the data, while precise, is considered uncertain is largely yet unexplored. If the query language is crisp or not, the issue is whether a particular data item matches a query term when it is not identical to the term.

Not withstanding that studies can exist indicating which errors are most likely in the data, essentially no information about the data can be assumed other than its value and perhaps the date entered. The best approach for accounting for uncertain data is the query language. A simple query language in which a user can indicate the degree of relaxation permitted to achieve a match is needed. Even with data having no uncertainty, such a query language would be useful.

2.2 With Imprecise Data

It was the precise enterprise and imprecise data that inspired one of the earliest seminal papers on uncertainty in databases [10]. The key notion is that while only one value applies to the enterprise, the database extension may contain a set. The classical approach is to reduce retrieval to 3-value logic [7] whether the query language is crisp or not. Each database object is *surely*, *maybe*, or *surely not* a response to the query.

The appropriate branch of fuzzy set theory that deals with problems of this nature is *possibility theory*. In possibility theory, the objective is – given that an object is a member of a fuzzy set, determine the possibility that a specific value applies. (Contrast this with – given a specific value, determine its "degree of membership" in a fuzzy set.)

The application to the precise world/imprecise data is obvious. The value in the database is a possibility distribution that is taken to mean the limits of knowledge concerning the actual value [11, 12]. The representation issues are illustrated in Figure 2. The two relations in the figure have the same schema but are not intended to contain equivalent data. The ultimate goal of the possibility-based model is to provide more information about the data retrieved in the *maybe* category.

Name	Age	Salary
Tom	36	{0.8/1900,1.0/2000,0.8/2100}
Mary	{1.0/26,0.8/27}	2200
⋮	⋮	⋮

(a) With Explicit Possibility Distributions

Name	Age	Salary
Tom	*Young*	2000
Mary	26	*moderate*
⋮	⋮	⋮

(b) With Implicit Possibility Distributions

Figure 2: Possibility-based Relations

In part (a) of the figure, the possibility distributions are explicitly shown. A more common implementation is given in part (b). Each possibility distribution is given a linguistic identifier. The actual distribution is given elsewhere in the database in the form of a relation having the name of the linguistic identifier.

Normally, the symbol Π is used to represent possibility distributions. Let $\Pi_{young} = \{1.0/22, 1.0/23, 0.8/24, 0.6/25, 0.4/26, \cdots\}$. The query

$$\sigma_{age=young}(R_{part(a)})$$

would return *Mary* with a possibility value of 0.4 computed from min(0.4, 1.0) where the values come from 0.4/26 in *young* and 1.0/26 in the tuple. The computation, in practice, is computationally difficult and requires an application of the extension principle. The extension principle requires, in effect, the comparison of every pair of terms in the query value and database value.

Bosc [2, 3] continues this research direction today along relational lines. Emphasis is on the semantics of quantification and development of a practical fuzzy version of SQL. From the perspective of the object-oriented model, Van Gysegham and De Caluwe [13] have made the most progress.

3 Vague Enterprises

Certain enterprises are not crisp. For example, the relationships among the world's languages can be fixed only in an arbitrary manner such as Russian is a Slavic language without contamination or tint from any other language group. In such enterprises, values are not necessarily crisp. Examples for which crisp values are not realistic include an international investment database containing, for each country, attributes such as "strength of judicial system", "extent of federal control over private enterprise", and "stability of government."

It is for this case that the earliest model developed by the authors [4, 5] applies. It, the similarity-based model, is illustrated in Figure 3(a). Attribute values, for which it is assumed no precise values exist, are represented as linguistic terms. Linguistic terms are themselves related to each other by a similarity relationship as shown in Figure 3(b).[1] The model is best applied to

[1] A similarity relationship is a generalization of an equivalence relation with counterparts to reflexivity, symmetry, and transitivity. Space, not to mention possible reader patience, prevents more complete specification.

enterprises (or parts of enterprises) in which the linguistic sets are finite and discrete although the model can be extended to continuous domains [6].

Name	Hobby	Disposition	Intelligence
Dale	*Sports*	*Jovial*	*Smart*
Jan	*Stamps*	*Lively*	*Below Average*
Don	*Music*	*Mild*	*Gifted*
⋮	⋮	⋮	⋮

(a) Relation

	Mild	*Jovial*	*Lively*
Mild	1.0	0.5	0.3
Jovial	0.5	1.0	0.3
Lively	0.3	0.3	1.0

(b) Similarity Relationship

Figure 3: Similarity-based Relations

With such enterprises and database extensions, crisp query languages make no sense. The concept of querying the similarity database is that equivalence is replaced with "similar to." One specifies a query with linguistic terms and indicates the minimum similarity or degree of matching that can be tolerated. For example,

$$\sigma_{disposition=mild}(R_{part(a)}) : Level \geq 0.5$$

will retrieve {0.5/Dale, 1.0/Don, $\cdots$ }.

The object-oriented model offers at least four possibilities (and problems) for representing vague enterprises. All are illustrated in Figure 4.

- The value of an attribute, e.g., the late 20th Century is a point in time
- The degree to which an object belongs to a class, e.g., Nelson Mandela is an instance of a liberal
- The compositional (also known as containment) hierarchy determined by reference attributes, e.g., Nelson Mandela supports open trading markets
- The class hierarchy, e.g., a liberal is a socialist

Incorporating these into one model is a daunting task but [1] attempts the first three and suggests an approach for the latter. In [9] the latter is tackled. The class hierarchy is the most troublesome to treat as uncertain because of the inheritance issues it raises. Some success is possible if the memberships in a linear hierarchy are monotonically increasing or decreasing.

4 Conclusions

The lack of commercial/industrial impact of the research in fuzzy databases has to be of concern. In fairness, it should be pointed out that other methods of uncertainty management have suffered

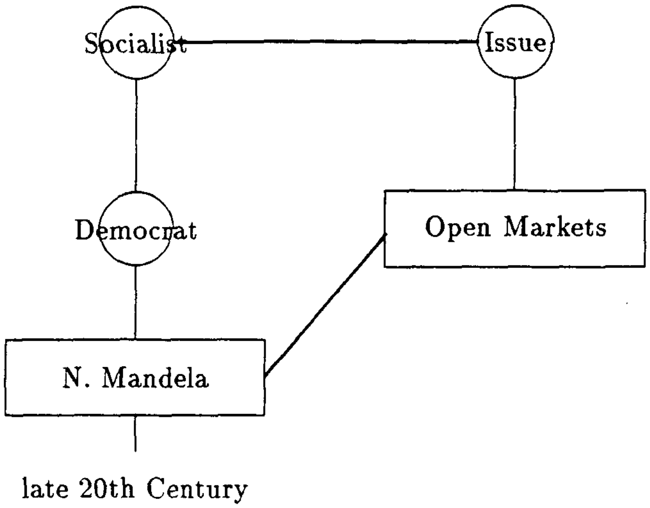

Figure 4: Object-oriented Database with Fuzzy Schema

similar fates. Most agree that databases, like any model, are imperfect realizations of the artifact modelled. Nevertheless, there is not yet on the horizon a system having potential wide-spread use that incorporates other than trivial means for representing or manipulating uncertainty.

The reluctance of vendors to utilize the advances that have been made thus far suggests that, for the time being, modest goals are in order. With the lack of acceptance of uncertainty management for precise enterprises, the acceptance of tools for vague enterprises is likely far away. From this standpoint, it seems likely that the work in [3, 13] is closest to the mark. There is an obvious lack of activity on the yet easier target – precise enterprise, precise but uncertain data, and noncrisp queries. We recognize, however, that some of the aforementioned research applies here where precise data is a special case of imprecise data.

References

[1] G. Bordogna, D. Lucarella, and G. Pasi. A fuzzy object oriented data model. In *Proc. IEEE International Conference on Fuzzy Systems*, pages 313–318, June 1994.

[2] P. Bosc, M. Galibourg, and G. Hamon. Fuzzy querying with sql: Extensions and implementation aspects. *Fuzzy Sets and Systems*, 28(3):333–349, Dec. 1988.

[3] P. Bosc and O. Pivert. Fuzzy querying in conventional databases. In L. Zadeh and J. Kacprzyk, editors, *Fuzzy Logic for the Management of Uncertainty*, pages 645–671. John Wiley and Sons, 1992.

[4] B. P. Buckles and F. E. Petry. A fuzzy representation of data for relational databases. *Fuzzy Sets and Systems*, 7:213–226, 1982.

[5] B. P. Buckles and F. E. Petry. Information-theoretic characterization of fuzzy relational databases. *IEEE Trans. on Systems, Man and Cybernetics*, 13(1):74–77, Feb. 1983.

[6] B. P. Buckles and F. E. Petry. Extension of the fuzzy database with fuzzy numbers. *Information Sciences*, 34(4):121–132, Apr. 1984.

[7] E. Codd. Extending the database relational model to capture more meaning. *ACM Trans. on Database Systems*, 4(4):397–434, July 1979.

[8] C. for Advanced DBMS Function. Third-generation data base system manifesto. *SIGMOD Record*, 19:31–44, Sept. 1990.

[9] R. George, B. P. Buckles, and F. E. Petry. Modeling class hierarchies in the fuzzy object-oriented model. *Fuzzy Sets and Systems*, 60(3):259–272, Dec. 1993.

[10] W. Lipski. On semantic issues connected with incomplete information databases. *ACM Trans. on Database Systems*, 4(3):262–296, May 1979.

[11] H. Prade and C. Testemale. Generalizing database relational algebra for the treatment of incomplete/uncertain information and vague queries. *Information Sciences*, 34:115–143, 1984.

[12] E. H. Ruspini. Possibility theory approaches for advanced information systems. *IEEE Computer*, 9(2):83–89, Feb. 1982.

[13] N. Van Gysegham, R. D. Caluwe, and R. Vandenberghe. UFO: Uncertainty and fuzziness in an object-oriented model. In *Proc. IEEE International Conference on Fuzzy Systems*, pages 773–778, Mar. 1993.