A DATABASE MODEL FOR MEDICAL CONSULTATION

M. Anvari

Division of Computer Science
Department of Electrical Engineering and Computer Science
University of California.
Berkeley, CA

Abstract The database model presented in this paper is suitable for application in which queries may require non-crisp references to certain attributes. The data item (attribute) values may be crisp or fuzzy. For instance, such adjectives as 'high' or 'normal' may be attribute values for the attribute blood pressure. A disease or a condition can be described by a number of symtoms which may be crisp alphanumeric values or fuzzy terms such as 'high' or 'normal'. A query into this database can retrieve diseases which have 'similar' symptoms. The similarity or 'indistinguishability' is a measure defined by the database user on the relations that describe a family of diseases. This database system in conjunction with a rule base can provide the framework for a medical consultation system.

چکیده مدل پایگاه اطلاعاتی ارائه شده در این مقاله برای کاربردهایی که در آنها جستجو از طریق صفات مبهم نیز می تواند انجام پذیرد مناسب می باشد. ارزشهای داده ها خود می تواند دقیق یا مبهم باشد. بعنوان مثال صفاتی از قبیل «زیاد» یا «معمولی» می تواند ارزش داده هائی چون «فشار خون» باشد. یک بیماری یا وضعیت می تواند توسط علائم دقیق با حروف آلفانومریک و یا مبهم با عباراتی چون «زیاد» و «معمولی» توصیف گردد. جستجو دراین پایگاه اطلاعاتی می تواند به بیماریهایی که علائم «مشابهی» دارند دستیابی پیدا نماید. میزان «تشابه» یا «عدم امکان تشخیص» توسط کاربر با روابطی که یک مجموعه از بیماریها را توصیف می نماید تعریف می گردد. این پایگاه اطلاعاتی بهمراه یک پایگاه قاعده می تواند چهارچوب یک سیستم تشخیص طبی را فراهم آورد.

INTRODUCTION

Uncertainty and imprecision are two fundamental properties of human discourse. They present themselves in the description of events, facts, knowledge and beliefs. Until recently, probability theory and statistics were nearly the only tools used in the formulation of uncertainty and imprecision. The publication of Zadeh's seminal paper [1] in 1965 and subsequent extensive research and publications in fuzzy set theory, logic and mathematics have changed this situation. They have provided new paradigms for the formulation of uncertainty and imprecision to researchers in many domains of scientific and technical inquiry. This paper introduces a fuzzy relational database model which supports fuzzy queries into the database whose relations may contain crisp or fuzzy attribute values. It will also describe how this model may be utilized, as the dynamic database of facts, in an expert/consultation system.

BACKGROUND

Traditional database systems (see Appendix 1 for a technical discussion of relational database systems) cannot deal with fuzzy queries or attribute values that are expressed in fuzzy terms. In the following paragraph, the relational database model described in Figure 1 of Appendix 1 is used to dilineate the differences between the traditional and the fuzzy handling of database attributes.

A fuzzy query contains an indefinite reference to an attribute. For example (see

Figure 1 in the Appendix), 'Retrieve names of older faculty members' is an indefinite reference to Date-of-Birth. Furthermore, fuzzy set theory introduces the concept of grade of membership (denoted by a number between 0 and 1) to deal with such indefinite references as the class of 'older Faculty Members' (OFM). Consider the three records of subset FACULTY of Database Instruction: 1) Brown, born 01-01-20; 2) Smith, born 01-01-45, and 3) Jones, born 01-01-53. 1), 2) and 3) are not all members of OFM to the same degree. The theory of fuzzy sets allows us to choose the grade of membership to OFM for Brown, Smith and Jones on the basis of our common sense understanding of the concept of being an older faculty member. We all agree that Brown belongs more than Smith who belongs more than Jones to OFM. Hence, one can choose the grades of membership to be 1. 0, 0.4, and 0.0 for Brown, Smith and Jones, respectively. We need not have universal agreement about these values. For details of the theory of fuzzy sets, see [1] and [2].

Table 1 below is a re-arranged version of a table found in [3], and is an example of a database used for professional reference. It gives a differential diagnosis of common causes of inflamed eye and expresses the attribute values in non-crisp terms. Compare this table presentation with the database entitled Instruction described in Appendix 1. The first column of Table 1 lists the four causes of the inflamed eye; the second column gives the relative occurrence of each cause. Columns three to eleven are headed by the symptoms of the inflamed eye. For instance, if <discharge> is 'watery' or 'purulent', <vision> is 'blurred', <pain> is 'moderate' (see last row in Table 1 for the remaining symptoms), then corneal trauma or infection is diagnosed. There are a few points to be analyzed here:

a. The attribute values appear in the form of adjectives (e.g.clear or diffuse), with or without an adverbial modifier. (e.g. usually), expressing the presence of an abnormal condition. Attribute values may also appear in a form that would indicate the absence of a condition (e.g. 'none', indicating the absence of <pain>).

b. The attribute values in Table 1 may

Table 1. Differential Diagnosis of Common Causes of Inflamed Eye

	Incidence	Discharge	Vision	Pain	Conjunctival Infection	Cornea	Pupil Size	Pupulary Light Responses	Intraocular Pressure	Smear
Acute Conjunctivitis	Extremely common	Modrate to copious	No effect on vision	None	Diffuse more toward fornices	Clear	Normal	Normal	Normal	Normal
Acute Iritis	Common	None	Slightly blurred	Moderate	Mainly circumcorneal	Usually Clear	Small	Poor	Normal	No Organisms
Acute Glaucoma	Uncommn	None	Markedly blurred	Severe	Diffuse	Steamy	Moderately and fixed	None	Elevated oraganism	No Organisms
Corneal Trauma or Infection	Common	Watery or Purulent	Usually blurred	Moderate to severe	Diffuse	Clarity change related to cause	Normal	Normal	Normal	Organisms found only in cornea due to infection

vary over a graded range (e.g. from <pain> ='none' to <pain> = 'severe'). In contrast, the attribute values are either identical or distinct in the classical relational model, as described in Appendix 1 for the relation FACULTY; no intermediate values exist. In the diagnosis case, the attribute values 'severe', 'moderate' and 'none' (plus possibly 'moderately severe', 'very severe', and other similar expressions) provide a range of semantic descriptions of the symptom <pain>.

In addition, two patients describing the intensity of eye pain (e.g. 'moderate') do not necessarily mean the same intensity. However, this borders on the sort of polemics that is beyond the scope of this paper.

In medicine, as in many other knowledge domains, subjective and qualitative terms are widely used to express facts or represent data. Due to the imprecise nature of the knowledge, we face a challenge to store and retrieve it and to reason with it.

FORMULATION OF A FUZZY RELATIONAL MODEL

A notion of fuzziness will be superimposed on a relation such as exemplified by Table 1. We assume that the attribute values are expressed in terms of linguistic modifiers, e.g. 'diffuse' or 'very severe'. They may also be numeric or take on crisp (nonfuzzy) values such as in the relation FACULTY.

Researchers in fuzzy relational database systems have developed various paradigms to deal with uncertainty and inexactness (see [4], [5], [6], and [7]). In this paper, the notion of distinguishability is used to measure the degree to which two values of an attribute are dissimilar. The distinguishability function for attribute A is a user-defined function

$$\mathrm{dis}_A\colon \mathrm{adom}(A)\ X\ \mathrm{adom}(A) \rightarrow [0,\ 1].$$

The number 0 is assigned to $\mathrm{dis}_A(x,y)$ if the attribute values x and y are identical; the number 1 is assigned if they are clearly distinguishable; and values between 0 and 1 reflect the graded assignment of values to the distinguishability of x and y.

Thus, $\mathrm{dis}_A(x, y)$ discriminates between attribute values x and y of attribute A. For instance, if three attribute values of <pain> are 'severe', 'very severe' and 'normal', one may define

dis_{pain}(severe, very severe) = 0.3
dis_{pain}(severe, normal) = 0.7
dis_{pain}(very severe, normal) = 0.9

Certain assumptions must be made regarding the behavior of dis_A:

a. For each attribute A there exists a particular attribute value N which corresponds to the normal state or absence of a condition. The values 'normal' for attributes <pain> or <pupil size> and the value 'none' for attribute <discharge> are three such particular values for attributes <pain> and <pupil size>, respectively.

Hence, $(\mathrm{dis}_A(x, N)$ provides a measure of dissimilarity between a condition x and the normal condition N.

b. $\mathrm{dis}_A(x,y) = \mathrm{dis}_A(y,x)$. In other words, the sequence in which attribute values appear in dis is immaterial.

c. $\mathrm{dis}_A(x,\ y) =< \mathrm{dis}\ _A(x,\ z) + {}_A(z,\ y)$ for x,y,z attribute values of A. In other words, dis follows the triangle inequality.

A distinguishability function over the relation scheme R, denoted by dis_R, is derived from the distinguishability function over the attributes A_1, ..., A_n of R. The scheme by which dis_R is determined is specified by the user; however, certain choices are preferred in that they allow useful properties of database operations and functional dependencies to carry over from traditional databases to the setting we proposed here. One such simple and natural scheme is to define the distinguishability $\mathrm{dis}_R(s, t)$ of tuples s and t by

$$\mathrm{dis}_R(s,\ t) = \max\ \mathrm{dis}\ _A(s(A),\ t(A))$$

over all A in R where $s(A)$ and $t(A)$ are values of attribute A in tuples s and t, respectively. Other possibilities include

$\mathrm{dis}_A(s, t)$ = root-mean-square of $\mathrm{dis}\ _A(s(A), t(A))$

which yields the Euclidean distance between two tuples s and t and $dis_A(s,t)$ = mean $dis_A(s(A), t(A))$ over all A in R

What often occurs in diagnosis is that two 'similar', but not identical, sets of signs and symptoms in two patients are regarded by the clinician as having the same cause. In our model, this corresponds to the presence of two distinct tuples which are indistinguishable with respect to a distinguishability function dis. Two tuples s and t are said to be equal with respect to the function dis if and only if $dis_R(s, t) =< d$ for some predefined threshold value d. This form of fuzzy equality of s and t is denoted by $s = t$. The tuples s and t are identical in the ordinary sense if and only if $d = 0$.

FUZZY QUERIES

A query on a relational database involves relational operations which include Boolean operations (i.e. union, intersection, set–theoretic difference and complement) and relational operations (i.e.select, project and join). Set membership of tuples in relations in this model takes on the following form: we say that a tuple t is in the relation r within the threshold d if and only if t is distinguishable by at most d from a tuple s belonging to r. This set–membership is denoted by t In_d r. Hence we have:

$$t \ In_d \ r \text{ if and only if } t =_d s \text{ for some } s \text{ in } r.$$

The notion of set membership of tuples in a relation is the basis of all other Boolean, set–theoretic and relational operations. In our example, the diagnosis of the cause of an inflamed eye involves matching the symptoms t in a patient with a tuple in Table 1. The symptoms t in a patient are specified by attribute values expressed in terms of linguistic expressions such as 'severe', 'moderate', etc. Hence a match must be made between the patient symptoms and the tuples in Table 1. An exact match is nearly impossible: therefore, the closest tuple in the table is the one which is least distinguishable from the symptoms t. In other words, t In_d r where r is the relation consisting of the last nine columns of Table 1. Hence, a query is equivalent to attempting a diagnosis.

FUNCTIONAL DEPENDENCIES AND RULE BASES

If X and Y are two sets of attributes in a relation scheme R, then a functional dependency $X \rightarrow Y$ in the conventional sense is specified by a set X of left–side attributes and a set Y of right–side attributes. we say that relation r satisfies this functional dependency if XY (the union of X and Y) is a subset of R and

$$t_1(X) = t_2(X) \text{ implies } t_1(Y) = t_2(Y)$$

for all tuples t_1 and t_2 in r. In other words, if the left–side attribute values are equal, then so are the right–side attribute values. In the context of our fuzzy relational database model, the notion of functional dependency requires an additional structure in the form of a monotone non–decreasing function $f:[0,1] \rightarrow [0,1]$. We say that the set of attributes Y is functionally dependent on the set of attributes X in the fuzzy sense if the following occurs: XY is a subset of R and for all tuples t_1 and t_2 in the relation r and all d in [0,1], whenever

$$t_1(X) =_d t_2(X),$$

we have

$$t_1(Y) =_{f(d)} t_2(Y).$$

In other words, if the left–side attribute values $t(x)$are distinguishable by at most d, then the right–side attribute balues are distinguishable by at most $f(d)$.

This notion of fuzzy functional dependency can be utilized in defining and constructing a rule base from a relation. This is achieved by defining a mapping between the content of a relation containing a functional dependency and a set of rules. Each rule would correspond to a tuple in the relation. The antecedent and the consequent of the rule correspond, respectively, to the left–side and the right–side of the functional dependency. For further details regarding Rule–based Expert systems see references [8] and [9].

For instance, the following rule corresponds to the tuple on Table 1 whose first entry is Acute Iritis:

If 1) the inflamed eye shows no and discharge
2) the vision is slightly blurred and
3) there is moderate pain and
4) the conjectival infection is mainly circumcorneal and
5) cornea is clear and
6) pupil size is small and
7) populary light response is poor and
8) intraocular pressure is normal and
9) smear shows no organisms

Then Acrte Iritis is diagnosed.

It must be noted that sentences 1-9 forming the antecedent of the rule are expressed in fuzzy terms. Symptoms and findings of an eye patient are also expressed in fuzzy terms. A clinician attempts to diagnose an abnormal eye condition by matching its symptoms and findings (the evidence) to the medical knowledge available to him such as expressed in Table 1. When medically available knowledge (in the form of rules and facts) and symptoms and findings (in the form of patient's medical data) are expressed in fuzzy terms, which is often the case, then the mechanical matching process becomes quite complicated. A knowledege-based system must perform the matching task to determine what physiopathological condition(s) of the eye has caused the current symptoms.

We can evaluate the distinguishability measure d between the symptoms of an inflamed eye (the target) and the corresponding antecedents 1-9 of the above rule stored in the rule base. The smaller the value of d, the more 'likely' for the diagnosis to be Acute Iritis. This measure of likelihood is expressed in terms of fuzzy functional dependency. The value of f(d) represents the closeness of the patient's condition to Acute Iritis given that the set of symptoms are within distinguishability measure d of the stored antecedents 1-9.

CONCLUSION

The relational data model outlined in this paper provides a vehicle for knowledge representation and manipulation in rule based consultation systems. The advantage of this model is that it merges the facts and the rules by using the concept of functional dependencies. The application of this model is not restricted to consultation systems. In situations where a decision can be based on a set of rules and facts which embody uncertainty, this model can be utilized as well.

REFERENCES

1. L.A. Zadah, *Information and Control* 8, 338-353 (1965).
2. L.A.Zadeh, *Information Sciences* 3, 177 (1971).
3. M.A. Krupp and M.J. Chatton, (eds.), Current Medical Diagnosis and Treatment Lange Medical Publications, Los Altos,CA, (1980).
4. B.P. Buckles and F.E. Petry. *Fuzzy Sets and Systems* 7, 213, (1982).
5. B.P. Buckles and F.E. Petry, Fuzzy databases and their applications. In: M.M. Gupta, and E. Sanchez, (eds.), Fuzzy Information and Decision Processes North-Holland, Amsterdam, (1982).
6. H.B. Potoczny,*Fuzzy Sets and Systems* 12, 231, (1984).
7. M. Zemankova-Leech and A. Kandel, Fuzzy Relational Data Bases- A Key to Expert Systems. Cologne: Verlag TUV.
8. F. Hayes-Roth, et al., Building Expert Systems Addison-Wesley, New York, (1983).
9. T. O'Shea and M. Eisenstadt, Artificial Intelligence Harper and Rowe, New York, (1984).
10. E.F., Codd *CACM* 13, 377, (1970).
11. D.F. Data, An Introduction to Database Systems, 4th edition Addison-Wesley, Reading, MA, (1986).
12. D. Maier, The Theory of Relational Databases Computer Science Press, Rockville, MD,(1983).
13. C. Zaniolo and M.A. Melkanoff, *ACM Transactions on Database Systems* 6, 1, (1981).
14. E.H. Shortliffe, Computer-Based Medical Consultaion: MYCIN American Elsevier, New York, (1976).

APPENDIX 1
Relational Database Systems

The relational data model was first developed by Codd [10]. In this model, entities and relationships between them are represented by relations (also called tables or flat files); a database is a group of related relations. Figure 1 describes a model relational database named Instruction, which uses the relations FACULTY, COURSE and SCHEDULE.

COURSE

Ticket-No	Dept	Course	Units
5432	Math	101	5
6543	Math	501	3
7654	CS	131	3

SCHEDULE

FacID	Ticket-No
123	5432
345	6543

FACULTY

FacID	FName	Dept	Date-of-Birth
123	Smith	Math	01-01-45
234	Brown	Physics	01-01-20
345	Jones	Math	01-01-53

Figure 1. Database Instruction

A relation scheme consists of a certain number of attributes each of which is defined over a domain. In other words, a relation scheme is a relation without data. For instance, the relation scheme FACULTY consists of the attributes FacID, FName, Dept and Data-of-Birth.

The relation scheme FACULTY and COURSE represents the two entities Faculty and course. The relation scheme SCHEDULE represents a relationship between those two entities, i.e.'which-Faculty-Teaches-what-Courses'. An instance of an entity is represented by a tuple consisting of a certain number of attribute values. Hence, <123, Smith, Math, 1-1-45> is an instance of the entity FACULTY (a tuple in the relation Faculty). A relation is a set of tuples. For details about the relational database models, refer to database texts, e.g. [11].

We generally define a relation scheme to be a set

$$R = [A_1, \cdots, A_n]$$

of attributes $A_1, \cdots, A_n$. A value of the attribute A comes from a set dom(A) (the domain of attribute A). In the case of attribute FName, dom (FName) is the set of all possible names (string of say 15 character). A tuple over relation scheme R is a mapping $t{:}r \rightarrow D$, where D is the union for A in R of dom(A), such that t(A) is in dom(A) for each A in R. Tuple t is often represented as

$$t = \langle t(A_1), \cdots, t(A_n) \rangle.$$

The domain of the relation scheme R is the set

$$\mathrm{dom}(R) = \mathrm{dom}(A_1)\, X \cdots X\ \mathrm{dom}(A_n)$$

of all possible tuples over R. Note that this ordered set contains all possible ways that attribute values of $A_1, \cdots, A_n$ can be juxtaposed. A relation on the relation scheme R is a finite set r of tuples over R. Note also that a relation r is a small subset of dom(R). For instance, dom (FACULTY) includes tuples such as <001, aircraft, eyeglasses, 01-01-99>, which is clearly not a bona fide tuple, we therefore define the concept of active domain of attribute A relative to relation r to be the smaller set

$$\mathrm{adom}(A, r) = \{a \text{ in } \mathrm{dom}(A){:}\ a = t(A) \text{ for some } t \text{ in } r\}.$$

The active domain of relation r is the set

$$\mathrm{adom}(R,\ r) = \mathrm{adom}(A_1,\ r)\, X \quad X\ \mathrm{adom}(A_n,\ r).$$

In the case of relation scheme FACULTY, the set adom (FACULTY, Faculty)
includes every tuple in the relation Faculty and all such :
tuples as <123, Brown, Math, 01-01-53>

which is not in the database.

If $X = B_1 \cdots B_m$ is a subset of R and t is a tuple over R, then the X-value of tuple t is the m-tuple

$$t(X) = \langle (B_1), \cdots, (B_m) \rangle.$$

For instance, if $x =$ FName, Date-of-Birth, then the x value of the tuple <123, Smith, Math, 01-01-45> is < Smith, 01-01-45 >.

A database is a finite collection of relations on a set of relation schemes. Hence, the database Instruction is the set of relations Faculty, Course and Schedule (see Figure 1). For a given database, we will want to have at hand, for each attribute A, the set of all domain elements appearing as A-values in any relation of the database. We will call this set the active domain of attribute A relative to the database and denote it by 'adom (A)'. Thus, adom (A) = the union of adom(A,r) over relations r in the database. For instance

adom(Dept, Course) = {Math, CS}
adom(Dept, Faculty) = {Math, Physics} and
adom(Dept, Schedule) = empty.

Hence,

adom(Dept) = {Math, Physics, CS}.

for relation scheme $R = A_1, \cdots, A_n$, we define the active domain of relation scheme R relative to the database as the set

$$\text{adom}(R) = \text{adom}(A_1) X \cdots X \text{ adom}(A_n).$$

References [12] and [13] provide theoretical treatment of relational database systems.