A Fuzzy Medical Data Model

FRIEDRICH STEIMANN and KLAUS-PETER ADLASSNIG,
Institut für Medizinische Computerwissenschaften,
Universität Wien,
Währinger Gürtel 18-20,
A-1090 Wien,
Austria

Abstract: The total computerisation of medical records is on its way. However, information systems developed for organisational and business applications are designed to fit in rigid, schematic environments which have very little in common with medical settings; here, the time at which an observation was made and the course of parameters play an essential role, and different kinds of uncertainty including vagueness, imprecision, and ignorance are the rule rather than the exception. With FuzzyBase we present a data model designed to suit medical requirements.

INTRODUCTION

In their widest definition, medical records comprise data of most different origin, quality, and significance. Originating from patient history, physical and clinical examinations, laboratory tests, on-line sampling through bedside monitors, and other sources, medical records contain information which can be sparse or dense, precise or vague, and reliable or uncertain. If computer-assisted evaluation of such a record is to take its properties into account, one prerequisite is that the entries be self-descriptive, i.e., contain sufficient information to allow for their appropriate use.

In the following we present FuzzyBase, a functional data model that integrates uncertainty of different origin uniformly through fuzzy sets. The model is intended to provide the theoretical basis for a dedicated medical record database. Its query language is designed to serve as an interface to both users and higher-level applications such as expert systems.

THE FUNCTIONAL VIEW

In our model, the only entries allowed in a medical record are *attributes*, named properties that describe a patient by associating values with him. While the fact that a patient exhibits a certain property is perpetual, the actual value of the attribute may change and, for practical reasons, may not be observable at all times.

As attributes change with time, they indicate a course, and it is often that course rather than the absolute value which conveys information crucial to draw appropriate conclusions. To be able to reconstruct the course from values, all attributes are explicitly time-related, i.e., every attribute has the time of observation associated with it. Figure 1

appeared in: R Trappl (ed) *Proceedings of the 12th European Meeting on Cybernetics and Systems Research* (World Scientific, Singapore 1994, Vol. 1) 271–278

depicts, in graphical form, typical attribute information: while on-line monitors allow to observe temperature quasi continuously, antibody concentration, being a laboratory test result, can only be obtained at certain points in time; chest pain in turn was stated by the patient to have lasted for a certain period, then it disappeared. Other information such as pictures or multi-valued attributes, although difficult to depict, may be modelled analogously.

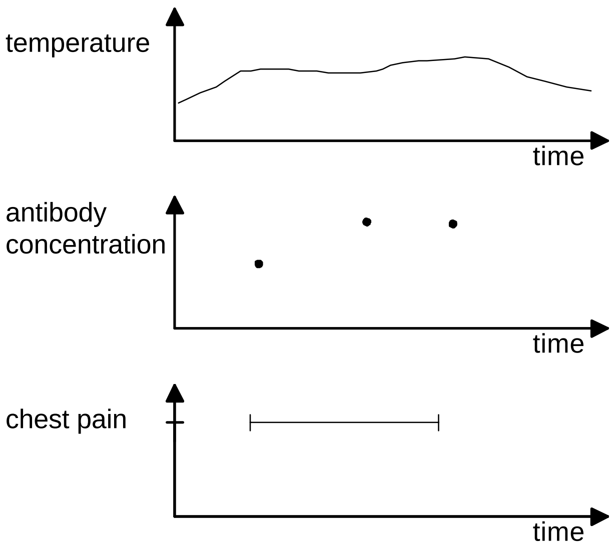

Fig. 1: Different kinds of time related information about a patient

Seen from a mathematical point of view, attributes can be modelled by a function from the patient and time space to the attribute value domain. Formally, every attribute a is then declared as a function $a : P \times T \to V_a$, where P is the set of patients, T represents time[1], and V_a is the range of the attribute. a is also often called the *attribute label*, while elements from V_a are then called *attribute values*.

Although functions model attributes naturally, their computational potential can hardly be exploited: for very few attributes only there is an actual prescription to derive their value as a function of time. For example, a patient's age can be calculated as the difference between the actual date and the date-of-birth, and in the trivial case of constant attributes such as sex the constant is all that is needed to define the function totally. In most cases, however, there is no effective model available that allows to compute the attribute value as a function of time—the system human body is just too complex. In practice the function is therefore defined in tabular form as obtained by observations made at discrete points in time, and the attribute values in between are unknown and can only be estimated with aid of heuristics.

As the basic entity of information we selected the *fact* which specifies one value at one time and is represented by an expression of the form

$$a(p,t) = v \qquad (1)$$

[1] Although different representations of time are possible, in the discourse of this paper we assume time to be represented by a totally ordered set of numerals, e.g., the real line.

appeared in: R Trappl (ed) *Proceedings of the 12th European Meeting on Cybernetics and Systems Research* (World Scientific, Singapore 1994, Vol. 1) 271–278

where a is an attribute label, p is a patient, t is a time, and v is a value. Note that by using facts an attribute can only be defined partially; for the adoption of total functions compare the discussion of this paper.

INTEGRATING UNCERTAINTY

Facts of the above kind cannot yet convey vague information: a precise attribute value is associated with a precise time. In realistic settings, however, neither time nor value are always precisely known. Instead, there is usually a certain amount of uncertainty adequately represented by a range from which the actual value is trusted to stem. As opposed to other models of uncertainty employing intervals, in FuzzyBase the boundaries of a range can be blurred rather than sharp to model a continuous transition from possible to impossible values.

Ranges of this kind are ideally modelled through fuzzy sets [1, 2]. To express that an unknown value x is known to be within a given range R represented by a fuzzy set, we write x:R.[2] x no longer stands for values of the universe, but it is restricted to members of R, i.e., $\mu_R(x) > 0$ holds.

Extending (1) with fuzzily restricted time and value then yields

$$a(p, t : T) = v : V \qquad (2)$$

meaning

$$\exists t, \mu_T(t) > 0 : \exists_1 v, \mu_V(v) > 0 : \mu_{a(p,t)=v}(p, t, v) = \min(\mu_T(t), \mu_V(v)) .$$

Note that the fuzzy set T is not interpreted as an interval throughout which a takes on value v, it rather denotes a time range at some point of which the association holds. Time and value argument do, however, have different semantics: while the existence of one value v excludes all other members of V from being value of a at the same time, v may very well be value of a at other times, too. E.g., if the temperature in the above example is assumed to be 39°C six days ago, it cannot be 38°C at the same time, although it may have remained constant for quite a while. This complies with a being a function: the same value may be associated with many times, but at no time two different values can be observed, which is natural.

Also note that (1) is equivalent to (2) if T and V are singletons (fuzzy sets with only one member) with a height of one.

Example:
A statement "temperature of patient p was elevated last week" is expressed by

$$\textit{temperature}(p, t : \textit{lastWeek}) = x : \textit{elevated},$$

where *lastWeek* and *elevated* are the labels of respective fuzzy sets. Figure 2 is a depiction of this: every pair of the time-temperature space has a grade assigned to which it complies with the statement.

---

[2] Note that fuzzy sets are also used to define possibility distributions [20] which serve a very similar purpose. We believe, however, that no detour via possibility distributions is necessary to explain fuzzy range restriction in a mathematical context.

appeared in: R Trappl (ed) *Proceedings of the 12th European Meeting on Cybernetics and Systems Research* (World Scientific, Singapore 1994, Vol. 1) 271–278

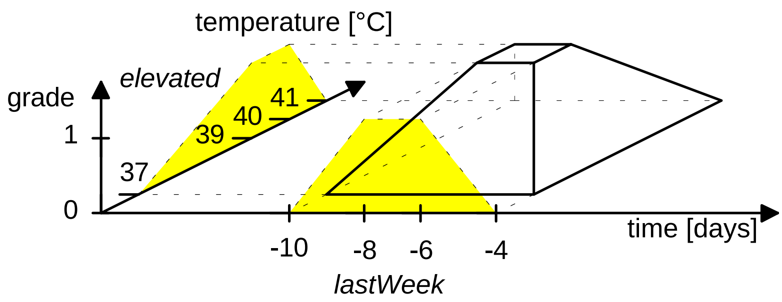

Fig. 2: Fuzzy function implementing "the temperature was elevated last week"

QUERYING

The querying mechanism of FuzzyBase is borrowed from Prolog which has gained considerable recognition as a relational database interface and query language. Because queries are syntactically identical to facts, information is simply retrieved by matching a query with the facts in the database. In short, a query and a fact match if the respective terms can be unified, i.e., made equal. Unification is achieved by substituting variables with terms, thereby restricting the set of values the variable stands for. Because in FuzzyBase variables are restricted to fuzzy sets, a special unification procedure must be employed.

In short, a query and a fact unify if and only if

- attribute label and patient identifier of query and fact are equal,
- their time ranges overlap, i.e., the intersection of the fuzzy sets representing the time ranges is non-empty, and
- the value ranges overlap.

Example:
A query "what was the temperature last week?" translates to *temperature*(*p*, *t* : *lastWeek*) = *x*, a query "when was the temperature elevated?" translates to *temperature*(*p*, *t*) = *x* : *elevated*, while "was the temperature elevated last month?" translates to *temperature*(*p*, *t* : *lastMonth*) = *x* : *elevated*. All queries unify with the fact depicted in Fig. 2.

Instead of Prolog's dual *yes/no* answer, FuzzyBase's unification returns a continuous degree of match which is defined to be the smaller of the heights of the intersecting ranges. Other degrees of match as described in the literature can also be employed. Analogously to Prolog's variable substitutions, FuzzyBase also returns the range restrictions of its query's arguments time and value. The range restrictions are an essential part of the answer: if time or value are restricted, then the rather unspecific query is being confined.

Example:
With a fact

$$temperature(p, 6{:}00) = 39$$

provided, the variables of a query

$$temperature(p,\ t : morning) = x : elevated$$

are restricted to t = 6:00 and x = 39.

Just like with PROLOG, backtracking yields alternative results in the case that more than one fact matches the query. However, unlike with PROLOG whose predicates implement relations, the functional model of FUZZYBASE imposes that no two mutually exclusive queries can be matched with the same fact. For example, taking the example of Fig. 2 as a fact, a query "was temperature above 39°C six days ago?" will yield an answer *yes* (degree of match = 1) and restrict the fuzzy set *elevated* accordingly. A subsequent query "was temperature between 37 and 38.5°C six days ago?" must be denied, as it contradicts the previous answer. If, however, it had been issued first, it would have been successful, while the second would have failed. To resolve this indeterminism, so-called assumptions are introduced: if a fact unifies with a query, its variable restrictions are memorised and applied to all subsequent queries. Without going in further detail, we state here that with this mechanism FUZZYBASE guarantees to maintain consistency with the function property of the facts.

OPEN WORLD ASSUMPTION

Often enough, an information is being queried for when there is no matching fact in the database. In PROLOG, if no match is possible, the answer is always *no*, i.e., the lack of information is interpreted as if there were evidence that it is not true (so-called *closed world assumption*). Without questioning this concept, in the clinical context missing information does not allow to draw such a general conclusion: it is very clear that the patient's attribute has a value, the problem is that it is not known. Ignorance is hence an extreme case of uncertainty where the attribute's range is restricted to V_a, i.e., not restricted at all.

We treat the lack of information as an implicit fact with the most general range restriction, so that a query to such an information always finds a fact it can unify with (open world assumption). After an according query, the (possibly restricted) fact is then temporarily stored as an assumption, and further queries to the same fact behave as if the assumption had been entered by the user. As a consequence, no contradictory assumptions about one attribute value at the same time can be made: of the two queries $temperature(p, 6{:}00) = 38$ and $temperature(p, 6{:}00) = 39$ one will always fail.

The main motivation behind this procedure is the fact that expert systems have to handle rules with large numbers of symptoms joined conjunctively in their conditions. It is then very probable that at least one of the attributes is not available, which usually means that the whole rule cannot be applied. Using FUZZYBASE as the underlying data model relieves the expert system from the meta-task of checking for the availability of attributes without trading it for the danger of inconsistencies.

It should be clear that all assumptions being made are part of the answer to a query—especially with long sequences of queries as they occur in sessions with an expert system this is to inform the user based on which assumptions the answer has been derived. Also, assumptions can interactively lead to new examinations which can then be entered into the database to make results more profound.

DERIVED ATTRIBUTES

An expert possesses knowledge suitable to draw simple conclusions from the actual findings contained in the record. Although the data model is explicitly not concerned with extending its expressiveness towards sophisticated expert systems, a strict border between such conclusions and diagnosis cannot be drawn. We decided to equip FUZZYBASE with facilities to define simple rules that implement tools every physician has at hand: arithmetic calculations, tables, graphs, and scores are examples that require no particular expertise to use. All these implement functions, mappings from a number of other (direct or derived) attribute values to a derived one, hence they fit smoothly into the functional framework of FUZZYBASE. Note that the introduction of derived attributes does not make information continuous—as long as the facts from which they are being derived are pointwisely defined, the derived attributes are also pointwise.

Example:

$$cardiacOutput(p, t) = heartRate(p, t) \times strokeVolume(p, t)$$

is a rule that allows to derive a patient's cardiac output. The time variable t is not restricted—the rule is applicable for all times. Prerequisite is, however, that both heart rate and stroke volume are specified for some overlapping ranges, and that the intersection also overlaps with the time restriction of the query.

Note how in the above rule the variable serves an additional purpose: while it is still a placeholder for an arbitrary time, it is also a symbol of co-reference, i.e., it enforces that the time range must be identical in all of its occurrences.

However, the example shows that resolution is not as simple as before: how is *elevated* in a query "was cardiac output elevated?", represented by $cardiacOutput(p, t) = elevated$, to be matched with the right-hand sidde of the rule? Certainly, the expression $heartRate(p, t) \times strokeVolume(p, t)$ could be evaluated, but then, how are the values of *heartRate* and *strokeVolume* to be restricted so that the result is *elevated* ? A detailed discussion of solutions to this problem goes beyond the scope of this paper.

DISCUSSION

As opposed to well-established general data models such as the relational, FUZZYBASE is restricted to one entity type: the patients. It is therefore not possible to model relations between patients and other entities, which is no drawback as such relations are not found in medical records. Instead, FUZZYBASE offers powerful features designed to meet medical needs: the time dependency of all attributes as well as the uniform integration of uncertain information overcome some of the major deficiencies of conventional databases employed in clinical settings.

appeared in: R Trappl (ed) *Proceedings of the 12th European Meeting on Cybernetics and Systems Research* (World Scientific, Singapore 1994, Vol. 1) 271–278

Related models are presented in [5,8,9,12,13,14]. There are several extensions of relational databases with time and uncertainty (e.g., [7,11,15–19]) and, analogously, of PROLOG [3,4,6,10], however, to our knowledge there is no uniform data model available that integrates the required features in a satisfactory manner.

Because of its deductive capabilities, it may be argued that FUZZYBASE implements an expert system—its main intent is, however, to provide a uniform interface to medical records, where standard procedures to derive attributes are integral part of the database rather than the application.

So far, FUZZYBASE's time model does only provide for pointwise definition of attribute values. However, such basic attributes as a patient's date-of-birth or sex are constant throughout the whole lifetime, while other observations are known to hold for a certain period of time, even though neither the duration nor begin or end of that period are precisely known. In a next step, the expressiveness of FUZZYBASE will therefore be extended to account for fuzzy intervals and thereby allow for definition of total functions in addition to partial ones.

REFERENCES

[1] Adlassnig, K.-P., "Fuzzy Set Theory in Medical Diagnosis", *IEEE Transactions on Systems, Man, and Cybernetics* **16** (1986) 260–265

[2] Adlassnig, K.-P., "Uniform Representation of Vagueness and Imprecision in Patient´s Medical Findings Using Fuzzy Sets", in *Proc. EMCSR 88*, Kluwer Academic Publishers, Dordrecht (1988) 685–692

[3] Baldwin, J.F. & Zhou, S.Q., "A Fuzzy Relational Inference Language", *Fuzzy Sets & Systems* **14** (1984) 155–174

[4] Brzoska, Ch., "Temporal Logic Programming with Metric and Past Operators Based on Constraint Logic Programming", *Interner Bericht* 2/93, Universität Karlsruhe, Fakultät für Informatik 1993

[5] Buneman, O.P. and Frankel, R.E., "FQL—A Functional Query Language", *Proc. 1979 ACM SIGMOD International Conference on the Management of Data* (1979)

[6] Ceri, S., Gottlob, G. & Tanca, L., *Logic Programming and Databases*, Springer-Verlag, Berlin-Heidelberg 1990

[7] Das, A.K., Tu, S.W., Purcell, G.P., and Musen, M.A., "An extended SQL for Temporal Data Management in Clinical Decision-Support Systems", *Proc. 16th Annual Symposium on Computer Applications in Medical Care* (1992) 128–132

[8] Dubois, D. & Prade, H., "Processing Fuzzy Temporal Knowledge", *IEEE Transactions on Systems, Man, and Cybernetics* **19:4** (1989) 729–744

[9] Housel, Waddle Yao, "The Functional Dependency Model for Logical Database Design", *Proc. 5th International Conference on Very Large Databases* (1979)

[10] Martin, T.P., Baldwin, J.F., & Pilsworth, B.W., "The Implementation of FPROLOG—A Fuzzy Prolog Interpreter", *Fuzzy Sets & Systems* **23** (1987) 119–129

[11] Nakajima, H., Sogoh, T., & Arao, M., "Fuzzy Database Language and Library-Fuzzy Extension to SQL", *Proc. Second IEEE International Conference on Fuzzy Systems* **1**, 477–482, San Francisco 1993

[12] Qiu, D. & Dudeck, J., "Temporal Propositions in a HIS-Oriented Expert System Shell", *Med. Inform.* **16:1** (1991) 65–76

[13] Shipman, D.W., "The Functional Data Model and the Data Language DAPLEX", *ACM Transactions on DBS*, **6:1** (1981)

[14] Shortliffe, E.H. & Buchanan, B.G., "A Model for Inexact Reasoning in Medicine", *Mathematical Biosciences* **23** (1975) 351–379

[15] Snodgrass, R., "Temporal Databases: Status and Research Directions", *ACM SIGMOD Record* **19:4** (1990) 83–89

[16] Snodgrass, R., "An Overview of the Temporal Query Language TQuel", Technical Report 92-22, Department of Computer Science, University of Arizona, Tucson 1992

[17] Snodgrass, R., "Temporal Databases", TempIS Technical Report No. 36, Department of Computer Science, University of Arizona, Tucson 1992

[18] Soo, M.D., "Bibliography on Temporal Databases", *ACM SIGMOD Record* **20:1** (1991) 14-23

[19] Tsichritz, D.C. & Lochovsky, F.H., *Data Models*, Software Series, Prentice Hall 1982

[20] Zadeh, L.A., "Fuzzy Sets as a Basis for a Theory of Possibility", *Fuzzy Sets & Systems* **1** (1978) 3–28

appeared in: R Trappl (ed) *Proceedings of the 12th European Meeting on Cybernetics and Systems Research* (World Scientific, Singapore 1994, Vol. 1) 271–278