Perspectives on the Integration of Fuzzy and Case-Based Reasoning Systems

Michael M. Richter

Perspectives on the Integration of Fuzzy and Case-Based Reasoning Systems

Michael M. Richter
University of Kaiserslautern
Center for Learning Systems and Applications (LSA)
PO-Box 3049
67653 Kaiserslautern, Germany
Email: richter@informatik.uni-kl.de

Abstract

We discuss relations and differences between fuzzy and case based reasoning methods in order to indicate possibilities for future research activities. We interpret the basic concepts of each approach in terms of the other one and discuss the computational methods in particular from an knowledge engineering point of view.

I. Preliminary Remarks

The intention of this note is to point out relations and differences between fuzzy set theory and case based reasoning, to discuss problems arising in this context and to indicate possibilities for future research activities. In fuzzy theory the notion of vaguely described sets is central. In case based reasoning the central notion of similarity deals also with some kind of vagueness. Both approaches have in common that they are not based on classical binary logic. They have raised considerable interest in various disciplines including computer science, mathematics and cognitive science and have led to many consequences and a rich theory. Both areas are also of high to practical interest. In particular fuzzy control plays an important commercial role but also case based reasoning has entered the commercial world.

There is much discussion about the various attempts to deal with these and related concepts about vagueness, uncertainty and so on. It is not intended here to contribute to these often controversial arguments but rather to discuss the relations on a computational and pragmatic level. The author also wants to avoid by all means the impression that any kind of evaluation of these areas is intended. Another observation is that both approaches have been developed relatively independently of each other. Despite the obvious analogies mentioned above not even the basic of each approach concepts (which we include in order to refer to them) have been interpreted in the terminology of the other approach. This is our starting point.

II. Basic terminology

II.1. Case-based reasoning.

We consider a set U called the universe which is equipped with some structure. A guiding example is where U is partitioned into nonempty subsets $K_1,...,K_n$. A classifier for such a partition is a mapping

$$c: U \longrightarrow \{1,...,n\}$$

which assigns to each $x \in U$ (the index of) its class. The classification problem asks for such a classifier. If we center on the task of classification where in the simplest situation we have a partition of U into two classes we can restrict our structure to (U, K, U\K).

For $x \in U$ the class of x is then denoted by K_x. This restriction is sufficient to explain many of the major concepts of CBR. A similarity measure on U is mapping

$$sim: U \times U \longrightarrow [0, 1]$$

with sim (x, x) = 1. Equivalently one can take distance measures for which d(x,x) = 0 is required. Often some additional properties are required. For simplicity we will sometimes assume the symmetry law sim (x, y) = sim (y, x). Sometimes the codomain of sim is extended to negative values, say [-1, 1]. This is, however, only of interest if negative similarities are treated different from positive ones in further arguments; we will come back to this point below.

If the objects are represented as attribute-value vectors) then the measures can sometimes be decomposed according to these attributes. A similarity measure on the domain of an attribute is called a local measure while the measure on the whole objects is called a global measure.

The most common measures are the Hamming measures H(a,b) which count the number of attributes on which a and b agree. A refinement is to introduce weights for each attribute which reflect their importance for the intended problem. Another refinement introduces a local similarity measure on the domain of each attribute; the global measure amalgamates them into a single value using again a weighted sum (or some more sophisticated function).

We identify U with the set of problems. A case is an ordered pair

$$(x, Sol)$$

where $x \in U$ and Sol is the solution of the problem denoted by x; the set of solutions (the "solution space") is denoted by Y. For the classification task Sol is just K_x and a case is of the form

$$(x, K_x).$$

In general the problem under consideration may have an arbitrary solution type. An example is when the solution is some action or decision. Certain actions are in the center of interest of fuzzy control.

One generally assumes that problems have unique solutions. Therefore it is justified two identify cases with elements of U. Then a case base is a subset $CB \subseteq U$. The intention is that the solution of the problem in question is available for each $x \in CB$. .

For $u \in U$ some $x \in CB$ is called a nearest neighbor of x:

$$NN (u, x) \text{ iff } sim (u, x) \geq sim (u, y) \text{ for all } y \in CB.$$

In our special situation the pair (CB, sim) classifies now all elements in U by

$$K_u = K_x,\ NN (u, x).$$

If the nearest neighbor is unique then the classification is unique. (CB, sim) is called a CBR-classifier; in general one calls it a problem solver. There are many variations of this, e.g. considering the k nearest neighbors etc.

II.2. Fuzzy sets.

Fuzzy subsets of A of U are denoted by $A \subseteq_f U$ and $\mu_A : U \longrightarrow [0, 1]$ is the membership function of A. A fuzzy partition of U into n fuzzy subsets is given by membership functions $\mu_1(x),\ldots,\mu_n(x)$ such that

$$\sum (\mu_i \mid 1 \leq i \leq n) = 1$$

holds for all $x \in U$. A fuzzy classifier for such a partition is mapping

$$c_f : U \longrightarrow [0, 1]^n$$

such that for

$$c_f(x) = (y_1(x),\ldots,y_n(x)) \text{ we have } \sum(y_i(x) \mid 1 \leq i \leq n) = 1.$$

One mostly considers the case where U is a finite set of m-dimensional real vectors. Then for each i

$$\sum (y_i(x) \mid x \in U) < 1$$

is also required.

Fuzzy partitions were introduced in [Ruspini 69] and fuzzy classifiers have been used in pattern recognition and cluster analysis, cf. e.g. [Bellman et al. 66].

A crisp fuzzy classifier satisfies in addition that for $c_f(x) = (y_1,...,y_n)$ for exactly one i we have $y_i = 1$ and $y_k = 0$ for $k \neq y$; hence crisp classifiers represent classifiers in the ordinary sense. Using a subset $CB \subseteq U$ of classified elements (a case-base in the CBR-terminology) and using the Euclidean norm as a distance measure one can as above define the nearest neighbor concept by

$$NN(u, x) \text{ iff } ||u - x|| \leq ||u - y|| \text{ for all } y \in CB$$

and obtain a crisp classifier.

The classical Boolean operations are replaced by real-valued functions which compute the membership function of a Boolean combination of fuzzy sets from membership values of the arguments. It is common to assume some axiomatic properties of these functions. For our further discussion it is convenient to list two examples.

a) t -norms f(x, y) (intended to compute $\mu_{A \cap B}$)

Axioms:

(T1) $f(x, y) = f(y, x)$

(T2) $f(x, f(y, z)) = f(f(x, y), z)$

(T3) $x \leq x', y \leq y' \Rightarrow f(x, y) \leq f(x', y')$

(T4) $f(x, 1) = x$

Typical t-norms are $f(x,y) = \min(x,y)$ or $f(x,y) = x \cdot y$

b) co-t-norms f(x, y) (intended to compute $\mu_{A \cup B}$).

Axioms:

(T1), (T2), (T3) and

(T4*) $f(x, 0) = x$

Typical co-t-norms are $f(x,y) = \max(x,y)$ or $f(x,y) = x+y - x \cdot y$.

Consequences are $f(x, 0) = 0$ for t-norms $f(x, 1) = 1$ for co-t-norms.

There are other fuzzy combination rules available which are fuzzy versions of general Boolean operators like different types of implication.

III. Similarities and fuzzy sets: First Observations.

For each similarity measure sim on U we define a fuzzy subset

$$SIM \subseteq_f V: = U \times U$$

by

$$\mu_{SIM} (x_1, x_2): = sim (x_1, x_2)$$

We can also associate to each $x \in U$ a fuzzy subset

$$F_x \subseteq_f U$$

by $\mu_x (y) = \mu_{F_x} (y) = sim (x,y)$ and symmetry of sim implies $\mu_x (y) = \mu_y (x)$. In this way we obtain, starting with a fixed sim, for each x some fuzzy subset view can be regarded how U is structured by sim from the viewpoint of x.

This treatment of similarity was essentially introduced in [Zadeh 71]. There similarity was treated as a fuzzy equivalence relation with the membership function μ_{SIM} from above. The transitivity law reads as

$$\mu_{SIM} (x,y) \geq sup_z (inf (\mu_{SIM} (x,z), \mu_{SIM} (z,y)).$$

The usual transitivity law is obtained by putting $\mu_{SIM} (x,y) = 1$. Also, for each x the "similarity class" (representing the fuzzy equivalence of x) was introduced as

$$\mu_{SIM,x} (y) = \mu_{SIM} (x,y).$$

Essentially the same is the treatment of fuzzy equality as a fuzzy equivalence relation

$$E(x,y) \subseteq_f V: = U \times U.$$

If one argument, say x = a, is fixed then this leads to a fuzzy set

$$E_a (y) = E(a,y) \subseteq_f U .$$

The nearest neighbor relation now reads as

$$NN (x, y) \mid \mu_x (y) \geq \mu_x (z), \forall z \in CB,$$

or equivalently

$$NN (x, y) \mid \mu_{SIM} (x, y) \geq \mu_{SIM} (x, z), \forall z \in CB$$

These considerations show that we can interpret the basic case based reasoning concepts the terms of the fuzzy set terminology. The other direction of an interpretation needs some addition. One aspect is easy. Suppose we have a fuzzy subset

$$S \subseteq_f V = U \times U$$

with $\mu_S (x, x) = 1$ then S defines obviously a similarity measure. But if we have simply a fuzzy set $K \subseteq_f U$ then we would need in addition a reference object in order to define a measure.

Such a reference object has to satisfy

$$\mu_K (x) = 1.$$

In this case we can define a similarity measure by

$$sim(x, y) = \mu_K (y)$$

Suppose there is a subset $CB \subseteq U$ such that for each $x \in CB$ we have some fuzzy subset $K_x \subseteq_f U$ with membership functions $\mu_x (y)$, for which $\mu_K (x) = 1$ holds. Then we can again define a measure on U x CB by

$$sim (y, x) = \mu_x (y), y \in U, x \in CB.$$

This is sufficient to define the nearest neighbor notion NN(y, x), $y \in U$, $x \in CB$. Suppose now U splits into two (unknown) classes K, $U\backslash K \subseteq U$ and the elements of CB are already classified. Then we may regard the fuzzy subsets

$$K_x \subseteq_f U, x \in K \text{ and } K_x \subseteq_f U, x \in U\backslash K$$

as "fuzzy versions" of K and U\K, resp. (i.e. again these are K and U\K from the viewpoint of x). Then $CB \cap K$ and $CB \cap (U\backslash K)$ can be regarded as the prototypes of K and U\K, resp., presented by CB (categories and their prototypes are discussed in section VIII).

These considerations show that the basic notions of fuzzy sets and case-based reasoning can be interpreted in terms of each other. Hence each problem and statement in one terminology can also expressed in the other terminology. This does not mean that both approaches are equivalent in each respect because they have developed individual computational models and techniques. We will now draw our attention to these aspects.

An important problem is connected with unknown values which ask either for completing the object description in an optimal way or working with incomplete descriptions. In CBR both aspects have been studied systematically; in the fuzzy approach this has not yet played a major role.

IV. Accumulation of similarities and combination of norms.

A basic task in case-based reasoning is to accumulate similarity measures. Suppose we have an attribute-value representation of objects $a = (a_1, \dots , a_n)$.

A local similarity measure sim_i then considers only the i-th coordinate of the objects. A global measure has to combine all local measures appropriately. Simple but common combination functions use weighted sums which give the so-called weighted Hamming measures.

Suppose now we represent the local measure by fuzzy sets as indicated above. We remark that although a local measure is concerned with a single attribute it can still be regarded as a measure on all of U (usual not a very good one). Strictly speaking, the combination of local measures extends the local measures to the product space.

In the terminology of fuzzy sets then a combination function applied to membership functions is needed. Suppose we want to accumulate sim_1 and sim_2 where $SIM_1 \subseteq_f V$ and $SIM_2 \subseteq_f V$ are the corresponding fuzzy sets. At a first glance the phrase "look for the accumulation of sim_1 and sim_2" would ask for the fuzzy intersection of SIM_1 and SIM_2 and would therefore require the application of some t-norm. The situation is, however, somewhat more involved which is due to the fact that it is not clear what we mean by the "accumulation of similarities". There are basically two ways to think of such an accumulation.

a) Each similarity measure contributes a nonnegative evidence to the proposition that two elements a and b are in the same class; sim (a, b) = 0 contributes just nothing and sim (a, b) = 1 guarantees for sure that a and b are in the same class.

b) Each similarity contributes also a negative evidence to the proposition that a and b are in the same class; sim (a, b) = 0 guarantees that a and b are in different classes.

If we have several measures (e.g. several local measures) then some may be of type a) and others of type b). The standard similarity measures (e.g. weighted Hamming measures and their variations) assume all type a) because then negative weight factors occur. In principal, there are no problems to accumulate measures entirely of type b) by considering also the complement. Difficulties arise if measures of type a) and type b) occur in the same context, see below

In terms of fuzzy operations we can state the following:

a) For the accumulation of sim_1, sim_2 of type a) some co-t-norm f is applied to SIM_1 and SIM_2.

b) For accumulating sim_1, sim_2 of type b) some t-norm f is applied to SIM_1 and SIM_2.

It is easy to see that at least on the present level of discussion the axioms (t1) - (T3), (T4) and (T4*) are meaningful. They do not cover, however, all situations discussed so far. If we e.g. consider fuzzy sets SIM derived from weighted Hamming measures axiom (T1) fails, i.e. conjunction and disjunction is not commutative. This means that fuzzy sets coming from very common measures in CBR need to treated in a more generalized manner.

Accumulating similarities of type a) has the consequence that even small contributions can only add something in favor for the similarity; the missing part is regarded as ignorance and not automatically as a contribution to the opposite. This point of view interprets similarities of x and y as evidences about the event that x and y belong to the same class in the sense of Dempster - Shafer. This is discussed in [Richter 95a] .

The viewpoint of measures of type b) is basically that of probability (although this is for many situations a too narrow view): The similarity of x and y represents the probability that x and y are in the same class. A consequence is that independent contributions to a similarity are combined multiplicatively. Fuzzy t-norms are in this context more related to probabilities and co-t-norms are more related to evidences

A problem arises when sim_1 is of type a) and sim_2 is of type b). At first, a nonsolvable conflict arises when $sim_1(a, b) = 1$ and $sim_2(a, b) = 0$, in this case we have a contradiction. As a consequence, the accumulation of sim_1 and sim_2 should be undefined. In evidence theory e.g. an analogous contradiction occurs if one evidence says with certainty 1 that $a \in U \setminus K$ and another evidences guarantees that $a \in K$; here again the accumulation of evidences is not defined.

Our above considerations suggest that

- sim_1 should be treated like an evidence
- sim_2 should be treated like a probability

Unfortunately, there is very little known about such an accumulation. In the PATDEX-system (see [Wess 95]) negative values for the measure have been introduced; the range of sim_{PAT} was [-1, 1]. This attempt has to be used with care, however. Suppose e.g. one uses weighted Hamming measures with negative coefficients without any restrictions. Then it can happen that sim(a,b) < 0 for a _ b which leads to the undesirable consequence sim(a,b) < sim(a,a

From the view point of fuzzy sets

- sim_1 ask for the application of a co-t-norm
- sim_2 ask for the application of a t-norm.

Here we have a related difficulty. One way is to apply one of the fuzzy versions of other Boolean operators (in particular employing negation or some of the various forms of implication)

V. Linguistic rules and cases.

In qualitative reasoning one distinguishes between quantitative and qualitative data (denoted by Quant and Qual, resp.). These are not absolute but relative notions. Usually we have an abstraction mapping

$$A: \text{Quant} \longrightarrow \text{Qual}$$

accompanied by a set of concretization mappings

$$C: \text{Qual} \longrightarrow \text{Quant}$$

which are compatible with A, i.e. A(C(x)) = x holds for all x in Qual.

There may be different qualitative levels for a fixed quantitative level. The abstraction mapping A transforms each predicate P on Quant into a predicate P_{qual} on Qual. In the context of the usual fuzzy terminology one refers to the predicates on Qual as to linguistic predicates or linguistic terms. This terminology refers furthermore to

abstraction mappings as to fuzzifications

and

concretization mappings as to defuzzifications.

A linguistic rule is simply of the form

$$P_1, P_2, ..., P_n \longrightarrow Sol$$

where the P_i are the premises and Sol is the conclusion. An important difference between qualitative reasoning and fuzzy theory now is that the qualitative predicates in

qualitative reasoning are interpreted in binary logic

while the

linguistic predicates are interpreted as fuzzy predicates.

For the intuitive meaning this is, however not very important; the expert who formulates the rules does normally not use a formalism. Technically the difference is much larger. In qualitative reasoning abstraction and concretization mappings have been developed only in an ad hoc manner while fuzzification and defuzzification mappings are treated much more systematically, in particular with respect to their computational aspects. On the other hand the success of a fuzzy system depends to a high degree on the possibility to encode the problem in numerical terms; often this takes much effort.

In fuzzy control linguistic rules are used essentially in the following way (which can be regarded as the "fuzzy problem solving method"): The premises describe the required state of the system while the conclusion describes the action to be taken. For each observation the degree of membership to the corresponding (fuzzy) premise is computed and an application of the t-norm in use computes the degree of membership of the whole observation to the premise. Using again the t-norm with the fuzzy conclusion the fuzzy set of the intended action is computed. It has to be noted here that for the premises the specific values of observations are used in order to compute the degree of their membership. The conclusion then is obtained as a fuzzy predicate. It is important to note that the membership function for the conclusion decreases when the degree of membership for the premises of the observation decreases. If several rules for the same action are present the co-t-norm used is applied for these rules in order to compute the final fuzzy set for the action in question. The final solution then is obtained by applying a defuzzification operator.

Fuzzy classification proceeds in principle in the same way. The knowledge about the classes is expressed in linguistic rules. The premises describe properties of the object to be classified and the conclusions yield properties of the class. Fuzzification transforms the premises into numerical values and defuzzification (often realized by the nearest neighbor method) finally returns a crisp classifier.

In applications of control theory such rules are iteratively used. The input values for the premises are obtained from the environment via sensors, the actions change the environment and so on. This demands high efficiency and asks therefore for simple computations.

Next we discuss a special situation which is important in fuzzy control and directly related to the CBR-approach. Suppose their are certain tuples for which the control action Sol has been correctly defined:

$$(a_1,...,a_n, Sol).$$

Such tuples may lead to linguistic rules of the form

"*If* x_1 *is almost equal to* a_1 *and ... and* x_n *is almost equal to* a_n *then the conclusion is Sol" .*

The intention is that values which are almost equal cannot be distinguished because of imprecise measurement or their difference is negligible for the intended action.

In the CBR terminology this reads as

"If x_1 *is similar to* a_1 *and ... and* x_n *is similar* a_n *then Sol",*

where “similar” is understood as “sufficiently similar”.

In fuzzy control “almost equal” is modeled as a fuzzy predicate. In particular, for each attribute A_i one looks at the subset

$$X_i = \{a|\ a \text{ occurs in one of the rules at the i-th coordinate}\}.$$

Each of these $a \in X_i$ now gives rise to a fuzzy predicate $E_{i,a}$ (x,x') on A_i of being “almost equal from the viewpoint of a” One should note that there is in general no uniform predicate E_i on A_i easily available and simple to formulate, it rather depends on the particular $a \in X_i$. It is, however, only a technical problem to merge the $E_{i,a}$(x,x') into a single E_i(x,x'). The selected $a \in X_i$ have to be chosen with care; they can be regarded as prototypical elements of E (cf. section VIII).

If there is also a fuzzy equality relation in the conclusion space we obtain for each fixed conclusion C in one of the rules a fuzzy predicate and we are back to the general situation described above. Technically this is done by extending the fuzzy equality relation to the product space

$$A_1 \times ... \times A_n \longrightarrow Y$$

where Y is the solution space (this requires, however, that the fuzzy equality relations on the different attribute sets are independent of each other).

Turning to CBR we assume that our objects are modeled as attribute-value-vectors Then a case is of the form

$$(a_1, ..., a_n, Sol)$$

where Sol is the desired conclusion (e.g. a class if we deal with classification). As already indicated such a case may be regarded as a rule:

If a_1 and...and a_n then Sol.

This rule is used as usually rules are used except:

1) The matching of actual situations is inexact; those situations match which are nearest neighbors w.r.t. the similarity measure.

2) The conclusion may be adapted (i.e. undergo a solution transformation).

The inexact match is closely related to the computation of the degree of membership to which the observations satisfy the premises. In particular, the fuzzy equality relation $E_i(x,x')$ is a local similarity measure.

The conclusion is not a fuzzy predicate but either a concrete conclusion or a solution on an abstract level; the latter would require a concretization mapping. The "fuzziness" of the conclusion is hidden in a twofold way:

1) The degree of similarity between the actual observations and the case is an indicator how reliable the conclusion is.

2) The way the solution of the case is adapted is another indication of how adequate the solution is.

VI. Background knowledge and knowledge container.

As each kind of knowledge background knowledge can be represented in an explicit or an implicit way. The implicit representation codes the knowledge in a more or less hidden way, e.g. in the terminology and the vocabulary used, in the way functions, algorithms and rules are defined etc. It depends mainly on the skill of the system designer and the user how this is done. An explicit representation requires a language with a clearly defined semantics (which should be closely related tom the intuitive understanding of the knowledge). In this language the knowledge is formulated.

In [Richter 95b] we introduced the concept of knowledge container and studied it for CBR. The knowledge containers are parts of the system in which knowledge can be represented and which can be treated independently of each other. In case-based reasoning four such containers can be identified:

1) The vocabulary
2) The similarity measure
3) The case base
4) The solution transformation.

In general CBR=systems the user can represent knowledge explicitly in at least three ways. AS an example, they are all realized in the INRECA-system (cf. [Althoff et al.95]):

(i) Exclusion rules: They describe constraints on the retrieved cases. This is some filtering; examples are restricting equalities as $x = a$ or inequalities as $x \leq a$ or $x \geq a$. These rules refine the similarity measure.

(ii) Completion rules: They specify how missing features of a case are inferred from known features of old cases in the case base. This refines again the similarity measure.

(iii) Adaptation rules: They specify the solution transformation. This describes how a solution from a similar case is adapted to a new case if the solution is not identically accepted.

The advantage of this explicit and easy to handle representation methods give the user a high flexibility to react on specific demands.

In a fuzzy system we find the following knowledge containers:

(1) The vocabulary
(2) The linguistic rules
(3) The membership functions
(4) The t-norms and the co-t-norms
(5) The defuzzification method
(6) The adaptation rules

All these containers contain compiled knowledge (i.e. has to undergo a compilation or rather coding procedure). Only (2) and to some degree (6) allow an explicit knowledge representation. Because linguistic rules have no precise semantics one has to specify, however, the remaining parts of the system in such a way that the intentions of the user are met.

VII. Optimization and Learning.

The goal of optimization in this context is to improve the performance of the system, This can be done in various ways. In principal we observe three major methods for improving performance:

(i) Control theory

(ii) Machine learning

(iii) Human interaction

Fuzzy control as a whole counts under (i) but in order to improve a fuzzy control system itself human skill has been used almost exclusively. It should also be mentioned that couplings with other methods like neural nets are possible where the fuzzy system gives the initialization and the learning takes place in the net.

In CBR in addition machine learning techniques (where we include hillclimbing and related methods) have been used in order to improve the similarity measure.

In both techniques there seems to be much space for employing learning and optimization methods.

VIII. Categories

The representation of classes by a case base and a similarity measure where decreasing similarities of elements of the case base correspond to a decreasing membership to the class has a tradition in cognitive science. Here one distinguishes between a concept and a category (cf. [Lackoff 87]). A concept in the traditional sense has a precise definition and each element either satisfies this definition and belongs to the concept or it does not. Categories need not to have definitions but have typical elements called prototypes. Such categories occur frequently in every day life and decision support systems should take this into account (what is not yet the case).

One kind of categories is of the form described above where the degree of membership is given by the similarity to the prototypes. In [Lackoff 87] it is argued that many important categories are not of this form and can in addition not be represented as fuzzy sets. Such a principal statement is probably too strong. What seems to more likely is that the fuzzy approach is not always suitable. In section V fuzzy equality was considered and mentioned that the membership function is often defined with respect to prototypical elements. So here we have an example how categories could be modeled using fuzzy sets.

In a category there is often no meaningful property which holds for all of its members. The way categories are structured is that one starts with the prototypical elements and reaches other elements by "links". Two elements are connected by a link if they have something in common with respect to a property which is considered as "important"; this is a user defined term depending on the domain and the task. Each member of the category should be reachable from the prototypical elements by a finite chain of links. To represent links one needs a

complex structure and a rich language to for all the intended aspects. This favors in principle symbolic representations over numerical encoding.
This has some resemblance in the way similarity measures are often defined: The weighted Hamming measures reflect the importance of the attributes. If properties correspond to attributes (which in principal can always be achieved) then links correspond to high similarities.
The complex structure of categories is related to problems in the area of terminology. There a term is not just simply described by a definition but rather by a structure called term record, The term record contains (on a formal or informal level) all relevant information about the term, see [Richter et al.95].

IX. A first discussion and comparison

IX.1 General observations

The techniques for fuzzy reasoning and case-based reasoning have been developed with somewhat different intentions and have therefore to some degree different merits and weakening. Our guideline is that one method favors numerical representations and methods while the other one puts more emphasis on the symbolic level. Basically numeric method result in higher efficiency while symbolic ones allow more flexibility in the system design. There is, however, not a sharp dividing line because each approach has elements of the other one too.

In fuzzy set theory one is mainly restricted to numerical entries rather than symbolic ones. This means that real numbers or n-dimensional real vectors represent the objects. For efficiency purposes the membership functions are piecewise linear in most situations. The efficiency issue comes from demands of control theory where rules have to be iterated, often under real time constraints. In addition, various (and efficient) defuzzification methods have been developed.

In CBR the intention is not to iterate the solution process but to apply it once, e.g. to a classification problem or to support a decision. These applications ask also for symbolic values and hence they play a much bigger role here. This in turn has the consequence that the demands on the accuracy usually are not as strong as when numerical values are involved.
Most of the major advantages of CBR are related to knowledge and software engineering issues as pointed out in [Maurer 96] and [Althoff et al.97].
The main point is here that software products are not regarded as static elements which remain unchanged through their life time; this is only for very dedicated

software products the case (which applies e.g. for certain systems in fuzzy control). In general such products undergo dynamic changes due to an unstable context or varying user demands. Also the reuse issue is important here where reuse is understood in non-identical situations. Two major aspects in modifying pieces of software are:

(1) Knowledge to be changed has to be represented explicitly and separated from the remaining knowledge.

(2) There is a knowledge container that carries knowledge which is only interpreted at run time.

From the discussion in section VI it follows that fuzzy systems are more of static character. In contrast CBR-systems allows more explicit knowledge representation and has the case base as a knowledge container which is interpreted at run time.

Abstract concepts in fuzzy systems occur in linguistic rules. Although the notions of similarity and abstraction have much in common their detailed interplay is somewhat delicate. A realization was obtained in the PARIS-system, see [Bergmann 96].

IX.2 Suggestions for integrating techniques

The integration of different techniques can be obtained on various levels of intensity. This reaches from hybrid systems where the individual methods are simple available to a seamless integration where the components are invisible to the user. At this point we will just indicate where and how the two techniques under consideration can supplement each other. Despite the differences listed so far there are still a number of promising possibilities for this.

(1) Improvement of fuzzy systems by CBR-techniques.

(1a) The experience in the CBR-domain can be used to extend numerical object descriptions to symbolic ones.

(1b) Fuzzy equality could be treated directly by similarity measures which allows to employ the experience with such measures.

(1c) CBR has detailed experience with missing values which could be employed in fuzzy systems.

(1d) If linguistic rules are difficult to acquire they could be replaced by cases. Such cases could also be used in machine learning procedures.

(2) Improvement of CBR-systems by fuzzy-techniques.

(2a) In applications often fuzzy predicates occur in a natural way. They are either treated using the detour via similarity measures (which is always possible, see section II) or transformed into a binary predicate. Using fuzzy membership functions this can be simplified as we will illustrate now.

(2b) Local as well as global similarity measure use often a fuzzy partition of their domain in order to fine structure the measure. This fine structure could be improved using the membership functions directly.

(2c) Object descriptions as well as attributes do not allow fuzzy notions although they are often obtained from the expert this way. As a consequence similarity measures have to be extended in order to allow numerical functions as arguments.

(2d) Also it is often appropriate to give the output as a vague predicate. This could either be interpreted by the user (e.g. the decision maker) individually or undergo some defuzzification method.

(2e) Fuzzy control itself could be used to improve similarity measures and hence supplement the machine learning methods employed so far.

Acknowledgment: The author thanks Ralph Bergmann for useful improvements. He also thanks the ICSI for the hospitality during his stay in the first two month of 1997.

Bibliography

[Althoff et al.95] K.-D.Althoff, E.Auriol, R.Barletta, M.Manago: A Review of Industrial Case-Based Reasoning Tools: Theory and Practice. AI Perspectives Report, Oxford: AI Intelligence.

[Althoff et al. 97] K.-D.Althoff, M.M.Richter, W.Wilke: In preparation.

[Bellman et al. 66] R.Bellman, R.Kalaba, L.Zadeh: Abstraction and Pattern Classification. J.Math.Anal.Appl. 13 (1966), p. 1-7.

[Bergmann 96] R.Bergmann: Effizientes Problemloesen durch flexible Wiederverwendung von Faellen auf verschiedenen Abstraktionsebenen. DISKI 138, infix-Verlag 1996.

[Lackoff 87] G.Lackoff: Women, Fire and Dangerous Things. The University of Chicago Press 1987.

[Maurer 96] F.Maurer:Current status of case-retrieval in engineering domains: an analysis from the knowledge engineering perspective. Knowledge-Based Systems 9(1996), p.83-91.

[Richter 95a] M.M.Richter: On the notion of similarity in case-based reasoning. Mathematical and Statistical Methods in Artificial Intelligence (ed. G. della Riccia et al.), Springer Verlag 1995, p. 171-184.

[Richter 95b] M.M.Richter: Invited talk at ICCBR'95, Sesimbra, Portugal 1995.

[Richter et al 95] M.M.Richter, G.Schmidt, M.Schneider: Terminology and Knowledge Representation in Complex Domains. Proc. of the GfKL, Basel 1995, Springer Verlag 1995.

[Ruspini 69] E.Ruspini: A new approach to clustering. Inform. Control 15 (1969), p. 22-38.

[Wess 95] S.Wess: Fallbasiertes Schliessen in wissensbasierten Systemen zur Entscheidungsunterstuetzung und Diagnostik. infix Verlag 1995.

[Zadeh 71]: L.Zadeh: Similarity Relations and Fuzzy Orderings. Inform. Sci. 3 (1971), p. 177-200.