In: Mathematical and Statistical Methods in Artificial Intelligence (eds. G. della Riccia, R. Kruse, R. Viertl), Springer Verlag 1995, pp. 171-184

On the Notion of Similarity in Case-Based Reasoning

Michael M. Richter
Kaiserslautern

ABSTRACT

The semantics of similarity measures is studied and reduced to the evidence theory of Dempster and Shafer. Applications are given for classification and configuration, the latter uses utility theory in addition.

In: Mathematical and Statistical Methods in Artificial Intelligence (eds. G. della Riccia, R. Kruse, R. Viertl), Springer Verlag 1995, pp. 171-184

1. INTRODUCTION

Case-Based Reasoning (CBR) is one of the areas in Artificial Intelligence which is maturing to a technology (cf. e.g.[12]). A widely accepted technology depends usually on factors of heterogeneous character including a solid theoretical foundation, a framework of appropiate terminology, well-developed engineering methods, and a large experience in practical applications. In this article, we will contribute to the clarification of the concept of similarity which is crucial for CBR. For this purpose, we will first repeat some basic facts about CBR.

CBR can deal in principle with almost unlimited types of problems. If such a type is chosen a case is a pair (P, S) where P is a problem of this type and S is a solution for P; a case base CB is a set of such cases. We assume that (P, S_1) and (P, S_2) implies $S_1 = S_2$. This means that the solutions depend functionally on the problems and allows us to identify the cases with the problems and to include situations where we have problems and no solutions.
As examples, we consider two problem classes (cf. e.g. [11], [6]):

a) Analytic problems as the classification of objects;
b) Synthetic problems as the configuration of technical devices or the design of plans.

The usual description of how CBR proceeds is:

1) Present an actual problem P_a.
2) Select a case (P, S) from the case base CB such that P and P_a are "similar".
3) Transform the solution S for P into a solution S_a for P_a.

This simple description contains already the most important aspects of CBR:

- the size and the structure of the case base
- the notion of similarity
- the retrieval problem for cases
- the notion of a solution transformation

One of the major difficulties, in particular for classification, comes from the fact that very often the problem is only partially described, i.e., we encounter a situation of incomplete information. In addition, the description may be noisy or uncertain.

For classification the solution transformation is often the identity which we will in the beginning assume here. Then the problem solving knowledge is contained in the case base and the specific similarity concept. The latter is usually given as a real valued function in order to express degrees of similarity. The literature is full of examples of such similarity measures and each running CBR system contains necessarily at least one measure. Sometimes the measure is not fixed and can be improved by a learning process, cf. e.g. the measure in PATDEX/2 (see [11]). The selection of a "similar" case from case base using a similarity measure μ is performed by applying the following principle:

In: Mathematical and Statistical Methods in Artificial Intelligence (eds. G. della Riccia, R. Kruse, R. Viertl), Springer Verlag 1995, pp. 171-184

Nearest Neighbor Principle NNP:

Given an actual problem P_a select a case (P,S) from CB such that $\mu(P_a, P)$ is maximal.

Notation: $P = NN(P_a, CB, \mu)$.

If the nearest neighbor is not uniquely defined then some selection procedure has to take place. In the sequel we will not consider such situations.

Although NNP is generally applied, it does not have a theoretical foundation like the Maximum Likelihood Principle in probability theory. In fact, for arbitrary measures μ other principles than just the nearest neighbor principle may be much more suitable.

Sometimes specific measures have a motivation coming from the problem situation, but in general the justification is just that it works quite well. To our knowledge, an attempt to give a formal semantics to similarity measures which justifies NNP has not been made. We will present an approach for specifiying such a semantics (or rather a "meaning") for similarity measures. This will result among others in an ideal measure reflecting precisely the available information when the selection has to take place. It is, however, not claimed that this is the only nor even the best approach; we rather hope to start a discussion on this topic.

2. SOME CONCEPTS FROM LOGIC

2.1 Classical Predicate Logic.

Next we will introduce some notions from logic in a way which is appropriate for our purposes.

We consider a class $\mathtt{M}$ of structures M of predicate logic,

$$M = \langle U, (R_i)_{i\in I}, (f_j)_{j\in J} \rangle$$

where U is the universe of of M, each R_i, is an n_i-place relation over U and each f_j is a partial n_j-ary function over U. Although neither U nor I or J are assumed to be fixed in $\mathtt{M}$, we require that there are U_0, I_0 and J_0 such that for all structures in $\mathtt{M}$

$$U_0 \subseteq U, I_0 \subseteq I \text{ and } J_0 \subseteq J$$

holds. This means that each model M has a reduct of the form

$$M_0 = \langle U_0, (R_i)_{i\in I_o}, (f_j)_{j\in J_o} \rangle$$

M_0 is called the nucleus of M.

In: Mathematical and Statistical Methods in Artificial Intelligence (eds. G. della Riccia, R. Kruse, R. Viertl), Springer Verlag 1995, pp. 171-184

In most of the intended applications the universe will be finite. The situations we want to investigate lead to the following kind of structures.

a) Classification:

$M = < U, R >, R \subseteq U$

This partitions U into the two sets R and U\R; using more predicates classification problems with n classes can be formulated. If the number relational of classes is a priori unknown, the signature needs to be extended in order to introduce more classes. Also R will usually be defined in terms of other relations and functions which again requires a larger signature.

b) Configuration:

M describes a (complex) technical device or a plan using relational and functional dependencies. Adding new parts to the device results in general in an extension of the universe U while adding new dependencies requires an extension of the signature.

For a) a problem is given by an element a of U and the corresponding solution is the determination of the class to which a belongs. For b) the problem is a set of conditions on the model and the solution is a description of a model which satisfies these conditions.

For each structure M, we denote the corresponding first order predicate language by L(M). L(M) is assumed to contain a constant for each $a \in U$. Furthermore we define

$$L(\mathtt{M}) = \bigcup (L(M) \mid M \in \mathtt{M})$$

We emphasize, however, that partial functions are admitted, mainly in order to cover incomplete attribute-value descriptions. In general, we do not distinguish notationally between the symbols in L(M) and the corresponding objects in M.

2.2 Alternative Model Descriptions

In most of our intended applications, the models of $\mathtt{M}$ are not presented in the terminology of predicate logic but in various other ways. We call the formalisms used for this purpose model description languages; they can be quite arbitrary.

Definition: A model description language LModD is given by

(i) A recursive set called *set of expressions*; this set will again be denoted by LModD.
(ii) A computable function Sem: LModD $\rightarrow \mathtt{p}(\mathtt{M})$ called a semantic function where $\mathtt{p}$ denotes the power set.

In: Mathematical and Statistical Methods in Artificial Intelligence (eds. G. della Riccia, R. Kruse, R. Viertl), Springer Verlag 1995, pp. 171-184

For $S \in$ LModD the set Sem(S) is called the set of models for S. Clearly Sem is a generalization of the usual semantics for predicate logic. The most important example for classification is:

The models are of the form (U, R), $R \subseteq U$

(a1) LModD = $\{ (CB, f) \mid CB \subseteq U, f: U \to CB, f\restriction CB = id\}$

The letters CB stand for "Case Base". If we assume that $R \cap CB$ is known, then Sem can be defined by reducing R to $R \cap CB$ using f:

$a \in R \;:\Leftrightarrow\; f(a) \in R.$

If the function f is only partially defined, the model given by Sem(CB, f) is not uniquely defined, and we obtain a set of models by the semantics function.

(a2) LModD = $\{(CB, \mu) \mid CB \subseteq U, \mu: U \times CB \to \mathbb{R}^+\}$

where $\mathbb{R}^+$ denotes the nonnegative reals.

μ is called a similarity measure which may meet some additional requirements. μ defines a function $f_\mu \to CB$ using the nearest neighbor principle NNP which defines a semantics as in (a1):

$f_\mu(a) := NN(a, CB, \mu)$

(b) LModD = L(M);

Syntactically, we put LModD = $L(\mathrm{M})$.

For the semantics let $\vdash_{\mathfrak{R}}$ be some deductive operator.

For $\varphi \in L(\mathrm{M})$ we define Sem(φ) by

$$(a_1,..., a_n) \in R_i \quad :\Leftrightarrow \quad \varphi \vdash_{\mathfrak{R}} R_i(a_1,..., a_n)$$

and

$$f(a_1,..., a_n) = b \quad :\Leftrightarrow \quad \varphi \vdash_{\mathfrak{R}} f(a_1,..., a_n) = b$$

Here the left sides of the equivalences take place in M while the right sides denote provability in $L(\mathrm{M})$ using $\vdash_{\mathfrak{R}}$. The operator $\vdash_{\mathfrak{R}}$ may, e.g., denote derivability from a certain rule system $\mathfrak{R}$. This operator may have access to a similarity measure μ; such operators can occur in the context of configuration.

In: Mathematical and Statistical Methods in Artificial Intelligence (eds. G. della Riccia, R. Kruse, R. Viertl), Springer Verlag 1995, pp. 171-184

3. SIMILARITY

3.1 Some similarity measures

In the sequel, we assume that our objects (i.e., the problems) $x \in U$ are given as real valued vectors $x = (x_1,..., x_n)$, $x_i \in D_i$ which, e.g., may result from an attribute-value representation; D_i is called the i-th domain. Some father of many similarity measures is the Hamming measure μ_H (for the simple case of values from {0,1}):

$$\mu_H (x, y) = n - \sum_{i=1}^{n} | x_i - y_i |$$

If the individual attributes are of different importance, weight functions are introduced:

$$\mu_H (x, y) = \sum_{i=1}^{n} g_i (x_i, y_i) ,$$

where the g_i are real valued functions. This covers also the cases of general real valued attributes and the presence of noise. A special case is where μ_H is a linear function; then we deal with weighted Hamming measures. The weights are real numbers g_i which are supposed to reflect the importance of the i-th attribute:

$$\mu_H (x, y) = \sum_{x_i = y_i} g_i$$

The Tversky-measure is much more general and of the form (see[10]):

$$\mu_T = \alpha \cdot f(A) - \beta \cdot f(B) - \gamma \cdot f(C)$$

where α, β, γ are real numbers, f a real-valued function and

$A = \{i \mid x_i = y_i\}$
$B = \{i \mid x_i = 1, y_i = 0\}$
$C = \{i \mid x_i = 0, y_i = 1\}$.

As indicated above, our objects $x \in U$ may only be partially described, i.e., the values of some x_i may be missing. This causes the serious problem to extend these measures appropriately. Additional problems arise if

- noise is present
- the values x_i are uncertain

In: Mathematical and Statistical Methods in Artificial Intelligence (eds. G. della Riccia, R. Kruse, R. Viertl), Springer Verlag 1995, pp. 171-184

- the values x_i are not independent
- a priori knowledge about the x_i is available.

There is of course the desire that the similarity measure reflects these aspects. As indicated in the introduction, we need a semantic interpretation of the real numbers which are values of the measure.

3.2 Similarity and Truth

We will deal here with problems of classification. The most striking difference between a similarity measure m and a truth evaluation function is that the first is real valued while the second is $\{0,1\}$-valued. It needs a device like the nearest neighbor principle NNP in order to obtain a binary decision from μ. There are some natural questions which arise in this context.
Suppose $a \in U$ and $x, y \in CB \subseteq U$ where x is the nearest neighbor of a in CB:

- what motivates to determine class (a): = class(x) instead class(a) := class(y)?
- which information is contained in the numerical value $\mu(a,x)$?
- which information is contained in the numbers $\mu(a,x)$ - $\mu(a,y)$ and $\mu(x,y)$?

A similarity does not only determine the nearest neighbor, it provides some additional services like:
- the elements of CB are arranged on an ordinal scale;
- this arrangement is attached with real numbers, i.e., we obtain really a cardinal scale.

We have to answer the question what the meaning of the scale is. All of the above questions are related to question what the values of μ have to do with an approximation of the truth of a statement about class(a). In other words, which precise information about the truth of the equation class(a) = class(x) is contained in the number $\mu(a,x)$? The answer to this question would assign a meaning, i.e., a semantics to μ.

3.3 Similarity and Evidence

To simplify the situation, we will assume that all attributes are independent and no a priori knowledge is present; then the only available information consists in the knowledge of some attribute values. Let $a \in U$ and $x \in CB$. If some a_i are already observed a first approach would be to define

$$\mu(a,x) = \text{Prob}((a, x) \mid \text{given observations})$$

It is, however, difficult to assign such a conditional probability in a satisfying manner if only a few attributes are observed.

A known attribute value a_i, however, is a piece of information which hints to the set

$$X_i = \{ x \in CB \mid x_i = a_i \}.$$

In: Mathematical and Statistical Methods in Artificial Intelligence (eds. G. della Riccia, R. Kruse, R. Viertl), Springer Verlag 1995, pp. 171-184

Following J. Kohlas (cf. [2], [3], [4]) this gives rise to a basic evidence measure m_i on CB, i.e., to a probability measure on the power set $\mathfrak{p}$(CB) provided we can quantify this hint on X_i by a real number g_i, $0 \le g_i \le 1$.

We define m_i by putting $m_i(X_i) = g_i$, $m_i(CB) = 1-g_i$, i.e., some evidence goes to X_i and the rest is ignorance. Therefore, the measure m_i has only two focal sets (i.e. sets with positive measure) and because no other knowledge is available, we cannot distinguish between the elements of X_i.

If more attributes values are observed the evidence measures can be accumulated using Dempster's rule (because of our independence assumption).

In general, Dempster's rule says for $X \neq \emptyset$ (cf. [1], [8]):

$$m_1 \oplus m_2(X) = \sum_{Y \cap Z = X} m_1(Y) \cdot m_2(Z) \cdot \frac{1}{1 - K}$$

$$\text{with } K = \sum_{Y \cap Z = \emptyset} m_1(Y) \cdot m_2(Z)$$

$$\text{and } m_1 \oplus m_2(\emptyset) = 0$$

For $K \neq 0$ we have conflicts, and for $K = 1$ the accumulation $m_1 \oplus m_2$ is undefined.

Now we introduce some additional notation:

Suppose $I = \{1,...,n\}$; assume $J \subseteq I$:

$X_J = \{x \in CB \mid x_i = a_i, i \in J\}$, $X_i = X_{\{i\}}$
$m_J = \bigoplus (m_i \mid i \in J)$, $m_i = m_{\{i\}}$

Note that after a series of observations the sets X_J are closed under intersections.

If $X_{J1} = X_{J2}$ for $J_1 \neq J_2$ we call it a multiplicity. Without multiplicities and conflicts, Dempster's rule simplifies and gives for $J' \subseteq J \subseteq I$

$$m_J(X_{J'}) = \prod_{i \in J'} g_i * \prod_{i \in J \setminus J'} (1-g_i)$$

In: Mathematical and Statistical Methods in Artificial Intelligence (eds. G. della Riccia, R. Kruse, R. Viertl), Springer Verlag 1995, pp. 171-184

$$= \sum_{J'' \subseteq J \setminus J'} \left(\prod_{i \in J} g_i \right) * (-1)^{|J''|} * \prod_{k \in J''} g_k$$

Also:

$$m_J(CB) = \prod_{i \in J \setminus J'} (1-g_i) = 1 - \sum_{J'' \subseteq J \setminus J'} (-1)^{|J''|} * \prod_{k \in J''} g_k$$

Some $x \in CB$ may be elements of several focal sets X. We now make the crucial assumption that each such membership contributes to the similarity of x and a according to the evidence measure of each X. This leeds to the following definition:

Definition: (i) $\nu_J(X) = \sum_{Y \supseteq X} m_J(Y)$, Y a focal set for m_J

(ii) $\nu_J(x) = \nu_J(X)$, X the minimal focal set containing $x \in U$ (which is uniquely defined).

(iii) $\mu_J^D(a,x) = \nu_J(x)$, where a is the actual case.

If noise is present, we can proceed as follows:

$X_i^{\varepsilon,\delta} = \{x \in CB \mid \varepsilon \leq |X_i - a_i| \leq \delta\}$,

$m_i^{\varepsilon,\delta}(X_i^{\varepsilon,\delta}) = g^{\varepsilon,\delta}$, $m_i^{\varepsilon,\delta}(CB) = 1 - \sum_{a,\delta} g_i^{\varepsilon,\delta}$

for $0 \leq \varepsilon < \delta \leq 1$; $g_i^{\varepsilon,\delta}$ are again real numbers.

If the source of the information for the attribute value a_i is unreliable then the g_i will also reflect this uncertainty. If more than one independent source confirms this value we can reflect this by the accumulation of evidences in the measure. We note that to our knowledge such situations have been neglected in CBR.

We call μ_J^D the Dempster (similarity) measure. If the attributes are not independent, the measure can also be defined, but it requires a more refined rule than Dempster' s rule (cf. e.g. [4]).

We obtain trivially $\nu_J(X_J) = 1$. We also have for $J' \subseteq J \subseteq I$ if no multiplicities and conflicts occur

In: Mathematical and Statistical Methods in Artificial Intelligence (eds. G. della Riccia, R. Kruse, R. Viertl), Springer Verlag 1995, pp. 171-184

$$\nu_J(X_{J'}) = \sum_{J'' \subseteq J'} \Big(\prod_{i \in J''} g_i \prod_{i \in J \setminus J''} (1-g_i)\Big)$$

Now suppose that $X_J = \emptyset$ but $X_{J'} \neq \emptyset$ for $J' = J \setminus \{ i \}$, some $i \in I$.

Neglecting renormalisation, we obtain for a minimal focal set $X_{J'}$:

$$\nu_J(X_{J'}) = \prod_{i \in J \setminus J'} (1 - g_i)$$

If $\sum_{i \in J} g_i = 1$ this gives

$$\nu_J(X_{J'}) = 1 - \sum_{i \in J \setminus J'} g_i + \sum_{i,j \in J \setminus J'} g_i g_j - \sum_{i,j,k \in J \setminus J'} g_i g_j g_k + - \ldots$$

$$= \sum_{i \in J'} g_i + \text{ terms of higher order}$$

or

$$\mu_J^D(a,x) = \sum_{x_i = a_i} g_i + \text{terms of higher order.}$$

This means that in this situation, the evidence measure coincides with a weighted Hamming measure up to a small error. Because the evidence measure is difficult to compute (cf. [5]) for the computational complexity of Dempster's rule), we obtain a good motivation for the use of Hamming measures from the viewpoint of efficiency.

If conflicts occur, a normalization has to take place, but this will not change the ordering of the neighbors of a and the cardinal scale is only changed by a constant factor.

If multiplicities are allowed, the situation is, however, not so easy.

We take an example:

$a = (1,1,1,1)$, $x = (1,1,0,0)$, $y = (0,1,1,0)$, $z = (0,0,1,1)$, $CB = \{x,y,z\}$,
$X_{12} = X_1 \cap CB = X_1 \cap X_2$

$I = J = \{1,...,4\}$

We obtain

$$\nu_I(X_{12}) = 1 + g_1 - g_3 - g_4 - g_1 (g_2 + g_3 + g_4) + g_3 g_4 + g_2 g_2 g_3 + g_2 g_2 g_4 - g_1 g_2 g_3 g_4$$

In: Mathematical and Statistical Methods in Artificial Intelligence (eds. G. della Riccia, R. Kruse, R. Viertl), Springer Verlag 1995, pp. 171-184

Our approach leads also to an indiscernilibity relation ≈ in the sense of the theory of rough sets:

$x \approx y \Leftrightarrow$ x and y are in the same focal sets.

If the available information is rich enough that the singletons are the only focal sets, then the evidence measure is a probability measure on CB. In such a situation, we have the desired formula mentioned at the beginning of 3.3.:

sim(a, b) = Prob(class(a) = class(b) | given observations)

In summary, the evidence measure μ_J^D can be seen as the measure reflecting exactly the given information. For this measure the Nearest Neighbor Priniciple is clearly justified because it is nothing than the Maximum Likelihood Principle (applied to the evidence measure m which is a probability measure on the power set of the case base). This similarity measure may in concrete situations be difficult to compute or to approximate. However, there is now a reason to employ the results of probability theory and statistics for such purposes. In practice, this will result in the design of adaption algorithms for the measure.

3.4 Evidence and Utility

In this section, we will finally sketch some aspects of similarity in the context of configuration and planning.
The notion of truth applies only partially to configuration. A configuration may or may not be correct (i.e., meet some requirements), but it may also be more or less optimal with respect to some specified preferences. There, the truth value has to be replaced by a pair

(α, β)

where α is a value measuring correctness while β measures the degree of optimality. In addition, we have also the solution transformation T which means that we have to question the semantics of (μ, T) as

$$\text{Semantics}(\mu, T) = (\alpha, \beta).$$

In order to consider a similarity measure μ, we will fix the transformation T for the rest of the paper. If we assume that T always checks for correctness, we have only to deal with the parameter β. This parameter is the form

$$\beta = f(\beta_1, \beta_2)$$

where β_1 measures the cost of T and β_2 measures the optimality of the solution. For the classification problems considered so far these costs were zero because T was the identity transformation. In the worst case, the case base contains no information and replanning takes place; then the costs are maximal.

In: Mathematical and Statistical Methods in Artificial Intelligence (eds. G. della Riccia, R. Kruse, R. Viertl), Springer Verlag 1995, pp. 171-184

Now another difference between classification and configuration enters the scenario. For classification, it is easy to check the correctness of the CB a posteriori. Therefore, CB contains only correctly classified cases, but in a case base for configuration, the cases usually will have suboptimal solutions. If a suboptimal solution is obtained from a similar case by applying T, this is not necessarily the result of an insufficent similarity or a bad transformation T, but may be entirely due to the fact that the solution of the case to which T was applied was not optimal. For technical reasons we therefore assume that all cases in CB have optimal solutions.

In the classical framework of utility theory one considers

- a set of situations $S = \{ S_i \mid i \in L\}$

- a set of actions $A = \{A_k \mid k \in K\}$

- a set of real valued utilities $\{ \mu_{ik} \mid i \in L, k \in K\}$ which measure the utility of action A_k in situation S_i.

If a probability distribution P over the set S is known, then the expected utility of A_k is

$$E_k = \sum_{i \in L} P(S_i)\, \mu_{ik}.$$

In practical situations the utility function is not given directly. What one has is usually a preference relation which implicitly defines a utility function (using the v. Neumann - Morgenstern theory).

In our framework the situations are the problems of the cases in CB, and the actions are the transformations carried out by T; the probabilities have to be replaced by evidences appropriately (see [4], [9]).

Suppose now that a is an actual problem. Using the same notation as in 3.3., we consider after the observation of some attributes (indexed by J) a minimal focal set $X \subseteq CB$ for which we have the accumulated evidence $\nu_J(X)$. We define for $x \in X \subseteq CB$

$u_{x,T}$:= utility of applying T to x where $u_{x,T}$ depends on the parameters $ß_1$,$ß_2$ introduced above.

Because all cases of X are indiscernable, it is reasonable to put

$$\mu_J^D(a, x) = \nu_J(x) \cdot \mu_{x,T} \quad \text{for } x \in X.$$

If the configuration task degenerates the classification, we obtain this as a special form of the result from 3.3.

Acknowledgement: The author is indebted to Christoph Globig for helpful discussions.

In: Mathematical and Statistical Methods in Artificial Intelligence (eds. G. della Riccia, R. Kruse, R. Viertl), Springer Verlag 1995, pp. 171-184

REFERENCES:

[1] Dempster, A.: Upper and Lower Probabilities Induced by a Multivalued Mapping, Anuals Math. Stat., 38 (1967), 325-339.

[2] Kohlas, J.: Modeling Uncertainty with Belief Functions in Numerical Models, Eur. J. of Oper. Res., 40 (1989), 377-388.

[3] Kohlas, J.: A Mathematical Theory of Hints, Inst. for Aut. and Oper. Res.,Univ. of Fribourg, Tech. Report, 173 (1990).

[4] Kohlas, J.and Monney, P.-A.: Theory of Evidence - A survey of its Mathematical Foundations, Applications and Computational Aspects, ZOR, 39 (1994), 35 - 68.

[5] Orponen, P.: Dempster's Rule of Combination is #P-complete, Artif. Intell., 44 (1990), 245 - 253.

[6] Paulokat, J. and Weß, S.: CABLAN - fallbasierte, nichtlineare, hierarchische Arbeitsplanung, in: Beiträge zum 2. Workshop AK-CBR (Ed. D. Janetzko, Th. Schult), Freiburg 1993.

[7] Richter, M. M. and Weß, S.: Similarity, Uncertainity and Case-Based Reasoning in PATDEX, in: Authoricated Reasoning. Essays in Honor of Woody Bledsoe (Ed. R. S. Boyer), Kluwer 1991, 249 - 265.

[8] Shafer, G.: A Mathematical Theory of Evidence, Princeton University Press 1967.

[9] Strat, T. M.: Decision Analysis Using Belief Functions, Int. J. Approximate Reasoning, 4 (1990), 391 - 418.

[10] Tversky, A.: Features of Similarity, Psych. Review, 84 (1977), 327 - 352.

[11] Weß, S.: PATDEX/2: Ein System zum adaptieren, fallfokussierenden Lernen in technischen Systemen, SEKI - Working Paper SW91/01, Dept. of Comp. Science, Univ. of Kaiserslautern 1991.

[12] Weß, S. and Althoff, K.D. and Richter, M. M.: Topics in Case-Based Reasoning. EWCBR '93, Selected Papers. Springer LNAI 837 (1994).