Toward a Probabilistic Formalization of Case-Based Inference

Eyke Hullermeier*
IRIT, Universite Paul Sabatier
118, route de Narbonne
31062 Toulouse, Prance

Abstract

We propose a formal framework for modelling case-based inference (CBI), which is a crucial part of the case-based reasoning (CBR) methodology. As a representation of the *similarity structure* of a system, the concept of a *similarity profile* is introduced. This concept makes it possible to formalize the CBR hypothesis that "similar problems have similar solutions" and to realize CBI in the form of constraint-based inference. In order to exploit the similarity structure more efficiently, a probabilistic generalization of the constraint-based view is developed. This formalization allows for realizing CBI in the context of *probabilistic reasoning* and *statistical inference* and, hence, makes a powerful methodological framework accessible to CBR. Within the generalized setting, a (formalized) CBR hypothesis corresponds to the assumption of a certain stochastic model, and a memory of cases can be seen as statistical data underlying the inference process. As a particular result we establish an approximate probabilistic reasoning scheme which generalizes the constraint-based approach.

1 Introduction

The approach of case-based reasoning (CBR) relies on the hypothesis that "similar problems have similar solutions." This assumption, which we subsequently refer to as the "CBR hypothesis," is the guiding principle underlying most CBR systems. Until recently, however, there have been only few attempts at formalizing this hypothesis [Dubois *et* a/., 1998; Esteva *et* a/., 1997; Plaza *et* a/., 1998] and, thus, at making an important step toward a theoretical foundation of CBR.

In this paper, we develop a formalization in which we proceed from a constraint-based interpretation of the CBR hypothesis, according to which the similarity of problems imposes a constraint on the similarity of associated solutions in form of a lower bound [Dubois *et* a/., 1998]. We then propose a generalization of this approach, according to which the similarity of problems allows for deriving conclusions about the (conditional) probability distribution of the similarity of associated solutions. A *probabilistic* formalization of the CBR hypothesis according to which "similar problems do *probably* have similar solutions" seems appropriate since it emphasizes the *heuristic* character of CBR and is suitable for modelling the "exception to the rule."

The formal model we are going to propose is not related to the complete methodological framework of case-based reasoning in the sense of the so-called "CBR cycle." Rather, we focus on CBR as *case-based inference* (CBI), which essentially corresponds to the REUSE process within the (informal) R^4 model of the CBR cycle [Aamodt and Plaza, 1994]. More precisely, we emphasize the idea of case-based reasoning as a *prediction* method [Dubois *et* a/., 1998; Faltings, 1997], which is close to the idea of *instance-based reasoning* [Aha *et* a/., 1991]. According to this point of view, the main task of CBI is to exploit past experience in the form of observed cases in order to predict or characterize the solution of a new problem. In this narrow sense, CBR may not even cover the complete process of problem solving, i.e., the process of ultimately *finding* the solution, but may only constitute the first part thereof. Typically, the characterization of the solution provided via CBI will be utilized by methods applied in subsequent stages of the overall problem solving procedure. Indeed, these subsequent stages, which roughly correspond to the REVISE part of the R^4 model, are often not directly "case-based" but make use of domain-specific knowledge or user input.

According to the point of view represented above, CBI has important aspects in common with statistical methods, and, more generally, with approaches to machine learning. An important difference, however, concerns *structural assumptions* about an underlying data generating process, which machine learning algorithms are based on, and which are generally represented in form of a *hypothesis space H.* As a simple example consider the case where H is given as a class of functions $h : X \to Y$. Each of these functions corresponds to a

*This work has been supported by a TMR research grant funded by the European Commission.

certain hypothesis which is considered for explaining observed data $\{(x_1,y_1),\ldots,(x_n,y_n)\}$. As in this example, the hypotheses *h* typically relate properties (attributes) of the *instances* $(x,y) \in X \times Y$. The CBR hypothesis, however, concerns the (supplementary) concept of *similarity,* which can be seen as a *derived* property related to *pairs* of instances. Thus, CBR makes structural assumptions about the data generating process not directly at the *system* or *instance level* but at the, say, *similarity level.* Seen from this perspective, the process of CBI should mainly take place in some kind of *similarity space* instead of *instance space* (such as, e.g., instance-based reasoning.) Consequently, a formalization of CBI, whether probabilistic or not, should proceed from this level.

The remaining part of the paper is organized as follows: Section 2 introduces the basic framework of case-based inference. CBI is realized in the form of constraint-based inference in Section 3. Section 4 is devoted to the probabilistic generalization of this approach. The paper is concluded with a summary in Section 5.

2 The Basic Framework

Within our framework, the primitive concept of a *case* is defined as a tuple consisting of a *situation* and a *result* or *outcome* associated with the situation.[1]

Definition 1 (CBI set-up). A CBI set-up is defined as a 6-tuple

$$\Sigma = \langle \mathcal{S}, \mathcal{R}, \varphi, \sigma_\mathcal{S}, \sigma_\mathcal{R}, \mathcal{M} \rangle ,$$

where *S is a countable set of situations, R is a set of results, and $\varphi : \mathcal{S} \to \mathcal{R}$ assigns results to situations. The functions $\sigma_\mathcal{S} : \mathcal{S} \times \mathcal{S} \to \mathbb{R}_{\geq 0}$ and $\sigma_\mathcal{R} : \mathcal{R} \times \mathcal{R} \to \mathbb{R}_{\geq 0}$ define similarity measures over the set of situations and the set of results, respectively. *M* is a finite memory

$$\mathcal{M} = \{\langle s_1, r_1\rangle, \langle s_2, r_2\rangle, \ldots, \langle s_n, r_n\rangle\}$$

of cases $c = \langle s, \varphi(s)\rangle \in \mathcal{S} \times \mathcal{R}$. □

We do not make particular assumptions concerning the characterization of situations or results, which will generally be marked using an *attribute-value representation.* As far as similarity measures are concerned we suppose them to be reflexive, symmetric and normalized in the sense that degrees of similarity are elements of the unit interval [0,1], where a value of 1 corresponds to perfect similarity. Observe that card$(\mathcal{S}) \leq \aleph_0$ implies the same to be true for the sets

$$D_\mathcal{S} := \{\sigma_\mathcal{S}(s,s') \mid s,s' \in \mathcal{S}\} \subset [0,1]$$

$$D_\mathcal{R} := \{\sigma_\mathcal{R}(\varphi(s),\varphi(s')) \mid s,s' \in \mathcal{S}\} \subset [0,1]$$

of actually attained similarity degrees.

Clearly, the assumption that a situation determines the associated outcome does not imply that the latter is *known* as soon as the situation is characterized. For example, let situations correspond to instances of a class of combinatorial optimization problems. Moreover, define the result associated with a situation as the (unique) optimal solution of the respective problem. Deriving this solution from the description of the problem might involve a computationally complex process. In this connection, we refer to *case-based inference* as a method supporting the overall process of problem solving by predicting the result associated with a certain situation. To this end, CBI performs according to the CBR *principle:* it exploits experience in form of precedent cases to which it "applies" background knowledge in form of the heuristic CBR hypothesis.

Definition 2 (CBI problem). A CBI problem is a tuple $\langle \Sigma, s_0 \rangle$ consisting of a CBI set-up £ and a new situation $s_0 \in \mathcal{S}$. The task is to exploit the similarity structure[2] of Σ in conjunction with observed cases in order to predict resp. characterize the result $r_0 = \varphi(s_0)$ associated with s_0. □

3 CBI as Constraint Propagation

The hypothesis of "similar situations having similar results" is interpreted in this section from a constraint-based point of view, according to which the similarity of situations constrains the similarity of the associated results (at a minimum level.)

Definition 3 (similarity profile). For a CBI set-up Σ, the function $h_\Sigma : D_\mathcal{S} \to [0,1]$ is defined by

$$h_\Sigma(x) := \inf_{s,s' \in \mathcal{S}, \sigma_\mathcal{S}(s,s')=x} \sigma_\mathcal{R}(\varphi(s),\varphi(s'))$$

and called the similarity profile of the set-up Σ. □

The similarity profile h_Σ is the "fingerprint" of the system $(\mathcal{S},\mathcal{R},\varphi)$ at the similarity level and (partly) represents the similarity structure of the set-up Σ. It can be seen as a *condensed* representation of knowledge concerning the *system structure* φ. Indeed, the domain and the range of h_Σ are one-dimensional, whereas *S* and R will generally be of higher dimension. Consequently, a hypothesis related to the similarity structure of a system will generally be less constraining than a hypothesis related to φ directly. On the other hand, a similarity profile has a relatively simple structure which facilitates the formulation, derivation, and adaptation of hypotheses.

Definition 4 (similarity hypothesis). A similarity hypothesis is identified by a function *h* : [0,1] -> [0,1] (and similarity measures $\sigma_\mathcal{S}, \sigma_\mathcal{R}$.)l The intended meaning of the hypothesis *h* (or, more precisely, the hypothesis $(h, \sigma_\mathcal{S}, \sigma_\mathcal{R})$) is that

$$(\sigma_\mathcal{S}(s,s') = x) \to (\sigma_\mathcal{R}(\varphi(s),\varphi(s')) \geq h(x)) \qquad (2)$$

holds true for all $s,s' \in \mathcal{S}$. A hypothesis *h* is called *stronger* than a hypothesis *h!* if *h!* < *h* and $h \not\leq h'$. We say that a CBI set-up Σ *satisfies* the hypothesis *h* if $h(x) \leq h_\Sigma(x)$ for all $x \in D_\mathcal{S}$. □

A similarity hypothesis *h* is thought of as an approximation of the similarity profile h and can be interpreted

[1]We prefer these slightly more general expressions to the terms "problem" and "solution."

[2]This expression will be defined formally in Section 4.

as a quantification of the CBR hypothesis for the set-up Σ. This hypothesis is often formulated as *"the more* similar two situations are, *the more* similar are the corresponding results." Returning to the concept of a similarity profile such as introduced above, it becomes obvious that this formulation implicitely makes a stronger assumption than the formulation used at the beginning of this section. Namely, it suggests the function h_Σ associated with a set-up Σ to be increasing, or at least non-decreasing. We, therefore, call h a *strict hypothesis* if it is a non-decreasing function.

Now, consider a CBI problem $\langle \Sigma, s_0 \rangle$, and suppose that Σ satisfies the hypothesis h. If the memory M contains a case (s, r) such that $s = s_0$, the outcome $r_0 = r$ can simply be retrieved. Otherwise, we can derive the following restriction:

$$r_0 \in C_h := \bigcap_{k=1,\dots,n} \mathcal{N}_{h(\sigma_S(s_0,s_k))}(r_k), \tag{3}$$

where the *α-neighborhood* of a result $r \in \mathcal{R}$ is defined as the set of all outcomes which are at least a-similar to r:

$$\mathcal{N}_\alpha(r) := \{r' \in \mathcal{R} \mid \sigma_\mathcal{R}(r,r') \geq \alpha\} \tag{4}$$

Thus, in connection with the constraint-based view, the task of CBI can be seen as one of deriving and representing the set C_h in (3), or an approximation thereof. This may become difficult if, e.g., the definition of the similarity R and, hence, the derivation of a neighborhood is complicated. The sets (4) may also become large, in which case they cannot be represented by simply enumerating their elements.

In the context of CBI it must generally be assumed that the similarity profile h_Σ of a CBI set-up Σ is unknown. Consequently, we cannot guarantee that Σ satisfies a certain hypothesis h. Nevertheless, taking for granted that h is indeed a (more or less) good approximation of h_Σ, it seems reasonable to utilize it for deriving a set C_h according to (3) as an approximation of C_{h_Σ} (while keeping the hypothetical character of h in mind.) This situation, which reflects the heuristic character of CBI as a problem solving method, is closely related to the aspect of *learning.* Within our framework, one reasonable way of realizing *case-based learning* is that of finding a good approximation of a similarity profile h_Σ. An obvious approach, for instance, is to start from a hypothesis space H and to look for the most favorable (e.g., strongest) among those hypotheses $h \in \mathcal{H}$ which are *consistent* with the memory M in the sense that (2) is satisfied for all $\langle s, \varphi(s)\rangle, \langle s', \varphi(s')\rangle \in \mathcal{M}$.

Remark 5. According to (3), the set C_h depends only on the *relation* of degrees of the (linearly ordered) similarity scales D_s and D_R, as specified by the hypothesis h. Hence, the measures σ_S and $\sigma_\mathcal{R}$ can be considered as *ordinal* concepts. □

The overall CBI process, as introduced in this section, is illustrated by Figure 1:

- In a first step, the problem $\langle \Sigma, s_0 \rangle$ is characterized at the *similarity level* by means of its *similarity structure,* consisting of the similarity profile h resp. a corresponding hypothesis h and the similarity structure z_s = Sstr(.M,so) of the (extended) memory *(M,so).* The latter can be thought of as the set of values $\{\sigma_S(s_0, s) \mid \langle s, r\rangle \in \mathcal{M}\}$. In fact, h_Σ resp. z_S can be seen as the "image" of the system $(\mathcal{S}, \mathcal{R}, \varphi)$ resp. the (extended) memory under the transformation defined by the similarity measures σ_S and $\sigma_\mathcal{R}$

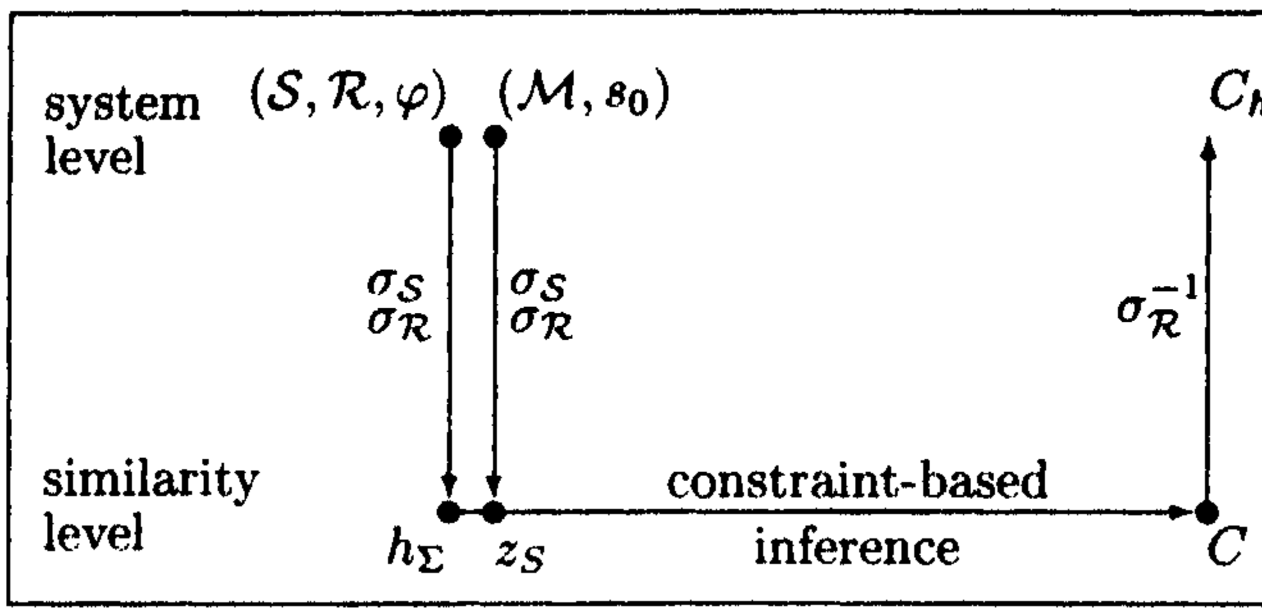

Figure 1: Illustration of the case-based inference process.

- The main step of the CBI process is then to utilize the similarity structure of the problem for constraining the unknown outcome r_0 at the similarity level. The corresponding constraints C are *implicit* in the sense that they refer to the derived (bilateral) property of similarity, not to the result itself.
- By applying the function $\sigma_\mathcal{R}^{-1} : \mathcal{R} \times [0,1] \to 2^\mathcal{R}$, which is inversely related to $\sigma_\mathcal{R}$ via $\sigma_\mathcal{R}^{-1}(r,\alpha) = \mathcal{N}_\alpha(r)$, to the observed outcomes r_k $(k = 1, \dots, n)$, the *similarity constraints* C are transformed into *constraints on outcomes,* which are combined via (3) to a constraint C_{h_Σ} resp. C_h at the system level.

Remark 6. Two characteristics of CBI, as introduced above, are worth mentioning. Firstly, CBI is *indirect* in the sense that constraints on outcomes are obtained via constraints on similarity degrees $\sigma_\mathcal{R}(r_0, r_k)$. Secondly, CBI is *local* in the sense that the rules (2) associated with a hypothesis h derive evidence concerning the value r_0 from single cases only. These pieces of evidence still have to be combined in order to obtain the constraint implied by the complete memory M. Within the deterministic framework of this section the combination of evidence is simply accomplished by (3). □

Of course, the more "convenient" the similarity structure of a set-up Σ is, the more successful CBI will be. Within our framework, we have quantified this convenience, i.e., the degree to which the CBR hypothesis holds true for the set-up Σ, by means of the similarity profile h_Σ. This quantification, however, may appear rather restrictive. The existence of some "exceptional" pairs of cases, for instance, might call for small values $h_\Sigma(x)$ of the similarity profile h_Σ in order to guarantee the validity of (3). Then, the predictions (3) which reflect the success of the CBI process might become imprecise even though the similarity structure of Σ is otherwise strongly developed. This motivates the probabilistic generalization of the constraint-based approach as proposed in the next section.

4 CBI as Probabilistic Reasoning

4.1 Basic probabilistic concepts

We now extend the definition of a CBI set-up such that the set S of situations is endowed with a probability measure μ_S (defined on 2^S.) This measure models the occurence of situations. Thus, we assume that the situations resp. the associated cases which constitute the memory M are chosen repeatedly (and independently) according to μ_S.[3] The assumption of this kind of probabilistic setting is a typical one for, e.g., machine learning, and can also be found in recent (more formally oriented) approaches to CBR [Bergmann and Wilke, 1998]. It is less restrictive than it might appear and should, therefore, not be overrated. In connection with probabilistic reasoning schemes, for instance, it will generally not be necessary to make explicit assumptions about us.

Now, consider a CBI problem $\langle \Sigma, s_0 \rangle$ with (1) the memory of the set-up Σ. According to our assumption, the sequence $(s_1, \ldots, s_n, s_0)$ of involved situations can be seen as the realization of a corresponding random sequence characterized by the measure

$$(\mu_S)^{n+1} := \underbrace{\mu_S \otimes \mu_S \otimes \ldots \otimes \mu_S}_{n+1 \text{ times}}. \quad (5)$$

This measure defines the (discrete) probability space $(\mathcal{S}^{n+1}, (\mu_S)^{n+1})$ underlying the CBI problem. Since CBR is particularly concerned with modelling the (similarity) relation between two cases let us consider a random tuple $(T, T') \in \mathcal{S} \times \mathcal{S}$ of situations. The random variable $Z = (X, Y)$, where X is the similarity $\sigma_{\mathcal{S}}(T, T')$ of the situations and Y denotes the similarity of the associated outcomes, is then defined on the probability space $(\mathcal{S} \times \mathcal{S}, \mu_S \otimes \mu_S)$ as the mapping

$$(s, s') \mapsto (\sigma_{\mathcal{S}}(s, s'), \sigma_{\mathcal{R}}(\varphi(s), \varphi(s')))$$

Let $\mu_Z := Z(\mu_S \otimes \mu_S)$ be the associated measure over $D_{\mathcal{S}} \times D_{\mathcal{R}}$, and define ux over Ds and μy over D_R in the same way. We use notations such as $(X = x)$ for events $X^{-1}(x)$ and $\mu_{Y|(X=x)}$ to denote corresponding conditional probabilities. Moreover, we make use of intuitive notations such as $P(Y = y | X = x)$ for $\mu_{Y|(X=x)}(y)$.

Definition 7 (probabilistic similarity profile). Consider a CBI set-up Σ and let $\mathcal{P}([0,1])$ denote the class of probability measures over [0, l].[4] The function

$$H_\Sigma : D_{\mathcal{S}} \to \mathcal{P}([0,1]), \, x \mapsto \mu_{Y|(X=x)}$$

is called the probabilistic similarity profile of Σ. □

The probabilistic similarity profile H provides a much more precise picture of the similarity structure of a CBI set-up Σ than a (deterministic) profile h_Σ does. For each degree of similarity $x \in D_S$ it specifies the probability distribution $\mu_{Y|(X=x)}$ of the similarity of results, i.e., of the random variable Y, given that the similarity of two situations is x. Compared to this, the function h_Σ specifies only the lower bound $\inf \{y \in D_{\mathcal{R}} \mid \mu_{Y|(X=x)}(y) > 0\}$

Definition 8 (stochastic dominance). For a probability measure $\mu \in \mathcal{P}([0,1])$ define the decumulative distribution function $G_\mu : [0,1] \to [0,1]$ by $G_\mu(x) := \mu([x,1])$. We write $\mu' \preceq \mu$ and say that the measure μ dominates the measure μ' stochastically if $G_{\mu'} \leq G_\mu$. □

Definition 9 (probabilistic hypothesis). A probabilistic! (similarity) hypothesis is identified by a function $H : [0,1] \to \mathcal{P}([0,1])$. Le Σ be a CBI set-up with probabilistic similarity profile H_Σ. Σ satisfies the hypothesis H if $H_\Sigma(x)$ dominates $H(x)$ stochastically for all $x \in D_S$. H is called *strict* if $H(x) \preceq H(x')$ for all $0 \leq x \leq x' \leq 1$. A hypothesis H' is called *stronger* than H if $H(x) \preceq H'(x)$ for all $x \in [0,1]$ and $H'(x_0) \not\preceq H(x_0)$ for at least one $x_0 \in [0,1]$ □

The $\preceq$-relation over P([0,1]) is a natural generalization of the $\leq$-relation over [0,1]. Again, a CBI set-up Σ satisfies a hypothesis H if the latter is "pessimistic" enough in the sense that it never over-estimates the probability that, for any $0 \leq \alpha \leq 1$, the degree of similarity of the outcomes associated with two situations is equal to or larger than α. The strict version of the CBR hypothesis now means that "the more similar two situations are, the larger is the probability that the associated results are at least α-similar."

Let us finally introduce the concept of a *generalized hypothesis* which turns out to be useful in connection with (approximate) probabilistic inference (cf. Section 4.2).

Definition 10 (generalized hypothesis). A generalized (similarity) hypothesis is identified by a function $G : [0,1] \to \mathcal{F}([0,1])$. Here, $\mathcal{F}([0,1])$ denotes the class of normalized uncertainty measures (fuzzy measures) over [0,1], i.e., the class of measures η such that $\eta(\emptyset) = 0$, $\eta([0,1]) = 1$, and $(A \subset B \subset [0,1]) \to (\eta(A) \leq \eta(B))$. □

In Section 4.2, a generalized hypothesis G will be used for modelling upper bounds of probability measures. Thus, it is associated with rules of the form

$$(\sigma_{\mathcal{S}}(T, T') = x) \to (\mu_{Y|(X=x)} \leq G(x)),$$

whereas a probabilistic hypothesis H defines rules

$$(\sigma_{\mathcal{S}}(T, T') = x) \to (\mu_{Y|(X=x)} = H(x)). \quad (6)$$

4.2 Probabilistic inference

It has already been mentioned that CBI is not based directly on the information provided at the system level. Rather, the concept of similarity, quantified in form of similarity functions $\sigma_{\mathcal{S}}$ and $\sigma_{\mathcal{R}}$, is exploited in order to transform this information and to represent it at the similarity level. The following definition specifies this kind of transformation.

Definition 11 (similarity structure). The similarity structure of the CBI problem $\langle \Sigma, s_0 \rangle$ is defined as the similarity profile $(H_\Sigma, \sigma_{\mathcal{S}}, \sigma_{\mathcal{R}})$ of Σ, together with the similarity structure $\mathrm{Sstr}(\mathcal{M}, s_0) := \{z_{ij} = (x_{ij}, y_{ij}) \mid 1 \leq i < j \leq n\} \cup \{x_j \mid 1 \leq j \leq n\}$ of the *extended memory* $(\mathcal{M}, s_0)$. Here, $x_{ij} := \sigma_{\mathcal{S}}(s_i, s_j)$, $y_{ij} := \sigma_{\mathcal{R}}(r_i, r_j)$, and

[3]This means that M is actually a sequence of not necessarily different cases.

[4]That is, the class of measures on $([0,1], \mathcal{B})$ with B the Borel σ-field over [0,1].

$x_j := \sigma_S(s_0, s_j)$. We generally assume H_Σ to be given and simply call Sstr(.M,so) the similarity structure of $\langle \Sigma, s_0 \rangle$. In order to take information about observed results into account, we define the *outcome structure* as Ostr$(\mathcal{M}, s_0) := \text{Sstr}(\mathcal{M}, s_0) \cup \{r_k \mid 1 \le k \le n\}$ □

The values $z_{ij} = (x_{ij}, y_{ij})$ and *Xj* define realizations of corresponding random variables $Z_{ij} = (X_{ij}, Y_{ij})$ and *Xj*. We combine these variables into one vector *Zs-* Hence, the similarity structure of a CBI problem can be thought of as a random variable *Zs* defined on the probability space $(S^{n+1}, (\mu_S)^{n+1})$. In the same way, a random variable *Zo* is associated with the outcome structure of a CBI problem.

The structure *Zs* can be seen as *statistical data at the similarity level.* The measure (5) determines the probability of the occurence of such structures completely. Thus, given the "observation" $z_S = \text{Sstr}(\mathcal{M}, s_0)$, it is principally possible to derive (conditional) probabilities such as, e.g.,

$$\mathrm{P}\,(R_0 = r \mid Z_S = z_S) \qquad (7)$$

or the *likelihood* $L(r) := \mathrm{P}\,(Z_S = z_S \mid R_0 = r)$ of results $r \in \mathcal{R}$ in connection with the unknown outcome, which is now treated as a random variable R_0. The *indirect* and *local* character of CBI (cf. Remark 6 and Figure 1,) however, has important consequences for probabilistic reasoning. Firstly, results such as (7) will not be derived directly. Rather, evidence concernig R_0 is obtained indirectly via evidence concerning similarity vectors $V = (Y_1, \ldots, Y_n) \in (D_\mathcal{R})^n$, where $Y_k = \sigma_S(R_0, r_k)$ Secondly, the *local* inference rules (6) do not allow for the direct specification of probabilities such as P(V = $v \mid Z_S = z_S$), which condition on a complete similarity structure as an event. This leads to the problem of combining probabilistic evidence derived from different cases. Due to the stochastic dependency of the random variables which constitute a similarity structure, this problem is more involved than the combination of evidence in the deterministic approach of Section 3. Thirdly, the transformation of a measure over $(D_\mathcal{R})^n$ in accordance with $\sigma_\mathcal{R}^{-1}$ does not necessarily yield a (unique) probability measure over R.

As a consequence of the characteristics just mentioned, probabilistic CBI will often be approximative. We can establish, for instance, the following result.[5]

Proposition 12. Consider a CBI problem $\langle \Sigma, s_0 \rangle$ and suppose the generalized hypothesis *G* to satisfy *G(x)* > $\bigvee_{r \in \mathcal{R}} \mu_{Y \mid (X = x, R = r)}$ for all $x \in D_S$, where $\mu \vee \mu'$ denotes the upper envelope of the measures μ and μ'. Moreover, let $\eta_x := G(x)$ for $x \in D_S$. The function

$$\overline{L}(r) := \min_{k=1,\ldots,n} \eta_{\sigma_S(s_0, s_k)}(\sigma_\mathcal{R}(r, r_k)) \qquad (8)$$

is an upper approximation of the likelihood function

$$L(r) := \mathrm{P}\,(Z_O = z_O \mid R_0 = r),$$

[6] The proof of this result can be found in the full version of this paper [Hiillermeier, 1999].

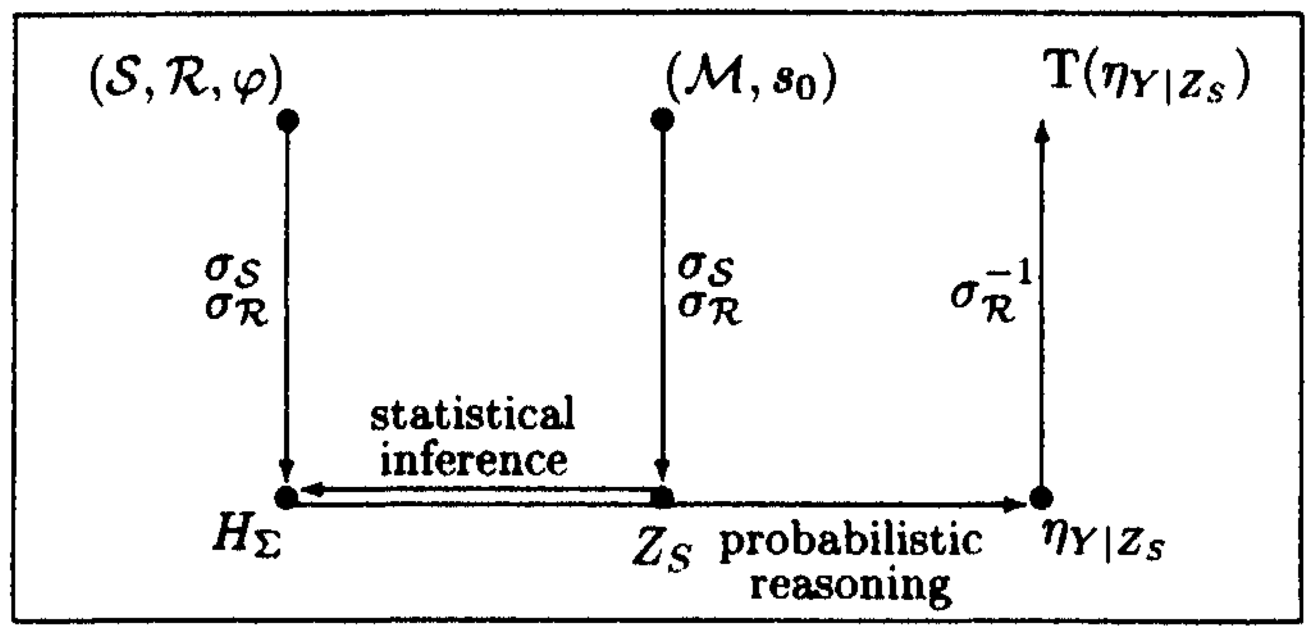

Figure 2: Illustration of the probabilistic CBI process.

$L(r) \le \overline{L}(r)$ for all $r \in \mathcal{R}$.

This proposition provides a natural generalization of the constraint-based inference scheme (3). Indeed, we recover (3) as a special case of (8) with $\eta_x|[0,1] \equiv 1_{[h(x),1]}$ Expression (8) also suggests to represent a generalized hypothesis *G* as a parametrized class of fuzzy measures such as, e.g., possibility measures and, hence, to realize probabilistic CBI by means of fuzzy set-based approximate reasoning techniques.

Figure 2 provides an overview of the probabilistic approach to CBI. Essentially, this approach realizes a process of *probabilistic reasoning in similarity space.* The similarity structure Z_S or the outcome structure Z_O of a CBI problem plays the role of the *statistical data.* A hypothesis *H* defines the *stochastic model* which explains the occurence of such structures and which underlies the reasoning process. The task of *case-based learning,* understood as the estimation of the similarity profile *H∑*, thus corresponds to *statistical inference.*

5 Summary

The main concern of this paper was to establish a framework for modelling CBI, rather than to derive formal results related to particular inference schemes. The following points deserve mentioning:

• We have introduced a formal framework in which the task of *case-based inference* has been defined as one of predicting resp. characterizing the outcome r_0 associated with a new situation S_0. The distinction between reasoning at the *instance level* and reasoning at the *similarity level* has been emphasized.

• We have adopted a *constraint-based view* of CBI, according to which the CBR hypothesis imposes constraints on the relation between the similarity of situations and the similarity of corresponding outcomes.

• The concept of a *similarity profile* has been introduced, which establishes a connection between the instance level and the similarity level, and represents the similarity structure of a CBI set-up. A *similarity hypothesis* is thought of as an approximation of the similarity profile and, hence, defines a quantification of the CBR hypothesis. It allows for realizing a constraint-based inference procedure.

• A probabilistic generalization of the constraint-based approach has been proposed. The similarity structure is now represented by means of a *probabilistic* similarity profile. It allows for replacing the constraint-based inference scheme by more general probabilistic reasoning procedures based on, e.g., the derivation of conditional probabilities or likelihood functions. An exemplary result concerning the approximation of a certain likelihood function has been established.

Finally, we would like to point out some additional aspects and make some remarks concerning related and future work.

• It should be noted that the probabilistic formalization developed here does not rely on very specific assumptions but emerges quite naturally as a generalization of the constraint-based approach in connection with the probabilistic modelling of the occurence of situations. It should also be stressed that this formalization does not correspond to a particular inference scheme. Rather, it provides the basic concepts for "translating" a CBI problem into one of probabilistic reasoning and case-based learning into statistical inference. This way, it makes the powerful methodology of statistics accessible to CBR. The investigation of particular inference schemes such as, e.g., those emerging from Proposition 12, is an important aspect of further research.

• The probabilistic point of view and the idea of (probabilistic) constraint-based inference guarantee for a clear semantics underlying our approach to CBI, in which *similarity* should be seen as an essential but at the same time *auxiliary* concept. Indeed, the inference procedure principally works with all pairs of similarity functions $\sigma_{\mathcal{S}}$ and $\sigma_{\mathcal{R}}$, each of which defines a certain similarity structure. Of course, the more suitably these functions are chosen, the better the inference results will be. The interpretation as an auxiliary concept contrasts with other formalizations of CBI [Dubois *et al,* 1998; Esteva *et al,* 1997; Plaza *et al,* 1998], in which similarity is awarded a considerable semantical meaning.

• Probabilistic models, particularly Bayesian networks, have been used in connection with case-based reasoning by several authors (see, e.g., [Aha and Chang, 1996]). As main differences between most of these approaches and our work let us emphasize two points. Firstly, the concept of similarity is often *derived* from that of probability or vice versa. In [Rodriguez *et al,* 1997], for instance, similarity is *interpreted* as a certain probability related to a classification task. Instead, probability and similarity are treated as different concepts in our model. Secondly, a probabilistic model is often related to the features (attributes) of cases (at the system level) *directly,* whereas our formalization proceeds from the similarity level. For example, the problem considered in [Rodriguez *et al,* 1997] is to estimate the (conditional) probability $\mathbf{P}(C_j \mid c)$ for a case $c = (f_1, \ldots, f_n)$ to belong to category C_j as a function of the features f_k. Seen from this perspective, such approaches are rather *case-based,* whereas our approach is *similarity-based:* inference results are derived from *similarity structures* at the similarity level instead of cases at the system level.

• The question of how to define suitable hypotheses for concrete applications is an important topic of future work. Closely related with this question is the realization of *case-based learning* as the adaptation of hypotheses. Furthermore, it seems interesting to explore the relation between probabilistic inference and (fuzzy set-based) approximate reasoning mentioned in Section 4.

• In this paper, CBI utilizes the complete memory *M.* Often, however, one will only take the most similar case(s) into account. The approach should therefore be generalized in this direction.

• The concepts developed in this paper have already been applied successfully to (repititive) combinatorial optimization problems. The thorough investigation of such applications is a further topic of future work.

References

[Aamodt and Plaza, 1994] A. Aamodt and E. Plaza. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications,* 7(l):39-59, 1994.

[Aha and Chang, 1996] D.W. Aha and L.W. Chang. Cooperative Bayesian and case-based reasoning for solving multiagent planning tasks. Technical Report AIC-96-005, Navy Center for Applied Research in AI, Naval Research Laboratory, U.S.A., 1996.

[Aha *et al,* 1991] D. Aha, D. Kibler, and M.K. Albert. Instance-based learning algorithms. *Machine Learning,* 6(l):37-66, 1991.

[Bergmann and Wilke, 1998] R. Bergmann and W. Wilke. Towards a new formal model of transformational adaptation in case-based reasoning. ECAI-98, pages 53-57, 1998.

[Dubois *et al,* 1998] D. Dubois, H. Prade, F. Esteva, P. Garcia, L. Godo, and R.L. de Mantaras. Fuzzy set modelling in case-based reasoning. *International Journal of Intelligent Systems,* 13:345-373, 1998.

[Esteva *et al,* 1997] F. Esteva, P. Garcia, L. Godo, and R. Rodriguez. A modal account of similarity-based reasoning. *International Journal of Approximate Reasoning,* 16:235 260, 1997.

[Faltings, 1997] B. Faltings. Probabilistic indexing for case-based prediction. In D.B. Leake and E. Plaza, editors, ICCBR-97, pages 611 622. 1997.

[Hiillermeier, 1999] E. Hiillermeier. A probabilistic approach to case-based inference. Technical Report 99-02 R, IRIT - Institut de Recherche en Informatique de Toulouse, Universite Paul Sabatier, January 1999.

[Plaza *et al,* 1998] E. Plaza, F. Esteva, P. Garcia, L. Godo, and R.L. de Mantaras. A logical approach to case-based reasoning using fuzzy similarity relations. *Journal of Information Sciences,* 106:105-122, 1998.

[Rodriguez *et al,* 1997] A.F. Rodriguez, S. Vandera, and L.E. Sucar. A probabilistic model for case-based reasoning. In D.B. Leake and E. Plaza, editors, ICCBR-97, pages 623 632, 1997.