To appear in *Information Sciences* (1997)

A Logical Approach to Case-Based Reasoning Using Fuzzy Similarity Relations ⋆

E. Plaza, F. Esteva, P. Garcia, L. Godo, R. López de Màntaras

IIIA – Institut d'Investigació en Intel·ligència Artificial
CSIC – Spanish Council for Scientific Research
Campus UAB, 08193 Bellaterra, Catalonia, Spain.
Vox: +34-3-5809570, Fax: +34-3-5809661
{enric | esteva | pere | godo | mantaras}@iiia.csic.es
http://www.iiia.csic.es

Abstract

This article approaches the formalization of inference in Case-based Reasoning (CBR) systems. CBR systems infer solutions of new problems on the basis of a precedent case that is, to some extent, similar to the current problem. Using the logics developed for similarity-based inference we characterize CBR systems defining what we call the Precedent-based Plausible Reasoning ($\mathcal{PPR}$) model. This model is based on the graded consequence relations named approximation entailment and proximity entailment. A modal interpretation is provided for the precedent-based inference where the plausibility is given by the graded possibility operator $\Diamond_\alpha$. The $\mathcal{PPR}$ model shows that both knowledge-intensive CBR systems and nearest neighbor algorithms share a common core formalism and that their difference is on whether or not (respectively) they use a general theory in addition to the precedent cases.

1 Introduction

Case-based reasoning (CBR) systems provide an interesting approach to integrate problem solving and learning in a variety of domains, and there is nowadays an increasing number of fielded applications and specialized software products [1]. However, no formal models exist for the basic inference

⋆ This research has been supported by the ANALOG Project CICYT 122-93 and the cooperation project CNRS-CSIC "Similarity-based Reasoning and Applications".

step performed in CBR: namely the inferences regarding a current problem that are based upon the similarity assessed between the current problem and a previously solved problem—what we will call the *precedent case*. Theoretical and empirical work has focused on the properties of specific measures for assessing similarity, on algorithms for learning to tune similarity measures, and on empirical comparison of different similarity measures empirically on a variety of domains. However, there is a lack of a formal model of the logical *warranty* that a precedent case and a similarity measure may offer to endorse inferences made upon a new problem case. In this paper we develop one such a model, called $\mathcal{PPR}$ for Precedent-based Plausible Reasoning.

1.1 Background

There are several approaches (like CBR systems and approximate reasoning systems) that share the following assumption.

Analogical Assumption Whenever a description D_1 is close to a description D_2 we can assume that what we can infer for D_2 is close to being true for D_1.

A direct way to characterize CBR systems is saying that they performs intradomain analogy [1]. That is to say, the basic assumption underlying CBR is that if our current problem C is similar to a precedent case P_i then the solution of C is similar to the (known) solution of P_i—or, alternatively, that it is plausible to assume that the solution of C is the solution of P_i. More formally:

Schema of Analogy If $D(P_i)$ is a description of a precedent case P_i and $S(P_i)$ is the solution of P_i, then if the description $D(C)$ of a current problem C is *similar* to $D(P_i)$ we can infer that $S(C)$, the solution of C, is *similar* (or *close*) to $S(P_i)$:

$$\frac{\begin{array}{c} D(P_i), S(P_i) \\ D(C) \sim D(P_i) \end{array}}{S(C) \sim S(P_i)}\,\text{analogy}$$

We will see that, in the $\mathcal{PPR}$ model, a precedent case will be modeled not as a conjunction $D(P_i) \wedge S(P_i)$ but as a certain kind of conditional $D(P_i) \Rightarrow S(P_i)$.

On the other hand, several approaches to approximate reasoning have addressed other inference rules based on the notion of *similarity* [7]. Graded

[1] Interdomain analogy is performed among cases pertaining to different domains—e. g. the analogy between electric circuits and hydraulic circuits.

consequence relations [5,6] have been proposed to model generalized modus ponens inference patterns like the following :

> "if A *extrapolatively* entails B, and we observe A', then, to some extent, it is plausible to conclude B whenever A' is *close* enough to A",

The term *extrapolative* entailment has to be understood as indicating that not only B follows from A but also propositions semantically close to B can be considered as approximate consequences of propositions semantically close to A. This kind of patterns has been the focus of research in the field of fuzzy logic, where, in general, the statement "if A extrapolatively entails B" has been modeled as a fuzzy rule whereas the facts A, B and A' are modeled as fuzzy sets (see for instance [18]). However, terms like "approximate" or " close", although fuzzy, denote notions of resemblance or proximity among propositions which may not be necessarily fuzzy. One way of proceeding is to equip the set of interpretations or possible worlds Ω with a fuzzy similarity relation $\mathcal{S}$, i.e. a reflexive, symmetric and t-norm transitive fuzzy relation ([17]). This kind of approach was introduced by Ruspini ([16]) who proposed a similarity-based semantics for fuzzy logic. Given a similarity relation $\mathcal{S}$ on the set Ω of interpretations of a boolean language $\mathcal{L}$, Ruspini proposed two measures, the *implication* and *consistency measures*, to account for the degree with which a proposition B is an approximate consequence of, or is consistent with, another proposition A, respectively. Namely,

$$I_{\mathcal{S}}(B \mid A) = Inf_{w_1 \models A} Sup_{w_2 \models B} \mathcal{S}(w_1, w_2)$$
$$C_{\mathcal{S}}(B \mid A) = Sup_{w_1 \models A} Sup_{w_2 \models B} \mathcal{S}(w_1, w_2)$$

where $\models$ denotes classical (crisp) entailment. This framework has been recently extended in [11] and [9] and compared to the possibilistic approach in [8] and [10]. See also [13] for another approach to similarity-based reasoning.

1.2 Proposal

We intend to provide a similarity-based semantics for CBR inference based on the proposals developed for fuzzy logic and approximate reasoning. The formal model we propose, called $\mathcal{PPR}$ for Precedent-based Plausible Reasoning, focuses on the core aspects of CBR systems, namely inference warranted by similarity to a precedent case. However, $\mathcal{PPR}$ is not intended as a formal model for the whole "CBR cycle" present in different CBR systems. The whole CBR cycle has been described in [1] as composed of four processes:

(i) RETRIEVE the most similar case or cases

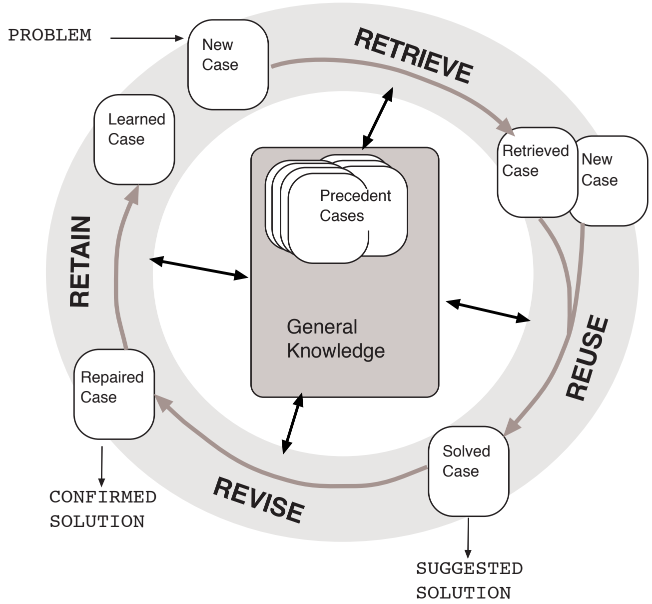

Fig. 1. The CBR cycle as described by [1].

(ii) REUSE the information and knowledge in that case to solve the problem
(iii) REVISE the proposed solution
(iv) RETAIN the parts of this experience likely to be useful for future problem solving

In figure 1, this cycle—called the R^4 model—is illustrated. Clearly similarity-based inference is involved in the RETRIEVE process, and is not necessarily involved in the REVISE and RETAIN processes. However, as we will show, the domain (or background) knowledge involved in the REUSE process will also be part of the $\mathcal{PPR}$ formal model.

As indicated in Fig. 1, general knowledge usually plays a part in this cycle, by supporting the CBR processes. This support may range from very weak (or none) to very strong, depending on the type of CBR method. By *general knowledge* we mean here general domain-dependent knowledge, as opposed to specific knowledge embodied by cases. For example, in diagnosing a patient by retrieving and reusing the case of a previous patient, an anatomical model together with causal relationships between pathological states may constitute the general knowledge used by a CBR system. A set of rules may play the same role.

The $\mathcal{PPR}$ model can be embedded in a wider model of CBR systems by means of biases. Biases are design decisions of a CBR system, usually based on knowledge about the application domain of the system. Biases express constraints imposed—or simplifications assumed—on the basic similarity-based inference model. Examples of this biases are: whether the CBR system is able to select one precedent or multiple precedents, whether the CBR system should infer a unique outcome or may deal with multiple plausible outcomes, etc. [14].

In summary, the intended interpretation of CBR inference in the $\mathcal{PPR}$ model is that of a plausible inference based on precedent cases. Indeed, the plausible inference will have a graded form, modeled in a graded modal logic framework, where the degree of plausibility is based on the degree of similarity between the current case and the precedent case.

The structure of the paper is as follows. First the models of similarity-based inference will be presented. Then, we will apply these formal models to CBR inference. We will show that our model allows to formally distinguish full-fledged CBR systems, that use precedents plus general domain knowledge, from CBR systems that use only precedent cases—so-called *instance-based* or *nearest-neighbour* algorithms. The paper closes with a discussion of the approach presented.

2 Models of similarity-based inference

The main goal of this section is twofold: to model analogical inference patterns like the ones presented in the introduction and to model what we understand by "close to" and "extrapolatively entails" as relations between classical (i.e. non-fuzzy) propositions. Both relations will be modeled through fuzzy similarity relations on the set of interpretations or possible worlds. Namely, if $\mathcal{L}$ is a classical propositional language (built upon a finite set of propositional variables), Ω is its corresponding (finite) set of interpretations or possible worlds, and $\otimes$ is a t-norm [2], then the intended modeling is built through a $\otimes$-similarity relation on Ω, i.e. a mapping $\mathcal{S} : \Omega \times \Omega \to [0,1]$ satisfying the separating ($\mathcal{S}(w,w') = 1$ iff $w = w'$), symmetric ($S(w,w') = S(w',w)$) and $\otimes$-transitive ($\mathcal{S}(w,w') \otimes \mathcal{S}(w',w'') \leq \mathcal{S}(w,w'')$) properties. In CBR systems this similarity is usually given with respect to every property or characteristic defining the cases. The issue of aggregating these similarities to obtain a global similarity is out of the scope of this paper [3].

In the rest of this section we will introduce two kinds of similarity-based consequence relations necessary for the $\mathcal{PPR}$ model of CBR, namely *approximate entailment* and *proximity entailment* (see [5,6] for further details). The first one intends to model the relation "to be close to" while the second one pro-

[2] A t-norm $\otimes$ is a binary operation in [0,1] which is associative, commutative, non–decreasing in both variables, and having 1 and 0 as neutral and absorbent elements respectively.

[3] You can find in [11] a study about the relation between a similarity in a product space and its projections in the case where the characteristics are equally relevant.

vides a logical account for an "extrapolative" conditional relationship. The approximate entailment will be used to model how close (or similar) is a current problem to a precedent case in CBR. The proximity entailment will be used to model to which extent the solution of a precedent case is a plausible solution for a current problem. The last subsection presents an interpretation of both entailments in a common modal framework which allows to describe inference in CBR systems as deduction in this logical setting.

2.1 Approximate entailment

Having a similarity relation $\mathcal{S}$ on Ω allows us to associate to any classical proposition p a fuzzy set p^* representing the approximation of p. This fuzzy set has the characteristic function $\mu_{p^*}(w) = Sup_{w' \models p}\mathcal{S}(w, w')$ defined on the set of interpretations W. The α-cuts of this fuzzy set,

$$[p^*]_\alpha = \{w \mid \mu_{p^*}(w) \geq \alpha\}$$

represent a nested family of approximations of the proposition p, i.e. $[p^*]_\alpha$ represents the set of worlds which are similar (or close) to some p–world at least to the degree α. Notice that if w belongs to the set $[p]$ of worlds that make p true, then $\mu_{p^*}(w) = 1$. From now on we will assume that a similarity $\mathcal{S}$ is given. Then it is natural to define that p α–approximately entails q whenever the set of p–worlds is in the α–approximation of q.

Definition 1 *A proposition p* approximately entails *a proposition q with degree α, written $p \models^\alpha_\mathcal{S} q$, if and only if each p-world belongs to the α-approximation of q, i.e. $p \models^\alpha_\mathcal{S} q$ iff $[p] \subseteq [q^*]_\alpha$.*

In the following we will omit the subscript $\mathcal{S}$ in the $\models^\alpha_\mathcal{S}$ notation when there is no ambiguity. When $\alpha > 0$, $p \models^\alpha q$ means that when p is true, q is close to being true, or in other words, p entails a proposition approximately equivalent to q. The meaning of this inference is made clear in Figure 2. The condition of this entailment relation can be also expressed using Ruspini's implication measure [16] as:

$$p \models^\alpha q \text{ iff } I_\mathcal{S}(q \mid p) = Min_{w \models p} Max_{w' \models q}\mathcal{S}(w, w') \geq \alpha$$

The main properties that the approximate entailment satisfies are the following [5] :

- $\models^1$ coincides with the classical consequence relation ($\models$).

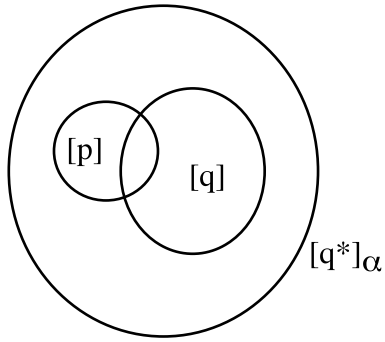

Fig. 2. Approximate entailment $p \models^{\alpha} q$.

- $\otimes$-Transitivity: if $p \models^{\alpha} r$ and $r \models^{\beta} q$ then $p \models^{\alpha \otimes \beta} q$ where $\otimes$ is the t-norm for which the underlying similarity relation is transitive.
- Reflexivity: for any α, $p \models^{\alpha} p$
- Right weakening: if $q \models r$ and $p \models^{\alpha} q$ then $p \models^{\alpha} r$
- Left strengthening: if $p \models r$ and $r \models^{\alpha} q$ then $p \models^{\alpha} q$ (monotonicity)
- Left OR: $p \vee q \models^{\alpha} r$ iff $p \models^{\alpha} r$ and $q \models^{\alpha} r$
- Right OR: If r has a single model, $r \models^{\alpha} p \vee q$ iff $r \models^{\alpha} p$ or $r \models^{\alpha} q$
- Consistency preservation: if $p \neq \bot$ then $p \models^{\alpha} \bot$ only when $\alpha = 0$.

A natural question about the similarity-based entailments is how to deal with some domain knowledge which is available under the form of a set K of formulas—or, equivalently, understood as a subset of worlds $E = [K]$ (called evidential set in [16]). Several extensions of the above approximate entailment can be envisaged (see [5]). A natural option is just to take the set K as a restriction on the set of p-worlds, and thus considering the extension $\models^{\alpha}_{K}$ of the approximate entailment, defined as follows:

$$p \models^{\alpha}_{K} q \;\text{ iff }\; K \wedge p \models^{\alpha} q$$

In other words, we have that $p \models^{\alpha}_{K} q$ iff $I_{\mathcal{S}}(q|K \wedge p) \geq \alpha$. This amounts to expressing that $[q]$ must be stretched to the degree α (at least) in order to encompass the models of K which are models of p, that is $[K] \cap [p] \subseteq [q]_{\alpha}$.

Although the approximate entailment $\models^{\alpha}_{K}$ verifies properties like Reflexivity, Right Weakening or Left Strengthening as $\models^{\alpha}$ does, it does not satisfy the previous transitivity property. Only the following *restricted* form of transivity holds: if $p \models^{\alpha}_{K} r$ and $r \models^{\beta}_{K} q$ then $p \models^{\alpha \otimes \beta}_{K} q$, provided that $r \models K$.

2.2 Proximity entailment

Now we turn our attention to the modeling of the *extrapolative* relationship "*p proximity* entails q" appearing in the inference pattern introduced in section 1.1 and that will play a major role in modeling the relationship between descriptions and solutions of precedent cases in next sections. As already mentioned, the intended meaning in such relations is the following: given a certain context under the form of a proposition K, p implies q in the classical sense, and moreover, if p is close to being true then q is close to being true as well. In other words, the neighborhood of the models of p should lie in the neighborhood of the models of q. Using the notion of approximation introduced in last section, this can be formally expressed by means of the inequality $\mu_{p^*}(w) \leq \mu_{q^*}(w)$, for any model w of K, i.e. $[K] \cap p^* \subseteq q^*$ in the sense of Zadeh's fuzzy set inclusion. We call this consequence relation *proximity entailment* and will be denoted as $p \models_K q$. Obviously, $p \models q$ iff for any $w \in K$, $I_{\mathcal{S}}(p|w) \leq I_{\mathcal{S}}(q|w)$, or equivalently, for any $w \in K$, $Max_{w' \models p} \mathcal{S}(w, w') \leq Max_{w' \models q} \mathcal{S}(w, w')$.

Notice that $\mu_{p^*}(w) \leq \mu_{q^*}(w)$ iff $\mu_{p^*}(w) \otimes\!\!\longrightarrow \mu_{q^*}(w) = 1$, where $\otimes\!\!\longrightarrow$ denotes the residuated implication corresponding to the t-norm $\otimes$, i.e. $x \otimes\!\!\longrightarrow y = sup\{c \in [0,1] \mid x \otimes c \leq y\}$, for all $x, y \in [0,1]$. Thus, using $\otimes\!\!\longrightarrow$, the notion of proximity entailment can be graded as well.

Definition 2 *The α-proximity entailment $p \models_K^\alpha q$ holds iff $[K] \subseteq [p^* \otimes\!\!\longrightarrow q^*]_\alpha$, where $[p^* \otimes\!\!\longrightarrow q^*]_\alpha = \{w \in \Omega \mid \mu_{p^*}(w) \otimes\!\!\longrightarrow \mu_{q^*}(w) \geq \alpha\}$*

The rationale for this definition is to model graded rules stating "the more p, the more q" and they are very close to the so-called *gradual rules* in fuzzy logic [4]. In terms of measures, this entailment relation is related to the conditional implication measure $J_{\mathcal{S},K}(q|p) = Min_{w \models K} I_{\mathcal{S}}(p|w) \otimes\!\!\longrightarrow I_{\mathcal{S}}(q|w)$, introduced in [10], in the sense that:

$$p \models_K^\alpha q \text{ iff } J_{\mathcal{S},K}(q|p) \geq \alpha$$

It is worth noticing that only in the case that $K = True$, the proximity ($\models_K^\alpha$) and the approximate ($\models^\alpha$) collapse (see [6]).

It is also interesting to remark that the entailment relation $\models_K^\alpha$ is $\otimes$-transitive, namely:

$$\text{If } p \models_K^\alpha r \text{ and } r \models_K^\beta q \text{ then } p \models_K^{\alpha \otimes \beta} q$$

Moreover, the set of relations $\{\models_K^\alpha\}_{\alpha \in [0,1]}$ is also nested, verifying $p \models_K^1 q$ iff classically $K \wedge p \models q$ and $\models_K^0$ being the universal relation. Finally, $\models_K^\alpha$ also satisfies the Left-OR property, i.e. $p \vee r \models_K^\alpha q$ iff $p \models_K^\alpha q$ and $r \models_K^\alpha q$.

So far, we have described two graded consequence relations separately. Now we show how both can be combined in order to model inference in CBR systems. As it is shown in [6], the previous definitions of both approximate and proximity entailments allow us to formulate the following sound inference pattern:

$$\frac{p \models^{\beta}_{K_1} q, \quad p' \models^{\alpha}_{K_2} p}{p' \models^{\alpha\otimes\beta}_{K_1 \wedge K_2} q} \tag{1}$$

This provides, in the framework of similarity-based inference, a sound logical account of the following *extrapolative syllogism*:

> "If p *proximity* entails q, given some extrapolative knowledge K_1, and if p' is close to p given a domain knowledge K_2, then p' is also close to q in the presence of both K_1 and K_2".

All these notions easily admit a natural interpretation in a common modal framework. Each similarity relation $\mathcal{S}$ on a set of possible worlds W induces a family of nested accessibility relations $\{R_\alpha\}_{\alpha\in[0,1]}$ such that

$$(w, w') \in R_\alpha \text{ iff } \mathcal{S}(w, w') \geq \alpha.$$

Modal logics are specially tailored to account for relations on the set of interpretations or possible worlds. If we enlarge the language by introducing, for every α, a pair of dual modal operators $(\Diamond_\alpha, \Box_\alpha)$, we can consider a *Similarity Kripke model* as a structure $\mathcal{M} = \langle W, \mathcal{S}, \| \ \| \rangle$ where:

(i) W is a non empty set of possible worlds,
(ii) $\mathcal{S}$ is a $\otimes$-similarity relation on W, for some t-norm $\otimes$,
(iii) $\| \ \|$ is a function that given an atomic formula F return the set $\|F\| \subseteq W$ where F is considered to be true.

Let us define then the satisfaction of modal formulas as follows: let w be a world in a model $\mathcal{M} = \langle W, \mathcal{S}, \| \ \| \rangle$ then:

$$(\mathcal{M}, w) \models \Diamond_\alpha A \text{ iff } I^{\mathcal{M}}_{\mathcal{S}}(A \mid w) \geq \alpha.$$

where $I^{\mathcal{M}}_{\mathcal{S}}(A \mid w) = sup_{(\mathcal{M},w')\models A}\mathcal{S}(w, w')$ is a free adaptation of the implication measure when the second variable is not a proposition but a world. Notice that, if W is finite, then $(\mathcal{M}, w) \models \Diamond_\alpha A$ if there exists a world w' such that $(w, w') \in R_\alpha$ and $(\mathcal{M}, w') \models A$.

Now, it is clear that the *approximate entailment* has a nice interpretation inside this modal framework. Namely, given a $\otimes$-similarity $\mathcal{S}$ on the set of interpretations Ω of the propositional language L, if p and q are non-modal formulas, then we have that

$$p \models_{\mathcal{S}}^{\alpha} q \text{ iff } \mathcal{M}_{\mathcal{L}} \models p \rightarrow \Diamond_{\alpha} q,$$
$$p \models_{\mathcal{S},K}^{\alpha} q \text{ iff } \mathcal{M}_{\mathcal{L}} \models K \rightarrow (p \rightarrow \Diamond_{\alpha} p),$$

where $\mathcal{M}_{\mathcal{L}} = \langle \Omega, \mathcal{S}, \| \ \| \rangle$. Since in $\mathcal{M}_{\mathcal{L}}$, worlds are interpretations we will use henceforth the simpler notation $\mathcal{M}_{\mathcal{L}} = \langle \Omega, \mathcal{S} \rangle$.

Concerning the *proximity entailment*, it is easy to check that $J_{\mathcal{S},K}(q \mid p) \geq \alpha$ holds iff, for any $\beta \in [0,1]$, $I_{\mathcal{S}}(p \mid w) \geq \beta$ implies $I_{\mathcal{S}}(q \mid w) \geq \beta \otimes \alpha$. Hence, if the range $G \subset [0,1]$ of the similarity relation $\mathcal{S}$ is finite, we are able to capture the proximity entailment in our modal framework as well. Indeed, the following relationship holds:

$$p \models\!\!\!\equiv_{\mathcal{S},K}^{\alpha} q \text{ iff } \mathcal{M}_{\mathcal{L}} \models K \rightarrow (\bigwedge_{\beta \in G} \Diamond_{\beta} p \rightarrow \Diamond_{\alpha \otimes \beta} q)$$

Obviously, we get into troubles when G is no longer finite. Then, in order to capture proximity entailments, we are led to introduce in our framework new conditional operators (binary modalities) $\Rightarrow_{\alpha}$ with the following semantics: a formula $A \Rightarrow_{\alpha} B$ is true in a world w of a similarity Kripke model $\mathcal{M} = \langle W, \mathcal{S}, \| \ \| \rangle$ if the conditional implication of B given A, taking w as extrapolative knowledge, is at least α. Formally:

$$(\mathcal{M}, w) \models A \Rightarrow_{\alpha} B \text{ iff } I_{\mathcal{S}}^{\mathcal{M}}(A \mid w) \otimes\!\!\rightarrow I_{\mathcal{S}}^{\mathcal{M}}(B \mid w) \geq \alpha$$

In this way, we capture the proximity entailment in the sense that the following equivalence holds:

$$p \models\!\!\!\equiv_{\mathcal{S},K}^{\alpha} q \text{ iff } \models_{\mathcal{M}_{\mathcal{L}}} K \rightarrow (p \Rightarrow_{\alpha} q).$$

Using this modal formalism, the extrapolative principle mentioned at the end of last subsection can be formalized as the following sound rule in $\mathcal{M}_{\mathcal{L}}$:

$$\frac{K_1 \rightarrow (A \Rightarrow_{\alpha} B), \quad K_2 \rightarrow (A' \rightarrow \Diamond_{\beta} A)}{(K_1 \wedge K_2) \rightarrow (A' \rightarrow \Diamond_{\alpha \otimes \beta} B).} \text{Analogical Extrapolation}$$

3 Models of Case-Based Reasoning

We will see in this section how the approximate and the proximity entailments can characterize precedent-based inference used in CBR systems. In order to do so we begin by establishing the language used in the $\mathcal{PPR}$ model to represent precedent cases.

Since we will understand cases just as a collection of ground instances of some predicates, possibly negated, we don't need a full language of predicate calculus, actually our language will be very simple. We start out from:

(i) a set of *sorts* of variables; a *type* is a tuple of sorts
(ii) a set of object *constants*, each one having its sort,
(iii) a finite set of predicate symbols $\Sigma = \{A_1, \ldots, A_n\}$, each one having a type, and
(iv) two connectives: $\neg$ (negation) and $\wedge$ (conjunction).

Predicate symbols stand for the different attributes used to describe cases in a certain domain. We assume that each predicate symbol A_i has associated a natural number $n_i \geq 1$ denoting its arity. Further, predicate symbols are classified in two disjoint subsets $\mathcal{O}$ and $\mathcal{I}$. The set $\mathcal{O}$ are called *outcomes* or *solutions*, while the set $\mathcal{I}$ are called *inputs* or *data*.

The *atomic formulas* will be of the form $A_i(c_1, ..., c_{n_i})$, where c_j are object constants, and as usual, *literals* will be either atomic formulas or their negation. Then *formulas* are built in the usual way from the atomic formulas and the connectives $\neg$ and $\wedge$. We will refer to this restricted language as $\mathcal{L}$. Actually, $\mathcal{L}$ is a simple propositional language since we do not deal with object variables. This simplifies the semantical notion of interpretation. Namely, the set Ω of interpretations of $\mathcal{L}$ is just the set of evaluations of the predicate instances taken as propositional variables.

Definition 3 *A* precedent P *is a pair* (C, S) *where* C *is a conjunction of literals of predicates belonging to* $\mathcal{I}$ *and* S *is a literal of a predicate belonging to* $\mathcal{O}$

In the framework of case-based reasoning, each precedent $\{P_i = (C_i, S_i)\}$ can be conceived as an *extrapolative* conditional relation $C_i \Rightarrow S_i$, i.e. given a situation which can be described, at some extent, by C_i we may infer that S_i is a plausible solution at some extent as well. Moreover, the closer is the situation to C_i, the more plausible is that S_i is a solution for that situation. When we have a new situation described by some conjunction of literals C, as is common in CBR literature, the task is to infer some plausible solution S taking into account the restrictions given by the precedents.

In practice, it is only available the similarity relation $\mathcal{S}_I$ on the *input space*, i.e. in the set of interpretations Ω_I of the sublanguage $\mathcal{L}_\mathcal{I}$ generated from predicates of $\mathcal{I}$, while the similarity relation $\mathcal{S}_O$ on the output space, i.e. in the set of interpretations Ω_O of the sublanguage $\mathcal{L}_\mathcal{O}$ generated from predicates of $\mathcal{O}$, is unknown. Therefore, given the similarity $\mathcal{S}_I$, we will consider similarity

Kripke structures $\mathcal{M}_{\mathcal{L}} = \langle \Omega, \mathcal{S} \rangle$ where Ω is the cartesian product $\Omega_I \times \Omega_O$ and the similarity $\mathcal{S}$ on Ω is the so-called *product similarity* $\mathcal{S} = \mathcal{S}_I \times \mathcal{S}_O$, for some similarity $\mathcal{S}_O$, and it is defined as:

$$\mathcal{S}((w_I^1, w_O^1), (w_I^2, w_O^2)) = min(\mathcal{S}_I((w_I^1, w_I^2), \mathcal{S}_O((w_O^1, w_O^2))$$

Then it is easy to show the following inequality:

$$I_{\mathcal{S}_I}(p \mid q) = I_{\mathcal{S}}(p \mid q)$$

for any propositions $p, q \in \mathcal{L}_{\mathcal{I}}$.

Our next task is to characterize inference performed a CBR system according to the $\mathcal{PPR}$ model.

Definition 4 *A $\mathcal{PPR}$ system is a structure $\langle B, \mathcal{S}_I, K \rangle$, where*

(i) $B = \{P_i = (C_i, S_i)\}_{i=1,\ldots,n}$ is a case base of precedents,
(ii) $\mathcal{S}_I$ is a similarity on the set of interpretations Ω_I of our sublanguage $\mathcal{L}_{\mathcal{I}}$ and
(iii) K is a set of $\mathcal{L}_{\mathcal{I}}$–formulas standing for the general knowledge the $\mathcal{PPR}$ system has about the application domain.

To characterize the inference in a $\mathcal{PPR}$ system $\langle B, \mathcal{S}, K \rangle$, using the modal approach introduced in section 2.3, we need first of all to extend our language $\mathcal{L}$ with modalities $\Diamond_\alpha$ and $\Rightarrow_\alpha$ as we did in section 2.3. Given a $\mathcal{PPR}$ system $\mathcal{P} = \langle B, \mathcal{S}_I, K \rangle$, let us denote by $\mathcal{C}_{\mathcal{P}}$ the class of Kripke structures $\mathcal{M} = \langle \Omega, \mathcal{S} \rangle$ such that $\mathcal{S}$ is a product similarity whose first component is $\mathcal{S}_I$.

Then given current description problem C, a $\mathcal{PPR}$ system $\langle B, \mathcal{S}_I, K \rangle$ provides us with two basic information components:

(i) for each precedent $P_i = (C_i, S_i) \in B$ we can compute, by means of the similarity $\mathcal{S}_I$, the implication degree $\alpha_i = I_{\mathcal{S}_I,K}(C_i \mid C)$, which measure how close is the current problem description C to any precedent description C_i, given the general knowledge K. Taking into account the above inequality, $I_{\mathcal{S},K}(C_i \mid C) \geq \alpha_i$ and thus we can assure that the formulas

$$K \rightarrow (C \rightarrow \Diamond_{\alpha_i} C_i)$$

are valid in the class $\mathcal{C}_{\mathcal{P}}$.

(ii) the case base of precedents B is given an *extrapolative interpretation*, i.e. B gives raise to a base of extrapolative rules $B^* = \{C_i \Rightarrow_{\beta_i} S_i \mid (C_i, S_i) \in$

$B\}$, for some suitable degrees β_i [4]. Hence, the formulas

$$B^* \to (C_i \Rightarrow_{\beta_i} S_i)$$

are also valid in the class $\mathcal{C}_\mathcal{P}$.

Therefore, applying the *analogical extrapolation* inference rule (see end of last section) we can derive the formulas

$$(K \wedge B^*) \to (C \to \Diamond_{\alpha_i \otimes \beta_i} S_i)$$

which will automatically be sound in the class $\mathcal{C}_\mathcal{P}$, for each $i = 1 \ldots n$.

We can summarize the whole inference process by the expression

$$\{K \to (C \to \Diamond_{\alpha_i} C_i)\}_{i=1\ldots n} \cup \{K, B^*, C\} \vdash \Diamond_{\alpha_i \otimes \beta_i} S_i \quad (2)$$

where $\vdash$ means deduction using only analogical extrapolation (taking $K_1 = B^*$ and $K_2 = K$) and modus ponens as inference rules. This provides the logical grounds for the next definition of plausible inference in the $\mathcal{PPR}$ model.

Definition 5 *Given a $\mathcal{PPR}$ system $\mathcal{P} = \langle B, \mathcal{S}_I, K\rangle$ and a current problem C, let $\alpha_i = I_{\mathcal{S}_I}(C_i \mid K \wedge C)$, $i = 1 \ldots n$, and let $B^* = \{C_i \Rightarrow_{\beta_i} S_i \mid (C_i, S_i) \in B\}$ the* extrapolative *interpretation of the case base of precedents B. Then we say that a solution S_i is inferred to be plausible at least at degree $\delta_i = \alpha_i \otimes \beta_i$.*

This notion of plausibility of a solution for a current case can be used to rank the solutions according to their maximum degree of plausibility. Notice that some precedents can share the same solution.

Definition 6 *Given a $\mathcal{PPR}$ system $\mathcal{P} = \langle B, \mathcal{S}_I, K\rangle$, together with some extrapolative interpretation of the precedents B^* as above, let us define, for any solution S in B the plausibility index as*

$$pl(S) = max\{\delta \mid \{K \to (C \to \Diamond_{\alpha_i} C_i)\}_{i=1\ldots n} \cup \{K, B^*, C\} \vdash \Diamond_\delta S\}.$$

Then we define the preorder $\leq_C$ induced by a current problem C on the set $Sol(B)$ of solutions of the case base of precedents B as follows: $S_i \leq_C S_j$ iff $pl(S_i) \leq pl(S_j)$

Notice that our representation of precedent cases $P_k = (C_k, S_k)$ in terms of extrapolative rules does not assume a priori that the relation becomes a crisp implication when the current case matches perfectly with C_k —i. e. that the set of precedents in the case base B is a set of problems with well known

[4] Note that in most CBR systems it is usual to model a case as a proximity entailment with degree 1.

solutions. Regarding the formal model of §2, granting this assumption would mean that the proximity entailment $\models^{\alpha}_{B*}$ has always $(\alpha = 1)$. This assumption is very usual because almost all CBR systems assume the precedents involve no imprecision. The ARC system [15] is an exception, since it worked with cases the solution of which was a set of clinical diagnosis with a plausibility degree . In this sense it is interesting to remark that the $\mathcal{PPR}$ model is a general one and allows to work either with precise or imprecise precedents, i.e. it is possible to incorporate a degree in the precedents without changes in the theory supporting the $\mathcal{PPR}$ model.

3.1 Instance-based learning

There is a family of techniques implemented in CBR systems that have only *casuistic* knowledge—i. e. they do not possess general knowledge about the domain they deal with, only a case base of precedents. These systems perform inferences warranted only by the case base and the similarity relation used. The k-nearest neighbor algorithms and the instance-based learning algorithm [12] pertain to this class of CBR systems since all the knowledge they use is casuistic—it is the knowledge embodied by the similarity used and set of cases considered. We will call these approaches *purely casuistic* $\mathcal{PPR}$ systems, and we will denote them by $\mathcal{PPR}^{\mathcal{I}} = \langle B, S, \emptyset \rangle$.

The lack of domain knowledge in $\mathcal{PPR}^{\mathcal{I}}$ amounts to say that K does not exists and thus the expression 2 of the inference process is now given by:

$$\{(C \rightarrow \Diamond_{\alpha_i^*} C_i)\}_{i=1 \ldots n} \cup \{B^*, C\} \vdash \Diamond_{\alpha_i^* \otimes \beta_i} S_i$$

where $\alpha_i^* = I_{\mathcal{S}_I}(C_i \mid C)$.

Definition 7 *For a $\mathcal{PPR}^{\mathcal{I}}$ system $\langle B, S_I, \emptyset \rangle$ plus some extrapolative interpretation of the precedents B^*, the plausibility index for any solutions S in C is*

$$pl(S) = max\{\delta \mid \{(C \rightarrow \Diamond_{\alpha_i} C_i)\}_{i=1 \ldots n} \cup \{B^*, C\} \vdash \Diamond_{\delta} S\}$$

and the preorder $\leq_C$ induced by a current problem C on the set of solutions of the case base of precedents B is as before: $S_i \leq_C S_j$ iff $pl(S_i) \leq pl(S_j)$

That is to say, the $\mathcal{PPR}$ model conceives BCR inference in two steps. First, the similarity between the current problem C and a precedent C_k is interpreted as $C \models^{\alpha} C_k$ iff $[C] \subset [C_k^*]_{\alpha}$, which means that when the current problem C is true then the current precedent C_k is close to being true. A visual representation is given in Figure 2 where the sphere of current precedent $[C_k]$ is enlarged to an α-degree $[C_k^*]_{\alpha}$ such that it encompasses the $[C]$ sphere.

Second, the similarity over cases in B is carried over to the space of solutions by means of the the extrapolative interpretation B^*, and finally yielding a partial order of plausibility over the solutions in $\mathcal{O}$.

3.2 CBR Inference with General Domain Knowledge

Purely casuistic CBR systems assume that the only represented knowledge is a specific collection of cases (in some language) with their solutions—plus a similarity relation. However, this assumption is quite strict and most CBR systems have also general knowledge about the domain of application [1]. The general knowledge K in CBR systems can take many forms and be represented in a number of languages. However, for our purposes we can characterize K as the ability of the CBR system to infer new propositions about a current problem C given the initial true propositions about C. That is to say, a case C (whether a precedent already solved or a problem to be solved) can be characterized as $C = C^O \cup C^D$ where C^O is a set of initially given propositions about C, and C^D is a set of propositions such that $K \wedge C^O \models C^D$—i. e, a set derived from K and C^O. In other words, cases are such that given some initially give propositions C^O then the rest of the propositions about C cannot have just any truth value—they are restricted to those established by K.

Regarding the $\mathcal{PPR}$ model, this amounts to say that K in a knowledge-intensive CBR system imposes some restrictions on the possible worlds we have to consider—namely, those that obey the restrictions imposed by K.

The main difference of general $\mathcal{PPR}$ systems with respect to $\mathcal{PPR}^\mathcal{I}$ systems is that now K plays a role in the $\mathcal{PPR}$ model. Namely, that a knowledge-intensive CBR system assumes the restrictions imposed by K on the possible worlds to be considered.

3.3 Overall Inference in CBR Systems

So far, we have dealt with the basic form of inference in a $\mathcal{PPR}$ system. Since a CBR system has a number n of precedents in the case base B, some overall mechanism has to be provided regarding which elementary inferences are actually performed and how they are aggregated into a global inference. A first overall mechanism is simply performing all possible elementary inferences and take the maximum α for each S_j. This exhaustive strategy is precisely the one portrayed in Definition 6 where the preorder over the solutions $S \in Sol(B)$ is induced by the plausibility index $pl(S)$. Other overall inference strategies may use t-conorms other than maximum. Since max is the smaller t-conorm, these overall strategies are aggregating the evidence from different precedents into the overall plausibility degree.

On the other hand, the most usual overall strategy taken by CBR systems is to take only one possible outcome by selecting the result S_j with a maximum α (a "minimum distance" criterion as is usually called). In fact, since there can be more than one solution with the same plausibility degree, this strategy yields a set of solutions $\{S_i\}$, where S_i is a maximal element with respect to the preorder $\langle Sol(B), \leq_C \rangle$.

Another common strategy is that of **k**-nearest neighbour methods where the case base is restricted to the **k** precedents closer to the current problem. That is to say, the CBR system selects the **k** precedents with higher α (according to $\leq_C$), takes the set $Sol_k(B) \subset Sol(B)$ of solutions associated with them, and builds a preorder $\langle Sol_k(B), \leq_C \rangle$.

There are a handful of other options for control regimes for inference in CBR systems but they can be seen as different restrictions upon the basic inference model of precedent-based plausible inference.

4 Discussion and Future Work

In this paper we have introduced $\mathcal{PPR}$ (Precedent-based Plausible Reasoning), a formal model for the similarity-based inference process that characterizes CBR systems. The $\mathcal{PPR}$ model is based on the notions of approximate entailment and proximity entailment. These notions are used, respectively, to model the similarity between a current problem and a precedent case and to model the implication relation between a precedent and its associate solution. Furthermore, a modal interpretation of CBR inference in terms of *α-plausible possibility* has been presented.

We have also shown that for the particular case of instance-based learning (i. e. for inference based only on precedents and a similarity relation but without general domain knowledge), what we call purely casuistic systems, modeled by $\mathcal{PPR}^{\mathcal{I}}$, are just a special case of $\mathcal{PPR}$ systems: they all share the same underlying inference principle, and the difference is whether or not that inference is improved by having general domain knowledge K. As a corollary, a clarification of the relationship of nearest neighbor and instance-based algorithms with CBR systems is possible in the $\mathcal{PPR}$ model. It is also interesting to notice that our model is general enough to encompass both crisp precedents and uncertain precedents

We have seen that CBR inference is based on a problem-centered preference ordering induced upon a set of cases. A similarity relation is the most common way to induce such a preference ordering, but it is not essential. Since a similarity relation can induce a preference ordering (and the converse is not true), and since the graded entailment relations are based on the preference ordering, it seems interesting to work directly in a formal model based on preference orderings. In fact, there are CBR systems that work directly with preferences and do not need explicit similarities. Examples of such systems are some CBR languages and application we have developed ([2,3]) that work with *partial* orders of preferences. In this approach, since the order is not total, instead of a numeric plausibility degree we obtain an ordinal plausibility degree in inference. Although the formal model for CBR with partial preferential orders is yet to be developed we think it will be based in the same formal apparatus used here.

The representation of precedent cases as extrapolative conditionals from case description propositions to the solution proposition is also interesting. This means we are considering precedent cases as *instantiated rules* and the similarity provides a mechanism to perform a "partial matching" of a current case against that rule. This view clarifies the role of *lazy learning* in CBR systems and the nature of CBR systems as Machine Learning systems. In *eager* learning methods like induction, a set of general rules are generated from a set of cases—independently of the future problems to solve. Then new problems are solved by *exact* matching against those induced rules. Lazy learning methods delay generalization until meeting with a new problem. In CBR systems a new kind of generalization is performed: from a precedent case $P_i = (C_i, S_i)$, the set of possible worlds $[C_i]$ is enlarged by means of a similarity until it encompasses the set of possible worlds $[C]$ corresponding to the current problem. The smaller the sphere, the closer the current case is to that precedent, and the more plausible is that S_i is a solution for C. In summary, since the generalization (that corresponds to the sphere) is delayed until the current problem case is encountered, CBR systems perform a form of lazy learning that does not need to construct a representation of the generalization. It is also worth to note that up to now lazy learning has been considered a feature of $\mathcal{PPR}^{\mathcal{I}}$ systems, while our model shows that any CBR system, even with complex inference patterns derived from domain knowledge, is indeed a lazy learner.

References

[1] Agnar Aamodt and Enric Plaza. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *Artificial Intelligence Communications*, 7(1):39–59, 1994. <url:http://www.iiia.csic.es/People/enric/AICom_ToC.html>.

[2] Josep Lluís Arcos and Enric Plaza. Integration of learning into a knowledge modelling framework. In Luc Steels, Guss Schreiber, and Walter Van de Velde, editors, *A Future for Knowledge Acquisition*, number 867 in Lecture Notes in Artificial Intelligence, pages 355–373. Springer-Verlag, 1994.

[3] Eva Armengol and Enric Plaza. Integrating induction in a case-based reasoner. In J. P. Haton, M. Keane, and M. Manago, editors, *Advances in Case-Based Reasoning*, number 984 in Lecture Notes in Artificial Intelligence, pages 3–17. Springer-Verlag, 1994.

[4] D. Dubois and H. Prade. Gradual inference rules in approximate reasoning. *Information Sciences*, 61:103–122, 1992.

[5] Didier Dubois, Francesc Esteva, Pere Garcia, Lluis Godo, and Henri Prade. Similarity-based consequente relations. In Christine Froidevaux and Jürg Kohlas, editors, *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, volume 946 of *Lecture Notes in Artificial Intelligence*, pages 168–174, Fribourg, Switzerland, July 1995. Springer.

[6] Didier Dubois, Pere Garcia, Francesc Esteva, Lluis Godo, and Henri Prade. A logical approach to interpolation based on similarity relations. *Int. Journal of Approximate Reasoning*, 1997. To appear, also available as IIIA Research Report 96-07.

[7] Didier Dubois and Henri Prade. Similarity-based approximate reasoning. In Jacek M. Zurada, Robert J. Marks II, and Charles J. Robinson, editors, *Computational Intelligence: Imitating Life*, pages 69–80. IEEE Press, 1994.

[8] Didier Dubois and Henri Prade. Comparison of two fuzzy set-based logics: similarity logic and possibilistic logic. In *Proc. FUZZIEEE'95*, pages 1219–1226, Yokohama, Japan, 1995. IEEE Press.

[9] Francesc Esteva, Pere Garcia, and Lluis Godo. On conditional in similarity logic. In B. Bouchon-Meunier, R.R. Yager, and L. A. Zadeh, editors, *Fuzzy Logic and Soft Computing*, pages 301–309. World Scientific, 1995.

[10] Francesc Esteva, Pere Garcia, and Lluis Godo. Relating and extending semantical approaches to possibilistic reasoning. *Int. J. of Approximate Reasoning*, 10:311–344, 1995.

[11] Francesc Esteva, Pere Garcia, Lluis Godo, Enrique Ruspini, and Llorenç Valverde. On similarity logic and the generalized modus ponen. In *IEEE International Conference on Fuzzy Systems*, volume 2, pages 1423–1427, Orlando USA, 1994. IEEE Press.

[12] D. Kibler and D. Aha. Learning representative exemplar concepts: An initial case study. In *Proc. Fourth International workshop on Machine Learning*, pages 24–30. Morgan Kaufmann, 1987.

[13] Frank Klawonn and Rudolf Kruse. Equality relations as a basis for fuzzy control. *Fuzzy Set and Systems*, 54:54–147, 1994.

[14] Enric Plaza. Bias in case-based reasoning systems. Technical report, IIIA, 1996.

[15] Enric Plaza and Ramon López de Mántaras. Learning typicality from fuzzy examples. In Z. Ras, M. Zemankova, and M. L. Emrich, editors, *Methodologies for Intelligent Systems 5*, pages 420–427. North Holland, 1990.

[16] Enrique Ruspini. On the semantics of fuzzy logic. *International Journal Of Approximate Reasoning*, (5):45–88, 1991.

[17] Enric Trillas and Llorenç Valverde. An inquiry on t-indistinguishability operators. In H. Skala et al, editor, *Aspects of Vagueness*, pages 231–256. Reidel, Dordrecht, 1984.

[18] Lofti A. Zadeh. A theory of approximate reasoning. In J. E. Hayes, D. Michie, and L.I. Mikulich, editors, *Machine Intelligence*, volume 9, pages 149–194. Wiley, 1979.