



**HAL**  
open science

## Case-Based Prediction - A Survey

Fadi Badra, Marie-Jeanne Lesot

► **To cite this version:**

| Fadi Badra, Marie-Jeanne Lesot. Case-Based Prediction - A Survey. 2022. hal-03859329

**HAL Id: hal-03859329**

**<https://hal.science/hal-03859329>**

Preprint submitted on 18 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Case-Based Prediction – A Survey

Fadi Badra<sup>a</sup>, Marie-Jeanne Lesot<sup>b</sup>

<sup>a</sup> *Université Sorbonne Paris Nord, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé, LIMICS, Sorbonne Université, INSERM, F-93000, Bobigny, France*

<sup>b</sup> *Sorbonne Université, CNRS, LIP6, F-75005 Paris, France*

---

## Abstract

This paper clarifies the relation between case-based prediction and analogical transfer. Case-based prediction consists in predicting the outcome associated with a new case directly from its comparison with a set of cases retrieved from a case base, by relying solely on a structured memory and some similarity measures. Analogical transfer is a cognitive process that allows to derive some new information about a target situation by applying a plausible inference principle, according to which if two situations are similar with respect to some criteria, then it is plausible that they are also similar with respect to other criteria. Case-based prediction algorithms are known to apply analogical transfer to make predictions, but the existing approaches are diverse, and developing a unified theory of case-based prediction remains a challenge. In this paper, we show that a common principle underlying case-based prediction methods is that they interpret the plausible inference as a transfer of similarity knowledge from a situation space to an outcome space. Among all potential outcomes, the predicted outcome is the one that optimizes this transfer, *i.e.*, that makes the similarities in the outcome space most compatible with the observed similarities in the situation space. Based on this observation, a systematic analysis of the different theories of case-based prediction is presented, where the approaches are distinguished according to the type of knowledge used to measure the compatibility

---

\*Corresponding author

*Email address:* badra@sorbonne-paris-nord.fr (Fadi Badra)

*URL:* limics.fr (Fadi Badra)

*Preprint submitted to International Journal of Approximate Reasoning*      *October 27, 2022*

between the two sets of similarity relations.

*Keywords:* case-based prediction, analogical transfer, similarity

---

## Introduction

Computational analogy is a subfield of computer science that aims at designing computational models of psychological and cognitive processes of analogical thinking [4, 45, 95]. Analogical reasoning is recognized to be at the core of human thought [46, 56, 57]. For instance, in medicine, analogies are commonly used by medical experts for their role in “explaining, naming and mediating knowledge” [92]. Among the different tasks that computational analogy systems implement, the transfer task implements a special type of plausible inference principle, according to which if two situations are similar according to some criteria, then it is plausible that they are similar according to some other criteria.

A review of the recent literature reveals a regain of interest in using analogical transfer e.g., for decision-making [15], preference learning [39], explainable AI [68], categorization [20], as well as to cope with insufficient training data [93], to guide language generation [80], to foster creativity [51], or to accelerate innovation [59]. The strengths of computational analogy methods include an ability to work with a small number of instances, to handle context, to produce explainable results, to leverage a structured memory, or to allow for creativity. Contrary to most machine learning approaches, no pre-trained model of the task at hand is required. Instead, the system solely relies on a structured memory and some similarity measures. Despite these obvious strengths, the theoretical study of analogical transfer in computer science has been mostly overlooked in the past years, and developing a unified theory remains a challenge. Setting a common ground for analogical transfer theories would require a better understanding of the common principles underlying the existing approaches.

Designing computer systems that implement a form of analogical reasoning has been studied since the 1980’s in the field of case-based reasoning (CBR) [96].

As a research field, CBR is mostly concerned with the knowledge-engineering aspects that need to be addressed when designing computational analogy systems. Case-based prediction (also called case-based *inference* [65]) algorithms are algorithms used in CBR to address supervised learning tasks such as classification or regression. The most popular case-based prediction algorithm is (by far) the  $k$ -Nearest Neighbor algorithm, but as we will see, other algorithms have been proposed, such as PossIBL [64], CCBI [66], WAPC [12] or, more recently, CoAT [6]. Case-based prediction algorithms are known to apply the plausible inference principle of analogical transfer, according to which similar situations have similar outcomes. However, there is no unified theory of case-based prediction, and the relation between these types of algorithms and the principles of computational analogy often remains unclear.

In this paper, we show that a common principle underlying case-based prediction methods is that they interpret the plausible inference principle of analogical transfer as a transfer of similarity knowledge from a situation space to an outcome space that the predicted outcome should optimize. Among all potential outcomes, the predicted outcome is the one that makes the similarities in the outcome space most compatible with the observed similarities in the situation space according to some compatibility measure. Based on this observation, a systematic analysis of the different theories of case-based prediction is presented, where the approaches are distinguished according to the type of knowledge used to measure the compatibility between the two sets of similarity relations.

The paper is organized as follows. The next section presents the transfer task of computational analogy systems. Sec. 2 shows that case-based prediction methods, which implement the transfer task for prediction purposes, interpret the plausible inference principle of analogical transfer as a transfer of similarity knowledge from a situation space to an outcome space. Sec. 3 presents a systematic analysis of existing case-based prediction theories, proposing a typology distinguishing between four main families of approaches. Sec. 4 to 7 describe in turn the most representative approaches for each of the four families. Sec. 8 concludes the paper and gives directions for future work.

## 1. The Transfer Task of Computational Analogy Systems

60 Computational analogy systems are often decomposed into different tasks in the literature [1, 46, 49, 75, 76]. Although decomposing the analogical inference in a succession of tasks is highly questionable from a cognitive science point of view, because analogical reasoning is rooted in perception, and involves complex cognitive processes that are often intertwined [19], making such a de-  
65 composition greatly simplifies the conception of computer systems, and allows for comparison.

The transfer task is the component of computational analogy systems that allows to make predictions. It implements analogical transfer, which is the part of the analogical reasoning process that allows to leverage a mapping with an  
70 analog retrieved from memory in order to derive some new information about the current situation [44, 47, 88]. It does so by applying a special type of plausible inference principle according to which if two situations are similar according to some criteria, then it is plausible that they are similar according to some other criteria. Such an inference process may serve different purposes [77]:

- 75 • It may be used for *prediction*, in order to complete the description of the new case. Examples include estimating a quantity [91], a preference [39], a ratio [60], a semantic relation [99], recommending a decision [33] or predicting the effect of a decision or a plan (for example to find a response plan for natural disasters [54] or to help an athlete choose a pace when  
80 running a marathon [85]).
- It may be used for *interpretation*, in order to borrow from an analog an explanation, or a justification, of the new situation. Examples include deriving explanations [68, 97], causal attributions [58], establishing a legal assessment [30], putting forth persuasive arguments in the context of ad-  
85 versarial reasoning [55, 74], or even tapping into the emotions that people gained from their own experience in order to sway a decision [48].
- It may also be used for *creativity*, in order to propose a novel solution for

the new situation, by adapting and combining past solutions. Examples include adapting cooking recipes in order to match some constraints [22],  
90 solving mathematical problems by adapting old solution procedures [89], proposing innovative ideas for designers [52], making content suggestions for authors of product reviews [18] or inventing new concepts by conceptual blending [38].

Apart from the transfer task, many other tasks are cited in the literature:  
95 given a new situation, the goal of the *retrieval* task is to draw from memory one (or many) situation(s) to compare it to. In the *mapping* task, the new situation is compared to the retrieved situation(s). This interpretation process involves a structural alignment between the new situation and the retrieved one(s) [43]. The *representation* task [75] produces an initial representation for the inputs.  
100 The *elaboration* task [41] enriches such representation with domain knowledge. The *re-representation* task [84] restructures the input representation to facilitate comparison. The *abstraction* task [50] produces a common abstraction of two inputs. The *validation* task [40] evaluates the quality and consistency of the results. The *memorization* task [57] updates the memory for later use.

105 In the next section, we focus on case-based prediction methods, which implement the transfer task for prediction purposes.

## 2. Case-Based Prediction

Developed in the domain of case-based reasoning, case-based prediction methods aim at predicting the outcome of a new case directly from its  
110 comparison with a set of cases retrieved from a case base, by relying solely on a structured memory and some similarity measures. In this section, we show that case-based prediction algorithms interpret analogical transfer as a transfer of similarity knowledge from the space of situations to the space of outcomes. They predict the outcome that makes the similarities on outcomes most  
115 compatible with the observed similarities on situations, and differ mainly by the

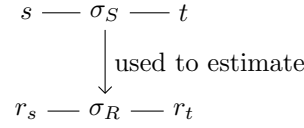


Figure 1: In the transfer task, a similarity relation  $\sigma_S$  is used to estimate another one,  $\sigma_R$ .

type of knowledge used to measure the compatibility between the two sets of similarity relations.

### 2.1. Definitions and Notations

A typical setting for case-based prediction is as follows. Let  $\mathcal{S}$  be an input  
120 space and  $\mathcal{R}$  an output space. An element of  $\mathcal{S}$  is called a *situation*, and an  
element of  $\mathcal{R}$  is called an *outcome*, or a result. A *case base* is a finite set  
 $CB = \{(s_1, r_1), \dots, (s_n, r_n)\}$  of elements in  $\mathcal{S} \times \mathcal{R}$ . For legibility, and abusing the  
notation, cases and outcomes are sometimes denoted with their corresponding  
situation as subscript: an element  $c_s = (s, r_s) \in CB$  is called a *source case*. In  
125 addition,  $\sigma_S$  and  $\sigma_R$  respectively denote similarity measures on situations and  
on outcomes. For a new case  $c_t = (t, r_t)$  whose outcome  $r_t$  is to be predicted,  
the case-based inference can be decomposed into three main tasks [49]:

- *Retrieval*: retrieve from  $CB$  a set of source cases  $\{c_s = (s, r_s)\}$ ;
- *Mapping*: for each retrieved situation  $s$ , estimate the similarity  $\sigma_S(s, t)$   
130 between  $s$  and the target situation  $t$ ;
- *Transfer*: estimate the similarities  $\sigma_R(r_s, r_t)$  on outcomes from the simi-  
larities  $\sigma_S(s, t)$  on situations.

In the transfer task, the similarities  $\sigma_R(r_s, r_t)$  on outcomes are estimated from  
the similarities  $\sigma_S(s, t)$  on situations, and the outcome  $r_t$  is (indirectly) deter-  
135 mined from these estimations (Fig. 1).

The goal of case-based prediction methods is to predict the outcome that  
makes analogical transfer most likely to succeed when the new case is compared  
to the retrieved cases. Intuitively, this means that the outcome  $r_t$  should be

chosen so that the outcome similarities reflect what the situation similarities are,  
 140 *i.e.*, one can rely on a similarity estimation between two situations to estimate  
 the similarity between their associated outcomes.

## 2.2. A Transfer of Similarity Knowledge

To find the most plausible outcome, case-based prediction methods search  
 for the outcome that makes the added similarities on outcomes most compatible  
 145 with the observed similarities on situations. More precisely, let  $\hat{c}_t = (t, r)$  denote  
 a potential new case formed by choosing the outcome  $r \in \mathcal{R}$  for the new case.  
 Adding the new case  $\hat{c}_t$  to the case base  $CB$  results in building two sets of  
 similarity relations:

- the situation similarities  $\Sigma_{\mathcal{S}}(t) = \{\sigma_{\mathcal{S}}(s, t) \mid c_s = (s, r_s) \in CB\}$
- 150 • the outcome similarities  $\Sigma_{\mathcal{R}}(r) = \{\sigma_{\mathcal{R}}(r_s, r) \mid c_s = (s, r_s) \in CB\}$

Among these two sets of similarity relations, only  $\Sigma_{\mathcal{R}}(r)$  depends on the choice  
 of  $r$ : changing the outcome  $r$  leads to new (possibly opposite) similarity rela-  
 tions. Case-based prediction methods search for the new case  $c_t = (t, r_t)$  for  
 which the similarity relations  $\Sigma_{\mathcal{R}}(r_t)$  in the outcome space is the most compat-  
 155 ible with the similarity relations  $\Sigma_{\mathcal{S}}(t)$  in the situation space.

As an example, consider the classification setting graphically represented on  
 Fig. 2. Situations are points of a 2D space, and the similarity  $\sigma_{\mathcal{S}}(s, t)$  between  
 two situations  $s$  and  $t$  is estimated from the Euclidean distance, e.g., by setting  
 $\sigma_{\mathcal{S}} = e^{-\|\cdot\|_2}$ , where  $\|\cdot\|_2$  is the Euclidean distance. Outcomes are classes (**blue**  
 160 or **red**), and the similarity  $\sigma_{\mathcal{R}}(r_s, r_t)$  between two outcomes  $r_s$  and  $r_t$  is given  
 by the discrete metric, which returns 1 if the two classes are the same, and 0  
 otherwise. Suppose that the case base  $CB$  contains a set of situations  $s$ , whose  
 outcome  $r_s$  is known, and we want to predict the outcome  $r_t$  of a new situation  $t$   
 (represented as  $\star$  on Fig. 2). When a potential new case  $\hat{c}_t = (t, r)$  is added  
 165 to the case base, the situation similarities  $\Sigma_{\mathcal{S}}(t)$  between the new situation  $t$   
 and each past situation  $s$  can be computed using  $\sigma_{\mathcal{S}}$ . However the outcome  
 similarities  $\Sigma_{\mathcal{R}}(r)$  depend on the outcome  $r$  chosen for  $t$ : if  $r = \mathbf{red}$ , then



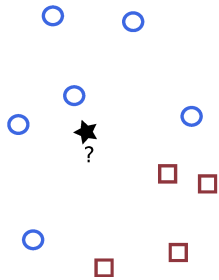


Figure 2: A simple classification setting. The goal is to predict the class (blue circle or red square) for a new situation represented by  $\star$  on the figure, by finding the class for which the added outcome similarities would be most compatible with the observed situation similarities.

the new case is similar to all cases whose outcome is also **red**, and dissimilar to the other cases. If  $r = \mathbf{blue}$ ,  $\Sigma_{\mathcal{R}}(r)$  contains opposite similarity relations.

170 Case-based prediction consists in finding the outcome  $r_t$  for which the added similarity relations  $\Sigma_{\mathcal{R}}(r_t)$  on classes are the most compatible with the observed similarity relations  $\Sigma_{\mathcal{S}}(t)$  between points, measured by the Euclidean distance. The idea is that if the similarities on outcomes are compatible with similarities on situations, then one can rely on similarities on situations (Euclidean distance)

175 to estimate similarities on outcomes (class membership), and analogical transfer succeeds.

This search for compatibility between the two sets of similarity relations can be seen as a transfer of similarity knowledge from the situation space to the outcome space. Indeed, each of the two similarity measures  $\sigma_{\mathcal{S}}$  and  $\sigma_{\mathcal{R}}$  can be

180 seen as a transformation of the underlying space that "groups together" similar points in the transformed space. In this view, case-based prediction is a search for an outcome  $r_t$  for which the situations transformed by  $\sigma_{\mathcal{S}}$  into similar points are also transformed by  $\sigma_{\mathcal{R}}$  into similar points. Recent work [9] further suggests that the success of the inference only requires that the similarity knowledge

185 is transferred locally, *i.e.*, that the inference succeeds once maximally similar situations in the situation space are associated with similar outcomes in the outcome space.

As we will see, the case-based prediction methods differ by the type of knowl-

edge used to express the compatibility requirement between the two similarity  
190 measures. The next section proposes a typology of case-based prediction meth-  
ods, where the methods are distinguished depending on the type of compatibility  
knowledge used to express this compatibility requirement.

### 3. Proposed Typology of Case-Based Prediction Methods

This section presents a systematic analysis of existing case-based prediction  
195 theories along the lines developed in Sec. 2.

#### 3.1. Methodology

As the goal of the survey is to outline the commonalities and differences be-  
tween the computational theories of case-based prediction, all works describing  
representation and knowledge engineering issues (such as, e.g., learning adap-  
200 tation rules) are excluded from the survey. We also exclude the works that  
study how to apply a theory of transfer to handle a particular type of input  
(e.g., decision problems [33]), to derive a particular type of information (e.g., a  
preference ordering [39, 69]), or to take into account domain knowledge [82].

However, we include a few case-based adaptation methods (see Sec. 6.2).  
205 Adaptation is the cognitive ability to envision a target solution that is dif-  
ferent from any previously encountered solution. Although adaptations may  
involve structural changes that go way beyond predicting a class or a value,  
the case-based adaptation approaches reported here appear relevant for this  
study because they are developed to predict a single value by applying a set of  
210 adaptation rules. Reporting such methods is also a way to show the intimate  
links between case-based adaptation and case-based prediction, and the fuzzy  
continuum that exists between these two domains of research.

#### 3.2. Discrimination Criteria

Many criteria can be used to categorize the different approaches. A few of  
215 them are reported here and applied in Sec. 3.3, leading to the four categories

discussed in turn in Sec. 4 to 7. Note that these distinctions are best viewed as fuzzy continua, rather than well-defined subtypes.

*Type of Compatibility Knowledge.* A decisive criterion is the type of knowledge that is used to measure the compatibility between  $\sigma_S$  and  $\sigma_R$ . Such compatibility knowledge may take the form of a set of adaptation rules, a continuity constraint, a compatibility indicator between cases or a global compatibility function, as discussed in the next sections.

*Prediction Strategy.* Depending on the type of compatibility knowledge that is used, different strategies are applied to evaluate which potential outcome is the most plausible. For example, adaptation rules are applied on the new case in a form of similarity-based reasoning, in order to derive an outcome that satisfies the consequent of the rule when the antecedent is verified (or approximately verified). Continuity constraints are used to exclude the outcomes that are not similar enough to the outcomes of the retrieved cases. Compatibility indicators are maximized in order to determine which potential new case is most compatible with the retrieved cases. A global compatibility function is maximized in order to determine which completed case base makes  $\sigma_R$  most compatible with  $\sigma_S$  on the case base as a whole.

*Knowledge-driven vs Data-driven Strategy.* The different approaches can be interpreted in a bipolar framework [94]. *Knowledge-driven* strategies consider the compatibility knowledge as negative information, by taking it as a constraint that the two sets of similarity relations should satisfy. Such constraint can be expressed e.g., as a fuzzy implication rule (as in some *transfer by constraint* approaches) or as a set of adaptation rules (as in some *transfer by approximate reasoning* approaches). The predicted outcome is the one for which the added similarity relations is the most consistent with the constraint(s). *Data-driven* strategies, on the contrary, consider an observed compatibility between  $\sigma_S(s, t)$  and  $\sigma_R(r_s, r)$  as positive information in favor of a potential outcome  $r \in \mathcal{R}$ . An indicator, such as a joint similarity measure (as in *transfer by evidence support*

245 approaches), or a global compatibility function (as in *transfer by global optimization* approaches), is used to aggregate the observed evidence in favor of each potential outcome, and the most plausible outcome is determined by a majority vote. This distinction between knowledge-driven and data-driven strategies is not always obvious, and some methods may be considered as belonging to both  
250 categories. For example, the global optimization method proposed in [6] defines an indicator that counts the number of times a set of continuity constraints are verified on a potential case base, and uses this indicator as positive information in favor of a potential outcome.

*Local vs Global Compatibility Estimation.* Assuming that  $\sigma_S$  and  $\sigma_R$  are defined  
255 on different sets of attributes, the compatibility between two similarity measures can not be evaluated *per se*, but only relatively to a given set of pairs of cases. In this respect, most approaches start with a set of local compatibility estimations, and then aggregate the results. Each potential new case  $\hat{c}_t$  is compared to the retrieved cases (which amounts to comparing the added similarity relations  
260  $\Sigma_S(\hat{c}_t)$  and  $\Sigma_R(\hat{c}_t)$  pairwise), and the results of this estimation is aggregated in order to determine the most plausible potential outcome. On the contrary, the *transfer by global optimization* approach performs a single global compatibility estimation. It considers the effect of a choice for  $r$  on the compatibility of  $\sigma_R$  with  $\sigma_S$  estimated on the case base as a whole, and takes into account in the  
265 compatibility estimation some pairs of cases in which  $c_t$  does *not* appear.

*Ordinal vs Numeric Strategies.* The approaches also differ by the way similarity relations are compared in order to produce a compatibility estimation. Some approaches such as *transfer by approximate reasoning* approaches compare the values of the two similarity measures pairwise. By contrast, some approaches  
270 only consider the similarity orderings to make predictions. Examples include some *transfer by constraint* approaches such as the credible case-based inference [66] or the *transfer by global optimization* method proposed in [6].

Approach	Compatibility knowledge	Prediction strategy
Evidence support	A joint similarity measure, that measures how compatible $\sigma_R$ is with $\sigma_S$ for a given pair of cases.	Find the case that is most compatible with the retrieved cases.
Continuity constraints	A set of continuity constraints, <i>i.e.</i> , rules that state that $\sigma_R$ should be compatible with $\sigma_S$ on each pair of cases.	Exclude the outcomes that are not similar enough to the outcomes of the retrieved cases.
Approximate reasoning	A set of rules of the form $(\sigma_S = \alpha) \rightarrow (\sigma_R = \beta)$ .	Make a majority vote on the outcomes derived from the rules.
Global optimization	A global function, that measures how compatible $\sigma_R$ is with $\sigma_S$ on the whole case base.	Optimize the global compatibility measure on the augmented case base.

Table 1: Proposed typology of case-based prediction theories.

### 3.3. Proposed Typology of Case-Based Prediction Theories

Four categories of methods are identified and summarized, respectively named  
 275 transfer by evidence support, transfer by constraint, transfer by approximate  
 reasoning, and transfer by global optimization. They differ by the type of com-  
 patibility knowledge that is used. The typology is shown in Tab. 1 and discussed  
 in turn in Sec. 4 to 7.

*Transfer by Evidence Support.* This type of data-driven approach consists in us-  
 280 ing a joint similarity measure to estimate for each pair of cases  $(c_s, \hat{c}_t)$  how com-  
 patible the similarity relation  $\sigma_R(r_s, r)$  is with the similarity relation  $\sigma_S(s, t)$ .  
 Examples include the  $k$ -Nearest Neighbor algorithm or the Possibilistic Instance-  
 Based Learning approach [13, 32, 64]. In these approaches, a new case is con-  
 sidered possible if the existence of a similar case is confirmed by observation.  
 285 The value of the joint similarity measure is interpreted as a degree of *confirma-*

tion, or *evidence support* that the new case is supported by the retrieved source cases. The predicted outcome  $r_t$  is the one for which the maximal compatibility is observed with a source case.

*Transfer by Continuity Constraints.* This strategy, which is more on the side of  
290 knowledge-driven approaches, consists in expressing the compatibility requirement between the two similarity measures  $\sigma_S$  and  $\sigma_R$  as a set of continuity constraints à la Lipschitz [13], for instance of the form  $\sigma_R(r_s, r_t) \geq h(\sigma_S(s, t))$ , where  $h$  is a transformation function that contains the provided information about the relation between  $\sigma_S$  and  $\sigma_R$ . Examples include similarity profiles [66]  
295 or fuzzy implication rules [67, 70, 72]. Such constraints are used to reduce the set of potential outcomes, excluding the ones that violate them. The predicted outcome is chosen among the potential outcomes that are consistent with all constraints.

*Transfer by Approximate Reasoning.* This type of approach consists in searching  
300 where the two similarity measures  $\sigma_S$  and  $\sigma_R$  align locally, and reason by similarity on the found alignments. Potential outcomes for the new case are derived in a rule-based approach by applying a set of rules of the form  $(\sigma_S = \alpha) \rightarrow (\sigma_R = \beta)$ , such as adaptation rules [7, 27, 73, 90], co-variations [5], or dependencies between analogical proportions [9, 17]. Some case-based adaptation approaches  
305 implement this strategy, as well as analogical proportion-based classifiers.

*Transfer by Global Optimization.* In most case-based prediction approaches, the compatibility of  $\sigma_R$  with  $\sigma_S$  is evaluated on the pair of cases  $(c_s, \hat{c}_t)$  for each retrieved case  $c_s$ , and the results are combined in order to find the most plausible outcome  $r$  for the new case. A recent work [6] proposes to define a  
310 global indicator that measures the compatibility of  $\sigma_R$  with  $\sigma_S$  on the whole case base. The prediction strategy consists in minimizing the value of this indicator on the augmented case base.

#### 4. Transfer by Evidence Support

In this type of approach, the compatibility of  $\sigma_R$  with  $\sigma_S$  is considered as  
315 a *positive* constraint, according to which the more similar two situations are,  
the more plausible it is that their outcomes are similar. The **compatibility  
knowledge** used in the inference takes the form of a compatibility indicator  $\sigma$   
between cases, defined as a joint similarity measure [2, 13, 32, 64]. The **pre-  
320 diction strategy** consists in choosing the new case  $c_t = (t, r_t)$  that is the  
most compatible with the retrieved source cases according to the compatibility  
indicator.

##### 4.1. General Principles

The compatibility indicator  $\sigma$  measures the compatibility of  $\sigma_R$  with  $\sigma_S$  on  
a pair of cases  $(c_s, \hat{c}_t)$ , and is used as an indicator of the plausibility of a new  
325 case  $\hat{c}_t$  when compared to a retrieved case  $c_s$ . These plausibility estimations are  
then aggregated on a selected set of source cases. The general idea is therefore  
to successively:

1. compare the potential new case  $\hat{c}_t$  to a set of source cases  $c_s$ ;
2. aggregate the values of the compatibility indicator  $\sigma$  for the pairs of cases  
330  $(c_s, \hat{c}_t)$ ;
3. predict the outcome  $r_t$  that makes the new case  $\hat{c}_t = (t, r)$  most compatible  
with the retrieved source cases.

In the rest of the section, two approaches of this category are described in  
more details: the  $k$ -Nearest Neighbors approach and the Possibilist Instance-  
335 Based Learning method.

##### 4.2. $k$ -Nearest Neighbors

In a classification setting, the  $k$ -Nearest Neighbor approach [2] makes a ma-  
jority vote among the classes of the  $k$  nearest neighbors of the target situation  $t$   
in order to predict its class  $r_t$ .

A source case  $c_s = (s, r_s)$  is considered compatible with a new case  $c_t = (t, r)$  if it is among the  $k$  nearest neighbors of  $c_t$  for  $\sigma_S$  (i.e.,  $\sigma_S(s, t) = 1$  if  $s$  is in a neighborhood  $\mathcal{N}_k(t)$  of  $t$ , and 0 otherwise), and belongs to the same class (i.e.,  $\sigma_R(r_s, r) = 1$  if the two classes are the same, and 0 otherwise). The compatibility indicator  $\sigma$  is the joint similarity measure:

$$\sigma(c_s, c_t) = \sigma_S(s, t) \cdot \sigma_R(r_s, r)$$

The values of the compatibility indicator  $\sigma$  are aggregated by summing over all retrieved source cases  $c_s$ , and the predicted outcome  $r_t$  is the one that makes the new case  $c_t$  most compatible with all source cases, so that

$$r_t = \arg \max_{r \in \mathcal{R}} \left( \sum_{c_s \in \text{CB}} \sigma_S(s, t) \cdot \sigma_R(r_s, r) \right)$$

#### 340 4.3. Possibilist Instance-Based Learning

A possibilist counterpart of the previous approach consists in considering a relaxed expression of the relation rule, of the form "the more similar two situations are, the more *possible* it is that their outcomes are similar" [13, 32, 36, 64, 70, 71]. Such a rule is formalized in the formal framework of possibility theory (see e.g. [37]), which constitutes an uncertainty modeling paradigm that generalizes the probability theory. More precisely, such a rule constrains the possibility distribution on potential new cases  $c_t = (t, r_t)$ , i.e. the possibility degrees  $\delta(c_t)$ : each source case  $c_s = (s, r_s)$  is then associated with the following constraint on the possibility distribution:

$$\delta(c_t) \geq \min\{\sigma_S(s, t), \sigma_R(r_s, r)\}$$

The constraint expresses that a lower bound of the degree of possibility  $\delta(c_t)$  of a new case  $c_t$  is given by the value of the joint similarity measure

$$\sigma(c_s, c_t) = \min\{\sigma_S(s, t), \sigma_R(r_s, r)\}$$



The possibility degree  $\delta(c_t)$  is interpreted as a degree of *confirmation*, or *evidence support* that the new case  $c_t$  is supported by the retrieved source cases. A new case is considered possible if the existence of a similar case is confirmed by observation, and data accumulation can only result in increasing the support for the new case. Therefore, the possibility degrees are aggregated using a principle of maximal informativeness:

$$\delta(c_t) \geq \max_{c_s \in CB} \sigma(c_s, c_t)$$

The predicted outcome  $r_t$  is the one that is most supported by the retrieved source cases, *i.e.*,

$$r_t = \arg \max_{r \in \mathcal{R}} \left( \max_{c_s \in CB} \min_r \{ \sigma_S(s, t), \sigma_R(r_s, r) \} \right)$$

## 5. Transfer by Constraint

A complementary, more knowledge-driven approach, consists in taking the compatibility of  $\sigma_R$  with  $\sigma_S$  as a *negative* constraint, that is used to exclude the outcomes that are not similar enough to the outcomes of a retrieved case.

345 The **compatibility knowledge** used in the inference is a set of continuity constraints such as similarity profiles [62, 65, 66], or fuzzy implication rules [31, 67, 71, 72]. Such a continuity constraint is interpreted as a negative information according to which it is not plausible to observe situations very dissimilar for  $\sigma_R$  when they are similar for  $\sigma_S$  [13]. The **prediction strategy** consists in using these constraints to exclude the potential outcomes that violate them.

350 For each retrieved case  $c_s = (s, r_s)$  and a potential new case  $\hat{c}_t = (t, r)$ , the compatibility of the outcome  $r$  with a continuity constraint is estimated by testing whether the pair  $(c_s, \hat{c}_t)$  satisfies the constraint or not. The resulting compatibility estimations are then aggregated on all retrieved cases  $c_s$ . The predicted

355 outcome  $r_t$  is then chosen among the outcomes  $r$  that are most compatible with the constraints.

### 5.1. Credible Case-Based Inference

Continuity constraints can be expressed by stating that if two situations are above a similarity level  $\alpha$  for  $\sigma_S$ , it is likely that their similarity for  $\sigma_R$  is greater or equal than a value  $\beta$ . The function  $h : [0, 1] \rightarrow [0, 1]$  which associates to each similarity level  $\alpha$  for  $\sigma_S$  a similarity level  $\beta$  for  $\sigma_R$  is called a *similarity profile* [62, 65, 66]. It is defined as  $h(\alpha) = \inf\{\sigma_R(r_s, r_{s'}) \mid \sigma_S(s, s') = \alpha\}$ . Assuming that  $h$  is known, one can compute for a new situation  $t$  the set

$$C(t) = \bigcap_{c_s \in CB} \{r : \sigma_R(r_s, r) \geq h(\sigma_S(s, t))\}$$

of *credible* solutions by taking, for each retrieved source case  $c_s$ , the set of outcomes  $r$  that would satisfy the constraint. The problem is then to learn the similarity profile  $h$ . In [67], this function is approximated by a step function, which is learned from the data. In [3], each hypothesis for  $h$  is a multi-category classifier. Determining the levels  $\beta$  of a similarity profile from the data is a task that is very sensitive to outliers. Two solutions are investigated in [65] to alleviate this issue: learning local similarity profiles for different regions of space, or weaken the concept of similarity profile by looking for levels that are “almost valid”, e.g., by defining a probabilistic similarity profile [62]. Besides, enough data is needed to learn the profiles. The credible case-based inference has been proven to be a special case of integrity constraint belief merging [21].

### 5.2. Gradual Rules

Gradual rules [34] are linguistically expressed as “the more  $X$  is  $A$ , the more  $Y$  is  $B$ ” and modelled in the formal framework of fuzzy set theory (see e.g. [35]):  $A$  and  $B$  are imprecisely defined concepts, modeled as fuzzy sets on the universes of  $X$  and  $Y$  respectively. The general semantics of such rules [34] is expressed in terms of membership degrees by  $B(Y) \geq A(X)$ , which is equivalent to a set of constraints of the form  $(X \in A_\alpha) \rightarrow (Y \in B_\alpha)$  stating that if the membership degree of  $X$  in  $A$  is at least  $\alpha$ , then it is guaranteed that the membership degree of  $Y$  in  $B$  is also at least  $\alpha$ . A specific case of such gradual rules is linguistically

expressed as "the higher  $X$ , the higher  $Y$ ", whose semantics directly applies to the numerical values taken by  $X$  and  $Y$ , instead of their membership degrees to the concepts  $A$  and  $B$ .

Such gradual rules can be used to express some continuity constraints that should be satisfied by a pair of cases  $(c_s, c_t)$ : the constraint "the more similar two situations are, the more similar are their associated outcomes" can be formally defined as

$$\sigma_R(r_s, r) \geq \sigma_S(s, t)$$

The transfer inference consists, as in the credible case-based inference, in choosing a potential outcome  $r_t$  among the ones that satisfy the constraint for all source cases, that is,

$$r_t \in \bigcap_{c_s \in CB} \{r : \sigma_R(r_s, r) \geq \sigma_S(s, t)\}$$

Although the approach can be refined by applying a non-decreasing function  $h : [0, 1] \rightarrow [0, 1]$  to define rules of the form  $\sigma_R(r_s, r) \geq h(\sigma_S(s, t))$ , the gradual rule approach is not very flexible because it is very sensitive to outliers. Indeed, an outcome  $r$  is ruled out of the set of potential results whenever the constraint  $\sigma_R(r_s, r) \geq \sigma_S(s, t)$  is not satisfied for at least *one* source case. This remark advocates for the use of a different type of fuzzy rule, such as a certainty rule, which allow for exceptional situations.

### 5.3. Certainty Rules

Certainty rules are linguistically expressed as "the more  $X$  is  $A$ , the more certain  $Y$  lies in  $B$ ". Their semantics is modeled in the possibility theory framework, formalized by the following constraint on the conditional possibility distribution  $\pi_{Y|X}$ :

$$\forall(u, v), \pi_{Y|X}(v | u) \leq \max(1 - A(u), B(v))$$

Such a constraint implies that  $1 - A(u)$  is an upper bound of the possibility that  
 390  $Y = v$  when  $v$  is not in the support of  $B$  (*i.e.*, when  $B(v) = 0$ ).

Thus the rule "the larger the similarity of two situations is, the more certain it is that the similarity of corresponding outcomes is large" can be formalized as

$$\pi(r | t) = \pi_{\sigma_R | \sigma_S}(\sigma_R(r_s, r) | \sigma_S(s, t)) \leq \max(1 - \sigma_S(s, t), \sigma_R(r_s, r))$$

This formalization takes the situation dissimilarity  $1 - \sigma_S(s, t)$  as an upper bound for the possibility of an outcome  $r$  when  $\sigma_R(r_s, r) = 0$ . If  $\sigma_S(s, t)$  is very small, the possibility bound can be large for very dissimilar outcomes. If, on the contrary,  $\sigma_S(s, t)$  is close to 1, the possibility bound can only be large for  
 395 very similar outcomes.

The possibility degree  $\pi(r | t)$  is interpreted as the degree to which the comparison of  $c_t$  with the retrieved source cases does not exclude the outcome  $r$  as a candidate. A new case  $c_t$  is considered possible if the application of a continuity constraint does not rule it out as having an outcome too dissimilar with the outcome of a retrieved source case, and data accumulation can only result in decreasing the possibility of certain outcomes. Therefore, the possibility degrees are aggregated using a principle of minimal specificity:

$$\pi(r | t) = \min_{c_s \in CB} (\max(1 - \sigma_S(s, t), \sigma_R(r_s, r)))$$

The predicted outcome  $r_t$  is the one that is the most possible given the retrieved source cases:

$$r_t = \arg \max_{r \in \mathcal{R}} (\min_{c_s \in CB} (\max(1 - \sigma_S(s, t), \sigma_R(r_s, r))))$$

## 6. Transfer by Approximate Reasoning

The third type of approach consists in searching where the two similarity measures  $\sigma_S$  and  $\sigma_R$  align locally, and reason by similarity on the found

alignments. The **compatibility knowledge** takes the form of a set of rules  
 400 providing information on relations between the similarity measures  $\sigma_S$  and  $\sigma_R$ ,  
 expressing that when  $\sigma_S$  takes value  $\alpha$ , the resulting similarity level for  $\sigma_R$  is  $\beta$ :  
 these rules can be written  $(\sigma_S = \alpha) \rightarrow (\sigma_R = \beta)$  and can be expressed in various  
 forms, such as adaptation rules [7, 27, 73, 90], dependencies between problem  
 and solution features [42], co-variations [5] or fuzzy rules [16] to name a few. The  
 405 **prediction strategy** consists in triggering the rules on pairs of cases involving  
 the new case using a kind of similarity-based inference, as detailed below, in  
 order to derive potential outcomes for the new case.

### 6.1. General Principles

A rule  $(\sigma_S = \alpha) \rightarrow (\sigma_R = \beta)$  is a piece of knowledge that states that  
 $\sigma_R$  is compatible with  $\sigma_S$  when  $\sigma_S$  takes the value  $\alpha$ , and that the resulting  
 similarity level for  $\sigma_R$  is  $\beta$ . Potential outcomes  $r$  for the new case are derived  
 by triggering such rules on pairs of cases involving the new case in a form  
 of similarity-based inference (SBI), by applying variants of the modus ponens  
 schema [11, 16, 26, 98]: for a retrieved case  $c_s = (s, r_s)$  and a potential new case  
 $\hat{c}_t = (t, r)$ , triggering the rule  $(\sigma_S = \alpha) \rightarrow (\sigma_R = \beta)$  on the pair of cases  $(c_s, \hat{c}_t)$   
 is of the form

$$\frac{(\sigma_S = \alpha) \rightarrow (\sigma_R = \beta) \quad \sigma_S(s, t) \approx \alpha}{\sigma_R(r_s, r) \approx \beta} \quad (\text{SBI})$$

This schema expresses that if the rule associates a level  $\beta$  for the similarity  $\sigma_R$   
 410 whenever the similarity level for  $\sigma_S$  is  $\alpha$ , and if the observed situation similarity  
 between the new case and a retrieved case is approximately  $\alpha$  (*i.e.*,  $\sigma_S(s, t) \approx \alpha$ ),  
 then the corresponding similarity on outcomes  $\sigma_R(r_s, r)$  is approximately  $\beta$ . The  
 concepts "approximately  $x$ " where  $x$  is a numerical value, are imprecisely defined  
 concepts that can be modeled in the formal framework of fuzzy set theory [79].

415 It is often the case that the similarity measures  $\sigma_S$  and  $\sigma_R$  are unknown, or  
 difficult to assess globally on the training data. One strategy then consists in  
 working with some local approximations  $\widetilde{\sigma}_S$  of  $\sigma_S$  and  $\widetilde{\sigma}_R$  of  $\sigma_R$  that are known

to be compatible for some pairs of cases of the case base. The resulting rules  $(\widetilde{\sigma}_S = \alpha) \rightarrow (\widetilde{\sigma}_R = \beta)$  are *adaptation rules* (see Sec. 6.2).

420 Some case-based prediction approaches such as [73] include a rule selection step prior to the inference, while others such as proportion-based analogical classifiers (see Sec. 6.4) trigger only one rule.

Each selected rule is triggered on a set of pairs of cases  $(c_s, c_t)$ , and the proposed outcome  $r_t$  for the new case  $c_t$  is obtained by a majority vote among the potential outcomes  $r$  derived from the rules:

$$r_t = \arg \max_{r \in \mathcal{R}} \sum_{(\sigma_S = \alpha) \rightarrow (\sigma_R = \beta)} |c_s \in CB \mid \sigma_S(s, t) \approx \alpha \text{ and } \sigma_R(r_s, r) \approx \beta|$$

In the rest of the section, three kind of approaches of this category are described in more details: rule-based adaptation, the analogical jump, and ana-  
425 logical proportion-based classification.

### 6.2. Rule-Based Adaptation

Adaptation rules are rules of the form  $(\widetilde{\sigma}_S = \alpha) \rightarrow (\widetilde{\sigma}_R = \beta)$ , where  $\widetilde{\sigma}_S$  and  $\widetilde{\sigma}_R$  are local approximations of  $\sigma_S$  and  $\sigma_R$ . The main difficulty when working with adaptation rules is that one needs to be able to learn the rules. They  
430 may be acquired from different sources such as a domain expert [81, 91], the user [8], or learned from data [7, 25, 27, 53, 73, 86]. In [91] for example, some adaptation rules are learned from the expert in the form of qualitative proportionalities  $y = qprop^+(x)$ . A qualitative proportionality is a qualitative constraint that indicates a co-monotony between two variables, such as "a larger  
435 apartment has a higher rent". The relationship between the two variables (for example here, `nb_rooms` and `price`) is assumed to be linear, and the ratio coefficient is learned by linear regression. If  $Q$  denotes the ratio coefficient, this amounts to defining a similarity  $\sigma_S$  on the values of the attribute `nb_rooms`, a similarity  $\sigma_R$  on the values of the attribute `price`, and applying the (SBI)  
440 inference schema with the hypothesis that  $\sigma_R$  increases linearly with  $\sigma_S$ , *i.e.*, with the hypothesis that  $\beta = Q\alpha$ . Another example is the work of [86], which

learns adaptation rules from data at runtime by considering only the pairs of situations  $(s, t)$  in which  $s$  differs from  $t$  only in the value of a single attribute (say, `nb_rooms` in the previous example), and retrieves from the case base another pair of situations  $(s_i, s_j)$  where the same difference is observed on this attribute. This approach amounts to learning a rule  $(\widetilde{\sigma}_S = \alpha) \rightarrow (\widetilde{\sigma}_R = \beta)$  by single instance induction, once the rule is verified on only one pair of cases.

A key question is to decide which rule can be considered as a valid piece of knowledge to be used in the similarity-based inference. Taking into account the support of the rule is important. Even a very specific rule, when learned with a support of 1 as in the previous paragraph, may be dubious, because it amounts to single instance induction. Several works [11, 17, 28, 91] have emphasized the idea that the rules  $(\widetilde{\sigma}_S = \alpha) \rightarrow (\widetilde{\sigma}_R = \beta)$  that are reasoned upon should be functional dependencies, *i.e.*, have a confidence value of 1 on any two pairs of cases of the case base. From this observation, [5] defined a *variation* as any function that associates a value to a pair of situations, and a *co-variation* as a functional dependency between variations. For example, the variation  $\text{nb\_rooms}^{\leq} : \mathcal{S} \times \mathcal{S} \mapsto \{0, 1\}$  maps a pair of apartments to 1 if the number of rooms increases, and to 0 otherwise, and the co-variation  $(\text{nb\_rooms}^{\leq} = 1) \rightarrow (\text{price}^{\leq} = 1)$  expresses that if the number of rooms of an apartment increases, then the price also increases. This rule can be used to draw the following similarity-based inference: if a pair of cases verifies  $\text{nb\_rooms}(s) \leq \text{nb\_rooms}(t)$ , then the antecedent of the rule is satisfied, so one can make the hypothesis that the rule applies, and that  $\text{price}(r_s) \leq \text{price}(r)$ . In this example, the inference schema SBI writes:

$$\frac{(\text{nb\_rooms}^{\leq} = 1) \rightarrow (\text{price}^{\leq} = 1) \quad \text{nb\_rooms}^{\leq}(s, t) = 1}{\text{price}^{\leq}(r_s, r) = 1}$$

Learning such co-variations from data corresponds to the task called gradual pattern mining, for which various meanings and approaches have been proposed [5, 14, 29, 63, 78].

### 6.3. The “Analogical Jump”

A crude version of the transfer by approximate reasoning approach was studied from a logical point of view in the 1980s. In their seminal work on the formalization of analogical transfer, the authors of [28] introduce an inference schema, called the “analogical jump” (AJ):

$$\frac{P(c_s) \quad P(c_t) \quad Q(c_s)}{Q(c_t)} \quad (\text{AJ})$$

According to this schema, analogical transfer consists in making the hypothesis that if  $c_s$  and  $c_t$  share some property  $P$ , and a property  $Q$  is true in  $c_s$ , then it is plausible that the property  $Q$  is also true in  $c_t$ . The problem identified by the authors is then to determine sufficient conditions for the inference (AJ) to be drawn. The authors remark that the rule should:

- (i) be weaker than a generalization rule  $\forall x P(x) \Rightarrow Q(x)$  (otherwise, the inference is simply deductive),
- (ii) on the contrary, be stronger than single instance induction, which would consist in applying the rule  $\forall x \forall y [(P(x) \wedge P(y) \wedge Q(x)) \Rightarrow Q(y)]$ , and
- (iii) take into account the level of similarity between  $c_s$  and  $c_t$ .

The analogical jump can be formalized as rule-based adaptation, considering the adaptation rule  $(\widetilde{\sigma}_S = 1) \rightarrow (\widetilde{\sigma}_R = 1)$ , with

$$\widetilde{\sigma}_S(s, t) = \mathbf{1}_P(s, t) = \begin{cases} 1 & \text{if both } P(s) \text{ and } P(t) \text{ hold} \\ 0 & \text{otherwise} \end{cases}$$

$$\widetilde{\sigma}_R(r_s, r) = \mathbf{1}_Q(r_s, r) = \begin{cases} 1 & \text{if both } Q(r_s) \text{ and } Q(r) \text{ hold} \\ 0 & \text{otherwise} \end{cases}$$

For a retrieved case  $c_s = (s, r_s)$  and a potential new case  $\hat{c}_t = (t, r)$ , drawing the inference schema (AJ) amounts to applying the following version of the (SBI)



inference schema:

$$\frac{(\widetilde{\sigma}_S = 1) \rightarrow (\widetilde{\sigma}_R = 1) \quad \widetilde{\sigma}_S(s, t) = 1}{\widetilde{\sigma}_R(r_s, r) = 1}$$

This inference schema states that if we know that the property  $Q$  is shared between two cases whenever the property  $P$  is shared ( $(\widetilde{\sigma}_S = 1) \rightarrow (\sigma_R = 1)$ ) and if it was observed that the two cases share the property  $P$  ( $\widetilde{\sigma}_S(s, t) = 1$ ) then  
 465 one can make the hypothesis that they also share the property  $Q$  ( $\sigma_R(r_s, r) = 1$ ).

The key question is to decide if it is legitimate to consider the rule  $(\widetilde{\sigma}_S = 1) \rightarrow (\widetilde{\sigma}_R = 1)$  as a valid piece of knowledge from which to derive that  $Q(c_t)$  holds. If one requires that the rule is verified on only one pair of cases, then it amounts to single instance induction. If one requires that the rule has a  
 470 confidence value of 1, then the rule  $\forall x P(x) \Rightarrow Q(x)$  is in particular valid on all source cases, and the SBI inference amounts to simple deduction.

#### 6.4. Analogical Proportion-Based Classification (APC)

This section reports the work presented in [9], that proposes to establish a correspondence between analogical proportion-based classification (denoted by  
 475 APC, in what follows) and case-based prediction, showing the former can be viewed as a special kind of the latter.

APC algorithms have shown competitive results in classification and recommendation tasks, see e.g. [12, 17, 23, 24, 61, 87]. They apply the principle of analogical reasoning [56], based on statements of the form "a is to b as c  
 480 is to d", called analogical proportion, and written  $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$ . More precisely, the analogical inference is applied in a classification setting to state that if an analogical proportion holds on the instance descriptions, then an analogical proportion can be inferred on their associated class labels: formally, denoting  $f$  the underlying, unknown, labelling function, one can derive from  
 485  $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$  that  $f(\mathbf{a}) : f(\mathbf{b}) :: f(\mathbf{c}) : f(\mathbf{d})$ . Let  $D$  be a data set containing a set of instances  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \dots$  with their associated labels  $f(\mathbf{a}), f(\mathbf{b}), f(\mathbf{c}), \dots$  To

predict the value  $f(x)$  for a new instance  $x$ , an analogical proportion-based classifier considers all triples  $(\mathbf{a}, \mathbf{b}, \mathbf{c}) \in D^3$  for which  $\mathbf{a} : \mathbf{b} :: \mathbf{c} : x$  holds, and the equation  $f(\mathbf{a}) : f(\mathbf{b}) :: f(\mathbf{c}) : y$  has a solution. This set of triples is called the

490 *analogical root* of  $x$  [60]. The predicted label for the new instance  $x$  is then the result of a majority vote among the potential solutions  $y$ . Yet it can be the case that the analogical root is empty: the previous classifier can then be extended to consider approximate analogy, relying on the notion of analogical dissimilarity [60]. The latter is defined as a function  $AD(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d})$  that quan-

495 tifies the extent to which the quadruplet is far from satisfying an analogical proportion:  $AD$  is such that  $AD(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) = 0$  iff  $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$  and satisfies constraints on argument permutation and a triangular inequality [12]. For real or Boolean values, it can for instance be defined as the sum of the componentwise

$AD(a, b, c, d) = \|(a - b) - (c - d)\|_1$ . If the analogical root of  $x$  is empty, the

500 search for potential solutions is extended to triples  $(\mathbf{a}, \mathbf{b}, \mathbf{c})$  with the  $k$  least values of  $AD(\mathbf{a}, \mathbf{b}, \mathbf{c}, x)$  and for which the equation  $f(\mathbf{a}) : f(\mathbf{b}) :: f(\mathbf{c}) : y$  has a solution. The predicted label is the result of a majority vote among the potential solutions  $y$ .

This correspondence between APC and case-based prediction is illustrated

505 by the diagram given in Fig. 3 that represents the APC in a similar view as case-based prediction, whose diagram is given in Fig. 1. More precisely, APC can be considered as applying a specific transfer by approximate reasoning method, where cases are *differences*, or *ratios* between two instances, and a single rule is triggered, that states that maximally similar situations should be associated

510 with maximally similar outcomes.

♦ When seen as a case-based prediction method, APC works by comparing some *ratios*  $\mathbf{a} : \mathbf{b}$  and  $f(\mathbf{a}) : f(\mathbf{b})$  between the instances and their respective labels. Assuming that both instances and labels are vectors, considering one-hot encoding for the classes, these ratios are represented by the differences  $s = \mathbf{a} - \mathbf{b}$  and  $r_s = f(\mathbf{a}) - f(\mathbf{b})$ . Let us denote by  $x \in D$  a new instance for which the class  $f(x)$  is to be predicted. Let  $C$  be the set of potential classes for  $f(x)$ , and

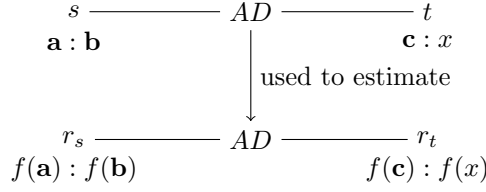


Figure 3: In APC methods, both situations and outcomes represent ratios.

$y \in C$ . The source case  $c_s$  and potential new case  $\hat{c}_t$  are of the following form:

$$\begin{aligned} c_s &= (\mathbf{a} - \mathbf{b}, f(\mathbf{a}) - f(\mathbf{b})) \\ \hat{c}_t &= (\mathbf{c} - x, f(\mathbf{c}) - y) \end{aligned} \quad (1)$$

where  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  are instances of  $D$ , and  $f(\mathbf{a})$ ,  $f(\mathbf{b})$ ,  $f(\mathbf{c})$  their associated classes.

The two similarity measures  $\sigma_S$  and  $\sigma_R$  are constructed from the analogical dissimilarity  $AD$ , by noticing that  $AD$  measures a distance  $AD(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) = \delta(\mathbf{a} - \mathbf{b}, \mathbf{c} - \mathbf{d})$  between two differences  $\mathbf{a} - \mathbf{b}$  and  $\mathbf{c} - \mathbf{d}$ . The similarity measures  $\sigma_S$  and  $\sigma_R$  are obtained by applying a strictly decreasing function to the distance  $\delta$ , e.g., by choosing  $\sigma_S = \sigma_R = e^{-\delta}$ . The similarity measure  $\sigma_S$  is such that the four instances  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$  form an analogical proportion iff  $\sigma_S(\mathbf{a} - \mathbf{b}, \mathbf{c} - \mathbf{d}) = 1$ . The similarity measure  $\sigma_R$  is such that the four instances  $f(\mathbf{a}), f(\mathbf{b}), f(\mathbf{c}), f(\mathbf{d})$  form an analogical proportion iff  $\sigma_R(f(\mathbf{a}) - f(\mathbf{b}), f(\mathbf{c}) - f(\mathbf{d})) = 1$ .

The transfer strategy can be interpreted as a transfer by approximate reasoning strategy when the prediction procedure is decomposed, as described in [83], an aggregation of the potential solutions  $y$  found for each instance  $\mathbf{c} \in D$  followed by a majority vote. In this view, the search for potential solutions  $y$  consists in successively:

- building the case base  $\{c_s = (s, r_s) = (\mathbf{a} - \mathbf{b}, f(\mathbf{a}) - f(\mathbf{b}))\}$
- enumerating all instances  $\mathbf{c}$ , and for each of them,
  - *Retrieval*: retrieve all source cases  $c_s = (s, r_s)$ ;
  - *Mapping*: compute the similarity  $\sigma_S(s, t)$  between  $s = \mathbf{a} - \mathbf{b}$  and

$$t = \mathbf{c} - x;$$

- 530 – *Transfer*: if  $\sigma_S(s, t) = 1$  holds (*i.e.*,  $\mathbf{a}, \mathbf{b}, \mathbf{c}, x$  are s.t.  $\mathbf{a} : \mathbf{b} :: \mathbf{c} : x$ ), find the solutions  $y$  such that  $\sigma_R(r_s, r) = 1$ , with  $r_s = f(\mathbf{a}) - f(\mathbf{b})$  and  $r = f(\mathbf{c}) - y$ .

This decision procedure thus considers all pairs  $(c_s, \hat{c}_t)$  that can be obtained from a triple  $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ , and searches for potential solutions  $y$  that can be inferred by applying the following similarity-based inference on a pair  $(c_s, \hat{c}_t)$ :

$$\frac{(\sigma_S = 1) \rightarrow (\sigma_R = 1) \quad \sigma_S(s, t) = 1}{\sigma_R(r_s, r) = 1}$$

The analogical root of  $x$  corresponds to the set of triples  $(\mathbf{a}, \mathbf{b}, \mathbf{c})$  for which the similarity-based inference allows to infer a solution  $y$ . The predicted solution  
535  $f(x)$  is the solution  $y$  that is inferred on the maximal number of pairs  $(c_s, \hat{c}_t)$  by triggering the rule.

If the analogical root of  $x$  is empty, analogical classifiers extend the search to triples with lowest analogical dissimilarity, *i.e.*, with highest value for the similarity  $\sigma_S$ . This amounts to relaxing the condition  $\sigma_S(s, t) = 1$  to the condition  $\sigma_S(s, t) \approx 1$ . The similarity-based inference becomes:

$$\frac{(\sigma_S = 1) \rightarrow (\sigma_R = 1) \quad \sigma_S(c_s, \hat{c}_t) \approx 1}{\sigma_R(c_s, \hat{c}_t) = 1}$$

Only the  $k$  solutions  $y$  that are derived from the rule  $(\sigma_S = 1) \rightarrow (\sigma_R = 1)$  with the highest values of  $\sigma_S(c_s, \hat{c}_t)$  are added to the solution set.

## 7. Transfer by Global Optimization

540 A recent work [6, 9, 10] proposes to define a global indicator that measures the compatibility of  $\sigma_R$  with  $\sigma_S$  on the whole case base. The **compatibility knowledge** takes the form of a global function  $\Gamma(\sigma_S, \sigma_R, CB)$ , that measures

the compatibility of  $\sigma_R$  with  $\sigma_S$  on any potential case base  $CB$ . The **prediction strategy** consists in completing the description of a case base in order to  
 545 minimize the value of the global indicator.

This principle is implemented in the **CoAT**, for Complexity-based Analogical Transfer, algorithm [6, 10]. In the **CoAT** method, the compatibility of  $\sigma_R$  with  $\sigma_S$  is measured from an ordinal point of view on the whole case base  $CB$ , by checking if  $\sigma_R$  orders the cases in the same manner as  $\sigma_S$ . The following continuity constraint is tested on each triple of cases  $(c_0, c_i, c_j)$ , with  $c_0 = (s_0, r_0)$ ,  $c_i = (s_i, r_i)$ , and  $c_j = (s_j, r_j)$ :

$$\text{if } \sigma_S(s_0, s_i) \geq \sigma_S(s_0, s_j), \text{ then } \sigma_R(r_0, r_i) \geq \sigma_R(r_0, r_j) \quad (C)$$

The constraint (C) expresses that anytime a situation  $s_i$  is more similar to a situation  $s_0$  than situation  $s_j$ , this order should be preserved on outcomes. A triple  $(c_0, c_i, c_j)$  does *not* satisfy the constraint if situation  $s_i$  is more similar to  $s_0$  than situation  $s_j$  for situations, but less similar for outcomes, *i.e.*, when  $\sigma_S(s_0, s_i) \geq \sigma_S(s_0, s_j)$  and  $\sigma_R(r_0, r_i) < \sigma_R(r_0, r_j)$ . Such a violation of the constraint is called an *inversion of similarity*. A global indicator  $\Gamma(\sigma_S, \sigma_R, CB)$  is introduced, that counts the total number of inversions of similarity observed on a case base  $CB$ :

$$\Gamma(\sigma_S, \sigma_R, CB) = |\{(s_0, r_0), (s_i, r_i), (s_j, r_j) \in CB \times CB \times CB \text{ such that } \sigma_S(s_0, s_i) \geq \sigma_S(s_0, s_j) \text{ and } \sigma_R(r_0, r_i) < \sigma_R(r_0, r_j)\}|$$

When the case base is fully known, except for the outcome  $r_t$  of one case  $c_t = (t, r_t)$ , the transfer inference consists in finding the outcome  $r_t$  that minimizes the value of the  $\Gamma$  indicator:

$$r_t = \arg \min_{r \in \mathcal{R}} \Gamma(\sigma_S, \sigma_R, CB \cup \{(t, r)\})$$

A main difference with other theories of case-based prediction lies in the

set of pairs considered to estimate the compatibility between the two similarity measures  $\sigma_S$  and  $\sigma_R$ . The compatibility estimator  $\Gamma$  considers all triples  $(c_0, c_i, c_j)$ , and checks for each triple if  $\sigma_R$  orders the two pairs of cases  $(c_0, c_i)$  and  $(c_0, c_j)$  in the same way that  $\sigma_S$  does. Therefore, some pairs of cases in which  $c_t$  does not appear are taken into account in the compatibility estimation. This is different from other theories of case-based prediction, in which the compatibility of  $\sigma_R$  with  $\sigma_S$  is estimated solely on the pair of cases  $(c_s, c_t)$  for each retrieved case  $c_s$  (i.e., the two sets of similarity relations  $\Sigma_S(t)$  and  $\Sigma_R(r)$  are compared two by two), and the results are combined in order to find the most plausible outcome  $r$  for the new case.

This approach also shares some commonalities with other approaches. The constraint  $C$  can be seen as a qualitative version of the continuity constraint  $\sigma_R(r_s, r_t) \geq h(\sigma_S(s, t))$  used in transfer by constraint methods (see Sec. 5). The global indicator  $\Gamma$  measures the extent to which this continuity constraint is verified on the whole case base. The inference also consists, as in transfer by evidence support approaches, in optimizing a global indicator. But the indicator is defined on all triples of cases, and not only on the pairs of cases involving  $c_t$ .

## 8. Conclusion and Further Work

This study constitutes the first survey of the wide and rich domain of case-based prediction. At the crossroad between the domains of case-based reasoning and computational analogy, this systematic analysis of the literature both contributes to developing a unifying theory of case-based prediction, and to setting a formal ground to a general theory of analogical transfer in computer science.

Case-based prediction methods are diverse, and therefore developing a unified theory is challenging. The present work makes an important contribution in that direction, by showing that all case-based prediction methods share a common principle, which is to interpret the plausible inference of analogical transfer as a transfer of similarity knowledge that the predicted outcome should optimize. In this respect, all approaches follow the same objective, which is to

find the outcome that makes the similarities in the outcome space most compatible with the observed similarities in the situation space. However, they differ in the way compatibility is measured: depending on the type of approach, the compatibility measure either takes the form of a joint similarity measure, a set of rules, some continuity constraints, or a global indicator. The prediction strategy varies accordingly: it consists either in maximizing an indicator, making a majority vote on the outcomes derived from a set of rules, or finding the outcome that is the most consistent with a set of continuity constraints.

By eliciting some shared principles among case-based prediction methods, this work also contributes to setting a formal ground to the theory of analogical transfer in computer science. Although case-based prediction methods only constitute a subset of analogical transfer methods, which are designed to apply analogical transfer to prediction tasks such as classification or regression tasks, these new insights suggest that it makes sense to model analogical transfer as a transfer of similarity knowledge between two description spaces. These theoretical advances help better understand the role of the similarity knowledge in the inference, and will allow for new developments in the study of analogical transfer. Such advances are needed because analogical transfer methods are gaining attention in many domains, and in particular in machine learning. Its inference principle, which consists in deriving new information from a set of comparisons with previous experiences, is attractive because it allows to produce inferences that are interpretable, take into account a memory of past experiences, allow for creativity, and take into account domain knowledge, context, and similarity.

Further work includes making an extensive survey of analogical transfer methods, that would encompass not only prediction tasks but also interpretation and creativity tasks. It would be interesting to study if the inference principles identified in this paper for case-based prediction methods also apply when analogical transfer is used e.g., for adaptation, or case-based explanation. Another research direction would consist in providing a shared implementation of the main case-based prediction algorithms. It would allow the different algorithms to be tested on real-world scenarii, and compared. Finally, the modeling

of analogical transfer as a transfer of similarity knowledge between two description spaces allows to address one major challenge, which is to learn a similarity measure that is adequate for a given transfer task.

## 610 **References**

- [1] Aamodt, A., & Plaza, E. (1994). Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communications*, 7, 39–59.
- [2] Aha, D. W. (1992). Tolerating noisy, irrelevant and novel attributes in instance-based learning algorithms. *Int. Journal of Man-Machine Studies*, 36, 267–287.
- [3] Anthony, M., & Ratsaby, J. (2015). A probabilistic approach to case-based inference. *Theoretical Computer Science*, 589, 61–75.
- [4] Antić, C. (2022). Analogical Proportions. [arXiv:2006.02854](https://arxiv.org/abs/2006.02854).
- 620 [5] Badra, F. (2016). Reasoning with co-variations. In *Proc. of the 17th Int. Conf. on Artificial Intelligence: Methodology, Systems, and Applications, AIMS A'16*.
- [6] Badra, F. (2020). A Dataset Complexity Measure for Analogical Transfer. In *Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'20* (pp. 1601–1607).
- 625 [7] Badra, F., Bendaoud, R., Bentebibel, R., Champin, P.-A., Cojan, J., Cordier, A., Desprès, S., Jean-Daubias, S., Lieber, J., Meilender, T., Mille, A., Nauer, E., Napoli, A., & Toussaint, Y. (2008). TAAABLE: Text Mining, Ontology Engineering, and Hierarchical Classification for Textual Case-Based Cooking. In M. Schaaf (Ed.), *Workshop Proc. of the 9th European Conf. on Case-Based Reasoning, ECCBR'08* (pp. 219–228).
- 630



- [8] Badra, F., Cordier, A., & Lieber, J. (2009). Opportunistic adaptation knowledge discovery. In *Case-Based Reasoning Research and Development* (pp. 60–74). Springer.
- 635 [9] Badra, F., & Lesot, M.-J. (2022). CoAT-APC: When Analogical Proportion-based Classification Meets Case-Based Prediction. In *ICCBR Analogies'22: Workshop on Analogies: From Theory to Applications at ICCBR-2022*. CEUR Workshop Proceedings (CEUR-WS.org).
- [10] Badra, F., & Lesot, M.-J. (2022). Theoretical and Experimental Study  
640 of a Complexity Measure for Analogical Transfer. In *Int. Conference on Case-based Reasoning (ICCBR)* (pp. 175–189).
- [11] Badra, F., Sedki, K., & Ugon, A. (2018). On the Role of Similarity in Analogical Transfer. In *Case-Based Reasoning Research and Development* (pp. 499–514). Springer volume 11156.
- 645 [12] Bayouhdh, S. (2007). Learning by Analogy: A Classification Rule for Binary and Nominal Data. In *Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'07* (pp. 678–683).
- [13] Beringer, J., & Hüllermeier, E. (2008). Case-based learning in a bipolar possibilistic framework. *Int. Journal of intelligent Systems*, 23, 1119–1134.
- 650 [14] Berzal, F., Cubero, J.-C., Sanchez, D., Vila, M.-A., & Serrano, J. M. (2007). An alternative approach to discover gradual dependencies. *Int. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems (IJUFKS)*, 15, 559–570.
- [15] Billingsley, R., Prade, H., Richard, G., & Williams, M.-A. (2017). Towards  
655 Analogy-Based Decision - A Proposal. In *Proc. of the Int. Conf. on Flexible Query Answering Systems, FQAS'17* (pp. 28–35). Springer volume 10333.
- [16] Bouchon-Meunier, B., & Valverde, L. (1999). A fuzzy approach to analogical reasoning. *Soft Computing*, 3, 141–147.

- [17] Bounhas, M., Prade, H., & Richard, G. (2017). Analogy-based classifiers  
660 for nominal or numerical data. *Int. Journal of Approximate Reasoning*, 91,  
36–55.
- [18] Bridge, D., & Healy, P. (2012). The GhostWriter-2.0 Case-Based Reasoning  
system for making content suggestions to the authors of product reviews.  
*Knowledge-Based Systems*, 29, 93–103.
- 665 [19] Chalmers, D. J., French, R. M., & Hofstadter, D. R. (1992). High-level  
perception, representation, and analogy: A critique of artificial intelligence  
methodology. *J. Exp. Theor. Artif. Intell.*, 4, 185–211.
- [20] Chen, C., Li, O., Tao, C., Barnett, A. J., Su, J., & Rudin, C. (2019). This  
looks like that: Deep learning for interpretable image recognition. In *Proc.*  
670 *of the 33rd Int. Conf. on Neural Information Processing Systems*. Curran  
Associates Inc.
- [21] Cojan, J., & Lieber, J. (2009). Belief Merging-Based Case Combination. In  
L. McGinty, & D. C. Wilson (Eds.), *Case-Based Reasoning Research and  
Development* (pp. 105–119). Springer volume 5650.
- 675 [22] Cordier, A., Dufour-Lussier, V., Lieber, J., Nauer, E., Badra, F., Cojan,  
J., Gaillard, E., Infante-Blanco, L., Molli, P., Napoli, A., & Skaf-Molli, H.  
(2014). Taaable: A case-based system for personalized cooking. In *Studies  
in Computational Intelligence*. volume 494.
- [23] Couceiro, M., Hug, N., Prade, H., & Richard, G. (2017). Analogy-  
680 preserving functions: A way to extend Boolean samples. In *Proc. of the  
Int. Joint Conf. on Artificial Intelligence, IJCAI'17* (pp. 1575–1581). Int.  
Joint Conferences on Artificial Intelligence Organization.
- [24] Couceiro, M., Hug, N., Prade, H., & Richard, G. (2018). Behavior of  
685 analogical inference w.r.t. Boolean functions. In *Proc. of the Int. Joint  
Conf. on Artificial Intelligence, IJCAI'18*.

- [25] Craw, S., Wiratunga, N., & Rowe, R. C. (2006). Learning adaptation knowledge to improve case-based reasoning. *Artificial Intelligence*, 170, 1175–1192.
- [26] Cross, V. V., & Sudkamp, T. A. (2002). *Similarity and Compatibility in Fuzzy Set Theory* volume 93. Springer.
- [27] d’Aquin, M., Badra, F., Lafrogne, S., Lieber, J., Napoli, A., & Szathmary, L. (2007). Case Base Mining for Adaptation Knowledge Acquisition. In *Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI’07* (pp. 750–755).
- [28] Davies, T. R., & Russell, S. J. (1987). A logical approach to reasoning by analogy. In *Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI’87* (pp. 264–270). [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- [29] Di Jorio, L., Laurent, A., & Teisseire, M. (2009). Mining frequent gradual itemsets from large databases. In *Proc. of the Int. Conf. on Intelligent Data Analysis, IDA’09*.
- [30] Dietrich, C., Schriml, S., Bergmann, R., & Raue, B. (2020). INWEND: Using CBR to automate legal assessment in the context of the EU General Data Protection Regulation. In *LWDA* (p. 10).
- [31] Dubois, D., Esteva, F., Garcia, P., Godo, L., Mántaras, R. L. D., & Prade, H. (1997). Fuzzy set-based models in case-based reasoning. In *Proc. of the 2nd Int. Conf. on Case-Based Reasoning, ICCBR’97* (p. 27).
- [32] Dubois, D., Hüllermeier, E., & Prade, H. (2000). Flexible control of case-based prediction in the framework of possibility theory. In *EWCCBR* (pp. 61–73). Springer.
- [33] Dubois, D., Hüllermeier, E., & Prade, H. (2006). Fuzzy methods for case-based recommendation and decision support. *Journal of Intelligent Information Systems*, 27, 95–115.

- [34] Dubois, D., & Prade, H. (1992). Gradual inference rules in approximate reasoning. *Information Sciences*, 61, 103–122.
- 715 [35] Dubois, D., & Prade, H. (Eds.) (2000). *Fundamentals of fuzzy sets* volume 7 of *The Handbooks of Fuzzy Sets Series*. Springer.
- [36] Dubois, D., & Prade, H. (2002). Fuzzy set-based methods in instance-based reasoning. *IEEE Trans. on Fuzzy Systems*, 10, 322–332.
- [37] Dubois, D., & Prade, H. (2002). Possibility theory, probability theory  
720 and multiple-valued logics: A clarification. *Annals of Mathematics and Artificial Intelligence*, 32, 35–66.
- [38] Eppe, M., Maclean, E., Confalonieri, R., Kutz, O., Schorlemmer, M., Plaza, E., & Kühnberger, K. U. (2018). A computational framework for conceptual blending. *Artificial Intelligence*, 256, 105–129.
- 725 [39] Fahandar, M. A., & Hüllermeier, E. (2019). Analogy-based preference learning with kernels. In *Proc. of the Joint German/Austrian Conf. on Artificial Intelligence KI'19* (pp. 1–15). Springer volume 11793 of *LNAI*.
- [40] Falkenhainer, B. (1987). An Examination of the Third Stage in the Analogy Process: Verification-Based Analogical Learning. In *Proc. of the 10th Int. Joint Conf. on Artificial Intelligence, IJCAI'87* (pp. 260–263). Morgan  
730 Kaufmann.
- [41] Fuchs, B., Lieber, J., Mille, A., & Napoli, A. (2006). Une première formalisation de la phase d'élaboration du raisonnement à partir de cas. In *14ème atelier francophone de raisonnement à partir de cas* (p. 13).
- 735 [42] Fuchs, B., Lieber, J., Mille, A., & Napoli, A. (2014). Differential adaptation: An operational approach to adaptation for solving numerical problems with CBR. *Knowledge-Based Systems*, 68, 103–114.
- [43] Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive science*, 7, 155–170.

- 740 [44] Gentner, D. (2010). Bootstrapping the Mind: Analogical Processes and Symbol Systems. *Cognitive Science*, 34, 752–775.
- [45] Gentner, D., & Forbus, K. D. (2011). Computational models of analogy. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2, 266–276.
- [46] Gentner, D., & Maravilla, F. (2018). Analogical Reasoning. In L. J. Ball, & V. A. Thompson (Eds.), *The Routledge Int. Handbook of Thinking and Reasoning*. Routledge.
- 745 [47] Gentner, D., Rattermann, M., & Forbus, K. (1993). The Roles of Similarity in Transfer: Separating Retrievability From Inferential Soundness. *Cognitive Psychology*, 25, 524–575.
- 750 [48] Goode, M. R., Dahl, D. W., & Moreau, C. P. (2010). The Effect of Experiential Analogies on Consumer Perceptions and Attitudes. *Journal of Marketing Research*, 47, 274–286.
- [49] Gust, H., Krumnack, U., Kühnberger, K., & Schwering, A. (2008). Analogical reasoning: A core of cognition. *KI - Künstliche Intelligenz*, 22, 8–12.
- 755 [50] Gust, H., Kühnberger, K. U., & Schmid, U. (2006). Metaphors and heuristic-driven theory projection (HDTP). *Theoretical Computer Science*, 354, 98–117.
- [51] Han, J., Shi, F., Chen, L., & Childs, P. R. (2018). A computational tool for creative idea generation based on analogical reasoning and ontology.
- 760 *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 32, 462–477.
- [52] Han, J., Shi, F., Chen, L., & Childs, P. R. (2018). A computational tool for creative idea generation based on analogical reasoning and ontology.
- 765 *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 32, 462–477.

- [53] Hanney, K., & Keane, M. T. (1997). The adaptation knowledge bottleneck: How to ease it by learning from cases. In *Case-Based Reasoning Research and Development* (pp. 359–370). Springer volume 1266.
- 770 [54] Hastings, J. D., Latchininsky, A. V., & Schell, S. P. (2010). CARMA : Scalability with approximate-model-based adaptation. In *Proc. of the Int. Congress on Environmental Modelling and Software*.
- [55] Heras, S., Jordán, J., Botti, V., & Julián, V. (2013). Case-based strategies for argumentation dialogues in agent societies. *Information Sciences*, 223, 1–30.
- 775 [56] Hofstadter, D. R., & Sander, E. (2013). *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*. Basic Books.
- [57] Holyoak, K. J. (2012). Analogy and Relational Reasoning. In *Analogy and Relational Reasoning The Oxford Handbook of Thinking and Reasoning* (pp. 234–259). Oxford University Press.
- 780 [58] Holyoak, K. J., Lee, H. S., & Lu, H. (2010). Analogical and category-based inference: A theoretical integration with Bayesian causal models. *Journal of Experimental Psychology: General*, 139, 702–727.
- [59] Hope, T., Chan, J., Kittur, A., & Shahaf, D. (2017). Accelerating Innovation Through Analogy Mining. In *Proc. of the 23rd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* (pp. 235–243). ACM.
- 785 [60] Hug, N., Prade, H., Richard, G., & Serrurier, M. (2016). Analogical classifiers: A theoretical perspective. In *Proc. of the 22nd European Conf. on Artificial Intelligence, ECAI'16* (pp. 689–697). volume 285.
- 790 [61] Hug, N., Prade, H., Richard, G., & Serrurier, M. (2019). Analogical proportion-based methods for recommendation – First investigations. *Fuzzy Sets and Systems*, 366, 110–132.

- [62] Hüllermeier, E. (1999). Toward a Probabilistic Formalization of Case-Based Inference. In *Proc. of the Int. Joint Conf. on Artificial Intelligence, IJ-CAI'99* (pp. 248–253). Morgan Kaufmann Publishers Inc.
- [63] Hüllermeier, E. (2002). Association rules for expressing gradual dependencies. In *Proc. of the 6th European Conf. on Principles of Data Mining and Knowledge Processing, PKDD'02* (pp. 200–211). Springer.
- [64] Hüllermeier, E. (2003). Possibilistic instance-based learning. *Artificial Intelligence*, 148, 335–383.
- [65] Hüllermeier, E. (2007). *Case-Based Approximate Reasoning* volume 44 of *Theory and Decision Library*. Springer.
- [66] Hüllermeier, E. (2007). Credible case-based inference using similarity profiles. *IEEE Transactions on Knowledge and Data Engineering*, 19, 847–858.
- [67] Hüllermeier, E. (2007). Fuzzy Set-Based Modeling of Case-Based Inference II. In *Case-Based Approximate Reasoning* chapter 6. (pp. 229–251).
- [68] Hüllermeier, E. (2020). Towards Analogy-Based Explanations in Machine Learning. In V. Torra, Y. Narukawa, J. Nin, & N. Agell (Eds.), *Modeling Decisions for Artificial Intelligence* (pp. 205–217). Springer volume 12256.
- [69] Hüllermeier, E., & Cheng, W. (2013). Preference-based CBR: General ideas and basic principles. In *Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'13* (pp. 3012–3016).
- [70] Hüllermeier, E., Dubois, D., & Prade, H. (2002). Knowledge-Based Extrapolation of Cases: A Possibilistic Approach. In J. Kacprzyk, B. Bouchon-Meunier, J. Gutiérrez-Ríos, L. Magdalena, & R. R. Yager (Eds.), *Technologies for Constructing Intelligent Systems 1* (pp. 377–390). Physica-Verlag HD volume 89.
- [71] Hüllermeier, E., Dubois, D., & Prade, H. (2002). Model adaptation in possibilistic instance-based reasoning. *IEEE Transactions on Fuzzy Systems*, 10, 333–339.

- [72] Hüllermeier, E., Dubois, D., Prade, H., Toulouse, D., & Sabatier, U. P. (1999). Fuzzy Rules in Case-Based Reasoning. In *In Conf. AFIA99 Raisonement à Partir de Cas* (pp. 45–54).
- [73] Jalali, V., Leake, D., & Forouzandehmehr, N. (2017). Learning and Applying Case Adaptation Rules for Classification: An Ensemble Approach. In *Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'17* (pp. 4874–4878). Int. Joint Conferences on Artificial Intelligence Organization.
- [74] Juthe, A. (2005). Argument by analogy. *Argumentation*, 19, 1–27.
- [75] Keane, M. T. (1993). The Cognitive Processes Underlying Complex Analogies:. *Ricerche Di Psicologia*, 17, 9–36.
- [76] Kokinov, B., & French, R. M. (2003). Computational Models of Analogy-Making. *Encyclopedia of Cognitive Science*, 1, 113–118.
- [77] Kolodner, J. L. (1992). An introduction to case-based reasoning. *Artificial Intelligence Review*, 6, 3–34.
- [78] Laurent, A., Lesot, M.-J., & Rifqi, M. (2009). Graank: Exploiting rank correlations for extracting gradual dependencies. In *Proc. of the Int. Conf. on Flexible Querying Systems, FQAS'09* (pp. 382–393). Springer volume 5822 of *LNAI*.
- [79] Lefort, S., Lesot, M.-J., Zibetti, E., Tijus, C., & Detyniecki, M. (2017). Interpretation of approximate numerical expressions: Computational model and empirical study. *Int. J. of Approximate Reasoning*, 82, 193–209.
- [80] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Proc. of the 34th Int. Conf. on Neural Information Processing Systems, NeurIPS'20*.



- [81] Lieber, J., d'Aquin, M., Badra, F., & Napoli, A. (2008). Modeling adaptation of breast cancer treatment decision protocols in the Kasimir project. *Applied Intelligence*, 28, 261–274.
- 850 [82] Lieber, J., Nauer, E., & Prade, H. (2021). When Revision-Based Case Adaptation Meets Analogical Extrapolation. In A. A. Sánchez-Ruiz, & M. W. Floyd (Eds.), *Case-Based Reasoning Research and Development* (pp. 156–170). Springer volume 12877.
- [83] Lieber, J., Nauer, E., & Prade, H. (2021). When Revision-Based Case  
855 Adaptation Meets Analogical Extrapolation. In A. A. Sánchez-Ruiz, & M. W. Floyd (Eds.), *Case-Based Reasoning Research and Development* (pp. 156–170). Springer volume 12877.
- [84] Lovett, A., & Forbus, K. (2017). Supplemental material for modeling visual problem solving as analogical reasoning. *Psychological Review*, 124, 60–90.
- 860 [85] McConnell, C., & Smyth, B. (2019). Going further with cases: Using case-based reasoning to recommend pacing strategies for ultra-marathon runners. In *Proc. of the Int. Conf. on Case-Based Reasoning, ICCBR'19* (pp. 155–171). volume 3.
- [86] McSherry, D. (1999). Demand-Driven Discovery of Adaptation Knowledge.  
865 In *Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'99* (pp. 222–227).
- [87] Miclet, L., Bayouhd, S., & Delhay, A. (2008). Analogical dissimilarity: Definition, algorithms and two experiments in machine learning. *Journal of Artificial Intelligence Research*, 32, 793–824.
- 870 [88] Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 510–529.

- [89] Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of experimental psychology. Learning, memory, and cognition*, 17, 398–415.
- 875
- [90] Ontañón, S., & Plaza, E. (2012). On knowledge transfer in case-based inference. In B. Díaz-Agudo, & I. Watson (Eds.), *ICCBR Proceedings* (pp. 312–326). Springer.
- [91] Paritosh, P. K., & Klenk, M. E. (2006). Cognitive processes in quantitative estimation: Analogical anchors and causal adjustment. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci-06)*.
- 880
- [92] Pena, G. P., & de Souza Andrade-Filho, J. (2010). Analogies in medicine: Valuable for learning, reasoning, remembering and naming. *Advances in Health Sciences Education*, 15, 609–619.
- [93] Peyre, J., Laptev, I., Schmid, C., & Sivic, J. (2019). Detecting unseen visual relations using analogies. In *IEEE/CVF Int. Conference on Computer Vision* (pp. 1981–1990).
- 885
- [94] Prade, H. (2016). Reasoning with Data - A New Challenge for AI? In S. Schockaert, & P. Senellart (Eds.), *Scalable Uncertainty Management* (pp. 274–288). Springer volume 9858.
- 890
- [95] Prade, H., & Richard, G. (2014). A Short Introduction to Computational Trends in Analogical Reasoning. In H. Prade, & G. Richard (Eds.), *Computational Approaches to Analogical Reasoning: Current Trends* (pp. 1–22). Springer volume 548.
- [96] Riesbeck, C. K., & Schank, R. C. (1989). *Inside Case-Based Reasoning*. Lawrence Erlbaum Associates.
- 895
- [97] Smyth, B., & Keane, M. T. (2022). A Few Good Counterfactuals: Generating Interpretable, Plausible and Diverse Counterfactual Explanations. In M. T. Keane, & N. Wiratunga (Eds.), *Case-Based Reasoning Research and Development* (pp. 18–32). Springer volume 13405.
- 900

[98] Turksen, I., & Zhong, Z. (Nov.-Dec./1988). An approximate analogical reasoning approach based on similarity measures. *IEEE Transactions on Systems, Man, and Cybernetics*, 18, 1049–1056.

[99] Washio, K., & Kato, T. (2018). Neural Latent Relational Analysis to  
905 Capture Lexical Semantic Relations in a Vector Space. In *Proc. of the Int. Conf. on Empirical Methods in Natural Language Processing, EMNLP'18* (pp. 594–600). Association for Computational Linguistics.