International Journal of Approximate Reasoning 158 (2023) 108920

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Case-based prediction – A survey

Fadi Badra [a,*], Marie-Jeanne Lesot [b]

[a] Université Sorbonne Paris Nord, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé, LIMICS, Sorbonne Université, INSERM, F-93000, Bobigny, France
[b] Sorbonne Université, CNRS, LIP6, F-75005 Paris, France

A R T I C L E I N F O

Article history:
Received 28 October 2022
Received in revised form 28 March 2023
Accepted 30 March 2023
Available online 5 April 2023

Keywords:
Case-based prediction
Analogical transfer
Similarity

A B S T R A C T

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 between the two sets of similarity relations.

1. Introduction

Computational analogy is a subfield of computer science that aims at designing computational models of psychological and cognitive processes of analogical thinking. Current research topics in this area go beyond the modeling of the analogical inference as a special type of reasoning (see *e.g.*, [48,91,99] for some surveys), and include the study of the algebraic properties of analogies [4,25], but also of the interactions between machine learning and analogy [86,87,103], the links between case-based and analogical reasoning [9,43,84], and the use of analogical reasoning for explainable AI [70].

Analogical reasoning is recognized to be at the core of human thought [49,58,59]. For instance, in medicine, analogies are commonly used by medical experts for their role in "explaining, naming and mediating knowledge" [96]. Analogical transfer is the part of the analogical reasoning process that allows to leverage a mapping with an analog retrieved from memory in order to derive some new information about the current situation. The new information is derived by implementing a

* Corresponding author.
E-mail address: badra@sorbonne-paris-nord.fr (F. Badra).
URL: https://limics.fr (F. Badra).

https://doi.org/10.1016/j.ijar.2023.108920

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 [41], categorization [20], as well as to cope with insufficient training data [97], to guide language generation [82], to foster creativity [54] or to accelerate innovation [61]. The strengths of analogical reasoning 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) [100]. 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* [67]) algorithms 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 [66], CCBI [68], 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. Among all potential outcomes, the predicted outcome is the one that optimizes this transfer, *i.e.*, 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. 3 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. 4 presents a systematic analysis of existing case-based prediction theories, proposing a typology distinguishing between four main families of approaches. Sec. 5 to 8 describe in turn the most representative approaches for each the four families. Sec. 9 concludes the paper and gives directions for future work.

2. The transfer task of analogical reasoning systems

Analogical reasoning systems are often decomposed into different tasks in the literature [1,49,52,77,78]. From a cognitive science point of view, it is questionable to decompose the analogical inference in a succession of tasks, because analogical reasoning is rooted in perception, and involves complex cognitive processes that are often interwined [19]. However, such a decomposition greatly simplifies the conception of computer systems and allows for comparison.

The transfer task is the component of analogical reasoning systems that allows to make predictions. It implements analogical transfer,[1] which is the part of the analogical reasoning process that allows to leverage a mapping with an analog retrieved from memory in order to derive some new information about the current situation [47,50,92]. 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 [79].

- It may be used for *prediction*, in order to complete the description of the new case. Examples include estimating a quantity [95], a preference [41], a ratio [62], a semantic relation [104], recommending a decision [35] or predicting the effect of a decision or a plan (for example to find a response plan for natural disasters [56] or to help an athlete choose a pace when running a marathon [88]).
- 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 [70,101], causal attributions [60], establishing a legal assessment [32], putting forth persuasive arguments in the context of adversarial reasoning [57,76], or even tapping into the emotions that people gained from their own experience in order to sway a decision [51].
- 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], solving mathemat-

[1] In cognitive psychology, the term analogical transfer is often used in a general sense to denote the whole analogical reasoning process. We use it here in a more restrictive manner, to denote the plausible inference that is triggered *after* the mapping has taken place, in order to derive some new information about the target situation.

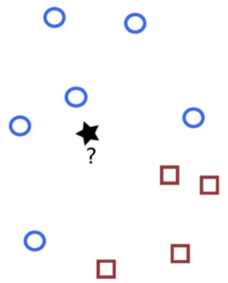

Fig. 1. A simple classification setting. The goal is to predict the class (blue circle or red square) for a new situation represented by $\star$ on the figure, by finding the class for which the added outcome similarities would be most compatible with the observed situation similarities.

ical problems by adapting old solution procedures [93], proposing innovative ideas for designers [54], making content suggestions for authors of product reviews [18] or inventing new concepts by conceptual blending [40].

Apart from the transfer task, many other tasks are cited in the literature: 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) [46]. The *representation* task [77] produces an initial representation for the inputs. The *elaboration* task [44] enriches such representation with domain knowledge. The *re-representation* task [85] restructures the input representation to facilitate comparison. The *abstraction* task [53] produces a common abstraction of two inputs. The *validation* task [42] evaluates the quality and consistency of the results. The *memorization* task [59] updates the memory for later use.

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

3. 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 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 compatible with the observed similarities on situations, and differ mainly by the type of knowledge used to measure the compatibility between the two sets of similarity relations.

3.1. Definitions and notation

A typical setting for case-based prediction is as follows. Let $\mathcal{S}$ be an input space and $\mathcal{R}$ an output space. An element of $\mathcal{S}$ is called a *situation*, an element of $\mathcal{R}$ is called an *outcome*, or a result, a couple of $\mathcal{S} \times \mathcal{R}$ is called a *case*. A *case base* is a finite set $CB = \{(s_1, r_1), \ldots, (s_n, r_n)\}$ of such pairs in $\mathcal{S} \times \mathcal{R}$. In addition, σ_S and σ_R respectively denote similarity measures on situations and on outcomes. The goal of the inference is to predict the outcome r_t of a new situation t. For readability, and abusing the notation, cases and outcomes are sometimes denoted by their corresponding situation as subscript: a case is denoted by $c_s = (s, r_s) \in CB$, and the new case is denoted by $c_t = (t, r_t)$.

As an example, consider the classification setting graphically represented on Fig. 1. Situations are points of a 2D space, *i.e.* $\mathcal{S} = \mathbb{R}^2$, and the similarity $\sigma_S(s, t)$ between two situations s and t is estimated from the Euclidean distance, *e.g.*, by setting $\sigma_S = e^{-\|\cdot\|_2}$, where $\|\cdot\|_2$ is the Euclidean distance. Outcomes are classes, $\mathcal{R} = \{\texttt{blue}, \texttt{red}\}$, and the similarity $\sigma_R(r_s, r_t)$ between two outcomes r_s and r_t is given by the discrete function that returns 1 if the two classes are the same, and 0 otherwise. A case base CB then contains a set of points s, whose class r_s is known, and the prediction task aims at determining the class r_t of a new point t (represented as $\star$ on Fig. 1).

3.2. A transfer of similarity knowledge

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 [52]:

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

The outcome r_t is finally (indirectly) determined from these estimations (Fig. 2).

$$\begin{tikzcd} s \arrow[r, dash] & \sigma_S \arrow[r, dash] \arrow[d, "\text{used to estimate}"] & t \\ r_s \arrow[r, dash] & \sigma_R \arrow[r, dash] & r_t \end{tikzcd}$$

Fig. 2. In the transfer task, a similarity relation σ_S is used to estimate another one, σ_R.

Case-based prediction methods search for an outcome r_t that would make the similarity relations in the outcome space most *compatible* with the similarity relations in the situation space. Informally, compatible means that a similarity relation between two situations can be used to estimate the corresponding similarity relation between outcomes; the formal definition is at the core of the variety of case-based predictions methods, as discussed in details in the next sections. 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 $\Sigma_{\mathcal{S}}(t) = \{\sigma_S(s,t) \mid c_s = (s, r_s) \in CB\}$
- the outcome similarities $\Sigma_{\mathcal{R}}(r) = \{\sigma_R(r_s, r) \mid c_s = (s, r_s) \in CB\}$

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 relations. The goal is to find the outcome $r \in \mathcal{R}$ for which the similarity relations $\Sigma_{\mathcal{R}}(r)$ are the most compatible with the similarity relations $\Sigma_{\mathcal{S}}(t)$. In the classification setting of Fig. 1, that would mean finding the class 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 rest of the paper will be devoted to showing how this notion of compatibility is defined and measured in the different approaches.

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 σ_S and σ_R can be 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 σ_S into similar points are also transformed by σ_R into similar points. Recent work [9] further suggests that the success of the inference only requires that the similarity knowledge 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 discussed in details in the following sections, case-based prediction methods differ by the type of knowledge used to express the compatibility requirement between the two similarity measures. The next section first proposes an overview, in the form of a typology, distinguishing the methods along this axis.

4. Proposed typology of case-based prediction methods

This section presents a systematic analysis of existing case-based prediction theories along the lines developed in Sec. 3.

4.1. Methodology

As the goal of the survey is to outline the commonalities and differences between the computational theories of case-based prediction, all works describing representation and knowledge engineering issues (such as, *e.g.*, learning adaptation rules [27,29]) 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 [35]), to derive a particular type of information (*e.g.*, a preference ordering [41,71]), or to take into account domain knowledge [84].

However, we include a few case-based adaptation methods (see Sec. 7.2). Adaptation is the cognitive ability to envision a target solution that is different from any previously encountered solution. Although adaptation 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 adaptation rules. Reporting such methods is also a way to show the intimate links between case-based adaptation and case-based prediction.

4.2. Discrimination criteria

Many criteria can be used to categorize the different approaches. A few of them are reported here and applied in Sec. 4.3, leading to the four categories discussed in turn in Sec. 5 to 8. 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 σ_S and σ_R: a set of adaptation rules, a continuity constraint, a compatibility indicator between cases or a global compatibility function, as discussed in the next sections.

Table 1
Proposed typology of case-based prediction theories.

Approach	Compatibility knowledge	Prediction strategy
Evidence support	A joint similarity measure, that measures how compatible σ_R is with σ_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.e.*, rules that state that σ_R should be compatible with σ_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 $(\sigma_S = \alpha) \rightarrow (\sigma_R = \beta)$.	Implement a majority vote on the outcomes derived from the rules.
Global optimization	A global function, that measures how compatible σ_R is with σ_S on the whole case base.	Optimize the global compatibility measure on the augmented case base.

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 a 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, when adding the new situation with its predicted outcome to the initial case base, makes σ_R most compatible with σ_S.

Knowledge-driven vs data-driven strategy. The different approaches can be interpreted in a bipolar framework [98]. *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 prediction strategy consists in predicting an outcome that does not violate 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* 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 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 σ_S and σ_R are defined 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 case pairs. 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 $\Sigma_{\mathcal{S}}(\hat{c}_t)$ and $\Sigma_{\mathcal{R}}(\hat{c}_t)$ pairwise), and the result 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 σ_R with σ_S estimated on the case base as a whole, and takes into account in the compatibility estimation some pairs of cases in which c_t does *not* appear.

Ordinal vs numerical strategies. The approaches also differ by the way similarity relations are compared in order to produce a compatibility estimation. Some of them, such as *transfer by approximate reasoning* approaches, compare the values of the two similarity measures pairwise, while others only consider the similarity orderings to make predictions. Examples of the latter include some *transfer by constraint* approaches such as the credible case-based inference [68] or the *transfer by global optimization* method proposed in [6].

4.3. Proposed typology of case-based prediction theories

Four categories of methods are identified and summarized, respectively named transfer by evidence support, transfer by constraint, transfer by approximate reasoning, and transfer by global optimization. They differ by the type of compatibility knowledge that is used. The typology is shown in Table 1 and discussed in turn in Sec. 5 to 8 that provide more formal descriptions of each category.

Transfer by evidence support. This type of data-driven approach consists in using a joint similarity measure to estimate for each pair of cases $(c_s, \hat{c}_t)$ how compatible the similarity relation $\sigma_R(r_s, r)$ is with the similarity relation $\sigma_S(s,t)$. Examples include the k-Nearest Neighbor algorithm and the Possibilistic Instance-Based Learning approach [13,34,66]. In these approaches, a new case is considered possible if the existence of a similar case is confirmed by observation. The value of the joint similarity measure is interpreted as a degree of *confirmation*, 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 follows the knowledge-driven approach, consists in expressing the compatibility requirement between the two similarity measures σ_S and σ_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 σ_S and σ_R. Examples include similarity profiles [68] or fuzzy implication rules [69,72,74]. 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 where the two similarity measures σ_S and σ_R align locally, and reason by similarity on these 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 [7,29,75,94], co-variations [5], or dependencies between analogical proportions [9,17]. Some case-based adaptation approaches implement this strategy, as well as analogical proportion-based classifiers.

Transfer by global optimization. In most case-based prediction approaches, the compatibility of σ_R with σ_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 global indicator that measures the compatibility of σ_R with σ_S on the whole case base. The prediction strategy consists in minimizing the value of this indicator on the augmented case base.

5. Transfer by evidence support

In this type of approach, the compatibility of σ_R with σ_S is considered as 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 σ between cases, defined as a joint similarity measure [2,13,34,66]. The *prediction 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.

5.1. General principles

The compatibility indicator σ measures the compatibility of σ_R with σ_S on a pair of cases $(c_s, \hat{c}_t)$, and is used as an indicator of the plausibility of a new 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 σ for the pairs of cases $(c_s, \hat{c}_t)$;
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 Possibilistic Instance-Based Learning method.

5.2. k-Nearest Neighbors

In a classification setting, the k-Nearest Neighbor approach [2] makes a majority 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 σ_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 σ 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 σ 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

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

5.3. Possibilistic Instance-Based Learning

A possibilistic 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" [13,34,38,66,72,73]. Such a rule is formalized in the formal framework of possibility theory (see *e.g.* [39]), 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.*,

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

6. Transfer by constraint

A complementary, more knowledge-driven approach, consists in taking the compatibility of σ_R with σ_S as a *negative* constraint, that is used to exclude the outcomes that are not similar enough to the outcomes of a retrieved case. The *compatibility knowledge* used in the inference is a set of continuity constraints such as similarity profiles [64,67,68], or fuzzy implication rules [33,69,73,74]. Such a continuity constraint is interpreted as a negative information according to which it is not plausible to observe situations very dissimilar for σ_R when they are similar for σ_S [13]. The *prediction strategy* consists in using these constraints to exclude the potential outcomes that violate them. 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 outcome r_t is then chosen among the outcomes r that are most compatible with the constraints.

6.1. Credible case-based inference

Continuity constraints can be expressed by stating that if two situations are above a similarity threshold α for σ_S, then it is likely that their similarity for σ_R is greater or equal than a value β. The function $h : [0,1] \longrightarrow [0,1]$ which associates to each similarity level α for σ_S a similarity level β for σ_R is called a *similarity profile* [64,67,68]. 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 [69], 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 β of a similarity profile from the data is a task that is very sensitive to outliers. Two solutions are investigated in [67] 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 [64]. 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].

6.2. Gradual rules

Gradual rules [36] are linguistically expressed as “the more X is A, the more Y is B” and modeled in the formal framework of fuzzy set theory (see *e.g.* [37]): 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 [36] is expressed in terms of membership degrees by $B(Y) \geq A(X)$, which is equivalent to a set of constraints of the form $(X \in A_\alpha) \longrightarrow (Y \in B_\alpha)$ stating that if the membership degree of X in A is at least α, then it is guaranteed that the membership degree of Y in B is also at least α. 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 $h : [0, 1] \longrightarrow [0, 1]$ 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.

6.3. 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 $\pi_{Y|X}$:

$$\forall (x, y) \in D_X \times D_Y, \quad \pi_{Y|X}(y \mid x) \leq \max(1 - A(x), B(y)),$$

where D_X and D_Y are the domains of X and Y, respectively.

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

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

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

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 very similar outcomes.

The possibility degree $\pi(r \mid t)$ 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:

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

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

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

7. Transfer by approximate reasoning

The third type of approach consists in searching where the two similarity measures σ_S and σ_R align locally, and reason by similarity on the found alignments. The *compatibility knowledge* takes the form of a set of rules providing information on relations between the similarity measures σ_S and σ_R, expressing that when σ_S takes value α, the resulting similarity level for σ_R is β: these rules can be written $(\sigma_S = \alpha) \to (\sigma_R = \beta)$ and can be expressed in various forms, such as adaptation rules [7,29,75,94], dependencies between problem and solution features [45], co-variations [5] or fuzzy rules [16] to name a few. The *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.

7.1. General principles

A rule $(\sigma_S = \alpha) \to (\sigma_R = \beta)$ is a piece of knowledge that states that σ_R is compatible with σ_S when σ_S takes the value α, and that the resulting similarity level for σ_R is β. 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 [11,16,28,102]: for a retrieved case $c_s = (s, r_s)$ and a potential new case $\hat{c}_t = (t, r)$, triggering the rule $(\sigma_S = \alpha) \to (\sigma_R = \beta)$ on the pair of cases $(c_s, \hat{c}_t)$ is of the form

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

This schema expresses that if the rule associates a level β for the similarity σ_R whenever the similarity level for σ_S is α, and if the observed situation similarity between the new case and a retrieved case is approximately α (*i.e.*, $\sigma_S(s,t) \approx \alpha$), then the corresponding similarity on outcomes $\sigma_R(r_s, r)$ is approximately β. 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 [81].

It is often the case that the similarity measures σ_S and σ_R are unknown, or difficult to assess globally on the training data. One strategy then consists in working with some local approximations $\widetilde{\sigma_S}$ and $\widetilde{\sigma_R}$ of σ_S and σ_R respectively, that are known to be compatible for some pairs of cases of the case base. The resulting rules $(\widetilde{\sigma_S} = \alpha) \to (\widetilde{\sigma_R} = \beta)$ are *adaptation rules* (see Sec. 7.2).

Some case-based prediction approaches such as [75] include a rule selection step prior to the inference, while others such as proportion-based analogical classifiers (see Sec. 7.4) trigger only one rule.

Each selected rule is triggered on a set of pairs of cases $(c_s, \hat{c}_t)$, 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:

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

In the rest of the section, three kinds of approaches of this category are described in more detail: rule-based adaptation, the analogical jump, and analogical proportion-based classification.

7.2. Rule-based adaptation

Adaptation rules are rules of the form $(\widetilde{\sigma_S} = \alpha) \to (\widetilde{\sigma_R} = \beta)$, where $\widetilde{\sigma_S}$ and $\widetilde{\sigma_R}$ are local approximations of σ_S and σ_R. The main difficulty when working with adaptation rules is that one needs to be able to learn the rules. They may be acquired from different sources such as a domain expert [83,95], the user [8], or learned from data [7,27,29,55,75,89]. In [95] 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 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 σ_S on the values of the attribute `nb_rooms`, a similarity σ_R on the values of the attribute `price`, and applying the (SBI) inference schema with the hypothesis that σ_R increases linearly with σ_S, *i.e.*, with the hypothesis that $\beta = Q\,\alpha$. Another example is the work of [89], 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 $(\widetilde{\sigma_S} = \alpha) \to (\widetilde{\sigma_R} = \beta)$ 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 [11,17,30,95] have emphasized the idea that the rules $(\widetilde{\sigma_S} = \alpha) \to (\widetilde{\sigma_R} = \beta)$ 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 $\mathtt{nb\_rooms}^{\leq} : \mathcal{S} \times \mathcal{S} \mapsto \{0, 1\}$ maps a pair of apartments to 1 if the number of rooms increases, and to 0 otherwise, and the co-variation $(\mathtt{nb\_rooms}^{\leq} = 1) \to (\mathtt{price}^{\leq} = 1)$ 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 $\texttt{nb\_rooms}(s) \leq \texttt{nb\_rooms}(t)$, then the antecedent of the rule is satisfied, so one can make the hypothesis that the rule applies, and that $\texttt{price}(r_s) \leq \texttt{price}(r)$. In this example, the inference schema (SBI) writes:

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

Learning such co-variations from data corresponds to the task called gradual pattern mining, for which various meanings and approaches have been proposed [5,14,31,65,80].

7.3. 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 [30] introduce an inference schema, called the "analogical jump" (AJ):

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

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 similarity-based inference schema (SBI):

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

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 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 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.

7.4. 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 APC, in what follows) and case-based prediction, showing the former can be viewed as a special kind of the latter.

APC algorithms have been applied to classification and recommendation tasks, see *e.g.* [12,17,23–26,63,90]. They apply the principle of analogical reasoning [58], based on statements of the form "**a** is to **b** as **c** 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, labeling function, one can derive from $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$ that $f(\mathbf{a}) : f(\mathbf{b}) :: f(\mathbf{c}) : f(\mathbf{d})$. Let D be a data set containing a set of instances **a**,**b**,**c**,… with their associated labels $f(\mathbf{a}), f(\mathbf{b}), f(\mathbf{c}), \ldots$ To predict the value $f(x)$ for a new instance x, an analogical proportion-based classifier considers all

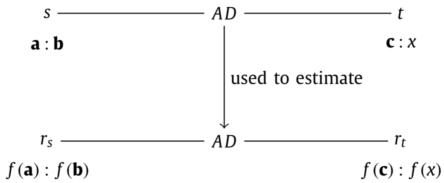

Fig. 3. In APC methods, both situations and outcomes represent ratios.

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 *analogical root* of x [62]. 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 [62]. The latter is defined as a function $AD(\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d})$ that quantifies 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 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 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. 2. 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 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 $y\in C$. The source case c_s and potential new case $\hat{c_t}$ are of the following form:

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

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 σ_S and σ_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 σ_S and σ_R are obtained by applying a strictly decreasing function to the distance δ, *e.g.*, by choosing $\sigma_S=\sigma_R=e^{-\delta}$. The similarity measure σ_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 σ_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 [84], 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 $t=\mathbf{c}-x$;
  - *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 $f(x)$ 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 σ_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.

8. Transfer by global optimization

A recent work [6,9,10] proposes to define a global indicator that measures the compatibility of σ_R with σ_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 σ_R with σ_S on any potential case base CB. The *prediction strategy* consists in completing the description of a case base in order to minimize the value of the global indicator.

This principle is implemented in the CoAT, for **Co**mplexity-based **A**nalogical **T**ransfer, algorithm [6,10]. In the CoAT method, the compatibility of σ_R with σ_S is measured from an ordinal point of view on the whole case base CB, by checking if σ_R orders the cases in the same manner as σ_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 $c_j=(s_j,r_j)$:

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

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:

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

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 Γ indicator:

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

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 σ_S and σ_R. The compatibility estimator Γ considers all triples (c_0,c_i,c_j), and checks for each triple if σ_R orders the two pairs of cases (c_0,c_i) and (c_0,c_j) in the same way that σ_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 σ_R with σ_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 $\Sigma_{\mathcal{S}}(t)$ and $\Sigma_{\mathcal{R}}(r)$ 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. 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. 6). The global indicator Γ 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.

9. Conclusion and further work

This study constitutes the first survey of the wide and rich domain of case-based prediction. At the intersection 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, some continuity constraints, a set of rules, or a global indicator. The prediction strategy varies accordingly: it consists either in maximizing an indicator, finding the outcome that is the most consistent with a set of continuity constraints, or making a majority vote on the outcomes derived from a set of rules.

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 to 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 scenarios, 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.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] A. Aamodt, E. Plaza, Case-based reasoning: foundational issues, methodological variations, and system approaches, AI Commun. 7 (1994) 39–59.
[2] D.W. Aha, Tolerating noisy, irrelevant and novel attributes in instance-based learning algorithms, Int. J. Man-Mach. Stud. 36 (1992) 267–287.
[3] M. Anthony, J. Ratsaby, A probabilistic approach to case-based inference, Theor. Comput. Sci. 589 (2015) 61–75.
[4] C. Antić, Analogical proportions, Ann. Math. Artif. Intell. 90 (2022) 595–644.
[5] F. Badra, Reasoning with co-variations, in: C. Dichev, G. Agre (Eds.), Artificial Intelligence: Methodology, Systems, and Applications, vol. 9883, Springer International Publishing, Cham, 2016, pp. 205–215.
[6] F. Badra, A dataset complexity measure for analogical transfer, in: Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'20, 2020, pp. 1601–1607.
[7] F. Badra, R. Bendaoud, R. Bentebibel, P.-A. Champin, J. Cojan, A. Cordier, S. Desprès, S. Jean-Daubias, J. Lieber, T. Meilender, A. Mille, E. Nauer, A. Napoli, Y. Toussaint, 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, 2008, pp. 219–228.
[8] F. Badra, A. Cordier, J. Lieber, Opportunistic adaptation knowledge discovery, in: Case-Based Reasoning Research and Development, Springer, 2009, pp. 60–74.
[9] F. Badra, M.-J. Lesot, 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), 2022.
[10] F. Badra, M.-J. Lesot, Theoretical and experimental study of a complexity measure for analogical transfer, in: Int. Conference on Case-based Reasoning (ICCBR), 2022, pp. 175–189.
[11] F. Badra, K. Sedki, A. Ugon, On the role of similarity in analogical transfer, in: Case-Based Reasoning Research and Development, vol. 11156, Springer, 2018, pp. 499–514.
[12] S. Bayoudh, Learning by analogy: a classification rule for binary and nominal data, in: Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'07, 2007, pp. 678–683.
[13] J. Beringer, E. Hüllermeier, Case-based learning in a bipolar possibilistic framework, Int. J. Intell. Syst. 23 (2008) 1119–1134.
[14] F. Berzal, J.-C. Cubero, D. Sanchez, M.-A. Vila, J.M. Serrano, An alternative approach to discover gradual dependencies, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 15 (2007) 559–570.
[15] R. Billingsley, H. Prade, G. Richard, M.-A. Williams, Towards analogy-based decision - a proposal, in: Proc. of the Int. Conf. on Flexible Query Answering Systems, FQAS'17, vol. 10333, Springer, 2017, pp. 28–35.
[16] B. Bouchon-Meunier, L. Valverde, A fuzzy approach to analogical reasoning, Soft Comput. 3 (1999) 141–147.
[17] M. Bounhas, H. Prade, G. Richard, Analogy-based classifiers for nominal or numerical data, Int. J. Approx. Reason. 91 (2017) 36–55.
[18] D. Bridge, P. Healy, The GhostWriter-2.0 case-based reasoning system for making content suggestions to the authors of product reviews, Knowl.-Based Syst. 29 (2012) 93–103.
[19] D.J. Chalmers, R.M. French, D.R. Hofstadter, High-level perception, representation, and analogy: a critique of artificial intelligence methodology, J. Exp. Theor. Artif. Intell. 4 (1992) 185–211.
[20] C. Chen, O. Li, C. Tao, A.J. Barnett, J. Su, C. Rudin, This looks like that: deep learning for interpretable image recognition, in: Proc. of the 33rd Int. Conf. on Neural Information Processing Systems, Curran Associates Inc., 2019, pp. 8930–8941.
[21] J. Cojan, J. Lieber, Belief merging-based case combination, in: L. McGinty, D.C. Wilson (Eds.), Case-Based Reasoning Research and Development, vol. 5650, Springer, 2009, pp. 105–119.

[22] A. Cordier, V. Dufour-Lussier, J. Lieber, E. Nauer, F. Badra, J. Cojan, E. Gaillard, L. Infante-Blanco, P. Molli, A. Napoli, H. Skaf-Molli, Taaable: a case-based system for personalized cooking, in: Studies in Computational Intelligence, vol. 494, 2014, pp. 121–162.
[23] M. Couceiro, N. Hug, H. Prade, G. Richard, Analogy-preserving functions: a way to extend Boolean samples, in: Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'17, Int. Joint Conferences on Artificial Intelligence Organization, 2017, pp. 1575–1581.
[24] M. Couceiro, N. Hug, H. Prade, G. Richard, Behavior of analogical inference w.r.t. Boolean functions, in: Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'18, 2018, pp. 2057–2063.
[25] M. Couceiro, E. Lehtonen, A Galois framework for the study of analogical classifiers, in: IARML@IJCAI-ECAI'2022: Workshop on the Interactions Between Analogical Reasoning and Machine Learning, at IJCAI-ECAI'2022, 2022.
[26] M. Couceiro, E. Lehtonen, L. Miclet, H. Prade, G. Richard, When nominal analogical proportions do not fail, in: J. Davis, K. Tabia (Eds.), Scalable Uncertainty Management, vol. 12322, Springer International Publishing, Cham, 2020, pp. 68–83.
[27] S. Craw, N. Wiratunga, R.C. Rowe, Learning adaptation knowledge to improve case-based reasoning, Artif. Intell. 170 (2006) 1175–1192.
[28] V.V. Cross, T.A. Sudkamp, Similarity and Compatibility in Fuzzy Set Theory, vol. 93, Springer, 2002.
[29] M. d'Aquin, F. Badra, S. Lafrogne, J. Lieber, A. Napoli, L. Szathmary, Case base mining for adaptation knowledge acquisition, in: Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'07, 2007, pp. 750–755.
[30] T.R. Davies, S.J. Russell, A logical approach to reasoning by analogy, in: Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'87, 1987, pp. 264–270.
[31] L. Di Jorio, A. Laurent, M. Teisseire, Mining frequent gradual itemsets from large databases, in: Proc. of the Int. Conf. on Intelligent Data Analysis, IDA'09, 2009, pp. 297–308.
[32] C. Dietrich, S. Schriml, R. Bergmann, B. Raue, INWEND: using CBR to automate legal assessment in the context of the EU General Data Protection Regulation, in: LWDA, 2020, pp. 192–201.
[33] D. Dubois, F. Esteva, P. Garcia, L. Godo, R.L.D. Màntaras, H. Prade, Fuzzy set-based models in case-based reasoning, in: Proc. of the 2nd Int. Conf. on Case-Based Reasoning, ICCBR'97, 1997.
[34] D. Dubois, E. Hüllermeier, H. Prade, Flexible control of case-based prediction in the framework of possibility theory, in: EWCBR, Springer, 2000, pp. 61–73.
[35] D. Dubois, E. Hüllermeier, H. Prade, Fuzzy methods for case-based recommendation and decision support, J. Intell. Inf. Syst. 27 (2006) 95–115.
[36] D. Dubois, H. Prade, Gradual inference rules in approximate reasoning, Inf. Sci. 61 (1992) 103–122.
[37] D. Dubois, H. Prade (Eds.), Fundamentals of Fuzzy Sets, The Handbooks of Fuzzy Sets Series, vol. 7, Springer, 2000.
[38] D. Dubois, H. Prade, Fuzzy set-based methods in instance-based reasoning, IEEE Trans. Fuzzy Syst. 10 (2002) 322–332.
[39] D. Dubois, H. Prade, Possibility theory, probability theory and multiple-valued logics: a clarification, Ann. Math. Artif. Intell. 32 (2002) 35–66.
[40] M. Eppe, E. Maclean, R. Confalonieri, O. Kutz, M. Schorlemmer, E. Plaza, K.U. Kühnberger, A computational framework for conceptual blending, Artif. Intell. 256 (2018) 105–129.
[41] M.A. Fahandar, E. Hüllermeier, Analogy-based preference learning with kernels, in: Proc. of the Joint German/Austrian Conf. on Artificial Intelligence KI'19, in: LNAI, vol. 11793, Springer, 2019, pp. 1–15.
[42] B. Falkenhainer, 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, Morgan Kaufmann, 1987, pp. 260–263.
[43] B. Fuchs, J. Lieber, L. Miclet, A. Mille, A. Napoli, H. Prade, G. Richard, Case-based reasoning, analogy, and interpolation, in: P. Marquis, O. Papini, H. Prade (Eds.), A Guided Tour of Artificial Intelligence Research, Springer International Publishing, Cham, 2020, pp. 307–339.
[44] B. Fuchs, J. Lieber, A. Mille, A. Napoli, Une première formalisation de la phase d'élaboration du raisonnement à partir de cas, in: 14ème atelier francophone de raisonnement à partir de cas, 2006, p. 13.
[45] B. Fuchs, J. Lieber, A. Mille, A. Napoli, Differential adaptation: an operational approach to adaptation for solving numerical problems with CBR, Knowl.-Based Syst. 68 (2014) 103–114.
[46] D. Gentner, Structure-mapping: a theoretical framework for analogy, Cogn. Sci. 7 (1983) 155–170.
[47] D. Gentner, Bootstrapping the mind: analogical processes and symbol systems, Cogn. Sci. 34 (2010) 752–775.
[48] D. Gentner, K.D. Forbus, Computational models of analogy, Wiley Interdiscip. Rev. Cogn. Sci. 2 (2011) 266–276.
[49] D. Gentner, F. Maravilla, Analogical reasoning, in: L.J. Ball, V.A. Thompson (Eds.), The Routledge Int. Handbook of Thinking and Reasoning, Routledge, 2018.
[50] D. Gentner, M. Rattermann, K. Forbus, The roles of similarity in transfer: separating retrievability from inferential soundness, Cogn. Psychol. 25 (1993) 524–575.
[51] M.R. Goode, D.W. Dahl, C.P. Moreau, The effect of experiential analogies on consumer perceptions and attitudes, J. Mark. Res. 47 (2010) 274–286.
[52] H. Gust, U. Krumnack, K. Kühnberger, A. Schwering, Analogical reasoning: a core of cognition, Künstl. Intell. 22 (2008) 8–12.
[53] H. Gust, K.U. Kühnberger, U. Schmid, Metaphors and heuristic-driven theory projection (HDTP), Theor. Comput. Sci. 354 (2006) 98–117.
[54] J. Han, F. Shi, L. Chen, P.R. Childs, A computational tool for creative idea generation based on analogical reasoning and ontology, Artif. Intell. Eng. Des. Anal. Manuf. 32 (2018) 462–477.
[55] K. Hanney, M.T. Keane, The adaptation knowledge bottleneck: how to ease it by learning from cases, in: Case-Based Reasoning Research and Development, vol. 1266, Springer, 1997, pp. 359–370.
[56] J.D. Hastings, A.V. Latchininsky, S.P. Schell, CARMA: scalability with approximate-model-based adaptation, in: Proc. of the Int. Congress on Environmental Modelling and Software, 2010.
[57] S. Heras, J. Jordán, V. Botti, V. Julián, Case-based strategies for argumentation dialogues in agent societies, Inf. Sci. 223 (2013) 1–30.
[58] D.R. Hofstadter, E. Sander, Surfaces and Essences: Analogy as the Fuel and Fire of Thinking, Basic Books, 2013.
[59] K.J. Holyoak, Analogy and relational reasoning, in: Analogy and Relational Reasoning, The Oxford Handbook of Thinking and Reasoning, Oxford University Press, 2012, pp. 234–259.
[60] K.J. Holyoak, H.S. Lee, H. Lu, Analogical and category-based inference: a theoretical integration with Bayesian causal models, J. Exp. Psychol. Gen. 139 (2010) 702–727.
[61] T. Hope, J. Chan, A. Kittur, D. Shahaf, Accelerating innovation through analogy mining, in: Proc. of the 23rd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, ACM, 2017, pp. 235–243.
[62] N. Hug, H. Prade, G. Richard, M. Serrurier, Analogical classifiers: a theoretical perspective, in: Proc. of the 22nd European Conf. on Artificial Intelligence, ECAI'16, vol. 285, 2016, pp. 689–697.
[63] N. Hug, H. Prade, G. Richard, M. Serrurier, Analogical proportion-based methods for recommendation – first investigations, Fuzzy Sets Syst. 366 (2019) 110–132.
[64] E. Hüllermeier, Toward a probabilistic formalization of case-based inference, in: Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'99, Morgan Kaufmann Publishers Inc., 1999, pp. 248–253.
[65] E. Hüllermeier, Association rules for expressing gradual dependencies, in: Proc. of the 6th European Conf. on Principles of Data Mining and Knowledge Processing, PKDD'02, Springer, 2002, pp. 200–211.
[66] E. Hüllermeier, Possibilistic instance-based learning, Artif. Intell. 148 (2003) 335–383.
[67] E. Hüllermeier, Case-Based Approximate Reasoning, Theory and Decision Library, vol. 44, Springer, 2007.

[68] E. Hüllermeier, Credible case-based inference using similarity profiles, IEEE Trans. Knowl. Data Eng. 19 (2007) 847–858.
[69] E. Hüllermeier, Fuzzy set-based modeling of case-based inference II, in: Case-Based Approximate Reasoning, 2007, pp. 229–251 (chapter 6).
[70] E. Hüllermeier, Towards analogy-based explanations in machine learning, in: V. Torra, Y. Narukawa, J. Nin, N. Agell (Eds.), Modeling Decisions for Artificial Intelligence, vol. 12256, Springer, 2020, pp. 205–217.
[71] E. Hüllermeier, W. Cheng, Preference-based CBR: general ideas and basic principles, in: Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'13, 2013, pp. 3012–3016.
[72] E. Hüllermeier, D. Dubois, H. Prade, 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, vol. 89, Physica-Verlag HD, 2002, pp. 377–390.
[73] E. Hüllermeier, D. Dubois, H. Prade, Model adaptation in possibilistic instance-based reasoning, IEEE Trans. Fuzzy Syst. 10 (2002) 333–339.
[74] E. Hüllermeier, D. Dubois, H. Prade, Fuzzy rules in case-based reasoning, in: Conférence Raisonnement à partir de Cas (RàpC 1999) @ Plateforme AFIA 1999, AFIA : Association française pour l'intelligence artificielle, Palaiseau, France, 1999, pp. 45–54.
[75] V. Jalali, D. Leake, N. Forouzandehmehr, Learning and applying case adaptation rules for classification: an ensemble approach, in: Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'17, Int. Joint Conferences on Artificial Intelligence Organization, 2017, pp. 4874–4878.
[76] A. Juthe, Argument by analogy, Argumentation 19 (2005) 1–27.
[77] M.T. Keane, The cognitive processes underlying complex analogies, Ric. Psicol. 17 (1993) 9–36.
[78] B. Kokinov, R.M. French, Computational models of analogy-making, in: Encyclopedia of Cognitive Science, vol. 1, 2003, pp. 113–118.
[79] J.L. Kolodner, An introduction to case-based reasoning, Artif. Intell. Rev. 6 (1992) 3–34.
[80] A. Laurent, M.-J. Lesot, M. Rifqi, Graank: exploiting rank correlations for extracting gradual dependencies, in: Proc. of the Int. Conf. on Flexible Querying Systems, FQAS'09, in: LNAI, vol. 5822, Springer, 2009, pp. 382–393.
[81] S. Lefort, M.-J. Lesot, E. Zibetti, C. Tijus, M. Detyniecki, Interpretation of approximate numerical expressions: computational model and empirical study, Int. J. Approx. Reason. 82 (2017) 193–209.
[82] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, D. Kiela, Retrieval-augmented generation for knowledge-intensive NLP tasks, in: Proc. of the 34th Int. Conf. on Neural Information Processing Systems, NeurIPS'20, 2020.
[83] J. Lieber, M. d'Aquin, F. Badra, A. Napoli, Modeling adaptation of breast cancer treatment decision protocols in the Kasimir project, Appl. Intell. 28 (2008) 261–274.
[84] J. Lieber, E. Nauer, H. Prade, When revision-based case adaptation meets analogical extrapolation, in: A.A. Sánchez-Ruiz, M.W. Floyd (Eds.), Case-Based Reasoning Research and Development, vol. 12877, Springer, 2021, pp. 156–170.
[85] A. Lovett, K. Forbus, Supplemental material for modeling visual problem solving as analogical reasoning, Psychol. Rev. 124 (2017) 60–90.
[86] E. Marquer, S. Alsaidi, A. Decker, P.-A. Murena, M. Couceiro, A deep learning approach to solving morphological analogies, in: M.T. Keane, N. Wiratunga (Eds.), Case-Based Reasoning Research and Development, vol. 13405, Springer International Publishing, Cham, 2022, pp. 159–174.
[87] E. Marquer, P.-A. Murena, M. Couceiro, Transferring learned models of morphological analogy, in: ICCBR Analogies'22: Workshop on Analogies: From Theory to Applications at ICCBR-2022, Nancy, France, 2022.
[88] C. McConnell, B. Smyth, 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, vol. 3, ICCBR'19, 2019, pp. 155–171.
[89] D. McSherry, Demand-driven discovery of adaptation knowledge, in: Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'99, 1999, pp. 222–227.
[90] L. Miclet, S. Bayoudh, A. Delhay, Analogical dissimilarity: definition, algorithms and two experiments in machine learning, J. Artif. Intell. Res. 32 (2008) 793–824.
[91] M. Mitchell, Abstraction and analogy-making in artificial intelligence, Ann. N.Y. Acad. Sci. 1505 (2021) 79–101, arXiv:2102.10717.
[92] L.R. Novick, Analogical transfer, problem similarity, and expertise, J. Exp. Psychol. Learn. Mem. Cogn. 14 (1988) 510–529.
[93] L.R. Novick, K.J. Holyoak, Mathematical problem solving by analogy, J. Exp. Psychol. Learn. Mem. Cogn. 17 (1991) 398–415.
[94] S. Ontañón, E. Plaza, On knowledge transfer in case-based inference, in: B. Díaz-Agudo, I. Watson (Eds.), ICCBR Proceedings, Springer, 2012, pp. 312–326.
[95] P.K. Paritosh, M.E. Klenk, Cognitive processes in quantitative estimation: analogical anchors and causal adjustment, in: Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci-06), 2006.
[96] G.P. Pena, J. de Souza Andrade-Filho, Analogies in medicine: valuable for learning, reasoning, remembering and naming, Adv. Health Sci. Educ. 15 (2010) 609–619.
[97] J. Peyre, I. Laptev, C. Schmid, J. Sivic, Detecting unseen visual relations using analogies, in: IEEE/CVF Int. Conference on Computer Vision, 2019, pp. 1981–1990.
[98] H. Prade, Reasoning with data - a new challenge for AI?, in: S. Schockaert, P. Senellart (Eds.), Scalable Uncertainty Management, vol. 9858, Springer, 2016, pp. 274–288.
[99] H. Prade, G. Richard, A short introduction to computational trends in analogical reasoning, in: H. Prade, G. Richard (Eds.), Computational Approaches to Analogical Reasoning: Current Trends, vol. 548, Springer, 2014, pp. 1–22.
[100] C.K. Riesbeck, R.C. Schank, Inside Case-Based Reasoning, Lawrence Erlbaum Associates, 1989.
[101] B. Smyth, M.T. Keane, A few good counterfactuals: generating interpretable, plausible and diverse counterfactual explanations, in: M.T. Keane, N. Wiratunga (Eds.), Case-Based Reasoning Research and Development, vol. 13405, Springer, 2022, pp. 18–32.
[102] I. Turksen, Z. Zhong, An approximate analogical reasoning approach based on similarity measures, IEEE Trans. Syst. Man Cybern. 18 (Nov.–Dec. 1988) 1049–1056.
[103] A. Ushio, L. Espinosa Anke, S. Schockaert, J. Camacho-Collados, BERT is to NLP what AlexNet is to CV: can pre-trained language models identify analogies?, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Association for Computational Linguistics, 2021, pp. 3609–3624.
[104] K. Washio, T. Kato, Neural latent relational analysis to capture lexical semantic relations in a vector space, in: Proc. of the Int. Conf. on Empirical Methods in Natural Language Processing, EMNLP'18, Association for Computational Linguistics, 2018, pp. 594–600.