CoAT-APC: When Analogical Proportion-based Classification Meets Case-based Prediction

Fadi Badra[1,*], Marie-Jeanne Lesot[2,*]

[1]Université Sorbonne Paris Nord, LIMICS, Sorbonne Université, INSERM, Bobigny, France
[2]Sorbonne Université, CNRS, LIP6, Paris, France

Abstract
This paper proposes to view analogical proportion-based classification as a special type of case-based prediction algorithm, in which (i) cases are *differences* between two instances, and (ii) only maximally similar cases are compared. It then proposes to tweak the CoAT case-based prediction algorithm in order to implement these two key design principles. The resulting analogical proportion-based classifier CoAT-APC shows a performance comparable to state-of-the-art analogical proportion-based classifiers, while implementing a different transfer strategy, based on the minimization of a dataset complexity measure, as opposed to a rule-based approach. Experimental results show the usefulness of combining these two design principles and suggest that the rule-based transfer strategy of analogical proportion-based classifiers has comparatively little impact on the performance of the system.

Keywords
analogical proportion-based classification, case-based prediction, interactions between case-based and analogical reasoning

1. Introduction

Case-based prediction [1, 2] consists in predicting the outcome (the label, in classification tasks) of a new case directly from its comparison with a set of cases retrieved from a case base and some similarity measures, without any attempt to learn a model of the observed data prior to the inference. In case-based prediction methods, cases are assumed to be pairs (situation, outcome) and the predicted outcome is the one that best enforces a compatibility requirement, according to which outcome similarities in the resulting case base should be compatible with the corresponding situation similarities. This principle can be expressed in numerous ways, leading to a large variety of case-based prediction algorithms, as discussed later in the paper. They mainly differ in their transfer strategy, that can for instance rely on rule-based or optimization-based approaches.

On the other hand, analogical proportion-based classification (abbreviated APC in the rest of the paper, see e.g. [3, 4, 5, 6, 7]) proposes to exploit the principle of analogical reasoning, based on statements of the form "**a** is to **b** as **c** is to **d**", to predict the label of new instances: when

ICCBR Analogies'22: Workshop on Analogies: from Theory to Applications at ICCBR-2022, September, 2022, Nancy, France
*Corresponding author.
badra@univ-paris13.fr (F. Badra); marie-jeanne.lesot@lip6.fr (M. Lesot)
0000-0002-2437-8230 (F. Badra); 0000-0002-3604-6647 (M. Lesot)

an analogical proportion holds on the instance descriptions, it is inferred that an analogical proportion also holds on their associated labels.

This paper proposes to compare these two classification paradigms and to study APC from the point of view of case-based prediction. More precisely, it shows that APC can be interpreted as a special kind of case-based prediction method, that applies a rule-based transfer strategy and implements the two following design principles:

[P1] Cases are *differences* between two instances of the considered instance set.

[P2] Only maximally similar cases are compared.

The second principle can be relaxed to comparing only the most similar cases, beyond the maximally similar ones, as discussed later in the paper.

The paper then investigates why analogical proportion-based classifiers exhibit better performance compared to other case-based prediction methods. To do so, a new case-based prediction method is proposed, called CoAT-APC, that tweaks the CoAT case-based prediction algorithm [8, 9] to implement the two principles [P1] and [P2] with the relaxed version of the latter. The resulting algorithm is an analogical proportion-based classifier, in which the rule-based transfer strategy is replaced by CoAT's optimization-based transfer strategy.

Experiments conducted on a variety of benchmark data sets, both of Boolean and numerical types, show that CoAT-APC offers performances comparable to state-of-the-art analogical proportion-based classifiers, and that complying with the two principles [P1] and [P2] significantly improves the performance of the original CoAT algorithm. This suggests that the two principles [P1] and [P2] do play a major role in their success, whereas their rule-based transfer strategy has little impact on their efficiency, and can be replaced without harm with another transfer strategy.

The paper is structured as follows: Section 2 recalls the main definitions about analogical proportion-based classifiers. Section 3 considers case-based prediction methods and in particular reviews their main prediction strategies. Section 4 shows that APC can be formulated as a case-based prediction method with rule-based transfer strategy. Section 5 presents the proposed exploitation of this view in the CoAT-APC algorithm, that modifies the CoAT case-based prediction method with optimization transfer strategy [8, 9] in order to implement the two key design principles of APC. Section 6 presents some experiments to validate the approach. Section 7 concludes the paper and discusses some directions for future work.

2. Analogical Proportion-based Classification

Analogical proportion-based classifiers have shown competitive results in classification and recommendation tasks, see e.g. [10, 3, 11, 12, 5]. They apply the principle of analogical reasoning [13], 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, labelling 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 $\mathbf{a}$,$\mathbf{b}$,$\mathbf{c}$, $\ldots$ 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 triples (**a**,**b**,**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 [4]. 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 [4]. 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 [10]. 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.

3. Case-Based Prediction: A Comparative Study

This section recaps the general principles of case-based prediction methods and reviews the main transfer strategies they rely on, proposing to distinguish between four categories discussed in turn in Sections 3.2 to 3.5.

3.1. Definition and Notations

Case-based prediction typically considers the following setting: $\mathcal{S}$ denotes an input space and $\mathcal{O}$ an output space. An element of $\mathcal{S}$ is called a *situation*, and an element of $\mathcal{O}$ is called an *outcome*, or a result. A finite set $CB = \{(s_1, r_1), \ldots, (s_n, r_n)\}$ of elements in $\mathcal{S} \times \mathcal{O}$ is called a *case base*. 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 addition, σ_S and σ_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, a common decomposition of the case-based inference involves three main tasks [14]:

- *Retrieval*: retrieve from CB a set of source cases $c_s = (s, r_s)$;
- *Mapping*: for each retrieved situation s, compute 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.

In the transfer task, as illustrated in the diagram of Fig. 1, a plausible inference is triggered in order to estimate the similarity $\sigma_R(r_s, r_t)$ on outcomes from the similarity $\sigma_S(s, t)$ on situations: it applies the principle according to which if two situations are similar, then it is plausible that their outcomes are also similar.

For a new situation t, case-based prediction is then a search, among all potential outcomes $r \in \mathcal{O}$, for the outcome r_t that makes the plausible inference most likely to succeed when the

$$\begin{array}{c} s \;—\; \sigma_S \;—\; t \\ \Big\downarrow \text{used to estimate} \\ r_s \;—\; \sigma_R \;—\; r_t \end{array}$$

Figure 1: Schematic view of the transfer task in case-based prediction: the similarity relation on situations used to estimate the similarity relation on outcomes.

new case is added to the source case to build the augmented case base $CB \cup \{(t, r_t)\}$. To find this outcome, case-based prediction methods express the plausible inference as a compatibility requirement on the resulting similarity relations: when the new case is compared to the retrieved cases, the outcome similarities should be compatible with the observed situation similarities. In the following, $\hat{c}_t = (t, r)$ denotes a potential new case formed by choosing the outcome $r \in \mathcal{O}$ for the new case.

Different prediction strategies can be found in the literature to express this compatibility requirement between the two similarity measures. The next sections propose to distinguish between four categories of transfer strategies, respectively named transfer by rule-based voting, by constraint, by evidence support and by optimization.

3.2. Transfer by Rule-based Voting

A first type of transfer strategy relies on 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 [15, 16, 17, 18], dependencies between problem and solution features [19], co-variations [20] or fuzzy rules [21] 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.

The proposed outcome r_t is obtained by a majority vote on the set of outcomes r derived from the rules. Triggering a rule consists in performing a similarity-based inference (SBI), applying variants of the modus ponens schema [21, 22, 23]: 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

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

It can be noted that 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 $\widetilde{\sigma_S}$ of σ_S and $\widetilde{\sigma_R}$ of σ_R 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*, that may be acquired from an expert [24], from the user [25] or learned from data [26, 27, 15, 16, 18].

3.3. Transfer by Continuity Constraints

Another strategy 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 [28], 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 [29], or gradual rules or certainty rules [30, 31, 32]. 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.

3.4. Transfer by Evidence Support

A more data-driven type of 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 or the possibilistic instance-based learning approach [28, 33, 34]. 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 would be observed with a source case.

3.5. Transfer by 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 [8] proposes to define a global indicator that measures the compatibility of σ_R with σ_S on the whole case base: the prediction then consists in minimizing the value of a dataset complexity indicator when augmented with the new case and its candidate associated outcome. This principle is implemented in the CoAT, for **Co**mplexity-based **A**nalogical **T**ransfer, algorithm [8, 9]. 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 $c_j = (s_j, r_j)$:

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

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:

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

$$\begin{tikzcd}
\underset{\mathbf{a}:\mathbf{b}}{s} \arrow[r, dash] & AD \arrow[r, dash] \arrow[d, "\text{used to estimate}"] & \underset{\mathbf{c}:x}{t} \\
\underset{f(\mathbf{a}):f(\mathbf{b})}{r_s} \arrow[r, dash] & AD \arrow[r, dash] & \underset{f(\mathbf{c}):f(x)}{r_t}
\end{tikzcd}$$

Figure 2: In analogical classifiers, both situations and outcomes represent ratios.

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:

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

4. Analogical Proportion-based Classification as a Case-Based Prediction Method

This section proposes to establish a correspondence between APC and case-based prediction, showing the former can be viewed as a special kind of the latter. This correspondence is illustrated by the diagram given in Fig. 2 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 rule-based voting method: first, cases are *differences* between two instances (principle [P1]), and a single rule is triggered, that states that maximally similar situations (principle [P2]) should be associated with maximally similar outcomes. This section makes explicit all components of the case-based prediction configuration that can be associated to a given analogical proportion-based classifier, discussing successively the considered case base and similarity measures, as well as the applied transfer strategy.

4.1. Case Base

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, these ratios are represented by the differences $s = \mathbf{a} - \mathbf{b}$ and $r_s = f(\mathbf{a}) - f(\mathbf{b})$ between two vectors. 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, represented as one-hot encoding vectors. APC implements the [P1] principle: cases are *differences* between instances of the data set D.

4.2. Similarity Measures

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 $\sigma_R(f(\mathbf{a}) - f(\mathbf{b}), f(\mathbf{c}) - f(\mathbf{d})) = 1$.

4.3. Transfer Strategy

When APC is viewed as a case-based prediction method, its transfer strategy is a rule-based voting strategy [35]. To see why, consider the decomposition described in [36] of the prediction procedure as 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:

1. enumerating all instances $\mathbf{c}$, and for each one of them,
2. *Retrieval*: retrieve all source cases $c_s = (s, r_s) = (\mathbf{a} - \mathbf{b}, f(\mathbf{a}) - f(\mathbf{b}))$;
3. *Mapping*: compute the similarity $\sigma_S(s, t)$ between $s = \mathbf{a} - \mathbf{b}$ and $t = \mathbf{c} - x$;
4. *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 $(c_s, \hat{c}_t)$:

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

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 was 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:

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

Only the k solutions y that were derived from the rule $(\sigma_S = 1) \to (\sigma_R = 1)$ with the highest values of $\sigma_S(c_s, \hat{c}_t)$ are added to the solution set. Therefore, when viewed as case-based prediction methods, analogical proportion-based classifiers implement the relaxed form of the [P2] principle (only most similar cases are compared).

Algorithm 1 CoAT-APC

inputs: D (data set), x (additional instance), C (set of potential classes for $f(x)$), σ_S, σ_R (situation and outcome similarity measures), k (number of neighbor instances), n (size of the case base).
output: the predicted value $f(x)$ for the new instance x
$\ell \leftarrow \{y : 0 \text{ for } y \in C\}$
$\mathcal{N}(x) \leftarrow$ the k instances $\mathbf{c} \in D$ that maximize $\sigma_S(\mathbf{c}, x)$
for $\mathbf{c} \in \mathcal{N}_k(x)$ **do**
  $t \leftarrow \mathbf{c} - x$
  $CB(\mathbf{c}) \leftarrow$ the n source cases $c_s = (s, r_s)$ that maximize $\sigma_S(s, t)$
  for all $y \in C$ **do**
    $r \leftarrow f(\mathbf{c}) - y$
    $\ell[y] \leftarrow \ell[y] + \Gamma(\sigma_S, \sigma_R, CB(\mathbf{c}) \cup \{(t, r)\})$
  end for
end for
$f(x) \leftarrow \arg\min_{y \in C} \ell[y]$
return $f(x)$

5. Tweaking CoAT to be an Analogical Proportion-based Classifier

The CoAT case-based prediction algorithm is modified in order to implement the two key principles [P1] and [P2] in its relaxed form. The resulting analogical proportion-based classifier, called CoAT-APC, implements a transfer strategy based on the minimization of a dataset complexity measure instead of being rule-based. Algo. 1 provides the pseudo-code description of the resulting CoAT-APC algorithm.

5.1. Proposed CoAT-APC Algorithm

Regarding the case base, the source cases c_s and potential new cases $\hat{c}_t$ are defined as in the previous section (Eq. 1) as differences between instances of the data set (principle [P1]).

Regarding the transfer strategy, CoAT-APC selects a set of instances $\mathbf{c} \in D$, and for each of them applying the CoAT method to evaluate the plausibility of each potential outcome (*i.e.*, class difference) $r = f(\mathbf{c}) - y$ associated with the target situation $t = \mathbf{c} - x$. The plausibility estimations obtained for each $y \in C$ are then aggregated to propose a solution $f(x)$. The decision procedure thus consists in successively:

1. finding the k instances $\mathbf{c}$ that maximize $\sigma_S(\mathbf{c}, x)$, and for each one of them,
2. forming a case base $CB(\mathbf{c})$ with the n source cases $s_c = (s, r_s)$ that maximize $\sigma_S(s, t)$, with $s = \mathbf{a} - \mathbf{b}$ and $t = \mathbf{c} - x$ (principle relaxed [P2]),
3. computing and storing $\Gamma(\sigma_S, \sigma_R, CB(\mathbf{c}) \cup \{(t, r)\})$ formed with $t = \mathbf{c} - x$ and $r = f(\mathbf{c}) - y$, for each potential solution $y \in C$,

Type	Dataset	Instances	Features	Classes
Discrete	balance	625	4	3
	car	1728	6	4
	monks1	432	6	2
	monks2	432	6	2
	monks3	432	6	2
	spect	238	23	2
	voting	435	16	2
	breastw	682	10	2
Continuous	iris	150	4	3
	pima	768	8	2
	user	258	5	4
	wine	178	4	3

Table 1
Data sets used for the experimental study of the proposed algorithm CoAT-APC and its variants.

4. aggregating the plausibility estimations by summing over $\mathbf{c}$:

$$f(x) = \arg\min_{y \in C} \sum_{\mathbf{c}} \Gamma(\sigma_S, \sigma_R, CB(\mathbf{c}) \cup \{(\mathbf{c} - x, f(\mathbf{c}) - y)\})$$

5.2. Computational Complexity Analysis

The first step (finding the k instances $\mathbf{c}$ that maximize $\sigma_S(\mathbf{c}, x)$) is in line with [11]. It was shown experimentally to slightly improve the results while greatly reducing the computational cost of the decision procedure. The second step (forming the case base $CB(\mathbf{c})$) requires to precompute and store all $|D|^2$ differences between instances of D (which is done beforehand), and then sort these differences at runtime by decreasing value of $\sigma_S(s, t)$. The sorting procedure is done in $\mathcal{O}(|D|^2 \texttt{log}(|D|))$, which may be the most costly part of the algorithm. The third step (computing Γ for each potential $y \in C$) can be done in $\mathcal{O}(n^2|C|)$, as shown in [9], which is tractable since optimal results are usually obtained with $n \leq 40$, as experimental results show. The last step consists in summing the plausibility estimations over $\mathbf{c}$ for each potential solution $y \in C$, and selecting the one that minimizes the sum. The overall computational complexity of the method is $\mathcal{O}(k \times (|D|^2 \texttt{log}(|D|) + n^2|C|))$.

6. Experiments

This section describes the experiments run to validate the proposed CoAT-APC algorithm and the strategy it relies on, and to determine the impact that the principles [P1] and [P2] have on the performance of the proposed case-based prediction instantiation of analogical proportion-based classification.

Algorithm 2 CoAT

inputs: D (data set), x (additional instance), C (set of potential classes for $f(x)$), σ_S, σ_R (situation and outcome similarity measures).
output: the predicted value $f(x)$ for the new instance x
$CB \leftarrow D$
$f(x) \leftarrow \arg\min_{y \in C} \Gamma(\sigma_S, \sigma_R, CB \cup \{(x, y)\})$
return $f(x)$

6.1. Experimental Protocol

The data sets used for classification are taken from the UCI repository[1], their characteristics are summarized in Table 1. They include 8 data sets with only nominal features and 4 data sets with only numerical features, in both cases associated with classification tasks with 2 to 4 classes. In all experiments, the two similarity measures σ_S and σ_R are fixed:

- $\sigma_S = e^{-ED}$, where $ED = \|\cdot\|_2$ is the standard Euclidean distance;
- $\sigma_R(u, v) = 1$ if $u = v$, and 0 otherwise.

Four algorithms are considered for comparison, applied to each data set D:

- CoAT: the case-based prediction algorithm, as of [9].The source cases are the instances of D, and the case base CB contains the whole data set D (see Algorithm 2).
- CoAT+[P1]: a modification of CoAT that implements principle [P1]. The source cases $c_s = (\mathbf{a} - \mathbf{b}, f(\mathbf{a}) - f(\mathbf{b}))$ are differences between instances of the data set D. The algorithm is the same as the one of CoAT-APC, but for each instance $\mathbf{c}$, the case base $CB(\mathbf{c})$ includes n randomly chosen source cases.
- CoAT+[P2]: a modification of CoAT that implements the relaxed form of principle [P2]. The source cases are instances of D, but the case base CB contains only the n instances $s \in D$ that maximize $\sigma_S(s, x)$ (see Algorithm 3).
- CoAT-APC: a combination of the two previous approaches. The source cases $c_s = (\mathbf{a} - \mathbf{b}, f(\mathbf{a}) - f(\mathbf{b}))$ are differences between instances of the data set D, and the case base $CB(\mathbf{c})$ contains the n source cases that maximize $\sigma_S(\mathbf{a} - \mathbf{b}, \mathbf{c} - x)$.

For each task, the performance is measured by the prediction accuracy, with 10-fold cross validation. The algorithm CoAT+[P2] is tested on each data set with a parameter n (the size of the case base CB) varying between 5 and $|D|$, by steps of 5. The algorithms CoAT+[P1] and CoAT-APC are tested on each data set for all pairs (k, n) with k varying between 3 and 51, by steps of 2, and n varying between 5 and 50 by steps of 5.

6.2. Results

Table 2 gives the classification results. For each data set, the best results considering standard deviations are marked in bold. When the two principles [P1] and [P2] are combined (algorithm

[1] https://archive.ics.uci.edu/ml/

Algorithm 3 CoAT+[P2]

inputs: D (data set), x (additional instance), C (set of potential classes for $f(x)$), σ_S, σ_R (situation and outcome similarity measures), n (size of the case base).
output: the predicted value $f(x)$ for the new instance x
$CB \leftarrow$ the n instances $c_s = (s, f(s)) \in D$ that maximize $\sigma_S(s, x)$
$f(x) \leftarrow \arg\min_{y \in C} \Gamma(\sigma_S, \sigma_R, CB \cup \{(x, y)\})$
return $f(x)$

Dataset	CoAT	CoAT+[P1]			CoAT+[P2]		CoAT-APC		
			k	n		n		k	n
balance	61.0% ± 5.45	77.1% ± 3.76	5	45	**92.8%**±2.44	180	**93.8%**±2.78	39	20
car	62.5% ± 2.51	81.3% ± 1.49	19	10	82.7% ± 2.61	85	**96.5%**±0.97	11	5
monks1	66.6% ± 6.85	91.9% ± 3.20	7	30	88.3% ± 5.38	20	**100%** ± 0.00	7	5
monks2	65.7% ± 0.35	74.4% ± 6.24	3	40	65.2% ± 1.45	55	**91.1%**±4.99	25	20
monks3	82.1% ± 7.09	**98.7%**±1.61	9	35	**97.1%**±2.57	45	**98.9%**±1.19	7	20
spect	**79.3%**±1.50	**82.7%**±9.46	11	15	**84.2%**±5.51	90	**82.0%**±6.99	47	35
voting	88.0% ± 3.96	**92.7%**±3.47	7	15	**94.0%**±3.24	65	**95.8%**±2.51	7	20
breastw	65.0% ± 0.33	67.1% ± 3.09	25	35	66.6% ± 1.88	400	**96.6%**±1.97	21	10
iris	**95.3%**±5.20	**98.6%**±2.67	17	40	**96.7%**±3.33	15	**99.3%**±2.00	27	20
pima	65.1% ± 3.42	**75.5%**±3.51	15	30	**75.4%**±3.91	270	**75.3%**±3.47	27	10
user	63.6% ± 4.59	31.9% ± 5.23	3	5	71.5% ± 6.55	25	**97.0%**±1.48	23	15
wine	64.9% ± 7.59	22.0% ± 9.47	3	10	60.6% ± 5.88	90	**94.4%**±4.31	51	5

Table 2
Classification results.

CoAT-APC), the resulting system offers a very good performance, and gives the best results for all data sets. This allows to validate the proposed approach and the integration of the analogical proportion principles into case-based predictions. When applied independently, the design principles [P1] and [P2] seem to have different impacts on the performance. Running CoAT with a case base restricted to the n cases that are most similar with the target situation (algorithm CoAT+[P2]) generally improves the performance, but not for all data sets (see e.g., `breastw`, `wine`, or `monks2`). Working on cases defined as differences between instances of D but with a random case base (algorithm CoAT+[P1]) often improves the performance as well, but for some data sets the performance results are surprisingly low (see e.g., `user`, `wine`, `breastw`, or even `monks2`). All tests were run using a fixed similarity measure σ_S, based on the Euclidean distance, which may not be optimal for all data sets. This may explain why the CoAT algorithm sometimes gives rather poor results. However, applying [P1] and [P2] design principles (CoAT-APC) greatly improves the performance of the classifier, even with this non-optimal similarity measure. In addition, it can be observed that the CoAT-APC algorithm obtains the best results for fairly low values of n, usually lower than 20. This parameter determines the size of the case base $CB(\mathbf{c})$. The computing time remains moderate: for the Balance Scale data set, with $k = 39$ and $n = 20$, predicting the class of a new instance takes 11.5 seconds on a current PC.

7. Conclusion and Future Works

This paper proposes an approach to bridge the gap between analogical proportion-based classifiers and case-based prediction algorithms, by showing that analogical proportion-based classification can be interpreted as a special kind of case-based prediction algorithm in which cases are differences between two instances of the data set, and maximally similar cases are compared to predict the class of a new instance. Results show that if these two design principles taken independently have an impact on the prediction performance of the case-based prediction system, they are especially powerful when combined, even when the prediction is done with a non-optimal similarity measure. On the contrary, the rule-based transfer strategy of analogical proportion-based classifiers seems to have a little impact on their efficiency: replacing it with a different transfer strategy, such as the CoAT's optimization strategy, leads to excellent performance results.

Future works will include comparing the CoAT-APC algorithm with state-of-the-art analogical proportion-based classification algorithms, both in terms of performance and computing time. More generally, there is a need for a shared implementation of the main case-based prediction algorithms, so that their performance can be compared on controlled benchmarks. The study also shows that as a case-based prediction algorithm, analogical proportion-based classifiers use a similarity measure constructed from the Euclidean distance. Future works will include learning a similarity measure that is more adequate to each considered case-based prediction task.

References

[1] E. Hüllermeier, Case-Based Approximate Reasoning, Theory and Decision Library B, Springer, 2007.

[2] D. Dubois, E. Hüllermeier, H. Prade, Fuzzy methods for case-based recommendation and decision support, Journal of Intelligent Information Systems 27 (2006) 95–115.

[3] L. Miclet, S. Bayoudh, A. Delhay, Analogical dissimilarity: Definition, algorithms and two experiments in machine learning, Journal of Artificial Intelligence Research 32 (2008) 793–824.

[4] N. Hug, H. Prade, G. Richard, M. Serrurier, Analogical classifiers: A theoretical perspective, in: Proc. of the 22nd European Conf. on Artificial Intelligence, ECAI, volume 285, 2016, pp. 689–697.

[5] N. Hug, H. Prade, G. Richard, M. Serrurier, Analogical proportion-based methods for recommendation – First investigations, Fuzzy Sets and Systems 366 (2019) 110–132.

[6] M. Bounhas, Logical Proportions for Classification and Preference Learning, HDR, Tunis University, 2021.

[7] H. Prade, G. Richard, Analogical proportions: why they are useful in AI, in: Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI, 2021, pp. 4568–4576.

[8] F. Badra, A Dataset Complexity Measure for Analogical Transfer, in: Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI, 2020, pp. 1601–1607.

[9] F. Badra, M.-J. Lesot, Theoretical and experimental study of a complexity measure for analogical transfer, in: Proc. of the Int. Conf. on Case-based Reasoning ICCBR, 2022.

[10] S. Bayoudh, Learning by analogy: a classification rule for binary and nominal data, in: Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI, 2007, pp. 678–683.

[11] M. Bounhas, H. Prade, G. Richard, Analogy-based classifiers for nominal or numerical data, Int. J. of Approximate Reasoning 91 (2017) 36–55.

[12] 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, 2017, pp. 1575–1581.

[13] D. R. Hofstadter, E. Sander, Surfaces and Essences: Analogy as the Fuel and Fire of Thinking, Basic Books, 2013.

[14] H. Gust, U. Krumnack, K. Kühnberger, A. Schwering, Analogical reasoning: A core of cognition., KI - Künstliche Intelligenz 22 (2008) 8–12.

[15] 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, 2007, pp. 750–755.

[16] F. Badra, R. Bendaoud, R. Bentebibel, P.-A. Champin, J. Cojan, A. Cordier, S. Després, 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: Workshop Proc. of the 9th European Conf. on Case-Based Reasoning, 2008, pp. 219–228.

[17] S. Ontañón, E. Plaza, On knowledge transfer in case-based inference, in: ICCBR Proceedings, Springer, 2012, pp. 312–326.

[18] 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, 2017, pp. 4874–4878.

[19] B. Fuchs, J. Lieber, A. Mille, A. Napoli, Differential adaptation: An operational approach to adaptation for solving numerical problems with CBR, Knowledge-Based Systems 68 (2014) 103–114.

[20] F. Badra, Reasoning with co-variations, in: Proc. of the 17th Int. Conf. on Artificial Intelligence: Methodology, Systems and Applications, AIMSA, 2016.

[21] B. Bouchon-Meunier, L. Valverde, A fuzzy approach to analogical reasoning, Soft Computing 3 (1999) 141–147.

[22] V. V. Cross, T. A. Sudkamp, Similarity and compatibility in fuzzy set theory, volume 93 of *Studies in Fuzziness and soft computing*, Springer, 2002.

[23] F. Badra, K. Sedki, A. Ugon, On the role of similarity in analogical transfer, in: Case-Based Reasoning Research and Development, volume 11156, Springer, 2018, pp. 499–514.

[24] J. Lieber, M. d'Aquin, F. Badra, A. Napoli, Modeling adaptation of breast cancer treatment decision protocols in the Kasimir project, Applied Intelligence 28 (2008) 261–274.

[25] F. Badra, A. Cordier, J. Lieber, Opportunistic adaptation knowledge discovery, in: Case-Based Reasoning Research and Development, Springer, 2009, pp. 60–74.

[26] K. Hanney, M. T. Keane, The adaptation knowledge bottleneck: How to ease it by learning from cases, in: Case-Based Reasoning Research and Development, volume 1266, Springer, 1997, pp. 359–370.

[27] S. Craw, N. Wiratunga, R. C. Rowe, Learning adaptation knowledge to improve case-based reasoning, Artificial Intelligence 170 (2006) 1175–1192.
[28] J. Beringer, E. Hüllermeier, Case-based learning in a bipolar possibilistic framework, International Journal of intelligent Systems 23 (2008) 1119–1134.
[29] E. Hüllermeier, Credible case-based inference using similarity profiles, IEEE Transactions on Knowledge and Data Engineering 19 (2007) 847–858.
[30] E. Hüllermeier, D. Dubois, H. Prade, Fuzzy Rules in Case-Based Reasoning, in: In Proc. of the Conf. AFIA99 Raisonnement à partir de cas, 1999, pp. 45–54.
[31] E. Hüllermeier, D. Dubois, H. Prade, Knowledge-based extrapolation of cases: a possibilistic approach, in: Technologies for Constructing Intelligent Systems 1, volume 89, Physica-Verlag HD, 2002, pp. 377–390.
[32] E. Hüllermeier, Case-Based Approximate Reasoning, Springer, 2007.
[33] 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.
[34] E. Hüllermeier, Possibilistic instance-based learning, Artificial Intelligence 148 (2003) 335–383.
[35] M. Bounhas, H. Prade, G. Richard, Analogical classification: a rule-based view, in: Proc. of the Int. Conf. o Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU, volume 443, Springer, 2014, pp. 485–495.
[36] J. Lieber, E. Nauer, H. Prade, When revision-based case adaptation meets analogical extrapolation, in: Case-Based Reasoning Research and Development, volume 12877, Springer, 2021, pp. 156–170.