Continuous Analogical Proportions-Based Classifier

Marouane Essid[1], Myriam Bounhas[1,2](✉), and Henri Prade[3]

[1] Larodec Lab, ISG de Tunis, Tunis, Tunisia
marouane.essid@gmail.com, myriam_bounhas@yahoo.fr
[2] Emirates College of Technology, Abu Dhabi, United Arab Emirates
[3] CNRS, IRIT, Université Paul Sabatier, Toulouse, France
prade@irit.fr

Abstract. Analogical proportions, often denoted $A : B :: C : D$, are statements of the form "A is to B as C is to D" that involve comparisons between items. They are at the basis of an inference mechanism that has been recognized as a suitable tool for classification and has led to a variety of analogical classifiers in the last decade. Given an object D to be classified, the basic idea of such classifiers is to look for triples of examples (A, B, C), in the learning set, that form an analogical proportion with D, on a maximum set of attributes. In the context of classification, objects A, B, C and D are assumed to be represented by vectors of feature values. Analogical inference relies on the fact that if a proportion $A : B :: C : D$ is valid, one of the four components of the proportion can be computed from the three others. Based on this principle, analogical classifiers have a cubic complexity due to the search for all possible triples in a learning set to make a single prediction. A special case of analogical proportions involving only three items A, B and C are called *continuous* analogical proportions and are of the form "A is to B as B is to C" (hence denoted $A : B :: B : C$). In this paper, we develop a new classification algorithm based on continuous analogical proportions and applied to numerical features. Focusing on pairs rather than triples, the proposed classifier enables us to compute an unknown midpoint item B given a pair of items (A, C). Experimental results of such classifier show an efficiency close to the previous analogy-based classifier while maintaining a reduced quadratic complexity.

Keywords: Classification · Analogical proportions · Continuous analogical proportions

1 Introduction

Reasoning by analogy establishes a parallel between two situations. More precisely, it enables us to relate two pairs of items (a, b) and (c, d) in such way that "a is to b as c is to d" on a comparison basis. This relationship, often noted $a : b :: c : d$, expresses a kind of equality between the two pairs, i.e., the

M.-J. Lesot et al. (Eds.): IPMU 2020, CCIS 1237, pp. 541–555, 2020.
https://doi.org/10.1007/978-3-030-50146-4_40

two items of the first pair are similar and differ in the same way as the two items of the second pair. The case of numerical (geometric) proportions where we have an equality between two ratios (i.e., $a/b = c/d$) is at the origin of the name "analogical proportions". Analogical proportions, when d is unknown, provides an extrapolation mechanism, which with numbers yields $d = (b \times c)/a$, and $d = b + c - a$ in case of arithmetic proportions (such that $a - b = c - d$). The analogical proportions-based extrapolation has been successfully applied to classification problems [4,8]. The main drawback of algorithms using analogical proportions is their cubic complexity.

A particular case of analogical proportions, named *continuous* analogical proportions, is obtained when the two central components are equal, namely they are statements of the form "a is to b as b is to c". In case of numerical proportions, if we assume that b is unknown, it can be expressed in terms of a and c as $b = \sqrt{a \times c}$ in the geometric case and $b = (a + c)/2$ in the arithmetic case. Note that similar inequalities hold in both cases: $min(a, c) \leq \sqrt{a \times c} \leq max(a, c)$ and $min(a, c) \leq (a + c)/2 \leq max(a, c)$. This means that the continuous analogical proportion induces a form of interpolation between a and c in the numerical case by involving an intermediary value that can be obtained from a and c. A continuous analogical proportions-based interpolation was recently proposed as a way of enlarging a training set (before applying some standard classification methods), and led to good results [2]. In contrast to extrapolation, interpolation with analogy-based classifiers has a quadratic complexity.

In this paper, we investigate the efficiency for classification of using such approach. The paper is organized as follows. Section 2 provides a short background on analogical proportions and more particularly on continuous ones. Then Sect. 3 surveys related work on analogical extrapolation. Section 4 presents the proposed interpolation approach for classification. Finally, Sect. 5 reports the results of our algorithm.

2 Background on Analogical Proportions

An analogical proportion is a relationship on X^4 between 4 items $A, B, C, D \in X$. This 4-tuple, when it forms an analogical proportion is denoted $A : B :: C : D$ and reads "A is to B as C is to D". Both relationships "is to" and "as" depend on the nature of X [9]. As it is the case for numerical proportions, the relation of analogy still holds when the pairs (A, B) and (C, D) are exchanged, or when central items B and C are permuted (see [11] for other properties). In the following subsections, we recall analogical proportions in the Boolean setting (i.e., $X \in \mathbb{B} = \{0, 1\}$)) and their extension for nominal and for real-valued settings (i.e., $X \in [0, 1]$), before considering the special case of *continuous* analogical proportions.

2.1 Analogical Proportions in the Boolean Setting

Let us consider four items A, B, C and D, respectively described by their binary values $a, b, c, d \in \mathbb{B} = \{0, 1\}$. Items A, B, C and D are in analogical proportion,

Table 1. Truth table for analogical proportion

a	b	c	d	$a : b :: c : d$
0	0	0	0	1
0	0	1	1	1
0	1	0	1	1
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

which is denoted $A : B :: C : D$ if and only if $a : b :: c : d$ holds true (it can also be written $a : b :: c : d = 1$ or simply $a : b :: c : d$). The truth table (Table 1) shows the six possible assignments for a 4-tuple to be in analogical proportion, out of sixteen possible configurations.

Boolean analogical proportions can be expressed by the logical formula:

$$a : b :: c : d = (a \wedge \neg b \equiv c \wedge \neg d) \wedge (\neg a \wedge b \equiv \neg c \wedge d) \tag{1}$$

See [10,12] for justification. This formula holds true for the 6 assignments shown in the truth table. It reads "a differs from b as c differs from d and b differs from a as d differs from c", which fits with the expected meaning of analogy. An equivalent formula is obtained by negating the two sides of the first and the second equivalence in formula (1):

$$a : b :: c : d = (a \rightarrow b \equiv c \rightarrow d) \wedge (b \rightarrow a \equiv d \rightarrow c) \tag{2}$$

Items are generally described by vectors of Boolean values rather than by a single value. A natural extension for vectors in $\{0,1\}^n$ of the form $\boldsymbol{x} = (x_1, \cdots, x_n)$ is obtained component-wise as follows:

$$\boldsymbol{a} : \boldsymbol{b} :: \boldsymbol{c} : \boldsymbol{d} \text{ iff } \forall i \in [1, n], a_i : b_i :: c_i : d_i \tag{3}$$

2.2 Nominal Extension

When a, b, c, d take their values in a finite set $\mathcal{D}$ (with more than 2 elements), we can derive three patterns of analogical proportions in the nominal case, from the six possible assignments for analogical proportions in the Boolean case. This generalization is thus defined by:

$$\begin{aligned} &a : b :: c : d = 1 \text{ iff } (a, b, c, d) \in \{(s, s, s, s), (s, t, s, t), (s, s, t, t) | s, t \in \mathcal{D}\} \\ &a : b :: c : d = 0 \textit{ otherwise} \end{aligned} \tag{4}$$

2.3 Multiple-Valued Extension

In case items are described by numerical attributes, it will be necessary to extend the logic modeling underlying analogical proportions in order to support a numerical setting. a, b, c, d are now real values normalized in the interval

$[0,1]$ and their analogical proportion $a : b :: c : d$ is extended from $\mathbb{B}^4$ to $[0,1]^4$. Analogical proportions are no longer valid or invalid but the extent to which they hold is now a matter of degree. For example, if a, b, c, d have 1, 0, 1 and 0.1 as values respectively, we expect that $a : b :: c : d$ has a high value (close to 1) since 0.1 is close to 0.

Table 2. Multi-valued extension

Operator	Extension
Negation: $\neg a$	$1 - a$
Implication: $a \rightarrow b$	$min(1, 1 - a + b)$
Conjunction: $a \wedge b$	$min(a, b)$
Equivalence: $a \equiv b$	$min(a \rightarrow b, b \rightarrow a) = 1 - \lvert a - b \rvert$

The extension of the logical expression of analogical proportions to the multiple-valued case requires the choice of appropriate connectives for preserving desirable properties [5]. To extend expression (2), conjunction, implication and equivalence operators are then replaced by the multiple valued connectives given in Table 2. This leads to the following expression P:

$$P(a,b,c,d) = a : b :: c : d = \begin{cases} 1 - |(a-b) - (c-d)|, \\ \quad \text{if } a \geq b \text{ and } c \geq d \text{ or } a \leq b \text{ and } c \leq d \\ 1 - max(|a-b|, |c-d|), \\ \quad \text{if } a \leq b \text{ and } c \geq d \text{ or } a \geq b \text{ and } c \leq d \end{cases} \tag{5}$$

When a, b, c, d are restricted to $\{0,1\}$, the last expression coincide with the definition for the Boolean case (given by (1)), which highlights the agreement between the extension and the original idea of analogical proportion. For the interval $[0,1]$, we have $P(a,b,c,d) = 1$ as soon as $a - b = c - d$ and as we expected, we get a high value for the 4-tuple $(1,0,1,0.1)$, indeed $1 : 0 :: 1 : 0.1 = 0.9$.

Moreover, since we have $|(1-a)-(1-b)| = |b-a| = |a-b|$, $|(1-a-(1-b)) - (1-c-(1-d))| = |(b-a)-(d-c)| = |(c-d)-(a-b)| = |(a-b)-(c-d)|$, and $1-s \geq 1-t \Leftrightarrow s \leq t$, it is easy to check a remarkable *code independence* property: $a : b :: c : d = (1-a) : (1-b) :: (1-c) : (1-d)$. Code independence means that 0 and 1 play symmetric roles, and it is the same to encode an attribute positively or negatively.

As items are commonly described by vectors, we can extend the notion of analogical proportion to vectors in $[0,1]^n$.

$$P(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}) = \frac{\sum_{i=1}^{n} P(a_i, b_i, c_i, d_i)}{n} \tag{6}$$

where $P(a_i, b_i, c_i, d_i)$ refers to expression (5)).

Let us observe that $P(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{d}) = 1$ (i.e. $\boldsymbol{a} : \boldsymbol{b} :: \boldsymbol{c} : \boldsymbol{d}$ holds) if and only if the analogical proportion holds perfectly on every component:

$$P(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{d}) = 1 \text{ iff } \forall i \in [1,n], P(a_i,b_i,c_i,d_i) = 1 \tag{7}$$

2.4 Inference with Analogical Proportions

Analogical proportion-based inference relies on a simple principle:if four Boolean vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ and $\boldsymbol{d}$ make a valid analogical proportion component-wise between their attribute values, then it is expected that their class labels also make a valid proportion [4].

$$\frac{\boldsymbol{a} : \boldsymbol{b} :: \boldsymbol{c} : \boldsymbol{d}}{cl(\boldsymbol{a}) : cl(\boldsymbol{b}) :: cl(\boldsymbol{c}) : cl(\boldsymbol{d})} \tag{8}$$

where $cl(\boldsymbol{x})$ denotes to the class value of $\boldsymbol{x}$.

It means that the classification of a Boolean vector $\boldsymbol{d}$ is only possible when the equation $cl(\boldsymbol{a}) : cl(\boldsymbol{b}) :: cl(\boldsymbol{c}) : x$ is solvable[1] (the classes of $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ are known as they belong to the sample set), and the analogical proportion $\boldsymbol{a} : \boldsymbol{b} :: \boldsymbol{c} : \boldsymbol{d}$ holds true. If these two criteria are met, we assign x to $cl(\boldsymbol{d})$.

In the numerical case, where $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{d}$ are 4 real-valued vectors over $[0,1]^n$ (the numerical values are previously normalized), the inference principle strictly clones the Boolean setting:

$$\frac{P(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{d}) = 1}{cl(\boldsymbol{a}) : cl(\boldsymbol{b}) :: cl(\boldsymbol{c}) : cl(\boldsymbol{d})} \tag{9}$$

In practice, the resulting degree $P(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{d})$ is rarely equal to 1 but should be close to 1. Therefore Eq. (9) has to be adapted for a proper implementation.

2.5 Continuous Analogical Proportions

Continuous analogical proportions, denoted $a : b :: b : c$, are ternary relations which are a special case of analogical proportions. This enables us to calculate b using a pair (a,c) only, rather than a triple as in the general case. In $\mathbb{B}$ the unique solutions of equations $0 : x :: x : 0$ and $1 : x :: x : 1$ are respectively $x = 0$ and $x = 1$, while $0 : x :: x : 1$ or $1 : x{::}x : 0$ have no solution.

Drawing the parallel with the Boolean case, we deduce that the only solvable equation for the nominal case is $s : x :: x : s$, having $x = s$ as solution, while $s : x :: x : t$ $(s \neq t)$ has no solution.

Contrary to these trivial cases, the multi-valued framework (Eq. (5)) is richer. We have

$$P(a,b,c) = a : b :: b : c = \begin{cases} 1 - |a+c-2b|, \\ \quad \text{if } a \geq b \text{ and } b \geq c \text{ or } a \leq b \text{ and } b \leq c \\ 1 - max(|a-b|, |b-c|), \\ \quad \text{if } a \leq b \text{ and } b \geq c \text{ or } a \geq b \text{ and } b \leq c \end{cases} \tag{10}$$

[1] Indeed the nominal equation $s : t :: t : x = 1$ has no solution if $s \neq t$.

We notice that for $b = (a+c)/2$, we have $a : b :: b : c = 1$ which fits the statement "A is to B as B is to C". As we expect, we get a higher value of analogy (closer to 1) as b tends to $(a+c)/2$. Computing continuous analogy for items described by vectors is exactly the same as for the general case (i.e., for real-valued setting $P(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \frac{\sum_{i=1}^{n} P(a_i, b_i, c_i)}{n}$).

Applying analogy-based inference for numerical values with continuous analogical proportions, we obtain:

$$\frac{P(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = 1}{cl(\boldsymbol{a}) : cl(\boldsymbol{b}) :: cl(\boldsymbol{b}) : cl(\boldsymbol{c})} \tag{11}$$

One may wonder if continuous analogical proportions could be efficient enough compared to general analogical proportions. As already said, $a : b :: c : d$ holds at degree 1 if and only if $a - b = c - d$ (from which one can extrapolate $d = c + b - a$). Now consider two continuous proportions: $a - b = b - c$ (which corresponds to the interpolation $b = (a + c)/2$) and $b - c = c - d$ (which gives the interpolation $c = (b + d)/2$). Adding each side of the two proportions yields $a - c = b - d$, which is equivalent to $a - b = c - d$. In this view, two intertwined interpolations may play the role of an extrapolation. However the above remark applies only to numerical values, but not to Boolean ones.

3 Related Works on Analogical Proportions and Classification

Continuous analogical proportions have been recently applied to enlarge a training set for classification by creating artificial examples [2]. A somewhat related idea can be found in Lieber et al. [6] which extended the paradigm of classical Case-Based Reasoning by either performing a restricted form of interpolation to link the current case to pairs of known cases, or by extrapolation exploiting triples of known cases.

In the classification context, the authors in [3] introduce a measure of oddness with respect to a class that is computed on the basis of pairs made of two nearest neighbors in the same class; this amounts to replace the two neighbors by a fictitious representative of the class. Moreover, some other works have exploited analogical proportions to deal with classification problems. Most noteworthy are those based on using analogical dissimilarity [1] and applied to binary and nominal data and later the analogy-based classifier [4] applied to binary, nominal and numerical data. In the following subsections, we especially review these two latter works as they seem the closest to the approach that we are developing in this paper.

3.1 Classification by Analogical Dissimilarity

Analogical dissimilarity between binary objects is a measure that quantifies how far a 4-tuple (a, b, c, d) is from being in an analogical proportion. This is equivalent to the minimum number of bits to change in a 4-tuple to achieve a perfect analogy, thus when a 4-tuple is in analogical proportion, its analogical dissimilarity is zero. So for the next three examples of 4-tuples, we have $AD(1, 1, 1, 1) = 0$, $AD(0, 1, 1, 1) = 1$ and finally $AD(0, 1, 1, 0) = 2$. In $\mathbb{B}$ the value of an analogical dissimilarity is in $[0, 2]$. When dealing with vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ and $\boldsymbol{d}$ in $\mathbb{B}^m$, analogical dissimilarity is defined as $\sum_{j=1}^{m} AD(a_j, b_j, c_j, d_j)$, in this case an analogical dissimilarity value belongs to the interval $[0, 2m]$.

A classifier based on analogical dissimilarity is proposed in [1]. Given a training set S, and a constant k specifying the number of the least dissimilar triples, the basic algorithm for classifying an instance $\boldsymbol{x} \notin S$ in a naive way, using analogical dissimilarities is as follows:

1. For each triple $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ having a solution for the class equation $cl(\boldsymbol{a}) : cl(\boldsymbol{b}) :: cl(\boldsymbol{c}) : x$, compute the analogical dissimilarity $AD(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{x})$.
2. Sort these triples by ascending order of their analogical dissimilarity $AD(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{x})$.
3. If the k-th triple of the list has the value p, then let the k'-th triple be the last triple of this list with the value p.
4. For the first k'-th triples, solve the class equation and apply a voting strategy on the obtained class labels.
5. Assign to $\boldsymbol{x}$, the winner class.

This procedure may be said naive since it looks for every possible triple from the training set S in order to compute the analogical dissimilarity $AD(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{x})$, therefore it has a complexity of $O(n^3)$, n being the number of instances in the training set. To optimize this procedure, the authors propose the algorithm FADANA which performs an off line pre-processing on the training set in order to speed up on line computation.

3.2 Analogical Proportions-Based Classifier

In a classification problem, objects A, B, C, D are assumed to be represented by vectors of attribute values, denoted $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}$. Based on the previously defined AP inference, analogical classification rely on the idea that, if vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ and $\boldsymbol{d}$ form a valid analogical proportion componentwise for all or for a large number of attributes (i.e., $\boldsymbol{a} : \boldsymbol{b} :: \boldsymbol{c} : \boldsymbol{d}$), this still continue hold for their corresponding class labels. Thus the analogical proportion between classes $cl(\boldsymbol{a}) : cl(\boldsymbol{b}) :: cl(\boldsymbol{c}) : x$ may serve for predicting the unknown class $x = cl(\boldsymbol{d})$ of the new instance $\boldsymbol{d}$ to be classified. This is done on the basis of triples $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ of examples in the sample set that form a valid analogical proportion with $\boldsymbol{d}$.

In a brute force way, AP-classifier proposed in [4], looks for *all* triples $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ in the training set whose class equation $cl(\boldsymbol{a}) : cl(\boldsymbol{b}) :: cl(\boldsymbol{c}) : x$ have a possible solution l . Then, for each of these triples, compute a truth value $P(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d})$ as the average of the truth values obtained in a componentwise manner using Eq. (5) (P can also be computed using the conservative extension, introduced in [5]). Finally, assign to $\boldsymbol{d}$ the class label having the highest value of P.

An optimized algorithm of this brute force procedure has been developed in [4] in which the authors rather search for suitable triples $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ by constraining $\boldsymbol{c}$ to be one of the k nearest neighbours of $\boldsymbol{d}$.

This algorithm processes as follows:

1. Look for each triple $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ in the training set s.t: $\boldsymbol{c} \in N_k(\boldsymbol{d})$.
2. Solve $cl(\boldsymbol{a}) : cl(\boldsymbol{b}) :: cl(\boldsymbol{c}) : x$.
3. If the previous analogical equation on classes has a solution l, increment the credit $credit(l)$ with $P(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d})$ as $credit(l) += P(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d})$.
4. Assign to $\boldsymbol{d}$ the class label having the highest credit as $cl(\boldsymbol{d}) = argmax_l(credit))$.

4 Continuous Analogical Proportions-Based Classifier

Extrapolation and interpolation have been recognized as suitable tools for prediction and classification [6]. Continuous analogical proportions rely on the idea that if three items $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ form a valid analogical proportion $\boldsymbol{a} : \boldsymbol{b} :: \boldsymbol{b} : \boldsymbol{c}$, this may establish the basic for interpolating $\boldsymbol{b}$ in case $\boldsymbol{a}$ and $\boldsymbol{c}$ are known. As introduced in Sect. 2, in the numerical case $\boldsymbol{b}$ can be considered as the *midpoint* of $(\boldsymbol{a}, \boldsymbol{c})$ and may simply be computed from $\boldsymbol{a}$ and $\boldsymbol{c}$.

In this section, we will show how *continuous analogical proportions* may help to develop an new classification algorithm dealing with numerical data and leading to a reduced complexity if compared to the previous Analogical Proportions-based classifiers.

4.1 Basic Procedure

Given a training set $S = \{(\boldsymbol{o_i}, cl(\boldsymbol{o_i})\}$, s.t. the class label $cl(\boldsymbol{o_i})$ is known for each $\boldsymbol{o_i} \in S$, the proposed algorithm aims to classify a new object $\boldsymbol{b} \notin S$ whose label $cl(\boldsymbol{b})$ is unknown. Objects are assumed to be described by numerical attribute values. The main idea is to predict the label $cl(\boldsymbol{b})$ by interpolating labels of other objects in the training set S. Unlike algorithms previously mentioned in Sect. 3, continuous analogical proportions-based interpolation enables us to perform prediction using pairs of examples instead of triples. The basic idea is to find all pairs $(\boldsymbol{a}, \boldsymbol{c}) \in S^2$ with known labels s.t. the equation $cl(\boldsymbol{a}) : x :: x : cl(\boldsymbol{c})$ has a solution l, l being a potential prediction for $cl(\boldsymbol{b})$. If this equation is solvable, we should also check that the continuous analogical proportion holds on each feature j. Indeed we have $\boldsymbol{a} : \boldsymbol{b} :: \boldsymbol{b} : \boldsymbol{c}$ if and only if $\forall j, \quad a_j : b_j :: b_j : c_j$ (i.e., for each feature j, b_j is being the exact midpoint of the pair (a_j, c_j), $b_j = (a_j + c_j)/2$).

As it is frequent to find multiple pairs $(\boldsymbol{a}, \boldsymbol{c})$ which may build a valid continuous analogical proportion with $\boldsymbol{b}$ with different solutions for the equation

$cl(\boldsymbol{a}) : x :: x : cl(\boldsymbol{c})$, it is necessary to set up a *voting* procedure to aggregate the potential labels for $\boldsymbol{b}$. This previous process can be described by the following procedure:

1. Find pairs $(\boldsymbol{a}, \boldsymbol{c})$ such that the equation $cl(\boldsymbol{a}) : x :: x : cl(\boldsymbol{c})$ has a valid solution l.
2. If the continuous analogical proportion $\boldsymbol{a} : \boldsymbol{b} :: \boldsymbol{b} : \boldsymbol{c}$ is also valid, increment the score $ScoreP(l)$ for label l.
3. Assign to $\boldsymbol{b}$ the label l having the highest $ScoreP$.

4.2 Algorithm

As already said, the simplest way is to consider pairs $(\boldsymbol{a}, \boldsymbol{c})$ for which the analogical equation $cl(\boldsymbol{a}) : x :: x : cl(\boldsymbol{c})$ is solvable and the analogical proportion $\boldsymbol{a} : \boldsymbol{b} :: \boldsymbol{b} : \boldsymbol{c}$ is valid.

However, unlike for Boolean features, where $a : b :: b : c$ may hold for many pairs (a, c), it is not really the case for numerical features. In fact, $P(a, b, c) = 1$ does not occur frequently. To deal with such situation in the numerical case, AP-classifiers [4] cumulate individual analogical credits $P(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d})$ to the amount $CreditP(l)$ each time the label l is a solution for the equation $cl(\boldsymbol{a}) : cl(\boldsymbol{b}) :: cl(\boldsymbol{c}) : x$. Even though learning from the entire sample space is often beneficial (in contrast to k-NN principle which is based on a local search during learning), considering all pairs for prediction may seem unreasonable as this could blur the results. Instead of blindly considering all pairs $(\boldsymbol{a}, \boldsymbol{c})$ for prediction, we suggest to adapt the analogical inference, defined by Eq. (9), in such way to consider only pairs $(\boldsymbol{a}, \boldsymbol{c})$ whose analogical score $P(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ exceeds a certain threshold θ.

$$\frac{P(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) > \theta}{cl(\boldsymbol{a}) : cl(\boldsymbol{b}) :: cl(\boldsymbol{b}) : cl(\boldsymbol{c})} \tag{12}$$

This threshold is fixed on an empirical basis. Determining which threshold fits better with each type of dataset is still has to be investigated. The case of unclassified instances may be more likely to happen because of a conflict between multiple classes (i.e., $max(ScoreP)$ is not unique) rather than because of no pairs were found to made a proper classification. That's why we propose to record the best analogical score $bestP(l)$, and even the number of pairs having this best value $vote(l)$ in order to avoid this conflicting situation.

Algorithm 1. CAP-classifier for numerical data

```
Input:a training set $S$, object b $\notin S$, *a threshold* $\theta$
for each label $l$ do $ScoreP(l) = 0$, $bestP(l) = 0$, $vote(l) = 0$ end for
for each pair $(\boldsymbol{a}, \boldsymbol{c}) \in S^2$ do
  if $cl(\boldsymbol{a}) : x :: x : cl(\boldsymbol{c})$ has solution $l$ then
    $p = P(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$
    if $p \geq \theta$ then
      $ScoreP(l) = ScoreP(l) + 1$
    else if $bestP(l) < p$ then
      $bestP(l) = p$
      $vote(l) = 1$
    else if $bestP(l) = p$ then
      $vote(l) = vote(l) + 1$
    end if
  end if
end for
$maxScore = max(ScoreP(l))$
if $unique(maxScore, ScoreP(l))$ then
  return $argmax_l(ScoreP(l))$
else
  $maxBest = max(bestP(l))$
  if $unique(maxBest, bestP(l))$ then
    return $argmax_l(bestP(l)), l \in argmax_l(ScoreP(l))$
  else
    return $argmax_l(vote(l)), l \in argmax_l(bestP(l)), l \in argmax_l(ScoreP(l))$
  end if
end if
```

5 Experimentations and Discussion

In this section, we aim to evaluate the efficiency of the proposed algorithm to classify numerical data. For this aim, we test the CAP-classifier on a variety of datasets from the U.C.I. machine learning repository [7], we provide its experimental results and compare them to the AP-classifier [4] as well as to the state of the art ML classifiers, especially, k-NN, C4.5, JRIP and SVM classifiers.

5.1 Datasets for Experiments

The experimentations are done on datasets from the U.C.I. machine learning repository [7]. Table 3 presents a brief description of the numerical datasets selected for this study. Datasets with numerical attributes must be normalized before testing to fit the multi-valued setting of analogical proportion. A numeric attribute value r is rescaled into the interval $[0, 1]$ as follows:

$$r_{rescaled} = \frac{r - r_{min}}{r_{max} - r_{min}}$$

r_{min} and r_{max} being the maximum and the minimum value of the attribute in the training set. We experiment over the following 9 datasets:

- "Diabetes", "W.B. Cancer", "Heart", "Ionosphere" are binary class datasets.
- "Iris", "Wine", "Sat.Image", "Ecoli" and "Segment" datasets are multiple class problems.

Table 3. Description of numeric datasets

Datasets	Instances	Numerical attrs.	Classes
Diabetes	768	8	2
W. B. Cancer	699	9	2
Heart	270	13	2
Ionosphere	351	34	2
Iris	150	4	3
Wine	178	13	3
Satellite Image	1090	36	6
Ecoli	336	7	8
Segment	1500	19	7

5.2 Testing Protocol

In terms of protocol, we apply a standard 10 fold cross-validation technique. As usual, the final accuracy is obtained by averaging the 10 different accuracies for each fold.

However, we have to tune the parameter θ of the CAP-classifier as well as parameter k for AP-classifier and the ones of the classical classifiers (with which we compare our approach) before performing this cross-validation.

For this end, in each fold we keep only the corresponding training set (i.e. which represents 90% of the full dataset). On this training set, we again perform an inner 10-fold cross-validation with diverse values of the parameter. We then select the parameter value providing the best accuracy. The tuned parameter is then used to perform the initial cross-validation. As expected, these tuned parameters change with the target dataset. To be sure that our results are stable enough, we run each algorithm (with the previous procedure) 5 times so we have 5 different parameter optimizations. The displayed parameter β is the average value over the 5 different values (one for each run). The results shown in Table 4 are the average values obtained from 5 rounds of this complete process.

5.3 Results for CAP-Classifiers

In order to evaluate the efficiency of our algorithm, we compare the average accuracy over five 10-fold CV to the following existing classification approaches:

Table 4. Results of *CAP*-classifier, *AP*-classifier and other ML classifiers obtained with the best parameter β

Datasets	CAP-classifier		AP-classifier		k-NN		C4.5		JRIP		SVM (RBF)		SVM (Poly)	
	acc.	β	acc.	β	acc.	β	acc.	β	acc.	β	acc.	β	acc.	β
Diabetes	72.81	0.906	73.28	11	73.42	11	74.73	0.2	74.63	5	77.37	(8192, 3.051E−5)	77.34	(0.5, 1)
Cancer	96.11	0.825	97.01	4	96.70	3	94.79	0.2	95.87	4	96.74	(2, 2)	96.92	(2, 1)
Heart	81.63	0.693	81.90	10	82.23	11	78.34	0.2	78.52	4	79.98	(32, 0.125)	83.77	(0.5, 1)
Ionosphere	86.44	0.887	90.55	1	88.80	1	89.56	0.1	89.01	5	94.70	(2, 2)	89.28	(0.03125, 2)
iris	95.73	0.913	94.89	5	94.88	3	94.25	0.2	93.65	6	94.13	(32768, 0.5)	96.13	(512, 1)
Wine	96.85	0.832	98.12	9	97.75	7	94.23	0,1	94.99	8	98.20	(32768, 2)	98.53	(2, 1)
Sat image	95.60	0.991	94.96	1	94.88	1	92.71	0.1	92.77	3	96.01	(8, 2)	95.11	(0.5, 4)
Ecoli	86.01	0.93	83.32	7	85.37	5	82.60	0.2	81.56	5	87.50	(2, 8)	87.50	(8, 1)
Segment	96.91	1	96.84	1	96.76	1	95.77	0.2	94.55	6	96.98	(2048, 0.125)	97.14	(8, 4)
Average	89.79		90.10		90.09		88.55		88.39		91.29		91.30	

- **IBk:** implements k-NN, using manhattan distance and the tuned parameter is the number of nearest neighbours during the inner cross-validation with the values $k = 1, 2, ..., 11$.
- **C4.5:** implements a generator of pruned or unpruned C4.5 decision tree. the tuned parameter is the confidence factor used for pruning with the values $C = 0.1, 0.2, ..., 0.5$.
- **JRip:** implements the rule learner RIPPER (Repeated Incremental Pruning to Produce Error Reduction) an optimized version of IREP. The number of optimization runs with the values $O = 2, 4, ..., 10$ is tuned during the inner cross-validation.
- **SVM:** an implementation of the Support Vector Machine classifier. We use SVM with both RBF and polynomial kernels and the tuned parameters are, successively *gamma* for the RBF Kernel, with $\gamma = 2^{15}, 2^{-13}, ..., 2^{3}$ and the degree for the polynomial kernel, $d = 1, 2, ..., 10$. The complexity parameter $C = 2^{-5}, 2^{-3}, ..., 2^{15}$ is also tuned.
- **AP-classifier:** implements the analogical proportions-based classifier with the tuned parameter k with k being the number of nearest neighbours $k = 1, 2, ..., 11$.
- **CAP-classifier:** We test the classifier and we tune the threshold θ with values $\theta = 0.5, 0.6, ..., 1$.

Results for AP-classifier as well as for classic ML classifiers are taken from [4], ML classifiers results are initially obtained by applying the free implementation of Weka software. Table 4 shows these experimental results.

Evaluation of CAP-Classifier and Comparison with Other ML Classifiers: If we analyse the results of CAP-classifier, we can conclude that:

- As expected, the threshold θ of the CAP-classifier change with the target dataset.
- The average θ is approximately equal to 0.89. This proves that CAP-classifier obtains its highest accuracy only if the selected pairs, useful for predicting the class label, are relatively in analogy with the item to be classified.

- For "Iris", "Ecoli", "Sat.Image" and "Segment" datasets, CAP-classifier performs better than AP-classifier, and even slightly better than SVM (polynomial kernel) on the "Sat.Image" dataset, which proves the ability of this classifier to deal with multi-class datasets (up to 8 class labels for these datasets).
- Moreover, we note that for most tested datasets, the optimized θ is close to 1. This fits our first intuition that CAP-classifier performs better when the selected pairs $(\boldsymbol{a}, \boldsymbol{c})$ form a valid continuous analogical proportion with $\boldsymbol{b}$ on *all* (case when $\theta = 1$) or maximum set of attributes (case when $\theta \approx 1$).
- CAP-classifier performs slightly less than AP-classifier for datasets "Diabetes", "Cancer" and "Ionosphere" which are binary classification problems. We may expect that extrapolation, involving triples of examples and thus larger set of the search space is more appropriate for prediction than interpolation using only pairs for such datasets. Identifying the type of data that fits better with each kind of approaches is subject to further instigation.
- For the rest of the datasets, CAP-classifier performs in the same way as the AP-classifier or k-NN. CAP-classifier achieves good results with a variety of datasets regardless the number of attributes (e.g., "Iris" with only 4 attributes, "Sat. image" with 36 attributes).
- As it may be expected, using triples of items for classification is more informative than pairs since more examples are compared against each other in this case. Even though, CAP-classifier performs approximately the same average accuracy as *AP-classifier* exploiting triples $(89,79\% \approx 90,10\%)$ while keeping a lower complexity if compared to classic AP-classifiers. These results highlight the interest of continuous analogical proportions for classification.

Nearest Neighbors Pairs. In this sub-section, we would like to investigate better the characteristics of the pairs used for classification. For this reason, we check if voting pairs $(\boldsymbol{a}, \boldsymbol{c})$ are close or not to the item $\boldsymbol{b}$ to be classified. To do that, we compute the proportion of pairs that are close to $\boldsymbol{b}$ among all voting pairs. If this proportion is rather low, we can conclude that the proposed algorithm is able to correctly classify examples $\boldsymbol{b}$ using pairs $(\boldsymbol{a}, \boldsymbol{c})$ for which $\boldsymbol{b}$ is just the midpoint of $\boldsymbol{a}$ and $\boldsymbol{c}$ without being necessarily in their proximity.

From a practical point, we adopt this strategy:

- Given an item $\boldsymbol{b}$ to be classified.
- Search for the k nearest neighbors $NN = \{n_1, n_2, ...n_k\}$ of $\boldsymbol{b}$. In practice, we consider to test with $k = 5, 10$.
- Compute the percentage of voting pairs $(\boldsymbol{a}, \boldsymbol{c})$ that are among the k nearest neighbors of $\boldsymbol{b}$, i.e. $min(D(\boldsymbol{a}, \boldsymbol{b}), D(\boldsymbol{b}, \boldsymbol{c})) \leq D(n_k, \boldsymbol{b})$, $D(x, y)$ being the distance between items x and y. If this percentage is low, it means that even if voting pairs $(\boldsymbol{a}, \boldsymbol{c})$ remain far to the item $\boldsymbol{b}$, the proposed interpolation-based approach succeeds to guess the correct label for $\boldsymbol{b}$.

The results are shown in Table 5. In this supplementary experiment, we only consider testing examples whose voting pairs $(\boldsymbol{a}, \boldsymbol{c})$ have a continuous analogical proportion $P(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ exceeding the threshold θ (see last column in Table 5).

From these results we can note:

- For $k = 5$ (first column), the proportion of pairs $(\boldsymbol{a}, \boldsymbol{c})$ (among those exceeding the threshold) that are in the neighborhood of $\boldsymbol{b}$ (those $(\boldsymbol{a}, \boldsymbol{c})$ that are closest to $\boldsymbol{b}$ than its neighbor n_5) is less than 10% for all tested datasets except for "Wine" which is little higher. This demonstrates that for these datasets, the CAP-classifier exploits the entire space of pairs for prediction, indeed most of examples are predicted thanks to pairs $(\boldsymbol{a}, \boldsymbol{c})$ that are located outside of the neighborhood of $\boldsymbol{b}$.
- Even when the number of nearest neighbors k is extended to 10, this proportion remains low for most of the datasets. Especially for "Diabetes" and "Ecoli", the percentage of pairs in the neighborhood of $\boldsymbol{b}$ is close to 5%. For other datasets, this percentage is less than 20%.
- Note that the behavior of our algorithm is quite different from the k-NN classifier. While this latter computes the similarity between the example $\boldsymbol{b}$ to be classified and those in the training set, then classifies this example in the same way as its closest neighbors, our algorithm evaluates to what extent $\boldsymbol{b}$ is in continuous analogy with the pairs in the training set (these pairs are not necessarily in the proximity), then classifies it as the winning class having the highest number of voting pairs.
- These last results show that voters $(\boldsymbol{a}, \boldsymbol{c})$ remain far from to the item $\boldsymbol{b}$ to be classified.

Table 5. Proportion of pairs $(\boldsymbol{a}, \boldsymbol{c})$ that are nearest neighbors to $\boldsymbol{b}$

Datasets	% of pairs $(\boldsymbol{a}, \boldsymbol{c})$ that are among the 5 neighbors of $\boldsymbol{b}$	% of pairs $(\boldsymbol{a}, \boldsymbol{c})$ that are among the 10 neighbors of $\boldsymbol{b}$	% of examples $\boldsymbol{b}$ for which $P(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) > \theta$
Diabetes	4.03%	5.98%	80.42%
Cancer	5.35%	8.29%	94.32%
Heart	6.85%	9.01%	95.04%
Ionosphere	5.53%	11.60%	63.17%
Iris	8.19%	14.67%	94.13%
Wine	14.65%	18.78%	87.85%
Ecoli	4.55%	6.88%	90.03%

6 Conclusion

This paper studies the ability of continuous analogical proportions, namely statements of the form *a is to b as b is to c*, to classify numerical data and presents a classification algorithm for this end. The basic idea of the proposed approach is to search for all pairs of items, in the training set, that build a continuous analogical proportion on *all* or most of the features with the item to be classified. An analogical value is computed for each of these pairs and only those pairs whose score exceeds a given threshold are kept and used for prediction. In case no such pairs could be found for each class label, the best pair having the highest analogical value is rather used. Finally, the class label with the best score is assigned to the example to be classified. Experimental results show the interest of the *CAP*-classifier for classifying numerical data. In particular the proposed algorithm may slightly outperform some state-of-the-art ML algorithms (such as: *k*-NN, C4.5 and JRIP), as well as the *AP*-classifier on some datasets. This leads to conclude that for classification, building analogical proportions with three objects (using continuous analogical proportions) instead of four enables to get an overall average accuracy close to that of previous *AP*-classifier while reducing the complexity to be quadratic instead of being cubic.

References

1. Bayoudh, S., Miclet, L., Delhay, A.: Learning by analogy: a classification rule for binary and nominal data. In: Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI 2007, Hyderabad, India, 6–12 January, pp. 678–683 (2007)
2. Bounhas, M., Prade, H.: An analogical interpolation method for enlarging a training dataset. In: Ben Amor, N., Quost, B., Theobald, M. (eds.) SUM 2019. LNCS (LNAI), vol. 11940, pp. 136–152. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-35514-2_11
3. Bounhas, M., Prade, H., Richard, G.: Oddness-based classification: a new way of exploiting neighbors. Int. J. Intell. Syst. **33**(12), 2379–2401 (2018)
4. Bounhas, M., Prade, H., Richard, G.: Analogy-based classifiers for nominal or numerical data. Int. J. Approx. Reason. **91**, 36–55 (2017)
5. Dubois, D., Prade, H., Richard, G.: Multiple-valued extensions of analogical proportions. Fuzzy Sets Syst. **292**, 193–202 (2016)
6. Lieber, J., Nauer, E., Prade, H., Richard, G.: Making the best of cases by approximation, interpolation and extrapolation. In: Cox, M.T., Funk, P., Begum, S. (eds.) ICCBR 2018. LNCS (LNAI), vol. 11156, pp. 580–596. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01081-2_38
7. Metz, J., Murphy, P.M.: UCI Repository (2000). ftp://ftp.ics.uci.edu/pub/machine-learning-databases
8. Miclet, L., Bayoudh, S., Delhay, A.: Analogical dissimilarity: definition, algorithms and two experiments in machine learning. J. Artif. Intell. Res. **32**, 793–824 (2008)
9. Miclet, L., Bayoudh, S., Delhay, A., Mouchére, H.: De l'utilisation de la proportion analogique en apprentissage artificiel. In: Actes des Journées Intelligence Artificielle Fondamentale, IAF 2007, Grenoble, 2–3 July 2007 (2007). http://www.cril.univ-artois.fr/konieczny/IAF07/
10. Miclet, L., Prade, H.: Handling analogical proportions in classical logic and fuzzy logics settings. In: Sossai, C., Chemello, G. (eds.) ECSQARU 2009. LNCS (LNAI), vol. 5590, pp. 638–650. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02906-6_55
11. Prade, H., Richard, G.: Analogical proportions: another logical view. In: Bramer, M., Ellis, R., Petridis, M. (eds.) Research and Development in Intelligent Systems, pp. 121–134. Springer, London (2010). https://doi.org/10.1007/978-1-84882-983-1_9
12. Prade, H., Richard, G.: Analogical proportions: from equality to inequality. Int. J. Approx. Reason. **101**, 234–254 (2018)