International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 4 (2012) pp. 648–658

Enforcing regularity by means of analogy-related proportions – A new approach to classification

Henri Prade[1] **Gilles Richard**[2] **Bing Yao**[1]

[1]IRIT, University of Toulouse
31062 Toulouse Cedex 9, France
prade@irit.fr

[2]British Institute of Technology & E-commerce
London, E7 9HZ, UK
grichard@bite.ac.uk

Abstract: **When classifying a new item, it seems reasonable to enforce a regularity principle stating, informally speaking, that the way the new item is classified should not introduce discrepancy with respect to the already classified items. In this paper, we propose to estimate this regularity or lack of discrepancy, by means of analogical proportions. An analogical proportion is a statement of the form "a is to b as c is to d", which may receive a Boolean logic interpretation, or more generally a multiple-valued logic interpretation. Two other related formal proportions named reverse analogy ("what a is to b is the reverse of what c is to d"), and paralogy ("what a and b have in common, c and d have it also") may be also used. These proportions relate items a, b, c, and d on the basis of their differences, or of their similarities. It provides a basis for proposing a plausible classification for an object d described in terms of a set of features, by considering three other already classified objects, say a, b, c, described with the same features, assuming that if some proportion holds for a sufficiently large number of features, it may hold on the allocation of the classes as well. This is the basis of a classification method which is tested on machine learning benchmarks for binary or multiple class problems with objects that have numerical features.** ***Keywords***: analogical proportion; multiple-valued logic; classification

I. Introduction

Analogical reasoning [11, 17] is generally viewed as a useful heuristic way to enhance reasoning processes in problem solving by transferring conclusions observed in a known situation to a new situation, which is only partially known, but which can be paralleled with the previous one. For instance, a problem $prob_1$ has a known solution sol_1, while one looks for a solution sol_2 for problem $prob_2$. Then, one may assume that sol_2 *is to* sol_1 *as* $prob_2$ *is to* $prob_1$ on the basis of some similarity or partial identity between $prob_1$ and $prob_2$. In the same way, an object 1 with description des_1 is in class cl_1, how to classify an object 2 with description des_2 close to des_1? These questions are at the basis of case-based reasoning [1], or of k-nearest neighbors methods [6, 14] for transduction of a class for the new item, on the basis of a repertory of known cases or examples. Then, analogy is understood as the identity or similarity of relations or properties between two ordered pairs, (a, b) and (c, d), as expressed in analogical proportions, which are statements of the form a *is to* b *as* c *is to* d, usually denoted $a : b :: c : d$. Thus, one may write for instance $cl_2 : cl_1 :: des_2 : des_1$ (also assuming $cl_2 : des_2 :: cl_1 : des_1$).

However, a more powerful use of analogical proportions has been recently envisaged by some authors [35, 34, 4, 18], which amounts to state $(des_2, cl_2) : (des_1, cl_1) :: (des_{1'}, cl_{1'}) : (des_{1''}, cl_{1''})$, where the object 2 to be classified is simultaneously paralleled with three other already classified objects 1, $1'$, and $1''$ rather than one. Intuitively, it has the merit of relating the change or the absence of change, in classification to the change or the absence of change in the descriptions of various objects based on different features. Indeed, it embeds a form of interpolation as in the following example [23, 20] where we try to estimate if a house h_2 is cheap (ch) or not ($\neg ch$) on the basis of the comparison with three other houses h_1, $h_{1'}$, and $h_{1''}$ that have equipment 1 (eq_1) or equipment 2 (eq_2), or not. Thus each house is described here in terms of three features: presence of equipment 1 (yes or no), presence of equipment 2 (yes or no), price (cheap or not cheap). Then, if we build the analogical proportion $h_1 : h_{1'} :: h_{1''} : h_2$, i.e., $(eq_1, eq_2, \neg ch) : (eq_1, \neg eq_2, ch) :: (\neg eq_1, eq_2, \neg ch) : (\neg eq_1, \neg eq_2, ?)$, one may expect to conclude that the '?' should be replaced by ch. Note that here h_2 is not identical to any of the three other houses.

A key issue is then to evaluate if an analogical proportion $a : b :: c : d$ holds or not, or holds to some degree, in case of binary, or more generally of multiple-valued features. Thus, a measure of analogical dissimilarity has been proposed [4, 18]. More recently, a propositional logic modeling of analogical proportions has been developed [20, 21], and extended to fuzzy logic for handling multiple-valued features [26]. Moreover, it has been also pointed out that two other types of proportion exist alongside analogical proportion [25, 24], namely i) *reverse analogy* which focuses on the differences between the items at hand as analogy does, but reverses the directions of changes, and the so-called ii) *paralogy*, denoted $a ; b :: c ; d$ which rather focuses on common properties and similarities, i.e. $a ; b :: c ; d$ holds if the (positive or negative) features common to a and b are also common to c and d.

In this paper, we present and discuss a classification procedure, first outlined in [27], entirely based on the use of analogical proportions (or of paralogical proportions), applied in a component-wise manner to vectors $(v_1, ..., v_i, ..., v_n, v_{n+1})$ describing features of objects together with their class v_{n+1}, where each component v_i for $i = 1, n$ encodes if a binary property P_i holds or not, or the extent to which it holds if P_i is gradual, for the item at hand. Intuitively speaking, the idea is that applying an analogy-related proportion on the class attribute to a 4-tuple made of three classified items and a new item, should echo the similarities and differences between the descriptions of these items, provided that the same proportion holds for most of the features used in these descriptions. In other words, the proportion will enforce some regularity in prescribing a change or an identity of class (depending on the situation). Obviously, our method will be sensitive to the quality of the set of examples and it could be useful to deal with "abnormal" cases before starting the classification process (see [3] for a comprehensive survey of what can be done).

The remaining of this article is organized as follows. The next section provides the necessary background on the logical modeling of the three analogy-related proportions, and their fuzzy (i.e., graded) extensions. Section III introduces the classification procedure, while Section IV reports results obtained on binary classes, or multiple classes benchmark problems with numerical features, and comments these results. In Section V, we report some other tests aiming at a comparison with the well-known k-NN approach. It appears that our results are not generally based on a use of the nearest neighbors of the target data to be classified. Section VI discusses the approach with respect to related works, while Section VII contrasts it with respect to other classification methods, leading to some new perspective on transduction (i.e., classification without inductive step). This paper is a revised and extended version of a conference paper [30].

II. Logical modeling of analogy-related proportions

An analogical proportion a is to b as c is to d is viewed in the following as a quaternary logical connective, and is shown to be closely related to two other logical proportions [27], called 'reverse analogy' and 'paralogy' respectively. A complete introduction to these proportions (and other related ones) can be found in [29].

A. 3 types of proportions

Analogy, reverse analogy and paralogy are formal relations involving 4 items a, b, c and d. A detailed investigation of their underlying semantics has been done in [25, 24]. They all satisfy symmetry when comparing a, b with c, d, but obey distinct characteristic postulates (e.g. 'central permutation' for the analogical proportion that requires the equivalence between $a : b :: c : d$ and $a : c :: b : d$, while the two other proportions are stable under two other permutations). To briefly introduce the three proportions, the most suitable starting point is to consider our items a, b, c, d as described by sets of binary features belonging to a universe X. In that case, each item can be viewed as a subset of X. An analogical proportion $a : b :: c : d$ focuses on *differences* and should hold when *the differences between* a *and* b *and between* c *and* d *are the same*. This corresponds to the condition [21] ($a \setminus b$ denotes the set difference $a \cap \overline{b}$):

$$a \setminus b = c \setminus d \text{ and } b \setminus a = d \setminus c \quad (A_{Set}) \qquad (1)$$

Reverse analogy denoted $a \,!\, b :: c \,!\, d$ amounts to exchange c and d, and expresses that the changes from a to b and from c to d (if any) are now in opposite directions:

$$a \setminus b = d \setminus c \text{ and } b \setminus a = c \setminus d \quad (R_{Set}) \qquad (2)$$

If instead of differences, we focus on *similarities*, we get a new proportion $a ; b :: c ; d$ called *paralogy*, which expresses that what a and b have in common, c and d have it also:

$$a \cap b = c \cap d \text{ and } \overline{a} \cap \overline{b} = \overline{c} \cap \overline{d} \quad (P_{Set}) \qquad (3)$$

B. Boolean model

A direct way to switch to the Boolean lattice $\mathbb{B} = \{0, 1\}$ is just to translate the set operators into Boolean connectors. Taking the complement of the set difference $\setminus$ yields the impli-cation $\rightarrow$, while $=$ is changed into $\equiv$, $\cup$ into $\vee$, $\cap$ into $\wedge$:

$$((a \rightarrow b) \equiv (c \rightarrow d)) \wedge ((b \rightarrow a) \equiv (d \rightarrow c)) \; (A_{Bool}) \quad (4)$$

$$((a \rightarrow b) \equiv (d \rightarrow c)) \wedge ((b \rightarrow a) \equiv (c \rightarrow d)) \; (R_{Bool}) \quad (5)$$

$$(a \wedge b \equiv c \wedge d) \wedge (a \vee b \equiv c \vee d) \; (P_{Bool}) \qquad (6)$$

In this paper, for sake of simplicity, we use the same notations for a proposition and its truth value, and for a connective and its associated truth function.

A simple, visual way to understand the difference between analogy, reverse analogy and paralogy is to examine their truth tables when they are considered as Boolean operators. Table 1 provides the 6 cases where each of the above three quaternary connectives holds true (among 16 possibilities). Note that the tables remain the same when exchanging 0 and 1 everywhere. This expresses that these proportions do not depend on a positive or a negative encoding of properties. The Boolean interpretation makes clear that $a : b :: b : a$ is false for analogy, $a \,!\, b :: a \,!\, b$ false for reverse analogy, and $a ; a :: b ; b$ false for paralogy, in agreement with their intuitive meaning. Note that the three proportions validate the pattern $a : a :: a : a$ and two of the above patterns (e.g., for analogy, $a : a :: b : b$ and $a : b :: a : b$ hold true, which respectively acknowledge the equivalence of identities and of changes having the same direction).

The following property, which holds in the Boolean setting (and thus in the set-based interpretation we start with), is easy to check on the truth tables, and establishes a strong link between the three kinds of proportion:

Table 1: Boolean model

Analogy				Reverse Ana.				Paralogy			
0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1
0	0	1	1	0	0	1	1	1	0	0	1
1	1	0	0	1	1	0	0	0	1	1	0
0	1	0	1	0	1	1	0	0	1	0	1
1	0	1	0	1	0	0	1	1	0	1	0

Property 1
- *$a \,!\, b :: c \,!\, d$ is a reverse analogy if and only if $a : b :: d : c$ is an analogy;*
- *$a \,;\, b :: c \,;\, d$ is a paralogy if and only if $a : d :: c : b$ is an analogy.*

The equations $a : b :: c : x$, $a \,!\, b :: c \,!\, x$, and $a \,;\, b :: c \,;\, x$ have not always a solution $x \in \{0, 1\}$ such that the corresponding proportion holds. For analogical proportions, the existence condition for a solution is $(a \equiv b) \vee (a \equiv c) = 1$ (which just states that $1 : 0 :: 0 : x$ and $0 : 1 :: 1 : x$ have no solutions). When a solution exists, it is unique and given by $x = a \equiv (b \equiv c)$ for the three proportions [21, 25], as first guessed from anthropological observations by Klein [15] (without distinguishing the three proportions).

It is worth noticing that the above logical view of analogy departs from more established views [13], which are more oriented towards analogy-making than towards analogy-checking (which is our primary interest in this paper).

C. Multiple-valued models

If we consider the Boolean expression of the analogical proportion given by formula (4), one may think of many possible multiple-valued extensions, depending on the operations chosen for modeling $\wedge$, $\equiv$, and $\to$. Moreover, a formula such as (4) can be written in many equivalent forms in Boolean logic. These forms are no longer necessarily equivalent in a non-Boolean setting (we use $[0, 1]$ as truth space). So it is important to make proper choices that are in agreement with the intended meaning of the considered proportion. Some properties seem very natural to preserve, such as
i) the independence with respect to the positive or negative encoding of properties (one may describe a price as the extent to which it is cheap, as well as it is not cheap), which leads to require that $\neg a : \neg b :: \neg c : \neg d$ holds if $a : b :: c : d$ holds (with $\neg a = 1 - a$);
ii) the knowledge of a and of the differences between a and b and between b and a, should enable us to recover b. Indeed in the Boolean case, we have

$$b = (a \wedge (a \to b)) \vee \neg(b \to a)$$

A careful analysis [26] of the requirements leads to choose

- i) the minimum operator for $\wedge$;
- ii) $s \equiv t = 1 - |s - t|$;
- iii) Lukasiewicz implication $s \to t = min(1, 1-s+t)$.

Note also that with these choices $s \equiv t = (s \to t) \wedge (t \to s)$.

This leads to the following expressions which both generalize the Boolean case to multiple-valued entries and introduce a graded view of the analogy-related proportions.

For analogy, we have
$a : b :: c : d =$
$1- \mid (a-b)-(c-d) \mid$ if $a \geq b \;\&\; c \geq d$, or $a \leq b \;\&\; c \leq d$
$1 - max(|a-b|, |c-d|)$ if $a \leq b \;\&\; c \geq d$ or $a \geq b \;\&\; c \leq d$
Thus, $a : b :: c : d$ is all the closer to 1 as the differences $(a - b)$ and $(c - d)$ have the same sign and have similar absolute values. Note that $1 : 0 :: c : d = 0$ as soon as $c \leq d$.

For reverse analogy, we have
$a \,!\, b :: c \,!\, d =$
$1- \mid (a-b)-(d-c) \mid$ if $a \leq b \;\&\; c \geq d$ or $a \geq b \;\&\; c \leq d$
$1 - max(|a-b|, |c-d|)$ if $a \geq b \;\&\; c \geq d$, or $a \leq b \;\&\; c \leq d$

The definition of *paralogy* does not involve any implication: it is built upon $\wedge$ and $\neg$ only (plus the operator $\equiv$):
$a;b::c;d = min(1-|(a \wedge b)-(c \wedge d)|, 1-|(a \vee b)-(c \vee d)|)$, with $a \vee b = 1 - (1 - a) \wedge (1 - b)$. Again we take $a \wedge b = min(a, b)$; see [26] for justifications.

However, with the above definitions, the relation between paralogy and analogy expressed by Property 1 is no longer true in the multiple-valued case.
With respect to equation solving, it can be shown that it exists x such that $a : b :: c : x = 1$ if and only if $x = c + b - a \in [0, 1]$, and when it exists, the solution is unique. Similar equations may be solved as well for the two other proportions. However, in the following the issue of solving equations will take place only in the Boolean setting for predicting class (a problem considered here as being of *binary* nature), while the expression of multiple-valued proportions will be used only for estimating to what extent the proportion holds for the non-binary features involved in the description of the items.

III. Classification as transduction

We consider a classification task, where each element of the problem universe is represented by a vector of n binary, or multiple-valued, features. Our input space is taken as $X = [0, 1]^n$. Thus, any numerical feature f_i (e.g. a price) should be normalized (as $(f_i - min\{f_i\})/(max\{f_i\} - min\{f_i\})$), while binary features are encoded in $\{0, 1\}$. Many-valued discrete features on ordered domains could be also consistently accommodated by the multiple-valued connectives we use. Moreover, each element of X (representing one or more items of our universe) belongs to a class $c \in \mathcal{C}$, with $|\mathcal{C}|$ finite and ≥ 2. Then, we try to classify an incoming item $x \in X$ in the output space $\mathcal{C}$. A set S of already classified items is supposed to be given, which is then just a subset of $X \times \mathcal{C}$, i.e., a set of labeled examples $(s, cl(s))$. In the following, it is assumed that the class of an element is unique. This means that $\nexists(x, c)$ and $(x, c') \in S$, with $c \neq c'$. This agrees with the hypothesis of the existence of an underlying classifying function cl from X to $\mathcal{C}$, this function being only known for the elements in S.

A. General principle

The basic idea underlying transduction [32, 10] is to try to predict the class of a new piece of data on the basis of the previously observed data S, without any attempt at inducing a generic model for the observed data (which would be then applied to the new piece of data in order to determine its class).
A simple example of transduction mechanism is the k-Nearest Neighbors method, where the class that is the most frequent among the k closest neighbors of x is attributed to x, i.e., when $k = 1$, $cl(x) = cl(argmin_{s|(s,cl(s))\in S} \mid dis(x,s))$, where dis is a distance. In the absence of external knowledge, the only thing that we have at our disposal for predicting the class of a new element x, beyond the simple exploitation of closeness relations, is the *observation of the behavior* of the data at hand, trying to relate the *variations of* $cl(x)$ *with the variation of the features describing* x. Since analogy-based proportions estimate the equivalence of two pairs of items on the basis of the similarities and/or the differences observed inside each pair, this constitutes a candidate tool for trying to take into account these variations and relating them to classification.
The idea is that when an analogy-related proportion holds for some observable features describing items a, b, c, and d, we can assume that the same proportion still holds between the other features of d, and the corresponding, observable and known features of a, b, and c. Namely, assuming that the items we are dealing with are vectors of the form $a = (a_1, \dots, a_n, a_{n+1}, \dots, a_{n+m})$ of truth values that encode the values of $n+m$ features associated to an item. Consider for the moment the case of Boolean truth values. Starting from a 4-tuple of vectors (a, b, c, d), we consider that if there is an analogy-related proportion T that holds between the first n components of these vectors, then this proportion should hold for the last m remaining components as well. This inference principle can be stated as below:

$$\frac{\forall i \in [1, n],\ \ T(a_i, b_i, c_i, d_i)}{\forall j \in [n+1, n+m],\ \ T(a_j, b_j, c_j, d_j)}$$

This is a generalized form of analogical reasoning, where we transfer knowledge from some components of our vectors to their remaining components. The fact that the proportions are independent w.r.t. the positive or negative encoding of the features guarantees that the same result is obtained whatever the encoding. The above pattern can be easily extended to graded truth values, by considering that if the proportion T holds at degree 1 (or more generally sufficiently holds) for the n first features, it should also hold in the same way for the m other features.

This idea, when applied to classification, provides the basis for a transduction mechanism, already discussed in [27] for binary classification and binary-valued features. In our approach, given a new instance d whose class is unknown, one looks for triples (a, b, c) of examples in the input space S such that some logical proportion T holds simultaneously between the n binary description features associated to (a, b, c, d) (i.e. T holds on the input space corresponding to the n first features in the above pattern) in order to predict the class of d by applying the same proportion T to the output space $\mathcal{C}$, i.e. the space of classes. Through the computation of the proportions, the approach focuses on the common features where the pairs (a, b) and (c, d) differ, and put it in relation with possible classification changes. See [27] for a detailed discussion in the binary case. Note that the approach includes the limit case where we have to classify an item whose representation (in the input space) is completely similar to an already classified item (i.e. $a_i = b_i = c_i = d_i$ for $i = 1, \dots, n$).

The simplest way for describing the output space is to use a description based on m binary-valued features if there are m possible classes. Then, given three items a, b, c different situations may be encountered regarding the way they are classified:

- $cl(a) = cl(b) = cl(c) = cl$. Thus $\exists! k \in \{1, \dots, m\}$ $a_{n+k} = b_{n+k} = c_{n+k} = 1$ and $\forall \ell \neq k, a_{n+\ell} = b_{n+\ell} = c_{n+\ell} = 0$. Then, by a straightforward application of any of the three analogy-related proportions we get $d_{n+k} = 1$ and $\forall \ell \neq k, d_{n+\ell} = 0$, i.e. $cl(d) = cl$.
- $cl(a) = cl(b) = cl$ and $cl(c) = cl'$. Thus $\exists! k \in \{1, \dots, m\}, \exists! k' \in \{1, \dots, m\}$ $a_{n+k} = b_{n+k} = 1$ and $\forall \ell \neq k, a_{n+\ell} = b_{n+\ell} = 0$, while $c_{n+k'} = 1$ and $\forall \ell \neq k', c_{n+\ell} = 0$. Then, we apply analogy or reverse analogy (provided that one of these two proportions holds for all the n description features), using the patterns $1 : /! \ 1 :: 0 : /! \ 0$[1] and $0 : /! \ 0 :: 1 : /! \ 1$ for the $(n + k)$th and the $(n + k')$th output features respectively, and $0 : /! \ 0 :: 0 : /! \ 0$ for the others. Thus we obtain $cl(d) = cl'$.

- $cl(a) = cl(c) = cl$ and $cl(b) = cl'$. Thus $\exists! k \in \{1, \dots, m\}, \exists! k' \in \{1, \dots, m\}$ $a_{n+k} = c_{n+k} = 1$ and $\forall \ell \neq k, a_{n+\ell} = c_{n+\ell} = 0$, while $b_{n+k'} = 1$ and $\forall \ell \neq k', b_{n+\ell} = 0$. Then, we apply analogy or paralogy (provided that one of these two proportions holds for all the n description features), using the patterns $1 : /; \ 0 :: 1 : /; \ 0$ and $0 : /; \ 1 :: 0 : /; \ 1$ for the $(n + k)$th and the $(n + k')$th output features respectively, and $0 : /; \ 0 :: 0 : /; \ 0$ for the others. Thus we obtain $cl(d) = cl'$.

- $cl(b) = cl(c) = cl$ and $cl(a) = cl'$. Thus $\exists! k \in \{1, \dots, m\}, \exists! k' \in \{1, \dots, m\}$ $b_{n+k} = c_{n+k} = 1$ and $\forall \ell \neq k, b_{n+\ell} = c_{n+\ell} = 0$, while $a_{n+k'} = 1$ and $\forall \ell \neq k', a_{n+\ell} = 0$. Then, we apply reverse analogy or paralogy (provided that one of these two proportions holds for all the n description features), using the patterns $0 \ !/; \ 1 :: 1 \ !/; \ 0$ and $1 \ !/; \ 0 :: 0 \ !/; \ 1$ for the $(n+k)$th and the $(n+k')$th output features respectively, and $0 \ !/; \ 0 :: 0 \ !/; \ 0$ for the others. Thus we obtain $cl(d) = cl'$.
In fact, the above procedure amounts to apply the following patterns

$$cl : /!/; \ cl :: cl : /!/; \ cl^2$$
$$cl : /! \ cl :: cl' : /! \ cl',$$
$$cl : /; \ cl' :: cl : /; \ cl',$$
$$cl \ !/; \ cl' :: cl' \ !/; \ cl,$$

[1] A compact notation for $1 : 1 :: 0 : 0$ or $1 \ ! \ 1 :: 0 \ ! \ 0$. The same kind of notation with obvious meaning is used in the following.

[2] Again a compact notation for $cl : cl :: cl : cl$ or $cl \ ! \ cl :: cl \ ! \ cl$ or $cl \ ; \ cl :: cl \ ; \ cl$.

to a unique multiple-valued feature corresponding to the identifier of the class.

Now consider the situation with three different classes $cl(a) = cl, cl(b) = cl', cl(c) = cl''$, i.e. $\exists! k \in \{1, \ldots, m\}, \exists! k' \in \{1, \ldots, m\}, \exists! k'' \in \{1, \ldots, m\}$ $a_{n+k} = 1$ and $\forall \ell \neq k, a_{n+\ell} = 0$, $b_{n+k'} = 1$ and $\forall \ell \neq k', b_{n+\ell} = 0$, and $c_{n+k"} = 1$ and $\forall \ell \neq k", c_{n+\ell} = 0$. It clearly leads to no classification solution, since the patterns $(1, 0, 0, x)$, $(0, 1, 0, x)$, $(0, 0, 1, x)$ cannot be simultaneously solved for a unique proportion. Moreover, the result $d_{n+k} = 1, d_{n+k'} = 1, d_{n+k"} = 1, d_{n+\ell} = 0, \forall \ell \neq k, k', k"$ which might be obtained with a mixed use of proportions, would not lead to any classification.[3]

B. An analogy-based transduction procedure

In the proposed classification method, we apply a strategy that is very different from k-NN methods. Indeed, the new item d to be classified is not just compared with classified items on a one-by-one basis. For the sake of uniformity, we continue to denote d the new item to be classified: its description d is known, but its class $cl(d)$ is unknown. First, we look for 3-tuples $(a, b, c) \in S^3$ such that the class equation $T(cl(a), cl(b), cl(c), x) = 1$ has a solution with respect to some analogy-related proportion T. This requires that $cl(a)$, $cl(b)$, and $cl(c)$ correspond to a *maximum of two* distinct classes, as explained in the previous subsection. Then either $cl(d) = cl(a) = cl(b) = cl(c)$, or there are two distinct classes, which we may encode by 1 and 0. Then, the corresponding class equation $T(cl(a), cl(b), cl(c), x) = 1$ should have a solution for the considered type T of proportion, in order the triple (a, b, c) be retained as potentially useful. Obviously the other triples (a, b, c) are useless for our objective because, whatever the coming d, they cannot constitute a logical proportion with d. This processing of the suitable set of triples can be done offline.

Choosing a fixed proportion T, when an item d has to be classified, we have to look, among the set of suitable triples, for the one(s) that seem(s) the most appropriate to predict the class $cl(d)$. For doing this, each suitable triple we consider is evaluated by means of the following vector $(T(a_1, b_1, c_1, d_1), \ldots, T(a_i, b_i, c_i, d_i), \ldots, T(a_n, b_n, c_n, d_n))$. Then the vectors (and thus the triples) are ordered in a lexicographic decreasing order[4]. Then we may choose for $cl(d)$ the class associated to the triple having the best evaluation, or the most frequent class among the k best triples. In this latter case, one might also consider the different classes as possible solutions, between which there is not enough information for making a proper choice.

Note that it should not be considered as a problem if some $T(a_i, b_i, c_i, d_i)$ are close to 0: indeed it should not be required that the proportion holds for *all* features (even approximately), since some features may turn to be irrelevant for the classification and have then no reason to exhibit any regularity with respect to it.

Several remarks may lead to some potential improvements of the basic procedure described above.

1. The evaluations $T(a_i, b_i, c_i, d_i)$ may be rounded to insure that vectors are compared on the basis of significant differences.
2. In case we use the k best evaluations for predicting the class, with $k > 1$, it would make sense that the k corresponding evaluation vectors remain sufficiently close, otherwise one may doubt of the meaningfulness of the triples having the poorest evaluations among the k best ones.
3. However, if there are not enough components in the best evaluation vector(s) that are close to 1, one may doubt that we have sufficient material among the available triples for predicting the class of d in a meaningful way.
4. The lexicographic ordering presupposes that all the features have the same importance. This may not be the case if some are less relevant (although the considered proportion holds for these features). Then one may think of using the discrimin ordering [7] instead of the leximin ordering (the components of each evaluation vector are no longer decreasingly reordered, and the comparison between two vectors is made on the basis of the minimum applied to the components where the two vectors are not equal). This is only a partial order that refines the Pareto ordering, and one may choose the most frequent class associated with the non dominated triples.

IV. Experimental Results

In order to validate the basic procedure (we have not tested any of the 4 types of refinements suggested above yet), we have tested the previous ideas on different data sets coming from the University of California, Irvine (UCI) machine learning repository [9]. We have only chosen classification problems where attributes are mainly quantitative (they may be defined on continuous universes, but for some of them (age for instance) the attribute domain is discrete). The number of attributes varies from 4 to 57. Some of the datasets (Iris and Pima for instance) have been extensively investigated, which allows us to have a clear picture of the performance of other approaches. Obviously, we have to normalize the attribute values to get numbers in $[0, 1]$ and to apply the method described at the beginning of Section III. We consider the normalized value of a given attribute a as its fuzzy truth value: we are then back to the theoretical framework previously described. We have implemented a Java program, without any rounding of the numbers (see remark 1 of the previous section). At this stage, we have used a brute force algorithm, which due to limited computation facilities, has somewhat limited the experimentations (w.r.t. the size of the samples). The protocol used is rather straightforward:

[3] However, an ordered, structured way of the set of classes may offer more possibilities. Assume for instance, we have four classes $very\ good$, $rather\ good$, $rather\ bad$, $very\ bad$, with the additional information that the distances between two successive labels in this ordered list are the same, one may then consider that an equation such that, e.g., $very\ good : rather\ bad :: rather\ good : x$ has $x = very\ bad$ as a solution from an analogy point of view. See [31] for a preliminary study of such an interpolation mechanism.

[4] $(u_1, \ldots, u_i, \ldots, u_n) >_{lexicographic} (v_1, \ldots, v_i, \ldots, v_n)$, *once the components of each vector have been decreasingly ordered*, iff $\exists j < n\ \forall i = 1, j\ u_i = v_i$ and $u_{j+1} > v_{j+1}$.

- In this first experimentation, we have left out reverse analogy and focused on the two other proportions.
- For each dataset, the classification procedure uses analogy, and then paralogy, in order to compare the respective accuracy rates.
- Depending on the size of the dataset, we experiment diverse sizes for the training sets.
- Given a normalized dataset S and a chosen size n, we randomly build up 10 training sets $train1, \ldots, train10$ with the same size n, each training set having the same repartition in terms of classes as the whole dataset.
- With each training set, we have tested the method on the whole dataset and we report the average accuracy rate for these 10 training sets in the given tables.

We divide our set of experiments between binary class and multiple-class (involving at least 3 classes) problems. Let us start with the binary classification experiments.

A. Binary classification

The target datasets are (in alphabetic order): Blood, Diagnostic (2 tasks), Pima, and Spam. We provide a brief description of each data set below and summarize their main characteristics in Table 2.

Blood: Data are initially coming from the Blood Transfusion Service Center (Hsin-Chu City-Taiwan) where 748 donors were randomly selected from the donor database. Each donor is identified with 4 numerical parameters plus a binary variable representing whether he/she donated blood (binary classification). Added to the repository in October 2008, this set has not been extensively investigated yet. In the case of Blood, we consider sample sets of size 50, 100 and 150. We have 3 batches of 10 tests.

Diagnosis: This small dataset records 120 patient data, each instance having 6 attributes, all of them being binary except the first one, recording the temperature, which is a real number. There are 2 binary output classes: each one deciding for one disease of the urinary system (see [5]). In our case, this will give us 2 classification tasks (that we denote Diag1 and Diag2), one for each disease to be predicted. Given the small size of the dataset, we consider sample sets of size 20, 30 and 40.

Pima: Diabetes diagnosis information for native American women of the Pima heritage. This data consists in diagnostic information for 768 women; 268 of these patients tested positive for diabetes, while 500 tested negative (binary classification). In the case of Pima, we consider sample sets of size 50, 100 and 150.

Spam: This is a standard collection of 4601 emails, each one identified with 57 numerical attributes and classified as spam or ham (binary classification). In that case, we proceed as previously with sample sets of size 50, 100 and 150. Due to the huge number of attributes (57), we had not enough resource to go for bigger training sets.

The results for binary classification tasks are in the corresponding Figures 1, 2 3, 4 and 5.

Table 2: Data sets for binary classification

	Blood	**Diag1**	**Diag2**	**Pima**	**Spam**
#attr.	4	6	6	8	57
#instances	748	120	120	768	4601
#classes	2	2	2	2	2

Figure. 1: Results for Blood data set

data set	Blood		
sample size	50	100	150
analogy	67	67	**72**
paralogy	67	68	70

Figure. 2: Results for Diag1 data set

data set	Diag1		
sample size	20	30	40
analogy	52.7	99	**99.7**
paralogy	89.8	98.7	98.7

Figure. 3: Results for Diag2 data set

data set	Diag2		
sample size	20	30	40
analogy	99.1	98.4	**100**
paralogy	98.9	98.6	**100**

Figure. 4: Results for Pima data set

data set	Pima		
sample size	50	100	150
analogy	66	67	70
paralogy	68	70	**71**

Figure. 5: Results for Spam data set

data set	Spam		
sample size	50	100	150
analogy	72	75	**77**
paralogy	73	75	76

B. Multi-class datasets

Let us consider now the multi-class problems. To start with this category of problems, the chosen target datasets are Iris (3 classes) and Image (7 classes). At this stage, we have not investigated the behavior of our methods when we have more classes.

Image: This image segmentation dataset has 2310 instances from seven outdoor images which are the classes. Each instance is described with 19 attributes. In the case of Image, we increase the size of the sample sets taking into account the huge size of the testing set and to have a representative sample of the 7 classes: we go for 70, 140 and 210 instances.

Iris: The Iris plant dataset is one of the oldest and most popular datasets in classification. It contains 150 instances from three classes: Iris-virginica (class 1), Iris-versicolor (class 2) and Iris-setosa (class 3), 50 instances each. Each instance has 4 numeric attributes. For Iris data set, because of its small size (150 instances) we go for sample sets of size 30, 60 and

Table 3: Data sets for multi-class problems

	Image	**Iris**
$\#attr.$	19	4
$\#instances$	2310	150
$\#classes$	7	3

Figure. 6: Results for Iris and Image data sets

data set	Iris			Image		
sample size	30	60	90	70	140	210
analogy	94	97	**99**	85	88	**90**
paralogy	86	94	95	68	69	65

C. Comments

Whatever the data set, it is not surprising that we get better results when we increase the size of the sample set, starting from a relatively small one. It is amazing to observe that, on the Spam data set, we get 78% of accuracy rate with only 150 observable instances among 4601. This makes our simple classifier not ridiculous knowing that the best rate on this kind of data is around 88% of accuracy.

Obviously, comparing different classifiers is not an easy task, not only because the experimental process can significantly differ, but also because it is not clear what has to be exactly compared. Here, we just give a rough comparison, based on the available figures and considering raw accuracy rates.

- Concerning the Blood data set, this is a relatively recent data set (2008) on which there is not much information available. Nevertheless, according to [37], the accuracy rate seems to be around 60% and we get around 68% (with paralogy).
- Concerning the Diagnostic dataset, we perform very well and this could be due to the fact that only one attribute is real, all the 5 remaining ones being binary. In that case, for the binary attributes, there is no rounding of the real number computation that we do not completely control.
- Concerning the Pima data set, the best results are around 77.7% of accuracy rate (obtained with SVM or LogDisc algorithms), while k-NN gives a little bit less than 72% and C4.5 decision trees gives 72%. We get 71% with only 150 training data, i.e. 1/5 of the full testing set, and without any optimization.
- Concerning the Spam data set, the best accuracy rates are around 88% (see [36]) and we get 77% for analogy and similarly for paralogy with a small training set of only 150 emails, i.e. less than 4% of the testing set including 4601 emails, which is quite encouraging.
- Concerning the Image data set where there are 7 classes, the work is more of a clustering problem than a pure classification one. We get an accuracy rate around 90% for a sample set of 210 instances. This is largely better than a random choice whose accuracy expectation is around 15%.
- Concerning the Iris data set, the best accuracy rates are around 99%, and we get this number with analogy.

It appears, on Diag1, Pima and Blood, that paralogy can from time to time perform better than analogy, but also perform very badly with Image where we have a loss of around 20% in terms of accuracy rate. This is not completely surprising because the two proportions capture different intuitions: it remains to understand exactly why this is the case here and if we can consider that some types of data are more suited to paralogy than analogy. This has to be investigated in the future. To conclude this short comparison, our proportion-based classifier is, in the worse case, less than 10% below the best known accuracy rates. For a preliminary test, this is not bad when we consider the fact that no optimization has been done, and that none of the remarks of Section III have been taken into account.

V. Further experiments

A second batch of experiments has been done to test the voting algorithm among the k "best" vectors, where instead of getting the class associated to the "best" vector ($k = 1$), we consider a vote among the 3 ($k = 3$) and 5 ($k = 5$) best vectors. To remain within the limit of a realistic computation time, we tried the smallest data set, namely Iris, using the same sample with 30 and 60 elements. The results are shown in Figure 7. It is not coming as a surprise that, whatever the size of the sample set (30 or 60), we improve the accuracy by increasing the number of voters (except for analogy and the 60 sample sets). According to Remark 2 of the previous section, it might be interesting to implement a more clever choice, other than the one based on a simple lexicographic decreasing (leximin ordered) comparison.

Figure. 7: Voting algorithm on Iris data set with n=30 and n=60

size of sample	30			60		
value of k	1	3	5	1	3	5
analogy	94	94.6	**95**	**96.5**	96	96
paralogy	86	93	94	93.8	95.3	95.5

Apart from the optimization of the basic procedure which would be worth investigating, it remains to understand the relationship between the approach and the well-known k-NN approach. Both approaches rely on common sense reasoning and in some sense, we may think that they are somewhat similar. Although our method does not explicitly make use of a distance on the dataset, it might be the case that the final classification step only considers neighbors of the target data d to predict the class of d. If this is the case, then our method would turn to be a simple variant of the k-NN algorithm. So we have to check if it is the case in general, or not (and explain now how we do it).

The question we want to answer is: given a target data d to be classified, does our technique make use of the "neighbors" of d to allocate a class to d? In other terms, if we consider our basic implementation where the "best triple" a, b, c decides for the class of d, do we have among the triple a, b, c, one element or more which is/are among let us say, the "10 nearest neighbors" of d ? To bring an answer to this question,

we need a notion of distance on the dataset. Since we use dataset involving mainly real valued attributes (normalized between 0 and 1), we choose to go for the most straightforward option: the Euclidean distance. This is a natural choice which does not take into consideration any pre-processing allowing, for instance, to introduce different weights for the attributes. We consider only the triple a, b, c leading to a successful classification, i.e., such that the unique solution of the class equation $cl(a) : cl(b) :: cl(c) : x$ (which is exactly the class allocated to the target data d) is the real class of d. Then we consider the percentage of these triples which include at least one among the k-nearest neighbors of d. We understand that, if this percentage is high, it means that our approach can be more or less considered as a variant of the k-NN approach. In this experience, we consider only the nearest neighbor (i.e. k=1) and we deal with analogical proportion.
We provide in Table 8 the percentage of successful classifications using the nearest neighbor of d among the best triple a, b, c. Despite the fact that this should be investigated more

Figure. 8: Link with k-nn for Diag1 and Diag2

training set	**Diag1**	**Diag2**
20	0	6.3
30	6.7	8
40	6.9	6.1

extensively, these first results suggest that the analogical proportion approach should not be considered just as a variant of a k-NN algorithm. This finding is consistent with observations already made in [18] for binary attributes classification problems, where a related approach (see next section) was able to outperform the k-NN method.

VI. Related works

The use of analogical proportions as an inferential tool for classification or machine learning purposes has not been considered by many researchers. As far as we know, neither the related proportions (reverse analogy and paralogy), nor the fuzzy extensions of the three analogy-related proportions have been investigated with such an aim in mind. Nevertheless, in [4, 18], the authors have developed a binary classifier on the basis of a very interesting notion that they call "analogical dissimilarity" and denote AD. For a given tuple (a, b, c, d), $AD(a, b, c, d)$ is a positive number which is zero if and only if the proportion $a : b :: c : d$ holds. In fact, $AD(a, b, c, d)$ quantifies how far is the tuple (a, b, c, d) from building an analogical proportion. Definitions of AD are available for both binary [4], and real valued features [18][5]. There is indeed some correlation between the analogical dissimilarity and the fuzzy truth value of an analogical proportion as defined in this paper, since when the truth value increases AD decreases. AD is easily extended to vectors of features just by adding the dissimilarities component-wise, possibly in a weighted way. The basis of the algorithm is then, given a new data d to be classified, to search for the triple (a, b, c) minimizing $AD(a, b, c, d)$ and then to allocate to d the class solution of the equation $cl(a) : cl(b) :: cl(c) : x$. Roughly speaking, the authors minimize a dissimilarity measure when we maximize a fuzzy truth value. Their method, when conveniently tuned with attribute weights and optimized with a voting algorithm to take into account more than one triple, provides very accurate results and seems to belong to the best classifiers on binary and nominal data sets (see [4] for comparative study). Recently in [12], the same authors have extended the notion of analogical proportion and analogical dissimilarity to tree-structures. When a tree represents the syntactic structure of a sentence, it is possible to predict the syntactic structure of a new sentence s_4, in analogical proportion with 3 other known sentences $s_1 : s_2 :: s_3 : s_4$, assuming their associated trees are in analogy as well. Solving the equation $t_1 : t_2 :: t_3 : x$ provides a solution for t_4. In 82% of the cases they deal with, the authors get an exact or almost exact restitution of the parsing tree. At this stage, it would be interesting to make a careful comparison, both at the theoretical and experimental levels, between the definition of "analogical dissimilarity" and the fuzzy truth value of an analogical proportion as defined in this paper. This might be also applied to the other proportions.
Beyond the scope of logic, [35, 19, 2] provide a large panel of algebraic interpretations for analogical proportions, from semi-groups to lattices, through words over finite alphabets and finite trees, even investigating a second order algebraic framework as in [33]. The practicality of the approach for natural language purpose has mainly been investigated in [34], using analogies on words and trees. They focus on supervised learning tasks, aimed at performing the lexical analysis of isolated word forms in diverse languages (English, German and Dutch). Their generic analogical inference procedure has shown promising generalization performance [34].
It is worth noticing that the proposed approach also encompasses the idea of *mean*, since the solution m of the analogical proportion $a : m :: m : b$ should be "half-way" between a and b. A non standard approach in machine learning [22] has recently investigated hybrid methods for estimating software development costs, using diverse forms of means (arithmetic, geometric and harmonic). This suggests, beyond a comparison between the two approaches, to specifically study classifiers where the item to be classified is considered as the mean term of a multiple-features analogical proportion. Such classifiers would obviously have a lower complexity than the approach described above where we have to look for triples a, b, and c.

VII. Towards a new perspective on transduction

In the machine learning literature, there is a variety of well-known algorithms; see e.g., [8] for a detailed comparison between diverse classification techniques (on data coming from a real private business). Among well-established learning techniques, ranging from basic k-NN algorithms to more sophisticated Bayesian approaches, and to SVM methods, which provide accurate results for classification problems. Among all these techniques, one can distinguish between

[5] However, it seems that experiments with AD-based classifiers have only been reported in the binary case, but several possible definitions exist for AD in the real-valued case [20].

- the ones that build a general predictive model of the data at hand which is then applied to the new incoming data to predict their class (this is the case for SVM, Bayesian, or neural networks methods in particular);
- the other ones, sometimes called "lazy methods", which only consider the data at hand to predict the class of the new data, without the intermediate step of building a global model. A typical example is the k-NN algorithm which relies on a kind of local view of the data, and considers its neighbors for classifying the new piece of data.
It is outside the scope of the paper to deeply examine the differences between these approaches and the proposed method. Nevertheless, at least from a conceptual viewpoint, the analogical method departs from the other ones in various respects, since
- we do not describe the problem using an uncertainty reasoning perspective and there is no statistics involved;
- we do not try to separate the classes in the best manner;
- we are not confined to a simple neighborhood analysis.
At this stage, it is worth noting that the proposed method, in its basic form, relies on simple logical computations aiming at comparing data on a quaternary basis, where

- no parameter has to be tuned,
- no weight has to be defined,
- no threshold has to be estimated,
- and finally no global estimate has to be computed.

Still, despite its simplicity, it appears that our results are comparable to the ones of classical classifiers. It is also interesting to note that our procedure does not need to compile multi-class problems into binary ones: proportion-based inference is naturally suitable whatever the number of classes is.
Clearly, the optimization of our method could be done by investigating some of the ideas previously mentioned at the end of Section III. Let us also briefly mention how we could integrate some numerical evaluation in our approach, leading to a completely different algorithm. Taking inspiration from other methods (transduction with SVM for instance), it would be possible to proceed in 2 steps (we consider a binary classification problem):
- allocate to the new data d to be classified the class 0. Then count the number of perfect analogical proportions we get in that case, i.e. the number C_0 of triples (a, b, c) such that $a : b :: c : d$ and $cl(a) : cl(b) :: cl(c) : 0$ hold.
- allocate to the new data d to be classified the class 1. Then compute as previously the associate number C_1 but replacing 0 by 1.

The rule to classify d would be:

```
if C_0==C_1 then cl(d)=undefined
            else {if C_0>C_1 then cl(d)=0
                                  else cl(d)=1;}
```

Obviously, this method could be extended to include paralogy, i.e. instead of only counting perfect analogies, we may count perfect analogies and perfect paralogies. It remains to implement this way to proceed to get a clear picture of its performance compared to our initial method. This new method is in fact more explicitly based on a way for privileging some kind of "regularity" regarding the way data are classified, since we choose to allocate to d the class that maximizes regularities (i.e. valid proportions) in the universe of data, considering proportions that hold as a pieces of evidence in favor of regularity.
As it is now clear, the use of analogical proportions (or paralogical ones) goes far beyond the standard k-NN approach, since similarities and differences between items take into account some regularities in the change in attribute values w.r.t. the way the corresponding vectors are classified. So far, in the component-wise evaluation of a 4-tuple of vectors, analogical proportions consider in the same way complete (approximate) i) similarities (a:a::a:a) of the 4 attribute values, ii) pairwise identities (a:a::b:b) and iii) identities of change (a:b::a:b). We may think that these different patterns do not play the same role and might be handled separately. This is another line of research for further investigation.

VIII. Conclusion

We have introduced a simple classification procedure that works by finding potential proportions that hold between available data and new data to be classified. Our inference principle is based on the assumption that a given proportion, holding between many attributes, should hold between classes. This assumption is the basis of analogical inference, and more precisely, of a proportion-based inference which may be based on analogical, reverse analogical and paralogical proportions. This principle, suitable for binary and graded proportions, allows us to deal with numerical data and multiple-class problems, without any restriction. Our experiments were primarily aiming at getting a better understanding of the process rather than to really compete with existing well-known classifiers. Obviously, a lot remains to be investigated in terms of experimentation, in terms of complexity optimization, and in terms of choice of the best vectors, etc. Nevertheless, our preliminary tests, relying on a basic implementation and tested on 6 standard UCI data sets, provide results which are often better than the ones obtained by many methods that are more sophisticated and rely on quite different ideas. Finally, analogy-based methods can be extended to deal with other "intelligent" tasks such as the ones suggested in [16] for intelligent assessments, or in [28] for automatic solving of IQ tests.

References

[1] A. Aamodt and E. Plaza. Case-based reasoning; foundational issues, methodological variations, and system approaches. *AICom, 7(1)*, pages 39–59, 1994.

[2] N. Barbot and L. Miclet. La proportion analogique dans les groupes: applications aux permutations et aux matrices. *IRISA, Rep. 1914*, 2009.

[3] A. M. Bartkowiak. Anomaly, novelty, one-class classification: A comprehensive introduction. *Int. J. of Computer Information Syst. and Industrial Management Appl.*, 3:061–071, 2011.

[4] S. Bayoudh, L. Miclet, and A. Delhay. A.: Learning by analogy: A classification rule for binary and nominal data. *In: Proc. Inter. Conf. on Artificial Intelligence IJCAI07*, pages 678–683, 2007.

[5] J. Czerniak and H. Zarzycki. Application of rough sets in the presumptive diagnosis of urinary system diseases. In *Artifical Inteligence and Security in Computing Systems, ACS'2002 9th International Conference Proceedings*, pages 41–51. Kluwer Academic Publishers, 2003.

[6] B. V. Dasarathy. *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*. Los Alamitos: IEEE Computer Society Press, 1990.

[7] D. Dubois, H. Fargier, and H. Prade. Refinements of the maximin approach to decision-making in fuzzy environment. *Fuzzy Sets and Systems*, 81:103–122, 1996.

[8] M. Eminagaoglu and S. Eren. Implementation and comparison of machine learning classifiers for information security risk analysis of a human resources department. *Int. J. of Computer Information Syst. and Industrial Management Appl.*, 3:391–398, 2011.

[9] A. Frank and A. Asuncion. UCI machine learning repository, 2010.

[10] A. Gammerman, V. Vovk, and V. Vapnik. Learning by transduction. In *Uncertainty in AI*, pages 148–155. Morgan Kaufmann, 1998.

[11] D. Gentner, K. J. Holyoak, and B.N. Kokinov. *(Eds.) The Analogical Mind: Perspectives from Cognitive Sciences*. MIT Press, 2001.

[12] A. Ben Hassena and L. Miclet. Analogical learning using dissimilarity between tree-structures. In *Proc. 19th Europ. Conf. on Artificial Intelligence (ECAI'10), Lisbon*, volume 215 of *Frontiers in AI and Applications*, pages 1039–1040. IOS Press, 2010.

[13] D. Hofstadter and M. Mitchell. The Copycat project: A model of mental fluidity and analogy-making. In D. et al. Hofstadter, editor, *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*, pages 205–267, New York, NY, 1995. Basic Books, Inc.

[14] E. Hüllermeier. *Case-Based Approximate Reasoning*. Theory and Decision Library. Springer-Verlag New York, 2007.

[15] S. Klein. Culture, mysticism & social structure and the calculation of behavior. In *Proc. Europ. Conf. in AI (ECAI)*, pages 141–146, 1982.

[16] J. Elmy Mat-jizat and E. McKay. Developing an instrument of assessment for ict literacy for trainee teachers: Preliminary findings. *Int. J. of Computer Information Syst. and Industrial Management Appl.*, 3:552–559, 2011.

[17] E. Melis and M. Veloso. Analogy in problem solving. In L. Farinas del Cerro, D. Gabbay, and H. J. Ohlbach, editors, *Handbook of Practical Reasoning: Computational and Theoretical Aspects*, volume 17 (1), pages 1–73. Oxford Univ. Press, 1998.

[18] L. Miclet, S. Bayoudh, and A. Delhay. Analogical dissimilarity: definition, algorithms and two experiments in machine learning. *JAIR, 32*, pages 793–824, 2008.

[19] L. Miclet and A. Delhay. Relation d'analogie et distance sur un alphabet defini par des traits. Technical Report 1632, IRISA, July 2004.

[20] L. Miclet and H. Prade. Logical definition of analogical proportion and its fuzzy extensions. In *Proc. Annual Meeting of the North Amer. Fuzzy Information Proc. Soc. (NAFIPS), New-York, May, 19-22*. IEEE, 2008.

[21] L. Miclet and H. Prade. Handling analogical proportions in classical logic and fuzzy logics settings. In *Proc. 10th ECSQARU, Verona*, volume LNCS 5590, pages 638–650. Springer, 2009.

[22] J. S. Pahariyaa, V. Ravia, M. Carra, and M. Vasua. Computational intelligence hybrids applied to software cost estimation. *Int. J. of Computer Information Syst. and Industrial Management Appl.*, 2:104–112, 2010.

[23] H. Prade. Proportion analogique et raisonnement à partir de cas. In *Rencontres Francophones sur la Logique Floue et ses Applications*, pages 296–303. Cépaduès, 2008.

[24] H. Prade and G. Richard. Analogical proportions: another logical view. In *Proc. 29th Ann. Inter. Conf. on Artif. Intellig. (AI'09), Cambridge UK*. Springer, 2009.

[25] H. Prade and G. Richard. Analogy, paralogy and reverse analogy: Postulates and inferences. In *Proc. 32nd Ann. Conf. on Artif. Intellig. (KI 2009), Paderborn, Sept. 15-18*, volume LNAI 5803, pages 306–314. Springer, 2009.

[26] H. Prade and G. Richard. Multiple-valued logic interpretations of analogical, reverse analogical, and paralogical proportions. In *Proc. 40th IEEE International Symposium on Multiple-Valued Logic, ISMVL 2010, Barcelona, Spain, 26-28 May 2010*, pages 258–263, 2010.

[27] H. Prade and G. Richard. Reasoning with logical proportions. In *Proc. 12th Inter. Conf. on Principles of Knowledge Representation and Reasoning, KR 2010, Toronto, Ontario, Canada, May 9-13, 2010 (F. Z. Lin, U. Sattler, M. Truszczynski, eds.)*, pages 546–55. AAAI Press, 2010.

[28] H. Prade and G. Richard. Analogy-making for solving iq tests: A logical view. In *Proc. Inter. Conf. on Case-Based Reasoning (ICCBR'11)*, pages 241–250. Greenwich University, London, Sept. 12-15, Springer, LNAI 6880, 2011.

[29] H. Prade and G. Richard. Logical handling of analogical proportions in commonsense and transductive reasoning. In *Proc. 3rd IEEE Inter. Conf. on Soft Computing and Pattern Recognition (SoCPaR'11)*, pages 561–566, DaLian, China, Oct. 14-16, 2011.

[30] H. Prade, G. Richard, and B. Yao. Classification by means of fuzzy analogy-related proportions – a preliminary report. In *Proc. 2nd IEEE Int. Conf. on Soft Computing and Pattern Recognition (SocPar'10), Evry*, pages 297–302. 2010.

[31] H. Prade and S. Schockaert. Completing rule bases in symbolic domains by analogy making. In *Proc. 7th Conf. Europ. Society for Fuzzy Logic and Technology (EUSFLAT), Aix-les-Bains, France, July 18-22*, pages 928–934, 2011.

[32] B. Russell. *The Problems of Philosophy. Chap. VI. On induction*. Home Univ. Libr.;Oxford Univ. Pr., 1959, 1912.

[33] U. Schmid, H. Gust, K. Kühnberger, and J. Burghardt. An algebraic framework for solving proportional and predictive analogies. *Eur. Conf. Cogn. Sci. 295-300*, 2003.

[34] N. Stroppa and F. Yvon. An analogical learner for morphological analysis. In *Proc. 9th Conf. Comput. Nat. Lang. Learning (CoNLL-2005)*, pages 120–127, 2005.

[35] N. Stroppa and F. Yvon. Analogical learning and formal proportions: Definitions and methodological issues. ENST Paris report, 2005.

[36] Yong Wang and I. H. Witten. Modeling for optimal probability prediction. In *In Proceedings of the Nineteenth International Conference on Machine Learning*, pages 650–657. Morgan Kaufmann, 2002.

[37] I-Cheng Yeh, King-Jang Yang, and Tao-Ming Ting. Knowledge discovery on rfm model using bernoulli sequence. *Expert Syst. & Appl.*, 36(3):5866–5871, 2009.

Author Biographies

Henri Prade is Research Director at CNRS (the French National Center for Scientific Research) and works at IRIT institute (University Paul Sabatier, Toulouse, France), member of the ADRIA team. His main research interests are in artificial intelligence, and more particularly in the use of possibility theory for uncertainty or preference representation in reasoning and decision. He has also extensively worked in fuzzy set theory, in possibilistic logic, and more recently in analogical reasoning.

Gilles Richard is Professor of Computer Sciences at University Paul Sabatier (Toulouse, France), member of the ADRIA team, and Visiting Professor at the British Institute of Technology and E-commerce (London, UK), where he is Head of Innovation and Technology. His current research interests are in the field of logical proportions and their applications to machine learning and inference. He has also worked on automatic theorem proving and logic programming.

Bing Yao is currently working as IT engineer for China Railway Signal & Communication shanghai engineering Co.,Ltd.(China). After achieving a BSc in Electronics Bing got his master in Artificial Intelligence and Robotics from University of Toulouse (France) in 2010. During his intership within the ADRIA team, he was the main responsible for the project implementation and development of the analogical classifier we investigate in this paper.