



HAL
open science

Classification by means of fuzzy analogy-related proportions – A preliminary report

Henri Prade, Gilles Richard, Bing Yao

► **To cite this version:**

Henri Prade, Gilles Richard, Bing Yao. Classification by means of fuzzy analogy-related proportions – A preliminary report. International Conference of Soft Computing and Pattern Recognition, Dec 2010, Cergy Pontoise / Paris, Unknown Region. pp.297–302. <hal-03349439>

HAL Id: hal-03349439

<https://hal.science/hal-03349439v1>

Submitted on 20 Sep 2021

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

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



HAL Authorization

Classification by means of fuzzy analogy-related proportions – A preliminary report

Henri Prade, Gilles Richard, Bing Yao

► **To cite this version:**

Henri Prade, Gilles Richard, Bing Yao. Classification by means of fuzzy analogy-related proportions – A preliminary report. International Conference of Soft Computing and Pattern Recognition, Dec 2010, Cergy Pontoise / Paris, Unknown Region. pp.297–302. hal-03349439

HAL Id: hal-03349439

<https://hal.archives-ouvertes.fr/hal-03349439>

Submitted on 20 Sep 2021

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

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

Classification by means of fuzzy analogy-related proportions – A preliminary report

Henri Prade
IRIT, University of Toulouse
31062 Toulouse Cedex 9, France
Email: prade@irit.fr

Gilles Richard
British Institute of Technology & E-commerce
London, E7 9HZ, UK
Email: grichard@bite.ac.uk

Bing Yao
IRIT, University of Toulouse
31062 Toulouse Cedex 9, France
Email: yaojing1@hotmail.com

Abstract—Boolean logic interpretations, as well as multiple-valued logic extensions, have been recently proposed for analogical proportions (i.e. statements of the form “ a is to b as c is to d ”), and for 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”). These proportions relate items a , b , c , and d on the basis of their differences, or of their similarities. This may provide a basis for proposing a plausible classification for an object d described in terms of a set of features, on the basis of three other already classified objects described on the same features, considering 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.

Index Terms—analogy; multi-valued; classification;

I. INTRODUCTION

Analogical reasoning [1], [2] 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 [3] or k -nearest neighbors methods [4], [5] when we have at our disposal a repertory of cases or examples. Then, analogy is understood as the identity or similarity of relations or properties between two ordered pairs, as 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 $cl_2 : cl_1 :: des_2 : des_1$.

However, a more powerful use of analogical proportions has been recently envisaged by some authors [6], [7], [8], [9], 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 [10] 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 $h_{1''}$ which have equipment 1 (eq_1) or equipment 2 (eq_2), or not. If we build the analogical proportion $(h_1, eq_1, eq_2, \neg ch) : (h_{1'}, eq_1, \neg eq_2, ch) :: (h_{1''}, \neg eq_1, eq_2, \neg ch) : (h_2, \neg eq_1, \neg eq_2, ?)$, 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 multiple-valued features. Thus, a measure of analogical dissimilarity has been proposed [8], [9]. More recently, a propositional logic modeling of analogical proportions has been developed [11], [12], and extended to fuzzy logic for handling multiple-valued features [13]. Moreover, it has been also pointed out that two other types of proportion exist alongside analogical proportion [14], [15], namely *reverse analogy* which focuses on the differences between the items at hand as analogy does but reverse the directions of changes, and the so-called *paralogy*, denoted $a ; b :: c ; d$ which rather focuses on common properties and similarities, i.e. $a ; b ; c ; d$ holds if the features common to a and b are also common to c and d .

In this paper, we investigate a classification procedure entirely based on the use of analogical proportions (or of paralogical proportions), applied in a component-wise manner to vectors $(v_1, \dots, v_i, \dots, v_n, v_{n+1})$ 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. 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 extensions. Section III introduces the classification procedure, while Section IV reports results obtained on binary classes, or multiple classes benchmark problems with numerical features. Section V discusses the approach with respect to related works, and contrasts it to other classification methods.

II. LOGICAL MODELING OF ANALOGY-RELATED PROPORTIONS

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 [14], [15]. They all satisfy symmetry when comparing a, b with c, d , but obey to 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 set 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 [12] ($a \setminus b$ denotes the set difference $a \cap \bar{b}$):

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

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 senses:

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

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:

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

The following property establishes a strong link between the three kinds of proportion:

Property 1: If A is an analogy, then the relation

- R defined as $R(a, b, c, d)$ iff $A(a, b, d, c)$ is a reverse analogy
- P defined as $P(a, b, c, d)$ iff $A(a, d, c, b)$ is a paralogy.

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 implication \rightarrow , while $=$ is changed into \equiv , \cup into \vee , \cap into \wedge :

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

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

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

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 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 I provides the truth values making the relations to hold (among 16 possibilities). In this interpretation, it is 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.

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

TABLE I
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 |

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 [12], [14], as first guessed from anthropological observations by Klein [16] (without distinguishing the three proportions).

It is worth noticing that the above logical view of analogy departs from more established views [17], which are more oriented towards analogy-making than analogy-checking.

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 of the operations chosen for modeling \wedge , \equiv , and \rightarrow . 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

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

A careful analysis [13] of the requirements leads to choose

- the minimum operator for \wedge ; - $s \equiv t = 1 - |s - t|$;
- Lukasiewicz implication $s \rightarrow t = \min(1, 1 - s + t)$. Note also that with these choices $s \equiv t = (s \rightarrow t) \wedge (t \rightarrow s)$.

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

(Analogy) $a : b :: c : d =$

$$1 - |(a - b) - (c - d)| \text{ if } a \geq b \ \& \ c \geq d, \text{ or } a \leq b \ \& \ c \leq d$$

$$1 - \max(|a - b|, |c - d|) \text{ if } a \leq b \ \& \ c \geq d \text{ or } a \geq b \ \& \ c \leq d$$

(Reverse analogy) $a ! b :: c ! d =$

$$1 - |(a - b) - (d - c)| \text{ if } a \leq b \ \& \ c \geq d \text{ or } a \geq b \ \& \ c \leq d$$

$$1 - \max(|a - b|, |c - d|) \text{ if } a \geq b \ \& \ c \geq d, \text{ or } a \leq b \ \& \ c \leq d$$

Clearly, it exists x such that $a : b :: c : x = 1$ if and only if $x = c + b - a \in [0, 1]$. When it exists, the solution is unique.

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 [13] for justifications.

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 $(f_i - \min\{f_i\})/(\max\{f_i\} - \min\{f_i\})$). This includes the particular cases of binary features, and of many-valued discrete features on ordered domains, which can 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 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 [18], [19] 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. 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 $k = 1$, $cl(x) = cl(\operatorname{argmin}_{s|(s, cl(s)) \in S} |dis(x, s)|)$, 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 $a = (a_1, \dots, a_n)$ of truth values that encode the values of n 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 p components of these vectors, then this proportion should hold for the last remaining components as well. This inference principle can be stated as below:

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

This is a generalized form of analogical reasoning, where we transfer knowledge from some components of our vectors

to their remaining components. This idea, when applied to classification, provides the basis for a transduction mechanism, already discussed in [20] for binary classification and 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 binary features associated to (a, b, c, d) (i.e. T holds on the input space) 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. 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. $a = b = c = d$).

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 [20] for a detailed discussion in the binary case. The fact that the proportions are independent w.r.t. the positive or negative encoding of the features guarantees that the same classification result will be obtained whatever the encoding.

B. Towards an algorithm

We apply a strategy very different from k-NN methods, since the new item d to be classified is not just compared with classified items on a one-to-one basis. For the sake of uniformity, we continue to denote d the new item to be classified: d is known, but $cl(d)$ is unknown. First, we look for 3-tuples $(a, b, c) \in S^3$ such that the class equation $T(cl(a), cl(b), cl(c), x)$ 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. Then either $cl(d) = cl(a) = cl(b) = cl(c)$, or there are two distinct classes, which we may encode 1 and 0, and the corresponding class equation should have a solution in order the triples to be retained. 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.

When an item d is to be classified, we have to look among the set of suitable triples (for a fixed T) for the one(s) that seem(s) the most appropriate for predicting the class $cl(d)$. For doing this, each suitable triple we consider is evaluated by means of the following vector $(T(a_1, b_1, c_1, d_1), \dots, T(a_i, b_i, c_i, d_i), \dots, T(a_n, b_n, c_n, d_n))$. Then the vectors (and thus the triples) are ordered in lexicographic decreasing order¹. 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 we have not enough information for making a proper choice. 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

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

TABLE II
THE DATA SETS

| | Pima | Spam | Blood | Image | Iris |
|------------|------|------|-------|-------|------|
| #attr. | 8 | 57 | 4 | 19 | 4 |
| #instances | 768 | 4601 | 748 | 2310 | 150 |
| #classes | 2 | 2 | 2 | 7 | 3 |

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 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 sensible way.

4. The lexicographic ordering presupposes that all the features have the same importance. This may not be the case if some are irrelevant. Then one may think of using the discrimin ordering instead (the vectors 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 previous ideas (we have only tested the basic procedure, but none of the 4 types of refinements suggested above), we test them on different data sets coming from the University of California, Irvine (UCI) machine learning repository. We have chosen 5 data sets which constitutes a representative panel of diverse classification contexts. The data sets are Pima, Spam, Blood, Iris and Image segmentation. The attributes are mainly quantitative and continuous but some of them (age for instance) are quantitative and discrete. The number of attributes varies from 4 to 57. In terms of classes, Pima, Spam and Blood are binary classification problems, while Iris and Image are multiple-class ones with 3 and 7 classes respectively. Moreover, Iris and Pima have been extensively investigated, which allows us to have a clear picture of the performance of other approaches. We provide below a brief description of each data set and we summarize their main characteristics in Table II.

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).

Spam: This is a standard collection of 4601 emails, each one identified with 57 numerical attributes and classified as spam or ham (binary classification).

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.

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

Image: This image segmentation dataset has 2310 instances from seven outdoor images which are the classes. Each instance is described with 19 attributes. 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 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 our limited computation facilities, somewhat limits the experimentations (w.r.t. the size of the samples).

Our protocol is quite simple: we choose a fixed size for the sampling sets and we consider 10 different sampling (training) sets of the same chosen size. For each sampling set, we run the algorithm and compute the accuracy rate on the full set of data (testing set). In the tables, we report the average accuracy rate over our 10 tests. This work is done for diverse sizes of sampling sets, in accordance with the size of the testing set.

- In the case of Pima, Spam, we consider sample sets of size 50, 100 and 150: we have then 3 batches of 10 tests.

- 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.

- For Iris data set, because of its small size (150 instances), we reduce the sample sets to 30, 60 and 90 for the test.

In this first experimentation, we leave out reverse analogy and we focus on analogy (first line of the tables) and paralogy (second line of the tables). The results are in the corresponding Figures 1, 2 and 3. Whatever the data set, it is not surprising

Fig. 1. Results for Pima and Spam data sets

| data set | Pima | | | Spam | | |
|-------------|------|-----|-----------|------|-----|-----------|
| sample size | 50 | 100 | 150 | 50 | 100 | 150 |
| analogy | 66 | 67 | 70 | 72 | 75 | 77 |
| paralogy | 68 | 70 | 71 | 73 | 75 | 76 |

Fig. 2. Results for Blood

| | | | |
|-------------|----|-----|-----------|
| sample size | 50 | 100 | 150 |
| analogy | 67 | 67 | 72 |
| paralogy | 67 | 68 | 70 |

Fig. 3. 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 |

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 Pima data set, the best results are around 77.7% of accuracy rate (obtained with SVM or LogDisc algorithm) where kNN gives a little bit less than 72% and C4.5 Decision Trees gives 72%. We give 71% with only 150 training data i.e. 1/5 of the full testing set, and without any optimization (source:<http://www.is.umk.pl/projects/datasets.html>).

- Concerning Spam data set, the best accuracy rates are around 88% (see [21]) 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 Blood data set, this is a relatively recent data set (2008) and then we did not find a lot of information. Nevertheless, according to [22], the accuracy rate seems to be around 60% and we get around 68% (with paralogy).

- Concerning Image data set where we have 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. Which 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 Pima and Blood, that paralogy can from time to time perform slightly better than analogy, and 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 optimisation has been done and none of the remarks of section III taken into account.

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 keep on within the limit of realistic time, we tried the smallest data set namely Iris using the same sample with 30 and 60 elements. The results are shown in Figure 4. It is not coming

Fig. 4. Voting algorithm on Iris data set with $n=30$ and $n=60$

| size of sample | 30 | | | 60 | | |
|----------------|----|------|-----------|-------------|------|------|
| | 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 |

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 would be interesting to implement a more clever choice, other than the one based on a simple lexicographic decreasing (leximin order) comparison.

V. DISCUSSION AND 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 [8], [9], the authors have developed a binary classifier on the basis of a very interesting concept that they call “analogical dissimilarity” denoted AD . For a given tuple (a, b, c, d) , $AD(a, b, c, d)$ is a positive number and is null iff $a : b :: c : d$ holds: in fact, $AD(a, b, c, d)$ quantifies how far are (a, b, c, d) from building an analogical proportion and definitions are available for both binary [8] or real valued features [9]². 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 componentwise, 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 [8] for comparative study). Recently in [23], the authors have extended the notion of analogical proportion and analogical dissimilarity to tree-structures. When a tree represent the syntactic structure a sentence, it is possible to predict the syntactic structure of a new sentence s_4 , in analogical proportion with 3 other known sentences $s_1 : s_2 :: s_3 : s_4$ and 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, [6], [24], [25] 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 [26]. The practicality of the approach for natural language purpose has mainly been investigated in [7],

²It seems that experiments with AD -based classifiers have only been reported in the binary case, and several possible definitions exist for AD in the real-valued case [11].

using analogies on words and trees. They focus on the task of supervised learning tasks, both 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 [7]).

In machine learning literature, there are different techniques, established for a long time, including basic kNN algorithms, Bayesian approaches, and SVM methods, providing accurate results for classification problems. It is outside the scope of the paper to deeply examine the differences with the proposed method. Nevertheless, at least from a conceptual viewpoint, the analogical method is different 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, our method is based on a simple logical computation, where no parameter has to be tuned, no weight has to be defined, no threshold has to be estimated. And, despite its simplicity, it appears that our results are comparable to the average 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. As it is 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 vectors are classified. So far, in the componentwise evaluation of a 4-tuples 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.

VI. 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 generally, 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 are 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 investigate in terms of experimentation, in terms of complexity optimization in terms of choice of the best vectors, etc. Nevertheless, our preliminary tests, relying on a basic implementation and tested on 5 standard UCI data sets, provide results which are better than a lot of much more sophisticated methods.

REFERENCES

- [1] D. Gentner, K. J. Holyoak, and B. Kokinov, (Eds.) *The Analogical Mind: Perspectives from Cognitive Sciences*. MIT Press, 2001.
- [2] E. Melis and M. Veloso, "Analogy in problem solving," in *Handbook of Practical Reasoning: Computational and Theoretical Aspects*. Oxford Univ. Press, 1998.
- [3] A. Aamodt and E. Plaza, "Case-based reasoning; foundational issues, methodological variations, and system approaches," *AICom*, 7(1), pp. 39–59, 1994.
- [4] B. V. Dasarathy, *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*. Los Alamitos: IEEE Computer Society Press, 1990.
- [5] E. Hullermeier, *Case-Based Approximate Reasoning*, ser. Theory and Decision Library. Springer-Verlag New York, 2007.
- [6] N. Stroppa and F. Yvon, "Analogical learning and formal proportions: Definitions and methodological issues," ENST Paris report, 2005.
- [7] —, "An analogical learner for morphological analysis," in *Proc. 9th Conf. Comput. Nat. Lang. Learning (CoNLL-2005)*, 2005, pp. 120–127.
- [8] S. Bayouduh, 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*, pp. 678–683, 2007.
- [9] L. Miclet, S. Bayouduh, and A. Delhay, "Analogical dissimilarity: definition, algorithms and two experiments in machine learning," *JAIR*, 32, pp. 793–824, 2008.
- [10] H. Prade, "Proportion analogique et raisonnement à partir de cas," in *Rencontres Francophones sur la Logique Floue et ses Applications*. Cepaduès, 2008, pp. 296–303.
- [11] 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.
- [12] —, "Handling analogical proportions in classical logic and fuzzy logics settings," in *Proc. 10th ECSQARU, Verona*, vol. LNCS 5590. Springer, 2009, pp. 638–650.
- [13] 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*, 2010, pp. 258–263.
- [14] —, "Analogy, paralogy and reverse analogy: Postulates and inferences," in *Proc. 32nd Ann. Conf. on Artif. Intellig. (KI 2009), Paderborn, Sept. 15-18*, vol. LNAI 5803. Springer, 2009, pp. 306–314.
- [15] —, "Analogical proportions: another logical view," in *Proc. 29th Ann. Inter. Conf. on Artif. Intellig. (AI'09), Cambridge UK*. Springer, 2009.
- [16] S. Klein, "Culture, mysticism & social structure and the calculation of behavior," in *Proc. Europ. Conf. in AI (ECAI)*, 1982, pp. 141–146.
- [17] D. Hofstadter and M. Mitchell, "The Copycat project: A model of mental fluidity and analogy-making," in *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*, D. a. Hofstadter, Ed. New York, NY: Basic Books, Inc., 1995, pp. 205–267.
- [18] B. Russell, *The Problems of Philosophy. Chap. VI. On induction*. Home Univ. Libr.;Oxford Univ. Pr., 1959; <http://www.ditext.B.com/russell/russell.html>, 1912.
- [19] A. Gammerman, V. Vovk, and V. Vapnik, "Learning by transduction," in *Uncertainty in AI*. Morgan Kaufmann, 1998, pp. 148–155.
- [20] 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.)*. AAAI Press, 2010.
- [21] Y. Wang and I. H. Witten, "Modeling for optimal probability prediction," in *In Proceedings of the Nineteenth International Conference on Machine Learning*. Morgan Kaufmann, 2002, pp. 650–657.
- [22] I.-C. Yeh, K.-J. Yang, and T.-M. Ting, "Knowledge discovery on rfm model using bernoulli sequence," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5866–5871, 2009.
- [23] A. B. Hassena and L. Miclet, "Analogical learning using dissimilarity between tree-structures," in *Proc. 19th European Conference on Artificial Intelligence ECAI2010, Lisbon, Portugal*, ser. Frontiers in AI and Applications, vol. 215. IOS Press, 2010, pp. 1039–1040.
- [24] L. Miclet and A. Delhay, "Relation d'analogie et distance sur un alphabet defini par des traits," *IRISA, Tech. Rep. 1632*, July 2004.
- [25] N. Barbot and L. Miclet, "La proportion analogique dans les groupes: applications aux permutations et aux matrices," *IRISA, Rep. 1914*, 2009.
- [26] 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.