



HAL
open science

Classification based on homogeneous logical proportions

Ronei Moraes, Liliane Machado, Henri Prade, Gilles Richard

► **To cite this version:**

Ronei Moraes, Liliane Machado, Henri Prade, Gilles Richard. Classification based on homogeneous logical proportions. 33rd International Conference on Innovative Techniques and Applications of Artificial Intelligence (SGAI 2013), Dec 2013, Cambridge, United Kingdom. pp.53-60, 10.1007/978-3-319-02621-3_4. hal-04082603

HAL Id: hal-04082603

<https://hal.science/hal-04082603>

Submitted on 26 Apr 2023

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.



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 12865

URL: http://dx.doi.org/10.1007/978-3-319-02621-3_4

To cite this version : Moraes, Ronei and Machado, Liliane and Prade, Henri and Richard, Gilles *Classification based on homogeneous logical proportions*. (2013) In: 33rd International Conference on Innovative Techniques and Applications of Artificial Intelligence (SGAI 2013), 12 December 2013 - 14 December 2013 (Cambridge, United Kingdom).

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Classification based on homogeneous logical proportions

Ronei M. Moraes, Liliane S. Machado, Henri Prade, and Gilles Richard

Abstract A classification method based on a measure of analogical dissimilarity has been proposed some years ago by Laurent Miclet and his colleagues, which was giving very good results. We restart this study on a slightly different basis. Instead of estimating analogical dissimilarity, we use the logical definition of an analogical proportion. We also consider other related logical proportions and their link with analogical proportion. The paper reports on an ongoing work and contributes to a comparative study of the logical proportions predictive accuracy on a set of standard benchmarks coming from UCI repository. Logical proportions constitute an interesting framework to deal with binary and/or nominal classification tasks without introducing any metrics or numerical weights.

1 Introduction

In problem solving, analogical reasoning [3, 5] is often viewed as a way to enhance reasoning processes by transferring conclusions observed in known situations to an only partially known situation. A particular instance of this reasoning style is derived from the notion of analogical proportion linking 4 situations or items a, b, c, d and stating that a is to b as c is to d , often denoted $a : b :: c : d$. This statement basically expresses that a differs from b as c differs from d and vice-versa. In a classification context where a, b, c are in a training set, i.e., their class $cl(a), cl(b), cl(c)$ are known, and d is a new object whose class is unknown, this analogical proportion leads to a common sense inference rule which can be stated as:

Ronei M. Moraes and Liliane S. Machado
LabTEVE, Federal University of Paraiba, Cidade Universitaria s/n, Joao Pessoa, Brazil, e-mail:
ronei@de.ufpb.br, liliane@di.ufpb.br

Henri Prade and Gilles Richard
IRIT, University of Toulouse, 118 route de Narbonne, 31062 Toulouse Cedex 09, France, e-mail:
prade@irit.fr, richard@irit.fr

if a is to b as c is to d, then cl(a) is to cl(b) as cl(c) is to cl(d)

In [13, 2, 6], this inference principle has been deeply investigated on diverse domains and even implemented, using a measure of analogical dissimilarity. More precisely, when dealing with classification problems as in [6, 12], the items a, b, c, d are generally represented as vectors of values. To fit with real life applications, the previous inference rule has to be (in some sense) weakened into: *if* an analogical proportion holds for a large number of properties between the four items, *then* an analogical proportion should also hold for the class of these items.

In the last past years, a logical model of analogical proportions has been proposed [7, 9, 10]. It has been shown that analogical proportions are a special case of so-called *logical proportions* [8]. Roughly speaking, a logical proportion between four terms a, b, c, d equates similarity or dissimilarity evaluations about the pair a, b with similarity or dissimilarity evaluations about the pair c, d . Among the set of existing logical proportions, four *homogeneous* ones (in the sense that they are equating similarities with similarities, or dissimilarities with dissimilarities), including the analogical proportion, enjoy remarkable properties and seem particularly attractive for completing missing values [10].

This paper suggests to use diverse homogeneous proportions to implement what could be called “logical proportion-based classifiers”. We make use of analogical proportion, but also of the three other homogeneous proportions. The implementation provides good results and suggests that logical proportions offer a suitable framework for classification tasks.

The paper is organized as follows. After a short background on Boolean analogical proportion and three related proportions in Section 2, we briefly recall in Section 3 the work of [2] and we describe the new algorithm we suggest. In Section 4, we report experiments on datasets coming from the UCI. Our results are analyzed and compared to those obtained in [2]. We conclude in Section 5.

2 Logical proportions: a short background

A *logical proportion* [9] is a particular type of Boolean expression $T(a, b, c, d)$ involving four variables a, b, c, d , (whose truth values belong to $\mathbb{B} = \{0, 1\}$). It is made of the conjunction of two distinct equivalences, involving a conjunction of variables a, b on one side, and a conjunction of variables c, d on the other side of \equiv , where each variable may be negated. Both $a \wedge \bar{b}$ and $\bar{a} \wedge b$ capture the idea of dissimilarity between a and b , while $a \wedge b$ and $\bar{a} \wedge \bar{b}$ capture the idea of similarity, positively and negatively. For instance, $(\bar{a}\bar{b} \equiv \bar{c}\bar{d}) \wedge (a\bar{b} \equiv \bar{c}d)$ ¹ is the expression of the *analogical proportion* [7]. As can be seen, analogical proportion uses only dissimilarities and could be informally read as *what is true for a and not for b is exactly what is true for c and not for d, and vice versa*. When a logical proportion does not mix similarities

¹ \bar{a} is a compact notation for the negation of a and $\bar{a}\bar{b}$ is short for $a \wedge \bar{b}$, and so on.

and dissimilarities in its definition, we call it *homogeneous*: For instance, analogical proportion is homogeneous.

More generally, it has been proved that there are 120 semantically distinct logical proportions that can be built [8]. Moreover, each logical proportion has exactly 6 lines leading to true (i.e., 1) in its truth table, the 10 remaining lines lead to false (i.e., 0). Logical proportions are quite rare in the world of quaternary Boolean formulas, since there are 8080 quaternary operators that are true for 6 and only 6 lines of this truth table. Inspired from the well known numerical proportions, two properties seem essential for defining the logical proportions that could be considered as the best counterparts to numerical proportions:

- Numerical proportions remain valid when (a, b) is exchanged with (c, d) . Similarly, logical proportions T should satisfy *symmetry* property:

$$T(a, b, c, d) \implies T(c, d, a, b) \quad (1)$$

- A valid numerical proportion does not depend on the representation of the numbers in a particular basis. In the same spirit, logical proportions should satisfy the *code independency* property:

$$T(a, b, c, d) \implies T(\bar{a}, \bar{b}, \bar{c}, \bar{d}) \quad (2)$$

insuring that the proportion T holds whatever the coding of falsity with 0 or 1, or if we prefer, T holds whatever we consider a property or its negation for describing the objects.

Only 4 among the 120 proportions satisfy the two previous properties [10]. These four proportions also satisfy other properties associated with the idea of proportion such as central permutation

$$T(a, b, c, d) \implies T(a, c, b, d) \quad (3)$$

In Table 1, we provide the definitions of the 4 homogeneous logical proportions and in Table 2, we give their truth tables, restricted to the 6 valuations leading to truth value 1.

Table 1 4 remarkable logical proportions: A, R, P, I

A	R	P	I
$(\bar{a}\bar{b} \equiv \bar{c}\bar{d}) \wedge (\bar{a}b \equiv \bar{c}d)$	$(\bar{a}\bar{b} \equiv \bar{c}d) \wedge (\bar{a}b \equiv \bar{c}\bar{d})$	$(ab \equiv cd) \wedge (\bar{a}\bar{b} \equiv \bar{c}\bar{d})$	$(ab \equiv \bar{c}\bar{d}) \wedge (\bar{a}\bar{b} \equiv cd)$

These proportions, denoted by $A(a, b, c, d)$, $R(a, b, c, d)$, $P(a, b, c, d)$ and $I(a, b, c, d)$, are called *analogy*, *reverse analogy*, *paralogy* and *inverse paralogy* [9], and respectively express that a (resp. b) differs from b (resp. a) as c (resp. d) differs from d (resp. c), that a is to b as d is to c , that what a and b have in common, c and d have it also, and that what a and b have in common, both c and d miss it). The 4 homogeneous proportions are linked together:

$$A(a, b, c, d) \equiv R(a, b, d, c) \equiv P(a, d, c, b) \equiv I(a, \bar{b}, \bar{c}, d). \quad (4)$$

The semantical properties of these 4 proportions have been investigated in [10].

Table 2 A, R, P, I: Boolean truth tables

A	R	P	I
0 0 0 0	0 0 0 0	0 0 0 0	0 0 1 1
1 1 1 1	1 1 1 1	1 1 1 1	1 1 0 0
0 0 1 1	0 0 1 1	0 1 1 0	0 1 1 0
1 1 0 0	1 1 0 0	1 0 0 1	1 0 0 1
0 1 0 1	0 1 1 0	0 1 0 1	0 1 0 1
1 0 1 0	1 0 0 1	1 0 1 0	1 0 1 0

The idea of proportion is closely related to the idea of extrapolation, i.e., to guess/compute a new value on the ground of existing values. In other words, if for some reason, it is believed or known that a proportion holds between four binary items, three of them being known, then one may try to infer the value of the fourth one. The problem can be stated as follows. Given a logical proportion T and a 3-tuple (a, b, c) , does there exist a Boolean value x such that $T(a, b, c, x) = 1$, and in that case, is this value unique? It is easy to see that there are always cases where the equation has no solution since the triple a, b, c may take $2^3 = 8$ values, while any proportion T is true only for 6 distinct 4-tuples. For instance, when we deal with analogy A , the equations $A(1, 0, 0, x)$ and $A(0, 1, 1, x)$ have no solution. And it has been proved, for instance that the analogical equation $A(a, b, c, x)$ is solvable iff $(a \equiv b) \vee (a \equiv c)$ holds. In that case, the unique solution is $x = a \equiv (b \equiv c)$. Similar results hold for the 3 remaining homogeneous proportions. As sketched in the introduction, this equation solving property can be the basis of a constructive classification rule over binary data by adopting the following inference rule: having 4 objects $\vec{a}, \vec{b}, \vec{c}, \vec{d}$ (three in the training set with classes $cl(\vec{a}), cl(\vec{b}), cl(\vec{c})$, the fourth being the object to be classified where $cl(\vec{d})$ is unknown:

$$\frac{\forall i \in [1, n], A(a_i, b_i, c_i, d_i)}{A(cl(\vec{a}), cl(\vec{b}), cl(\vec{c}), cl(\vec{d}))}$$

Then, if the equation $A(cl(\vec{a}), cl(\vec{b}), cl(\vec{c}), x)$ is solvable, we can allocate to $cl(\vec{d})$ the solution of this equation. Obviously, A could be replaced with R, P or I . Let us investigate an implementation of this principle.

3 Analogical dissimilarity and algorithms

Learning by analogy, as presented in [2] is a lazy learning technique which uses a measure of *analogical dissimilarity* between four objects. It estimates how far four situations are from being in analogical proportion. Roughly speaking, the analogical dissimilarity ad between 4 Boolean values is the minimum number of bits that have to be switched to get a proper analogy. For instance $ad(1,0,1,0) = 0$, $ad(1,0,1,1) = 1$ and $ad(1,0,0,1) = 2$. Thus, $A(a,b,c,d)$ holds if and only if $ad(a,b,c,d) = 0$. Moreover ad differentiates two types of cases where analogy does not hold, namely the 8 cases with an odd number of 0 and an odd number of 1 among the 4 Boolean values, such as $ad(0,0,0,1) = 1$ or $ad(0,1,1,1) = 1$, and the two cases $ad(0,1,1,0) = ad(1,0,0,1) = 2$.

When, instead of having 4 Boolean values, we deal with 4 Boolean vectors in \mathbb{B}^n , we add the ad evaluations componentwise to get the analogical dissimilarity between 4 Boolean vectors, which leads to an integer belonging to the interval $[0, 2n]$. This number is a numerical measure of how far the 4 vectors are from building, componentwise, a complete analogy. It has been used in [2] to implement a classification algorithm where we have as input a training set S of classified items, a new item \vec{d} to be classified, and an integer k . The algorithm proceeds as follows:

Step 1: Compute the analogical dissimilarity ad between \vec{d} and all the triples in S^3 that produce a solution for the class of \vec{d} .

Step 2: Sort these n triples by the increasing value of ad wrt with \vec{d} .

Step 3: If the k -th triple has the integer value p for ad , then let k' be the greatest integer such that the k' -th triple has the value p .

Step 4: Solve the k' analogical equations on the label of the class. Take the winner of the k' votes and allocate this winner as the class of \vec{d} .

This simple approach provides remarkable results and, in most of the cases, outperforms the best known algorithms (see [6]). Our approach is a bit different and we consider a simpler analogical dissimilarity that we denote ad^* which, on binary values a, b, c, d , is equal to 0 when $A(a, b, c, d)$ holds, and is 1 otherwise. If we consider $A(a, b, c, d)$ as a numerical value 0 or 1, the function:

$$ad^*(a, b, c, d) = (a - b - c + d)^2)^{1/2}, \text{ for } a, b, c, d \in \{0, 1\}, \quad (5)$$

it is nothing more than $ad^*(a, b, c, d) = 1 - A(a, b, c, d)$. Of course, other functions can be used and can provide the same results, as for instance: $ad^{*1}(a, b, c, d) = |a - b - c + d|$.

It means that we do not differentiate the two types of cases where analogy does not hold. Considering Boolean vectors in \mathbb{B}^n , we add the ad^* evaluations componentwise, to get a final integer belonging to the interval $[0, n]$ (instead of $[0, 2n]$ in the previous case). Namely,

$$\text{ad}^*(\vec{a}^k, \vec{b}^k, \vec{c}^k, \vec{d}) = \sum_{i=1,n} \text{ad}^*(a_i^k, b_i^k, c_i^k, d_i). \quad (6)$$

Our algorithm is as follows, taking as input a training set S , the item to be classified \vec{d} , and a integer p , with $0 \leq p \leq n$, which means for each quadruple analyzed, we accept at most p features for which the considered proportion does not hold true.

Step 1: Compute $\text{ad}^*(\vec{a}^k, \vec{b}^k, \vec{c}^k, \vec{d})$ between \vec{d} and all triples $(\vec{a}^k, \vec{b}^k, \vec{c}^k) \in S^3$.

Step 2: If $\text{ad}^*(\vec{a}^k, \vec{b}^k, \vec{c}^k, \vec{d}) \leq p$, solve the class analogical equation. If there is a solution, consider this solution as a vote for the predicted class of \vec{d} .

Step 3: Take the winner of the votes and allocate it as the class of \vec{d} .

In terms of worst case complexity, our algorithm is still cubic, due to step 1. But, it avoids the sorting step of the previous algorithm, leading to a lower complexity than the one of [2]. These options are implemented and the results are shown below.

4 First experiments and discussion

To ensure a fair comparison with the approach in [2], we use the same first 4 datasets with binary and nominal attributes, all coming from the University of California at Irvine repository [1]. We consider the following datasets.

- MONK 1,2 and 3 problems (denoted MO.1, MO.2 and MO.3). MONK3 problem has noise added.
- SPECT heart data (SP. for short).

To binarize, we use the standard technique, replacing a nominal attribute having n different values with n binary attributes. We summarize the datasets structure and our results in Table 3.

As can be seen, the results obtained are very similar to those obtained in [6], although we are using a counterpart ad^* of the logical view of the analogical proportion for estimating the analogical dissimilarity of four vectors, instead of the more discriminating measure introduced in [6]. Note that for very small values of p some instances \vec{d} may not find a triple for predicting a class for them. This may lead to a low accuracy, which then increases when p is less requiring. For too large values of p then the accuracy generally decreases since the number of features for which an analogical proportion holds becomes too small and then we allow “poor quality” triples to vote.

Besides, instead of using $\text{ad}^*(a, b, c, d) = 1 - A(a, b, c, d)$, one may replace A by reverse analogy R , or paralogy P , or inverse paralogy I . We have checked that the results remain the same if we use R , or P instead of A . This is in perfect agreement with the existence of permutations changing one proportion into another as highlight in equation (4).

Table 3 Results

Information	MO.1	MO.2	MO.3	SP.
nb. of nominal attributes	7	7	7	22
nb. of binary attributes	15	15	15	22
nb. of training instances	124	169	122	80
nb. of test instances	432	432	432	172
nb. of class	2	2	2	2
Our algorithm accuracy (best results)	97.22	99.77	97.22	57.22
Best value of p	1	0	2	4
Miclet <i>et al.</i> best result [6] ^a	98	100	96	58
Decision Table	97.22	66.21	97.22	71.12
PART	92.59	74.31	98.15	81.28
Multilayer Perceptron	100.00	100.00	93.52	72.73
Logistic	71.06	61.57	97.22	66.31
BayesNet	71.06	61.11	97.22	75.94

^a Results were provided by [6] as presented here.

However, when considering $\text{ad}^*(a, b, c, d) = 1 - I(a, b, c, d)$, first results indicate that the accuracy decreases substantially. This does not really come as a surprise, since looking at the truth table of I , we then allow the simultaneous presence of patterns such as $(1, 0, 1, 0)$ and $(1, 0, 0, 1)$, in complete disagreement with the measure of analogical dissimilarity ad , even if I would apparently agree with the idea that objects that share (almost) nothing should belong to different classes. Ultimately, when a triple (a, b, c) belongs to S^3 , it does not imply that the triple (a, \bar{b}, \bar{c}) belongs to S^3 , and then, thanks again to equation 4, the set of voters for I may be different from the set of voters of A , R and P , for which the set is the same.

We made comparisons with some classifiers found in the literature and for these ones, we used the Weka package [4] with default values. It is possible to note the Multilayer Perceptron achieved the best accuracy for MO.1 and MO.2 databases, but its performance is lower for the MO.3 database. PART classifier provides best classification for MO.3 and SP. databases, but its performance is not so good for MO.2. Decision Table provides good classifications for MO.1 and MO.2 databases. Similar results were found by [6]. Logistic and BayesNet classifiers have good performance only for MO.3 database. So, for these four databases and this set of classifiers, only our algorithm and [6] achieved good results for three databases. It shows that our algorithm is a competitive approach for classification tasks of databases with binary and nominal features.

5 Concluding remarks

This preliminary study shows that the logical view of analogical proportion leads to good results in classification. More experiments currently under progress should

bring more light on the behaviour of the classification process with regard to p . Using the same type of algorithm and the extension to multi-valued logics as done in [11] will allow us to deal with numerical values and continuous attributes.

The algorithm, when compared with others found in the literature, provides competitive results for three of the four benchmarks considered. New comparisons with other classifiers and other databases are under progress to better understand the applicability and the competitiveness of the approach.

Acknowledgements This work is supported by CNPq, processes 246939/2012-5, 246938/2012-9, 310561/2012-4, 310470/2012-9 and INCT-MACC (process 181813/2010-6).

References

1. Bache, K., Lichman, M.: UCI machine learning repository (2013). URL <http://archive.ics.uci.edu/ml>
2. Bayouhd, S., Miclet, L., Delhay, A.: Learning by analogy: A classification rule for binary and nominal data. Proc. Inter. Conf. on Artificial Intelligence IJCAI07 pp. 678–683 (2007)
3. Gentner, D., Holyoak, K.J., Kokinov, B.N.: The Analogical Mind: Perspectives from Cognitive Science, Cognitive Science, and Philosophy. MIT Press, Cambridge, MA (2001)
4. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: An update. SIGKDD Explorations **11**(1), 10 – 18 (2009)
5. Melis, E., Veloso, M.: Analogy in problem solving. In: Handbook of Practical Reasoning: Computational and Theoretical Aspects. Oxford Univ. Press (1998)
6. Miclet, L., Bayouhd, S., Delhay, A.: Analogical dissimilarity: definition, algorithms and two experiments in machine learning. JAIR, 32 pp. 793–824 (2008)
7. Miclet, L., Prade, H.: Handling analogical proportions in classical logic and fuzzy logics settings. In: Proc. 10th Eur. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'09), Verona, pp. 638–650. Springer, LNCS 5590 (2009)
8. Prade, H., Richard, G.: Logical proportions - typology and roadmap. In: E. Hüllermeier, R. Kruse, F. Hoffmann (eds.) Computational Intelligence for Knowledge-Based Systems Design: Proc. 13th Inter. Conf. on Information Processing and Management of Uncertainty (IPMU'10), Dortmund, June 28 - July 2, LNCS, vol. 6178, pp. 757–767. Springer (2010)
9. Prade, H., Richard, G.: Reasoning with logical proportions. In: Proc. 12th Inter. Conf. on Principles of Knowledge Representation and Reasoning (KR'10), Toronto, Ontario, Canada, May 9-13, (F. Z. Lin, U. Sattler, M. Truszczynski, eds.), pp. 545–555. AAAI Press (2010)
10. Prade, H., Richard, G.: Homogeneous logical proportions: Their uniqueness and their role in similarity-based prediction. In: G. Brewka, T. Eiter, S.A. McIlraith (eds.) Proc. 13th Inter. Conf. on Principles of Knowledge Representation and Reasoning (KR'12), Roma, June 10-14, pp. 402–412. AAAI Press (2012)
11. Prade, H., Richard, G.: Analogical proportions and multiple-valued logics. In: L.C. van der Gaag (ed.) Proc. 12th Europ. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'13), Utrecht, July 8-10, LNCS, vol. 7958, pp. 497–509. Springer (2013)
12. Prade, H., Richard, G., Yao, B.: Enforcing regularity by means of analogy-related proportions—a new approach to classification. International Journal of Computer Information Systems and Industrial Management Applications **4**, 648–658 (2012)
13. Stroppa, N., Yvon, F.: Analogical learning and formal proportions: Definitions and methodological issues. ENST Paris report (2005)