Supervised Classification Using Homogeneous Logical Proportions for Binary and Nominal Features

Ronei Moraes, Liliane Machado, Henri Prade, Gilles Richard

Supervised Classification Using Homogeneous Logical Proportions for Binary and Nominal Features

Ronei M. Moraes[1,2], Liliane S. Machado[1,2], Henri Prade[2], and Gilles Richard[2]

[1] LabTEVE Federal University of Paraiba,
Cidade Universitaria s/n, Joao Pessoa, Brazil
[2] IRIT University of Toulouse, 118 route de Narbonne,
31062 Toulouse Cedex 09, France
ronei@de.ufpb.br, liliane@di.ufpb.br, {prade,richard}@irit.fr

Abstract. The notion of homogeneous logical proportions has been recently introduced in close relation with the idea of analogical proportion. The four homogeneous proportions have intuitive meanings, which can be related with classification tasks. In this paper, we proposed a supervised classification algorithm using homogeneous logical proportions and provide results for all. A final comparison with previous works using similar methodologies and with other classifiers is provided.

Keywords: supervised classification, analogical proportion, analogical dissimilarity.

1 Introduction

Numerical as well as analogical proportions are used since the ancient Greeks. However, this is only recently that logical models for analogical proportions were laid bare [14]. *Analogical proportion* was proposed first and *reverse analogical proportion* and *paralogical proportion* came after (proposed for Boolean [10] and multiple-valued [12] features). The last logical proportion we consider here, named *inverse paralogical proportion*, and the characterization of those four proportions as homogeneous logical proportions were presented recently [14] for the Boolean case.

A particular instance of analogical reasoning is based on the notion of analogical proportion (or analogy, or for short A) linking four situations or items a, b, c, d. It focuses on the differences between a and b and between c and d that are the same. The reverse analogical proportion (or reverse analogy, or for short R) does the same, but reverses the directions of changes, i.e. differences between a and b, are the same as between d and c. The paralogical proportion (or paralogy, or for short P) focuses on similarities, unlike previous proportions. It points out that similarities between a and b and between c and d are the same. The inverse paralogical proportion (or inverse paralogy, or for short I) focuses on similarities also, and expresses that what a and b have in common, c and

d do not have it, and vice versa. Those proportions have intuitive meanings, which can be related with classification tasks. However, in the recent literature, only one (analogical proportion) was studied for classification tasks and has been shown as a competitive supervised classifier [7], which, from descriptions of known objects (their features and classes), classifies another object, whose class is unknown. This reasoning is similar to the one used by k-NN classifiers [3] or in case-based reasoning [5]. For instance, in general pattern recognition terms, and using analogical proportions, from three objects (obj1, obj2, obj3), whose descriptions are made of a finite number of features (desc1, desc2, desc3), we want to classify a fourth object (obj4), with description desc4. Then, if an analogical proportion A(desc1, desc2, desc3, desc4) holds, it may be possible to suggest a class for obj4. Besides, an analogical dissimilarity (AD) measure was proposed by [2], which can be used on Boolean and also multi-valued features. This measure is able to assess the truthfulness of an analogical proportion [7].

In this paper, using a modified version of analogical dissimilarity, we provide some relations among analogy, reverse analogy, paralogy and inverse paralogy. It is shown that, using those relations, all those proportions can be computed in a simplified way and an algorithm for supervised classification is provided. Moreover, we present some results for binary and/or nominal classification tasks, using databases from UCI repository and also a comparison with other methodologies.

2 Characterization of Homogeneous Proportions

Analogy, reverse analogy, paralogy and inverse paralogy are formal relations T involving 4 items a, b, c and d. In this study, T is a Boolean proportion. This means that a, b, c, d are Boolean variables and can be pointwisely generalized by vectors of Boolean variables. A detailed investigation with respect to the basic semantics has been done in [10] [11] [14]. They obey different characteristic postulated [12]. For $a, b, c, d \in \{0, 1\}$ the considered properties are: a) *Reflexivity*, which validates the proportion $T(a, b, a, b)$; b) *Reverse reflexivity*, which validates the proportion $T(a, b, b, a)$; c) *Identity*, which validates the proportion $T(a, a, b, b)$; d) *Full identity*, which validates the proportion $T(a, a, a, a)$; e) *Symmetry*, which requires the equivalence between $T(a, b, c, d) \Leftrightarrow T(c, d, a, b)$; f) *Central permutation*, which requires the equivalence between $T(a, b, c, d) \Leftrightarrow T(a, c, b, d)$; g) *Code independency*, which requires the equivalence between $T(a, b, c, d) \Leftrightarrow T(\neg a, \neg b, \neg c, \neg d)$. It should be noted that all homogeneous proportions satisfy the symmetry and code independency properties [14]. However, some of the other properties are not satisfied by all four proportions (see Table 1). As shown in [13], there is close relation among three of those proportions:

Proposition 1: $R(a, b, c, d)$ is a reverse analogy if and only if $A(a, b, d, c)$ is an analogy, i. e. $R(a, b, c, d) \Leftrightarrow A(a, b, d, c)$; $P(a, b, c, d)$ is a paralogy if and only if A(a, d, c, b) is an analogy, i. e. $P(a, b, c, d) \Leftrightarrow A(a, d, c, b)$.

Prade and Richard [14] established further relations among those four proportions, with respect to analogy, through permutations and negation:

Proposition 2: $R(a, b, c, d)$ is a reverse analogy if and only if $A(a, b, \neg c, \neg d)$ is an analogy, i. e. $R(a, b, c, d) \Leftrightarrow A(a, b, \neg c, \neg d)$; $P(a, b, c, d)$ is a paralogy if and only if $A(a, \neg b, c, \neg d)$ is an analogy, i. e. $P(a, b, c, d) \Leftrightarrow A(a, \neg b, c, \neg d)$; $I(a, b, c, d)$ is an inverse paralogy if and only if $A(a, \neg b, \neg c, d)$ is an analogy, i. e. $I(a, b, c, d) \Leftrightarrow A(a, \neg b, \neg c, d)$.

Table 1. Homogeneous proportions and their properties

Properties	A	R	P	I
Reflexivity	✓	–	✓	–
Reverse reflexivity	–	✓	✓	–
Identity	✓	✓	–	–
Full identity	✓	✓	✓	–
Symmetry	✓	✓	✓	✓
Central permutation	✓	–	–	✓
Code independency	✓	✓	✓	✓

To introduce formally the four proportions, let us consider the items a, b, c, d as described by sets of binary features, which belong to the universe X, i. e. each item is viewed as a subset of X. An analogical proportion, denoted by $A(a, b, c, d)$ focuses on the differences and should hold when the differences between a and b, and c and d are the same [8]:

$$a \wedge \neg b = c \wedge \neg d \text{ and } \neg a \wedge b = \neg c \wedge d \tag{1}$$

Reverse analogy, denoted by $R(a, b, c, d)$ exchanges c and d, with respect to analogy and expresses that the changes from a to b in relation to c to d (if any) are now in opposite directions:

$$a \wedge \neg b = \neg c \wedge d \text{ and } \neg a \wedge b = c \wedge \neg d \tag{2}$$

If instead of differences, we focus on similarities, we have a different proportion, denoted by $P(a, b, c, d)$ and named paralogy, which expresses that a and b have in common, c and d have it too:

$$a \wedge b = c \wedge d \text{ and } \neg a \wedge \neg b = \neg c \wedge \neg d \tag{3}$$

The inverse paralogy [14], denoted by $I(a, b, c, d)$ focuses on similarities also and expresses a and b have in common, c and d do not have it, and vice versa:

$$a \wedge b = \neg c \wedge \neg d \text{ and } \neg a \wedge \neg b = c \wedge d \tag{4}$$

Table 2 presents the Boolean truth table for each logical proportion presented above. We can note that there are only 6 situations in which the logical value is true for each proportion. Observing this Table, it is easy to see that there is no situation where analogy, reverse analogy, paralogy and inverse paralogy hold true together. However, there are some cases in which all proportions hold false. These relations will be discussed in details in Section 4, after we provide some relevant details about the measure of analogical dissimilarity.

Table 2. Boolean truth tables for Proportions: A, R, P and I:

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

These four logical proportions are called homogeneous proportions because they are true only for patterns having an even number of 1 (and thus an even number of 0), due to the fact they are strongly linked together by the relations:

$$A(a, b, c, d) \equiv R(a, b, d, c) \equiv P(a, d, c, b) \equiv I(a, \overline{b}, \overline{c}, d), \tag{5}$$

Their semantical properties have been extensively investigated in [14].

3 Analogical Dissimilarity Measure and an Approximation Function

An analogical dissimilarity (AD) measure was proposed by [6] for binary and by [2][7] for nominal data, using a binary encoding. So, for both cases the binary definition is appropriate and for this reason we present the definition for AD only for this case.

Definition 1. The analogical dissimilarity among four binary values ($u, v, w, x \in \{0, 1\}$) is given by the following truth table [7]:

u	0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1
v	0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1
w	0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1
x	0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1
$AD(u, v, w, x)$	0 1 1 0 1 0 2 1 1 2 0 1 0 1 1 0

As can be seen, AD is a function whose domain is $\{0, 1\}^4$, but its range is $\{0, 1, 2\}$. This is not completely in agreement with the homogeneous logical proportions, whose domains and ranges are Boolean. However, AD is consistent with the analogical proportion when we have $AD(u, v, w, x) = 0$ [6].

Definition 2. The AD measure among four objects ($u, v, w, t \in \{0, 1\}^n$) of a finite set X defined by binary features is the sum of the values of analogical dissimilarities for each feature [7].

Using the Definition 2 and encoding nominal variables as binary ones, it is possible to use AD for nominal variables as well. In [6][7] procedures to compute binary analogical proportions on $\mathbb{R}^n$ and on sequences, are presented. However, in order to obtain an efficient algorithm for AD computing that fits with all the homogeneous proportions mentioned in Section 2, it is necessary to use a mathematical function which approximates and replaces the AD definition.

The algorithm should also use the relationships given above in order to provide the values of all proportions from the values given by computing an approximation function for AD for each one. Thus, let the function $AD^* : \{0,1\}^4 \Rightarrow \{0,1\}$:

$$AD^*(a,b,c,d) = \left[(a-b-c+d)^2\right]^{1/2}, \text{ for } a,b,c,d \in \{0,1\} \qquad (6)$$

be an approximation for computing the function AD, according to Definition 1. Obviously, AD^* can also be written: $AD^*(a,b,c,d) = |a-b-c+d|$. That function has interesting properties: The codomain for AD^* is $\{0,1\}$. Using the binary values $\{0,1\}$ as truth values, the function AD^* given by (6) is able to provide exact estimation for 14 values among 16 values of the AD table for the binary case. The exceptions are the cases in which $AD = 2$. Moreover, even if AD^* is less discriminating, it is closely related to the value of the proportions.

4 Relations

As mentioned before, each logical proportion holds true in only 6 situations, which can be seen in the Table 2. However, they are not true in the same situations. As pointed out by [13], in the Boolean interpretation and according to their intuitive meaning, the pattern $T(a,b,b,a)$ is false for analogy; $T(a,b,a,b)$ is false for reverse analogy, T(a,a,b,b) is false for paralogy and $T(a,a,a,a)$ is false for inverse paralogy. For any of these proportions in the binary case, it is possible to find a relationship between the values provided by the proportions and the value provided by AD measure. In the case of Analogy and AD, it was announced as a property of AD [6]. More precisely, we have $AD = 0 \Leftrightarrow AD^* = 0$ and $AD \in \{1,2\} \Leftrightarrow AD^* = 1$, and besides, we have the relationship with respect to analogical proportion:

$$A(a,b,c,d) = 1 - AD^*(a,b,c,d) \qquad (7)$$

For the Reverse Analogy and Paralogy we use Proposition 1 for finding the following relations:

$$R(a,b,c,d) = 1 - AD^*(a,b,d,c) \qquad (8)$$

$$P(a,b,c,d) = 1 - AD^*(a,d,c,b) \qquad (9)$$

The same can be done for Inverse Paralogy, using Proposition 2:

$$I(a,b,c,d) = 1 - AD^*(a,\neg b,\neg c,d) \qquad (10)$$

These relations allow us to implement an algorithm using AD^* for computing to what extent the four proportions hold, in a new and simpler way than in [2].

5 Algorithm for Supervised Classification Using AD^*

The algorithm presented in this paper partly relies on a previous work [2]. The first difference of what is proposed here is the possibility to use any of the four proportions mentioned above. The second one is a change the ordering of execution of the steps of the previous algorithm, in order to take advantage of AD^* and to avoid the step of sorting the partial results. It improves the computational performance of the proposed classifier, with respect to [2]. It is important to remark that the algorithm proposed by [2] uses AD to implement analogical proportion only. The task is to classify m objects (obj_i), whose descriptions are made in terms of a finite number N_f of binary features ($desc_i$), where $i = 1, ..., m$, using AD^*, into a finite and known number of classes in set C. A collection S of such objects is available, with their descriptions and respective classes in C, for the training of the classifier. Let x be a new object, not belonging to set S, and for which we desire to assign a class. The algorithm depends on a parameter p, with $0 \leq p \leq N_f$, which means that for each 4-tuple of description vectors analyzed, we accept that the proportion is false for at most p features. In other words, among the total number of features N_f, it is intended that for each 4-tuple (triple in S^3 and x) analyzed, at least $(N_f - p)$ features yield perfect proportions. We call this rule "*Rule of maximum p dissimilar 4-tuples*". The algorithm consists in the following steps:

Step 1: Given a vector x, find all triples in S^3 which satisfies the sum of $AD^* \leq p$ and $0 \leq p \leq N_f$. Store those n triplets (a_j, b_j, c_j), with $j = 1, ..., n < m$).

Step 2: Solve the proportion equations (for A, R, P or I) on the label of the class of x and the n triples. Compute the solution as a vote for a class $c \in C$, obtained by solving the logical proportion on the classes (when there is solution, which means that the objects corresponding to a triple should belong to 1 or 2 classes only in order to have a useful triple for predicting the class).

Step 3: Assign the winner of votes among the n results, as the class for x.

It is worth noting when using AD^* and changing one step in the algorithm of [2], that it is no longer necessary to sort partial results anymore. So, the parameter k used in the algorithm in [2] does not have any role here. A new parameter p was added in the Step 1 to control the maximum number of dissimilar 4-tuples which the user can accept in each processing.

6 Results

In order to analyze the behaviour of the proposed algorithm, we applied it to some databases from UCI Repository [1]. These databases were used also by [7] and for this reason are used here to provide a comparison with the results obtained in this paper: The SPECT (SP.) database is related to heart data, with 22 binary features and contained with 80 samples for training and 187 for test. The databases MONK 1,2 and 3 (MO.1, MO.2 and MO.3) are related to classification problems with six nominal features. The databases contains 124, 169 and 122 samples for training, respectively and all of them 432 samples for test. The MO.3 database is corrupted with noise. All three MONK databases were binarized using standard techniques, i. e., replacing a nominal feature with n different values with n binary features. The results are summarized in Table 3. The first five lines give some characteristics of the databases used. In the following we indicate the best percentage of correct classification obtained for each proportion and for which value of p this is obtained. The last line shows the best results presented by [7] for comparison. It is worth noting in Table 3, all results for A, R and P are the same for all databases and all of them using the same value for p. In fact this is due to a property of three of the homogeneous proportions (A, P and R) which can be seen in equation (5). The difference among them is just a permutation of the elements in the 4-tuple. So, any of these three proportions proportions can be used for classification and it provides the same final results. However, this is not valid for proportion I, because beyond the permutation of elements, it is necessary also to perform negation for two of them [9].

Table 3. Results

	$SP.$	$MO.1$	$MO.2$	$MO.3$
number of nominal attributes	22	7	7	7
number of binary attributes	22	15	15	15
number of training instances	80	169	122	124
number of test instances	172	432	432	432
number. of class	2	2	2	2
Best results (%)				
Analogy (A):	$58(p = 5)$	$98(p = 2)$	$100(p = 1)$	$98(p = 3)$
Reverse Analogy (R):	$58(p = 5)$	$98(p = 2)$	$100(p = 1)$	$98(p = 3)$
Paralogy (P):	$58(p = 5)$	$98(p = 2)$	$100(p = 1)$	$98(p = 3)$
Inverse Paralogy (I):	$11(p = 8)$	$79(p = 8)$	$62(p = 6)$	$83(p = 6)$
Comparisons (%)				
Miclet *et al.* [7]	$58(k = 100)$	$98(k = 100)$	$100(k = 100)$	$96(k = 100)$
Decision Table	72	98	67	98
PART	82	93	75	99
Multilayer Perceptron	73	100	100	94
IBk (k=1)	59	91	69	89
Naive Bayes	75	72	61	98
J48 (C4.5)	76	100	71	100

It is possible to see in Table 3, that proportion I achieved results of lower quality than the other proportions and with values of p higher. In comparison with results provided by [7], for three among the four databases, the results obtained by the new algorithm were the same. However, the new algorithm provided better results for the MO.3 database, which is corrupted by noise. We made some comparisons with classifiers found in the literature and we used the Weka package [4] with default values. It is possible to note the Multilayer Perceptron provides best classification for MO.1 and MO.2. J48 Decesion Tree achieved the complete accuracy for MO.1 and MO.3. PART provides best classification for SP. However, our algorithm provides competitive results for three of those databases.

7 Conclusion and Future Works

This paper proposes a new algorithm for supervised classification using homogeneous logical proportions. The algorithm was presented and we provided a short discussion in relation to the previous one proposed by [2], which implemented only one of those proportions. Implementation results are presented for all homogeneous proportions, as well as a final comparison with a previous paper using similar methodologies and other classifiers too. The results achieved showed that proportions A, R and P provide the same results when it is used with the same value for p. This way, they yielded the same best results in comparison with results reported in [7] (which used AD and analogical proportion). The new algorithm provides results as good as the previous one or better, as it was the case of MO.3 (which is a database corrupted with noise).

As future works, we intent to improve comparisons using other classifiers and by considering more databases in order to extend our study about the relative performance of our approach with respect to other classifiers. Another issue is to determine how value p influences classification results.

Acknowledgement. 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, `http://archive.ics.uci.edu/ml` (2013)
2. Bayoudh, S., Miclet, L., Delhay, A.: Learning by analogy: A classification rule for binary and nominal data. In: Proc. Inter. Conf. on Artificial Intelligence, IJCAI 2007, pp. 678–683 (2007)
3. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley-Interscience (2001)
4. Hall, M., et al.: The Weka data mining software: An update. SIGKDD Explorations 11, 10–18 (2009)
5. Hüllermeier, E.: Case-Based Approximate Reasoning. Theory and Decision Library. Springer, New York (2007)
6. Miclet, L., Delhay, A.: Analogical Dissimilarity: definition, algorithms and first experiments in machine learning. Technical Report 5694, IRISA (September 2005)
7. Miclet, L., Bayoudh, S., Delhay, A.: Analogical dissimilarity: definition, algorithms and two experiments in machine learning. JAIR 32, 793–824 (2008)
8. Miclet, L., Prade, H.: Handling analogical proportions in classical logic and fuzzy logics settings. In: Sossai, C., Chemello, G. (eds.) ECSQARU 2009. LNCS, vol. 5590, pp. 638–650. Springer, Heidelberg (2009)
9. Moraes, R.M., Machado, L.S., Prade, H., Richard, G.: Classification based on homogeneous logical proportions. Proc. 33th Int. Conf. on Artificial Intelligence (AI 2013), Cambridge (to appear, 2013)
10. Prade, H., Richard, G.: Analogy, paralogy and reverse analogy: Postulates and inferences. In: Mertsching, B., Hund, M., Aziz, Z. (eds.) KI 2009. LNCS, vol. 5803, pp. 306–314. Springer, Heidelberg (2009)

11. Prade, H., Richard, G.: Analogical proportions: another logical view. In: Proc. 29th Int. Conf. Artif. Intellig (AI 2009), Cambridge, pp. 121–134. Springer (2009)
12. Prade, H., Richard, G.: Multiple-valued logic interpretations of analogical, reverse analogical, and paralogical proportions. In: Proc. 40th IEEE International Symp. on Multiple-Valued Logic (ISMVL 2010), Barcelona, pp. 258–263 (2010)
13. Prade, H., Richard, G., Yao, B.: Enforcing regularity by means of analogy-related proportions - A new approach to classification. Int. Jour. Computer Information Systems and Industrial Management Applications 4, 648–658 (2012)
14. Prade, H., Richard, G.: Homogeneous Logical Proportions: Their Uniqueness and Their Role in Similarity-Based Prediction. In: Proc. XIII International Conference on Principles of Knowledge Representation and Reasoning, pp. 402–412 (2012)