Learning by Analogy : a Classification Rule for Binary and Nominal Data

Sabri Bayoudh and **Laurent Miclet** and **Arnaud Delhay**
IRISA/CORDIAL – Universit de Rennes 1, France
sabri.bayoudh, laurent.miclet, arnaud.delhay @enssat.fr

Abstract

This paper deals with learning to classify by using an approximation of the analogical proportion between four objects. These objects are described by binary and nominal attributes. Firstly, the paper recalls what is an analogical proportion between four objects, then it introduces a measure called ”analogical dissimilarity”, reflecting how close four objects are from being in an analogical proportion. Secondly, it presents an analogical instance-based learning method and describes a fast algorithm. Thirdly, a technique to assign a set of weights to the attributes of the objects is given: a weight is chosen according to the type of the analogical proportion involved. The weights are obtained from the learning sample. Then, some results of the method are presented. They compare favorably to standard classification techniques on six benchmarks. Finally, the relevance and complexity of the method are discussed.

Keywords : *instance-based learning, learning by analogy, analogical proportion, analogical dissimilarity.*

1 Introduction

The aim of this article is to present a new non-parametric classification rule for objects described by binary and nominal attributes. Some definitions and original algorithms will be presented to explain this method, and results will be given. The principle of non-parametric (or *instance-based*) learning is to keep the whole learning set as a base for the decision instead of inducing an explicit classification concept. The best known and simplest instance-based classification rule is the k-nearest neighbors (*k-nn*) rule, which requires only a metric space. This rule, in its simplest form, computes the distance between the object to classify and all the instances in the learning set and considers the k nearest instances as electors of their class [DUDA *et al.*, 2001]. The class of the new object is the class which has the majority given by the k electors.

Learning by analogy, as we present it here, is also an instance-based technique which uses the concept of *analogical dissimilarity* between four objects (three in the training set, the fourth being the object to classify). One property of the analogical dissimilarity between four objects is that it is null when the objects are in *analogical proportion*. Let us take a toy example to explain what is an analogical proportion and how to learn with it. Let *tomcat* be the object to classify into *Feline* or *Ruminant*. Suppose that there are only three instances in the training set: *calf*, *bull* and *kitten*. The animals are described by four binary attributes (1 stands for $TRUE$ and 0 for $FALSE$) which are *has claws(HC)*, *is an adult(IA)*, *is a male(IM)*, *feeds by suckling(FS)*.

Animals	*HC*	*IA*	*IM*	*FS*	class
calf	0	0	0	1	*Ruminant*
bull	0	1	1	0	*Ruminant*
kitten	1	0	0	1	*Feline*
tomcat	1	1	1	0	?

For each attribute, we notice that there is an analogical proportion between the three objects, taken in this order, from the training set and the object *tomcat*, since the four binary values have one of the forms below (we will see later what are all the possible analogical proportions between binary values).

0 is to 0 as 1 is to 1 (*HC*)
0 is to 1 as 0 is to 1 (*IA* and *IM*)
1 is to 0 as 1 is to 0 (*FS*)

Thus, the four objects are in analogical proportion since all their binary attributes are in analogical proportion :

calf **is to** *bull* **as** *kitten* **is to** *tomcat*

We have to emphasize two points here. Firstly, taking these three objects (triplet) in a different order will not produce anymore an analogical proportion with *tomcat*, except when *bull* and *kitten* are exchanged. Secondly, to make full use of the training set, we must consider also the triplets containing two or three identical objects. Consequently, a training set of m objects produces m^3 different triplets. In this example, it is easy to verify that only two triplets among the 27 are in analogical proportion with *tomcat*.

For the triplet (*calf*, *bull*, *kitten*), the resolution of an analogical equation on the classes gives the class of *tomcat*:

Ruminant **is to** *Ruminant* **as** *Feline* **is to** *class(tomcat)*

which solves in : *class(tomcat) = Feline*.

The second triplet (*calf*, *kitten*, *bull*) in analogy with *tomcat* gives the same result. Hence, the classification by analogy of *tomcat* on this learning set is *Feline*.

Notice that the 1-nearest neighbor technique, using the Hamming distance, concludes in this example that *class(tomcat) = Ruminant*. The strength of the analogical classifier stems from the fact that it can use objects in the training set belonging to some class ω_1 to conclude that the unknown class object is ω_2. In terms of distance, an analogical classifier can give importance to objects that are far from the object to be classified. It actually makes a different use of the information in the training set than distance classifiers or decision trees, because it uses a (restricted) form of analogical reasoning [GENTNER *et al.*, 2001]. Is the use of analogical proportions actually relevant for classification? The goal of this article is to investigate in this direction, for data sets described by binary and nominal data.

In Section 2, we give a formal definition of the notion of analogical proportion and we introduce that of analogical dissimilarity. In Section 3, we present a naive algorithm of learning a classification rule followed by a faster version, which makes use of the properties of analogical dissimilarity. In Section 4, we propose to learn from the set of instances three different weights for each attribute, and to use one of them to modify the computation of the analogical dissimilarity (a weight is chosen among the three according to the type of analogical proportion which is involved). We give results on eight data bases from the UCI ML repository [NEWMAN *et al.*, 1998] and compare with various standard classification algorithms in Sect. 5. Finally, Section 6 discusses these results and the complexity of the algorithm.

2 Analogical Proportion and Analogical Dissimilarity

2.1 Analogical Proportion between Four Binary Objects

Generally speaking, an *analogical proportion* on a set X is a relation between four elements of X, i.e. a subset of X^4, which writes $a : b :: c : d$ and reads:

$$a \textbf{ is to } b \textbf{ as } c \textbf{ is to } d$$

The elements a, b, c and d are said to be *in analogical proportion* [DELHAY and MICLET, 2005]. As defined by Lepage [LEPAGE and ANDO, 1996][LEPAGE, 1996], an analogical proportion $a : b :: c : d$ implies two other analogical proportions:

Symmetry of the "as" relation: $c : d :: a : b$

Exchange of the means: $a : c :: b : d$

And the following property is also required :

Determinism: if $a : a :: b : x$ *then* $x = b$

From the first two properties one can deduce five more equivalent analogical proportions, which gives eight equivalent ways to write that the objects a, b, c and d are in analogical proportion :

	$c : a :: d : b$
$c : d :: a : b$	$d : b :: c : a$
$a : c :: b : d$	$d : c :: b : a$
$a : b :: c : d$	$b : a :: d : c$
	$b : d :: a : c$

When X is the boolean set $\mathbb{B} = \{0,1\}$ there are only six 4-tuples in analogical proportion among the sixteen possibilities (and two special cases *(a)* and *(b)* in the other ten, as we shall see in the next paragraph):

In analogical proportion	Not in analogical proportion	
0 : 0 :: 0 : 0	0 : 0 :: 0 : 1	
0 : 0 :: 1 : 1	0 : 0 :: 1 : 0	
0 : 1 :: 0 : 1	0 : 1 :: 0 : 0	
1 : 0 :: 1 : 0	0 : 1 :: 1 : 0	*(a)*
1 : 1 :: 0 : 0	0 : 1 :: 1 : 1	
1 : 1 :: 1 : 1	1 : 0 :: 0 : 0	
	1 : 0 :: 0 : 1	*(b)*
	1 : 0 :: 1 : 1	
	1 : 1 :: 0 : 1	
	1 : 1 :: 1 : 0	

Now, if we take objects described by m binary attributes, we can easily construct an analogical proportion on $\mathbb{B}^m$.

Definition 1 *Four objects in $\mathbb{B}^m$ are in analogical proportion if and only if all their attributes are in analogical proportion:*

$$a : b :: c : d \Leftrightarrow a_j : b_j :: c_j : d_j \quad \forall\, 1 \leq j \leq m$$

With this definition, it is straightforward to verify that the above properties of analogical proportion are still verified.

2.2 Analogical Dissimilarity between Binary Objects

Up to now, four objects are either in analogical proportion or not. Here we introduce a new notion to measure, when they are not in analogical proportion, how far are four objects from being in analogical proportion. This measure is called Analogical Dissimilarity (AD). Another measure has already been presented in the literature, for natural langage processing, in the framework of semantic analogy [TURNEY, 2005] by considering similarity relations between two pairs of words.

In $\mathbb{B}$, we give the value of 1 to every four-tuple which is not in analogical proportion, except for *(a)* and *(b)*, which take the value of 2. 1 and 2 are the number of binary values to switch to produce an analogical proportion.

This definition leads to the following properties:

Property 1

1. $\forall u, v, w, x \in \mathbb{B},\quad AD(u, v, w, x) = 0 \Leftrightarrow u : v :: w : x$
2. $\forall u, v, w, x \in \mathbb{B}$,
   $AD(u, v, w, x) = AD(w, x, u, v) = AD(u, w, v, x)$
3. $\forall u, v, w, x, z, t \in \mathbb{B}$,
   $AD(u, v, z, t) \leq AD(u, v, w, x) + AD(w, x, z, t)$
4. *In general,* $\forall u, v, w, x \in \mathbb{B}$ *:*
   $AD(u, v, w, x) \neq AD(v, u, w, x)$

In $\mathbb{B}^m$, we define the analogical dissimilarity $AD(a, b, c, d)$ as $\sum_{j=1}^{m} AD(a_j, b_j, c_j, d_j)$, and the four properties above still hold true [MICLET and DELHAY, 2004].

$h(a)$: $h(b)$ :: $h(c)$: $h(x)$	resolution
ω_0 : ω_0 :: ω_0 : ?	$h(x) = \omega_0$
ω_1 : ω_0 :: ω_1 : ?	
ω_1 : ω_1 :: ω_1 : ?	$h(x) = \omega_1$
ω_0 : ω_1 :: ω_0 : ?	

Table 1: Possible configurations of a triplets

2.3 Extension to Nominal Attributes

To cope with nominal data, two approaches are possible:

- The first one (*one-per-value encoding*) consists in splitting the nominal attribute. As a result, an n valued nominal attribute is replaced by n binary attributes with exactly one at 1.
- The second approach consists in keeping the nominal attribute as a single attribute. This requires to define an analogical proportion on the values of the attribute.

The second approach can be used when there is some knowledge about the nominal values, like an order relation or a measure of distance between the values. Since it is not the case in the data sets that we have worked on, we have chosen to use the one-per-value encoding to treat the nominal attributes.

3 A Classification Rule by Analogical Dissimilarity

3.1 Principle

Let $\mathcal{S} = \{(o_i, h(o_i)) \mid 1 \leq i \leq m\}$ be a learning set, where $h(o_i)$ is the class of the object o_i. The objects are defined by binary attributes. Let x be an object not in $\mathcal{S}$. The learning problem is to find the class of x, using the learning set $\mathcal{S}$. To do this, we define a learning rule based on the concept of analogical dissimilarity depending on an integer k, which could be called the k *least dissimilar triplets* rule. It consists of the following steps:

1. Compute the analogical dissimilarity between x and all the n triplets in $\mathcal{S}$ which produce a solution for the class of x.
2. Sort these n triplets by the increasing value of their AD when associated with x.
3. If the k-th object has the integer value p, then let k' be the greatest integer such that the k'-th object has the value p.
4. Solve the k' analogical equations on the label of the class. Take the winner of the votes among the k' results.

To explain, we firstly consider the case where there are only two classes ω_0 and ω_1. An example with 3 classes will follow. Point 1 means that we retain only the triplets which have one of the four[1] following configuration for their class (see Table 1). We ignore the triplets that do not lead to an equation with a trivial solution on classes :

[1]There are actually two more, each one equivalent to one of the four (by exchange of the mean objects).

$h(a)$: $h(b)$ :: $h(c)$: $h(x)$
ω_0 : ω_1 :: ω_1 : ?
ω_1 : ω_0 :: ω_0 : ?

Point 2 is comparable to the k nearest neighbors method. Since AD takes integer values, there are in general several triplets for which $AD = 0$, $AD = 1$, etc. That is why point 3 increases the value of k to take into account all the triplets with the same value for AD. Finally, point 4 uses the same voting technique as the k-*nn* rule.

Multiclass problem: we proceed in the same way as in the two classes problem, keeping only the triplets which solve on the class label, for example:

Analogical equations
with a solution
ω_1 : ω_3 :: ω_1 : ? $\Rightarrow h(x) = \omega_3$
ω_4 : ω_4 :: ω_2 : ? $\Rightarrow h(x) = \omega_2$
with no solution
ω_1 : ω_3 :: ω_0 : ? $\Rightarrow h(x) = ?$
ω_0 : ω_1 :: ω_3 : ? $\Rightarrow h(x) = ?$

Missing values problem: we consider a missing value as a value for which the distance to each valued nominal attribute is the same. Hence, when splitting a nominal attribute, a missing value would take the value 0 in all the split binary attributes instead of taking the value 1 in exactly one of them (Sect. 2.3).

Example

Let $\mathcal{S} = \{(a, \omega_0), (b, \omega_0), (c, \omega_1), (d, \omega_1), (e, \omega_2)\}$ be a set of five labelled objects and let $x \notin \mathcal{S}$ be some object to be classified. According to the analogical proportion axioms, there is only 75 $(= (Card(\mathcal{S})^3 + Card(\mathcal{S})^2)/2)$ non-equivalent analogical equations among $125 (= Card(\mathcal{S})^3)$ equations that can be formed between three objects from $\mathcal{S}$ and x. Table (2) shows only the first 14 lines after sorting with regard to some arbitrarily analogical dissimilarity. The following table gives the classification of an object x according to k :

k	1 2 3 4 5 6 7
k'	1 3 3 5 5 7 7
votes for x	1 1 1 ? ? 2 2

3.2 A Fast Algorithm : FADANA*

As shown in Sect. 3.1, this learning by analogical proportion technique needs to find the k least dissimilar analogical proportion triplets in $\mathcal{S}$. The naive way is to compute the analogical dissimilarity of every possible triplet from the learning set with the unknown class object as the fourth object of the analogical proportion. This method has a complexity of $O(m^3)$ based on the number of all the non equivalent triplets that can be constructed from the learning set $\mathcal{S}$. To make it faster, we use the algorithm FADANA* (FAst search of the least Dissimilar ANAlogy) inspired from LAESA [Moreno-Seco *et al.*, 2000,] for the 1-*nn* classification rule. This algorithm speeds up the computation on line but has to make off line some preprocessing of the learning set. The number of off line computations depends on the number of "base prototypes" (complexity of $O(bp*m^2)$). The more base prototypes

o_1	o_2	o_3	$h(o_1)$	$h(o_2)$	$h(o_3)$	$h(x)$	AD	k
b	a	d	ω_0	ω_0	ω_1	ω_1	0	1
b	d	e	ω_0	ω_1	ω_2	$\perp$	1	
c	d	e	ω_1	ω_1	ω_2	ω_2	1	2
a	b	d	ω_0	ω_0	ω_1	ω_1	1	3
c	a	e	ω_1	ω_0	ω_2	$\perp$	2	
d	c	e	ω_1	ω_1	ω_2	ω_2	2	4
d	b	c	ω_1	ω_0	ω_1	ω_0	2	5
a	c	e	ω_0	ω_1	ω_2	$\perp$	2	
a	c	c	ω_0	ω_1	ω_1	$\perp$	3	
a	b	e	ω_0	ω_0	ω_2	ω_2	3	6
b	a	e	ω_0	ω_0	ω_2	ω_2	3	7
b	c	d	ω_0	ω_1	ω_1	$\perp$	3	
c	c	c	ω_1	ω_1	ω_1	ω_1	4	8
a	a	c	ω_0	ω_0	ω_1	ω_1	4	9
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table 2: An example of classification by analogical dissimilarity. Analogical proportions whose analogical resolution on classes have no solution (represented by $\perp$) are not taken into account. AD is short for $AD(o_1, o_2, o_3, x)$.

we have the more computations we do off line and the less we do on line. To show its efficiency, we give on the SPECT data base (Sect. 5.1) the performance of FADANA* (the naive method would need 512 000 AD computations).

bp	1	2	5	10	20	50	100	200
nAD	83365	41753	6459	1084	404	299	283	275

bp is the number of base prototypes, and nAD is the number of AD computations in the online part.

4 Weighting the Attributes for an Analogical Dissimilarity Decision Rule

The basic idea in weighting is that all the attributes do not have the same importance in the classification. The idea of selecting interesting attributes for analogy is not new. In [TURNEY, 2005] a discrimination is also done by keeping the most frequent patterns in words. Therefore, one should give greater importance to the attributes that are actually discriminant. However, in an analogical classification, there are several ways to find the class of the unknown element. Let us take again the preceding two class problem example (1). We notice that there is two ways to decide between the class ω_0 and the class ω_1 (there is also a third possible configuration which is equivalent to the second by exchange of the means (Sect. 2.1)). We therefore have to take into account the equation used to find the class. This is why we define a set of weights for each attribute, according to the number of classes. These sets are stored in an *analogical weighting matrix* (Sect. 4.1).

4.1 Weighting Matrix

Definition 2 *An analogical weighting matrix (W) is a three dimensional array. The first dimension is for the attributes, the second one is for the class of the first element in an analogical proportion and the third one is for the class of the last element in an analogical proportion. The analogical proportion weighting matrix is a $d \times C \times C$ matrix, where d is the number of attributes and C is the number of classes.*

For a given attribute a_k of rank k, the element W_{kij} of the matrix indicates which weight must be given to a_k when encountered in an analogical proportion on classes whose first element is ω_i, and for which ω_j is computed as the solution.

Hence, for the attribute a_k:

		Last element (decision)	
		class ω_i	class ω_j
First element	class ω_i	W_{kii}	W_{kij}
	class ω_j	W_{kji}	W_{kjj}

Since we only take into account the triplets that give a solution on the class decision, all the possible situations are of one of the three patterns:

Possible patterns	First element	Decision class
$\omega_i : \omega_i :: \omega_j : \omega_j$	ω_i	ω_j
$\omega_i : \omega_j :: \omega_i : \omega_j$	ω_i	ω_j
$\omega_i : \omega_i :: \omega_i : \omega_i$	ω_i	ω_i

This remark gives us a way to compute the values W_{kij} from the learning set.

4.2 Learning the Weights from the Training Sample

The goal is now to fill the three dimensional analogical weighting matrix using the learning set. We estimate W_{kij} by the frequency that the attribute k is in an analogical proportion with the first element class ω_i, and solves in class ω_j.

Firstly, we tabulate the splitting of every attribute a_k on the classes ω_i:

	...	class ω_i	...
$a_k = 0$	...	n_{0i}	...
$a_k = 1$	...	n_{1i}	...

where a_k is the attribute k and n_{0i} (resp. n_{1i}) is the number of objects in the class i that have the value 0 (resp. 1) for the binary attribute k. Hence,

$$\sum_{k=0}^{1} \sum_{i=1}^{C} n_{ki} = m$$

m is the number of objects in the training set.
Secondly, we compute W_{kij} by estimating the probability to find a correct analogical proportion on attribute k with first element class ω_i which solves in class ω_j.
In the following table we show all the possible ways of having an analogical proportion on the binary attribute k. 0_i (resp. 1_i) is the 0 (resp. 1) value of the attribute k that has class ω_i.

1^{st}	$0_i : 0_i :: 0_j : 0_j$
2^{sd}	$0_i : 1_i :: 0_j : 1_j$
3^{rd}	$0_i : 0_i :: 1_j : 1_j$
4^{th}	$1_i : 1_i :: 1_j : 1_j$
5^{th}	$1_i : 0_i :: 1_j : 0_j$
6^{th}	$1_i : 1_i :: 0_j : 0_j$

$\mathcal{P}_k(1^{st})$ estimates the probability that the first analogical proportion in the table above occurs.

$$\mathcal{P}_k(1^{st}) \quad = \quad n_{0i}n_{0i}n_{0j}n_{0j}/m^4$$
$$\vdots$$

From $W_{kij} = \mathcal{P}_k(1^{st}) + \ldots + \mathcal{P}_k(6^{th})$, we compute

$$W_{kij} = \left((n_{0i}^2+n_{1i}^2)(n_{0j}^2+n_{1j}^2)+2*n_{0i}n_{0j}n_{1i}n_{1j}\right)/(6*m^4)$$

The decision algorithm (Sect. 3.1) is only modified at point 1, which turns into *Weighted Analogical Proportion Classifier* ($WAPC$):

- Given x, find all the n triplets in $\mathcal{S}$ which can produce a solution for the class of x. For every triplet among these n, say (a, b, c), consider the class ω_i of the first element a and the class ω_j of the solution. Compute the analogical dissimilarity between x and this triplet with the weighted AD:

$$AD(a, b, c, x) = \sum_{k=1}^{d} W_{kij} AD(a_k, b_k, c_k, x_k)$$

Otherwise, if point 1 is not modified, the method is called *Analogical Proportion Classifier* (APC).

5 Experiments and Results

5.1 Experiment Protocol

We have applied the weighted analogical proportion classifier ($WAPC$) to eight classical data bases, with binary and nominal attributes, of the UCI Repository [NEWMAN *et al.*, 1998].

MONK 1,**2** and **3** Problems (*MO.1*, *MO.2* and *MO.3*), MONK3 problem has noise added. **SPECT** heart data (*SP.*). **Balance-Scale** (*B.S*) and **Hayes Roth** (*H.R*) database, both multiclass database. **Breast-W** (*Br.*) and **Mushroom** (*Mu.*), both data sets contain missing values.

In order to measure the efficiency of $WAPC$, we have applied some standard classifiers to the same databases, and we have also applied APC to point out the contribution of the weighting matrix (Sect.4.1). We give here the parameters used for the comparison method in Table 3:

- **Decision Table**: the number of non improving decision tables to consider before abandoning the search is 5.
- **Id3**: unpruned decision tree, no missing values allowed.
- **Part**: partial C4.5 decision tree in each iteration and turns the "best" leaf into a rule, One-per-value encoding.
- **Multi layer Perceptron**: back propagation training, One-per-value encoding, one hidden layer with (number of classes+number of attributes)/2 nodes.
- **LMT** ('logistic model trees'): classification trees with logistic regression functions at the leaves, One-per-value encoding.
- **IB1**: Nearest-Neighbor classifier with normalized Euclidean distance.
- **IBk (k=10)**: k Nearest-Neighbor classifier with distance weighting (weight by 1/distance).
- **IB1 (k=5)**[Guide, 2004]: k Nearest-Neighbors classifier with attributes weighting, similarity computed as weighted overlap, relevance weights computed with gain ratio.

5.2 Results

We have worked with the WEKA package [WITTEN and FRANK, 2005] and with TiMBL [Guide, 2004], choosing 7 various classification rules on the same data from WEKA and the last classifier from TiMBL. Some are well fit to binary data, like ID3, PART, Decision Table. Others, like IB1 or Multilayer Perceptron, are more adapted to numerical and noisy data. The results are given in Table 3. We have arbitrarily taken $k = 100$ for our two rules. We draw the following conclusions from this preliminary comparative study: firstly, according to the good classification rate of $WAPC$ in *Br.* and *Mu.* databases, we can say that $WAPC$ handles the missing values well. Secondly, $WAPC$ seems to belong to the best classifiers for the *B.S* and *H.R* databases, which confirms that $WAPC$ deals well with multiclass problems. Thirdly, as shown by the good classification rate of $WAPC$ in the *MO.3* problem, $WAPC$ handles well noisy data. Finally, the results on *MO.* and *B.S* database are exactly the same with the weighted decision rule $WAPC$ than with APC. This is due to the fact that all AD that are computed up to $k = 100$ are of null value. But on the other data bases, the weighting is quite effective.

6 Discussion and FutureWork

Besides its original point of view, the weighted analogical proportion classifier seems, after these preliminary experiments, to belong to the best classifiers on binary and nominal data, at least for small training sets. The reason is that it can profit from regularities in alternated values in the classes in contrast to decision trees or metric methods which directly correlate the attributes with the classes. This means that an analogical proportion between four objects in the same class is reinforced when taking into account a new attribute with values $(0, 1, 0, 1)$ on the four objects but is decreased if the values are $(0, 1, 1, 0)$.

We believe that there is still ample room for progress in this technique. In particular, we intend to investigate a more precise mode of weighting the attributes according to the type of analogical proportion in which they are involved, and also to test $WAPC$ with numerical attributes. Obviously, more work has to be done on the computational aspects. Even with the FADANA* method (Sect. 3.2), the decision process still takes too much time for realistic computation on large data sets.

Another interesting question is the limits of this type of classification. We already know how to define an analogical proportion and an analogical dissimilarity on numerical objects, and we have investigated on sequences of symbolic or numerical objects [DELHAY and MICLET, 2005]. Still, we do not know in which cases an analogical proportion between classes is related to that of objects that constitute the classes

Methods	*MO.1*	*MO.2*	*MO.3*	*SP.*	*B.S*	*Br.*	*H.R*	*Mu.*
number of nominal attributes	7	7	7	22	4	9	4	22
number of binary attributes	15	15	15	22	20	90	15	121
number of train instances	124	169	122	80	187	35	66	81
number of test instances	432	432	432	172	438	664	66	8043
number of classes	2	2	2	2	3	2	4	2
WAPC ($k = 100$)	**98%**	**100%**	**96%**	**79%**	**86%**	**96%**	**82%**	**98%**
APC ($k = 100$)	**98%**	**100%**	**96%**	58%	**86%**	**91%**	74%	**97%**
Decision Table	**100%**	64%	**97%**	65%	67%	86%	42%	**99%**
Id3	78%	65%	**94%**	71%	54%	mv	71%	mv
PART	93%	78%	**98%**	**81%**	76%	88%	**82%**	**94%**
Multi layer Perceptron	**100%**	**100%**	**94%**	73%	**89%**	**96%**	77%	**96%**
LMT	94%	76%	**97%**	**77%**	**89%**	88%	**83%**	**94%**
IB1	79%	74%	83%	**80%**	62%	**96%**	56%	**98%**
IBk ($k = 10$)	81%	7%	**93%**	57%	82%	86%	61%	91%
IB1 ($k = 5$)	73%	59%	**97%**	65%	78%	**95%**	**80%**	**97%**

Table 3: Comparison Table between $WAPC$ and other classical classifiers on eight data sets. The best classification rates, within a significance level = 5%, are in Boldface.

in the learning sample. Are there data for which an analogical proportion classification is "natural" and data for which it does not make sense?

7 Conclusion

In this article, we have shown a new method for binary and nominal data classification. We use the definition of an analogical proportion between four objects. The technique of the classifier is based on the new notion of analogical dissimilarity and on the resolution of analogical equations on the class labels.

We also have shown the importance of weighting the attributes according to the type of analogical proportion in which they are involved. This has led us to good results in all situations, compared to off the shelf standard classifiers, on six classical data bases.

References

[DELHAY and MICLET, 2005] A. DELHAY and L. MICLET. Analogie entre séquences : Définitions, calcul et utilisation en apprentissage supervisé. *Revue d'Intelligence Artificielle.*, 19:683–712, 2005.

[DUDA *et al.*, 2001] R. DUDA, P. HART, and D. STORK. *Pattern classification*. Wiley, 2001.

[GENTNER *et al.*, 2001] D. GENTNER, K. J. HOLYOAK, and B. KOKINOV. *The analogical mind: Perspectives from cognitive science*. MIT Press, Cambridge, MA, 2001.

[Guide, 2004] Version Reference Guide. Timbl: Tilburg memory-based learner, 2004.

[LEPAGE and ANDO, 1996] Y. LEPAGE and S. ANDO. Saussurian analogy: a theoretical account and its application. In *Proceedings of COLING-96*, pages 717–722, København, August 1996.

[LEPAGE, 1996] Y. LEPAGE. Ambiguities in analysis by analogy. In *Proceedings of MIDDIM-96, post–COLING seminar on interactive desambiguation*, pages 93–100, Grenoble, 1996.

[MICLET and DELHAY, 2004] L. MICLET and A. DELHAY. Relation d'analogie et distance sur un alphabet défini par des traits. Technical Report 5244, INRIA, July 2004. in French.

[NEWMAN *et al.*, 1998] D.J. NEWMAN, S. HETTICH, C.L. BLAKE, and C.J. MERZ. UCI repository of machine learning databases, 1998.

[MORENO-SECO *et al.*, 2000] F. MORENO-SECO, J. ONCINA-CARRATALÁ, and L. MICO-ANDRES. *Improving the Linear Approximating and Eliminating Search Algorithm (LAESA) Error Rates*, page 43. IOS Press, 2000. Pattern Recognition and Applications. Frontiers in Artificial Intelligence and Applications,.

[TURNEY, 2005] Peter D. TURNEY. Measuring semantic similarity by latent relational analysis. *Proceedings Nineteenth International Joint Conference on Artificial Intelligence (IJCAI-05)*, 05:1136, 2005.

[WITTEN and FRANK, 2005] Ian H. WITTEN and Eibe FRANK. *Data Mining: Practical machine learning tools and techniques, 2nd Edition*. Morgan Kaufmann Publishers, 2005.