

RESUMEN DE TESIS DOCTORAL

A Unified Methodology to Evaluate Supervised and Non-Supervised Classification Algorithms *Una Metodología Unificada para la Evaluación de Algoritmos de Clasificación tanto Supervisados como No-Supervisados*

Graduated: Salvador Godoy Calderón

Centro de Investigación en Computación-IPN

Av. Juan de Dios Bátiz s/n esq. Miguel Othón Mendizábal C. P. 07738 México D. F.

sgodoyc@cic.ipn.mx

Graduated in June 19, 2006

Advisor: José Francisco Martínez Trinidad

Instituto Nacional de Astrofísica Óptica y Electrónica

Luis Enrique Erro 1, Sta. Ma. Tonantzintla, 72840 Puebla

fmartine@inaoep.mx

Co-Advisor Manuel S. Lazo-Cortes

Instituto de Cibernética, Matemáticas y Física

Ministerio de Ciencia, Tecnología y Medio Ambiente

E No. 309 esq. A 15 Vedado

Ciudad de la Habana, Cuba C. P. 10400

mlazo@icmf.inf.cu

Co-Advisor Juan Luis Díaz de León Santiago

Centro de Investigación en Computación-IPN

Av. Juan de Dios Bátiz s/n esq. Miguel Othón Mendizábal C. P. 07738 México D. F.

jdiaz@cic.ipn.mx

Abstract

There is presently no unified methodology that allows the evaluation of supervised or non-supervised classification algorithms. Supervised problems are evaluated through quality functions while non-supervised problems are evaluated through several structural indexes. In both cases a lot of useful information remains hidden or is not considered by the evaluation method, such as the quality of the sample or the structural change generated by the classification algorithm. This work proposes a unified methodology that can be used to evaluate both type of classification problems. This new methodology yields a larger amount of information to the evaluator regarding the quality of the initial sample, when it exists, and regarding the change produced by the classification algorithm in the case of non-supervised classification problems. It also offers the added possibility of making comparative evaluations with different algorithms

Keywords: Supervised Classification, Non-Supervised Classification, Evaluation of algorithms, Methodologies.

Resumen

Actualmente no existe una metodología que permita la evaluación de algoritmos de clasificación tanto supervisados como no-supervisados. Los algoritmos aplicados a problemas supervisados se evalúan mediante funciones de calidad mientras que los algoritmos aplicados a problemas no-supervisados se evalúan mediante diversos índices estructurales. En ambos casos mucha información útil permanece oculta o no es considerada por el método de evaluación. En este trabajo se propone una metodología unificada que puede ser usada para evaluar ambos tipos de problemas de clasificación. Esta nueva metodología entrega una mayor cantidad de información al evaluador acerca de la calidad de la muestra inicial, cuando ésta existe y acerca de el cambio producido por el algoritmo de clasificación in el caso de problemas no-supervisados. También ofrece la posibilidad de realizar evaluaciones comparativas con diferentes algoritmos.

Palabras Claves: Clasificación Supervisada, Clasificación No-Supervisada, Evaluación de algoritmos, Metodologías

1 Introduction

While working in the pattern recognition area, whether in field applications or in research, it is common to have to evaluate the result of a classification process. [1][2]. On many occasions the objective of such evaluation is, either to find out the behavior of the classification algorithm used or to establish the pertinence of the application of such algorithm to the type of problem being evaluated. Classification problems may be shown in two different ways [4] known as: Supervised Problems and Non-Supervised Problems. Unfortunately, nowadays there is no methodology that allows us to evaluate the result of an algorithm under the same criteria.

A classification problem is informally called Supervised when there is previous knowledge on the classes or categories into which objects or patterns being studied are classified and also, each one of these classes contains in it at least one previously classified pattern. This means that a problem is Supervised when a sample of previously classified patterns in each of the categories to be considered is available. Such a sample is called Control Sample, Supervision Sample or Learning Information.

A classification problem is considered Non-Supervised when such previous knowledge does not exist. Similarly, a classification problem is considered Non-Supervised when such previous knowledge does not exist. In that case, the problem shows itself as a universe of patterns without structure that must be classified, but the number and nature of the classes to be built are part of the initial definitions necessary to solve the problem. For this reason there is not an initial sample for Non-Supervised problems. Some times researchers talk about a third form a classification problem known as a Partially-Supervised problem which is an intermediate state between the two previous types of problems. A classification problem is considered Partially-Supervised when the previous knowledge regarding the nature of its solution is partial, e.g., when some, but not all of the classes in which the objects will be classified are known. Likewise, some, but not all of the known classes contain previously classified patterns.

The methods used to evaluate the algorithms that solve classification problems change considerably due to the intrinsic characteristics of each type of problem [3].

2 Traditional Evaluation Methods

In order to evaluate Supervised problems, the classification algorithm is applied to a previously studied problem and the result is compared with a previously known solution considered as valid [5][6]. This comparison is made through Quality Functions that generate a score, which is typically a real number and which synthesizes the evaluation of the problem and the effectiveness of the classification algorithm. Regardless of the type of problem: with multi – classification, with fuzzy classes or with absence of information, quality functions take into consideration the following criteria:

1. The amount patterns classified by the algorithm.
2. The correctness and precision of the degrees of membership assigned to the patterns in each class.
3. The amount of patterns for which the algorithm does not assign a membership to any class (abstentions).
4. A different weight for each type of mistake made.

The quality function to be used in the evaluation of a specific problem depends to a large extent on the conditions and semantics of the problem, so there is an infinite amount of possible quality functions. In many cases, simple quality functions, such as the following one, are applied [3] [7] [10]:

Let A be a supervised classification algorithm and let $\Phi(A)$ be the quality function that evaluates it and which expression is:

$$\Phi(A) = \frac{x}{x + y + z}$$

where x is the number of patterns correctly classified by the algorithm,

y is the number of patterns erroneously classified,
 z is the number of abstentions.

Other times, much more detailed quality functions, such as the following, are applied.

$$\Phi(A) = \frac{1}{n} \left(\sum_{i=1}^k \sum_{j=1}^k \alpha_{ij} E_{ij} + \sum_{s=1}^k \beta_s A_s \right)$$

where: n is the total number of patterns in the control sample
 k is the number of classes in the problem

α_{ij} is the amount of objects that belong to class i , mistakenly classified in class j .

E_{ij} is the specific weighting of the mistake counted in α_{ij}

β_s is the amount of objects h that belong to class i in which the algorithms refrained from classifying.

A_s is the specific weight of the error counted in β_s

Regardless of how complex the selected quality function may be, the result of the evaluation is always expressed with only one number which hides the details of the analysis made and the specific reasons for the assigned evaluation.

In the case of Non-Supervised problems, there is no explicit formula to evaluate the quality of the classification algorithm. But, opposite to what happens in the supervised case, the idea of measuring the quality of the resulting covering in terms of its structural conditions [9] is quite common. This is known in literature as Analysis of Clusters. Given the fact that on many occasions it is not possible to anticipate the optimum number of classes that will be formed, it is necessary to validate each one of the clusters made and to decide if the selection of the number k of classes was right or if the selection must be modified and the patterns must be re-classified. The validation is made through one or more Structural Indexes that evaluate different properties of the clusters defined by the classification algorithm and determine if there is a better option according to the evaluated parameters. Many structural properties can be evaluated in a covering, but commonly aspects considered include the compacting of clusters, the separation between clusters, the max and min degree of membership of each cluster, etc. (See [10],[13] and [14])

Several indexes have been proposed to evaluate partitions and coverings. Three of the more widely used are the Partition Coefficient and the Entropy index proposed by Bezdek [9], like the Xie-Beni index [10]. Let us examine each one of them.

For a Non-Supervised classification problem, with n patterns and with k being the pre-determined number of classes to be formed, Bezdek defines the Partition Coefficient (PC) in [9] as:

$$PC = \frac{\sum_{i=1}^n \sum_{j=1}^k \mu_{ij}}{n}$$

With μ_{ij} being the membership of pattern i to class j .

Under the same assumptions, Bezdek also defines the Partition Entropy (PE) as:

$$EP = \frac{-\sum_{i=1}^n \sum_{j=1}^k (\mu_{ij} \log(\mu_{ij}))}{n}$$

The main disadvantage of these indexes, as Bezdek himself states in [9] is that they evaluate each class by considering exclusively the degrees of membership assigned to the patterns and not their (geometric) structure or the structure of the whole covering.

X. L. Xie and G. Beni proposed an index that evaluated two structural aspects: the compactation and the separation of the classes [10]. For them, an optimum partition is that which has a strong compactation and a noticeable separation between clusters. Therefore, the index they proposed and which takes its name after them is made up as follows:

The Compactation, Ψ of the e la partition is calculated as:

$$\Psi = \sum_{i=1}^k \sum_{j=1}^n \mu_{ij} \|x_j - v_i\|^2$$

with v_i being the centroid of each class. The second factor then represents the Euclidian Norm of the distance between each object and the corresponding centroid in the covering. The Separation, Ξ between classes, is calculated as:

$$\Xi = n \left(\min_{i \neq k} \|v_i - v_k\|^2 \right)$$

Lastly, the Xie-Beni (XB) index is formed as the quotient of these two quantities:

$$XB = \frac{\Psi}{\Xi}$$

Like in the case of Supervised problems, all of these Structural Indexes limit their evaluation to only one number which, in this case, represents the quality of the structuring in the solution covering generated by the classification algorithm.

Most authors don't even consider Partially Supervised problems as a different category of problems [12]. These problems are treated as Supervised in what regards the evaluation of the classification algorithms. Therefore in the rest of this paper no explicit reference will be made to Partially Supervised problems and the same conditions of Supervised problems will be assumed for them.

3 Advantages and Disadvantages of Traditional Methods

The evident advantage of evaluating Supervised problems through quality functions is the flexibility of the latter. The researcher can build a quality function as thorough as the problem requires, and one that can encompass situations of very different kinds, such as the abstentions of the classifying algorithm or the different weighing for each type of error made in the assignment of the membership. In return for this, the way of evaluating Supervised problems has some evident disadvantages. The most noticeable one is the need of having a previously known solution for the problem being evaluated. This requirement makes it impossible to evaluate problems for which such a solution is not available and even more: the consideration of such solution as "the correct solution" may cause important biases for the evaluation process. There are two main reasons for these biases in the evaluation: first, the quality of the control sample used for the evaluated classification algorithm and for the previously known solution considered as correct is not analyzed. This lack of evaluation of the control sample used for a Supervised problem seriously limits the ability to judge the action of

the classifying algorithm. Second: the quality of the structure induced in the solution coverage by the evaluated algorithm is not measured.

It is not hard to imagine that a very well built sample (with the more representative patterns of each class) may induce the generation of the same solution even for less effective algorithms for the conditions of the problem being evaluated, while a poorly built sample (with patterns not very representative of each class) may induce errors or abstentions in the algorithms based on the similarity of the patterns to generate the classification. These are not clear criteria to select a specific solution for a problem and to consider it as correct and as a point of reference to evaluate other algorithms. Should the methodology include any type of measurement of the structure of the solution covering generated by the evaluated algorithm, the evaluation would not depend so much of the quality of the control sample. Nonetheless, the quality function is limited to comparing the membership to each of the classes assigned by the classifying algorithm to each pattern. Lastly, notice that most of the classification algorithms (both for Supervised and for Non-Supervised problems) are based on a specific form of measuring the similarity between two patterns. The criterion or set of criteria through which the similarity is measured is called *Analogy Function Between Patterns* and it is evident that in spite of the fact that this function is the most important element for the algorithm, it is in no way considered by the evaluation methodology for Supervised problems. In summary, the following disadvantages may be noticed:

1. The quality of the control sample is not measured.
2. The structural quality of the solution covering is not measured.
3. The *Analogy Function between Patterns* is not involved in the evaluation.

The classic method to evaluate Non-Supervised problems has very different characteristics. The evaluation is made based on the quality of the structure of the solution yielded by the algorithm instead of comparing it with a previously known solution. This is by far the most evident advantage of this method. In general, the elements considered to make the evaluation are precisely those which are not considered in Supervised problems. These are evaluation methods which are radically different in both cases, but the diverse conditions of each type of problem do not allow the indiscriminate use of the respective methods. Nevertheless, in both cases the evaluation of the algorithm is reduced in its expression to only one number which generally hides more information than the one it gives, because it does not allow an analysis of the specific situation of a pattern or category. Therefore, the list of disadvantages of classical methods may be completed as follows:

4. The evaluation is synthesized in only one number which does not allow alternative interpretation.
5. The evaluation methods are not unified for all types of problems.

All the information above leads us to ask the following question: Is it possible to devise an evaluation methodology that can overcome the deficiencies found in the present methods and produces unified criteria to evaluate classification algorithms applied to any type of problem?

4 The Main Definitions

Before presenting the methodology proposed we now introduce the three most important theoretical concepts on which the design and methodology are based. These concepts are: the Covering, the Classification Problem and the Classification Algorithm. For a more detailed description (see [12]).

Let Ω be a known universe of objects under study and let $O \subseteq \Omega$,

Definition #1.- A *Covering* of O is a tuple $(O, \mathfrak{R}, \delta, Q, \pi, Cc, f)$ where O , \mathfrak{R} and Q (called *Structural Sets*) are respectively sets of Objects, Descriptive Features for the objects and Classes. Components δ and π (called *Structural Relations*) are, respectively, *Description and Membership functional relations*. The first one describes the objects of O in terms of the features in \mathfrak{R} and the second one assigns to each o_i object a membership to each of the C_j classes. Last, Cc and f are respectively a set of *Comparison criteria* and the *Analogy Function Between*. (see [12]).

According to the definition given above, the following special types of coverings may be characterized:

Table 1 Different types of Coverings

| A Covering is called: | If the following condition is satisfied: |
|---------------------------|--|
| Total Partial Blind | Every object belongs to a category There is an object that does not belong to any category No object belongs to any category |
| Stringent Flexible | All categories are not empty There is an empty category |

Definition #2.- A Classification Problem is a tuple of the form (Z_0, Θ) where:

Z_0 is a *Partial covering* which we wish to transform in a *Total* one.

Θ is a Restrictions Set imposed over the solution of the problem.

- A problem is *Supervised* if and only if its initial covering is Stringent.
- A problem is *Partially Supervised* if and only if its initial covering is Flexible.
- A problem is *Non-supervised* if and only if its initial covering is Blind.

Definition #3.- A Classification Algorithm is an algorithm of the form $A(P) = Z_1$ such that, it takes a Classification Problem (in any of its forms) and delivers a Total Final Covering which is the solution to this problem.

5 Proposed Methodology

In designing this evaluation methodology two general objectives were established: 1) to safe keep the advantages of each classical method, but to overcome their disadvantages, and 2) to generate a unified methodology for all types of classification problem. According to the definitions in the previous section, the methodology designed to evaluate classification algorithms is based on the structural comparison between the Initial Covering of a problem and its Final Covering generated as a solution by the classification algorithm. Such comparison can always be made, even in cases in which one of the two compared coverings is a blind covering (as is the case in Non-Supervised problems). The measurement of all types of properties in the covering that involve the membership of patterns to classes and the similarity between them, in accordance with the analogy function between patterns is considered as a structural comparison.

The proposed evaluation methodology obviously starts with the application of the classifying algorithm to the problem being solved. From that moment on, the evaluation process is developed in the following three stages:

Stage #1 (Structural Analysis of the coverings) during this stage the initial and final coverings of the problem are analyzed separately, calculating the same set of structural properties for each of them. These properties are discussed in detail in a later section. The analysis takes place at three levels for each covering:

- **Level of Objects:** The structural properties are calculated for each object, making reference to each class in the covering.
- **Level of Classes:** The values corresponding to each of the structural properties in the patterns that form the support of each class are accumulated and averaged.
- **Level of the Covering:** The indexes for the structural properties for the covering under study are calculated.

Stage #2 (Comparison between Coverings) The difference in value of each structural property calculated for each covering during the previous stage is determined. The calculated set of differences is called *Differences Tuple* and it is the score assigned to the classification algorithm. This tuple expresses the structural change generated by the classifying algorithm in the initial covering of the problem.

Stage #3 (Interpretation) Once we have the partial results of each of the previous stages, particularly those corresponding to the three levels of structural analysis of the coverings, the researcher interprets the obtained score.

Unlike the classical methods, this one refrains from reducing all the evaluation process to only one final score that hides the details involved in the evaluation process. The partial results obtained in each stage are valuable sources of information for the researcher, where he can study the particular situations regarding the problem being solved. Another distinctive characteristic of this methodology is the fact that it is useful independently of the quantity and specification of the structural properties that we want to calculate during the first stage. Sometimes the researcher may be interested in using a specific set of structural properties, according to the characteristics of the problem under study. For this reason, the methodology described above was introduced without any reference to the specific properties used in the analysis of coverings. In this sense, the set of structural properties that have been used and are described in the next section are shown for the sole purpose of clarifying all of the elements involved in the methodology. Nonetheless, the researcher is free to use the set of properties that he deems to be more adequate for his particular study.

6 Application Details

For the structural analysis stage, only four properties, considered as determining factors in the structure of a covering, are calculated:

- The *Tipicity* (T) of an o_i object, with respect to a C_j class, understood as the degree to which the object is representative of such class is calculated as follows:

$$T(o_i, C_j) = \frac{\sum_{o_s \neq o_i} f(o_i, o_s) * \pi(o_s, C_j)}{|Sop(C_j)|}$$

where:

- $f(o_i, o_s)$ is the Analogy function between patterns.
- $\pi(o_s, C_j)$ is the Membership of the o_i object to the C_j class
- $|Sop(C_j)|$ is the cardinality of the support of the C_j class

- The *Contrast* (C) of an o_i object with respect to a C_j class, understood as the degree to which the object is representative of all of the other classes in the covering:

$$C(o_i, C_j) = \frac{\sum_{C_s \neq C_j} T(o_i, C_s)}{k - 1}$$

where:

- $T(o_i, C_s)$ is the Tipicity of the o_i object, in the C_j class
- k is the total number of classes in the covering

- The *Discrimination Error* (\mathcal{E}) of an O_i object with regard to a C_j class, understood as the degree of confusion of the object in the covering:

$$\mathcal{E}(O_i, C_j) = \sum_{C_s \neq C_j} \pi(O_i, C_{js})$$

where:

$\pi(O_i, C_{js})$ is the degree of membership of O_i to the intersection of the C_j and C_s classes.

- The *Characterization Error* (\mathcal{Y}) of an O_i object with regard to a C_j class, understood as the difference between the membership of the object to the class and its tipicity in this same class:

$$\mathcal{Y}(O_i, C_j) = \left| \pi(O_i, C_j) - T(O_i, C_j) \right|$$

During the analysis at the level of the classes, each of these structural properties is averaged in the analyzed class. During the analysis at the level of the covering, the structural indexes corresponding to each property are calculated. In every case, the index is calculated as one minus the corresponding property averaged in the whole covering.

Striving to give this methodology the same flexibility shown by the quality functions in Supervised problems, a special technique for the structural analysis of the coverings during the first stage was developed. This technique consists of adding to each covering an additional class which represents the complementary set for the rest of the classes in the covering and then calculating all the structural properties, also regarding this class. In the initial covering of a problem all of the patterns that are not classified will be considered to have maximum membership to the complementary class. This technique allows the proposed analysis to account for the abstentions incurred by the classification algorithm although, evidently, without achieving the same degree of flexibility achieved by the quality functions.

7 Conclusions

Comparison between the initial and final coverings of a problem allow the evaluation of the behavior of the classifying algorithm independently from other circumstantial factors in the problem, such as the quality of the control sample in the case of Supervised problems. Thanks to the definitions previously established, such comparison is a common element between Supervised and Non-Supervised problems and unifies the evaluation methodology.

The specification of what is meant by structural properties allows us to include in the analysis of the coverings both, the basic elements considered by the quality functions (membership assigned to each pattern in each class), as well as those considered by most of the structural indexes with which Non-Supervised problems are evaluated. At the same time, the main advantages of classic methodologies are avoided. Notoriously, the discussed methodology does not require a previously known solution to the problem or evaluates the algorithm by considering such solution as a reference point.

The flexibility of the discussed methodology is manifested in two main aspects: first, the possibility to change the set of structural properties to be used during the analysis of the coverings. Second, the possibility of accounting for the abstentions of the classifying algorithm by using the complementary class technique.

References

1. *Pattern Recognition. Concepts, Methods and Applications* Springer, Portugal, 2001.
2. J. Ruiz Shulcloper, and M.A. Abidi. *Logical Combinatorial Pattern Recognition: A Review*. Recent Research Developments in Pattern Recognition, Pub. Transworld Research Networks, USA, 2002, 133-176.
3. J. Ruiz Shulcloper, Eduardo A. Cabrera, Manuel Lazo Cortés *Introducción al Reconocimiento de Patrones (Enfoque Lógico-Combinatorio)* CINVESTAV-IPN Serie Verde No. 51, México, 1995.
4. J. F. Martínez-Trinidad, and A. Guzman-Arenas. "The logical combinatorial approach to Pattern Recognition an overview through selected works". *Pattern Recognition*, 34, 4, 2001, 1-11.
5. J.C. Bezdek *Pattern Recognition with fuzzy Objective Function Algorithms* Plenum Press, NY, 1981.
6. A.K. Jain, R.C. Dubes *Algorithms for Clustering* Prentice-Hall, Englewood Cliffs, NJ, 1998.
7. A.K. Jain, M.N. Murty, P.J. Flynn "Data Clustering: a review" *ACM Computing Surveys*, 31(3), 1999.
8. J.C. Bezdek "Numerical taxonomy with fuzzy sets" *Journal of Mathematical Biology*, 1, 1974.
9. J.C. Bezdek "Cluster validity with fuzzy sets" *Journal of Cybernetics*, 3, 1974.
10. X.L. Xie, G. Beni "A validity measure for fuzzy clustering" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8), 1991.
11. M. Lazo-Cortés Models based on Testor theory for feature selection and supervised classification with non classical object descriptions PhD. Tesis, Universidad Central de las Villas, Cuba, 1994. (In Spanish)
12. S. Godoy-Calderón, M. Lazo-Cortés, J.F. Martínez-Trinidad "A non-classical view of Coverings and its implications in the formalization of Pattern Recognition problems" *WSEAS Transactions on Mathematics*, 2, 1-2, 2003, 60-66.
13. Dae-Won Kim, Kwang H. Lee, Doheon Lee On cluster validity index for estimation of the optimal number of fuzzy clusters *Pattern Recognition* 37, 2004.
14. H. Lee-Kwang, Y.S. Song, K.M. Lee Similarity measure between fuzzy sets and between elements *Fuzzy Sets and Systems* 62, 1994.



Salvador Godoy Calderón. Graduated in Computer Engineering in 1992 at ITAM- Mexico, Master in Sciences received in 1994 from CINVESTAV and Doctorate in Computer Sciences in 2006 from Center for Computing Research of the National Polytechnic Institute (CIC, IPN), Mexico His research interests Pattern Recognition, Testor Theory, Logic and Formal Systems.



José Francisco Martínez Trinidad. Received his B.S. degree in Computer Science from Physics and Mathematics School of the Autonomous University of Puebla (BUAP), Mexico in 1995, his M.Sc. degree in Computer Science from the faculty of Computers Science of the Autonomous University of Puebla, Mexico in 1997 and his Ph.D. degree in the Center for Computing Research of the National Polytechnic Institute (CIC, IPN), Mexico in 2000. Professor Martinez-Trinidad edited/authored four books and over fifty papers, on subjects related to Pattern Recognition.



Manuel S. Lazo-Cortes. (1955) is a Senior Researcher of the Institute for Cybernetics, Mathematics and Physics (ICIMAF) in Havana, Cuba. He obtained his B. S. (1979) and Ph.D. (1994) degrees in Mathematics from Universidad Central de Las Villas (UCLV), Cuba. He was Professor at UCLV for 15 years and he was Visiting Professor at the Center for Research and Advanced Studies (Mexico City) several times. He is now Director of ICIMAF. His main interests are testor based algorithms and unsupervised classifiers based on similarity functions.



Dr Juan Luis Díaz de León is an engineer from Regional Institute of Technology at Veracruz. He obtained his MSc, and PhD on Electical Engineering at CINVESTAV-IPN. He was Director of Technology at the National Public Security Sistem for the Mexican Goberment.He is Full Professor. He was Director of the Center for Computing Research in the IPN. His main interests are: Computer Vision, Artificial Inteligence, Automatic Control, Mobile Robotics and Pattern Recognition.