A Method of Solving Pattern or Image Recognition Problems

Article · July 1995

Source: CiteSeer

2 authors:

Ventzeslav Valev
University of North Florida

Petia Radeva
University of Barcelona

A Method of Solving Pattern or Image Recognition Problems by Learning Boolean Formulas

Ventzeslav Valev

Institute of Mathematics
Bulgarian Academy of Sciences
1090 Sofia, P.O.Box 373, Bulgaria

Petia Radeva

Faculty of Sciences, UPIIA
Autonomous University of Barcelona
08193 Barcelona, Spain

Abstract

A method of solving supervised pattern recognition problems based on the model of learning Boolean formulas is suggested. It is proved that this method of learning is of NP-complexity. An efficient learning procedure using some tools of combinatorics and graph theory is proposed. The suggested method differs from those known in its diminished number of computational operations. The results obtained are applied to supervised image recognition problems.

1 Introduction

As a rule all models of solving pattern recognition problems use the concepts of similarity and dissimilarity to a certain extent. These concepts are involved in learning or classification rules. The pattern recognition problems may be solved through various models which require only positive examples or positive and negative examples (counterexamples) for a conceptual class. We will deal with a model which requires either positive examples and counterexamples in the sense given below. Each class of a finite number of classes is defined by a finite number of positive examples. The classes should not intersect. If we consider a certain class, then the positive examples of the remaining classes can be treated like negative examples for that class. Note that negative examples can help us to circumscribe the desired conceptual class.

Experience gained in solving a number of practical pattern recognition problems shows that the information needed for an object classification is included either in unique features or in combinations of them (like syndromes in medicine). In Russian literature such a combination of features is called a Non-Reducible Descriptor (NRD). The latter notation will be used further on.

Each object (positive example) creates a set of NRD. The NRD sets for all the positive examples of a certain class form the NRD set for this class. Taking into account this approach to pattern recognition problems, learning may be treated as a process of creating the NRD sets for the classes. Learning is based on the dissimilarity concept. The latter is formed by creating a dissimilarity matrix for the description of a given object (positive example) through comparing its description with the descriptions of all the objects from the remaining classes (negative examples).

Each NRD can be presented in various ways. For example, every NRD may be expressed by writing the sequences of the values of the object features it refers to. Another way of expressing an NRD is by a conjunction. Let $x_1,...,x_n$ be Boolean variables through which the description of an object is created. Let us create an NRD including some of these features. Then the NRD for this object can be expressed by its conjunction. If variable x takes the value of 0, then the conjunction includes the negation of variable x. If the value is 1, then the conjunction includes variable x. The conjunctions obtained for each object of the investigated class (NRD set) may be connected by disjunctions. Thus the NRD set obtained for each object included in learning may be presented by a Boolean formula in disjunctive normal form (DNF).

The method of creating NRD was first used in [1] and it is known under the name "Kora". It is based on the exhausted search method. The number of calculations is limited for the sake of limiting the NRD length. Later some other mathematical methods of creating NRD have been used. They have been based on some tools of Boolean logic such as creating the set of all nonreducible coverings of a Boolean matrix or creating the reduced DNF of a Boolean function [2, 3].

The problem of creating the all nonreducible coverings may be reduced to the problem of finding the reduced DNF of a monotone Boolean function by means of a conjunctive normal form (CNF). The problem of finding the reduced DNF of a partial Boolean function equal to 1 for the objects of the class under consideration and to 0 - for the objects from the remaining classes, is solved by means of obtaining the reduced DNF of a defined everywhere Boolean function, given by a perfect CNF. One of the disadvantages of these methods is that the conjunctions created which are not included in the reduced DNF of the partial Boolean function (i.e. those which are not NRD), should be removed. The efficiency of these methods depends also on the correlation between the number of features and the number of objects included in learning.

We have mentioned that there exist some learning models for which no counterexamples are needed. For instance, the model of learning by examples suggested for learning Boolean formulas [4, 5] comprises counterexamples in an unexplicit form. All the examples which are not positive are treated like counterexamples. The DNF learning is accomplished by removing the conjunctions (from all 3^k conjunctions, $k \leq n$) which are not included in the reduced k-DNF (i.e the conjunctions whose length is at most k and which are not included in the NRD set for positive examples).

In this paper we suggest a method for solving the problem of creating the NRD set for a certain object (positive example) through some tools of combinatorics and graph theory. We have shown that this problem is equivalent to the problem of finding all the different unit submatrices of (0,1)-matrix for which the condition of covering the columns holds. We prove that this problem is of NP-complexity. The NRD set for each object is found directly without removing some of the conjunctions created, i.e. those which do not belong to the reduced DNF.

The results obtained for pattern recognition problems are applied to supervised image recognition problems. Next we will consider the cases when learning examples are Boolean vectors with pattern recognition problems and Boolean matrices - with image recognition problems. The present paper deals with solving supervised pattern or image recognition problems by means of Boolean formulas.

2 A Method of Learning Boolean Formulas

Let us consider the following problem of supervised pattern recognition under the assumptions given in [7]. Let M be a set of objects. It is known that set M is a sum of finite number of subsets $K_1, \ldots, K_l$ which do not intersect. These subsets are called classes. The division of set M into classes is not completely defined. Only its subset $M' \subset M$ and the fact that set M' is divided into l classes are known. Each class K_j, $j = 1, \ldots, l$ is defined by its objects. Each object Q is given by its description $I(Q) = \{t_j\}, j = 1, \ldots, n$, where t_j in the general case may take values from sets: $\{0, 1, \ldots, k-1\}$, k - integer, $k \geq 2$; R, where R is the set of all real numbers. The information given so far for the classes is called training information.

The pattern recognition problem for an arbitrary object $Q \in M$ consists of determining the values of the predicates (properties) $Q \in K_j$, $j = 1, \ldots, l$ through the training information and the description $I(Q)$.

The descriptions of objects $Q_1, \ldots, Q_i, \ldots, Q_m$ are given by a system of features $x_1, \ldots, x_n$ comprising a sequence of values $t_{i\,1}, \ldots, t_{i\,n}$ for these features. A training table $T = (t_{i\,j})$, $i = 1, \ldots, m$, $j = 1, \ldots, n$ is created where line i is the description of object Q_i. We consider the case when features $x_1, \ldots, x_n$ are Boolean variables.

Definition 1. If some features of a certain object take values which cannot be found in the descriptions of the objects of the remaining classes, then such a sequence of values is called a descriptor.

Definition 2. If a certain descriptor loses its property of a descriptor when reduced by a feature value chosen arbitrarily, then it is called a Non-Reducible Descriptor (NRD) [2, 3].

Each NRD is of minimal length. That means that each NRD cannot be reduced more but each NRD may distinguish its own description from the descriptions of the objects belonging to the remaining classes. From the condition that the classes do not intersect and from Definition 2 it follows that the description of each object possesses at least one NRD. In particular it may be the description itself.

Let us consider the problem of obtaining the NRD set of an object $Q_r \in K_j$, $j = 1, \ldots, l$. Let the number of objects Q_i which do not belong to K_j be m'.

Definition 3. The dissimilarity matrix for a given object $Q_r \in K_j$ is the Boolean matrix $L = (l_{i\,j})$, $i = 1, \ldots, m'$, $j = 1, \ldots, n$ obtained as follows:

$$l_{i\,j} = \begin{cases} 1, & \text{if } t_{r\,j} \neq t_{i\,j}, \\ 0, & \text{otherwise}, \end{cases}$$

where $t_{r\,j}$ and $t_{i\,j}$ are respectively values of the features of objects $Q_r \in K_j$ and $Q_i \notin K_j$.

Let dissimilarity matrix L_r be created for object $Q_r \in K_j$.

Definition 4. Two elements l_{i_1,j_1} and l_{i_2,j_2} belonging to dissimilarity matrix L_r are called compatible, if:

i) $l_{i_1,j_1} = l_{i_2,j_2} = 1$ for $i_1 \neq i_2$ and $j_1 \neq j_2$,
ii) $l_{i_1,j_2} = l_{i_2,j_1} = 0$.

Definition 5. Elements $l_{i_1,j_1}, \ldots, l_{i_d,j_d}$ are called a Sequence of Compatible Elements (SCE), if:

i) for $d = 1$, $l_{i_1,j_1} = 1$,
ii) for $d > 1$, each pair of elements is a pair of compatible elements.

Definition 6. The number of elements d of a SCE (NRD) is called a rank of this SCE (NRD) and it is denoted by SCE^d (NRD^d).

From Definition 5 it follows that elements $l_{i_1,j_1}, \ldots, l_{i_d,j_d}$ are compatible iff the matrix formed by lines $i_1, \ldots, i_d$ and columns $j_1, \ldots, j_d$ comprises only one unit in each line and each column.

The problem of obtaining the SCE set in a dissimilarity matrix L_r may be reduced to the problem of creating all the different matrices $L_r^{'}$ of the following form:

$$L_r' = \begin{bmatrix} E_d & P_1 \\ P_2 & P_3 \end{bmatrix}, \qquad (1)$$

by all possible dislocations of columns and lines in matrix L_r. The submatrix E_d is a unit submatrix of size $(d \times d)$. Submatrices P_1,P_2 and P_3 are such that further dislocations in their lines and columns cannot result in obtaining another larger unit submatrix comprising E_d, i.e. submatrix E_d is the maximal unit submatrix. We have to point out that this problem always has a solution since the arbitrarily chosen dissimilarity matrix L_r comprises at least one unit in each line. The last statement follows from the condition that $K_i \cap K_j = \emptyset$ for $i \neq j$; $i,j = 1, \ldots, l$. Next follow two theorems with no proofs.

Theorem 1. The problem of transforming an arbitrary $(0,1)$-matrix L_r into a matrix of type (1) is of NP-complexity.

Theorem 2. Object Q_r has NRD^d $t_{r,j_1}, \ldots, t_{r,j_d}$ iff dissimilarity matrix L_r comprises:
i) lines $i_1, \ldots, i_d$ whose elements $l_{i_1,j_1}, \ldots, l_{i_d,j_d}$ form SCE^d,
ii) columns $j_1, \ldots, j_d$ form a covering of matrix L_r.

The method of creating the NRD set of a certain object Q_r may be expressed by the following procedure steps:
a) find out the SCE set in dissimilarity matrix L_r,
b) disregard those SCE which are equivalent to the SCE obtained,
c) check the condition of covering for each SCE.

The problem of finding out the SCE set in dissimilarity matrix L_r is accomplished by a non-oriented graph juxtaposed with L_r. The algorithm created uses some tools of combinatorics and graph theory. Some theorems illustrate that NRD sets for multidimensional pattern recognition problems can be created through decomposition of the training tables. The computations may be parallel.

The proposed method of creating Boolean formulas (NRD set) differs from those known in its diminished number of computational operations. The computational complexity of this method compared to other methods [2, 3, 4, 5] does not depend on the fact whether $m \gg n$ or $m \ll n$ in the training table. The NRD set for each object is found directly without removing some of the conjunctions created. The computational complexity depends on the number of units in the dissimilarity matrix. Their number depends on the degree of dissimilarity of the descriptions of the objects of the different classes. The greater the number of zeros in the dissimilarity matrix is (the greater the rank of NRD is), the more efficient this method is compared to the methods mentioned above. Therefore, the proposed method of learning will be more efficient whenever the descriptions of the objects belonging to different classes do not distinguish very much. Note that "even more helpful are counterexamples that are "near misses" - that is, negative examples that just barely fail to be positive examples" [6].

The classification rules may include elements of the NRD sets in different ways. For example, the classification may be accomplished by searching elements of the NRD sets for the classes in the description of the recognized object through a voting procedure [7]. The simplest modification of a classification rule is: a vote is given for a recognized object if its description comprises an NRD. These votes are calculated for all the classes and the recognized object is related to some of them by using the rule of maximum. The classification rules may comprise also threshold values.

3 Application to Image Recognition Problems

A number of methods of solving image recognition problems use various transformations, for example, shape analysis, segmentation, contour tracing or thinning. The results obtained by these transformations are used to define the features space as well as to determine the concepts of similarity and dissimilarity between the descriptions of the different image classes. Usually, the features selected for recognition are binary vectors or matrices with the same dimensionality as the images. The problem of the minimal set of features needed often remains unsolved.

The results obtained for learning Boolean formulas of vector descriptions of objects may be applied to solving the supervised image recognition problems. Each object with a matrix data organization is treated as an image. The case of binary images is considered. Thus learning would consist of creating the set of non-reducible descriptors called here Non-Reducible Fragments (NRF) for each image of the learning sequence.

The dissimilarity matrix may be created in various ways which determines the complexity of the calculations. For that purpose each matrix description is transformed to one or more vector descriptions by taking into account the matrix as a whole or parts of it. According to this assumption a dissimilarity matrix might be created for the image as a whole, for a column, for a row, for a fixed window (for example 3×3) or for a certain given region of the image. Note that these parts of images might shift over the image. That results in creating all the elements of the NRF set (where each element is of a rank up to $(m \times n)$) if the image is considered as a whole or in creating its subsets (where each element has a smaller rank) if parts of the image are considered.

Since the images may be processed in parts then the calculation complexity of creating NRF sets will not influence essentially on the dimensionality of the images. The classification rules used for images recognition are like those used for solving pattern recognition problems.

A numerical example illustrates the application of the method proposed to learning Romanian numerals. The description of each symbol is a binary matrix of size 5×5. For example, the NRF set obtained for symbol M (when creating the dissimilarity matrix of image M as a whole) includes 105 elements (conjunctions) whose ranks is 2 or 3. The calculations are accomplished by a program system developed and implemented on IBM PC and compatibles.

4 Conclusions

This method using binary descriptions may be generalized and applied to arbitrary numerical descriptions. Another approach is using the method for transforming the arbitrarily numerical descriptions into k-values, k - integer, $k \geq 2$. Then it is possible to obtain binary descriptions if the considered problem has a solution in $\{0, 1\}$ [8]. If transformation results in obtaining $k > 2$, then tools of k-valued logic can be used as well. From the definition of dissimilarity matrix it follows that the suggested approach can be used also for symbolic or mixed symbolic-numerical descriptions of the objects studied.

When features are numerous some of them might depend on others. The use of methods of decreasing the dimensionality of the features space, for example, through transforming the features space or by disregarding the non-informative features always results in a partial loss of information. Let us point out that it is not necessary to decrease the number of features when using the method of learning Boolean formulas since the interrelationship between them is given by the notion of non-reducible descriptor.

Learning based on creating Boolean formulas can be used also for solving the following problems: finding the weight of a feature, finding out the weight of an object or an image as well. Note that the Boolean formulas obtained through the proposed method can be treated as classification rules and can be used either for validation, verification, minimization, etc. of existing classification rules or for generating new classification rules in rule-based recognition systems.

The present paper is partly supported by the Bulgarian Ministry of Education and Science under contract I 24/1991.

References

[1] M.N.Vajnzvag, *Learning algorithms in pattern recognition "Kora", In: Learning algorithms in pattern recognition,* Sov.radio, Moscow, 1973, (in Russian).

[2] L.V.Baskakova and J.I.Zhuravlev, "A model of pattern recognition algorithms with representative sequences and systems of supporting sets," *USSR Comp. Math. Math. Phys.*, Vol. 21, pp. 1264-1275, 1981, (in Russian).

[3] E.V.Djukova, *Pattern recognition algorithms of the "Kora" type, In: Pattern recognition, classification, forecasting - Mathematical techniques and their applications, Issue 2,* Nauka, Moscow, 1989, (in Russian).

[4] L.G.Valiant, "A theory of the learnable," *Commun. ACM*, Vol 27, N 11, pp. 1134-1142, 1984.

[5] H.Shvaytser, "Learnable and nonlearnable visual concepts," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol.24, N 5, pp. 459-466, 1990.

[6] T.G.Dietterch and R.S.Michalski, *A comparative review of select methods for learning from examples, In: R.S.Michalski, J.G.Carbonel and T.M. Mitchell (Eds.), Machine learning: An artificial intelligence approach,* Tioga, Palo Alto, CA, 1983.

[7] J.I.Zhuravlev, "On the algebraic approach to pattern recognition problems or classification," *Probl. Cybernet.*, Vol. 33, pp. 5-68, 1978, (in Russian).

[8] V.Valev and J.I.Zhuravlev, "Integer-valued problems of transforming the training tables in k-valued code in pattern recognition problems," *Pattern Recognition*, Vol.24, N 4, pp. 283-288, 1991.