Generalized Non–Reducible Descriptors

Ventzeslav Valev, IAPR Fellow
Department of Computer Science,
Fac. Sciences, St. Louis University,
St. Louis, MO 63156, USA
email: valevv@slu.edu

Bülent Sankur
Department of Electrical
and Electronic Engineering,
Bogazici University,
80815 Bebek, Istanbul, Turkiy

Petia Radeva
Computer Vision Center
Bellaterra (Barcelona),
08193 Spain
email: petia@cvc.uab.es

Abstract

This paper provide a generalization of non-reducible descriptors. Non-reducible descriptors are used in supervised pattern recognition problems when the pattern descriptions consist of Boolean variables. This generalization extends the concept of distance between patterns of different classes. A mathematical model to construct generalized non-reducible descriptors, a computational procedure, and numerical examples are discussed.

Keywords: Supervised Pattern Recognition, Machine Learning, Descriptors, Non-Reducible Descriptors, Generalized Non-Reducible Descriptors

1. Introduction

In this paper, we address the supervised pattern recognition problem where the mathematical model is based on Boolean formulas. A typical example of a pattern recognition problem with binary features would be a medical diagnosis based on the presence or absence of a number of symptoms. The minimal combination of such features is called a *Non-Reducible Descriptor* (NRD) [7].

The formulation of the *supervised pattern recognition problem* is as follows. Let M be a set of patterns, each pattern denoted by Q. The set M is a union of a finite number of l disjoint subsets $M = \cup_{j=1}^{l} K_j, K_i \cap K_j = \emptyset, i \neq j$, which are called *classes*. The set M is not completely known. The only information known is the training set $M' \subset M$ containing m elements and how M' is divided into l classes. The pattern recognition problem for an arbitrarily chosen pattern $Q \in M \backslash M'$ consists of determining the values of the predicates $Q \in K_j, j = 1, \ldots, l$ using the training set and the description of the pattern Q. We will consider the case when the training set is given by the training table $T_{m,n,l} = [t_{i,j}], i = 1, \ldots, m; j = 1, \ldots, n$, where the line i corresponds to the description of the pattern $Q_i, i = 1, \ldots, m$.

The *descriptor* of a certain pattern is a sequence of values of its features that makes it different from the descriptions of patterns of the remaining classes. An Non-Reducible Descriptor (NRD) is a descriptor of minimal length. In other words, if any of its arbitrarily chosen feature values are disregarded, then this NRD is no longer a descriptor. Thus, the NRD discriminates the pattern from all patterns of the remaining classes.

In this approach the training stage refers to the process of constructing NRDs for all the classes (NRD sets). The process of construction of NRDs inherently contains the process of feature selection and reduction. In this paper the NRD concept is extended to *k-Non-Reducible Descriptors* (k-NRDs). The k-NRDs are also called *Generalized Non-Reducible Descriptors* (GNRD). The k-NRD is a descriptor for which the Hamming distance to all the patterns from the remaining classes is at least k.

The proposed technique has some similarities to the n-tuples techniques for OCR feature extraction in handprinted character recognition [2], [3], [6]. An n-tuple is a collection of n different binary features that "fit" the descriptions of some patterns and do not "fit" the descriptions of other patterns from the training set. Therefore, the n-tuple is designed to dichotomize a set of patterns. In other words an n-tuple is associated with the presence or absence of a specific black and white pixel configuration in a given pattern.

The approach to finding NRDs and GNRDs differs from the approach to extract n-tuples in two aspects. First, NRDs and GNRDs are constructed for a given pattern and are properties of that pattern. NRDs and GNRDs discriminate that pattern from all the patterns of the remaining classes. The n-tuples dichotomize the training set on two subsets. The second difference is related to the length of the descriptors. In the n-tuples, the value of n is experimentally found and all n-tuples have the same length. In contrast, the NRD is a descriptor with minimal length and hence, different NRDs for a given pattern may have different length. The GNRD is also a descriptor of minimal length. The lengths of NRDs and GNRDs are found automatically during the

Proceedings of the International Conference on Pattern Recognition (ICPR'00)
1051-4651/00 $10.00 @ 2000 IEEE

process of their construction.

2. Non-Reducible Descriptors

Definition 1. *Let $Q_r = (t_{r1}, t_{r2}, \dots, t_{rn})$. The sequence $(t_{rj_1}, t_{rj_2}, \dots, t_{rj_d})$, $j_d \leq n$ is called a descriptor for pattern $Q_r \in K_i$ if there does not exist any pattern $Q_s \in K_p$, $p = 1, 2, \dots, i-1, i+1, \dots, l$ in the training table with the same sequence.*

Definition 2. *A given descriptor is called a Non-Reducible Descriptor (NRD) if none of its arbitrarily chosen subsets is a descriptor.*

Definition 2 means that if an arbitrarily chosen feature is removed, then this descriptor loses its property of descriptor. Therefore, the NRD is a descriptor of minimal length. Next, we assume that the NRD of pattern Q_r is given by its indices $j_1, \dots, j_d$. Let us consider the problem of obtaining the NRD set of a given pattern $Q_r \in K_i, i = 1, \dots, l$. Let the number of patterns which do not belong to K_i be m'.

Definition 3. *The dissimilarity matrix for a pattern $Q_r \in K_i$ is a binary matrix $L_r = [l_{v\,j}]$; $v = 1, \dots, m'$, $j = 1, \dots, n$ obtained as follows:*

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

where $t_{r\,j}$ and $t_{v\,j}$ are the values of feature j of $Q_r \in K_i$ and $Q_v \notin K_i$, respectively.

Note that from the condition $K_i \cap K_j = \emptyset$ for $i \neq j$; $i, j = 1, \dots, l$ it follows that every row of the matrix L_r contains at least one unit.

Definition 4. *The number of elements d in an NRD is called a rank and is denoted by NRD ^{d}.*

Definition 5. *Columns $j_1, j_2, \dots, j_d$ of an arbitrary $\{0,1\}$-matrix M of dimension $(m \times n)$ form a covering of matrix M if there does not exist row p, $p = 1, 2, \dots, m$, such that $m_{p,j_q} = 0$ for $q = 1, 2, \dots, d$.*

Theorem 1. [7] *A pattern Q_r possesses an NRD ^{d} iff*

a) *there exists a finite number of permutations of rows and columns that transform the dissimilarity matrix L_r into a matrix:*

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

where the submatrix E_d is a identity submatrix of dimension $(d \times d)$ and further permutations cannot result in obtaining a larger identity submatrix comprising E_d;

b) *the covering condition of submatrix P_2 holds.*

E_d is the maximal identity submatrix of dimension $(d \times d)$, where d is the rank of the constructed NRD. The indices of the columns of E_d define the indices of Boolean variables included in NRD ^{d} expressed as a conjunction with rank d. This problem always has a solution since any dissimilarity matrix L_r contains at least one unit in each row due to our assumption of class disjointness.

3. Generalized Non-Reducible Descriptors

Let the training table $T_{n,m,l}$ be given. Let the pattern $Q_r \in K_i$. Let d and k be positive integers, where $d \in [3, n]$ and $k \in [2, d-1]$, and $k < d$.

Definition 6. *A descriptor $(t_{r\,j_1}, t_{r\,j_2}, \dots, t_{r\,j_d})$ of pattern Q_r is called a k-Non-Reducible Descriptor (k-NRD) if none of its arbitrarily chosen subsets, containing $(d-k)$ elements is a descriptor.*

Definition 7. *The number of elements d in a k-NRD is called a rank and is denoted by k-NRD ^{d}.*

The following properties hold:

Property 1. *The rank d of each k-NRD ^{d} is minimal with respect to the parameter k.*

Let k-NRD ^{d} be constructed on the set of features $\{j_1, j_2, \dots, j_d\}$. This property means that there is no $(k-1)$-NRD ^{d} on the same set of features $\{j_1, j_2, \dots, j_d\}$.

Property 2. *For each k-NRD ^{d} there exists d in number different $(k-1)$-NRD $^{d-1}$.*

Definition 8. *Columns $j_1, j_2, \dots, j_d$ of $\{0,1\}$-matrix M of dimension $(m \times n)$ form a k-covering of matrix M if in each row p, $p = 1, 2, \dots, m$, there are at least k units.*

Let us remove from the training table $T_{n,m,l}$ all rows belonging to class K_i with exception the row r. Let us divide the remaining rows on two classes so that the first class is represented only by pattern Q_r, and the second class is the union of all other classes $K_j, j = 1, \dots, l$, $j \neq i$. Let the second class contain m' patterns.

Let $s = \binom{d}{k}$, where $k < d$, $k = 2, 3, \dots, d-1$. We assume that $m' \geq s$. Analogous to Theorem 1, the following theorem can be proven.

Theorem 2. *A pattern Q_r possesses a k-NRD ^{d} iff*

a) *there exists a finite number of permutations of rows and columns that transform the dissimilarity matrix L_r into a matrix:*

$$L_r^{''} = \begin{bmatrix} F_d & R_1 \\ R_2 & R_3 \end{bmatrix}, \qquad (2)$$

where the submatrix F_d is of dimension $(s \times d)$ in which k units are distributed in d positions in all possible ways (s in numbers), and $k < d$;

b) *the k-covering condition of submatrix R_2 holds.*

The k-covering condition of matrix R_2 is necessary to keep the pattern Q_r in a Hamming distance at least k far from all patterns that are not involved in the construction of k-NRD ^{d}. The computational procedure for construction the k-NRD ^{d} set for pattern Q_r follows from Theorem 2:

1. Find out in L_r all the different matrices F_d with the given property by all the possible permutations of rows and columns.

2. Check the k-covering condition for matrix F_d.

The indices of the columns in matrix F_d define the indices of Boolean variables included in the k-NRD ^{d}. If $k = 1$ then an NDR ^{d} is obtained. Therefore, the k-NRD ^{d} is an NRD ^{d} for which the distance to the remaining classes is at least k, $k \geq 1$.

We will discuss the suggested procedure. Let the minimal values of $d = 3$ and $k = 2$ be given. By permuting the rows and the columns, all the possible matrices F_d are constructed whenever they exist. If no k-NRD ^{d} is constructed for a given d and k, then the value of k is increased, $k = 3, \ldots, d-1$ and the procedure is repeated. The value of d is increased after exhausting k. In other words the procedure to obtain the set of k-NRD ^{d} is to construct all 1-NRD and, if possible, to extend them to k-NRD ^{d} for $k = 2, 3, \ldots,$ until the maximal value of k is achieved. Note that the problem for constructing k-NRD ^{d} for $k > 1$ does not always have a solution.

Example 1. Given the following training table $T_{7,14,2}$:

$$T_{7,14,2} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Let the patterns $Q_1, Q_2 \in K_1$ and $Q_3, \ldots, Q_{14} \in K_2$. The sequence $(t_{1,4}, t_{1,5}, t_{1,6}, t_{1,7}) = (0, 1, 0, 0)$ is a descriptor and $(t_{1,4}, t_{1,5}, t_{1,6}) = (0, 1, 0)$ is an NRD^3 of pattern Q_1, and $(t_{2,3}, t_{2,4}, t_{2,5}) = (0, 0, 1)$ is an NRD^3 of pattern Q_2. These two NRDs may be expressed respectively by conjunctions $\bar{x}_4 x_5 \bar{x}_6$ and $\bar{x}_3 \bar{x}_4 x_5$. The dissimilarity matrix L_1 for the pattern $Q_1 \in K_1$ is as follows:

$$L_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

If dissimilarity matrix L_1 is subjected to permutations by rows and columns, then matrix $L_1^{''}$ will be as follows:

$$L_1^{''} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

A submatrix F_5 with dimension (10×5) is obtained in the upper left corner of matrix $L_1^{''}$. In F_5 two units are distributed in all 10 possible ways in 5 columns. Therefore, according to Theorem 2, elements $(t_{1,1}, t_{1,2}, t_{1,3}, t_{1,6}, t_{1,7})$ form a 2-NRD ^{5} for the pattern Q_1. It is easy to see that if two arbitrarily chosen features are removed, then this descriptor loses its property of being descriptor. The Hamming distance between the constructed 2-NRD ^{5} and all the descriptions of patterns from the class K_2 involving only the sequence of features $(1, 2, 3, 6, 7)$ is at least 2. The obtained 2-NRD ^{5} can be expressed as the conjunction $x_1 \bar{x}_2 x_3 \bar{x}_6 \bar{x}_7$.

Analogous to NRD, the problem for transforming matrix L_r into a matrix $L_r^{''}$ of kind (2) belongs to the class of NP-complete problems. The proof is based on a polynomial reducing the d-clique problem to our problem. The computational complexity of the method depends on the number of units in the dissimilarity matrix. The number of units represents the degree of dissimilarity between patterns. Therefore, the closer the descriptions between patterns belonging to different classes are, the more efficient the proposed method of learning Boolean formulas will be. Note that "even more helpful are counterexamples that are 'near misses' - that is, negative examples that just barely fail to be positive examples" [1], [4].

4. Decision rules

The decision rule may include constructed NRDs and GNRDs in different ways. The decision may be accomplished by searching for true conjunctions in the description of recognized patterns through a voting procedure. A vote is given for a recognized pattern Q if its description comprises an NRD or GNRD. Votes are counted for all the classes. The simplest decision rule is the majority vote.

Let $Q = (x_1, x_2, \ldots, x_n)$ be the description of a recognized pattern. Let $m_1, m_2, \ldots, m_l$ be the number of patterns in $K_1, K_2, \ldots, K_l$, respectively. Let $n_1, n_2, \ldots, n_l$ be the number of votes (true conjunctions) given for the pattern Q from classes $K_1, K_2, \ldots, K_l$ respectively. The rule of maximum r can be expressed as:

$$r = \begin{cases} Q \in K_j, & \text{if } \max\left(\frac{n_1}{m_1}, \frac{n_2}{m_2}, \ldots, \frac{n_l}{m_l}\right) = \frac{n_j}{m_j}, \\ \text{undefined}, & \text{otherwise.} \end{cases}$$

Example 2. Let the Arabic numerals be given by the following representations as (5×5)-matrices [5]:

```
00100   01110   01110   00110   11111
01100   10001   10001   01010   10000
00100   00010   00110   10010   11110
00100   01000   10001   11111   00001
01110   11111   01110   00010   11110
    01111   11111   01110   01110   01110
    10000   00010   10001   10001   10001
    11110   00100   01110   01111   10001
    10001   01000   10001   00001   10001
    01110   10000   01110   11110   01110
```

The results of the training stage are:

Numeral	1	2	3	4	5
Number of NRDs	179	190	204	121	136
Numeral	6	7	8	9	0
Number of NRDs	142	176	9	140	116

The Boolean formula constructed for the numeral 8 is:

$$\begin{aligned} f(8) \; = \; & \bar{x}_{1,5}x_{3,2}\bar{x}_{3,5} \vee \bar{x}_{1,5}x_{3,2}x_{4,1} \vee \bar{x}_{1,5}x_{3,2}\bar{x}_{5,1} \vee \\ & x_{2,5}x_{3,2}\bar{x}_{3,5} \vee x_{2,5}x_{3,2}x_{4,1} \vee x_{2,5}x_{3,2}\bar{x}_{5,1} \vee \\ & \bar{x}_{3,1}x_{3,2}\bar{x}_{3,5} \vee \bar{x}_{3,1}x_{3,2}\bar{x}_{5,1} \vee \bar{x}_{3,1}x_{3,2}x_{4,1}. \end{aligned}$$

Let us assume that a few binary elements in the descriptions of the numerals are changed due to noise. This will reflect on the decision rule on the following manner. Let p_i conjunctions fail on the numeral i and another q_i conjunctions vote wrongly for numeral i from numerals different than i. If the noise of recognized patterns is moderate, then due to the big number of NRDs (GNRDs), the decision rule of maximum still correctly recognizes patterns. Hence, due to the multitude of NRDs (GNRDs), the recognition system will be robust against moderate noise and distortions.

5. Conclusions

The NRD and GNRD concepts can be easily extended to non-binary features. For example, when the features take value from the set of real numbers R, the dissimilarity matrix L_r can be constructed as follows: 1, if $|t_{r\,j} - t_{v\,j}| > \varepsilon_j$ and 0, otherwise, where ε_j is a chosen threshold. Another approach to obtain binary features is by transforming the description of the training set into k different values, where k is an integer, $k \geq 2$, [8]. If $k > 2$ is obtained, then NRDs (GNRDs) may be expressed using the tools of the k-valued logic. Another possible extension of NRD and GNRD concepts deals with fuzzy theory.

Potential applications of the proposed approach lie in many fields, including medicine, molecular biology (for example, protein family classification), social sciences, etc.

6. Acknowledgments

This work was partly supported by the grant of the Scientific and Technical Research Council of Turkey (TÜBİTAK) within the framework of the NATO-CP Advanced Fellowship Program, 1997 and by CICYT of Ministerio de Educacion y Ciencias of Spain under project TIC98-1100. The authors thank Dr. Eric Rouchka for his useful comments and Ms. Brenda Rainey for her secretarial assistance.

References

[1] T. Dietterich and R. Michalski. *A Comparative Review of Select Methods for Learning from Examples. In: Machine Learning: An Artificial Intelligence Approach, R.S. Michalski, J.G. Carbonell, and T.M. Mitchell, (Eds.)*. TIOGA Publishing, Palo Alto, CA, 1983.

[2] J. Dz-Mou, M. Krishnamoorthy, and A. Shapira. n-tuple features for ocr revisited. *IEEE Trans. PAMI*, 18(7):734–745, 1996.

[3] M. Levine. Feature extraction: A survey. *Proceedings of the IEEE*, 57(8):1391–1407, 1969.

[4] R. Michalski, J. Carbonell, and T. Mitchell. *Machine Learning: An Artificial Intelligence Approach, Vol.II*. Morgan Kaufmann, Los Altos, CA, 1986.

[5] G. Nagy. *Feature Extraction on Binary Patterns, In: Machine Recognition of Patterns*. A.K. Agarwala (Ed.) IEEE Press, New York, 1997.

[6] J. Ullmann. Experiments with the n-tuple method of pattern recognition. *IEEE Trans. on Computers*, pages 1135–1137, December 1969.

[7] V. Valev and P. Radeva. On the determining of non-reducible descriptors for multidimensional pattern recognition problems. *Pattern Recognition and Image Analysis*, 3(3):258–265, 1993.

[8] V. Valev and Y. Zhuravlev. Integer-valued problems of transforming the training tables in k-valued code in pattern recognition problems. *Pattern Recognition*, 24(4):283–288, 1991.