PERGAMON

Available at www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Pattern Recognition 37 (2004) 1809 – 1815

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Generalized non-reducible descriptors

Ventzeslav Valev[a,*], Bülent Sankur[b]

[a] Department of Computer Science, Parks College of Engineering and Aviation, Saint Louis University, 3450 Lindell Boulevard, St. Louis, MO 63103, USA

[b] Department of Electrical and Electronic Engineering, Bogazici University, 80815 Bebek - Istanbul, Turkey

Received 10 June 2003; received in revised form 22 March 2004; accepted 22 March 2004

Abstract

This paper provides a generalization of non-reducible descriptors by extending the concept of distance between patterns of different classes. Generalized non-reducible descriptors are used in supervised pattern recognition problems where the feature vectors consist of Boolean variables. Generalized non-reducible descriptors are expressed as conjunctions and correspond to syndromes in medical diagnosis. Generalized non-reducible descriptors minimize the number of operations in the decision rules. 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; Syndromes

1. Introduction

As a rule, all models of solving pattern recognition problems use the concepts of similarity or dissimilarity to a certain extent. These concepts are involved in the mathematical models of learning procedures. In this paper, we address the supervised pattern recognition problem where the mathematical model is based on Boolean formulas. In other words, the information needed for pattern classification is included in various combinations of binary features. 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 formulation of the supervised pattern recognition problem is as follows. Let M be a set of patterns, where each pattern is denoted by Q. Set M is a union of a finite number of l disjoint subsets $M=\bigcup_{j=1}^{l} K_j$, $K_i \cap K_j = \emptyset$, $i \neq j$, which are called classes. 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. Each class K_j, $j=1,\dots,l$ is defined by its patterns Q. The supervised pattern recognition problem for an arbitrarily chosen pattern $Q \in M - M'$ consists of determining the values of the predicates $Q \in K_j$, $j=1,\dots,l$ using the training set and the description of pattern Q.

We will consider the case when the training set is given as a training table $T_{m,n,l}=[t_{i,j}]$, $i=1,\dots,m$; $j=1,\dots,n$, where line i corresponds to the description of pattern Q_i, $i=1,\dots,m$ in the system of features $x_1,\dots,x_n$. Supervised pattern recognition problems involving binary features are those in which each pattern Q may be represented by an n-dimensional vector of the form $(t_1,t_2,\dots,t_n)$, where $t_j \in \{0,1\}$ for $j=1,2,\dots,n$. Note that the training table is a $\{0,1\}$-matrix for this problem.

It is important to realize that each row of the training table may contain redundant information. Therefore, it may be worthwhile to remove such redundant information. The descriptor of a certain pattern in a particular class is a sequence of values of its features that makes it different from the descriptions of patterns of the remaining classes. A descriptor that has no redundant information is called a non-reducible descriptor (NRD). An NRD is a descriptor

* Corresponding author. Tel.: +1-314-977-8436; fax: +1-314-977-8390.

E-mail address: valevv@slu.edu (V. Valev).

0031-3203/$30.00 © 2004 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2004.03.008

of minimal length. Thus, the NRD discriminates the pattern from all patterns of the remaining classes and cannot be reduced. In other words, if any of its arbitrarily chosen feature values are disregarded, then this NRD is no longer a descriptor.

In this approach the training stage refers to the process of constructing NRDs for all the classes. 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-NRD). The k-NRDs are also called generalized non-reducible descriptors (GNRD). The k-NRD is a descriptor for which the Hamming distance to the patterns of the remaining classes is at least k.

The proposed technique has some similarities to the N-tuples techniques for OCR feature extraction in hand-printed character recognition [1–3]. 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. For binary patterns, it was proven that the problem of finding a distinguishing tuple is NP-complete task [1].

The approach to finding NRDs and GNRDs differs from the approach to extract N-tuples in two aspects. First, NRDs (GNRDs) are constructed for a given pattern and are properties of that pattern. An NRD (GNRD) discriminates that pattern from all 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 also is a descriptor of minimal length. The length of the NRD (GNRD) is automatically found during the process of its construction.

The paper is organized in the following manner. In Section 2, we present definitions and methods related to NRDs. The mathematical model and computational procedure to construct a GNRD are discussed in Section 3. The computational complexity of the model is discussed as well. Finally, in Section 4 we illustrate the usefulness of our model with an application from recognition of Arabic numerals.

2. Non-reducible descriptors

Let us consider the case in which the description of each pattern Q is given as a sequence $(t_1, t_2, \ldots, t_n)$. The members of this sequence correspond to the presence or absence of features $x_1, x_2, \ldots, x_n$, so that each $t_i \in \{0, 1\}$, $i = 1, \ldots, n$. In other words, if $t_i = 1$, then it corresponds to the presence of the feature x_i, which will be represented by the occurrence of the feature x_i in the descriptor, and if $t_i = 0$ then it corresponds to the absence of the feature x_i which will be represented by the occurrence of the feature $\bar{x}_i$ in the descriptor. For example, if $(t_1, t_3, t_7) = (1, 0, 1)$ is a descriptor for an object, then its representation is given by $x_1\bar{x}_3x_7$.

For ease of reference, we include here some definitions similar to the ones introduced in previous related work [4–6]. We will also describe briefly the computational procedure of [5,6] for the construction of all NRDs for pattern recognition problems with binary features. This will enable us to present the procedure for problems with generalized non-reducible descriptors in a simple and straightforward manner.

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

Definition 2. A given descriptor is called a *non-reducible descriptor* (NRD) if none of its arbitrarily chosen proper subsequences is a descriptor.

Definition 2 means that if an arbitrarily chosen feature is removed, then this descriptor loses its property as a descriptor. Therefore, an NRD is a descriptor of minimal length. Next, we assume that the NRD of pattern Q_r is given by its indices $j_1, \ldots, j_d$. Let us consider the problem of obtaining the NRD set of a given pattern $Q_r \in K_i$, $i = 1, \ldots, l$. 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 $m' \times n$ binary matrix $L_r = [l_{vj}]$ obtained as follows:

$$l_{vj} = \begin{cases} 1 & \text{if } t_{rj} \neq t_{vj}, \\ 0 & \text{otherwise,} \end{cases}$$

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

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

Definition 4. The number of features d in an NRD is called its *rank* and is denoted by NRD^d.

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

Theorem 1 (Valev and Radeva [4]). *The problem for constructing all NRDs for an arbitrarily pattern Q_r is* *equivalent to permuting the rows and columns of the dissimilarity matrix L_r to obtain a matrix L_r' of order $m' \times n$ of the form*

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

satisfying the following properties:

(a) *submatrix E_d is an identity submatrix of order d, and no further permutations of rows and columns of L_r will result in a larger identity submatrix comprising E_d*;
(b) *the columns of the submatrix P_2 form a covering of P_2; in other words, each row of P_2 has at least one unit.*

Therefore, E_d is the maximal identity submatrix, where d is the rank of the constructed NRD. Note that the above problem always has a solution because each row of the dissimilarity matrix L_r must contain at least one unit due to the pairwise disjointedness of the classes.

As promised earlier, the above result immediately yields the following computational procedure for obtaining all NRDs for a given pattern Q_r, which is also due to Refs. [5,6].

(a) By permuting the rows and columns of the dissimilarity matrix L_r, obtain all possible different maximal identity submatrices E_d.
(b) Corresponding to each submatrix E_d thus obtained, check if the columns of P_2 satisfy the covering condition.

The order d of the maximal identity submatrix E_d is the rank of the NRD constructed. The indices corresponding to the columns of E_d also correspond to the indices of the Boolean variables to be included appropriately in the NRD. Thus, an NRD of rank d can be obtained.

The problem for construction of NRD always has a solution since any dissimilarity matrix L_r contains at least one unit in each row due to our assumption of class disjointedness. Note that the problem of transforming matrix L_r into a matrix L_r' belongs to the class of NP-complete problems [7].

3. Generalized non-reducible descriptors

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

Definition 6. A descriptor $(t_{rj_1}, t_{rj_2}, \ldots, t_{rj_d})$ of pattern Q_r is called a *k-non-reducible descriptor*, if none of its arbitrarily chosen proper subsets, containing $d-k$, $(k > 0)$ features is a descriptor.

Definition 7. The number of features d in a k-NRD is called its *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 $\{j_1, j_2, \ldots, j_d\}$. This property means that there is no $(k-1)$-NRD^d on the same set of features $\{j_1, j_2, \ldots, j_d\}$.

Property 2. For each k-NRD^d there are d different $(k-1)$-NRDs^{d-1}.

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

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

Let $s = \binom{d}{k}$, where $k < d$, $k = 2, 3, \ldots, d-1$. We assume that $m' \geqslant s$. Analogous to Theorem 1, we can formulate the following theorem.

Theorem 2 (Valev et al. [8]). *The problem of constructing all k-NRD^d for an arbitrary pattern Q_r is equivalent to permuting the rows and columns of the dissimilarity matrix L_r to obtain a matrix L_r'' of the form*

$$L_r'' = \begin{bmatrix} F_d & R_1 \\ R_2 & R_3 \end{bmatrix}$$

satisfying the following properties:

(a) *F_d is a submatrix of order $(s \times d)$ in which k units $(k < d)$ are distributed in d positions in all possible s in number ways*;
(b) *the columns of submatrix R_2 form at least k-covering of R_2; in other words, each row of R_2 has at least k units.*

The k-covering condition of matrix R_2 is necessary in order 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. There are no conditions on submatrices R_1 and R_3. The computational procedure for the construction of the k-NRD^d set for pattern Q_r consists of two steps:

(a) By permuting the rows and columns of the dissimilarity matrix L_r, obtain all possible different submatrices F_d.
(b) Corresponding to each F_d thus obtained, check if the columns of R_2 satisfy the k-covering condition.

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, $k \geqslant 1$.

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. The existence of matrix L_r'' depends on the distribution of units in matrix L_r. If the matrix L_r'' does not exist, this may limit the potential application of the method.

Example. Let the following training table $T_{14,7,2}$ be given:

$$T_{14,7,2} = \left[\begin{array}{ccccccc} 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ \hline 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array}\right].$$

Let patterns $Q_1, Q_2 \in K_1$ and $Q_3,\ldots,Q_{14} \in K_2$. 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. Since all subsequences of these two NRDs appear in objects of class K_2, it is clear that these NRDs satisfy Definition 2. This means that if from the given NRD an arbitrarily chosen feature is removed, then this NRD loses its property as a descriptor. 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 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 permutation σ by rows:

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 2 & 7 & 12 & 10 & 11 & 4 & 6 & 1 & 9 & 5 & 3 & 8 \end{bmatrix}$$

and permutation μ by columns:

$$\mu = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 2 & 3 & 6 & 7 & 4 & 5 \end{bmatrix},$$

then matrix L_1'' will be as follows:

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

Matrix L_1'' is obtained by checking all possible permutations of rows and columns of matrix L_1. Submatrix F_5 of order (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. Columns of matrix R_2 form a 3-covering of R_2. Therefore, according to Theorem 2, features $(t_{1,1}, t_{1,2}, t_{1,3}, t_{1,6}, t_{1,7})$ form a 2-NRD^5 for pattern Q_1.

It is easy to see that if two arbitrarily chosen features are removed, then this descriptor loses its property as a descriptor. The Hamming distance between the constructed 2-NRD[5] and all the descriptions of patterns from class K_2 involving only the features with indices $(1, 2, 3, 6, 7)$ is at least 2. The 2-NRD[5] obtained can be expressed as conjunction $x_1\bar{x}_2x_3\bar{x}_6\bar{x}_7$.

In Ref. [7], it has been shown that the problem of constructing all NRDs for a pattern using the dissimilarity matrix approach is in the class of NP-complete problems. However, it has been reduced to the well-known d-clique problem in graph theory [9], for which there exists a sub-exponential algorithm requiring $O(d^{\log_2 d})$ computing time [10]. The problem of transforming matrix L_r into a matrix L_r'' also belongs to the class of NP-complete problems. The proof is based on a polynomial reducing the d-clique problem to our problem. Therefore, the problem of transforming an arbitrary $\{0, 1\}$-matrix L_r into a matrix L_r'' can also be solved in satisfactory time, although the problem is NP-complete. For example, if $d = 50$, then the problem can be solved with approximately 15 billions operations, which can be achieved by contemporary computers. 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 the classes. 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” [11,12].

4. An application

The decision rule may include elements from the sets of NRDs in different ways. For example, the decision may be accomplished by searching for elements from the NRD sets in the description of recognized patterns through a voting procedure. A vote is given for a recognized pattern if its description comprises an NRD. This is performed by checking the conjunctions (NRDs or GNRDs) in Q. These votes are counted for all classes. The simplest decision rule is the rule of 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, respectively, given for pattern Q from classes $K_1, K_2, \ldots, K_l$. The decision rule r of the majority vote can be expressed as

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

A modification of the decision rule could include number of the NRDs and the GNRDs for each pattern and their ranks.

Let the Arabic numerals be given by the following representations as (5×5)-matrices [13]:

0 0 1 0 0	0 1 1 1 0	0 1 1 1 0	0 0 1 1 0	1 1 1 1 1	0 1 1 1 1	1 1 1 1 1	0 1 1 1 0	0 1 1 1 0	0 1 1 1 0
0 1 1 0 0	1 0 0 0 1	1 0 0 0 1	0 1 0 1 0	1 0 0 0 0	1 0 0 0 0	0 0 0 1 0	1 0 0 0 1	1 0 0 0 1	1 0 0 0 1
0 0 1 0 0	0 0 0 1 0	0 0 1 1 0	1 0 0 1 0	1 1 1 1 0	1 1 1 1 0	0 0 1 0 0	0 1 1 1 0	0 1 1 1 1	1 0 0 0 1
0 0 1 0 0	0 1 0 0 0	1 0 0 0 1	1 1 1 1 1	0 0 0 0 1	1 0 0 0 1	0 1 0 0 0	1 0 0 0 1	0 0 0 0 1	1 0 0 0 1
0 1 1 1 0	1 1 1 1 1	0 1 1 1 0	0 0 0 1 0	1 1 1 1 0	0 1 1 1 0	1 0 0 0 0	0 1 1 1 0	1 1 1 1 0	0 1 1 1 0

In this example $m = 10$, $n = 25$, and number of classes $l = 10$. The results of the training procedure are given in the next table:

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

For example, 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 some conjunctions fail on the numeral i and some other conjunctions vote wrongly for numeral i from numerals different from i. If the noise of recognized patterns is moderate, then due to the big number of NRDs (GNRDs), the decision rule of majority vote will still correctly recognize patterns. In other words, due to the multitude of NRDs (GNRDs), the recognition system will be robust against moderate noise and distortions.

5. Conclusions

In this paper, we have shown how the dissimilarity-matrix model for pattern recognition problem with binary features may be used to construct generalized non-reducible descriptors. The applicability of the model to a classification problem of Arabic numerals has been presented as an illustration.

The NRD and GNRD concept 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:

$$l_{vj} = \begin{cases} 1 & \text{if } |t_{rj} - t_{vj}| \geqslant \varepsilon_j, \\ 0 & \text{otherwise,} \end{cases}$$

where ε_j is a chosen threshold value for feature x_j $j = 1, 2, \ldots, n$. Another approach to obtain binary features is by transforming the description of the training set into k different values, where k is an integer, $k \geqslant 2$, [14]. If $k > 2$ is obtained as a result of this transformation, then the NRD may be expressed using the tools of the k-valued logic.

Another method for constructing an NRD set for multidimensional pattern recognition problems is based on decomposition of the training table and is given in Ref. [4]. The problem of the asymptotic behavior of the number of non-reducible descriptors and their rank as a function of the dimension of dissimilarity matrix L_r is open.

Some applications of the proposed pattern-recognition model based on the learning of Boolean formulas in computer vision problems such as recognition of Arabic numerals in the case when their descriptions are given by primitive elements or ECG are proposed in Refs. [15] and [5], respectively. Another possible extension of NRD and GNRD concept deals with fuzzy features [16]. Potential applications of the proposed approach lie in many fields, including medicine, molecular biology (for example, protein family classification), social sciences, etc.

References

[1] Jung Dz-Mou, M.S. Krishnamoorthy, A. Shapira, N-tuple features for OCR revisited, IEEE Trans. Pattern Anal. Mach. Intell. 18 (7) (1996) 734–745.

[2] M.D. Levine, Feature extraction: a survey, Proc. IEEE 57 (8) (1969) 1391–1407.

[3] J.R. Ullmann, Experiments with the n-tuple method of pattern recognition, IEEE Trans. Comput. (1969) 1135–1137.

[4] V. Valev, P. Radeva, On the determining of non-reducible descriptors for multidimensional pattern recognition problems, Pattern Recognition Image Anal.: Adv. Math. Theory Appl. 3 (3) (1993) 258–265.

[5] V. Valev, P. Radeva, Construction of Boolean decision rules for ECG recognition by non-reducible descriptors, Proceedings of 13th International Conference on Pattern Recognition, Vol. B, Vienna, Austria, August, 1996, pp. 111–115.

[6] V. Valev, P. Radeva, A method of solving pattern or image recognition problem by learning Boolean formulas, Proceedings of 11th International Conference on Pattern Recognition, Vol. II, Hague, Netherlands, August 30–September 3, 1992, pp. 359–362.

[7] V. Valev, Construction of Boolean classification rules and their applications in computer vision problems, Mach. Graph. Vision 5 (1/2) (1996) 5–23.

[8] V. Valev, P. Radeva, B. Sankur, Generalized non-reducible descriptors, Proceedings of 15th International Conference on Pattern Recognition, Barcelona, Spain, September, 2000, Vol. 2, pp. 394–397.

[9] M. Garey, D. Johnson, Computers and Intractability. A Guide to the Theory of NP-Completeness, Bell Teleph. Lab., New Jersey, 1979.

[10] T. Zeugmann, Parallel algorithms, in: A. Kent, J.G. Williams (Eds.), Encyclopedia of Computer Science and Technology, Marcel Dekker: Portland, OR Vol. 21(6), 1990, pp. 223–244.

[11] T.G. Dietterich, R.S. Michalski, A comparative review of select methods for learning from examples, in: R.S. Michalski, J.G. Carbonel, T.M. Mitchell (Eds.), Machine Learning I: An Artificial Intelligence Approach, Tioga Publ., Palo Alto, CA, 1983.

[12] R.S. Michalski, J.G. Carbonell, T.M. Mitchell, Machine Learning II, Morgan Kaufmann, Los Altos, CA, 1986.

[13] G. Nagy, Feature extraction on binary patterns, in: A.K. Agarwala (Ed.), Machine Recognition of Patterns, IEEE Press, New York, 1977, pp. 124–129.

[14] V. Valev, Yu.I. Zhuravlev, Integer-valued problems of transforming the training tables in k-valued code in pattern recognition problems, Pattern Recognition 24 (4) (1991) 283–288.

[15] V. Valev, P. Radeva, Structural pattern recognition by non-reducible descriptors, Shape, Structure and Pattern Recognition, World Scientific Publishers, Singapore, 1995, pp. 221–230.

[16] V. Valev, A. Asaithambi, Fuzzy non-reducible descriptors, Mach. Graph. Vision 12 (3) (2003) 353–361.

About the Author—VENTZESLAV VALEV graduated from the Technical University of Wroclaw, Poland, in Computer Science, and from Wroclaw University, Poland, in Applied Mathematics in 1970 and 1974, respectively. He obtained his Ph.D. degree in Computer Science from the Computer Center of the Academy of Sciences of Russia in Moscow in 1979 and the Doctor of Science degree in Informatics from the Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences, Sofia, Bulgaria, in 1995. He joined the Department of Artificial Intelligence of the Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences in Sofia in 1971 where until 1997 he worked as associate professor. Since 1999, Dr. Valev is Associate Professor at the Department of Computer Science, Saint Louis University in Saint Louis, Missouri, USA. Dr. Valev is a Fellow of the International Association for Pattern Recognition (IAPR). His research interests include pattern recognition, computer vision, machine learning, and data mining. He has authored or coauthored more than 50 journal and conference papers in the above areas.

About the Author—BÜLENT SANKUR received his B.S. degree in Electrical Engineering from Robert College, Istanbul and completed his M.Sc. and Ph.D. degrees at Rensselaer Polytechnic Institute, New York, USA. He has been active at Bogaziçi University in the Department of Electric–Electronic Engineering in establishing curricula and laboratories, and guiding research in the areas of digital signal processing, image and video compression, and multimedia systems. He was the chairman of the ITC& #8217;96: International Telecommunications Conference and the technical co-chairman of ICASSP& #8217;2000. Dr. Sankur has held visiting positions at the University of Ottawa, Canada; Istanbul Technical University; Technical University of Delft, The Netherlands; and Ecole Nationale Supérieure des Telecommunications, France.