Pattern Recognition Letters 45 (2014) 106–114

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

From binary features to Non-Reducible Descriptors in supervised pattern recognition problems ☆

Ventzeslav Valev *, IAPR Fellow

School of Computing, University of North Florida, Jacksonville, FL 32224, USA

ARTICLE INFO

Article history:
Received 11 September 2013
Available online 27 March 2014

Keywords:
Machine learning of Boolean formulas
Non-Reducible Descriptors
Binary feature selection problem
Combining classifiers
Recognition of Arabic numerals
Recognition of ECG

ABSTRACT

The present paper explores the supervised pattern recognition problem when binary features are used in pattern descriptions. The concept of Non-Reducible Descriptors (NRDs) for binary features is introduced. NRDs are descriptors of patterns that do not contain any redundant information. They correspond to syndromes in medical diagnosis and are represented as conjunctions. The proposed approach is based on learning Boolean formulas. Combinatorial and decision-tree computational procedures for construction of all NRDs for a pattern are presented. The computational complexity of the proposed approach is discussed. The process of construction of all NRDs and the obtained NRDs are used for solving the binary feature selection problem. A procedure for combining classifiers is presented. The proposed approach is illustrated with applications for recognition of Arabic numerals in different graphical representations and recognition of *QRS* complexes in electrocardiograms. The obtained results are discussed.

1. Introduction

Many supervised pattern recognition applications deal with binary features. The information needed for pattern classification is generally 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. In medicine the minimal combination of such features is called a syndrome. A 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. A descriptor that has no redundant information is called a Non-Reducible Descriptor (NRD). In this approach the machine learning procedure corresponds to the process of construction of all NRDs for all classes.

The NRD concept was developed in the middle of the fifties by Cheguis and Yablonskii [1,2], who were working on mathematical methods for detecting faults in electrical circuits. The descriptor concept related to pattern recognition problems was introduced in 1966 by Dmitriev et al. [3]. The idea was based on the works of Cheguis and Yablonskii, and applied to the solution of classification problems in geology. The descriptor is defined as a subset of features differentiating objects from different classes.

The proposed approach has some similarities to the n-tuples approach for OCR feature extraction in the hand-printed character recognition, first published in [4–6]. An n-tuple is a collection of n different binary features that "fits" the descriptions of some patterns and does 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 configuration of features in a given pattern. For binary patterns, it was proven that the problem of finding a distinguishing tuple is a *NP*-complete task [4].

The approach of finding NRDs differs from the n-tuples approach in two aspects. First, NRDs are constructed for a given pattern and are properties of that pattern. An NRD recognizes that pattern from all patterns of the remaining classes, whereas 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 length of the NRD is automatically found during the process of its construction.

The rest of the paper is organized as follows. In Section 2 the definitions of descriptor and Non-Reducible Descriptor are introduced. A combinatorial approach for NRDs construction is proposed. This

☆ This paper has been recommended for acceptance by Eckart Michaelsen.
* Tel.: +1 314 660 7137. Associate member of the Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences, Sofia, Bulgaria.
E-mail addresses: valev@math.bas.bg, ventzeslav.valev@gmail.com

http://dx.doi.org/10.1016/j.patrec.2014.03.010
0167-8655/

approach is based on the dissimilarity matrix concept. A computational procedure for construction of all NRDs for a given pattern is described. Instead of the combinatorial approach for the construction of NRDs, a decision-tree approach is proposed. The procedure resembles the depth-first search approach. In Section 3 the computational complexity of the proposed model is discussed. In Section 4 a new mathematical approach for solving the binary feature selection problem using NRDs is described. Procedures for computing the weights of the features and the weights of the NRDs are presented as well. In Section 5 a computational procedure for combining classifiers based on the computed votes given for different classes is described. An illustration of the proposed approach for recognition of Arabic numerals given in different graphical representations is presented in Section 6. In Section 7 another illustration of the proposed approach for recognition of *QRS* complexes in electrocardiograms (ECG) is presented as well. Section 8 summarizes the obtained results and directions for future work are presented.

2. Non-Reducible Descriptors

Let us assume that a phenomenon to be studied using the available information is in the form of *patterns*. Let us denote by M the set of all such patterns Q. M is viewed as a union of a finite number of subsets $K_1, K_2, \dots, K_l$ which are called *classes*. Let us assume that the classes do not intersect, however their convex hulls can overlap. The available information pertains only to the partitioning of some subset $M' \subset M$, called the *training set*. Let us assume that there are m patterns in M', which are divided into l classes, and each pattern Q is described by n features. This information is organized as a table called the *training table*, denoted by $T_{m,n,l}$, assuming that there are m_1 patterns in the class K_1, m_2 patterns in the class $K_2, \dots, m_l$ patterns in the class K_l. The first m_1 rows of $T_{m,n,l}$ will correspond to the patterns in K_1, the next m_2 rows will correspond to the patterns in K_2, and so on. The *supervised pattern recognition problem* is formulated as follows. Using the training set, the class membership of patterns in the training set, and the description Q, assign a pattern $Q \in M \setminus M'$ to one of the classes $K_1, \dots, K_l$.

Without loss of generality, let us consider the supervised pattern recognition problem with two classes K_1 and K_2. Let each pattern Q be described with binary features in the training table as a sequence $t_1, t_2, \dots, t_n$. The members of this sequence correspond to the presence or absence of features $t_1, t_2, \dots, t_n$, so that each $t_i \in \{0, 1\}$, $i = 1, \dots, n$. In other words, if $t_i = 1$, then it corresponds to the presence of the feature t_i, which will be represented by the occurrence of the feature t_i in the descriptor, and if $t_i = 0$ then it corresponds to the absence of the feature t_i which will be represented by the occurrence of the feature $\bar{t}_i$ in the descriptor. It is important to realize that each row of the training table may contain redundant information. Therefore, it will be useful to remove such redundant information. Let us introduce the following Definitions [7].

Definition 1. Let $Q_r = (t_{r,1}, t_{r,2}, \dots, t_{r,n})$. The subsequence $(t_{r,j_1}, t_{r,j_2}, \dots, t_{r,j_d}), j_d \leqslant n$ is called a *descriptor* for the pattern $Q_r \in K_i$ if no other pattern $Q_s \in K_p,\ p = 1, 2, \dots, i-1, i+1, \dots, l$ exists in the training table T with the same subsequence.

Therefore, the descriptor of a certain pattern in a given class is a sequence of values of its features that makes it different from the descriptions of patterns of the remaining classes. From the condition that classes do not intersect it follows that the description of each pattern Q satisfies the Definition 1, i.e. it is a descriptor.

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

Let us point out that the NRD does not have redundant information. The NRD recognize the pattern from all the patterns of the remaining classes and cannot be reduced. Definition 2 means that if an arbitrarily chosen feature is removed, then this descriptor is no longer a descriptor. Therefore, an NRD is a descriptor of minimal length.

As a rule, each pattern Q creates a set of Non-Reducible Descriptors (NRDs). From Definitions 1 and 2 it follows that the NRDs constructed for a given object could intersect, but one NRD cannot be inserted into another NRD.

Example 1. Let us consider the following training table

$$T_{9,7,2} = \begin{array}{c} \begin{matrix} t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & t_7 \end{matrix} \\ \left[\begin{array}{ccccccc} 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right] \end{array},$$

where Q_1, $Q_2 \in K_1$ and $Q_3, \dots, Q_9 \in K_2$. From Definitions 1 and 2 it follows that $(t_{1,4}, t_{1,5}, t_{1,6}, t_{1,7}) = (1, 1, 0, 1)$ is a descriptor for the pattern Q_1. Also, $(t_{1,4}, t_{1,5}, t_{1,7}) = (1, 1, 1)$ is an NRD for the pattern Q_1, and $(t_{2,2}, t_{2,3}, t_{2,7}) = (0, 0, 0)$ is an NRD for the pattern Q_2. The NRDs for the patterns Q_1 and Q_2 can be expressed by the conjunctions $t_4 t_5 t_7$ and $\bar{t}_2 \bar{t}_3 \bar{t}_7$, respectively.

2.1. Construction of Non-Reducible Descriptors using a combinatorial approach

Let us consider a combinatorial approach for construction of NRDs based on the following dissimilarity matrix concept [7]. Let us assume that the NRD of pattern Q_r is given by the following sequence of features $(t_{j_1}, t_{j_2}, \dots, t_{j_d})$. Let us consider the problem of obtaining all NRD for 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 $m' \times n$ binary matrix $L_r = [l_{vj}]$ where each element is 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 the feature j for $Q_r \in K_i$ and $Q_v \notin K_i$, respectively.