ON STRUCTURAL RECOGNITION WITH LOGIC AND DISCRETE ANALYSIS
Levon Aslanyan, Hasmik Sahakyan

Abstract: The paper addresses issues of special style structuring of learning set in pattern recognition area. Above the regular means of ranking of objects and properties, which also use the structure of learning set, the logic separation hypotheses is treated over the multi value features area, which structures the learning set and tries to recover more valuable relations for better recognition. Algorithmically the model is equivalent to constructing the reduced disjunctive normal form of Boolean functions. The multi valued case which is considered is as harder as the binary case but it uses approximately the same structures.

Keywords: Learning, Boolean function, logic separation.

ACM Classification Keywords: F.2.2 Nonnumerical Algorithms and Problems: Computations on discrete structures

Introduction
Pattern recognition undergoes an important developing for many years. As a research area this is not unimodular like classic mathematical sciences, it has a long history of establishment. Inside the theory there are a number of sub disciplines such as – feature selection, object and feature ranking, analogy measures, supervised and unsupervised classification, etc. The same time pattern recognition is indeed an integrated theory studying object descriptions and their classification models. This is a collection of mathematical, statistical, heuristic and inductive techniques of fundamental role in executing the intellectual tasks, typical for a human being, – but on computers.

In many applied problems with multidimensional experimental data the object description is often non-classical, that is, - not exclusively in terms of only numerical or only categorical features, but simultaneously by both kinds of values. Sometimes, the missing value is introduced so that finally we deal with mixed and incomplete descriptions of objects as elements of Cartesian product of feature values, without any algebraic, logical or topological properties assumed in applied area. How then, do we select in these cases the most informative features, classify a new object given a (training) sample or find the relationships between all objects based on a certain measure of similarity? Logic Combinatorial Algebraic Pattern Recognition is a research area formed since 70’s, that uses a mix of discrete descriptors – and similarities, separation, frequencies, integration and corrections, optimization, and solves the whole spectrum of pattern recognition tasks.

This approach is originated by the work [Dmitriev et al, 1966] that transfers the engineering domain technique of tests for electrical schemes [Chegis, Yablonskii, 1958] to the feature selection and object classification area. The applied task of [Dmitriev et al, 1966] is prognosis for mineral resources. In a basic model authors consider that all features are Boolean. Later formal extensions with different kinds of features appeared, as we do it below for the logic separation analysis. Consider the general form of learning set data, $L$:
### FEATURES

<table>
<thead>
<tr>
<th>OBJECTS</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>...</th>
<th>$x_n$</th>
<th>CLASSES</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{i1}^i$</td>
<td>$a_{i1}^j$</td>
<td>$a_{i1}^k$</td>
<td>...</td>
<td>$a_{in}^i$</td>
<td>$C_1$</td>
</tr>
<tr>
<td>$a_{i1}^j$</td>
<td>$a_{i2}^j$</td>
<td>$a_{i2}^j$</td>
<td>...</td>
<td>$a_{jn}^j$</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>$a_{i1}^m$</td>
<td>$a_{i1}^m$</td>
<td>$a_{i2}^m$</td>
<td>...</td>
<td>$a_{im}^m$</td>
<td>$C_m$</td>
</tr>
<tr>
<td>$a_{i1}^m$</td>
<td>$a_{i1}^m$</td>
<td>$a_{i2}^m$</td>
<td>...</td>
<td>$a_{in}^m$</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Features $x_1, x_2, \ldots, x_n$ are categorical or numerical properties represented by their domains of values $M_1, M_2, \ldots, M_n$. Categorical feature takes values from a set $H_z^i = \{s_1, s_2, \ldots, s_t\}$, and numerical features are from metric spaces assuming the following two types:

- **a)** $H_{k,z} = \{k, k+1, \ldots, r\}$, where $k, r$ are nonnegative integers, and $k < r$.
- **b)** $H_{k,r} = \{\alpha : \alpha \in (k, r)\}$, where $(k, r)$ is an interval on the real number line, and $k < r$.

Distances between the space elements are usual (numerical). This choice of primary/elementary attribute spaces reflects the common/usual situation that exists in application areas of pattern recognition and classification tasks.

A space $M$ in which metric spaces $M_i, i = 1, \ldots, n$, are of $H_z^i, H_{k,z}$, and/or $H_{k,r}$ types, we call $n$ dimensional attribute space.

Test (testor) theory [Dmitriev et al., 1966] is based on a restriction of feature sets and learning elements - with preservation of the basic learning set property - preservation of membership to different classes. A features subset $T = \{x_1, x_2, \ldots, x_n\}$ is called a testor of $L$, if projection of $L$ on $T$ keeps the classes nonintersecting (learning set property). The testor is irreducible when no one $x_{ij}$ feature may be eliminated with conservation of the learning set property. Further a feature ranking is made taking into consideration the frequency of belonging $x_{ij}$ to the testors (irreducible testors). Other measure is introduced taking into account frequency of testors on classes. Testor based supervised classification algorithms are constructed by use of frequency similarity measures. Which is the structural property used from the learning set? This is the pair wise difference of learning set elements from different classes (learning set property). We may suppose that this is part or consequence of the well accepted compactness hypothesis.

On formal basis testor technology is well visible on binary tables. Now it is known that constructing all irreducible testors is an NP hard problem. Approximations are studied as well. There is a high similarity with association rule mining models, especially in part of learning of monotone set structures — be it with frequent itemsets in associative rule mining or irreducible tests and testers in this theory.

Consider $n$-vector $\tau = (\tau_1, \tau_2, \cdots, \tau_n)$, that indicates the types of elementary attribute spaces of space $M$. In further consideration without loss of generality we may suppose that $\tau_j \in \{z, r\}$ (this may be coded by binary 0,1
values). Hence \( M = H_{d_1} \times H_{d_2} \times \cdots \times H_{d_n} \). Then the space \( M \) can be adjusted by a pair of n-vectors \( \vec{c} = (c_1, c_2, \ldots, c_n) \) and \( \vec{d} = (d_1, d_2, \ldots, d_n) \) above the \( \vec{r} = (r_1, r_2, \ldots, r_n) \). n-dimensional attribute space \( M \) is called sub-space of \( M \), if its each metric sub-space \( M_i, i = 1, \ldots, n \) is a sub-space of corresponding to it in \( M \), metric space \( M_i, i = 1, \ldots, n \). Obviously such space \( M \) can be presented by a pair of vectors \( \vec{c'} = (c'_1, \ldots, c'_n) \), \( \vec{d'} = (d'_1, \ldots, d'_n) \), where \( c'_i \geq c_i \) and \( d'_i \leq d_i \), \( i = 1, \ldots, n \). We call the number of positions with \( c'_i \neq c_i \) the dimension of \( M \).

[Vaintsvaig, 1973] formulated one of the basic concept in pattern recognition – the KORA algorithm. KORA is constructing elementary conjunctions that intersect with only one class \( C_i \) of learning set \( L \). Contrary, [Aslanyan, 1975] considers all irreducible conjunction forms that intersects with only one class \( C_i \) of learning set \( L \). This is not the generating idea of this work but is the consequence of the Logic Separation Principle. It is to think, that the Logic Separation Principle is some kind of completion of the well known Compactness hypotheses. The work [Aslanyan, 1975], factually for the first time, is considering learning set elements in a non separated/isolated manner. The concept is used that an element spreads its “similarity”, reduced with distance measure of course, which interrupts facing the different class object. An extension may consider not only pairs of learning set elements – one spreading the similarity measure and the second interrupting that - but also arbitrary subsets/fragments of learning set. Several comments: logically, it is evident that the best learning algorithm must be suited best to the learning set itself (at least the learning set is reconstructable by the information on learning set used by algorithm). This also may use the recognition hypothesis when available. The same learning set fragments play a crucial role in estimating the choice of algorithm by the given learning data. The technical solution of Logical Separation (LS) is by Reduced Disjunctive Normal Forms (RDNF), which answers to all issues, - implementation, complexity, interpretation. Further is important to mention that advanced data mining technique IREP (Incremental Reduced Error Pruning) finds its theoretical interpretation in terms of LS framework mentioned above.

Sub-spaces of \( M \) we will also call intervals of \( M \). Similarly to the testor theory the irreducible subspace cover technology is well visible on binary tables. The mentioned two visible formalisms, in addition to the basic principles by Yu. Zhuravlev [Dmitriev et al, 1966], that appeared in early 70’s, present the basics of Logic Combinatorial Algebraic Pattern Recognition research area.

**Formal extensions**

This point considers the LS model with similarity spread and interruption by learning set elements. We try to extend the basic constructions to the case of complicated attribute spaces. Let us introduce partially defined characteristic functions \( f_i(x_1, x_2, \ldots, x_n), \ i = 1, \ldots, m \), which separate some class \( L_i \) from the rest of classes, \( L_j, j \neq i, j = 1, \ldots, n \). The separation is based on the information about the classification given by subsets \( L_i \), \( i = 1, \ldots, m \) in learning set.

\[
\tilde{f}_i = f(\alpha_1, \alpha_2, \ldots, \alpha_n) =
\begin{cases}
1, & \text{if } \tilde{\alpha} \in L_i \\
0, & \text{if } \tilde{\alpha} \in L \setminus L_i \\
\text{undefined}, & \text{if } \tilde{\alpha} \in M \setminus L
\end{cases}
\]
This system of \( m \) partly defined functions will be considered in an analogous way to the class of the algebra of logic (Boolean) functions. For system \( \tilde{F}_i, i = 1, \ldots, m \) we will use analogous constructions [Yablinski, 1958], to the concepts that are well known for functions of the algebra of logic.

The subspace \( \mathcal{M} \) of \( \mathcal{M} \) we call an interval of the function \( \tilde{f}(x_1, \ldots, x_n) \) if

1) \( \mathcal{M} = \mathcal{M} \setminus \bigcup_{j=1}^{m} L_j \)

2) \( \mathcal{M} \cap L_j = \emptyset \)

Let \( \tilde{U} \) be an interval of \( \tilde{f}(x_1, \ldots, x_n) \). Then \( \tilde{U} \) is called the maximal interval if there is no other interval \( \tilde{V} \) of \( \tilde{f}(x_1, \ldots, x_n) \), such that \( \tilde{V} \subseteq \tilde{U} \).

We denote by \( U_i, \ i = 1, \ldots, p \), the set of all maximal intervals of \( \tilde{f}(x_1, \ldots, x_n) \). By analogy to the similar constructions of the algebra of logic, we call the set of \( U_i \) reduced interval normal form /or interval cover form/ of \( \tilde{f}(x_1, \ldots, x_n) \).

Further, for each function \( \tilde{f}(x_1, \ldots, x_n), \ i = 1, \ldots, m \) we need to construct \( U_i \), - the set of its all maximal intervals. This question is addressed by [Aslanyan, 1975] in the particular case when \( M_i \equiv H^r_{k,i}, \ i = 1, \ldots, m \).

Other methods designed for the mentioned constructions in the class of the algebra of logic are given in [Yablinski, 1958] /for \( k \)-valued logic/. But the learning set \( L \) in our tasks is very specific – it consists of a finite and very low number of points, and therefore growing type methods analogous to those given in [Yablinski, 1958] for constructing reduced interval forms, are less efficient than the step by step constructions given by sets \( L_i, \ i = 1, \ldots, m \). Notice also that if among the metric spaces \( M_i \) there exists a space of the type \( H^r_{k,r} \), then the direct use of methods from [Yablinski, 1958] is simply impossible. Below we describe the geometric constructions that appear in step by step procedures for reduced interval covers of systems of partial separation functions.

Our goal is to demonstrate that there is a high similarity with the binary case with additions, that constructed intervals are accompanied with small size descriptive data.

**Reduced interval covers of systems if partial separation functions**

Before starting to construct reduced interval forms \( U_i \) of \( \tilde{f}(x_1, x_2, \ldots, x_n) \) let us assume that intervals \( \{k,r\} \) of individual directions defined above /specifically the real value case/ include their endpoints in a following way: point is inclusive iff it is a boarder point of corresponding \( M_i \) and doesn’t match to one of the considered vertices of the learning set. Different coding is possible. For \( H^r_{k,r} \) we may use the instant intervals, - all end points included. For \( H^r_{k,r} \) we use the scheme that the end point is an exclusive real value. We need this for only internal point indication, and end points of \( M_i \) are always inclusive when doesn’t match to one of the considered vertices of learning set. Algorithmically, exclusive end point is very easy to code – the use of negative value of corresponding coordinate may code this case.

Thus we consider functions \( \tilde{f}(x_1, \ldots, x_n) \) which separate the sets \( L_j \) from the complementary sets \( L_j, \ j \neq i, \ j = 1, \ldots, m \). Construct sets \( U_i \) - the sets that consist of corresponding maximal intervals. Let \( \tilde{P}_1, \tilde{P}_2, \ldots, \tilde{P}_{k_i} \) be points of \( M \), that belong to \( L_j, \ j \neq i, \ j = 1, \ldots, m \), where \( k_i = m - (m_i - 1) \). Let \( \tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_{n_i} \) be points...
of \( L_i \), where \( r_i = m_i - m_{i-1} \) (\( m_i \) are defined by table of learning set given above as the last enumerated index in the class \( L_i \)).

In general we are given also the point \( \vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{M} \) that is to be recognized/arranged into one of the classes \( C_i, i = 1, \ldots, m \).

Consider the first point \( \vec{\beta}_1 = (\beta_{11}, \beta_{12}, \ldots, \beta_{1n}) \). Then an interval \( \vec{U} \) will be maximal out of \( \{\vec{\beta}_i\} \) if and only if at least one of its elementary intervals \( \{c'_i, d'_i\}, i = 1, \ldots, n \) has no intersection with the corresponding \( \beta_{1i} \), and all these intervals are maximal in \( \{c_i, d_i\} \) out of the corresponding \( \beta_{1i} \). Besides, it is clear that all but one elementary intervals from \( \{c'_i, d'_i\}, i = 1, \ldots, n \) should be coincide with the corresponding \( \{c_i, d_i\} \), and the presence of exactly one interval \( \{c'_i, d'_i\} \) which does not contain \( \beta_{1i} \), - is sufficient. Let \( s_1, \ldots, s_k \) be the coordinates of \( \vec{\beta}_1 \) such that the following relations take place simultaneously:

\[
\beta_{1i} = c_{si} \quad \text{and} \quad \beta_{1i} = d_{si} \quad \text{for} \quad j = 1, \ldots, k.
\]

Let \( s_{k+1}, \ldots, s_n \) - are the rest of coordinates of \( \vec{\beta}_1 \). In accordance with the reasoning above, we make the following table for maximal intervals after \( \vec{\beta}_1 \):

\[
\{c_{s_1, d_{s_1}}, \{c_{s_2, d_{s_2}}, \ldots, \{c_{s_n, d_{s_n}}\}
\]

Here several cases are possible. \( \nu = 1 \) and \( c_i = c_i, d_i = d_i, i = 1, \ldots, n \) if \( \vec{\beta}_1 \) doesn't prick interval \( \vec{U} \). This is when at least in one direction, \( \beta_{1i} \) is out of \( \{c_i, d_i\} \) or when \( \beta_{1i} \) is the repeated (or boarder) point. This case is not possible in step 1 but can appear later in iterations. Other special cases appear when \( \vec{U} \) became one dimensional. The general construction is as follows in supposition that \( \vec{U} \) is of dimension \( n \):

\[
c_{tv} = \begin{cases} \begin{align*}
c_{s_j} & \text{for } t = 2j - 1 \text{ and } r = s_j \text{ if } \tau_{s_j} = 0 \\
\beta_{s_j} + 1 & \text{for } t = 2j \text{ and } r = s_j \text{ if } \tau_{s_j} = 0 \\
c_{s_j} & \text{for } t = 2j - 1 \text{ and } r = s_j \text{ if } \tau_{s_j} = 1 \\
\beta_{s_j} & \text{for } t = 2j \text{ and } r = s_j \text{ if } \tau_{s_j} = 1 \\
c_{s_j} & \text{for } t = k + j \text{ and } r = s_j \text{ if } d_{s_j} = \beta_{s_j} \\
c_{s_j} + 1 & \text{for } t = k + j \text{ and } r = s_j \text{ if } c_{s_j} = \beta_{s_j} \\
\end{align*} \end{cases}
\]

\[
d_{tv} = \begin{cases} \begin{align*}
\beta_{s_j} - 1 & \text{for } t = 2j - 1 \text{ and } r = s_j \text{ if } \tau_{s_j} = 0 \\
d_{s_j} & \text{for } t = 2j \text{ and } r = s_j \text{ if } \tau_{s_j} = 0 \\
\beta_{s_j} & \text{for } t = 2j - 1 \text{ and } r = s_j \text{ if } \tau_{s_j} = 1 \\
d_{s_j} & \text{for } t = 2j \text{ and } r = s_j \text{ if } \tau_{s_j} = 1 \\
d_{s_j} & \text{for } t = k + j \text{ and } r = s_j \text{ if } d_{s_j} = \beta_{s_j} \\
d_{s_j} & \text{for } t = k + j \text{ and } r = s_j \text{ if } c_{s_j} = \beta_{s_j} \\
\end{align*} \end{cases}
\]

\[
\]
It is easy to check that the rows of the constructed above table are all maximal intervals in $M$ out of the point $\vec{p}_i$. For $s_{k_1},\ldots,s_{k_n}$ just change of the code sign on positive coordinates is applied. Assume that we have constructed a table whose rows are all maximal intervals in $M$ out of $\{\vec{p}_1,\vec{p}_2,\ldots,\vec{p}_m\}$. Then for constructing an analogous table specifying the set $S(\vec{p}_i,\vec{p}_{i_1},\ldots,\vec{p}_{i_n})$, it is sufficient to do the following: for each row of constructed in the previous stage table that specifies the maximal intervals of $M$ out of $\{\vec{p}_1,\vec{p}_2,\ldots,\vec{p}_{i_n}\}$, we construct the set of its maximal subintervals out of $\vec{p}_i$, - and list them as strokes of corresponding tables. It is easy to check that among the maximal intervals out of $\{\vec{p}_i,\vec{p}_2,\ldots,\vec{p}_{i_n}\}$, those that have no intersection with $\vec{p}_i$, will not be changed, and thus the construction of subintervals out of $\vec{p}_i$ is similar to the construction of maximal intervals of $M$ out of $\vec{p}_i$, given in the first step, and because instead of $M$ we can deal with an interval which is maximal out of $\{\vec{p}_1,\vec{p}_2,\ldots,\vec{p}_{i_n}\}$, and instead of $\vec{p}_i$, we take $\vec{p}_i$. We take into consideration the dimension and fixed coordinates of current interval to prick. It is worth to mention that all new intervals that we get in $i$th step, are pair wise different and are not contained in each other. Therefore for the final construction of $S(\vec{p}_i,\vec{p}_{i_1},\ldots,\vec{p}_{i_n})$, all we have to do is to remove from the set of all constructed intervals, those intervals constructed in $i$th step, which are included in others. The algorithm is complete.

Further, for recognition, we will consider the sets $S_i(\vec{r}_i)$ and $S_i(\vec{c}_i,\vec{r}_i)$ of all maximal intervals of $M$ out of $\{\vec{c}_i,\vec{c}_2,\ldots,\vec{c}_n\}$ passing trough the points $\vec{r}_i$ and $\vec{c}_i$ and $\vec{r}_i$, respectively. The construction of these sets $S_i(\vec{r}_i)$ and $S_i(\vec{c}_i,\vec{r}_i)$ can be done either in an analogue way as the construction of $S_i$, or they might be derived from $S_i$ by choosing those maximal intervals that pass trough the points $\vec{r}_i$ and $\vec{c}_i$ and $\vec{r}_i$, respectively.

Let $\vec{u}=(u_1,u_2,\ldots,u_n)$ is a point from $M$, where each coordinate $u_i$ is equal either to $c_i$ or to $d_{i}$. There are $2^n$ such points assuming that $c_i \neq d_{i}$, $i=1,\ldots,n$. Let $\vec{u},\vec{u}_1,\ldots,\vec{u}_n$ denote these points. We call them corner/angular points of $M$. Let $\vec{c}_i \in M$. We consider intervals of the form $\vec{U}(\vec{c}_i,\vec{u})$, $i=1,\ldots,2^n$ in $M$. We call them intervals of directions $\vec{u}_i$, outgoing from the point $\vec{c}_i$. Such way we get in $M$ a system of independent directions, - outgoing from $\vec{c}_i$, which simplifies the study of such point $\vec{c}_i$. This study is fallen into iterations, studying the corresponding pictures/situations in spaces $\vec{U}(\vec{c}_i,\vec{u}_i)$, $i=1,\ldots,2^n$, where $\vec{c}_i$ itself is a corner point of these spaces.

Clearly the above constructions are valid for any subspace (intervals) of $M$. These are useful also for studying $S_i$ sets of maximal intervals of the function $\vec{f}_i$, $i=1,\ldots,m$, as well as for discovering interrelations of points of $M$ and constructions given by the sets $S_i$.

**Conclusion**

The description above is an attempt to interlink several basic ideas of Logic Combinatorial pattern recognition. In the point of view given it appears that ideas are around the recovering more valid relations in the learning set. Learning set (plus the global hypotheses is there is one) is the only information about the classes and its best use is related to selecting its characteristic fragments and constructing the classification algorithms on this basis. Two examples considered are the testor scheme with pair of elements from different classes that are different, and logic separation with similarity spread – interruption fragment. These basic ideas were further initiated as the associated rule generation and incremental reduced error pruning schemes in Data Mining theory. After this
methodological discussion the paper gives the algorithm of constructing the reduced interval covers for systems of discrete functions that just demonstrate the complexity of tasks and the similarity with the binary case.

Bibliography


Authors’ Information

Levon Aslanyan – Head of Department, Institute for Informatics and Automation Problems, NAS RA, P. Sevak St. 1, Yerevan 14, Armenia, e-mail: lasl@sci.am

Hasmik Sahakyan – Leading Researcher, Institute for Informatics and Automation Problems, NAS RA, P. Sevak St. 1, Yerevan 14, Armenia, e-mail: hasmik@ipia.sci.am