LOGIC BASED PATTERN RECOGNITION - ONTOLOGY CONTENT (1) [1]

Levon Aslanyan, Juan Castellanos

Abstract: Pattern recognition (classification) algorithmic models and related structures were considered and discussed since 70s: – one, which is formally related to the similarity treatment and so - to the discrete isoperimetric property, and the second, - logic based and introduced in terms of Reduced Disjunctive Normal Forms of Boolean Functions. A series of properties of structures appearing in Logical Models are listed and interpreted. This brings new knowledge on formalisms and ontology when a logic based hypothesis is the model base for Pattern Recognition (classification).

ACM Classification Keywords: I.5.1 Pattern Recognition: Models

1. Introduction

Pattern Recognition is in reasonable formalization (ontology) of informal relations between objects visible/measurable properties and of object classification by an automatic or a learnable procedure. Among the means of formalization (hypotheses) - metric and logic based ones are the content of series of articles started by the current one. The stage of pattern recognition algorithmic design in 70s dealt with algorithmic models – which are huge parametric structures, combined with diverse optimization tools. Algorithmic Models cover and integrate wide groups of existing algorithms, integrating their definitions, and multiplying their resolution power. Well known example of this kind is algorithmic model of estimation of analogies (AEA) given by Yu. I. Zhuravlev [1]. This model is based indirectly on compactness hypothesis, which is theoretically related to the well known discrete isoperimetric problem (3). The optimization problem of isoperimetry is a separate theoretical issue and its pattern recognition implementations are linked alternatively to the general ideas of potential functions [4]. We present the logical separation (LSA) algorithmic model, as it is described below, to be one of the generalizations of algorithmic model of estimation of analogies. For AEA models a number of useful combinatorial formulas (algorithms) to calculate the analogy measure of objects and of objects and classes were proven [2]. These are the basic values for the decision making rules in AEA. In these models large number of parameters appears, being consecutively approximated using the appropriate optimization procedures. For this reason, a special control set besides the learning set is considered having the same formal structure as the learning set. Considering classification correctness conditions for the set of given objects by the decision procedure we get a system of restrictions/inequalities, which may not be consistent. In the simplest case a system of linear inequalities appear and then we receive a problem of approximating the maximal consistent subsystem of this basic requirements system. In terms of Boolean functions this is equivalent to the well known optimization problem of determining of one of the maximal upper zeros of a Monotone Boolean function when it is given by an operator.

LSA is based on implementation of additional logical treatments on learning set elements, and above the AEA specific metric considerations. Some formalization of additional properties on classification in this case is related to the terms of Boolean functions and especially - to the reduced disjunctive normal forms of them. Let us consider a set of logical variables (properties) $x_1, x_2, \ldots, x_n$ and let we have two types/classes for classification: K_1 and K_2. Let $\beta \in K_1$, and $\gamma \in K_2$, and α is an unknown object in the sense of classification. We say, that γ is separated by the information of β for α if $\beta \oplus \gamma \le \beta \oplus \alpha$, where $\oplus$ is summation by $mod\, 2$ operation. After this assumption we get, that the reduced disjunctive normal forms of two complementary partially defined Boolean functions describe the structure of information enlargement of the learning set. This construction is extending the model of estimation of analogies. It was shown that the logical separators divide the object sets into three subsets, where only one of them needs the treatment by AEA. This set is large enough for almost all weakly

[1] The research is supported partly by INTAS: **04-77-7173** project, http://www.intas.be

defined Boolean functions, but for the functions with the property of compactness it is small. Let, for $0 \le k_0 < k_1 \le n$ F_{n,k_0,k_1} is the set of all Boolean functions consisting of pair of k_0 and $n-k_1$ spheres centered at 0 and 1 respectively as the sets of zeros and ones of the function. On the remainder part of vertices of n-cube the assignment/evaluation of the functions are arbitrary. This functions satisfy the compactness assumptions, and their quantity is not less than $2^{\varepsilon(n)2^n}$ for an appropriate $\varepsilon(n) \to 0$ with $n \to 0$. For these functions, also, it is enough learning set, consisting of any $n2^{n-\varepsilon(n)\sqrt{n}}$ or more arbitrary points for recovering the full classification by means of logical separators procedure. This is an example of postulations considered. The given one is relating the metric and logic structures and suppositions, although separately oriented postulations are listed as. The follow up articles will describe the mixed hierarchy of recognition metric-logic interpretable hypotheses, which helps to allocate classification algorithms to the application problems.

2. Logic Based Model

Solving the main problem of pattern recognition or classification assumes that indirect or informal information or data on classification $K_1, K_2, ..., K_l$ is given. Often this information is in form of appropriate conditions in an analogy to the compactness hypothesis, which in the very common shapes assumes, that given a metric in the space of all objects M and that closer values of classification predicate K corresponds to the pairs of "near" objects of M. We assume that objects of M are coded - characterized by the collections of values of n properties $x_1, x_2, ..., x_n$. Then each object is identified with the corresponding point of the n-dimensional characteristic space. So under the compactness of classes we assume the geometrical compactness of sets of points in the characteristic space, which corresponds to the elements of classes $K_1, K_2, ..., K_l$ and the consecutive adjustments of this property can be given in the following descriptive form: closer neighborhoods of class elements belong to the same class; the distance increase from a class element increases the class change probability; for elements pairs of different classes there exist simple paths in three parts – classes and a limited transition area in the middle.

Above we already considered the general formalization models of hypothesis by metrics and by logic. More formalizations move to more restricted sets of allowable classifications and in this regard it is extremely important to determine the level of formalisms applied. During the practical classification problem arrangements it is to check the satisfaction level of the application problem to the metric and/or logic hypothesis. Resolution is conditioned by the properties of the given learning set $\bigcup_{i=1}^{l} \mathrm{M}_i$. On the other hand there are more different conditions and methods of classification, which are very far in similarity to the model of compactness. These structures require and use other formalisms, providing the solution tools to the wide amounts of practical of pattern recognition problems. Such are the classes of algorithms of estimation of analogies, test's algorithms [2] potential function methods [4], etc.

Note that the arbitrary pattern recognition class problems can be reduced to the others, with the satisfaction of compactness type hypothesis. However this doesn't mean that the compactness hypothesis is universal because the pattern recognition problem's solution for the given space or creation of appropriate transformations to the other problems are the equivalent problems.

Now let us formulate the condition F_0, which will formalize the additional to the compactness hypothesis properties of classes. We'll consider the case of two classes ($l = 2$) intending the formalism simplifications. Particularly, in case of completing of partially defined predicate P, we will base on condition F_0. We'll apply a correspondence of considered object properties and the set of binary variables $x_1, x_2, ..., x_n$, and the same time between the partial predicate P - and it's characteristic function $f(x_1, x_2, ..., x_n)$, and will solve the modeled problem of determining (completing) of the target Boolean function F.

Let $\tilde{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_n) \in \mathrm{M}$.

Consider the determination (completion) of function f in $\tilde{\alpha}$. Take the arbitrary $\tilde{\gamma}$, $\tilde{\gamma} \in \mathrm{M}_1$. If there exists such a point $\tilde{\beta}$, $\tilde{\beta} \in \mathrm{M}_0$, that $\tilde{\gamma} \oplus \tilde{\beta} \le \tilde{\gamma} \oplus \tilde{\alpha}$ (so the $\tilde{\beta}$ is different of $\tilde{\gamma}$ on a subset of the set of properties, where are different $\tilde{\alpha}$ and $\tilde{\gamma}$), then we conclude that $\tilde{\beta}$ logically separates $\tilde{\alpha}$ from $\tilde{\gamma}$, and the information, that $f(\tilde{\gamma}) = 1$ doesn't affect on the determination of the value of the function f in the point $\tilde{\alpha}$ by 1. In the opposite case we'll call $\tilde{\alpha}$ allowable in respect to the point $\tilde{\gamma}$ and to the set M_1 and decide, that information $f(\tilde{\gamma}) = 1$ influence on the determination of $\tilde{\alpha}$ by one, and the real measure of that is given by the value of the object similarity measures.

Consider the following classes of points of the n--dimensional unit cube:

N_0^f -- the set of all $\tilde{\alpha} \in \mathrm{M}$, which are allowable for the set M_0 and not allowable for M_1,

N_1^f -- the set of all $\tilde{\alpha} \in \mathrm{M}$, which are allowable for the set M_1 and not allowable for M_0,

N_2^f -- the set of all $\tilde{\alpha} \in \mathrm{M}$, which are not allowable for the sets M_0 and M_1,

N_3^f -- the set of all $\tilde{\alpha} \in \mathrm{M}$, which are allowable for both the M_0 and M_1.

[3] pointed out the general relation of condition F_0 with the notion of the reduced disjunctive normal form of Boolean functions. To see this relation let us consider the functions f and its negation $\bar{f}$, and let $\Re_f$, $\Re_{\bar{f}}$ correspondingly are the reduced forms for these functions. Denote by:

M_0^f --the collection of all points $\tilde{\alpha}$ for which $(\Re_f)_{\tilde{\alpha}} = 0$, $(\Re_{\bar{f}})_{\tilde{\alpha}} = 1$,

M_1^f --the collection of all points $\tilde{\alpha}$ for which $(\Re_f)_{\tilde{\alpha}} = 1$, $(\Re_{\bar{f}})_{\tilde{\alpha}} = 0$,

M_2^f --the collection of all points $\tilde{\alpha}$ for which $(\Re_f)_{\tilde{\alpha}} = 0$, $(\Re_{\bar{f}})_{\tilde{\alpha}} = 0$,

M_3^f --the collection of all points $\tilde{\alpha}$ for which $(\Re_f)_{\tilde{\alpha}} = 1$, $(\Re_{\bar{f}})_{\tilde{\alpha}} = 1$,

Proposition 1. $\mathrm{N}_i^f \equiv \mathrm{M}_i^f, i = 0.1.2.3.$

Proposition 2. If $\mathrm{M}_0 \cup \mathrm{M}_1 \ne 0$, then M_2^f is empty, in opposite case $\mathrm{M}_2^f \equiv \mathrm{M}$.

It is simply to prove this and some of the consecutive propositions and by this reason we omit the complete proofs and give the main idea of that. So, to prove proposition 2 consider an arbitrary point $\tilde{\alpha} \in \mathrm{M}$. If $\mathrm{M}_0 \cup \mathrm{M}_1 \ne 0$, then let us take the distance of $\tilde{\alpha}$ to the set $\mathrm{M}_0 \cup \mathrm{M}_1$ (which equals the minimal possible distance of $\tilde{\alpha}$ from any of the points of $\mathrm{M}_0 \cup \mathrm{M}_1 \ne 0$), which is in some point $\tilde{\beta} \in \mathrm{M}_0 \cup \mathrm{M}_1$. Suppose, without loss of generality, that $\tilde{\beta} \in \mathrm{M}_0$. Then the interval (binary subcube) $\mathrm{E}(\tilde{\alpha}, \tilde{\beta})$, constructed on base of points $\tilde{\alpha}$ and $\tilde{\beta}$ does not contain points from the set M_1. From here, on base of definition of reduced disjunctive normal form implies, that the point $\tilde{\alpha} \in \mathrm{M}$ is allowable in respect to the set M_0.

Proposition 3. If f_0 is an appropriate completion of function f, constructed on base of condition F_0, then $\forall \tilde{\alpha} \in \mathrm{M}_0^f (f_0(\tilde{\alpha}) = 0)$ and $\forall \tilde{\beta} \in \mathrm{M}_1^f (f_0(\tilde{\beta}) = 1)$.

Proposition 4. $\mathrm{M}_0 \subseteq \mathrm{M}_0^f$ and $\mathrm{M}_1 \subseteq \mathrm{M}_1^f$.

As a consequence from these two propositions we get, that the arbitrary completion of function f , which is made on base of condition F_0 , constructs the function, allowable in respect of f . In terms of pattern recognition problems this means that arbitrary methods of recognition, which are based on the condition F_0, couldn't be "false" on the elements of the learning set $\mathrm{M}_0 \cup \mathrm{M}_1$. Write out the minimal completions of the function f , constructed on base of the condition F_0:

$$f_1(x_1,x_2,...,x_n)=\begin{cases} 0, & if\ (x_1,x_2,...,x_n)\in \mathrm{M}_0^f \\ 1 & if\ (x_1,x_2,...,x_n)\in \mathrm{M}_1^f \\ is\ not\ determined & if\ (x_1,x_2,...,x_n)\in \mathrm{M}\setminus(\mathrm{M}_0^f\cup \mathrm{M}_1^f)=\mathrm{M}_3^f \end{cases}$$

So we get some "enlargement" for the basic function f . There arose a question -- might f_1 be the new starting point (learning set, function) for the completion on base of condition F_0, and how close we can approach by this steps the final goal? The answer gives the

Proposition 5. If f_{i+1} is completion of partial function f_i, constructed on base of condition F_0, $i=1,2,...$, then $f_1 \equiv f_k, k=1,2,...$.

Let us now analyze the conditions, related to the successful continuation on base of F_0 of a partial Boolean function (that is the case of the solvable problems). Let f -- is a partially defined Boolean function and $\varphi_1,\varphi_2,...,\varphi_\tau$ -- all that functions of class $P_2(n)$ which might appear as a continuation of function f , constructed by the given assumptions. Then we are interested in conditions, when extension f_1 is allowable in respect to each of functions $\varphi_1,\varphi_2,...,\varphi_\tau$.

Consider the function f_0, defined in the following way:

$$f_0(x_1,x_2,...,x_n)=\begin{cases} 0, & if\ \varphi_i(x_1,x_2,...,x_n)=0, i=1,2,...,\tau \\ 1, & if\ \varphi_i(x_1,x_2,...,x_n)=1, i=1,2,...,\tau \\ is\ not\ defined & in\ other\ cases \end{cases}$$

Denote by $\mathrm{M}_0(f_0)$ and $\mathrm{M}_1(f_0)$ sets of all n--cube vertices, where function f_0 achieves values 0 and 1 respectively. Then our requirement can be formulated as the following: $\mathrm{M}_0^f \subseteq \mathrm{M}_0(f_0)$ and $\mathrm{M}_1^f \subseteq \mathrm{M}_1(f_0)$. Here $\mathrm{M}_3^f=\mathrm{M}\setminus(\mathrm{M}_0^f\cup\mathrm{M}_1^f)$ and $\mathrm{M}_3^f \supseteq \mathrm{M}\setminus(\mathrm{M}_0(f_0)\cup\mathrm{M}_1(f_0))$ so that this partial continuation doesn't violate the continuality of starting function to the each of functions $\varphi_1,\varphi_2,...,\varphi_\tau$. It is to mention that the conditions $\mathrm{M}_0^f \subseteq \mathrm{M}_0(f_0)$ and $\mathrm{M}_1^f \subseteq \mathrm{M}_1(f_0)$ are not convenient, which is related to the applied information on the final goal (the functions $\varphi_1,\varphi_2,...,\varphi_\tau$). Supposing the case of continuation for needs of pattern recognition problems let us show that practically useful conditions of the given type might be formulated.

Consider the structural behavior, when $n\to\infty$ and suppose a parameter $\theta\to 0$ given as. Suppose $f_0 \in P_2(n)$ (note, that the results below are true in more general cases and in more general forms). Here are some preliminary results.

1. Consider the concept $H_k^-(f_0)$ introduced by Glagolev [7]. $H_k^-(f_0)$ equals the number of vertices $\tilde{\alpha}\in E^n$, where $f_0(\tilde{\alpha})=1$, and which are covered by (involved in) maximal intervals of function f_0 of sizes not exceeding k . It was proven [7] that for almost all functions $f_0 \in P_2(n)$ $H_k^-(f_0)=o(2^n)$ when $n\to\infty$ and $k\le k_1=\log\log n-1$.

2. We'll say, [5] that the cube vertices prick the intervals including these vertexes. The set A of vertices of n-dimensional unit cube is a pricking set for the set B_k -all of the k -size intervals, if each k -subcube is pricked at least by one of the vertices of A. Denote by $K(n,k)$ the minimal number of vertices, forming a pricking set for k -subcubes. By [5] $2^{n-k} \le K(n,k) \le (n+1)2^{n-k}$. We will use the upper bound by this formulae but in our case $k \le k_1 = log\, log\, n - 1$ and a better estimation is possible as follows [4] (an extended survey on pricking is included in [6]). Let us denote by $A_i(\tilde{\alpha})$ the set of all of n-cube vertices, which lay in respect to the given vertex $\tilde{\alpha}$ on layers with numbers $\tau, \tau \equiv i(mod\, k+1)$, $i = 0,1,...,k, k \le n$. Let E^k -is an arbitrary k -subcube of an n-cube. Points of subcube E^k are placed exactly in the $k+1$ consecutive layers of E^n in respect to it's arbitrary vertex $\tilde{\alpha}$. It is correct to post the

Proposition 6. Each of the sets $A_i(\tilde{\alpha}), \tilde{\alpha} \in E^n, i = 0,1,...,k$ are pricking for the set B_k -all of the k -subcubes of n-cube, and $2^{n-k} \le K(n,k) \le 2^n / k+1$.

Proposition 7. F_0 implemented in continuation of almost all functions $f_0 \in P_2(n)$ yields the accuracy, tending to 1 as $n \to \infty$, if for the initial function f holds the condition $\mathrm{M}_0 \cup \mathrm{M}_1 \supseteq A_i(\tilde{\alpha})$ at least for any $i, i = 1,2,...$ and vertices $\tilde{\alpha} \in E^n$, where $A_i(\tilde{\alpha})$ is constructed for a $k \le [log\, log\, n] - 1$.

Note, that the proposiition 7 postulation is constructive, and it implies to the "sufficient" learning set, consisted no more that from $2^n / log\, log\, n$ points (which is $o(2^n)$) as $n \to \infty$. However, basically, in the pattern recognition problems it is impossible to obtain the learning set arbitrarily. Often it is formed as a random collection of any fixed size of the main collection of the studying objects.

Conclusion

Logic Separation is an element of pattern recognition hypotheses and formalisms. Structures appear in this relation based and introduced in terms of Reduced Disjunctive Normal Forms of Boolean Functions. An initial set of properties of these structures were introduced in propositions 1-7.

Bibliography

1. Zhuravlev Yu. I. On an algorithmic approach to the problems of recognition and classification. Problemi Kibernetiki, 33, (1978) 5--68.
2. Zhuravlev Yu. I. Selected Scientific Works, Publishing House Magister, Moscow, (1998) 417p.
3. Aslanyan L. H. On a pattern recognition method based on the separation by the disjunctive normal forms. Kibernetika, 5, (1975), 103--110.
4. Vapnik V. and Chervonenkis A. Theory of Pattern Recognition. “Nauka”, 1974.
5. Aslanyan L. H. The Discrete Isoperimetric Problem and Related Extremal Problems for Discrete Spaces. Problemy Kibernetiki, 36, (1979), 85--128.
6. Nechiporuk E. I. On topological principles of self-correcting. Problemy Kibernetiki, 21, (1969), 5--102.
7. Graham N., Harary F., Livingston M. and Stout Q. Subcube Fault-Tolerance in Hypercubes. Information and Computation 102 (1993), pp. 280{314.
8. Glagolev V. V. Some Estimations of D.N.F. for functions of Algebra of Logic. Problemy Kibernetiki, 19, (1967), 75--94.

Authors' Information

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

Juan Castellanos – Dpto. Intelligencia Artificial – Facultad de Informatica – Universidad Politecnica de Madrid 28660 Boadilla del Monte – Madrid, Spain; e-mail: J.Castellanos@fi.upm.es