Associative-Structural Analysis of 0-1 Data using Monotone Systems [*]

G. Hencsey

We construct a variational model of search for similar objects in data file. The model is applied to propose a new approach to the analysis of complex structured data.

1. Introduction

The "object × attribute" matrix is very common form of empirical data, for which many methods of analysis have been developed. These methods can be divided into two categories – integral and structural. With integral methods, processing produces a relatively small number of variables characterizing the data matrix as a whole. Structural methods partition the matrix into submatrices and perform integral analysis of each submatrix. Many variants of this structural analysis are described in Hartigan's book [1]. In Soviet literature, the monograph [2] is devoted to this topic.

An important feature of both groups of methods is that the objects and the attributes (the rows and the columns of the data matrix), although treated as two distinct sets, are not interpreted as active, interacting elements. The quantity x_{ij} corresponding to object i and attribute j in the data matrix $X = \|x_{ij}\|$ is treated as passive label with index j assigned to object i.

There is, of course, a different active interpretation of the data matrix, which treats the quantity x_{ij} as the measure (the weight) of interaction between element i of one set (the set of objects), and the element j of another set (the set of attributes). This interpretation clearly contributes new insights to data processing problems and methods: we can study the steams of interaction weights to the arcs of a complete bipartite "object × attribute" graph.

The aim of the present paper is to describe one of the applications of this active interpretation of the "object × attribute" matrix for the case of 0-1 data.

---

[*] Budapest. Hungary. Translated from Avtomatica i Telemekhanika, No. 2, pp. 137 – 141, February, 1987. Original article submitted April 11, 1986.

The proposed method will be called associative-structural, since it relies on preliminary identification, for each object, of some special subset of objects, which we call the set of associative images of the object. Then the objects are analyzed using the associative image sets, rather the original attributes. This approach implicitly borrows ideas from Zadeh's fuzzy set theory [3] and Arabie's additive classification [4].

An essential component of the proposed approach is the use of the method of monotone systems [5] [1], which produces a globally optimal solution.

2. Constructing the Set of Associative Images of an Object

Let $X = \|x_{ij}\|$ be an input matrix of N rows (objects) and m columns (attributes). The attributes are Boolean, with values 0 and 1 only. With each object i we associate a submatrix $Y^i = \|y^i_{rk}\|$, comprising the attribute columns of the original matrix X in which the object i has 1-s, i.e., $y^i_{rk} = x_{rk}$, if $x_{ik} = 1$ (if $x_{ik} = 0$, the corresponding column k is not included in Y^i).

On the set of rows of each matrix Y^i define a monotone system $\langle W, \pi^i \rangle$:

$$\pi^i(r,H) = \alpha \cdot |Y^i_H| - (1-\alpha) \cdot |y^i_r|, \tag{1}$$

where $|y^i_r|$ is the number 1's in row r of the matrix Y^i, and $Y^i_H = \bigcup y^i_r$ is an artificial row defined for each subset of rows $H \subseteq W$ ($1 \geq \alpha \geq 0$): W is the set of rows ($|W| = N$).

The introduction of the system (1) makes it possible [6] to effectively partition the set W into two nonintersecting subsets G^i and $\overline{G}^i = W \setminus G^i$, the first consisting mainly of 0's and the second mainly of 1's. This partition is a solution of some extremal problem

$$\left(G^i : \max_{H \subseteq W} \min_{r \in H} \pi^i(r,H) \right).$$

Clearly, for any α we have the inclusion $i \in \overline{G}^i$. This follows directly from the equality $y^i_{ik} \equiv 1$, which holds for all $k = 1,\ldots,m$ by definition of Y^i.

[1] Better called "the method of monotonic systems" as it turns out the term "Monotone System" was already occupied in "Reliability Theory", see Sheldon M. Ross, "Introduction to Probability Models", Fourth Ed., Academic Press, Inc., pp. 406-407, jm.

The set of associative images of the object i is the set $\overline{G}^i$, which contains a relatively large number of 1's in the same columns, where 1's occupy their positions in row i of the matrix X. Having solved N problems to find the associative image sets of all the rows of the matrix X, we arrange the solutions in the form of a square 0-1 matrix $Z = \|z_{ij}\|$ of dimension $N \times N$. In this matrix, row i corresponds to an object whose associative image set is the subset of columns of Z with 1's in row i. In other words, the element z_{ij} is an indicator (if $z_{ij} = 1$) that the object j is an associative image of the object i.

The columns of the matrix Z are used as new parameters. Note that if the ℓ-th column object is an associative image for some set H of row objects, then $z_{i\ell} = 1$ for all $i \in H$.

3. Classification and Its Structure Graph

The classification scheme proposed in this section consists of three stages.

The first stage applies to the matrix Z the method of linguistic analysis of 0-1 matrices described in [6]. It partitions the matrix columns into a prespecified number k of submatrices and at the same time partitions the matrix rows into two classes in each of the k column submatrices. The row partition in each column submatrix is effected by introducing the monotone system (1) and finding the corresponding extremum subsets G and $\overline{G}$.

The subset Z^q of column objects is called the defining part (or the nucleus) of the q-th class in the classification, and the set $\overline{G}(Z^q)$ of the row objects of Z corresponding to this nucleus is called its fuzzy part (or hull). On the whole, the class q is defined as the union $S_q = Z^q \cup \overline{G}(Z^q)$. The sought classification is a family of k subsets $S = \{S_1,...,S_k\}$. The index set $\{1,2,...,k\}$ is denoted by $I(S)$. This classification in general has intersecting classes. However, the nuclei of its classes are nonintersecting: they constitute a partition of the given set of objects.

The structure of the classification S is the graph $\Gamma(S,V)$, with vertices corresponding to the elements of S (the classes S_q) and the arcs defined by the rule $v_{qt} = (S_q, S_t)$, if $Z^q \subset \overline{G}(Z^t)$.

The implication induced by the arc v_{qt} corresponds to the following inference: "membership of an object in the class S_q implies its membership also in the class S_t." This interpretation is consistent with the notion of "dependence" (succession) of properties as subsets of some finite set, introduced in [7-9]. The representation of implication by the arc v_{qt} differs from these notions in two important respects. First, it is satisfied with some likelihood and not with logical necessity and, second, it is the outcome of a special analysis of the structure of the original 0-1 matrix (in distinction from [7-9], where it is defined on the original matrix). In [7-9], the introduction of implication generates the method of data analysis (i.e., it is the input of the data analysis algorithm), whereas in our framework it is a form that enables us to interpret the results of analysis (i.e., it is the output of the algorithm).

The structure $\Gamma(S,V)$ is called well-formed if for each $t=1,...,k$ the following condition is satisfied: $v_{qt} \in V$ implies that the vertex t has a loop $v_{tt} \in V$. In other words, the implication v_{qt} in a well-formed structure is true if and only if the nucleus Z^t is contained in its hull $\overline{G}(Z^t)$; the membership of an object in the nucleus of a class implies that it is located inside the hull of that class $\left(Z^t \subseteq \overline{G}(Z^t)\right)$.

This concept is useful in two respects. First, it is constructive, and therefore enables us to identify well-formed subgraphs in a given classification. Second, assuming well-formedness, it leads to a natural operation of aggregation of structures.

Definition. The graph $\Gamma(S',V')$ is an aggregated structure of the well-formed graph $\Gamma(S,V)$ if

1) S' is a partition of S into subsets (classes) each corresponding to a complete subgraph of the graph $\Gamma(S,V)$;
2) $v_{qt}' \in V'$ implies that S'_q contains a vertex $S_\alpha \in S'$ and S'_t contains a vertex $S_\beta \in S'$ such that $v_{\alpha\beta} \in V$ in $\Gamma(S,V)$.

If in the aggregated classification S' we define the nucleus and the hull of the macro-class S'_q as the union of the respective nuclei and hulls of all the original classes and hulls from S' entering S'_q then the aggregated graph $\Gamma(S',V')$ is again a well-formed graph. Since this construction may produce new complete subgraphs, which previously were not observed in the graph $\Gamma(S,V)$, we may apply the aggregation operation to the new graph $\Gamma(S',V')$ and generate a graph $\Gamma(S'',V'')$ of the next higher level of aggregation.

This process of successive aggregation has a natural stopping condition – a situation when no further aggregation is possible. This process may be viewed as a variant of the agglomerative procedure [1], which starts with some initial classification into k classes and ends with a classification into k^* classes ($k^* \le k$).

It is significant that on each level we have a classification representable in two forms: nuclear, when the classes are nonintersecting, and fuzzy, with intersections; the nuclear forms have their own agglomerative tree and the fuzzy forms have a different tree. The relation formalized by the structure describes the relationship between these trees [2] $\Gamma(\tilde{S},\tilde{V})$.

In conclusion, we should stress that the agglomerative process described above essentially depends on the partition of the current graph $\Gamma(\tilde{S},\tilde{V})$ into complete subgraphs (in general, this partition is not unique). The existence of different agglomerative procedures, combined with the free parameters α in (1) and $k = |I(S)|$, enables us to tune the general scheme of analysis to the solution of various applied problems.

4. Conclusion

In this paper we have focused on certain aspects and concepts of the data structuring problem, such as interaction of objects and parameters, the set of associative images of an object, the structure and aggregation of classifications. The proposed construction demonstrates the usefulness of these concepts. It is easily extended from 0-1 to general numerical data; the associative image set can be constructed not only for an object but also for a parameter, and then comparison of two image sets may lead to a new interaction scheme of objects and attributes. Other classification algorithms may be generated by the basic classification S. Of special interest is the intersection structure of the sets $\overline{G}(Z^1),...,\overline{G}(Z^k)$; the aggregation procedure may be defined in a different way. We did not aim at an exhaustive study of these alternatives; rather we used a single example in order to demonstrate how, given little prior information, we can apply the system approach to analyze the data bank and the structural approach to characterize the data.

[2] The tilde ~ is used to emphasize that in general this is some intermediate level of aggregation.

In this framework, the classification problem emerges to the forefront, and the performance measure of the classes is no longer a functional, as in the main body of classification research [3] [2], but rather simplicity of the structure induced by the classification and the comprehensibility of the classification structure on the logical level.

The method of monotone systems implemented in this framework provides a fairly general tool [5] for aggregated description of a priori poorly structured data, by considering the interaction between the elements of the relevant sets.

LITERATURE CITED

1. J. A. Hartigan, Clustering Algorithms, *Wiley*, New York (1975).
2. E. M. Braverman and I. B. Muchnik, Structural Methods of Empirical Data Processing [in Russian], *Nauka*, Moscow (1983).
3. L. A. Zadeh, "A fuzzy theoretical interpretation of linguistic hedges," *J. Cybernetics*, *2*, 4-39 (1972).
4. P. A. Arabie, S. A. Boorman, and P. R. Levitt, "Constructing block models: how and why," *J. Math. Psychol.*, *17*, No. 1, 21-63 (1978).
5. J. E. Mullat, "Extremal subsystems of monotone systems, I," *Avtomat. Telenekh.*, No. 5, 130-139 (1976), http://www.datalaundering.com/download/extrem01.pdf .
6. E. N. Kuznetsov, I. B. Muchnik, and L. V. Shvartser, "Local transformations of monotone systems, I," *Avtomat. Telemekh.*, No. 11, 76-84 (1985), http://www.datalaundering.com/download/ltransfo01.pdf .
7. A. A. Plotkin, "A measure of independence of classifications," *Avtomat. Telemekh.*, No. 4, 97-104 (1980).
8. A. A. Plotkin, "Hierarchical systems of subsets," *Avtomat. Telemekh.*, No. 5, 141-147, (1981).
9. A. A. Plotkin, "On a class of dependence relations and quantitative evaluation of the degree of dependence," Avtomat. Telemekh., No.4, 148-156 (1983).

[3] Functionals do have a role in the construction of the basic classification S.