Control
and Cybernetics
VOL. 18 (1989) No. 1

On global optimality in cluster-wise regression

by

JAN W. OWSIŃSKI

Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warszawa, POLAND

The paper refers to the problem of finding „local linear models", understood in a global manner, i.e. involving determination of the „best" number of local models, the models themselves and the sub-samples related to these models. The problem can be formulated as a clustering problem with, however, predefined solution structure, and, if a classical linear regression framework for local models is adopted, also predefined local objective function. The paper reviews the basic questions pertaining to the thus defined clustering problem, both typical to clustering problems in general (globality of solution, numerical efficiency) and specific for the type of problems considered (non-uniqueness, statistical validity). Then, a number of existing approaches, whether ad hoc or specific, are reviewed, and their relation to questions outlined before is discussed. A proposal for an objective function implying globally optimal solutions in the sense mentioned and a method referring to it is presented.

KEYWORDS: clustering, regression, cluster-wise identification, global optimality.

1. Intruduction

Assume a general clustering problem, i.e.: for a set I of indices i, $i \in I$, $I = \{1,...,n\}$, denoting n objects characterized by vectors x_i, and/or by values d_{ij} or s_{ij}, which are interpreted as, respectively, „distances" and „proximities", to find such a partition p^* of I that objects belonging to the same clusters — subsets of I be possibly „close" or „alike", while those belonging to different clusters be possibly „far" of „dissimilar". A partition P of I is a set $P = \{A_q\}_{q=1}^{p}$ of clusters A_q, such that

$$\bigcup_{q=1}^{P} A_q = I \tag{1}$$

and, according to the standard most widely accepted, and used throughout this paper:

$$A_q \cap A_{q'} = \emptyset, \qquad \forall q \neq q' \in K \tag{2}$$

where $q \in K = \{1,...,p\}$, i.e. usually no overlapping and no fuzziness.

In the construction of cluster analysis methods it is of primary importance to define the intra- and inter-cluster „distance" and „proximity" measures, and, eventually, their overall aggregate. There are a number of methods which, in defining these measures, refer to distances or proximities with regard to certain points, such as cetroids or medians, whose interpretation might often be to „represent" the clusters.

Hence, if aggregate distance between clusters is denoted $D(A_q, A_q')$, and that within a cluster, for simplicity, is denoted $D(A_q)$, then for the methods which refer to a „representative" there is

$$D(A_q) = \tilde{D}(A_q, \tilde{x}(A_q)) \tag{3}$$

where $\tilde{x}_q(A_q)$ is equivalent to description of an (actual or dummy) representative object or a set of objects, $\tilde{x}(A_q) \subset A_q$. Quite often, in fact,

$$\tilde{D}(A_q,\tilde{x}(A_q)) = D(A_q,\tilde{x}(A_q)) \tag{4}$$

In general, it is possible that solution structures involving „representatives" take the form $(P,X) = \{(A_q,\tilde{x})\}_{q=1}^{p}$. In most practical cases, however, $\tilde{x}_q$ can be uniquely determined on the basis of A_q, so that it is sufficient to refer to P.

It should be strongly emphasized that application of a (3), together with definition of $\tilde{x}(A_q)$ is, in an explicit of implicit manner, strongly related to an a priori assertion as to the „nature" of the set of objects. Most often this assertion points towards the spherical shape of „natural" and — therefore, par force, also — algorithmically delineated clusters. This may reflect a wish of getting away from other cluster-shape biasses displayed by, e.g., hierarchic agglomerative methods.

Note that from the point of view of the solution structure implied by minimization of an aggregate of (3), the „resulting distance" of two objects i and j, is, provided appropriate definition of $d(.,.)$ is available,

$$\tilde{d}_{ij} = \min_q L(d(x_i,\tilde{x}(A_q)), d(x_j,\tilde{x}(A_q))) \tag{5}$$

where, as usual, $d(a,b) = d(b,a)$, and L is nondecreasing with respect to both arguments.

It is easy to see that such clustering problem solution constructions provide a far less opportunity for an intuitive assessment and analysis than, for instance,

the classical progressive merger procedures, which, quite often, do only refer to the matrix F_d of $\{d_{ij}\}$. This phenomenon is, however, inevitable as the solution structures become more complex.

The solution structure implied by (1) — (3) is seen as the one in which clusters are determined together with their „representatives", which can also be referred to as cluster „models", i.e. models of cluster-proper objects. Thus, for instance, in biological taxonomy, clusters A_q may represent families, while cluster-wise models $\tilde{x}(A_q)$ — those species, which are the „most characteristic" for the families, the model-species being either real or dummy ones. Introduction of the notion of cluster-wise model does not necessarily cut out of the cluster analysis domain a well defined area of problem formulations, i.e. assumed solution structures. It makes a room, however, for a class of problems, which refer more directly to the notion of model.

2. Problem outline

2.1. The alternative

In particular, a problem may arise of determining, for a given data set, a solution structure sometimes called „local linear models", see e.g. Diday (1986). Within this domain a particular class can be distinguished. Assume, for instance, that the data matrix X containing elements x_{ik}, such that $[x_{i1},...,x_{im}] = x_i$, is generated by a process in which, roughly, vectors $x_{\cdot 1},...,x_{\cdot m-1}$, $x_{\cdot k} = [x_{1k}, x_{2k},...,x_{ik},...,x_{mk}]^T$, describe various „input" vector values of the proces, while $x_{\cdot m}$ decribe corresponding „output" values. (The multioutput case is just and extension of this one, provided there are no inter-output dependences.) Thus, vectors x_i, $i \in I$, represent various „point" descriptions of the functioning of a process. Matrix X has therefore the structure as below:

$$X = \left| \begin{array}{ccccc} x_{11} & \cdots & x_{1k} & \cdots & x_{1m} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \cdots & x_{ik} & \cdots & x_{im} \\ \vdots & & \vdots & & \vdots \\ x_{n1} & \cdots & x_{nk} & \cdots & x_{nm} \end{array} \right|$$

vectors x_i of „objects", i.e. descriptions „input-output"

output vector $x_{\cdot m}$

input value vectors $x_{\cdot k}$, $k = 1,...,m-1$

The alternative that one is confronted with is as follows:

a. To treat all descriptions as generated by virtually the same (unchanged) process, and differing therefore only by some random error, so that all point descriptions can be aggregated into a one, approximate model, constituting

full description of the process functioning, allowing estimation of other x_{*m} for given $x_{*_1},\ldots,x_{*m-1}$.

b. To reject the hypothesis of the same generative process, stating, instead, that the individual descriptions x_i differ so much that it becomes impossible to determine a valid model for a hypothetical process, and hence the descriptions can rather be treated as separate objects; no projections are feasible.

c. To assume that the descriptions do not, in fact, represent the functioning of the same unchanged process, and that therefore no valid overall aggregate model for the whole I can be determined, but that it is possible to define subsets of I, corresponding to distinct processes or process modes, for which valid models can be obtained.

In reality, variant c of the above alternative may also take the form:

d. To compare the quality and validity of the process model obtained for the whole I and for its, optimally chosen, subsets.

Structure of this alternative is presented in Fig. 1:

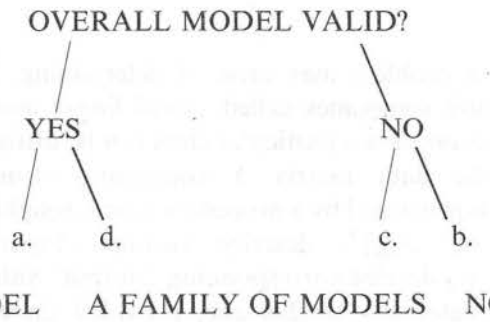

Fig. 1. Structure of the model adoption alternative

The present paper looks at the variants d and c from the point of view of numerical feasiblility and capacity of obtaining theoretically justified results.

2.2. Problem formulation

Assume that the process model relating x_{*_m} to $x_{*_1},\ldots,x_{*_{m-1}}$ has in general the linear form, i.e.

$$x_{*m} = f(x_{*1},\ldots,x_{*m-1}) = ax_*^T \tag{6}$$

where $a = [a_1,\ldots,a_{m-1},a_m]$, and $x_* = [x_{*1},\ldots,x_{*m-1},1]$. Determination of a from x_i subject to errors, implies, in case of one overall model, that

$$x_{im} = f(x_{i_1},\ldots,x_{im-1}) + e_i = a\hat{x}_i^T + e_i \tag{7}$$

where $\hat{x}_i = [x_{i1},\ldots,x_{im-1},1]$. In case of „local linear modelling", there is

$$x_{im}^q = f^q(x_{i1},\ldots,x_{im-1}) + e_i^q = a^q\hat{x}_i^T + e_i^q \tag{8}$$

i.e. depending on the model, values

$$x_{im}^q - e_i^q = \hat{x}_{im}^q = a^q\hat{x}_i^T \tag{9}$$

will be obtained, for $q \in K$.

Assume further that models a^q are obtained via the classical regression technique for clusters $A_q \subset I$, i.e. within the clusters the sum of error squares is minimized:

$$\min_{a^q \in R^m} \sum_{i \in A_q} (x_{im} - a^q\hat{x}_i^T)^2 = \min_{a^q \in R^m} D(A_q, a_q) \tag{10}$$

and the standard regression equation can be applied, provided all usual conditions hold, that is

$$a^q = (X_q^T X_q)^{-1} X_q^T X_{qm} \tag{11}$$

where X_q is a submatrix of $[\hat{x}_i]_{i=1}^n$ composed of vectors $\hat{x}_i$ such that $i \in A_q$, and X_{qm} is a vector of values x_{im}, for which $i \in A_q$.

Now, the initially stated general clustering problem can be reformulated for the above outlined case. Thus, a partition p^* of I is sought which would ensure minimization of intra-cluster distances according to (10), preserving, however, the fundamental feature previously mentioned, namely that of distinguishing essentially differing clusters and their models.

Thus, the announced cluster-wise identification problem was reduced to the cluster-wise linear multiple regression (CLR) problem, which, though, retains all the essential features of the more general one.

2.3. The issues

Since the problem was yet formulated in such a way that its properties should still be made more precise, a number of issues must at this point be discussed.

First, there is the usual globality issue. Denote

$$\min_{a^q \in R^m} D(A_q, a^q) = D_{\min}(A_q) \tag{12}$$

Then,

$$\min_P \sum_{q=1}^{p} D_{\min}(A_q) \overset{\text{def}}{=} \min_P Q_D(P) = \underset{\sim}{Q}_D(P^*_{\underset{\sim}{Q}_D}) \tag{13}$$

where, naturally, $P^*_{\underline{Q}D} = I$ i.e. minimization of the sum of intra-cluster aggregate distances (10), leads to trivial solution $P^*_{\underline{Q}D} = I$, that is $p = n$, $A_q = q$, and indeterminate a^q, e.g. $a^q = \left[\frac{x_{im}}{x_{i1}}, 0,...,0\right]$, for which $\min_P \underline{Q}_D(P) = \underline{Q}_D(I) = 0$.

The most frequent way around the globality issue is to analyse (13) for a definite class of P, e.g. for the P's with a given p, since $\min \underline{Q}_D(P)$ is monotone with regard to p, and most functions of similar nature are also. In fact, an important group of methods reside upon precise definitions of the class of P considered. It can be seen, though, that even for thus limited subsets of the space of partitions only local solutions can usually be reached, not ensuring globality with regard to a class of P considered.

As is pointed out further on in this paper, it is feasible to construct a method which operates on an objective function based upon (10), and which would not yield a monotone dependence of $\underline{Q}^*$ upon f. Such a method ought namely to take into account, in accordance with the general problem formulation presented, both intra-cluster, i.e. (10), and inter-cluster distances.

In a number of cases, however, the choice of $\{A_q\}$ is made with complete disregard of (10), which is used only to obtain a^q minimizing it for otherwise given A_q. Thus, for instance, one can first use a progressive merger procedure to obtain P and then determine $\{a^q\}_{q=1}^p$ via (11). A consequent application of such an approach would require a choice or design of such merger criteria which would be justified from the point of view of (10). Otherwise, and this is presently mostly the case, there is little reason to accept thus obtained solutions.

Hence, for both these types of methods proper formulation of objective functions seems to be crucial.

There is, however, another group of approaches, which are, in fact historically the oldest ones within the CLR domain, and which stem directly from the classical regression analysis. These approaches are meant to solve quite specific, and, seemingly, simpler problem of finding such division points in a sequence of values of an independent variable that a regression model, having $m = 2$, preserves its validity between two adjacent division points, but changes at each of these division points. (This fact is referred to as „regime change" of the process.) In many cases existence of just one division point ($p = 2$) is assumed. The source and main interpretation of this problem is obvious: change of a model (say, a growth model) with time, although other interpretations are quite as feasible, e.g. change of a consumption model along the per capita income axis. Note that the independent, model shift defining, variable may enter the model or just stand as an index.

In spite of apparent simplicity of the problem, approaches aimed at solving it for larger p and m get very soon quite difficult to devise, because their purpose is not just numerical operation on the data table X, in abstracto of its inter-

pretation, but rather an in-depth statistical analysis, based upon the rules of statistical inference. Thus, construction of appropriate precise tests gets prohibitively difficult as p and m increase above 2. That is why for more general cases, e.g. $p \gg 2$, heuristic approaches are proposed, combining statistical analysis with clustering. Hence, the methodological loop closes, indicating clustering as the proper algorithmical framework for multi-regime regression analysis.

At this point a more general remark is due. In such complex data settings as those implied by the CLR problem the classical statistical analysis is very difficult to carry out. Statistical analysis is virtually hindered by lack of a priori knowledge of assignment of objects to hypothetical subpopulations. It should be emphasized that not only the precise form, but also the very kind of hypotheses which can be forwarded depend upon the contents and dimensions of subpopulations. These hypotheses, therefore, need not be of the same kind for various clusters. Thus, any reasonable analytical operation to be performed on such data sets has to be somehow related to cluster analysis.

3. Classification of approaches

3.1. A summary

Along the lines of reasoning already deployed Table 1 presents a range of problem formulations pertaining to the question taken up in this paper together with samples of appropriate references. It can easily be seen that the set of approaches and techniques available require quite an analytic effort priori to application of any given method, so that the proper choice of method or methods is made.

3.2. The „algorithm"

A quasi-algorithm of such a choice of methods, based upon the ranges of applicability and validity of the existing ones, is outlined below. The numbers appearing here correspond to those of Table 1.

A. Is approximation-type formulation more suitable? If Yes, go to 9. If No, go to B.

B. Is the case unidimensional? If Yes, go to C. Only two submodels? If No, go to D. If Yes, go to 2 or 3. If No, go to 3 or 4.

D. Is there a basis for a good guess as to the „right" partition and models? If Yes, go to 6. If No, go to 5 and 1: Is the result satisfactory? If Yes, Quit. If No, go to E.

E. Maximum number of trials not attained? If Yes, go to D. If No, go to 7 and/or 8.

Strictly statistical methods fail for multidimensional cases with unknown

Table 1

Problem formulations and exemplary references

No	General problem formulation / approach	Remarks / Limitations	Illustrative references
1	Testing of significance of differences between given regresnion dependences	Analytical, and not synthetic approach, low number of alternatives considered	Chow (1960), Quandt (1960), Hald (1962), Williams (1967), Caliński a. Malec (1976), Malec (1980)
2	Testing whether a parameter shifts	Usually one independent variable and two regimes	Page (1955, 1957), Kander a. Zacks (1966), Brown a. Durbin (1968), Hinkley (1969), Fender (1973)
3	Finding of a breaking point over a time-like variable, perhaps together with the parameters of resulting regressions	As above	Quandt (1958, 1972), Robinson (1964), Mustafi (1968), Farley a. Hinich (1970), Ferreira (1975), Kiefer (1978, 1980), Schmidt (1982)
4	Identification of multiple regimes along a time-like variable	One independent variable, sequental regimes	McGee a. Carleton (1970)
5	Cluster-wise regression: first cluster, then regress	Little, if any regression-wise justification for cluster; numerical efficiency	Fakiner, Krieger a. Rohmeier (1977) Owsiński (1984 a,b)
6	Cluster-wise regression: exchange and „center-and-reallocate" approaches	Local optima depending upon initial partitions	Späth (1979, 1981, 1982, 1983), Charles (1977), Diday et al. (1980)
7	Sample breakdown for better regression	More robust, and not varying models	Hinich a. Talwar (1975)
8	Cross-validation	As above	Picard a. Cook (1984)
9	Piecewise curve fitting	Similar as in 4. above	Charles a. Lechevallier (1979), Dorofeyuk, Ibragimli a. Movsuniov (1976), Bacon a. Watts (1971)

numbers of subpopulations/submodels. In these circumstances the methods generally referred to as „center and reallocate" turn out to be the strongest, especially those of the „nuées dynamiques" kind, but also, for more precisely difined situations — more usual exchange algorithms (No 6 in Table 1, see, for instance, a clear-cut case study by Späth (1986)). Obviously, all these yield local solutions and not the global ones. The term „local" is used here to point out the fact that not only the final result of a procedure working depends upon the initial partition-and-model structure (P^0 $\{a^q\}^0$), i.e. locality of the procedure itself, but also, that the objective functions used have the feature that their optimum values for consecutive p are monotone, i.e. locality of the objective function, that is, problem formulation, itself. (Note, though, that in some methods, like e.g. those based upon the „nuées dynamiques" idea, it is possible to find a partition-and-model solution better than the initial one and having p different from the initial one, in particular, $p^* \leqslant p^0$).

Once a procedure stopped it may be advisable to apply a sort of statistical test for significance of differences among the obtained models (No.1 of Table 1). It should be remembered, however that simultaneous comparison is necessary and that a variety of basic statistical assumptions would have to be used and/or verified (population-proper distribution,..). In fact, prior to application of a test to the solution obtained it would be proper to check whether for the data set at hand and with a given intra-cluster measure, say like in (10), there exists at all a possibility of obtaining significant model differences. This, though, may turn out to be quite difficult if no extra information on the data set is available.

The latter remark applies equally to the choice of $(P^0, \{a^q\}^0)$. Indeed, there may by some more or less ad hoc, more or less effective, techniques for determination of the starting point, usually referring to overall regression and/or to regression-related data set analyses (see Nos. 7 and 8, Table 1), but none of these is sufficiently effective to allow disregarding of possible guesses based upon experience and intuition of people knowing the problem at hand.

Thus, the globality issue is still to be resolved on both the problem formulation („objective function") and solution procedure levels.

It should be noted that the globality issue is compounded with the numerical efficiency one. Putting aside the „regime change" formulations, which face problems other than numerical efficiency, it seems obvious that approaches of the type „cluster first, regress after" are much quicker and simpler than the „center and reallocate" approaches, even if the first ones are complemented with validation procedures (which are also applicable to the latter approaches, anyway). Hence, there arises a need for developing such approaches that will combine globality of formulation and, possibly, procedure, with numerical efficiency of the usual progressive merger procedures, which is attained by the „cluster first, regress after" approaches.

4. Suggested approach

The paper presents just an outline for the approach that would solve the globality issue, first at the problem formulation level.

4.1. The fundamentals

The approach proposed is based upon the general philosophy sketched out in Owsiński (1984), this time applied to multi-model identification. According to previous remarks, the general philosophy mentioned resides upon the postulate of simultaneous consideration of intra-cluster and inter-cluster measures (distances,...). Thus, along the lines set out in Owsiński (1984) the proper objective function, to be minimized over the space of P, would have the form

$$Q_D^S(P) = \underline{Q}_D(P) + \bar{Q}_S(P) \tag{14}$$

where $\underline{Q}_D(P)$ is, in analogy to (13), an aggregate of intracluster distances and $\bar{Q}_S(P)$ is the corresponding aggregate of inter-cluster similarities (proximities). Similarity values are by definition taken non-negative and decreasing with distance values for the same objects. Formula (14) omits the reference to models $\{a^q\}$ assuming that they can be uniquely defined for a given P. $Q_D^S(P)$ is being minimized, while its „dual":

$$Q_S^D(P) = \underline{Q}_S(P) + \bar{Q}_D(P) \tag{15}$$

is being maximized, both under the constraint of P being a partition. It is relatively easy to establish conditions under which

$$\arg\max_P Q_D^S(P) = \arg\min_P Q_S^D(P) = P^{opt} \tag{16}$$

In particular, for established forms of Q, in which only D and S (i.e. distances and proximities) are exchanged in order to obtain $\underline{Q}_D$, $\underline{Q}_S$ and $\bar{Q}_D$, $\bar{Q}_S$, it is sufficient to apply a linear transformation of distance to proximity and vice versa in order for (16) to be true. Now, the situation depicted in (13) would not occur, since both (14) and (15) can be easily made to imply globally optimal P^{opt}.

Additional, algorithmically convenient and interpretatively facilitating device introduced in this approach is the use of the weighing coefficient r, $r \in [0, 1]$, so that

$$Q_D^S(P, r) = r\underline{Q}_D(P) + (1 - r)\bar{Q}_S(P) \tag{17}$$

and the algorithm can proceed by optimizing $Q_S^D(P, r)$ for r starting from 1 and going to 0. Thereby $Q_S^{D\,opt}(r)$ and corresponding $P^{opt}(r)$ are obtained. For this algorithm to be really simple, though, certain limitations must be set upon the classes of changes that P's can undergo with decreasing r. It is namely obvious that the interval [0, 1] can be divided into subintervals $[r^t, r^{t-1}] = \Delta r^{t-1}$, where endpoints overlap, and $\bigcup_t^{T+1}$, $\Delta r^t = [0,1]$, $P^{opt}(r)$ being constant within Δr^t and changing only at their endpoints. (There is, obviously, $P^{opt}(r^0) = I$, $P^{opt}(r^T) = \{I\}$ and $r^{T+1} = 0$.) Thus, changes at finite number of r^t would be of a predefined nature.

4.2. Global objective function for CLR

The same analysis as in 4.1, given after Owsiński (1984), applies to the objective function defined as in (13), for $\underline{Q}_D(P)$, with

$$\bar{Q}_S(P) = \sum_{q} \sum_{q' > q} (\bar{R} - D(A_{q'}, a^{q'})) \tag{18}$$

where $\bar{R}$ is a sufficiently large positive constant, chosen on the basis of the potential range of values of $D(A_q, a^q)$ and considerations related to the balance of Q_D and Q_S in the overall Q's. Transformation from distance to proximity can also be moved from the cluster proper to the object level.

Thus, according to (14), the minimized objective function is:

$$Q_D^S(P) = \sum_{q} D(A_q, a^q) + \sum_{q} \sum_{q' > q} (\bar{R} - D(A_{q'}, a^{q'})) \tag{19}$$

and, according to (17), its parametrized variant:

$$Q_D^S(P, r) = r \sum_{q} D(A_q, a^q) + (1 - r) \sum_{q} \sum_{q' > q} (\bar{R} - D(A_{q'}, a^{q'})) \tag{20}$$

Similarly, for the „dual", maximized, objective function, one has

$$Q_S^D(P) = \sum_{q} (\underline{R} - D(A_q, a^q)) + \sum_{q} \sum_{q' > q} D(A_{q'}, a^{q'})) \tag{21}$$

and therefore

$$Q_S^D(P, r) = r \sum_{q} (\underline{R} - D(A_q, a^q)) + (1 - r) \sum_{q} \sum_{q' > q} D(A_{q'}, a^{q'})) \tag{22}$$

with, again, $\underline{R}$ sufficiently large. Obviously, a^q can be determined according to $D_{min}(A_q)$, which would then substitute $D(A_q, a^q)$ in (19) – (22).

It is easy to see that formulations (19) – (22) do exactly follow the recipe of simultaneous consideration of inter and intra- measures. With that respect the choice of $\bar{R}$ and $\underline{R}$ is crucial. As mentioned, transformation of distance into proximity can be moved over to the object, rather than cluster, level. Besides that, the very form of this transformation, shown here, is just an exemplification.

Obviously, thus defined Q_D^S and Q_S^D can imply globally optimal P's under proper choice of $\bar{R}$ and $\underline{R}$ (or, more generally — of distance-proximity transformation). This particular exemplification gives $P^{opt}(r)$ ranging from the equivalent trivial partitions, obtained as optimal for $r = 1$, for which all the $D(A_q, a^q) = 0$ and minimal p is determined by the relation between n and m, to just two clusters, card $P^{opt}(0) = 2$, when Q_D is maximized or Q_S is minimized. In between, there is a room for finding optimum P for such r value or segment, which, due to the a priori considerations would be considered proper for the given distance-proximity transformation and/or $\bar{R}$ and $\underline{R}$ (or vice versa: the transformation can be sought such that the optimum required would occur at $r = 0.5$).

4.3. Principles of the procedure

As mentioned, the general outline of the principles for the procedure is as follows.

First, the dual objective functions considered cannot be optimized in a straightforward manner. Certain suboptimizing heuristics have therefore to be applied. When applying them, care should be taken when n/m is low, i.e. of the order of m. Namely, the alternative which arises is: while low n/m allow more exhaustive search in the space of partitions, it incurs greater overall error in case of erroneous assignment of particular objects i to cluster A_q. Moreover, larger n/m facilitate one of the „branches" of the procedure. This procedure starts, theoretically with $D(A_q, a^q) = 0$, for $r^0 = 1$. Such a starting point would be, though, awkward for further steps. That is why a number of starting partitions $P^{0\,l}$ is generated, with card $P^{0\,l} \in (n/2m, n/m)$. To these partitions a reallocation algorithm is applied as given in Späth (1983) or Diday (1986), but implemented for functions (20) or (22). Resulting partitions, $\hat{P}^{0\,l}$, are subject to an operation, say, merger, at $r^1 < r^0$, as determined through simple algebraic manipulations from (20) or (22), see e.g. Owsiński (1984 b), on the basis of cluster-level measures D and their proximity-like transformations. Thus, partitions $P^{1\,l}$ are established, anew subject to reallocation, which yields $\hat{P}^{1\,l}$ etc.

First results obtained with the new objective function and with the variants of the procedure outlined, differing by

* the choice od reallocation algorithm,
* class of the r^t-driven operations, and
* the forms of distance-proximity transformation,

are promising enough to go further in this direction.

5. Conclusions

The main issue in multi-model identification turns out to be the globality of solutions. This issue appears on the level of problem formulation and on the level of solution finding. The paper asesses this situation and proposes a clear way out for the level of problem formulation. Some work is still to be done on the solution finding procedures, which refer partly to existing methods based upon matrix calculus.

6. References

[1] BACON D.W., WATTS D.G. Estimating the Transition Between Two Intersecting Straight Lines. *Biometrika* **58** (1971), 3, 524-534.

[2] BROWN and DURBIN (1968) Methods for investigating whether a regression relationship is constant over time. European Congress of Statistics, Amsterdam, typescript.

[3] CALIŃSKI T., MALEC E. (1976) Comparing multiple regression equations (in Polish). Annals of Agricultural Academy in Poznań, issue 5, p. 165.

[4] CHARLES C. (1977) Regression typologique. Rapport de recherche IRIA, No.257, October 1977.

[5] CHARLES C., LECHEVALLIER Y. (1979) Pattern Recognition by a Piecewise Polynomial Approximations with Variable Joints. Rapport de Recherche IRIA laboria, No.338, January 1979.

[6] CHOW G. (1960) Test of equality between two sets of coefficients in two linear regressions. *Econometrica* **28**, pp. 561-605.

[7] DIDAY E. et al. (1980) Optimisation en classification automatique. Le Chesnay. INRIA.

[8] DIDAY E. (1986) Canonical analysis from the automatic classification point of view. *Control a. Cynernetics,* **15,** 2, pp. 115-138.

[9] DOROFEYUK A.A., IBRAGIMLI Sh.D., MOVSIUNOV V.G. (1976) The use of the statistical equivalence criterion in the piecewise approximation problem (in Rusian). *Avtomatika i Telemekhanika*, No.7.

[10] FAKINER H., KRIEGER, E., ROHMEIER H. (1977) Regional differenzierte Analyse und Prognose des Wasserbedarfs der privaten Haushalte in der Bundesrepublik Deutschland. In: H. Späth, ed. Fallstudien Cluster Analyse. München, Oldenbourg.

[11] FARLEY J.V., HINICH M.J. A test for shifting slope coefficient in a linear model. *JASA* **65**, 331, pp. 1320-1329.

[12] FEDER P. (1973) On Asymptotic Distribution Theory in Segmented Regression Problems — Identified Case. Tech. Report, Statistical Dept. of Yale University.

[13] FERREIRA R.E. (1975) A Bayesian analysis of a switching model: known number of regimes. *JASA* **70**, 350, pp. 370-374.

[14] GOLDFELD S.M., QUANDT R.E. (1972) Nonlinear Methods in Econometrics. Amsterdam, North Holland.

[15] GOLDFELD S.M., QUANDT R.E. (1973) A Markov model of switching regressions. *Journal of Econometrics* **1**, pp.3.16.

[16] HALD (1962) Statistical Theory with Engineering Applications. New York, Wiley.

[17] HINICH M.J., TALWAR P.P. (1975) A simple method of robust regression. *JASA* **70**, 349, pp. 113-119.

[18] HINKLEY D.V. (1969) Inference about the intersection in two-phase regression. *Biometrika* **56**, No.3, pp. 495-504.

[19] KANDER Z., ZACKS (1966): Test procedures for possible changes in parameters of statistical distribution occurring at unknown time points. *Annals of Math. Statistics* **37**, October 1966, pp. 1196-1210.
[20] KIEFER N.M. (1978) Discrete parameter variation: efficient estimation of a switching regression model. *Econometrica* **46**, No.2, pp. 427-434.
[21] KIEFER N.M. (1980) A note on switching regressions and logistic discrimination. *Econometrica* **48**, No.4, pp. 1065-1069.
[22] MALEC E. (1980) Analysis of spatial differentiation of regression dependences (in Polish), In: Taxonomic Methods in Geography. Polish Academy of Sciences, Poznań Section, Geography Series, vol.5. Warszawa-Poznań, Polish Scientific Publishers.
[23] MANTEL M. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, pp. 209-220.
[24] MC GEE V.E., CARLETON W.T. (1970) Piecewise regression. *JASA* **65**, 331, pp. 1109-1124.
[25] MUSTAFI C.K. (1968) Inference problems about parameters which are subjected to changes over time. *Annals of Math. Statistics* **30**, June 1968, pp. 840-854.
[26] OWSIŃSKI J.W. (1984a) On aggregation in large scale models through a global clustering technique. In: A. Straszak, ed., Large Scale Systems: Theory and Applications. London, Pergamon Press for IFAC.
[27] OWSIŃSKI J.W. (1984b) On a quasi-objective global clustering method. In: E. Diday et al., eds., Data Analysis and Informatics III. Amsterdam, North-Holland.
[28] PAGE E.S. (1955) A test for a change in a parameter occurring at an unknown point. *Biometrika* **42**, pp. 523-527.
[29] PAGE E.S. (1957) On problems in which a change in parameters occurs at an unknown point. *Biometrika* **44**, pp. 248-252.
[30] PICARD R.D., COOK R.D. (1984) Cross-validation of regression models. *JASA* **79**, 387, pp. 575-583.
[31] QUANDT R.E. (1958) The estimation of the parameters of a linear regression system obeying two regimes. *JASA* **54**, 284, pp. 873-880.
[32] QUANDT R.E. (1960) Test of the hypothesis that a linear regression system obeys two separate regimes. *JASA* **55**, pp. 324-330.
[33] QUANDT R.E. (1972) A new approach to estimating switching regressions. *JASA* **67**, 338, pp. 306-310.
[34] ROBISON D.E. (1964) Estimates for the points of intersection of two polynomial regressions. *JASA* **59**, 306, pp.214-224.
[35] SAGUEZ and SANTINI (1979) Formalisation du problème de la régression par morceaux en termes de contrôle des systèmes. Rapport de Recherche IRIA.
[36] SCHMIDT P. (1982) An improved version of the Quandt-Ramsey MGF estimator for mixtures of normal distributions and switching regressions. *Econometrica* **50**, 2, pp. 501-516.
[37] SPÄTH H. (1979) Clusterwise linear regression. *Computing* **22**, pp. 367-373.
[38] SPÄTH H. (1981) Correction to: Clusterwise linear regression. *Computing* **26**, p. 275.
[39] SPÄTH H. (1982) A fast algorithm for clusterwise linear regression. *Computing* **29**, pp. 175-181.
[40] SPÄTH H. (1983) Cluster-Formation und -Analyse. München Oldenburg.
[41] SPÄTH H. (1986) Clusterwise linear least squares versus least absolute deviations regression: A numerical comparison for a case study. In: Classification as a tool of research. W. Gaul and M. Schader (eds.). Elsevier, Amsterdam.
[42] WILLIAMS E.J. (1967) Regression Analysis. New York, Wiley.

Received, October 1988.

O optymalności globalnej w regresji dla skupień

Artykuł dotyczy zagadnienia „liniowych modeli lokalnych" w perspektywie globalnej, tj. zagadnienia obejmującego: wyznaczenie „najlepszej" liczby modeli lokalnych, tych modeli i podzbiorów próbek odnoszących się do tych modeli. Tak sformułowane zagadnienie może być przedstawione jako zadanie analizy skupień, z zadaną z góry strukturą rozwiązania, a także, w pewnej mierze, postacią funkcji celu, odpowiadającą zadaniu regresji liniowej. Artykuł zawiera przegląd podstawowych kwestii odnoszących się do tak postawionego zadania, właściwych zarówno dla wszystkich zadań analizy skupień (globalność rozwiązań, sprawność numeryczna), jak i dla rozważanego zadania (niejednoznaczność, własności statystyczne). Podano także przykłady metod rozwiązania tego zadania, wraz z ich charakterystyką w odniesieniu do podanych kwestii. Na końcu zaproponowano postać funkcji celu implikującą rozwiązania globalnie optymalne i zarys odpowiedniej metody.

О глобальной оптимальности в регрессии для кластеров

Статья касается вопроса „линейных локальных моделей" в глобальной перспективе, т.е. вопроса охватывающего: определение „наилучшего" числа локальных моделей, этих моделей и подмножеств выборки, относящейся к этим моделям. Так сформулированный вопрос может рассматриваться как задача кластерного анализа, с заранее заданной структурой решения, а также в некоторой мере, видом функции цели, соответствующим задаче линейной регрессии. Статья содержит обзор проблем, относящихся к так поставленной задаче, свойственных как для всех задач кластерного анализа (глобальность решений, численная эффективность), так и для рассматриваемой задачи (неоднозначность, статистические свойства). Даются также примеры методов решения этой задачи, вместе с их характеристикой в отношении рассмотренных проблем. В заключение предлагается вид функции цели, имплицирующий глобально оптимальные решения и зарисовка соответствующего метода.