Acta Psychologica 68 (1988) 171-182
North-Holland

SYSTEMATIC RESEARCH INTO HUMAN BEHAVIOR IN MULTIATTRIBUTE OBJECT CLASSIFICATION PROBLEMS *

O.I. LARICHEV, H.M. MOSHKOVICH and S.B. REBRIK
Institute for Systems Studies, Moscow, USSR

This paper presents results of two series of experiments analyzing the behavior of four groups of subjects: senior students, senior pupils, members of the editorial board and physicians taking part in constructing expert systems. The complexity of a classification problem in each experiment was determined by three parameters: the number of attributes characterizing the evaluated objects; the number of estimates, on ordinal attribute scales; the number of decision classes to which the considered objects should be assigned. The results show that there are limits to human capacities in a wide range of intellectual problems. These constraints are objective and determined by the specifics of human information processing system.

Introduction

Numerous studies (Simon 1978; Tversky 1969, 1972) indicate that a decision maker's behavior depends not only on intellect, zeal, available information, and luck, but it also is affected by the specifics of human thinking that are common for all people. Although there are decision problems that decision makers manage, there also are problems that they cannot manage. Obviously, there must be some line that separates these two sets. It would be interesting to know where one can draw this line and upon what it depends.

Our research focused on the common problem of classification in multidimensional situations. For example, an R&D programme leader deciding upon which projects to incorporate into the program on the bases of their characteristics; a physician diagnosing a disease on the basis of patient's symptoms; an engineer determining the fault in a complex technical system on the basis of its errors. In these examples the decision maker has to assign an object which possesses a set of

* Requests for reprints should be sent to O.I. Larichev, Institute for Systems Studies (VNIISI), Pr. 60 let Octyabrja-9, 117312 Moscow, USSR.

characteristics (defined by several attributes) to one of several decision classes. In other words, the decision maker copes with a problem of multidimensional expert classification. This kind of problem has arisen in recent years in the construction of knowledge bases in expert systems.

In spite of the fact that this kind of decision problem is widespread and of practical importance it has not been subjected to systematic research.

General experimental pattern

This paper presents results of two series of experiments analyzing the behavior of four groups of subjects: senior students, senior pupils, members of the editorial board of a large research institute and physicians taking part in constructing expert systems.

The complexity of a classification problem in each experiment was determined by three parameters: the number of attributes (criteria, characteristics), N, characterizing the evaluated objects; the number of estimates, Wi $(i = 1, \ldots, n)$, on ordinal attribute scales (the estimates are ranked from best to worst); the number of decision classes, P, to which the considered objects should be assigned. The hypothesis was that behavior varies as a function of changes in these problem variables.

All feasible estimate combinations with respect to different attributes determine the entire set of possible descriptions of objects. In the course of each experiment, except for the last, the subjects were requested to examine all objects and assign each of them to one of the prescribed decision classes. Consider, for example, a problem handled by the students in experiment 5. They had to classify descriptions of cooperative apartments – an object familiar to the subjects. They were offered the following evaluation attributes: (1) Layout and size of auxiliary premises, (2) Interior trim, (3) Aesthetics, (4) Extent of environmental pollution and noise, (5) Price.

To each attribute a scale of three verbal estimates was allotted ordered with respect to quality from the first to the third one. Thus, for example, the following scale was used for the 'Layout' attribute:

(1) Layout is very comfortable, auxiliary premises and kitchen are rather spacious.

(2) Layout is comfortable, auxiliary premises and kitchen are small.
(3) Layout is inconvenient, kitchen is small, there are no auxiliary premises.

Thus we have $N = 5$ and $Wi = 3$ $(i = 1, \ldots, n)$. It can be easily seen that the combinations of different attribute estimates provide a complete set of feasible objects. In this case, the number of feasible objects is: $Q = 3^5 = 243$. A description of 243 hypothetical apartments was presented at random to each subject, who then assigned each combination to one of the following decision classes:

(1) Apartment is good and completely satisfies you.
(2) Apartment is satisfactory, but there are a lot of defects.
(3) Apartment does not suit you.

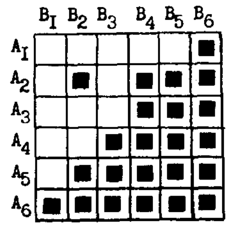
Variables used for assessing the subjects' behavior

Assignment of an object to a certain class, under decision class ranking (the first class is better than the second one, etc.), and ordinal attributes estimate scales, imposes certain constraints on the remaining set of objects. Thus, for example, objects dominating (with respect to estimates) some object cannot be referred to a lower class, while the objects it dominates cannot be placed in a higher class. Violation of these constraints by a subject was considered an error in his or her process of classification. The notion of error in classifying is somewhat similar to the notion of intransitivity under a pairwise comparison of multiattribute objects; both are determined by inconsistencies in the successive decision maker's judgments.

The subjects' behavior was evaluated against three criteria:

(1) Number of inconsistencies

The subjects' job was to assign the objects (combinations of attributes estimates) to ordered classes. Fig. 1 shows an extremely simple version of this problem – assignment to either of the two classes (the first class is better than the second one) of the estimate combinations using two attributes, A and B (the first estimates are better; the estimates on scales are ranked with respect to quality). Fig. 1 presents a hypothetical division into two classes (empty blocks, the 1st class; shaded blocks, the 2nd class). Obviously, the estimate of block A_2B_2 is in conflict with the estimates of blocks A_2B_3, A_3B_2, A_4B_2, A_3B_3. Hence, there are four inconsistencies in the assignments.



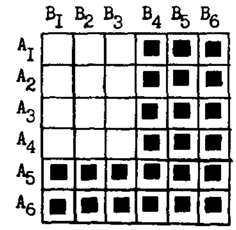


Fig. 1 (*left*). The hypothetical division into two classes with contradictions.

Fig. 2 (*right*). The hypothetical division into two classes without contradictions.

(2) Number of substitutions (errors)

Along with the number of contradictions, information also is provided by the number of changes that must be made in the subject's answers in order to make the classification consistent. Thus, the assignments presented in fig. 1 need only one alteration, assignment of a different class (the 1st), to the combination A_2B_2. This replacement makes the classification consistent. The number of required substitutions indicates the number of assignment errors made by the subject.

(3) Complexity of boundaries between classes

This characteristic as we have suggested in an earlier paper (Larichev and Moshkovich 1980), implies the complexity of the rules the subject employed in class assignment. Thus, the line between classes in fig. 2 is very simple as the subject has replaced attributes by constraints. His decision rule was very simple in this case: combinations with estimates superior to A_5 and B_4 are referred to the first class.

The boundary between classes in fig. 1 is much more complex. It can be seen that it is described by five combinations of estimates using two attributes. There are two reasons for the substitution of an attribute by constraints (see figs. 1 and 2). First of all, some of the subjects view the original problem not as a multiattribute one but as more simple one-attribute problem constrained by other attributes (it is not without reason that in their experiments A. Tversky (Tversky 1969) and D. Russo (Russo and Rosen 1975) selected subjects who used all attributes).

As we shall see later, one and the same person may start using constraints instead of attributes as the problem becomes complicated. As is well known, the successive introduction of constraints instead of attributes (Tversky 1972) is cognitively very simple.

In line with the aforementioned criteria, a level of requirements to the performance quality was set which was behind the judgment as to whether the subject succeeded in managing the classification problem. It is common knowledge that people often make errors in information processing operations. No two errors are alike, however. As is shown in fig. 1, errors made far from the boundary entail a larger number of inconsistencies. These errors are, as a rule, quite obvious. They do not prevent the drawing of boundaries between decision classes. On the other hand the errors made in the vicinity of boundaries are quite another matter. Thus, if the subject assigns block A_2B_3 in fig. 1 to the second class, there would be only one inconsistency (regarding block A_3B_3 belonging to the first class) and the question is whether block A_2B_3 should be referred to the first class or A_3B_3 to the second class. Hence, errors near and on the boundary are especially dangerous because they change the boundary between classes, and with a large number of such errors it is impossible to set boundaries between classes.

Accordingly, the number of errors made on the boundary or nearby – at a distance of a unit (a change for one estimate against any attribute converts the combination into a boundary element) from the boundary – served as a value of the first criterion determining whether the subject managed the classification problem. It was agreed that the subject manages a problem if and only if the number of errors he makes near the boundaries does not exceed two.

The complexity of the boundary reflecting the complexity of the subjects' decision rules served as the second criterion determining whether the subject succeeded in solving the classification problem. Precisely: the boundary elements between classes were required to contain at least one or two elements representing combinations of attributes estimates. To put it another way that if the subject converted all attributes into constraints and turned the problem into 'elimination by aspects' then he failed in managing the problem. Indeed, the multiattribute classification problem simply vanishes in the latter case.

Description of experiments

It is necessary to distinguish between the experiments involving people with almost no decision-making experience (students, pupils – the first series of experiments) and those involving professionals handling real-life problems (second series of experiments).

The first group of subjects had ample opportunities for modifying the classification problem parameters and terms of the experiment. Students (experiments from 1 to 12) classified cooperative apartments by deciding to what extent the presented alternatives satisfied them, and pupils (experiments 13 and 14) classified higher educational establishments with regard to their suitability for entering after graduation from school. The second group of people, when classification was a real decision-making problem, were almost completely deprived of the opportunities for modifying the problem parameters because the experimental pattern corresponded to a real problem. The second series of experiments involved members of an editorial board, who estimated the quality of preprints submitted for publication (experiments 15 and 16), and physicians who diagnosed suspected diseases given a certain set of symptoms (experiment 17).

Results of the 1st series of experiments

Data on the average number of errors made by the subjects during 100 classification runs in each of the 14 experiments of the first series are shown in table 1. As is seen

Table 1
The parameters of the classification problem and major results in various sets of experiments.

Srl. no. of exp.	Number of subjects	*N*	*W*	*P*	*Q*	*E*	% [a]
1st series of experiments							
1	9	7	2	5	128	9.5	11
2	9	7	2	4	128	6.5	0
3 *	19	7	2	3	128	6.5	37
4	15	5	3	4	243	9.7	13
5 *	20	5	3	3	243	5.8	35
6 *	24	5	3	2	243	5.0	46
7	20	4	4	4	256	8.8	10
8	20	4	4	3	256	6.2	20
9 *	9	4	4	2	256	3	67
10	10	3	5	5	125	17	0
11	10	3	5	5	125	8.8	9
12 *	10	3	5	3	125	5.1	60
13	16	5	3	4	243	9.8	19
14 *	16	5	3	2	243	3.5	73
2nd series of experiments							
15 *	9	5	3	2	243	3.3	88
16	4	5	3	4	243	1.3	25
17 *	6	8	3	4	3^8	5.0	100

Note: *N* – number of attributes; *W* – number of divisions on their estimate scales; *P* – number of classes; *Q* – number of classified objects; *E* – average number of errors made
[a] Percentage of subjects accomplishing the task;
* Complexity of this problem is within human capabilities.

from the table, the average number largely depends on the complexity of the classification problem. An ANOVA on the number of errors as a function of parameters N and Wi showed that under a constant $N:W$ ratio the number of errors (substitutions) depends on the number of decision classes (P).

The average time for assigning an object to a class amounted to 14 sec. An additional analysis of classification quality, conducted for each subject for the criteria of quantity and quality of errors and complexity of the decision rule, made it possible to determine whether the subject managed the task properly. The percentage of subjects who succeeded in accomplishing the task in each experiment is shown in table 1. Thus, in experiment 9, for example, for $N = 4$ (number of attributes), $Wi = 4$, $i = 1, \ldots, n$ (number of estimates on ordinal scales), $p = 2$ (number of decision classes) 67% subjects managed the job (the average number of substitutions or errors was 3). For the same $N = 4$ and $Wi = 4$ but for $P = 4$, 90% subjects failed to accomplish the task and there was a marked increase in the number of inconsistencies and errors (the average number of substitutions was 8.8).

Such values of N, W, and P were defined under which a considerable percentage of the subjects failed to manage the task. It was conventionally agreed that if at least one third of the group of 10–15 subjects successfully managed the task, the classification problem of the given complexity was within human capacities. The percentage of subjects successfully managing the task is given in table 1. The experiments, for which in conformity with our criteria, the complexity of classification problem is within the subjects' capacities, are marked with an asterisk (*).

Discussion of results of the 1st series of experiments

The experimental results confirmed the hypothesis that there are certain 'limits' to the subjects' capacities in multiattribute classification problems. The available results indicate that for some parameter values the number of inconsistencies and substitutions sharply increase. The subjects fail in managing the problem, and their answers make it impossible to draw a line between the classes. What is behind the phenomenon? A special analysis was conducted for this purpose. As was noted earlier, the results of the subjects' efforts may be represented in the form of elements of boundaries between the classes. The elements indicate that all alternatives dominating them relate to higher classes, and those they dominate – relate to lower classes. Following the reduction of the subjects' answers to a consistent form, it is easy to define boundaries between classes with the algorithms suggested by Larichev et al. (1986). The boundaries are characterized by both the quantity of the boundary elements and their complexity: by the number of estimates in the boundary element different from the first ones. A substantive analysis of the boundary elements and a comparison with the outcome of the written protocol analysis showed that the elements are classified into structural units (chunks). A set of several boundary elements usually reflects the more general rules which are highly meaningful for the subjects. Such rules may be defined by grouping the boundary elements with respect to their proximity (with respect to the content of similar estimates). Such rules are generally rather easy to remember. For example, one of the rules used by the students was that if the apartment is in an industrial district and expensive, it does not suit (3rd class) unless it has a very comfortable layout. A set of such kinds of rules is reflected in the boundary elements between classes by combining them in structural units of information. The strategies employed by the subjects may be represented as a set of such rules. In two experiments (4 and 13) where the subjects failed, and in two experiments (6 and 14) where they succeeded in solving the classification problem, each subject was allotted rules to be used in classification. The analysis showed that when the subjects managed the task the number of rules they used did not exceed eight. In cases when they failed, a formal analysis revealed a much larger number of rules. An average number of rules used by the subjects in experiments 4 and 13 amounted to 12, and in experiments 6 and 14 to five.

The most suitable explanation for the above data is probably as follows. In assigning an alternative to some class or other, the subject has to keep all rules in short-term memory, constituting structural units of information (chunks) he operates. As is known, the volume of short-term memory is limited. Different papers (Miller 1956; Simon 1981) indicate that it does not exceed 5–7 structural units of information (blocks), and they may differ in size.

When subjects employed 9 or fewer rules (structural units of information) for classification, they managed the task. If more, then a part of rules were abundant for the operating short-term memory which sharply increased the number of errors and inconsistencies.

The average time expended on referring an object to some class indirectly confirms this assumption. An average time of one classification in the first series of experiments amounted to 14 sec. And since the time needed for registering the information in the long-term memory is about 5 sec (Simon 1981), a conclusion suggests itself that no active informational exchange between short-term and long-term memory takes place during classification. According to Simon (1978), one operation in short-term memory takes about 100 msec. Incomparability of this time with that required for the access to the long-term memory forces the information processing system in short-term memory to minimize communications with the long-range memory slowing down the rate of information processing by about two to three orders of magnitude.

Analysis and discussion of results of the second series of experiments

The major purpose of these experiments was to see how professionals managed classification problems and to what extent the results obtained with the groups of students and pupils may be related to real-life classification problems.

Three experiments were conducted in which a multiattribute classification problem was combined with the occupational tasks of the subjects. In experiment 15 the subjects classified preprints pending publication ($N = 5$, $Wi = 3$, $P = 2$). The average number of errors is given in table 1. The analysis showed that all but one subject (members of the editorial board) managed the classification problem with two classes of decision. The analysis of boundaries between the classes indicated that no subject made use of more than five rules during classification, combining the boundary elements in structural units of information.

We expected that as the problem becomes more complicated (experiment 16, 4 decision classes) the number of errors should increase. The number of errors, however, turned out to decrease (see table 1). Of special interest in this connection is the analysis of subjects' strategies. The analysis of boundaries between classes showed that as the problem became more complicated, the subjects simplified their strategies. As a result, two out of four subjects, according to the 'complexity of decision rule' criterion, failed to solve the classification problem in spite of the small number of errors made. The number of rules used in classifying did not exceed the volume of short-term memory. It is worth noting that although the substantive content of the classification problem was familiar to the subjects, (preprint classification), it was presented in a form of description of their attributes estimates, that was unusual to them.

In the last experiment of this series we wanted to see how professionals handle the common, recurring problems of multiattribute classification. The experiment was carried out with a group of physicians taking part in the construction of expert systems. On the basis of a set of symptoms, the physicians assigned the degree of probability of appendicitis. The parameters of the classification problem they coped with are shown in table 1. It is clear that the number of estimate combinations, under the given parameters, is too big for a physician to directly estimate all significant combinations. Therefore, in contrast to other experiments, for eliciting information use was made of the CLASS system (Larichev et al. 1986) of construction of complete and consistent knowledge bases. In working with this system the physician assessed a combination of estimates, presented to him (as in the other experiments). Only a subset of possible combinations was presented and each answer of the expert extended to other estimate combinations with respect to dominance relations. Note that all estimate combinations were classified (directly or indirectly through dominance relation) several times with the aim of checking the expert for consistency. As soon as a contradiction emerged it was immediately presented to the subject who, while removing the contradictions, refined his preferences. Thus, the CLASS system helped the subjects handle the more complex problems. We wanted to determine as to how many rules the physicians used in producing their judgments. The analysis of the boundary elements between classes made it possible to identify the structural units of information reflecting the number and complexity of the decision rules employed in classification. The analysis showed that not more than nine rules were used for solution of the problems. The number of errors and the complexity of the decision rules suggests that all subjects succeeded in solving the classification problem whose complexity exceeded the limits of the subjects' capacities demonstrated in other experiments. We believe, that the reason is that the CLASS system increased the subjects' capabilities by presenting the inconsistent answers to them for correction as soon as they emerged. In the recurring classification problems, which is the case for physicians, the rules and structural units of information reflecting them (chunks), are developed over time on concrete cases that occur in the course of their professional activities. It is interesting to note that the size of structural units of information, which the physicians used in classification (up to 8–10 elements), were very large in comparison to those used by the other subjects. Utilization of larger structural units during classification allows the subjects to group the classification rules, thereby reducing the load on short-term memory.

Thus, we may conclude that the behavior of experienced decision makers differs from that of ordinary people when handling increasingly complicated problems. In solving new, non-recurring classification problems, whose complexity exceeds their capacities, they try, primarily, to be consistent. In doing so, they simplify their task by discarding some attributes from consideration and transferring them to constraints. Having substantially simplified the original problem they in fact solve a different problem that is more amenable to human information processing capabilities. However, when handling recurring problems of multiattribute classification, the decision maker employs more complex rules and copes with problems which are beyond the capabilities exhibited when solving new, unique problems. Here, too, the decision maker's capabilities are not unlimited. The number of structural units of information (chunks) employed in these tasks does not exceed the volume of short-term memory.

General discussion

To interprete the obtained results, it would be helpful to turn to the decision problems classification suggested in Larichev (1984). Depending on the clarity of the various attributes in the choice problem, the authors distinguish the following problems: those with either subjective or objective models; those that are either recurring or unique; those that are either holistic or attributes expert choice. The limitations of short-term memory, common to all people, manifest themselves in different ways in these problems because the structural units of information, employed by people in these problems differ both in structure and size.

In conformity with the presented classification the results of the first 16 experiments relate to unique problems of attributes-expert choice with subjective models.

What do the obtained results then indicate when applied to problems of this type? First of all, they show the existence of distinct limits to human capabilities in multidimensional classification problems. With experienced decision makers these limits are hidden by their ability to greatly simplify the problem and applying a strategy of 'sequential elimination by aspects' (Tversky 1972), which considerably reduces the load on short-term memory, thereby distorting the problem. When applied to new problems of attributes-expert classification, the obtained results lead to a conclusion about human capabilities in solving 90 classification problems differing in complexity. Thus, experiment 3

Table 2
The marginal number of attributes under which the subjects still manage solution of new multiattributes-classification problems.

Number of estimates on ordinal scales	Number of decision classes			
	2	3	4	5
2	7–8	6–7	4–5	3
3	5–6	3–4	2–3	2
4	2–3			

indicates that people are capable of breaking the objects into three classes characterized by seven attributes and two divisions on estimate scales. This leads to the conclusion that 12 easier problems are within human capabilities.

The results of human capabilities in coping with problems of differing complexity are brought together in table 2 which shows the marginal number of attributes under which the subjects still manage multiattribute classification problems (Larichev et al. 1986). The impossibility of solving problems that are more complex than the ones specified in table 2 manifests itself in numerous contradictions, errors, or simplifications of the problem to the detriment of its content.

As for the recurring attributes-expert classification problems, the decision makers' capacities are extended due to utilization of larger structural units of information reflecting their professional skills. However, in those problems too the number of structural units of information employed did not exceed nine.

The question arises as to what extent people are incapable of overcoming this limit. Is there any opportunity to actually extend their capabilities in solving complex multiattribute problems? We believe, there is. We have mentioned earlier that with special man–machine systems, like our CLASS system, increasing human capabilities is quite feasible. There are, however, other opportunities that make no use of computers. One is to substitute parallel problems with a large load on short-term memory by sequential problems. In this first place, it is worth mentioning the hierarchical decision rules when classifications hierarchy can be used (Larichev et al. 1986). A necessary prerequisite for this is comprehension of notions used at each level of the hierarchy. Of course, this approach is far from universal, so in each case it should be used in a creative manner. We also conducted special experiments to test the efficiency of different strategies of classification of problem solution. Thus, utilization of a sample classification strategy, when classification proceeds stepwise (first, all objects of the first class are selected, then of the 2nd class, etc.), reduced the number of errors by 1.5 times as compared with a standard classification when subjects refer an object to some or other class following a sequential consideration of the list.

The results show that there are limits to human capacities in a wide range of intellectual problems. These constraints are objective and determined by the specifics of the human information processing system. In our opinion, such limits can also be determined in other decision problems. What is needed are new behavior evaluation criteria and new experiments.

References

Larichev, O.I., 1984. Psychological validation of decision methods. Journal of Applied Systems Analysis, 11. 37–46.

Larichev, O.I. and H.M. Moshkovich, 1980. 'On possibilities of eliciting consistent estimates of multidimensional alternatives from people'. In: Descriptive research into decision making procedures under multiple criteria. Moscow: Institute for Systems Studies. pp. 60–67.

Larichev, O.I. and H.M. Moshkovich, 1986. Direct classification problems in decision making. Transactions of the Academy of Sciences 287. 567–570.

Larichev, O.I., H.M. Moshkovich and E.M. Furems, 1986. Decision support system 'CLASS'. In: B. Brehmer, H. Jungerman, P. Lourens and G. Sevon (eds.), New direction in research on decision making. Amsterdam: North-Holland. pp. 305–315.

Larichev, O.I., V.S. Naginskaya and A.I. Mechitov, 1984. Interactive procedure of alternative industrial building choice'. In: Procedures of multiattribute objects estimation. Moscow: Institute for Systems Studies. pp. 84–93.

Miller, G., 1956. The magical number seven plus or minus two: some limits on our capacity for processing information. Psychological Review 63, 81–97.

Russo, I. and L. Rosen, 1975. An eye fixation analysis of multialternative choice. Memory and Cognition 3. 267–276.

Simon, H.A., 1978. 'Information-processing theory of human problem solving'. In: W.K. Estes (ed.), Handbook of learning and cognitive processing, Vol. 5: Human information processing. Hillsdale, NJ: Erlbaum. pp. 271–295.

Simon, H.A., 1981. The sciences of the artificial (2nd ed.). London: MIT.

Tversky, A., 1969. Intransitivity of preferences. Psychological Review 76, 31–48.

Tversky, A., 1972. Elimination by aspects: a theory of choice. Psychological Review 79, 281–299.