Techniques for reducing dimensionality of attribute space

A B Petrovsky¹,²,³

¹ Federal Research Center “Computer Science and Control”, Russian Academy of Sciences, 9, pr. 60 Letiya Octyabrya, Moscow, 117312, Russian Federation
² Volgograd State Technical University, 28, Lenina Avenue, Volgograd, 400005, Russian Federation
³ Belgorod State Technological University Named After V.G. Shukhov, 46, ul. Kostyukova, Belgorod, 308012, Russian Federation

E-mail: a.b.petrovsky@gmail.com

Abstract. The paper describes techniques for reducing the dimensionality of attribute space where objects are represented as vectors, tuples and multisets with numerical and/or verbal characteristics. In these tools, many initial attributes are aggregated into a single integral index or several composite indicators with small scales of qualitative estimates. Aggregation of indicators includes various methods for transformation of attributes and their scales. Reducing the dimensionality of attribute space allows us to simplify the solution of applied problems, in particular, problems of multiple criteria choice, and explain the obtained results.

1. Introduction

The popular decision-making methods [2-7] are poorly suitable for solving problems of a strategic and unique choice, in which there are very few objects, and a lot (tens or hundreds) of object characteristics. In real situations, when objects are described by numerous attributes, it is difficult to rank or classify objects, select the best one because, as a rule, many objects are not comparable formally. Examples of such objects are configurations of complex technical systems, schemes of transportation networks, places of airports or power plants, routes of gas or oil pipelines, and the like. Additional difficulties arise for ill-structured problems combining quantitative and qualitative dependencies, the modelling of which is either very hard or impossible in principle.

We can facilitate the object choice in a large space of attributes and reduce data loss in the following ways: using psychologically correct operations to obtain information from decision-makers and experts; reducing the dimension of attribute space. It has been experimentally confirmed that, due to the peculiarities of human physical memory, a person easier compares objects with a small number of indicators, makes fewer mistakes when indicators have not numerical, but verbal scales [4, 5].

The reduction of dimension of the object description is a decrease of number of the indicators, which characterize the properties, state or functioning of objects. The reduction includes some data transformations that combine the set of initial attributes $K_1, \ldots, K_n$ into smaller sets of new intermediate attributes $L_1, \ldots, L_m, \ldots$ and final attributes $N_1, \ldots, N_l$. Formally, we can write these transformations as

$$K_1, \ldots, K_n \rightarrow L_1, \ldots, L_m \rightarrow \ldots \rightarrow N_1, \ldots, N_l.$$ (1)
The initial attribute $K_i$ has the scale $X_i = \{x_i^1, \ldots, x_i^h\}$, $i = 1, \ldots, n$, the intermediate attribute $L_j$ has the scale $Y_j = \{y_j^1, \ldots, y_j^j\}$, $j = 1, \ldots, m$, the final attribute $N_k$ has the scale $Z_k = \{z_k^1, \ldots, z_k^l\}$, $k = 1, \ldots, l$, $l < m < n$. The scale of each indicator has numerical (point, interval) or verbal evaluation grades.

Usually, decreasing the dimension of attribute space is a non-formalized multi-stage procedure based on the knowledge, experience and intuition of a decision-maker or expert. An actor constructs a hierarchical system of attributes as an aggregation tree, defines the rules for the conversion of indicators, and specifies the structure, number and scales of new indicators. Verbal grades of scales characterize desirable new properties of compared objects. The attributes are aggregated consecutively step by step. At each level of hierarchy, including the highest level, the actor determines which of attributes is considered as the independent final indicator and which ones are combined into particular intermediate indicators. The upper level of the aggregation tree can consist of several final indicators that implement the idea of multi-criteria choice, or be a single integral index that implements the idea of holistic choice [7].

Methods for shortening the space dimension depend on characteristics of aggregating attributes and features of rating scales. However, almost all of them operate with numerical attributes [1-3, 11]. This paper describes techniques, which allows reducing a large number of initial attributes of objects into a single integral index or several composite indicators with small verbal scales. An aggregation and transformation of characteristics can significantly diminish the complexity of original applied problems, simplify the problem solution and reasonable explanation of the obtained results, particularly for problems of individual and group multi-criteria choice.

2. HISCRA-M method

Traditionally, we associate an object (alternative, variant) $O_p$, $p = 1, \ldots, q$ with a vector or tuple $x_p = (x_{p1}, \ldots, x_{pm})$. A component $x_{pi}$ is a value of the attribute $K_i$ equal to $x_{pi}$, $i = 1, \ldots, h$, if all attributes $K_1, \ldots, K_n$ have the same scale, $X = \{x^1, \ldots, x^h\}$, or equal to $x_{pi}$, $e = 1, \ldots, h$, if each attribute $K_i$ has its own scale $X_i = \{x_i^1, \ldots, x_i^h\}, i = 1, \ldots, n$. The vector/tuple $x_p$ is a point of the n-dimensional space $X_1 \times \ldots \times X_n$.

Different copies (versions, exemplars) $O_p^{<s)}, s = 1, \ldots, t$ of the object $O_p$ arise, for example, when the object is evaluated by $t$ experts upon many criteria $K_1, \ldots, K_n$, or the characteristics of the object are calculated $t$ times by several methods $K_1, \ldots, K_n$, or measured $t$ times by several tools $K_1, \ldots, K_n$. In such cases, the object $O_p$ does not correspond to a single vector/tuple, but a group of $t$ vectors/tuples $\{x_p^{<1)}, \ldots, x_p^{<t}\}$ where $x_p^{<s)} = (x_{p1}^{<s)}, \ldots, x_{pm}^{<s)})$ describes one of the versions $O_p^{<s)}$ of the object $O_p$. In the $n$-dimensional attribute space $X_1 \times \ldots \times X_n$, the object $O_p$ is now represented not as a single point $x_p$, but as a group (“cloud”) consisting of $t$ points $\{x_p^{<1)}, \ldots, x_p^{<t}\}$.

It is important that the group of vectors/tuples $x_p^{<1)}, \ldots, x_p^{<t}$, representing the object $O_p$, should be considered on the whole. Moreover, generally speaking, values of the same characteristics describing different versions $O_p^{<s)}$ of the object $O_p$ (estimates of different experts, characteristics measured by different methods or tools) can be similar, different, or even contradictory [7, 9]. This, in turn, can lead to incomparability of vectors/tuples $x_p^{<s)}$, representing one and the same object $O_p$. Thus, the collection of multi-attribute objects $O_1, \ldots, O_q$, each of which corresponds to its own “cloud” consisting of $t$ different points, is quite difficult to analyze. Therefore, it is desirable to simplify the description and aggregation of such objects.

For numerical attributes $K_1, \ldots, K_n$, it is easiest to represent each object $O_p$ with a single vector $x_p^{con} = (x_{p1}^{con}, \ldots, x_{pm}^{con})$, components of which are determined by additional formal conditions or substantial considerations. For example, $x_p^{con}$ could be a vector that is the center of group; the vector closest to all vectors of group; a vector with total, averaged, or weighted values of components of the vectors $x_p^{<1)}, \ldots, x_p^{<t}$ describing versions $O_p^{<1)}, \ldots, O_p^{<t}$ of the object $O_p$. In the case of symbolic, verbal, or mixed attributes $K_1, \ldots, K_n$, a group of tuples, representing versions of any object, cannot be replaced, even in principle, by a single tuple with total, averaged, weighted, mixed values of components, since such mathematical operations on such variables are not feasible.

In the reduced $l$-dimensional space $Z_1 \times \ldots \times Z_l$ of final indicators, the object $O_p$ correspond to a vector/tuple $z_p = (z_{p1}, \ldots, z_{pl})$, or a group of vectors/tuples $\{z_p^{<1)}, \ldots, z_p^{<s)}\}$, where $z_p^{<s)} = (z_{p1}^{<s)}, \ldots, z_{pl}^{<s)})$.
A component \( z_{pk} \) is equal to \( z_{pk}' \), \( e_1 = 1, \ldots, q_i \), if all indicators \( N_1, \ldots, N_i \) have the same scale \( Z = \{ z^1, \ldots, z^d \} \), or equal to \( z_{pk}' \), \( e_2 = 1, \ldots, q_i \), if each indicator \( N_i \) has its own scale \( Z_k = \{ z^{1_k}, \ldots, z^{d_k} \}, k = 1, \ldots, l. \)

The HISCRAM (Hierarchical Structuring Criteria and Attributes by Many experts) method allows us reducing the dimension of the attribute space and building several hierarchical aggregation trees when multi-attribute objects are represented by vectors/tuples [8, 9]. The task (1) takes the form:

\[
X_1 \times \ldots \times X_n \to Y_1 \times \ldots \times Y_m \to \ldots \to Z_1 \times \ldots \times Z_l \tag{2}
\]

The dimension of the corresponding attribute space is defined as the power of the direct product of numerical or verbal scales, grades of which are components of vectors/tuples. We use two tools for data transformations: a reduction of attribute scale and an aggregation of attributes.

The reduction of attribute scale is aimed at decreasing the number of gradations due to combining several grades of any attribute into a single new grade. The transition to a smaller number of grades on the scales is the transformation (2) in the form:

\[
X_1 \times \ldots \times X_n \to R_1 \times \ldots \times R_m
\]

where \( R_i = \{ r_1^i, \ldots, r_{d_i}^i \} \) is the shortened scale of the attribute \( K_i, |R_i| = d_i < h_i = |X_i|, i = 1, \ldots, n. \) For example, several gradations \( x_1^{a_1}, x_1^{a_2}, \ldots, x_1^{a_n} \) of the original scale \( X_1 \) of the attribute \( K_1 \) are combined into a single gradation \( r_1^{a_0} \) of the shortened scale \( R_1.\) It is desirable to form shortened scales so that they consist of a small number (2-4) of grades, which have certain content for the actor.

In the reduced space of attributes \( K_1, \ldots, K_n \) with scales \( R_1, \ldots, R_n, \) the object \( O_i \) correspond to a vector/tuple \( r_p = (r_{p1}, \ldots, r_{pn}), \) or a group of vectors/tuples \( \{ r_{p1}^{a_1}, \ldots, r_{pn}^{a_1} \}, \) where \( r_{pi}^{a_1} = (r_{p1}^{a_1}, \ldots, r_{pn}^{a_1}). \) A component \( r_{pi} \) is a value of the attribute \( K_i \) equal to \( r_{pi}^{a_0}, a = 1, \ldots, d_i, \) if all attributes \( K_1, \ldots, K_n \) have the same scale \( R = \{ r_1^{a_0}, \ldots, r_n^{a_0} \}, \) or equal to \( r_{pi}^{a_0}, a = 1, \ldots, d_i, \) if each attribute \( K_i \) has its own scale \( R_i = \{ r_1^i, \ldots, r_{d_i}^i \}, i = 1, \ldots, n. \)

The aggregation of attributes is aimed at decreasing the number of attributes due to combining several attributes into a single new attribute. We shall call this new attribute a composite indicator. The transition to a smaller number of attributes is the transformation (2) in the form:

\[
Y_1 \times Y_2 \times \ldots \times Y_c \to Z_u
\]

where \( Y_j = \{ y_1^j, \ldots, y_{k_j}^j \} \) is the scale of the original attribute \( L_{ji}, j = a, b, \ldots, c, Z_u = \{ z_u^1, \ldots, z_u^{l_u} \} \) is the scale of the composite indicator \( N_u, |Z_u| = f_u < g_a + g_b + \ldots + g_c = |Y_a| + |Y_b| + \ldots + |Y_c|. \) It is recommended to combine 2-4 original attributes in a composite indicator with a small (2-4 grades) scale. In practical tasks, it is convenient to build scales of all indicators so that they have the same number of grades, that is \( g_a = g_b = \ldots = g_c = f_u = d, \) and each grade of the composite indicator scale consists of similar grades of the combined attribute scales. For instance, all high grades on the scales of original attributes generate the high grade on the scale of composite indicator, all middle grades generate the middle grade, and all low grades generate the low grade.

We shall construct an aggregation tree of composite indicators from unified blocks specified by the actor. Each block at any hierarchical level of tree is considered as a multi-criteria ordinal classification problem. The classified objects are tuples \( (y_1^{a_c}, y_2^{b_c}, \ldots, y_c^{b_c}) \) of original attributes \( L_{a}, L_{b}, \ldots, L_{c}, \) and the decision class is a verbal grade \( z_u^{a_c} \) on the rating scale of a composite indicator \( N_u [7-9]. \) Finally, we can build grades of the integral index.

Rating scales of composite indicators can be formed using different methods. The simplest method is the tuple stratification, when the multi-attribute space is cut with parallel hyper-planes [2]. Each layer (stratum) consists of combinations of the homogeneous (for example, with the fixed sum of grade numbers) original attributes and represents any generalized grade on the scale of composite indicator. Methods of the verbal decision analysis are more complicated [5]. The ORCLASS (ORdinal CLASSification) method builds a complete and consistent classification of all tuples of original attributes. Decision classes are defined by their boundaries and correspond to grades of composite indicator scale. The ZAPROS (the abbreviation of Russian words: CLosed PRocedures
nearby Reference Situation) method allows us to construct a joint ordinal scale of composite indicator from original attributes. The actor determines a number of scale grades (layers or classes) and can apply different techniques simultaneously at various hierarchical levels.

3. SOCRATES method

A multiset or a set with repetitions is a convenient mathematical model for representing and comparing objects, which are described by many numerical and/or verbal attributes and presented in several versions that differ in the values of their characteristics [7-9].

When all attributes $K_1, ..., K_n$ have the same rating scale $X = \{x^1, ..., x^h\}$ of numerical and/or verbal grades, we associate the object $O_p$, $p = 1, ..., q$ with a multiset:

$$A_p = \{k_{Ap}(x^1)x^1, ..., k_{Ap}(x^h)x^h\}$$

over the set $X = \{x^1, ..., x^h\}$. When each attribute $K_i$ has its own rating scale $X_i = \{x_i^1, ..., x_i^h\}$, $i = 1, ..., n$, we associate the object $O_p$, $p = 1, ..., q$ with a multiset:

$$A_p = \{k_{Ap}(x_i^1)x_i^1, ..., k_{Ap}(x_i^h)x_i^h\}$$

over the set $X = X_1 \cup ... \cup X_n = \{x_1^1, ..., x_1^h, ..., x_n^1, ..., x_n^h\}$, which consists of $n$ groups of attributes and combines all grades on the scales of all attributes. In (3) or (4), the multiplicity $k_{Ap}(x^i)$ or $k_{Ap}(x_i^i)$ shows how many times the estimate $x^i \in X$, $e = 1, ..., h$, or $x_i^e \in X_i$, $e = 1, ..., h_i$ is present in the description of the object $O_p$.

A variety of operations on multisets provides the ability to group multi-attribute objects in different ways [7, 9]. A multiset $A$ that represents the group of objects can be formed by the sum $A = \sum A_i$, $k_A(x^i) = \sum k_{A_i}(x^i)$, union $A = \bigcup A_i$, $k_A(x^i) = \max k_{A_i}(x^i)$, intersection $A = \bigcap A_i$, $k_A(x^i) = \min k_{A_i}(x^i)$ of multisets $A_i$ that describe the grouping objects, or by one of the linear combinations of operations on multisets $A_i$ such as $A = \sum c_i A_i$, $A = \bigcup c_i A_i$, $A = \bigcap c_i A_i$, $c_i > 0$ is an integer. The copy (version) $O_p \circ \circ \circ \circ$ corresponds to a multiset $A_p \circ \circ \circ \circ$ of the form (3) or (4) with the multiplicities $k_{Ap} \circ \circ \circ \circ (x^i)$ or $k_{Ap} \circ \circ \circ \circ (x_i^i)$. We will form the multiset $A_p$ as a weighted sum of multisets as follows:

$$A_p = c^{15} A_p \circ \circ \circ \circ + ... + c^{15} A_p \circ \circ \circ \circ$$

where the multiplicity of $A_p$ is calculated by the rule $k_{Ap} \circ \circ \circ \circ (x^i) = \sum c^{15} k_{Ap} \circ \circ \circ \circ (x^i)$, and the coefficient $c^{15}$ characterizes the significance (expert competence, measurement accuracy) of the object version $O_p \circ \circ \circ \circ$.

The SOCRATES (ShOrtening CRiteria and ATtributES) method allows us reducing the dimension of the attribute space and building several hierarchical aggregation trees when multi-attribute objects are represented by multisets [10]. The task (1) takes the form:

$$X_1 \cup ... \cup X_n \rightarrow Y_1 \cup ... \cup Y_m \rightarrow ... \rightarrow Z_1 \cup ... \cup Z_t.$$  \hspace{1cm} (5)

The dimension of the corresponding attribute space is defined as the power of the union of numerical or verbal scales, grades of which are elements of multisets. We use two tools for data transformations: a reduction of attribute scale and an aggregation of attributes.

The reduction of attribute scale to a smaller number of grades is the transformation (5) as follows:

$$X_1 \cup ... \cup X_n \rightarrow R_1 \cup ... \cup R_n,$$

where $R_i = \{r_i^1, ..., r_i^{d_i}\}$ is the shortened scale of the attribute $K_i$, $|R_i| = d_i < h_i = |X_i|$, $i = 1, ..., n$. For example, several gradations $x_i^{o_1}, x_i^{o_2}, ..., x_i^{o_{d_i}}$ of the original scale $X_i$ of the attribute $K_i$ are combined into a single grade $r_i^{o_i}$, $o_i = 1, ..., d_i$ of the shortened scale $R_i$. It is desirable to form shortened scales so that they consist of a small number (2-4) of grades, which have certain content for the actor.

In the reduced space of attributes $K_1, ..., K_n$ with rating scales $Q_1, ..., Q_n$, the object $O_p$ correspond to a multiset of the form (3) or (4):

$$B_p = \{k_{Bp}(r_1^{o_1})r_1^{o_1}, ..., k_{Bp}(r_n^{o_n})r_n^{o_n}\}$$

over the set $R_1 \cup ... \cup R_n$ of grades of the shortened scales. The multiplicity of the element $r_i^{o_i}$ of
the multiset $B_p$ is determined by the sum of multiplicities of the combined grades $x_1^e, x_2^e, \ldots, x_{k_p}^e$:

$$k_{B_p}(x_i^e) = k_{A_p}(x_i^e) + k_{A_p}(x_i^e) + \ldots + k_{A_p}(x_i^e).$$

The aggregation of attributes is the transition (5) to a smaller number of composite indicators:

$$Y_1 \cup Y_2 \cup \ldots \cup Y_t \rightarrow Z,$$

where $Y_j = \{y_1^1, \ldots, y_{k_j}^e\}$ is the scale of the original attribute $L_j, j = a, b, \ldots, c$, and $Z_a = \{z_1^1, \ldots, z_{k_c}^e\}$ is the scale of the composite indicator $N_a$, where $|Z_a| = g_a + g_b + \ldots + g_c = |Y_a| + |Y_b| + \ldots + |Y_c|$. It is recommended to combine 2-4 original attributes in a composite indicator with a small (2-4 grades) scale. In practical tasks, it is convenient to build scales of all indicators so that they have the same number of grades, that is $g_a = g_b = \ldots = g_c = f_u = d$, and each grade of the composite indicator scale consists of similar grades of the combined attribute scales.

In the space of original attributes $L_1, \ldots, L_m$, let the object $O_p, p = 1, \ldots, q$ be defined by a multiset:

$$I_p = \{k_{B_p}(y_1^1)y_1^1, \ldots, k_{B_p}(y_{k_j}^e)y_{k_j}^e, \ldots; k_{B_p}(y_1^1)y_1^1, \ldots, k_{B_p}(y_{k_j}^e)y_{k_j}^e\}$$

over the set $Y_1 \cup \ldots \cup Y_m$ of grade scales. In the reduced space of composite indicators $N_1, \ldots, N_t$, the object $O_p$ correspond to a multiset of the form (3):

$$J_p = \{k_{B_p}(z_1^1)z_1^1, \ldots, k_{B_p}(z_{k_i}^d)z_{k_i}^d, \ldots; k_{B_p}(z_1^1)z_1^1, \ldots, k_{B_p}(z_{k_i}^d)z_{k_i}^d\}$$

over the set $Z_1 \cup \ldots \cup Z_t$ of grade scales. Here, all scales $Y_j = \{y_1^1, \ldots, y_{k_j}^e\}, j = 1, \ldots, m$, and $Z_k = \{z_1^1, \ldots, z_{k_d}^e\}, k = 1, \ldots, d$ have the same number $d$ of grades. The multiplicity of the element $z_{k_e}^e, e = 1, \ldots, d$ of the multiset $J_p$ is determined by the sum of multiplicities of the combined elements $y_{a}^e, y_{b}^e, \ldots, y_{c}^e$:

$$k_{B_p}(z_{k_e}^e) = k_{B_p}(y_{a}^e) + k_{B_p}(y_{b}^e) + \ldots + k_{B_p}(y_{c}^e).$$

Finally, we can build grades of the integral index.

4. Conclusion

The HISCRA-M and SOCRATES methods for reducing the dimension of the attributes space have certain universality, as it allows to operate simultaneously with both verbal (qualitative) and numerical (quantitative) data. An attractive feature of the methods is the possibility to use them in combination with various decision-making methods and information processing technologies. And most important, the initially available information is not distorted or lost.

In the HISCRA-M and SOCRATES methods, we form several aggregation trees with different rating scales of composite indicators in order to reduce the influence of the methods’ features and increase the validity of the results. We shall consider such hierarchical systems of indicators as different judgements or viewpoints of several experts. So, at the conclusive stage, every object will be presented in several versions (copies) with different sets of estimates. Thus the considered problem is transformed into a problem of group multi-attribute choice.

The HISCRA-M and SOCRATES methods are easily integrated into the original multi-stage technology PAKS (Progressive Aggregation of the Classified Situations) and multi-method technology PAKS-M (Progressive Aggregation of the Classified Situations by many Methods) [8, 9] for solving problems of multi-criteria choice in large-dimensional spaces. These technologies provide greater validity for choosing the most preferable object because the problem considered is solved by several decision methods An understandable explanation of the obtained results helps a decision-maker to find the most suitable scheme for aggregating attributes, or to apply several schemes together.

Technologies for solving multi-criteria choice problems in large-dimensional spaces were used to select a prospective personal computing complex, evaluate the results of scientific research, rate organizations by the activity effectiveness [8, 9]. The use of the HISCRA-M and SOCRATES methods will vastly reduce the complexity, labor costs and time of solving similar practical problems.
5. Acknowledgments

This work was supported by the Russian Foundation for Basic Research (the projects 17-29-07021, 18-07-00132, 18-07-00280, 19-29-01047).

References
[1] Ayvazyan S A, Bukhshtaber V M, Enyukov I S, and Meshalkin L D 1989 Applied Statistics. Classification and Reduction of Dimension (Moscow: Finansy i statistika)
[2] Glotov V A and Pavel’ev V V 1984 Vector Stratification (Moscow: Nauka)
[3] Hartigan J A 1975 Clustering Algorithms (New York: Wiley)
[4] Kahneman D, Slovik P, and Tversky A 1982 Decision-Making in Uncertainty: Heuristics and Biases (Cambridge: Cambridge University Press)
[5] Larichev O I 2006 Verbal Decision Analysis (Moscow: Nauka)
[6] Noghin VD 2018 Reduction of the Pareto Set. An Axiomatic Approach (Springer International Publishing AG)
[7] Petrovsky A B 2009 Decision Making Theory (Moscow: Publishing Center “Academiya”)
[8] Petrovsky A B and Lobanov V N 2015 Sci. Tech. Inf. Proc. 42(5) 76-86
[9] Petrovsky A B 2019 Group Verbal Decision Analysis (Moscow: Nauka)
[10] Petrovsky A B 2020 Art. Intel. and Dec. Making 2 63-77
[11] Samet H 2006 Foundation of Multidimensional and Metric Data Structures (Boston: Elsevier)