Multiple Criteria Project Selection Based on Contradictory Sorting Rules

Alexey B.Petrovsky

Institute for System Analysis
Russian Academy of Sciences
Prospect 60 Let Octyabrya 9
Moscow 117312, Russia
pab@isa.ru

Abstract: This paper considers the techniques for selecting a collection of projects which are estimated by several experts with many qualitative criteria and sorted beforehand into some classes by diverse individual sorting rules. These tools are based on the theory of multiset metric spaces and allow to generate classes of such objects and define a general classification rule which approximates the set of contradictory sorting rules.

1 Introduction

There are practical tasks where plurality and redundancy of data that characterize objects, alternatives, situations, and their properties are peculiar. Let $X=\{A_1,...,A_k\}$ be a collection of k projects submitted for the program. Suppose that these projects are evaluated beforehand by n experts by m qualitative criteria $Q_1,Q_2,...,Q_m$ and sorted into some classes $X_1,...,X_f$. For instance, the questionnaire for the project estimation includes the following criteria: Q_1 "The project contribution to the program goals", Q_2 "Long -range value of the project", Q_3 "Novelty in the approach to solve the task", Q_4 "Qualification level for the team", Q_5 " Resources available for the project realization", Q_6 "Character of the project results" [La89].

Each criterion has a nominative or ordered scale of verbal estimates. The scale of the criterion Q_4 "Qualification level for the team" can look like this:

q_4^1 – the team is one of the best by the experience and qualification level;
q_4^2 – the team has the experience and qualification level sufficient for the project realization;
q_4^3 – the team has the experience and qualification level insufficient for the project realization;
q_4^4 – an experience and qualification level of the team are unknown.

Several experts evaluate each project with all criteria $Q_1,...,Q_m$ and make one of the following recommendations:

r_1 – to approve the project;

r_2 – to reject the project;

r_3 – to consider the project later after improving.

Obviously, one and the same project can be differently evaluated by different experts. The expert recommendations for the project approval may also coincide or not.

The project description consists of several groups of attributes $G=\{Q_1,...,Q_m,R\}$. Each group $Q_s=\{q_s^{e_s}\}$ (s=1,…,m; e_s=1,…,h_s) is the attribute family that expresses the project properties. The group of attributes $R=\{r_t\}$ is the set of expert individual sorting rules, which are assigned each project into the specific class X_t (t=1,…, f) and may be similar or discordant. Since each project is estimated by several independent experts, some of attributes $q_s^{e_s}$, r_t can be repeated several times. So, for the project selection it is desirable to construct a simple general classification rule that approximates the variety of expert rules and estimates.

This paper describes an approach to generating the classes of such multiple criteria projects and defining boundaries between classes. The technique for project classification and construction of the general decision rule is based on representing projects with multisets and searching for the best decomposition of multisets in metric spaces.

2 Problem of Object Classification

Any classification deals with combining the initial collection of objects into several groups or sorting them out of the predefined categories. Information about the object properties can be presented with the set of attributes whose values are numerical and/or verbal. >From the point of formal logic, a procedure for object classification can be written as a sequence of the following decision rules:

$$\text{IF} \;\; \langle \text{conditions} \rangle, \text{THEN} \;\; \langle \text{decision} \rangle. \qquad (1)$$

There are direct and indirect classifications. The direct classification is an enumeration of objects within the class. So in this case, the term conditions includes the names of objects or the list of attribute values that describe the objects. The indirect classification is based on properties common for the class. And the term conditions expresses relations between different attributes and/or their values. The term decision signifies that an object belongs to a specific class.

When objects are sorted by many experts, there is a family of decision rules which may be similar, diverse, or contradictory. Individual sorting rules are coincident or similar when the objects with identical or resembling values of attributes are included in the same class. Contradictory rules assign weakly discernible objects into diverse classes.

Inconsistencies of individual rules may be caused, for instance, by errors in expert classification of objects, incoherence between expert' estimates of objects and decision classes, intransitivity of expert judgements, and by other reasons. Note that knowledge bases of expert systems are built in the same manner.

If the number of objects and attributes is rather small, then decision rules are reviewed and utilized relatively easily. The greater the family of decision rules, the more difficult the analysis of these rules. In this case the problem arises: how to generate the simple approximating rule(s) which would maximally coincide with the individual sorting rules. The final decision rule would include a minimal number of attributes and assign objects into the given classes with admitted accuracy. The construction of generalized decision rule allows also to discover divergences in the initial contradictory sorting rules and to correct them if necessary.

The classification of multi-attribute qualitative objects has some additional peculiarities. Firstly, the amount, complexity and peculiarities of information necessary to specify qualitative objects are essentially larger and more varied then that for quantitative objects. Secondly, multiplicity and redundancy of factors, that express the substance of the considered problem, are possible. Thirdly, the multi-attribute space and indexes of similarity/difference between objects are to be chosen corresponding to the qualitative nature of object properties. And finally, in order to classify qualitative objects, a lot of verbal and numerical data are to be taken into consideration simultaneously and processed without unfounded transformations (like “averaging”, “mixing”, “weighting” attributes, and so on). So, the special procedures to collect and process these kinds of data are needed [LM97], [Pe94], [Pe97].

3 Multiset Model for Presentation of Multiple Criteria Projects

A multiset (also called a bag) or set with repeating elements is a very convenient mathematical model in order to present and analyze a collection of objects, that are described with many qualitative attributes and can exist in several copies with various values of attributes. Unlike the set, each element may occur in the multiset more than once [Kn69], [Ya86], [Pe94], [Pe97]. Let $G=\{g_1,g_2,...,g_j,...\}$ be a crisp set, where all elements g_j are different. A is called a multiset drawn from G if A can be presented by the set of pairs as $A=\{n_A(g)\ g\}$, where $n_A(g)$ is called a counting function. This function defines the number of occurrences of the element g in the multiset A, and n_A: G $\ N^+$. The element g is said to be a member of the multiset A ($g\ \ A$), and there are k copies of g in A, if $n_A(g)=k>0$. If $n_A(g)=0$, then $g\ \ A$. The sign denotes that n_A copies of element g occur within the multiset A. When $n_A(g)$ is equal to $_A(g)=\{0,1\}$, the multiset A becomes an ordinary set. The set $G=\{g_j\}$ is said to be a generic domain for the collection of multisets X, if all multisets from X are composed from the elements of G. The multiset is called: an empty multiset , if $n\ (g)=0$ for all $g\ \ G$; a maximum multiset U, if $n_U(g)=\max\ n_A(g)$ for all multisets $A\ \ U$. The cardinality of multiset A is a total number of all elements in the multiset $|A| = \sum_{g \in G} n_A(g)$; and the dimensionality of multiset A is a number of different elements in the multiset $[A] = \sum_{g \in G} \chi_A(g)$.

Define the following operations under multisets:

a union of multisets $A \cup B = \{g \mid n_{A \cup B}(g) = \max(n_A(g), n_B(g))\}$;

an intersection of multisets $A \cap B = \{g \mid n_{A \cap B}(g) = \min(n_A(g), n_B(g))\}$;

a sum of multisets $A+B = \{g \mid n_{A+B}(g) = n_A(g) + n_B(g)\}$;

a difference of multisets $A - B = \{g \mid n_{A-B}(g) = n_A(g) - n_{A \cap B}(g)\}$;

a symmetrical difference of multisets $A \Delta B = \{g \mid n_{A \Delta B}(g) = |n_A(g) - n_B(g)|\}$;

a multiplication of multiset on scalar k $k \times A = \{g \mid n_{k \times A}(g) = k n_A(g), k>0, \text{integer}\}$;

a product of multisets $A \times B = \{g \mid n_{A \times B}(g) = n_A(g) \times n_B(g)\}$;

a complement of multiset $\overline{A} = \{g \mid n_{\overline{A}}(g) = n_U(g) - n_A(g)\}$.

The metric spaces of multisets were introduced in [Pe94]. Different metric spaces (X,d) can be determined for the same collection of objects by introducing various types of distances $d(A,B)$. The following metrics can exist in multiset spaces:

$$d_0(A,B) = m(A \Delta B); \quad d_1(A,B) = m(A \Delta B)/m(U); \quad d_2(A,B) = m(A \Delta B)/m(A \cup B).$$

Here, $m(A)$ is the measure of multiset A that can be determined in various ways, for instance, as a linear combination of counting functions $m(A) = \sum_j w_j n_A(g_j)$, $w_j > 0$. Functions $d_1(A,B)$ and $d_2(A,B)$ satisfy the normalization condition $0 \le d(A,B) \le 1$. Note, that due to the continuity of the multiset measure, the distance $d_2(A,B)$ is undefined for $A=B=\emptyset$. So, $d_2(\emptyset,\emptyset)=0$ by the definition.

In our case, the multiple criteria project A_i ($i=1,\ldots,k$) can be presented as the following multiset $A_i \in X$ drawn from the domain $G=\{Q_1,\ldots,Q_m,R\}=\{g_j\}$:

$$A_i = \{(n_i(g_j) \circ g_j)\} = \{(n_i(q_s^{e_s}) \circ q_s^{e_s}), (n_i(r_t) \circ r_t)\}. \tag{2}$$

Here $n_i(g_j)$ is a number of attribute g_j ($j=1,\ldots, h$, $h=h_1+\ldots+h_m+f$) or, in other words, a number of experts who have estimated the project A_i with the attribute g_j. The arguments in the formula (2) are associated with the decision rule (1) as follows: the various combinations of attribute values $n_i(q_s^{e_s})$ correspond to the term "conditions"; the term "decision" reflects that the project A_i belongs to the class X_t, that is $A_i \in X_t$ iff $n_i(r_t) > \sum_{p \ne t} n_i(r_p)$. So the problem of project selection can be considered as the problem of multiset classification in a metric space.

In order to simplify the problem assume that the collection of projects $X=\{A_1,...,A_k\}$ is to be sorted only in two classes X_a and X_b. In this case the collection X can be represented as the following decompositions of multisets:

$$X = \sum_{t=a,b} X_t = \sum_{t=a,b} \left(\sum_{s=1}^{m} Q_{st} + R_t \right), \quad Q_{st} = \sum_{e_s=1}^{h_s} Q_{st}^{e_s}, \quad Q_{st}^{e_s} = \sum_{i \in I_{st}^{e_s}} A_i, \quad R_t = \sum_{i \in I_{rt}} A_i, \tag{3}$$

where $I_{st}^{e_s} = I_s^{e_s} \cap I_t$; $I_{rt}=I_r \cap I_t$; I_t is the subset of indexes i for $A_i \in X_t$ with $n_i(r_t) > n_i(r_p)$, $p \neq t$; $I_s^{e_s}$ is the subset of indexes i for $A_i \in X$ with $n_i(q_s^{e_s}) \neq 0$, $n_i(q_v^{e_v})=0$, $v \neq s$, $n_i(r_t)=0$; I_r is the subset of indexes i for $A_i \in X$ with $n_i(r_t) \neq 0$, $n_i(q_s^{e_s})=0$.

The demand to sort objects into two classes is not the principle restriction. Whenever objects are to be classified into more than two classes, it is possible to divide the collection X into two groups, then into subgroups, and so on. For instance, competitive projects can be classified into projects approved and not approved, then the not approved projects can be divided into projects rejected and considered later, and so on.

4 Approximation of Contradictory Classification Rules

The main idea of approximating a large family of contradictory sorting rules with the compact decision algorithm or simple decision rule can be formulated as follows. The pairs of new multisets are generated in the metric space of multisets (X,d) for every group of attributes $Q_1,...,Q_m,R$. The multisets within each pair are to be spaced at the maximum distance, and be mostly coincident with the initial expert sorting of projects in the classes X_a and X_b. Combinations of attributes that define the pairs of the generated multisets produce the generalized decision rule.

Obviously, the decomposition $R=\{R_a,R_b\}$ is the best partition of the project collection $X=\{A_i\}$ into the classes X_a and X_b. The distance between multisets R_a,R_b in the metric space (X,d) is maximum and equal to

$$d(R_a,R_b) = \max d(R_a,R_b) = d^*. \tag{4}$$

In the case of the ideal classification without inconsistencies of individual sorting rules, the maximum distances are equal to $d_0^*=kn$, $d_1^*=1/h$, $d_2^*=1$.

The problem of how to approximate diverse rules for sorting a collection of multiple criteria projects is transformed into the problem of how to find the best binary decompositions $Q_s=\{Q_{sa},Q_{sb}\}$, where multisets Q_{sa},Q_{sb} are to be at the maximum distance in the metric space (X,d). In other words, the following m optimization problems should be solved:

$$d(Q_{sa}, Q_{sb}) \quad \max d(Q_{sa}, Q_{sb}) = d(Q_{sa}^{*}, Q_{sb}^{*}). \tag{5}$$

Here $Q_s^{*}=\{Q_{sa}^{*}, Q_{sb}^{*}\}$ is the best binary decomposition of the multiset Q_s. The solution of each optimization problem (5) can be presented as the sum of submultisets $Q_{st}^{*1}+Q_{st}^{*2}$ $(t=a,b)$. The boundary between submultisets Q_{st}^{*1} and Q_{st}^{*2} is determined by the so-called approximating attribute q_s^{*}. Combinations of approximating attributes q_s^{*} for various numbers s of attribute groups define conditions for assigning the project A_i X into a certain class X_t.

Approximating attributes q_s^{*} can be ordered according to values of distances $d(Q_{sa}^{*}, Q_{sb}^{*})$. Then, attributes q_s^{*}, which occupy first places in this ranking, are to be included in the generalized decision rule. The nearer the distances $d(Q_{sa}^{*}, Q_{sb}^{*})$ to the maximum distance d^{*}, the more accurate the approximation of individual sorting rules. The accuracy of sorting rules approximation can be estimated by the expression

$${}_{s} = d(Q_{sa}^{*}, Q_{sb}^{*})/d^{*}, \tag{6}$$

where the distance d^{*} is determined by (4). Approximating attributes q_s^{*}, which provide the demanded accuracy of approximation ${}_{s} \quad {}_{0}$, are included in the generalized decision rule (1) for project classification.

Relations between the collection of objects $X=\{A_i\}$ and the set of their attributes $G=\{g_j\}$ can be expressed by the matrix $C=||n_i(g_j)||$. The matrix C is often used in data analysis, pattern recognition and called the "object-attribute" table, information table or decision table. In our case, information on properties of the multi-attribute objects A_i and information that the object A_i belongs to a certain decision class can be presented as the decision table C, that has a dimension $k \quad h$, and consists of $2(m+1)$ boxes which correspond to multisets Q_{sa}, Q_{sb} and R_a, R_b (k is a number of objects, h is a number of object attributes, $m+1$ is a number of attributes groups). The reduced decision table $C'=||n_i'(g_j)||$ has a dimension $2 \quad h$, and consists of two rows $n_a'(g_j)$ and $n_b'(g_j)$ which correspond to the classes X_a and X_b.

The procedure for generating the generalized decision rule can be summarized as follows.

1. Compute the decision table $C = ||n_i(g_j)||$ of dimension $k \quad h$, that presents to the collection of objects $X=\{A_i\}$ and consists of $2(m+1)$ boxes.
2. Combine the objects A_i related to the given classes X_a, X_b using the formula (3). Obtain the reduced decision table $C'=||n_i'(g_j)||$ of dimension $2 \quad h$, that corresponds to the specified classes X_a and X_b.
3. Solve the optimization problem (5) for every binary decomposition Q_s and find the approximating attributes q_s^{*} in every s-th box of the reduced matrix C'.
4. Range the approximating attributes q_s^{*} by the values of distances $d(Q_{sa}^{*}, Q_{sb}^{*})$.
5. Select the attributes q_s^{*} that provide the demanded value of approximation accuracy ${}_{s} \quad {}_{0}$. The set of corresponding attributes $\{q_s^{*}\}$ forms the generalized decision rule for sorting the objects.

The proposed method for approximation of contradictory sorting rules was tested on the base of expert decisions related to the State Scientific and Technological Program. The following general decision rules for selecting competitive projects were found:
"The team must be one of the best or have the experience and qualification level sufficient for the project realization" (the estimates q_4^1 or q_4^2; the approximation accuracy $_s$ 0,65);
"The project is to be very important or important for achievement of the major program goals, the team must be one of the best or have the experience, qualification level, and resources sufficient for the project realization" (the estimates q_1^1 or q_1^2, and q_4^1 or q_4^2, and q_5^1 or q_5^2; the approximation accuracy d_s 0,55).
The last decision rule completely coincides with the rule mentioned in [La89].

5 Conclusions

In this paper, we have suggested the tools for classifying a collection of objects represented by many qualitative attributes, when a lot of copies of objects and a set of diverse sorting rules can exist. These tools are based on the theory of multiset metric spaces. The multiset approach allows to discover, present and utilize the available information contained in the object descriptions, to interpret the results of classification and their peculiarities, especially in the case of individual sorting rules and objects' properties inconsistencies. Some of the techniques proposed here were applied to prepare and analyze decisions related to the real-life cases of science management.

This work is supported partially by the Russian Foundation for Basic Research (project N99-01-00476).

Bibliography

[Kn69] Knuth, D.E.: The Art of Computer Programming. Vol.2. Seminumerical Algorithms. Addison-Wesley, Reading, MA, 1969.

[LM97] Larichev, O.I.; Moshkovich, E.M.: Verbal Decision Analysis for Unstructured Problems. Kluwer Academic Publishers, Boston, 1997.

[La89] Larichev, O.I.; Prokhorov, A.S.; Petrovsky, A.B.; Sternin, M.Yu.; Shepelev, G.I.: The Experience of Planning of the Basic Research on the Competitive Base. Vestnik of the USSR Academy of Sciences, 1989,7, pp. 51-61 (in Russian).

[Pe94] Petrovsky, A.B.: An Axiomatic Approach to Metrization of Multiset Space. In (Tzeng, G.H.; Wang, H.F.; Wen, U.P.; Yu, P.L. Eds.): Multiple Criteria Decision Making. Springer-Verlag, New York, 1994; pp. 129-140.

[Pe97] Petrovsky, A.B.: Structuring Techniques in Multiset Spaces. In (Fandel,G.; Gal, T. Eds.): Multiple Criteria Decision Making. Lecture Notes in Economics and Mathematical Systems, N448. Springer-Verlag, Berlin, 1997; pp. 174-184.

[Ya86] Yager, R.R.: On the Theory of Bags. International Journal of General Systems, 1986, 13, pp. 23-37.