p-ISSN 1607-3274 Радіоелектроніка, інформатика, управління. 2024. № 4
e-ISSN 2313-688X Radio Electronics, Computer Science, Control. 2024. № 4

UDC 519.85

SOLUTION OF A MULTICRITERIA ASSIGNMENT PROBLEM USING A CATEGORICAL EFFICIENCY CRITERION

Novozhylova M. V. – Dr. Sc., Professor, Head of the Department of Computer Science and Information Technologies, O. M. Beketov National University of Urban Economy in Kharkiv, Ukraine.

Karpenko M. Yu. – PhD, Associate Professor of the Department of Computer Science and Information Technologies, O. M. Beketov National University of Urban Economy in Kharkiv, Kharkiv, Ukraine.

ABSTRACT

Context. The paper considers a problem of assigning a set of employees to a finite set of operations in a multicriteria statement, under condition of a hierarchical structure of a partial efficiency criterion of performing a set of operations, being presented in such a way that each employee possesses a finite set of competencies and each operation has a finite set of characteristics. Numerical and categorical data types are provided for the use as exogenous parameters of the problem. The relevance of the assignment problem being considered is determined by an extremely wide range of practical applications, both in the classical statements and new modifications, the high demand for which is constantly generated by the dynamically developing economic environment. At the same time, a critically smaller number of scientific publications propose means of modeling and solving multi-criteria assignment problems, despite the importance of this type of problems in decision-making, both in theoretical and practical aspects. In general, in conditions of lack of information, the exogenous parameters of the problem cannot be specified in numerical form, therefore there is a need to use categorical data with further numerical coding.

Objective. The goal of the work is to build a multicriteria mathematical model and, on this basis, carry out a numerical study of the optimization assignment problem, taking into account a hierarchical structure of a partial efficiency criterion of the selection of «operation – employee» pairs.

Method. The study proposes a novel method of solving the assignment problem that implemented as a multi-stage process, which includes the stage of transformation of exogenous parameters of the model, given by categorical variables, based on the implementation of the Pareto principle and logistic mapping, the stage of constructing linear scalarization of the efficiency and the cost criteria.

KEYWORDS: mathematical and computer modeling of an assignment problem, multicriteria optimization, Pareto set, categorical parameters, logistic curve.

ABBREVIATIONS

NP-complexity is an nondeterministic polynomial complexity;

DM is a decision maker.

NOMENCLATURE

i is an index of operation;

j is an index of employee;

k is an index of operation characteristics;

I is a number of operations;

J is a number of employees;

K is a number of operation characteristics;

A is a matrix of weighting coefficients;

B^i is a binary K×3 matrix;

b^i_{kn} is an element of binary matrix B^i;

c_{ij} is a cost of performing the i-th operation by the j-th employee;

C is a cost matrix;

C^H is a normalized cost matrix;

c^{norm}_{ij} is an element of the normalized cost matrix;

$e_{ij}(v_i, s_j)$ is a similarity measure of vectors (v_i, s_j) ;

$c^{\min}$ the minimum value of cost matrix elements;

$c^{\max}$ the maximum value of cost matrix elements;

E is a performance effectivity matrix;

$G(C,X)$ is a cost criterion of the problem;

$f(M)$ is a function of digitalization of categorical variables;

K_E is an effectivity function;

$\hat{K}_E$ is an approximation of an effectivity function;

K_C is a cost function;

$\boldsymbol{M}$ is a set of categorical values**;**

N_1 is a value from the interval (0,1);

N_2 is a value from the interval (0,1), $N_1 < N_2$;

N is a number of labels in dummy coding;

R^2 is a reliability of approximation;

s_j is a vector of competence levels of an j-th employee;

s_{jk} is a level of competence of an j-th employee regarding k-th characteristic of operation;

S is a matrix of levels of employees' competencies;

x_{ij} is a binary endogenous variable of the assignment problem;

X is a matrix of endogenous variables of the assignment problem;

v_i is a vector of operation θ_i characteristics;

v_{ik} is a k-th characteristic of operation θ_i;

w_i is a ordered set of operation θ_i characteristics;

α_i is a assessment vector of operation θ_i characteristics;

α_{ik} is a k-th component of the vector α_i;

Θ is a set of operations;

θ_i is an i-th operation;

Ξ is a team of employees;

ω_j is a j-th employee;

DOI 10.15588/1607-3274-2024-4-7

λ_1, λ_2 is a weighting coefficients of partial criteria in a assignment problem;

$\Phi(E,X)$ is an efficiency criterion to select optimal pairs (θ_i, ω_j);

$\times$ is a cartesian product of sets.

INTRODUCTION

Assignment problems as a scientific direction propose instrumental tools to optimal use of production capacities and resources for the livelihood of consumers that is one of the urgent problems today. To solve this problem is even more important when functioning of enterprises is under pressure of hostile external environment and all the production and economic activities are forced to take place in the conditions of an emergency situation, and the conduct of continued military operations.

Thus assingment problems both in the classical statement and in new modifications belong to the most widespread optimization problems of resource distribution, which arise in business practice, in particular in the urban economy of large cities, as well as when providing services to the population, individuals and legal entities in the field of retail, medical care, hotel business, transport and repair services, supply organization, etc., and the demand for which is constantly generated by life itself.

The obvious practical and theoretical importance of the problem for the development of the theory of algorithms and the theory of optimization in general at least because a quadratic assignment problem belongs to the class of combinatorial optimization problems and it is NP-complex – causes continiuos interest of scientists.

Without losing generality let's consider this problem in the statement about the assignment of employeers having various qualifications for operational, repairing, monitoring activities that relate to one or a group of complex technical or social objects.

Thus, let there be a set $\Theta = \{\theta_i\}$, $i=1,...,I$, of certain operations (objects) and a team (set) $\Xi = \{\omega_j\}$, $j=1,...,J$, of employers (resources), in general $I \neq J$.

We need to assign at least one operation to each j-th employee, and only one employee to each i-th operation. For each such pair (θ_i, ω_j) the cost c_{ij} of performing an operation θ_i by an employer ω_j is determined. According to the classical formulation, the objective function models the total costs. It is necessary to find the minimum value of objective function.

Assignment problems admit of several generalizations. We will further consider multicriteria problems as a natural generalization of the classical statement. In multicriteria problems, a partial cost criterion is considered along with a partial criterion of the effectivity of operation performance [1].

Thus, along with the cost of performing operations, the effectivity function of performing operations is also considered, which, depending on the specific statement, can be both a dimensionless and a dimensional value (monetary equivalent, quantity of products, etc.).

However, when solving practical problems, situations arise when the effetivity functional of the pair (θ_i, ω_j) is a vector and not a scalar. In addition, the exogenous parameters of the effectivity function can be specified as categorical characteristics in case of uncertainty or lack of input information.

In this study, the multicriteria assignment problem is considered, provided that in addition to the cost c_{ij} of assignment j-th employee for a certain operation θ_i, a concept of categorical vector that measures effectivity E_{ij} to perform i-th operation by j-th employee is proposed.

The object of the study is the process of solving a multi-criteria assignment problem in conditions of lack of input information.

The subject of the study is an optimization method for solving a multi-criteria assignment problem with categorical exogenous parameters.

The purpose of this work is to build a mathematical model and, on this basis, to carry out a numerical study of the optimization assignment problem taking into account matrix representation of the effectivity criterion to select optimal pairs (θ_i, ω_j) on the Cartesian product of sets $\Theta \times \Xi$.

To achieve this goal, the following tasks have been solved further:

– to develop and substantiate the methodology of numerical coding of categorical data;

– to determine a means of reducing a multicriteria problem to a set of scalar optimization problems;

– to carry out the software implementation of the model based on the creation of a software simulator, and to build the Pareto set of the assignment problem;

– to justify the possibility of generalizing the proposed tools to other classes of assignment problems.

We chose Python as the working programming language, which, thanks to the availability of numerous scientific libraries and cloud computing tools, allows to create the necessary tools for the implementation of numerical experiments.

1 PROBLEM STATEMENT

Let for each operation θ_i of the set $\Theta = \{\theta_i\}$, $i=1,...,I$, be given an ordered set

$$v_i = (v_{i1}, v_{i2},\ldots,v_{ik},\ldots, v_{iK}),\ i=1,...,I,$$

containing assessments of operation characteristics and each employee ω_j of team $\Xi = \{\omega_j\}$, $j=1,...,J$ owns an ordered set of values

$$s_j = (s_{j1}, s_{j2},\ldots, s_{jk},\ldots, s_{jK}),\ j=1,...,J,$$

of competence levels, respectively.

Moreover, each value v_{ik} of i-th operation characteristics v_i as well as value s_{jk} of j-th employee competence level s_j is defined as a categorical parameter.

For each pair (θ_i, ω_j), defined on the Cartesian product of sets $\Theta \times \Xi$, the cost c_{ij} of performing an

DOI 10.15588/1607-3274-2024-4-7

operation θ_i by an employer ω_j has been set. So, as input data, we also have a matrix of values

$$C= ||c_{ij}||_{i=1,...,I, j=1,...,J}.$$

As the endogenous variables of the assignment problem under consideration, we will take the binary matrix $X= ||x_{ij}||_{i=1,...,I, j=1,...,J}$, where x_{ij} is the sign of the assignment: x_{ij} = 1 of j-th employee for the i-th operation and x_{ij} = 0 otherwise.

Then the assignment problem in the two-criteria statement is to define the optimal values of two performance criteria, namely the cost of execution $G(C,X)$ and the effectivity of execution $\Phi(E,X)$ of operations:

$$\{ G(C,X) \to \min, \Phi(E,X) \to \max \} \quad (1)$$

subject to

$$\sum_{j=1}^{J} x_{ij} \le 1, i = 1,...,I;$$
$$\sum_{i=1}^{I} x_{ij} = 1, j = 1,...,J; \quad (2)$$
$$x_{ij} \in \{0,1\}, i=1,...,I, j=1,...,J.$$

The cost $G(C,X)$ of performing operations of the set Θ is given by the formula

$$G(E.X) = \sum_{i=1}^{I} \sum_{j=1}^{J} c_{ij} x_{ij} .$$

The efficiency $\Phi(E,X)$ is given as follows

$$\Phi(E,X) = \sum_{i=1}^{I} \sum_{j=1}^{J} e_{ij} x_{ij} ,$$

where the elements $e_{ij}(v_i, s_j)$ of the effectivity matrix E

$$E = || e_{ij}(v_i, s_j) ||_{i=1,...,I, j=1,...,J}$$

are functional values defining similarity of the vectors (v_i, s_j).

In other words, functional $e_{ii}(v_i, s_i)$ represents numerical measure of how well j-th employee fits to i-th operation.

2 REVIEW OF THE LITERATURE

Various approaches to solve both deterministic assignment problems and problems under conditions of uncertainty in scalar and multicriteria formulations are proposed by researchers in [1–5] and other studies. At the same time, not only the namely assignment problems have been considered, but also derived classes of problems, for example, the linear problem of ranking alternatives [2]. The total cost and quality of work, the number of performers and the time of work are partial objective functions. An interesting result is the formulation of the assignment problem in fuzzy set terms [1].

E. Acar and H. S. Aplak [4] examine the problem of dynamic distribution of resources for the service provision on the set of locations in urban areas. Service requests occur spontaneously, resources are distributed dynamically. Real-world examples of such applications include dispatching traffic police officers to accident sites and dispatching mechanics to repair sites. It is proposed to use the so-called policy gradient approach, according to which performers are distributed by location in order to minimize the delay.

It is generally not possible to describe exogenous parameters of the mathematical model in numerical form for practical assignment problems. Therefore, certain characteristics of the research objects are presented in a categorical form. Sets of values containing a limited number of separate categories are called categorical (qualitative) ones, while the possible values are parameter levels as it proposed by Kondruk N. E. [6]. At the same time, there is a problem of developing effective distance metrics and measures of similarity for such features, in particular categorical ordered features.

When choosing an approach to the digitization of categorical data from the set of assignment pronlem (1) – (2) input data, a comparative analysis of available classical methods of numerical coding [7] was carried out with the following results:

1. Uniform coding (One-Hot coding) is a process of converting categorical variables into a numerical representation in a binary format, where each category is represented by a binary vector.

Applying the concept of uniform coding to a vector of categorical variables $v_i = (v_{i1}, v_{i2}, v_{iK})$ generates a binary matrix $B^i = \{b^i_{kn}\}_{k=1,...,K, n=1,...,3}$, where the k-th row of the matrix B^i corresponds to the categorical variable v_{ik} and contains only one unit, that is, it satisfies the condition

$$\sum_{i=1}^{I} b^i_{kn} = 1, b^i_{kn} \in \{0,1\}.$$

2. Dummy Encoding is a slight improvement over uniform encoding, which uses N–1 functions to represent N labels/categories.

3. Ordinal Encoding is used when the categories in the variable have a natural order. In this method, categories are assigned a numerical value based on their order, such as 1, 2, 3, etc.

4. Binary Encoding is similar to One-Hot encoding, but instead of creating a separate column for each category, the latter is represented as binary number.

5. Count Encoding is a method of coding categorical variables by counting the frequency of occurrence of a category in a data set.

6. Target Encoding is a more advanced encoding method used to work with categorical features of high

DOI 10.15588/1607-3274-2024-4-7

power, that is, features with many unique categories. An average target value for each category is calculated, which in turn is used to replace the categorical feature.

Given the conditions of Proposition 1, the most adequate for the case of ordered categorical variables among those mentioned are ordinal or target coding methods.

The Python implementation of these methods, allowing encoding categorical data, is performed using the LabelEncoder and OneHotEncoder encoders from SciKit Learn scientific library [8].

Solving multicriteria assignment problems does not guarantee the generation of a single, best solution, therefore it is important to create and use software tools for constructing a set of Pareto-optimal solutions [9, 10] of problems of this class and conducting numerical experiments.

Currently, such tools are offered as part of a wide range of software products – so-called solvers, i.e. hardware and software for solving complex optimization problems. According to the set of conditions, solvers go through the solution options and choose the best one. They are convenient modern means for creating an information environment to model and solve assignment problems in various fields: in logistics when building the optimal delivery routes and calculating the optimal loading of vehicles; in the retail, medical service, hotel business when generating employee work shift schedules; in industry when solving the ussies of optimizing product warehouses, taking into account the available raw materials.

The market of software products includes commercial developments by IBM or Gurobi, as well as Open Source solvers: OR-Tools, COIN-OR, SCIP, GLPK and others [11–14]. Among the latter, the most developed is the Google OR-Tools solver [11], which is used in the future when creating a software simulator.

3 MATERIALS AND METHODS

The statement of the multi-riteria assignment problem (1)–(2) contains numerical (in particular, binary) and categorical types of parameters.

Therefore, at the first stage of processing the mathematical model (1) – (2), there is a need to unify its exogenous parameters.

Let us assume that the characteristics of a certain operation θ_i constitute a vector of ordered categorical multi-positional variables $v_i = (v_{i1}, v_{i2},\ldots,v_{ik},\ldots, v_{iK})$, where each variable v_{ik} takes a value on a set of M categorical values

$$M = \{\text{“low”}, \text{“medium”}, \text{“high”}\}.$$

Similar representation has been introduced for vectors s_j.

Taking into account the structure of categorical parameters here we propose a novel approach to the digitization of categorical data different from the well-known ones discussed in the previous section.

The approarch involves constructing a function $f(M)$ of the form:

$$M \xrightarrow{f} [0,1], \tag{3}$$

defined on the set M. The range of the function $f(M)$ is a unit segment.

A method of digitalization of categorical parameters being proposed in this study is based on the implementation of the well-known Pareto principle [15] and functional-cost analysis [16]. The Pareto principle is one of the universal principles of optimization, whose informal definition is as follows: “20% of the effort provides the opportunity to obtain 80% of the result, and the rest – 80% of the effort – gives only 20% of the result” [17]. In general, an 80/20 ratio often occurs, but does not necessarily have to be exactly 80/20. It can be 90/10, 70/30, etc., which also confirms the observation formulated above.

First of all, we will apply interval estimates of categorical variables, considering the coding problem as continuous. Categorical values must then be encoded by numerical half-open intervals of the form

$$\begin{array}{l} M = \\ \{\text{“low”}, \quad \text{“medium”}, \quad \text{“high”}\}. \\ \quad\;\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow \\ \{\;[0, N_1), \quad [N_1, N_2), \quad [N_2, 1]\;\} \end{array} \tag{4}$$

We apply the Pareto principle to determine the distribution of the employees’ competence levels and construct the function $f(M)$ of the form (3).

Proposition 1. Based on the Pareto principle, we consider the following distribution of the employees’ competence levels:

$$[0, 0.2), [0.2, 0.8), [0.8,1]. \tag{5}$$

Thus, we will assume (Table 1) that the «high» level of a competence s_{jk} is estimated by the segment [0.8, 1] and means the immediate readiness of proven successful application of knowledge and skills in practice.

The “medium” level of competence means the readiness to apply existing knowledge and skills along with the need for additional training that does not have a critical impact on the quality and time of operation performance, and finally, the “low” level designates the absence or insufficient availability of knowledge that requires additional specialized training.

Table 1 – Interval coding of ordered categorical data

№ of category	Categorical value	Numerical value
1	low	0… < 0.2
2	medium	0.2...< 0.8
3	high	0.8...1

DOI 10.15588/1607-3274-2024-4-7

Proposition 2. The distribution (3) for the digitization of the characteristics v_{ik} which are inherent to i-th operation from the set Θ can be explained as follows: the "high" level of the characteristic v_{ik} as a resource of the operation is estimated by the segment [0.8, 1] and means the critical importance of the presence of this resource, the "medium" level means an optional requirement that does not have a critical impact on the execution time of the operation with acceptable quality and, finally, "low" level makes it clear that the presence of this resource is desirable to support the quality assurance of the operation.

In favor of this hypothesis there is the ideology of functional-cost analysis [16]. Within the framework of this paradigm we know that a qualitative dependence of effectivity K_E on cost K_C is described the same way for many technical, economic, and social processes.

Cost K_C can be perceived, for example, as the time or cost of retraining and upgrading the skills of employeers.

Such a dependence is well described by a logistic (S-shaped) curve of a kind (Fig. 1):

$$K_E = \frac{1}{1+\exp(-K_c)}. \quad (6)$$

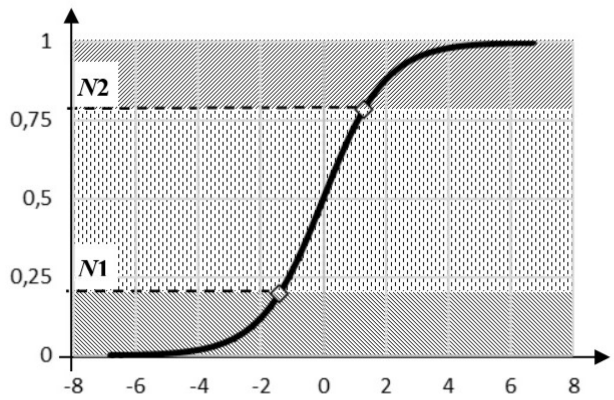

Figure 1 – Decomposition of the logistic curve "effectivity K_E – cost K_C"

Fig. 1 shows a graphical presentation of dependence (6) (to within its own affine transformation) together with a horizontal decomposition of the logistic curve into three segments belonging to

– Zone I : $[0, N_1]$,
– Zone II : $[N_1, N_2]$,
– Zone III : $[N_2, 1]$, $N_1 < N_2$.

Table 1 contains the results concerning approximation of the logistic curve parts belonging to Zones I – III.

Table 1 – Approximation of logistic curve

№ of Zone	Type of approximation dependence $\hat{K}_E$	R^2
I	$0.0074\hat{K}_c^2 - 0.0555\hat{K}_c + 0.093$	0.934
II	$0.2261\, K_C - 1.76$	0.998
III	$0.0087\hat{K}_c^2 - 0.2768\hat{K}_c + 1.201$	0.943

Thus, estimates of the reliability R^2 of the approximating dependence $\hat{K}_E$ confirm the use of such a decomposition. At the same time, horizontal boundary lines of the zones I–III (Fig. 1) practically coincide with the boundaries of the coding ranges (5) of ordered categorical data: $N_1 \approx 0.198$, $N_2 \approx 0.785$.

In this connection, it is interesting to turn to the analysis of the second derivative of the logistic function (6), which has the form:

$$\frac{d^2 K_E}{dK_c{}^2} = \frac{2\exp(-2K_c) - \exp(-K_c)}{(1+\exp(-K_c))^3}.$$

On the graph of the second derivative of the logistic function (Fig. 2), two extrema can be distinguished, the arguments of which coincide with the border values N_1 and N_2.

So, in terms of the main research problem, it can be interpreted as follows: the improvement of the employee's qualification, that is, the improvement of his efficiency, within the limits of a low level, occurs quickly at insignificant costs. We can see that the second derivative is positive and increases, that is, the growth rate of function increases.

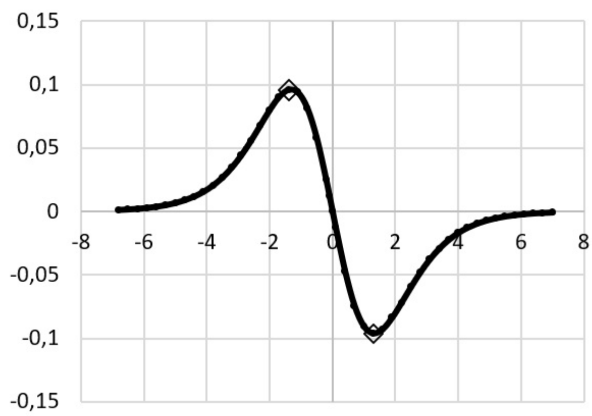

Figure 2 – Second derivative of the logistic curve "effectivity K_E – cost K_C"

Within the average level, the second derivative linearly decreases, so that the growth rate of the efficiency function decreases.

This means that more costs are required to improve efficiency within the medium range.

DOI 10.15588/1607-3274-2024-4-7

In the zone of a high level, the second derivative of the logistic function (6) from negative values approaches 0, that is, the rate of change of the function becomes constant.

The type of approximation dependence (Table 1) is convenient to use in the case of the need to implement continuous numerical coding of categorical data.

Let us investigate a scalarization procedure for of partial efficiency criteria, being considered.

In the problem under consideration, the functional $\Phi(E, X)$ of operational efficiency is dimensionless, while the cost $G(C,X)$ of operations according to the statement is a dimensional parameter. Therefore, when choosing a scalarization of partial criteria further, we will normalize the cost matrix C and obtain a normalized matrix $C^H = \| c_{ij}^{norm} \|$ according to the formula

$$c_{ij}^{norm} = \frac{c_{ij} - c^{\min}}{c^{\max} - c^{\min}}.$$

Carrying out numerical coding of categorical characteristics $v_i = (v_{i1}, v_{i2},\ldots,v_{ik},\ldots, v_{iK})$ of operations and levels of competences $s_i = (s_{i1}, s_{i2},\ldots, s_{ik},\ldots, s_{iK})$ of employers according to the scale of Table 1 makes it possible to consider the formalization of the following observation. In practice, certain properties of some operation are more important for its development than others, regardless of the level of qualification features or, in general, the quantity and quality of the necessary resources.

Proposition 3. Let us assume that the characteristics of each i-th operation make up the following vector $v_i = (v_{i1}, v_{i2}, v_{iK})$, which assumes a partial or complete ordering by its importance:

$$v_i \xrightarrow{\alpha_i} w_i\,,$$

where the arguments of the vector w_i are the components of the vector v_i, weighted by the components of the vector of preferences

$$\alpha_i=\{\alpha_{i1}, \alpha_{i2},\ldots \alpha_{iK}\}.$$

As an initial approximation for each i-th operation, a vector of preferences $\alpha_i=\{\alpha_{i1}, \alpha_{i2},\ldots \alpha_{iK}\}$. can be considered, which specifies the ordering of the operation's characteristics by importance and is calculated according to the formulas

$$\alpha_{ik} \in \left\{ m\left(\sum_{k=1}^{K} k\right)^{-1} \right\}_{m=1,2,\ldots,K},\ \sum_{k=1}^{K} \alpha_{ik} = 1, i = 1,2,\ldots,I.$$

In total, the vectors of preferences form a matrix $A=\{\alpha_{ik}\}_{i=1,2,\ldots,I,\ k=1,2,\ldots,K}$ of weighting factors for a set of operations Θ. Accordingly, the vectors $s_j = (s_{j1}, s_{j2}, \ldots, s_{jK})$ of the employers' competencies are also ordered using the matrix A, forming the vectors $r_j= (\alpha_{i1}r_{i1}, \alpha_{i2}r_{i2},\ldots \alpha_{iK}r_{iK}).$).

Convolution of this information can be done by several means. This study develops an approach in which the efficiency e_{ij} is estimated by the value of the cosine of the angle ϕ between K-dimensional vectors v_i and s_j using the scalar product $\langle v_i, s_j \rangle$ of these vectors of the form:

$$e_{ij} = \cos\phi_{ij} = \frac{\langle v_i, s_j \rangle}{|v_i||s_j|}.$$

Thus, as a result of the performed transformations, it is possible to apply a scalarization of the partial criteria $G(C,X)$ and $\Phi(E,X)$.

In so doing we find that the mathematical model of the multicriteria assignment problem being considered takes the form

$$\max\left(\lambda_1 \sum_{i=1}^{I} \sum_{j=1}^{J} e_{ij} x_{ij} - \lambda_2 \sum_{i=1}^{I} \sum_{j=1}^{J} c_{ij}^{norm} x_{ij} \right) \quad (7)$$

subject to restrictions (2) and additional conditions regarding weighting factors λ_1, λ_2:

$$\lambda_1, \lambda_2 \geq 0,\ \lambda_1 + \lambda_2 = 1.$$

4 EXPERIMENTS

To conduct numerical experiments based on the obtained mathematical model of the multicriteria assignment problem, such software development tools as Google Colab [18] and Google OR-tools [11] were involved in the work (Fig. 3).

```
# Importing necessary libraries
# Google OR-tools and Python
from ortools.linear_solver import
pywraplp
import numpy as np
import matplotlib.pyplot as plt
import random
from numpy.linalg import norm
```

Figure 3 – Application of Google OR-Tools libraries and scientific Python libraries for modeling

Google OR-Tools is a set of open source software (Apache 2.0 license) developed by Google for solving optimization problems, including mathematical programming, vehicle routing, assignment problems, and others. Google OR-Tools contains a set of ready-made components written in such well-known programming languages as C++, Java, .NET and Python

The assignment problem was solved using the MIPSolver shell, which is a component of OR-Tools,

DOI 10.15588/1607-3274-2024-4-7

designed for linear programming and mixed integer programming (MIP) problems. MIPSolver interacts with the SCIP tool, a platform for solving integer problems with constraints (CIP) and mixed integer nonlinear problems (Fig. 4)

```
# Creating MIP solver, that calls SCIP
# backend
solver =
pywraplp.Solver.CreateSolver("SCIP")
```

Figure 4 – Call a MIPSolver

The stages of numerical modeling and solving the assignment problem in the Google Colab cloud environment using OR-Tools are presented in Fig. 5.

The experiments were conducted on a set of input data $\{I, J, K, C, S\}$ (Block 1 in Fig. 5).

At the same time, at the choice of the DM, the parameters of the dimensionality of the problem and the matrices of efficiency E and cost C can be randomly set by calling the Python functions random.randint and random.random() or read from a data file.

Blocks 2 and 3 of the information model being described are preparatory for calling required functions from OR-Tools arsenal.

Blocks 4–6 serve to describe problem independent (endogenous) variables, constraints, the specification of linear scalarization of two partial criteria of the original problem: efficiency criteria and the cost ones, after that and the transfer of this information is carried out to MIPSolver. At the same time, we consider that the relation $I \geq J$ is fulfilled.

Further, the function solver.Solve() call (Block 7) implements the process of solving two-criterion assignment problem with the given values of the weighting coefficients λ_1 and λ_2.

5 RESULTS

The last stage of modeling (Block 8) is intended to construct and analyse a set of Pareto-optimal solutions of the two-criteria assignment problem under consideration.

The set of Pareto-optimal solutions, which is the result of solving the two-criteria assignment problem for definite set of exogenous parameters $\{I, J, K, C, S\}$, is constructed in the space of partial criteria based on the variation of their weighting coefficients λ_1 and λ_2 accordingly.

Thus, the procedure for numerical modeling of the assignment problem necessarily involves the visualization of a set of Pareto-optimal solutions in the space of partial target criteria, an example of which is shown in Fig. 6.

Block 1
```
# Setting input data -
# exogenous parameters:
num_workers #J
num_tasks #I
num_charact #K
efficiency_workers #S[][]
charact_tasks #V[][]
cost_workers #C[][]
```

Block 2
```
# Standardization of the
# cost matrix: cost_workers
```

Block 3
```
# Digitization of exogenous
# categorical parameters:y_dot[][]
```

Block 4
```
# Problem endogenous variables
x[i,j] = solver.IntVar(0, 1, "")
```

Block 5
```
# Problem constraints
solver.Add(solver.Sum([x[i, j]
for j in range(num_tasks)]) <= 1)
solver.Add(solver.Sum([x[i, j]
for i in range(num_workers)]) == 1)
```

Block 6
```
# Objective functional
# Weight1: λ1, Weight1: λ2
objective_terms.append((Weight1*
y_dot[][] - Weight2*cost_workers[][])*
x[][])
```

Block 7
```
# Solving the problem
status = solver.Solve()
```

Block 8
```
# Analysis of Pareto-optimal
# solutions
# Definining "indifference" ranges
# of partial objective criteria
```

Figure 5 – Numerical modeling of the assignment problem in the Google OR-Tools environment

The analysis of a set of Pareto-optimal solutions enables the DM to make flexible decisions regarding the final choice.

Numerous numerical experiments with various data sets provide the basis for the following constructive remarks.

Remark 1. Values of partial effectivity and cost criteria are unchanged (indifferent) on certain ranges of values of weight coefficients λ_1, λ_2, despite different values of the generalized criterion (7).

DOI 10.15588/1607-3274-2024-4-7

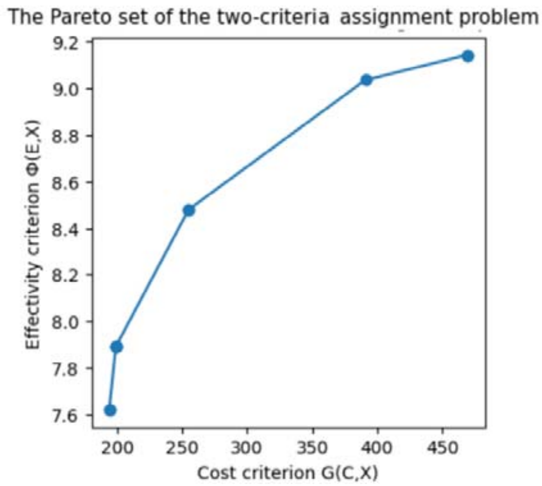

Figure 6 – Construction of the Pareto set for one of sets {*I*, *J*, *K*, *C*, *S*} of exogenous parameters

This result is statistically significant, because it was observed in a critically large part of the numerical experiments. An illustration of this conclusion is shown in the Fig. 6, which provides an example of solving a certain assignment problem, where the ranges for determining the weighting factors λ_1, λ_2, contained 10 values each. And only five pairs of them generated different results. In other words, sets of λ_1, λ_2 values {(0;1), (0.1;0.9)}, {(0.2;0.8)}, {(0.3;0.7),(0.4;0.6)}, {(0.5;0.5), (0.6;0.4), (0.7;0.3), (0.8;0.2)} {(0.9;0.1), (1;0)}, enclosed in curly braces, generate the same results respectively. In Fig. 6 different values pairs of partial criteria are shown by dots (markers) – generated different results. The solid line in Fig.6 represents an approximate shape of the Pareto set.

Remark 2. Further processing of information regarding the set of Pareto-optimal solutions obtained allows to define set of ranges, which are characterized by relative indifference of partial criteria values. In so doing the DM can apply of a certain concession on the values of the partial criteria which he considers as acceptable.

6 DISCUSSIONS

The proposed mathematical model of the two-criterion assignment problem, unlike the classical formulation, takes into account the possibility of specifying exogenous parameters with categorical values, which expands the range of model applications. Introducing logistic mapping into consideration at the stage of digitization of categorical variables, in contrast to known methods, the analysis of which is carried out in the section “Review of the literature”, allows for finer adjustment to the specifics of the problem and, if necessary, to apply elements of continuous analysis.

The proposed algorithm and software form a simple flexible and reproducible environment for the numerical study of multi-criteria assignment problems. The presented calculation results confirm the validity and efficiency of the developed mathematical model and solution method for multi-criteria assignment problem being considered.

7 CONCLUSIONS

Modeling and numerical research of the multi-criteria assignment problem was carried out under the condition of vector presentation of the efficiency criterion and the presence of categorical exogenous parameters.

The scientific novelty.

We proposed a novel method of solving the assignment problem that implemented as a multi-stage process, which includes the stage of transformation of exogenous parameters of the model, given by categorical variables, based on the implementation of the Pareto principle and logistic mapping, the stage of constructing linear scalarization of the efficiency and the cost criteria.

A generalization of the multi-criteria assignment problem has been developed, where the partial criterion for the performance of *i*-th operation by *j*-th employee is defined as the scalar product of the digitized competences of the employee and the characteristics of the operation.

The practical significance.

We applied cloud software development tools of Google Colab and Google OR TOOLS as well as Python scientific libraries to conduct numerical experiments on the basis of the mathematical model proposed for the main research problem.

Numerical experiments conducted on multiple sets of input data using this optimization mathematical model showed that the proposed constructive means of mathematical modeling and information environment developed in this study provide flexible methodology for researching other types of optimization assignment problems.

Prospects for further research consist in developing the proposed tools in two directions. First, it is an adaptation of the proposed model by considering exogenous parameters of assignmet efectivity as functions of time and/or cost. The second direction of research is in developing the solution method to reach refined approximation of the set of Pareto-optimal solutions for considered multi-criteria assignment problem.

ACKNOWLEDGMENTS

The study was performed within the framework of the state-budget research “City’s information infrastructure modeling and optimizatione” of O.M. Beketov National University of Urban Economy in Kharkiv (state registration number 0121U109215).

REFERENCES

1. Song M., Cheng L. Solving the reliability-oriented generalized assignment problem by Lagrangian relaxation and Alternating Direction Method of Multipliers, *Expert Systems with Applications*, 2022, Vol. 205, P. 117644. DOI: 10.1016/j.eswa.2022.117644
2. Mehlawat M. K., Gupta P., Pedrycz W. A New Possibilistic Optimization Model for Multiple Criteria Assignment Problem, *IEEE Transactions on Fuzzy Systems*, 2018, Vol. 26, № 4, pp. 1775–1788. DOI: 10.1109/TFUZZ.2017.2751006
3. Borovicka A. Possible Modifications of the Multiple Criteria Assignment Method, *Czech Economic Review,* 2013, Vol. 7, pp. 55–67.

DOI 10.15588/1607-3274-2024-4-7

4. Acar E., Aplak H. S. A Model Proposal for a Multi-Objective and Multi-Criteria Vehicle Assignment Problem: An Application for a Security Organization, *Mathematical and Computational Applications*, 2016, Vol. 21(4), P. 39. DOI: 10.3390/mca21040039
5. Yan Y., Deng Y., Cui S. et al. A policy gradient approach to solving dynamic assignment problem for on-site service delivery, *Transportation Research Part E: Logistics and Transportation Review,* 2023, Vol. 178, P. 103260 DOI: 10.1016/j.tre.2023.103260
6. Kondruk N. E. Methods for determining similarity of categorical ordered data, *Radio Electronics, Computer Science, Control*, 2023, № 2, pp. 31–36. DOI: 10.15588/1607-3274-2023-2-4.
7. Dahouda M. K., Joe I. A Deep-Learned Embedding Technique for Categorical Features Encoding, *IEEE Access,* 2021, Vol. 9, pp. 114381–114391. DOI: 10.1109/ACCESS.2021.3104357.
8. Scikit-learn 1.4.2 [Electronic resource]. Access mode: https://pypi.org/project/scikit-learn/
9. Dosantos P. S., Bouchet A., Mariñas-Collado I. et al. OPSBC: A method to sort Pareto-optimal sets of solutions in multi-objective problems, *Expert Systems With Applications*, 2024, Vol. 250, P. 123803. DOI: 10.1016ps://doi.org//j.eswa.2024.123803.
10. Makarova V.V. Pryntsyp Pareto v konteksti orhanizatsii staloho ahrarnoho zemlekorystuvannia, *Pryazovskyi ekonomichnyi visnyk,* 2020. Vyp. 1(18), pp. 220–224. DOI: https://doi.org/10.32840/2522-4263/2020-1-39
11. Route. Schedule. Plan. Assign. Pack. Solve. OR-Tools is fast and portable software for combinatorial optimization [Electronic resource]. Access mode: https://developers.google.com/optimization.
12. COIN-OR [Electronic resource]. Access mode: https://www.coin-or.org/
13. SCIP. Solving Constraint Integer Programs. [Electronic resource]. – Access mode: https://www.scipopt.org/
14. GLPK (GNU Linear Programming Kit) [Electronic resource]. – Access mode: https://www.gnu.org/software/glpk/
15. Koh R. Princip 80/20 ta 92 fundamental'nih zakoni prirodi. Nauka uspihu. Kiyv, Vidavnichij centr "KM-BUKS", 2019, 360 p.
16. Petrov E. H., Novozhylova M. V., Hrebennik I. V. Metody i zasoby pryiniattia rishen u sotsialno-ekonomichnykh systemakh. Kyiv, Tekhnika, 2003, 240 p.
17. Blackman I., Chan E. Pareto principle plus statistic methodology in establishing a cost-estimating model, *International CIB World Building: 19th Congress, Brisbane, 05–09 May 2013: proceedings*. Brisbane, Queensland University of Technology, 2013, pp. 1–14.
18. Google-Colab [Electronic resource]. Access mode:https://colab.research.google.com/?utm_source=scs-index.

Received 15.05.2024.
Accepted 09.10.2024.

УДК 519.85

РОЗВ'ЯЗАННЯ БАГАТОКРИТЕРІАЛЬНОЇ ЗАДАЧІ ПРО ПРИЗНАЧЕННЯ ЗА УМОВИ КАТЕГОРІАЛЬНОГО КРИТЕРІЮ ЕФЕКТИВНОСТІ

Новожилова М. В. – д-р фіз.-мат. наук, професор, завідувач кафедри комп'ютерних наук та інформаційних технологій, Харківський національний університет міського господарства імені О. М. Бекетова, Україна.

Карпенко М. Ю. – канд. техн. наук, доцент, доцент кафедри комп'ютерних наук та інформаційних технологій, Харківський національний університет міського господарства імені О. М. Бекетова, Україна.

АНОТАЦІЯ

Актуальність. В роботі розглядається задача про призначення множини виконавців на скінчений набір операцій в багатокритеріальній постановці за умови ієрархічної структури частинного критерію ефективності виконання набору операцій, поданого таким чином, що кожний виконавець володіє скінченим набором компетентностей і кожна операція має скінчений набір характеристик. В якості екзогенних параметрів задачі передбачено застосування числових та категоріальних типів даних. Актуальність задачі про призначення, що розглядається, обумовлюється надзвичайно широким спектром практичних застосувань як в класичній постановці, так і нових модифікаціях, запит на які постійно генерується економічним середовищем, що динамічно розвивається. При цьому критично менша кількість наукових публікацій присвячена засобам моделювання та розв'язання саме багатокритеріальних задач про призначення, незважаючи на важливість задач такого типу в прийнятті рішень, як в теоретичному, так і в практичному аспектах. В загальному випадку в умовах нестачі інформації екзогенні параметри задачі не можуть бути задані в числовій формі, тому виникає потреба застосування категоріальних даних з подальшим числовим кодуванням.

Метою роботи є побудова математичної моделі та проведення на цій основі чисельного дослідження оптимізаційної задачі про призначення з урахуванням можливості ієрархічної структури частинного критерію ефективності вибору пар «операція – виконавець».

Метод розв'язання задачі – це багатоетапний процес, що включає етап трансформації екзогенних параметрів моделі, заданих категоріальними змінними, на основі втілення принципу Парето та логістичного відображення, етап побудови лінійної згортки частинних критеріїв ефективності та вартості виконання робіт.

Результати роботи. Проведено системологічний аналіз існуючих підходів та запропоновано методику числового кодування категоріальних даних. Обгрунтовано засіб зведення багатокритеріальної за постановкою задачі до набору однокритеріальних оптимізаційних задач на основі лінійної згортки та функціонально-вартісного аналізу, побудовано множину Парето основної задачі дослідження.

Висновки. Проведене моделювання та розв'язання задачі про призначення на основі створення програмного симулятора із застосуванням солверу Google OR-Tools підтвердило можливість узагальнення запропонованих інструментальних засобів на інші класи задач про призначення.

КЛЮЧОВІ СЛОВА: математичне та комп'ютерне моделювання задачі про призначення, багатокритеріальна оптимізація, множина Парето, категоріальні параметри, логістична крива.

DOI 10.15588/1607-3274-2024-4-7

ЛІТЕРАТУРА

1. Song M., Solving the reliability-oriented generalized assignment problem by Lagrangian relaxation and Alternating Direction Method of Multipliers / M. Song, L. Cheng // Expert Systems with Applications. – 2022. – Vol. 205. – P. 117644. DOI: 10.1016/j.eswa.2022.117644
2. Mehlawat M. K. A New Possibilistic Optimization Model for Multiple Criteria Assignment Problem / M. K. Mehlawat, P. Gupta, W. Pedrycz // IEEE Transactions on Fuzzy Systems. – 2018. – Vol. 26, №4. – P. 1775–1788. DOI: 10.1109/TFUZZ.2017.2751006
3. Borovicka A. Possible Modifications of the Multiple Criteria Assignment Method / A. Borovicka // Czech Economic Review.– 2013. – Vol. 7. – P. 55–67.
4. Acar E. A Model Proposal for a Multi-Objective and Multi-Criteria Vehicle Assignment Problem: An Application for a Security Organization / E. Acar, H. S. Aplak // Mathematical and Computational Applications. – 2016. – Vol. 21(4), P. 39. DOI: 10.3390/mca21040039
5. A policy gradient approach to solving dynamic assignment problem for on-site service delivery / [Y. Yan, Y. Deng, S. Cui et al.] // Transportation Research Part E: Logistics and Transportation Review. – 2023. – Vol. 178, P. 103260. DOI: 10.1016/j.tre.2023.103260
6. Kondruk N. E. Methods for determining similarity of categorical ordered data / N. E. Kondruk // Radio Electronics, Computer Science, Control. – 2023. – № 2. – P. 31–36. DOI: 10.15588/1607-3274-2023-2-4.
7. Dahouda M. K. A Deep-Learned Embedding Technique for Categorical Features Encoding / M. K. Dahouda, I. Joe // IEEE Access. – 2021. – Vol. 9. – P. 114381–114391. DOI: 10.1109/ACCESS.2021.3104357.
8. Scikit-Learn 1.4.2 [Electronic resource]. – Access mode: https://pypi.org/project/scikit-learn/
9. OPSBC: A method to sort Pareto-optimal sets of solutions in multi-objective problems / [P. S. Dosantos, A. Bouchet, I. Mariñas-Collado et al.] // Expert Systems With Applications. – 2024. – Vol. 250. – P. 123803 DOI: https://doi.org/10.1016/j.eswa.2024.123803.
10. Макарова В.В. Принцип Парето в контексті організації сталого аграрного землекористування / В. В. Макарова // Приазовський економічний вісник. – 2020. – Вип. 1(18). – С. 220–224. DOI: https://doi.org/10.32840/2522-4263/2020-1-39.
11. Route. Schedule. Plan. Assign. Pack. Solve. OR-Tools is fast and portable software for combinatorial optimization [Electronic resource]. – Access mode: https://developers.google.com/optimization.
12. COIN-OR [Electronic resource]. – Access mode: https://www.coin-or.org/
13. SCIP. Solving Constraint Integer Programs. [Electronic resource]. – Access mode: https://www.scipopt.org/
14. GLPK (GNU Linear Programming Kit) [Electronic resource]. – Access mode: https://www.gnu.org/software/glpk/
15. Кох Р. Принцип 80/20 та 92 фундаментальних закони природи. Наука успіху / Р. Кох. – Київ : Видавничий центр «КМ-БУКС», 2019. – 360 с.
16. Петров Е. Г. Методи і засоби прийняття рішень у соціально-економічних системах / Е. Г. Петров, М. В. Новожилова, І. В. Гребеннік. – Київ : Техніка, 2003. – 240 с.
17. Blackman I. Pareto principle plus statistic methodology in establishing a cost-estimating model / I. Blackman, E. Chan // International CIB World Building: 19th Congress, Brisbane, 05–09 May 2013: proceedings. – Brisbane : Queensland University of Technology, 2013 – P. 1–14.
18. Google-Colab [Electronic resource]. – Access mode:https://colab.research.google.com/?utm_source=scs-index.

DOI 10.15588/1607-3274-2024-4-7