Modern mathematical methods for the formation of an optimal set of fulfilled requirements with restriction on resource in big-data analysis

D. G. Fisenko, V.S. Mozgovoy,
Limited Liability Company "Research Institute Of Economics And Management Organization In The Gas Industry", St. Staraya Basmannaya, 20, p. 8, Moscow, 105066
E-mail: fisenko.dima@yandex.ru

Abstract. The article is considered a solution for the mathematical problem for the formation of the optimal set of fulfilled requirements with taking into account the different restriction. The use of the mathematical programming method for solving this problem is demonstrated and the methods for formalizing the conditions of the problem related with certain properties of the requirements under consideration are proposed. The method for analyzing the risks for the generated sets is proposed. The class of problems presented in the article relates to the field of big-data metrology.

1. Introduction

The article studies problem of making a decision to formation of the set of fulfilled requirements with limited resources. To determine the problem can be applied many different approaches including greedy algorithm or dynamic programming method but the linear-integer programming method is most efficient [1,2]. In this task binary variables illustrate the condition of requirement i.e. indicate that each of the proposed requirements had to be included in the set.

2. Problem statement

Consider the N requirements which realization are planned in the next T periods. It is planned to reach the maximum or minimum on any indicator by using limited resource S_1… S_T. It is necessary to indicate which requirements have to be started, in what period of time it should happened and the value of target. To determine the task by using linear integer programming it is necessary to provide conditions of the task in terms of formal target function system and inequality which should be maximized or minimized. In this case target function will be an index A and time-varying index B_1… B_T will be a limitation. The condition of the task in the formalized form will be:

\[
\begin{aligned}
B_1 &\leq S_1 \\
\vdots &\\
B_T &\leq S_T \\
A &\to \text{max}
\end{aligned}
\]  

This task could be solved by using linear programming methods if all the indexes in the set of requirements would have the additive property i.e. linearly depend on relevant indexes of the considered requirements. It is convenient to use the set of the binary variables x_1… x_T to determine the timeline of requirements implementation. In this system the variable that appropriate for the period when requirements started equal to one and other variables equal to zero. In this case the branch and bound method is the most efficient to find the solution [3].

One requirement cannot be implement twice therefore for any requirement this condition should be done:
When the number of requirement and its indicators it is known it is possible to calculate the contribution that this number contributes to full range.

The data on requirements are showed in the table as:

| Table 1. Dynamic and static indicators |
|----------------------------------------|
| Period                  | Indicator A | Indicator B |
|-------------------------|-------------|-------------|
|                         | -           | 1           | 2           | ... | T  |
| Requirement 1           | $a_1$       | $b_{11}$    | $b_{12}$    | ... | $b_{1T}$ |
| Requirement 2           | $a_2$       | $b_{21}$    | $b_{22}$    | ... | $b_{2T}$ |
| ...                     | ...         | ...         | ...         | ... | ...     |

Here is the indicator $A$ is static, i.e. it is attributed to requirement as a whole and remained in time but the indicator $B$ is dynamic, i.e. has its own value in each period. Contribution of each requirement into static indicator set of implementation requirements depends on whether the requirement it is included into the set:

$$A_n = a_n \sum_{t=1}^{T} x_{nt}$$  \(3\)

However, if we talk about the indicator which reduces its own significance depending on period (discounting) then with each period the contribution of requirement needs to decrease by a certain coefficient (discount rate) $D$:

$$A_n = a_n \sum_{t=1}^{T} \frac{x_{nt}}{(1 + D)^{t-1}}$$  \(4\)

In the case of dynamic indicators, the requirement contributes to each period depending on which period its realization will start. If it is delayed for several periods so value of the contribution of requirement is shifting accordingly:

| Table 2. Model of three dynamic requirements calculating for three periods |
|-----------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Period                     | 1               | 2               | 3               | ...             |
| Contribution               |                 |                 |                 |                 |
| Requirement 1              | $b_{11}$        | $b_{12}$        | $b_{13}$        | ...             |
| Requirement 2              | -               | $b_{21}$        | $b_{22}$        | ...             |
| Requirement 3              | -               | -               | $b_{31}$        | ...             |
| Total:                     | $b_{11}$        | $b_{12} + b_{21}$| $b_{13} + b_{22} + b_{31}$ | ...         |

The contribution of requirement in the dynamic indicator of the set of performing requirements during the period $t$ in general terms can be expressed as:
\[ B_{n \tau} = \sum_{\tau=1}^{\tau} b_{n \tau-\tau+1} x_{n \tau} \]  

(5)

Contribution to discounted dynamic indicators which are subject to decreasing significance depending on the execution period (discounting) can also be calculated by using a formula similar to (4):

\[ B_{n \tau} = \sum_{\tau=1}^{\tau} \frac{b_{n \tau-\tau+1} x_{n \tau}}{(1 + D)^{\tau-1}} \]  

(6)

3. Preprocessing phase

When the set of requirements is formatted there should also take into account certain conditions which could be associated with the implementation of each requirement separately:

In cases when requirement cannot be started later than certain period \( t_{max} \), besides the condition (2), additional limitation on it imposed:

\[ \sum t \cdot x_{n \tau} \leq t_{max} \]  

(7)

Some requirements can be imposed the commitments which means that requirement must be fulfilled a certain period.

To take this aspect into account a penalty function should be introduced and it should not exceed a certain maximum. Contribution in this function will make only requirements excluded from the set.

\[ R_n = a_n \left( 1 - \sum x_{n \tau} \right) \]  

(8)

If it is necessary to start implement the requirement in the certain period \( \tau \), limitations from this period will also lead to violations so consequently expression (9) takes the form:

\[ R_n = a_n \left( 1 - x_{n \tau} \right) \]  

(9)

In case of long-term consequences of breach of obligations, the dynamic penalty function could be introduced and its contribution of requirement for each period:

\[ R_{n \tau} = b_{n \tau} \left( 1 - x_{n \tau} \right) \]  

(10)

In this article requirements are viewed as indivisible objects but in practice they presented as a combination of the tasks performed. If part of the task is already completed it is necessary to complete them even if it is planned to suspend implementation the rest of the requirement. In order to take account of such task it is necessary to know their indicators:
Table 3. Example of already started tasks

| Period | Tasks that started | Requirement |
|--------|-------------------|-------------|
| 1      |                   | 1           |
| 2      |                   | 2           |
| …      |                   | …           |
| T      |                   | T           |

Even if this requirement was excluded from the set its contributions into the considered indicators will be equal \( b'_{n_1} \) ... \( b'_{n_T} \) but if it is incorporated then given the temporary shift the contributions \( (b_{n_1} - b'_{n_1}) \) ... \( (b_{n_T} - b'_{n_T}) \) would be added to considered indicators. The final contribution of these requirements might be expressed as:

\[
B_{n_t} = \sum_{t=1}^{T} (b_{n_{t-\tau+1}} - b'_{n_{t-\tau+1}})x_{n_t} + b'_{n_t}
\]  

(11)

4. Formalized task

To define the set of indicators it is necessary to sum the contributions from all participating in the requirements considered. As the result the system of inequality that characterizing task limitations besides the conditions (2) and (7) may contain expresses presented below.

The limitation at simple and static indicators which exposed to decrease the importance (discounting):

\[
\sum_{n=1}^{N} \sum_{t=1}^{T} a_n x_{n_t} \leq A_n \quad \sum_{n=1}^{N} \sum_{t=1}^{T} \frac{a_n x_{n_t}}{(1 + D)^{t-1}} \leq A_n
\]  

(12)

The limitation at simple and static indicators which exposed to decrease the importance (discounting):

\[
\sum_{t=1}^{T} \sum_{n=1}^{N} b_{n_{t-\tau+1}} x_{n_t} \leq B_{n_t} \quad \sum_{t=1}^{T} \sum_{n=1}^{N} \frac{b_{n_{t-\tau+1}} x_{n_t}}{(1 + D)^{t-1}} \leq B_{n_t}
\]  

(13)

Penalty function for requirement that should be start in the period \( t \):

\[
\sum_{n=1}^{N} b_{n_t} x_{n_t} \geq R_t - \sum_{n=1}^{N} b_{n_t}
\]  

(14)

The limitation at dynamic indicators for requirements which tasks have already begun to be completed:

\[
\sum_{t=1}^{T} \sum_{n=1}^{N} (b_{n_{t-\tau+1}} - b'_{n_{t-\tau+1}})x_{n_{t}} \leq B_{n_t} - \sum_{n=1}^{N} b'_{n_t}
\]  

(15)

As a target function may be used any statistic indicator or dynamic indicator summarized by periods:
According to linear programming method result of the solution composed of these expresses will be the value of state variables for each of the requirements under consideration which allows to form the optimal set in given conditions.

Risk indicators of implement the requirements have no additivity property therefore they cannot be used as objective function or as limitation in solving optimization task. However, they are important criterions in comparing set of various options of the set of carried out by requirements among themselves therefore they need to be evaluated.

5. Risk analysis

In the current task the indicators that reflecting the result of implementation the requirement is probabilistic with a given normal distribution function. It follows that relevant indicators of the set also are not defined consequently there is a risk that the results of realization the set of requirements would be below expectations.

In accordance with Central limit theorem the sum of a sufficiently large number $N$ independent random values have a distribution close to normal with an average and standard deviation equal:

$$M_p = \frac{1}{N} \sum M_i \quad \sigma_p = \sqrt{\sum \sigma_i^2} \quad (17)$$

Numerical experiments have shown that in order to adequately apply the Central limit theorem it is enough to include in the set 5-6 requirements comparable to each other in scale. In this case the risk that probabilistic indicator of the set of carried out by requirements $A$ turns out to be negative equal:

$$P(A < 0) = N(M_p, \sigma_p^2, 0) \quad (18)$$

If it is about indicator of the set of requirements reflecting the income so expected losses in disadvantageous outcome may be expressed as:

$$L = \int_{-\infty}^{0} N^{-1}(M_p, \sigma_p^2, x) \, dx \quad (19)$$
6. Conclusions

This article provides the formalization of the task of seeking the optimal set of carried out by requirements with using linear integer programming. Expressions which reflecting basic limitations of target indicators, penalty function breach of obligations as well as taken into account the features of requirement which tasks started earlier than the moment of formation of the set were formulated. For formed sets of requirements carried out was presented the risk assessment method related with their realization.

References

[1] Wagner G. Fundamentals of operations research. / - M.: World, 1972. 1,2,3 volume.
[2] Hamdy A. Taha Operations Research: An introduction. – M.: PH «Williams», 2005.
[3] Land A., Doig A. An automatic method of solving discrete programming problems // Econometrica. – V. 28. – 1960.
[4] U..V. Litvin, V.S. Kulyk, A.V. Kirpichnikov and others. Methodology for risk assessment of project portfolios geological prospecting for oil and gas company// Economy and management of the oil and gas complex– 4(172).2019