Analysis of characteristics of the local search method in the process of solving the knapsack optimization problem in the decision support systems

P A Rahman¹ and V E Bolnokin²

¹Department of Automated Technological and information Systems, Ufa State Petroleum Technological University, 2, October Avenue, Sterlitamak, 453118, Russian Federation
²Mechanical Engineering Research Institute of the Russian Academy of Sciences, 4, Pionerskaya street, Korolyov, 141070, Russian Federation

E-mail: pavelar@yandex.ru, vitalybolnokin@yandex.ru

Abstract. This scientific paper deals with the knapsack optimization task and solving method based on the local search in the boolean space. The obtained by the author analytical formulas for assessment of the upper bound for the amount of enumeration of points in the boolean space and lower bound for the quality of suboptimal solutions in the process of solving the knapsack optimization problem by the local search method are also discussed. Finally, the results of experimental research of the amount of enumeration of points in the boolean space and quality of suboptimal solutions for the different initial conditions of the knapsack optimization task, which confirmed the analytical formulas, are also given.

1. Introduction

In present days the decision problems and methods and tools for finding the solutions [1, 2] play key role as in life of persons, so in business processes of enterprises. Moreover, the modern world often deals with the discrete indivisible objects, and, accordingly, with the solving methods based on the enumeration of the finite set of solution variants in the multidimensional discrete space [3, 4]. Accordingly, the mathematical models of the discrete optimization problems and their solving methods are quite urgent in the modern world.

It is known, that in general case the global optimum of the discrete optimization problem [5], including the well-known knapsack optimization task [6], can be found only via exhaust enumeration of all points of the multidimensional discrete space. Accordingly, in practice the approximate solving methods are used in order to reduce the amount of enumerated points. One of them is the local search method [7, 8], which provides directed enumeration in the search space without guarantee to find the global optimum. In most cases, the local search method allows to find some suboptimal solution at which the value of the criterion function of the discrete optimization problem can be much worse than in the global optimum.

In such situation the assessment of the lower bound for the quality of suboptimal solutions is quite important task. The lower bound for the quality of suboptimal solutions allows us to estimate the
maximum possible «loss» in the value of the criterion function compared to its value in the unknown
global optimum. In particular, if we deal with the decision problems in economics, this can help us to assess
the greatest possible lack of profit in case of the worst suboptimal solution.

The value of quality of suboptimal solution depends entirely on the initial conditions of the discrete
optimization task. However, there is lower bound for the quality of suboptimal solutions, below which the
quality does not fall for any initial conditions of the task with the given dimension and additional
parameters of solving method.

Moreover, the assessment of upper bound for the amount of enumeration of points in the discrete
space is also very important, because it is directly related with the time of finding the solution of the
discrete optimization task in the worst case. The amount of enumeration also completely depends on
the dimension and initial conditions of the discrete optimization task. However, there is upper bound
for the amount of enumeration, above which it does not rise for any initial conditions of the task with
the given dimension and additional parameters of the solving method.

Within the scope of research in the field of combinatorial analysis [9, 10], scientific problem of estimation
of the upper bound for the amount of enumeration and lower bound for the quality of suboptimal solutions in the process of solving the knapsack optimization task by the local search method was raised and solved by the author.

2. The mathematical model of the knapsack optimization task and its solving method

One of the most prominent examples of the discrete optimization tasks is the knapsack optimization
task [5, 6]. The essence of the task is as follows: a set of items \{P_j\}, j = 1..n, with the different
weights \(a_j \geq 0\) and values \(c_j \geq 0\), and the knapsack with the limited capacity \(b \geq 0\) are given. It is
necessary to find a subset of the items, for which the total value of the items placed in the knapsack is
maximal and total weight of the items is not greater than capacity of the knapsack.

Accordingly, the mathematical model of the knapsack optimization task is as follow:

\[
\begin{align*}
L &= \sum_{j=1}^{n} c_j x_j \rightarrow \max; \\
\forall j \in [1,n] : x_j \in \{0,1\}; \\
a_j \geq 0; b \geq 0; c_j \geq 0.
\end{align*}
\]

Where, \(L\) is the criterion function of the knapsack optimization task. It shows the total value of the
items placed in the knapsack.

The global optimum of the task for any initial conditions can be reliably found only via exhaust
search – complete enumeration of all \(2^n\) points in the boolean search space. Therefore, to reduce the
amount of enumeration the approximate methods, which give some suboptimal solutions, are used. One
of the well-known approximate solving methods is the local search method [7, 8].

To simplify further discussion of the local search method let us at first introduce several important
definitions of the combinatorial analysis.

Let the \(n\)-dimensional boolean space consisting of the set of points \(X = (x_1, \ldots, x_n)\) with boolean
coordinates \(x_j \in \{0,1\}, j = 1..n\), is given.

Definition 1. \(D(X_A, X_B)\) is the distance between the points \(X_A\) and \(X_B\), numerically equal to the
number of the coordinates distinguishing these points. According to the Hamming metric, the distance
can be calculated by the following formula: \(D(X_A, X_B) = \sum_{j=1}^{n} |x_j^{(A)} - x_j^{(B)}|\).

Definition 2. \(S(X_0)\) is the \(n\)-dimensional search sphere with the radius \(r\) and centre \(X_0\), consisting of
a set of points, which differ from the point \(X_0\) by exactly \(r\) coordinates.

In other words, \(\forall X \in S(X_0) \rightarrow D(X, X_0) = r\).
Definition 3. \( Z_i(X_0) \) is the \( n \)-dimensional search zone with the radius \( r \) and centre \( X_0 \), consisting of all search spheres with the radii from 0 to \( r \).
In other words, \( Z_i(X_0) = S_0(X_0) \cup ... \cup S_r(X_0) \).
Now let us briefly discuss the local search method for solving the knapsack optimization task.

1) At first, some starting point \( X_0 \) is selected and the criterion function is calculated in it. The starting point is taken as the current optimum \( X_{OPT} \) and centre of the search zone \( Z_i(X_{OPT}) \).

2) Next, all points \( X \) in the zone \( Z_i(X_{OPT}) \) are enumerated, value of criterion function is calculated in them (in the base version of the local search method the radius of the search zones is not specified and it is assumed to be 1). Two search strategies are possible during the enumeration of points:

- Enumeration of all points in the zone \( Z_i(X_{OPT}) \) and searching of the best \( X^* \) point, in which value of the criterion function is better than in all other points of the zone (including the current centre of zone). This point is taken as a new current optimum and centre of the new search zone.
- Enumeration of the points in the zone \( Z_i(X_{OPT}) \) until the «first improvement», when during the process of enumeration of points the first point, in which value of the criterion function is better than in the previous point, is taken as a new current optimum and centre of the new search zone.

3) For the both strategies, the algorithm terminates when there is no point can be found in the search zone, in which value of the criterion function is better than in the centre of the search zone. It is obvious, that the second search strategy provides faster solving of the knapsack optimization tasks, because in this case less amount of points in the search zones are enumerated. However, the quality of solution may be worse than in case of using the first search strategy.

Thus, we overviewed the base version of the local search method. Within this scientific research we consider the local search method in the following modification:

- The point \( X_0 = (0, 0, ..., 0) \) is used as the starting point.
- The radius of the search zones is not assumed to be one, and can be given as an additional parameter \( r \). This allows us to control the size of the search zones.
- For enumeration of points in the search zones the first search strategy is used. However, the current search zone \( Z_i(X_c) \) (except the first search zone) always contains a part of points from the previous search zone \( Z_i(X_{PC}) \), because \( D(X_c, X_{PC}) \leq r \). Therefore, the modified local search method ignores the points, which belongs to the previous zone, and checks only the points \( X^* \in Z_i(X_c) \) for which \( D(X^*, X_{PC}) > r \). This allows us reduce the amount of enumeration. It should be noted, that in the first search zone the method always checks all points of the zone.

Finally, to use the local search method as the solving method for the knapsack optimization task, let us convert the mathematical model of the task (1) into the following form:

\[
\begin{align*}
F &= \sum_{j=1}^{n} c_j x_j - G \cdot \max\{0, \sum_{j=1}^{n} a_j x_j - b\} \rightarrow \max; \\
& \quad \forall j \in [1, n]: x_j \in [0,1].
\end{align*}
\]  

(2)

In the converted mathematical model (2), the constraint of the knapsack optimization task is absent explicitly. However, the constraint is taken into account with the help of the penalty function \( G \cdot \max\{0, \sum_{j=1}^{n} a_j x_j - b\}, \ G > 0 \). When we introduced the knapsack optimization task, it was noted, that the parameters \( a_j, b \) and \( c_j \) of the task are non-negative. Accordingly, the adjusted criterion function \( F \) in this case is always non-negative for all “acceptable” points, in which the constraint is not violated. In case of “unacceptable” points, the adjusted criterion function \( F \) can be both negative or non-
negative, it depends on the coefficient $G$. Therefore, it is better to select the coefficient $G$ large enough to make the adjusted criterion function $F$ be negative for any “unacceptable” point. After that, the task can be solved with the use of the local search method.

Figure 1 shows the algorithm for the modified local search method using the additionally given radius $r$ of the search zones. The algorithm uses the following main and auxiliary designations:

$q$ is the radius of the current search sphere in the current search zone;
$X_C$ is the centre of the current search zone (current optimum);
$X_{PC}$ is the centre of the previous search zone (previous optimum);
$X^*$ is the current best point in the current search zone;
$F^*$ is the value of the criterion function in the point $X^*$;
Flag1 indicates that the algorithm is working with the first search zone, and, respectively, all points of the search zone should be checked without exception;
Flag2 indicates that the better point in the search zone was successfully found.

**Figure 1.** Diagram of the algorithm of the modified local search method.
3. The upper bound for the amount of enumeration of points

Within the research work the author analyzed the upper bound for the amount of enumeration of points in the boolean space when solving the knapsack optimization task by the local search method with the given dimension \( n \) of the task and radius \( r \), \( 1 \leq r \leq n \), of the search zones.

Under the amount of enumeration \( V \) the total number of the points enumerated by the local search method in the process of solving the knapsack optimization task is understood.

As a result, the following analytical formula for estimation of the upper bound \( V_{\text{max}} \) for the amount of enumeration was obtained by the author:

\[
V_{\text{max}} = \sum_{p=0}^{r} C_n^p \cdot \left[ \frac{n}{r} \cdot \sum_{q=0}^{\lfloor n/2 \rfloor-1} \sum_{r=0}^{p} C_n^q C_{n-r}^{p-q} \right].
\]  

(3)

Within the research work the author also developed a software implementation of the local search method and carried out the experimental research that confirmed the obtained analytical estimation of the upper bound for the amount of enumeration.

The separate series of the experiments for the various cases of the dimension \( n \) and radius \( r \) with 10000 experiments in each series for the different random values for the parameters \( b \), \( a_j \), and \( c_j \) of the knapsack optimization task were carried out. In each experiment the task was solved by the modified local search method, and the amount of enumerated points during the solving process was counted. After that, the highest value among the values of the amount of enumeration in all experiments in the series was chosen.

The following values were used for the parameters \( b \), \( a_j \), and \( c_j \):

- \( b = \) random integer number in the range from 1 to 1000;
- \( a_j = \) random integer numbers in the range from 1 to \( b \);
- \( c_j = \) random integer numbers in the range from 1 and 1000;

Table 1 given below shows the results of the experimental research of the highest values of the amount of enumeration for the given \( n \) and \( r \), among all the values of the amount of enumeration for the different random values of \( b \), \( a_j \), and \( c_j \).

| \( n \) | \( r \) | \( 1 \) | \( 2 \) | \( 3 \) | \( 4 \) |
|-------|-------|-------|-------|-------|-------|
| 2     |       | 5     |       |       |       |
| 3     |       | 10    | 9     |       |       |
| 4     |       | 17    | 17    | 17    |       |
| 5     |       | 26    | 34    | 36    | 33    |
| 6     |       | 37    | 52    | 74    | 69    |
| 7     |       | 50    | 89    | 148   | 145   |
| 8     |       | 65    | 121   | 238   | 273   |
| 9     |       | 82    | 186   | 388   | 546   |

Table 1. The highest values of the amount of enumeration.

The results of experiments show that the highest values of the amount of enumeration are close to the analytical estimations of the upper bound for the amount of enumeration and, most importantly, they are not greater than analytical estimations.

4. The lower bound for the quality of suboptimal solutions

Within the research work the author also analyzed the lower bound for the quality of suboptimal solutions when solving the knapsack optimization task by the local search method with the given dimension \( n \) of the task and radius \( r \), \( 1 \leq r \leq n \), of the search zones.
Under the quality of the suboptimal solution $\xi$ the ratio of value of the criterion function $\bar{L}$ in the suboptimal solution to value of the criterion function $L^*$ in the global optimum of the knapsack optimization task is understood. In other words, $\xi = \bar{L} / L^*$.

As a result, the following analytical formula for estimation of the lower bound $\xi_{mn}$ of the quality of solutions was obtained by the author:

$$
\xi_{mn} = \begin{cases} 
\frac{r}{n-1}, & 1 \leq r \leq n-2; \\
1, & n-1 \leq r \leq n.
\end{cases}
$$

(4)

Within the research work the author also developed a software implementation of the local search method and carried out the experimental research that confirmed the obtained analytical estimation of the lower bound for the quality of suboptimal solutions.

The separate series of the experiments for the various cases of the dimension $n$ and radius $r$ with 10000 experiments in each series for the different random values for the parameters $b, a_j$, and $c_j$ of the knapsack optimization task were carried out. In each experiment the task was solved by the local search method and also by the exhaust search to find the global optimum, and quality of solution was calculated. After that, the lowest value among the values of the quality of solutions in all experiments in the series was chosen.

The following values were used for the parameters $b, a_j$, and $c_j$:

- $b =$ random integer number in the range from 1 to 1000;
- $a_j =$ random integer numbers in the range from 1 and $b$;
- $c_j =$ random integer numbers in the range from 1 and 1000;

Table 2 given below shows the results of the experimental research of the lowest values of the quality of solutions for the given $n$ and $r$, among all the values of the quality of solutions for the different random values of $b, a_j$, and $c_j$.

| $n$ | 1    | 2    | 3    | 4    |
|-----|------|------|------|------|
| 2   | 1.000| –    | –    | –    |
| 3   | 0.508| 1.000| –    | –    |
| 4   | 0.344| 0.675| 1.000| –    |
| 5   | 0.262| 0.516| 0.761| 1.000|
| 6   | 0.212| 0.407| 0.615| 0.821|
| 7   | 0.171| 0.345| 0.521| 0.673|
| 8   | 0.149| 0.292| 0.434| 0.579|
| 9   | 0.128| 0.261| 0.381| 0.512|

The results of experiments show that the lowest values of the quality of solutions are close to the analytical estimations of the lower bound for the quality of solutions and, most importantly, not lower than the analytical estimations.

5. Conclusion

Thus, within this scientific paper the knapsack optimization task and its solving method based on the local search in the boolean space are discussed.

The obtained by the author formulas for estimation of the upper bound for the amount of enumeration of points in the boolean space and lower bound for the quality of suboptimal solutions in the process of solving the knapsack optimization task by the local search method are also discussed.
Finally, the results of experimental research of the amount of enumeration of points in the boolean space and quality of suboptimal solutions for the different initial conditions of the knapsack optimization task, which confirmed the analytical formulas, are also presented. The obtained scientific results were used by the author for development of the specialized optimization software for the resource allocation in computer networks and laboratory works for studying the discrete optimization methods for the students of technical specialties.

Acknowledgements
The author is grateful to professor I. I. Ladygin from Moscow Power Engineering Institute for scientific support and theoretical base in the field of discrete optimization.

References
[1] Sharda R, Delen D and Turban E 2014 Business Intelligence and Analytics: Systems for Decision Support (Pearson Education)
[2] Kravets O Ja, Preobrazhenskiy A P, Kocheagarov A V, Choporov O N and Bolnokin V E 2018 International Journal on Information Technologies and Security 10(4) 45-56
[3] Gallier J 2011 Discrete Mathematics (New York: Springer-Verlag)
[4] Novikov F A 2009 Discrete mathematics for programmers (Saint-Petersburg: Piter)
[5] Cook W J, Cunningham W H, Pulleyblank W R and Schrijver A 1998 Combinatorial Optimization (New York: Wiley)
[6] Kellerer H, Pferschy U, Pisinger D and 2004 Knapsack Problems (Springer-Verlag)
[7] Michiels W, Aarts E and Korst J 2007 Theoretical Aspects of Local Search (Springer-Verlag)
[8] Battiti R, Mauro B and Franco M 2008 Reactive Search and Intelligent Optimization (Springer-Verlag)
[9] Rahman P A 2018 J. Phys.: Conf. Ser. 1015 022015
[10] Rahman P A 2018 J. Phys.: Conf. Ser. 1015 022016