Dependence of the Average Number of Solutions in the Knapsack Problem on the Parameters of the Constraint Domain

Vladimir K. Leontiev [1,2], Eduard N. Gordeev [1]

[1] *Bauman Moscow State Technical University, 5/1 2nd Baymanskay ul., Moscow, 105005, Russia*
[2] *Dorodnicyn Computing Centre of RAS, 119133, Moscow, 40 Vavilova ul., Moscow, 119133, Russia*

Abstract

The knapsack problem is used in many mathematical models, including in the field of information security. Solving a problem (finding the optimal knapsack load) or even answering the question about the existence of a valid solution is an NP-complete problem. In this regard, the question of finding the power of the set of acceptable solutions or estimates of this power is relevant. The paper analyzes the combinatorial aspects of this problem based on the method of generating functions. Formulas and estimates for the number of solutions and the average number of solutions depending on the coefficients of the constraint vector are obtained. On their basis, computational algorithms for finding these values can be constructed. All this can be used to assess the adequacy and quality of the original mathematical model.

Keywords

Knapsack problem, search for an optimal solution, optimization problem, feasible solutions, generating functions, estimates of the number of solutions, NP-completeness

1. Introduction

Use The knapsack problem with Boolean variables has the form:

$$\sum_{j=1}^{n} c_j x_j \to \max ;$$

$$\sum_{i=1}^{n} a_i x_i \le b , \qquad (1)$$

where $x=(x_1,,x_n)$ is an n-dimensional Boolean vector.

In what follows, we will assume that all parameters of the problem under consideration are non-negative integers. The set of feasible solutions to this problem V_b is the set of Boolean vectors satisfying the inequality

$$\sum_{i=1}^{n} a_i x_i \le b . \qquad (2)$$

The volume V_b is the number $|V_b|$ feasible solutions to problem (2).

This is a classical combinatorial optimization problem (see [1], [2]). This work is a continuation of the research carried out in [3], [4]. Various aspects of the knapsack problem related to the topic of this work were studied, for example, in [2], [5], [6]. The approach proposed by the authors has no direct analogs, as can be seen, for example, from the most detailed survey monograph devoted to the knapsack problem [7].

The importance of this topic from the point of view of cryptography is confirmed by numerous works in specialized journals.

BIT-2021: XI International Scientific and Technical Conference on Secure Information Technologies, April 6-7, 2021, Moscow, Russia
EMAIL: vkleontiev@yandex.ru (A. 1); werhorn@yandex.ru (A. 2)
ORCID: 0000-0001-5700-1950 (A. 1); 0000-0002-7766-3772 (A. 2)

For example, in the works of G.V. Balakin [8] and [9] consider specific classes of systems of Boolean equations (of recurrent type) and their application in cryptography.

In the work of A.M. Zubkov [10] the moment characteristics of the weights of vectors in random binary linear codes are described in terms of the properties of special systems of equations.

Linearization of systems of Boolean equations is a method for solving systems, which consists in replacing all monomials of degree higher than the first with new variables, solving the resulting linear system and then checking the obtained solutions for correctness. This, for example, is the subject of the articles [11], [12].

A group of scientists led by N. Courtois proposed improvements to XL4 and XSL5 of the linearization method for cases when the number of equations in the system is not enough to effectively apply linearization in the classical form [13]. The essence of these methods is to supplement the system with new equations that do not change the set of solutions to the system, but increase the size of the system and the rank of the linearized system. Later N. Courtois and G.V. Bard [14] proposed another method based on the ElimLin linearization method.

As it was said in the annotation, in our work we consider the problem of finding the number of feasible solutions. The results along this path are illustrated by works [15-17]. And in the applied works of A.S. Meluzov [18] and [19], a software package was developed and implemented to solve the above problem.

This article also discusses the problem of parametrization of the system on its right side. The authors are not aware of any analogues of the approach presented here. To some extent, both classical algorithms of parametric linear programming and, for example, work [20], where linear equations of the Boolean type with a "distorted" right-hand side are investigated, are related to it.

The paper proposes a method for constructing a set containing the desired vector with a probability not less than a given one, and estimates the cardinality of this set. Theoretical calculations of the parameters of the method are illustrated by the results of experiments. This approach uses a probabilistic setting, while the combinatorial apparatus is used here.

This work consists of an introduction and three sections. The main lemma is given in the next section. In the third section, we consider the question of the average number of solutions depending on the values of the right-hand side, and then in the section following this we consider the case of a hypercube of bounded volume.

Some definitions, concepts and methods of proof were previously used by the authors in [3-6].

In what follows, we will assume that all the parameters of the problem under consideration, the numbers $c_1,\ldots,c_n;\ a_1,\ldots,a_n;\ b$ - non-negative integers.

2. Generating functions for the number of feasible solutions

The answer to the question about the existence of solutions to the problem under consideration is already an NP-complete problem. Therefore, finding the number of solutions to the knapsack problem is currently carried out either by exhaustive search algorithms, or is estimated from heuristic methods.

In both cases, knowledge of the average number of solutions can be used both for constructing algorithms and for modifying them. In addition, formulas for finding the average number of solutions depending on the parameters of the problem can be used to assess the feasibility of using both exhaustive search and heuristic algorithms.

For example, we need to find at least one solution to an individual problem that lies in a class of problems with a fixed right-hand side. We find the average number of solutions in this class. If it is large, then this can serve as a justification for the expediency of using heuristic or probabilistic methods. Otherwise, the exhaustive search algorithm is most likely more reasonable.

The main method that is used in the work is the method of generating functions.

First, we directly express the volume of feasible solutions using the generating function for

$$P_b(z_1,\ldots,z_n)=\sum_{x\in V_b} z_1^{a_1x_1}z_2^{a_2x_2}\ldots z_n^{a_nx_n}.$$

The following lemma was proved in [4].

Lemma 1. The following formula is valid

$$\sum_{b=0}^{\infty} P_b(z_1,...,z_n)u^b = \frac{(1+(z_1u)^{a_1})...(1+(z_nu)^{a_n}).}{1-u} \tag{3}$$

Consequence. Let 0 <ρ <1. Then equality holds for the volume of the region of feasible solutions:

$$|V_b| = \frac{1}{2\pi i} \oint_{|u|=\rho} \frac{(1+u^{a_1})...(1+u^{a_n})}{(1-u)u^{b+1}} du \,. \tag{4}$$

3. Average number of feasible solutions

Let us now consider the question of the average value of the volume of solutions.
Let be

$$|\overline{V_b}| = \frac{1}{(b+1)^n} \sum_{\substack{\{a_1,...,a_n\} \\ 0\le a_i \le b, 1\le i\le n}} |V_b| \,. \tag{5}$$

Recall that our variables are Boolean. The number b is fixed, and the dimension n is also fixed. Therefore, each coefficient of the constraint vector varies from 0 to b, i.e. takes n + 1 values. Hence (5) follows.

Of course, the question of the average number of solutions is rather theoretical in nature. However, it sheds light on the combinatorics of the problem when investigating the interdependence of b and n.

Below we give several statements that give various formulas (calculation algorithms) for the average number of solutions. As an illustration of the use of the method of generating functions, their proofs are presented.

Theorem 1. The following formula is valid

$$|\overline{V_b}| = \frac{1}{(b+1)^n} \sum_{k=0}^{n} C_n^k C_b^{n-k} (b+2)^k \,. \tag{6}$$

Proof. Directly from (4) and (5) we have

$$|\overline{V_b}| = \frac{1}{(b+1)^n} \frac{1}{2\pi i} \oint_{|u|=\rho} \left\{ \frac{\sum_{a_1=1}^{b}(1+u^{a_1})...\sum_{a_n=1}^{b}(1+u^{a_n})}{(1-u)u^{b+1}} \right\} du \,.$$

Further, we note that, firstly,

$$\sum_{a_k=0}^{b}(1+u^{a_k}) = 2+(1+u)+...+(1+u^b) = 2+b+\frac{u-u^{b+1}}{1-u}.$$

Secondly, $(1-u^b)^{n-k} = 1-C_{n-k}^1 u^b + C_{n-k}^2 u^{2b} - ... = \sum_{i=0}^{n-k}(-1)^i C_{n-k}^i u^{ib}$.

Therefore, further we have

$$|\overline{V_b}| = \frac{1}{(b+1)^n}\frac{1}{2\pi i}\oint\limits_{|u|=\rho}\left\{\frac{\left(b+2+\dfrac{u-u^{b+1}}{1-u}\right)^n}{(1-u)u^{b+1}}\right\}du = \frac{1}{(b+1)^n}\frac{1}{2\pi i}\oint\limits_{|u|=\rho}\left\{\frac{\left((b+2)(1-u)+u(1-u^b)\right)^n}{(1-u)^{n+1}u^{b+1}}\right\}du =$$

$$= \frac{1}{(b+1)^n}\sum_{k=0}^{n} C_n^k (b+2)^k \frac{1}{2\pi i}\oint\limits_{|u|=\rho}\left\{\left(\frac{u}{1-u}\right)^{n-k}\frac{(1-u^b)^{n-k}}{u^{b+1}}\frac{1}{1-u}\right\}du .$$

Now note that

$$(1-u^b)^{n-k} = 1 - C_{n-k}^1 u^b + C_{n-k}^2 u^{2b} - \ldots = \sum_{i=0}^{n-k}(-1)^i C_{n-k}^i u^{ib} . \tag{7}$$

Therefore, from (7) it follows:

$$\frac{1}{2\pi i}\oint\limits_{|u|=\rho}\left\{\frac{(1-u)^{-n+k-1}(1-u^b)^{n-k}}{u^{b+1+k-n}}\right\}du = \frac{1}{2\pi i}\oint\limits_{|u|=\rho}\left\{\frac{(1-u)^{-n+k-1}}{u^{b+1+k-n}}\right\}du = (-1)^{b+k-n}C_{-(n-k+1)}^{b+k-n}.$$

After obvious transformations, we obtain

$$(-1)^{b+k-n}C_{-n+k-1}^{b+k-n} = (-1)^{b+k-n}C_{-(n-k+1)}^{b+k-n} = C_b^{b+k-n} = C_b^{n-k} . \tag{8}$$

But now the assertion of the theorem follows from (7) and (8).
The theorem is proved.

Let's look at this expression from the other side.

Theorem 2. The following formula is valid

$$|\overline{V_b}| = \frac{1}{(b+1)^n}\sum_{k=0}^{n} C_n^k C_{b+k}^b (b+1)^{n-k} . \tag{9}$$

Proof. Let's apply the method of coefficients to the calculation of the sum.

$$|\overline{V_b}| = \frac{1}{(b+1)^n}\sum_{k=0} C_n^k C_b^{n-k}(b+2)^k .$$

We get a chain of ratios:

$$|\overline{V_b}| = \frac{1}{(b+1)^n}\sum_{k=0}^{n} C_n^k (b+2)^k \underset{u}{Coef}\left\{(1+u)^b u^{-b+n-k-1}\right\} = \frac{1}{(b+1)^n}\underset{u}{Coef}\left\{\frac{(1+u)^b}{u^{b-n+1}}\sum_{k=0}^{n} C_n^k\left(\frac{b+2}{u}\right)^k\right\} =$$

$$= \frac{1}{(b+1)^n}\underset{u}{Coef}\left\{\frac{(1+u)^b}{u^{b-n+1}}(1+\left(\frac{b+2}{u}\right)^n\right\} = \frac{1}{(b+1)^n}\underset{u}{Coef}\left\{\frac{(1+u)^b}{u^{b+1}}(u+b+2)^n\right\} =$$

$$= \frac{1}{(b+1)^n}\underset{u}{Coef}\left\{\frac{(1+u)^b}{u^{b+1}}((1+u)+(b+1))^n\right\} = \frac{1}{(b+)^n}\underset{u}{Coef}\left\{\frac{\sum_{k=0}^{n} C_n^k (1+u)^k (b+1)^{n-k}}{u^{b+1}}\right\}(1+u)^b =$$

$$= \frac{1}{(b+1)^n}\sum_{k=0}^{n} C_n^k (b+1)^{n-k}\underset{u}{Coef}\left\{\frac{(1+u)^{b+k}}{u^{b+1}}\right\}(1+u)^b = \frac{1}{(b+1)^n}\sum_{k=0}^{n} C_n^k C_{b+k}^b (b+1)^{n-k} .$$

The theorem is proved.

Examples.

1. Let n=b=2, then we have 9 restrictions: 1) $x_1+x_2\le 2$; 2) $2x_1+x_2\le 2$; 3) $x_1+2x_2\le 2$; 4) $2x_1+2x_2\le 2$. . And 4 more with one variable, as well as with two zero coefficients for variables. The last 5 have four solutions, and for the first four we have: $|V_1|=4$; $|V_2|=3$; $|V_3|=3$; $|V_4|=3$.

   Directly we get: $|\overline{V_b}| = 33/9 = 11/3$.

   The same is obtained from (9):

$$|\overline{V_b}| = \frac{1}{(b+1)^n}\sum_{k=0}^{n} C_n^k C_{b+k}^b (b+1)^{n-k} = \frac{1}{3^2}\sum_{k=0}^{2} C_2^k C_{2+k}^2 3^{2-k} =$$

$$= \frac{1}{3^2}(C_2^0 C_2^2 3^2 + C_2^1 C_3^2 3^1 + C_2^2 C_4^2 3^0) = 33/9 = 11/3.$$

2. Let b is arbitrary and n=3. The kind of restrictions is obvious. Direct calculation is difficult, but from (9) we immediately obtain:

$$|\overline{V_b}| = \frac{1}{(b+1)^n}\sum_{k=0}^{n} C_n^k C_{b+k}^b (b+1)^{n-k} = \frac{1}{(b+1)^3}\sum_{k=0}^{3} C_3^k C_{b+k}^b (b+1)^{n-k} =$$

$$= \frac{1}{(b+1)^3}((b+1)^3 + 3(b+1)^3 + 3(b+2)(b+1)^2/2 + (b+1)(b+1)(b+1)/6).$$

   This is asymptotically equals $5\frac{2}{3}$.

4. Average number of solutions to the knapsack problem on a hypercube of a fixed size

Let us now consider the question of the average value of the volume of feasible solutions on a hypercube of a fixed size.

Let be

$$t_b(a_1,...,a_n) = \frac{1}{2\pi i}\oint_{|u|=\rho} \frac{(1+u^{a_1})...(1+u^{a_n})}{(1-u)u^{b+1}}du,$$

$0 \le a_i \le c, i = 1,...,n$ and $\bar{t}_c$ is the average value for the volume of feasible solutions on a hypercube of a fixed size, i.e.

$$\bar{t}_c = \frac{1}{(c+1)^n}\sum_{\substack{\{a_1,...,a_n\} \\ 0\le a_i \le c, 1\le i\le n}} t_b(a_1,...,a_n).$$

Theorem 3. The formula is valid

$$\bar{t}_c = \frac{1}{(c+1)^n}\sum_{k=0}^{n} C_n^k (c+2)^k \sum_{r=0}^{n-k} (-1)^r C_{n-k}^r C_{b-rc}^{n-k}.$$

Proof. Directly from (4) and (6) we have

$$\bar{t}_c = \frac{1}{(c+1)^n}\frac{1}{2\pi i}\oint_{|u|=\rho}\left\{\frac{\sum_{a_1=1}^{c}(1+u^{a_1})\ldots\sum_{a_n=1}^{c}(1+u^{a_n})}{(1-u)u^{b+1}}\right\}du.$$

Further, note that, $\sum_{a_k=0}^{c}(1+u^{a_k}) = 2+(1+u)+\ldots+(1+u^c) = 2+b+\frac{u-u^{c+1}}{1-u}$.

Therefore, we have

$$\bar{t}_c = \frac{1}{(c+1)^n}\frac{1}{2\pi i}\oint_{|u|=\rho}\left\{\frac{\left(c+2+\frac{u-u^{c+1}}{1-u}\right)^n}{(1-u)u^{b+1}}\right\}du = \frac{1}{(c+1)^n}\frac{1}{2\pi i}\oint_{|u|=\rho}\left\{\frac{\left((c+2)(1-u)+u(1-u^c)\right)^n}{(1-u)^{n+1}u^{b+1}}\right\}du =$$

$$= \frac{1}{(c+1)^n}\sum_{k=0}^{n}C_n^k(c+2)^k\frac{1}{2\pi i}\oint_{|u|=\rho}\left\{\left(\frac{u}{1-u}\right)^{n-k}\frac{(1-u^c)^{n-k}}{u^{b+1}}\frac{1}{1-u}\right\}du.$$

Now we will find the sum of the residues to calculate the integral

$$\frac{1}{2\pi i}\oint_{|u|=\rho}\left\{\frac{(1-u)^{-n+k-1}(1-u^c)^{n-k}}{u^{b+1+k-n}}\right\}du.$$

Now, note that, $(1-u^c)^{n-k} = 1-C_{n-k}^1u^c+C_{n-k}^2u^{2c}-\ldots = \sum_{i=0}^{n-k}(-1)^iC_{n-k}^iu^{ic}$.

Therefore

$$\bar{t}_c = \frac{1}{(c+1)^n}\sum_{k=0}^{n}C_n^k(c+2)^k\frac{1}{2\pi i}\oint_{|u|=\rho}\left\{\frac{(1-u)^{k-n-1}}{u^{b+k-n+1}}\sum_{r=0}^{n-k}(-1)^rC_{n-k}^ru^{rc}\right\}du =$$

$$= \frac{1}{(c+1)^n}\sum_{k=0}^{n}C_n^k(c+2)^k\sum_{r=0}^{n-k}(-1)^rC_{n-k}^r\frac{1}{2\pi i}\oint_{|u|=\rho}\left\{\frac{(1-u)^{k-n-1}}{u^{b+k-n+1-rc}}\right\}du.$$

Notice, that

$$\frac{1}{2\pi i}\oint_{|u|=\rho}\left\{\frac{(1-u)^{-n+k-1}}{u^{b+1+k-n-\kappa c}}\right\}du = \begin{cases}0, b-rc\le 0\\ (-1)^{k-n+b-rc}C_{-(n-k+1)}^{b-rc-n+k}, b-rc\ge n-k.\end{cases}$$

In the general case, we obtain

$$\bar{t}_c = \frac{1}{(c+1)^n}\sum_{k=0}^{n}C_n^k(c+2)^k\sum_{r=0}^{n-k}(-1)^rC_{n-k}^r(-1)^{k-n+b-rc}C_{-(n-k+1)}^{b-rc-n+k} = \frac{1}{(c+1)^n}\sum_{k=0}^{n}C_n^k(c+2)^k\sum_{r=0}^{n-k}(-1)^rC_{n-k}^rC_{b-rc}^{b-rc-n+k} =$$

$$= \frac{1}{(c+1)^n}\sum_{k=0}^{n}C_n^k(c+2)^k\sum_{r=0}^{n-k}(-1)^rC_{n-k}^rC_{b-rc}^{n-k}.$$

The theorem is proved.

5. Conclusions

The method of generating functions can be successfully applied to the analysis of problems of a "combinatorial" nature. In the above study, with its help, new formulas were obtained for the number of olutions and the average number of solutions in the knapsack problem.

They can be improved and refined by considering equations not of a general, but of a special form. It is these problems that arise in specific applied areas, in particular, in mathematical models of information security, taking into account the real features of the original formulations.

Therefore, the results presented here can serve as a basis for further research.

The reliability of the results follows from the correctness of the definitions and proofs of the theorems.

6. Acknowledgements

The work was carried out with the support of a grant from the Russian Academy of Sciences: RFBR grant 20-01-00645.

.

7. References

[1] H. Papadimitriou, S. Steiglitz, Combinatorial optimization. Mir, Moscow, 1989.
[2] S.P. Gorshkov S. P., A.V. Tarasov , The complexity of solving systems of Boolean equations. Kurs, Moscow 2017.
[3] V.K. Leontiev, E.N. Gordeev, Generating functions in the knapsack problem, Reports of the Academy of Sciences 481 (2018) 478-480. doi: 10.31857/S086956520002139-5.
[4] V.K. Leontiev, E.N. Gordeev, On some combinatorial properties of the backpack problem, Journal of computational mathematics and mathematical physics 59 (2019) 1439-1447. doi: 10.1134/S0044466919080076.
[5] V.K. Leontiev, Boolean polynomials and linear transformations, Reports of the Academy of Sciences 425 (2009) 478-480.
[6] V.K. Leontiev, Combinatorics and information. Part 1. Combinatorial analysis, Moscow, MFTI, 2015, 174 p.
[7] H. Kellerer, U. Pferschy, D. Pisinger, Knapsack problems, Berlin, Springer, 2004.
[8] G.V. Balakin, On solving some classes of systems of Boolean equations of recurrent type, Mathematical questions of cryptography 4-1 (2013) 5-25. doi: 10.4213/mvk71.
[9] G.V. Balakin, On the possibility of partial recovery of some sequences from observations, Mathematical questions of cryptography 4-4 (2013) 7-25. doi: https://doi.org/10.4213/mvk97.
[10] A.M. Zubkov, V.I. Kruglov, Moment characteristics of vector weights in random binary linear codes, Mathematical questions of cryptography 3-4 (2013) 55-70. doi: https://doi.org/10.4213/mvk67.
[11] J.-C Faug're, A new efficient algorithm for computation Grebner bases (F4), Journal of pure and applied algebra 139 (1999) Issues 1-3 61-88. doi: https://doi.org/10.1016/S0022-4049(99)00005-5.
[12] J.-C Faug're, A new efficient algorithm for computation Grebner e o bases without reduction to zero (F5), Proceedings of the 2002 international symposium on Symbolic and algebraic computation 2002, Universit´e de Lille, France, ACM Press, 2002, pp. 75–83. doi: https://doi.org/10.1145/780506.780516.
[13] N. Courtois, J. Pieprzyk, Cryptanalysis of block chiphers with overdened systems of equations, Proc. 8th Int. Conf. on the Theory and Application of Cryptology and Information Security. Springer, Berlin, 2002, pp. 267–287.

[14] N. Courtois, G.V. Bard, Algebraic cryptanalysis of the data encryption standard, IMA International Conference on Cryptography and Coding Theory. Lecture Notes in Computer Science Springer-Verlag, Berlin, 2007, pp. 152-169. doi: https://doi.org/10.1007/3-540-36178-2_17.
[15] F. Massacci, L. Marraro, Logical Cryptanalysis as a SAT Problem, Journal of Automated Reasoning 24 (2000) 165-203. doi:10.1023/A:1006326723002.
[16] C. Fiorini, E. Martinelli, F. Massacci, How to fake an RSA signature by encoding modular root nding as a SAT problem, Discrete Applied Mathematics 130 (2003) 101-127.
[17] I. Mironov , L. Zhang, Applications of SAT Solvers to Cryptanalysis of Hash Functions, in: A. Biere , C.P. Gomes (Eds), Theory and Applications of Satisfiability Testing - SAT 2006, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, Vol. 4121, 2006, pp. 102-115. doi: https://doi.org/10.1007/11814948_13
[18] A.S. Meluzov, Construction of effective algorithms for solving systems of polynomial Boolean equations by testing a part of variables, Discrete Math. Appl., 21 (2011) 381–395. doi: 10.1515/DMA.2011.024.
[19] A.S. Meluzov, The use of associative information processing for constructing algorithms for solving systems of Boolean equations, Mathematics and Mathematical Physics 50(2010) 1925-1940. doi:10.1134/S0965542510110151.
[20] E.K. Alekseev, I.V. Oshkin, V.O. Popov, S.V. Smyshlyaev, Solving systems of linear Boolean equations with noisy right-hand sides over the reals, Discrete Math. Appl., 28 (2018) 1-5. doi: 10.1515/dma-2018-0001.