Generation Methods for Multidimensional Knapsack Problems and their Implications

Raymond R. Hill
Chaitr S. Hiremath
Department of Biomedical, Industrial, and Human Factors Engineering
Wright State University
Dayton, Ohio, 45435, U.S.A.

ABSTRACT

Although there are a variety of heuristics developed and applied to the variants of the binary knapsack problem, the testing of these heuristics are based on poorly defined test problems. This paper reviews the various types of knapsack problems, considers how test problems have been generated and depicts via empirical results the implications of using poorly formed test problems for empirical testing.

Keywords: knapsack problems, problem generation, multiple dimensional knapsack

1 INTRODUCTION

A variety of heuristic approaches have been devised for the knapsack problem (KP) and its variants. These approaches are generally empirically examined using test problems. The focus of this paper is on the type of test problems used, how they are generated, with emphasis on the test problem characteristics embedded across the range of test problems available. We demonstrate the implications of ill-devised test problems using multidimensional knapsack problems.

The organization of this paper is as follows. Section 2 gives a brief overview of KP and its variants and summarizes the empirical studies involving the variants of KP. Section 3 presents the results of a constraint coefficient analysis of the standard test problem sets. Section 4 provides the results of an empirical study explicitly demonstrating the implications of ill-devised test problems. Paper summary and concluding remarks are provided in section 5.

2 BACKGROUND

The KP is a non-trivial binary integer programming problem with a single constraint. While useful in its own right, industrial applications find the need for satisfying additional constraints such as urgency of requests, priority and time windows of the requests, and packaging with different weight and volume requirements, or the filling of multiple container types. These lead to variants and extensions of knapsack problems we examine: Multi-Dimensional Knapsack Problems (MDKP), Multiple Knapsack Problem (MKP), Multiple Choice Knapsack Problems (MCKP), Multiple-choice Multidimensional Knapsack Problems (MMKP).

2.1 The Multi-Dimensional Knapsack Problems (MDKP)

A set of n items are packed in m knapsacks with capacities c_i. Each item j has a profit p_j and weight w_{ij} associated with placing that item into knapsack i. The objective of the problem is to maximize the total profit of the selected items. The MDKP problem is formulated as:
Maximize,

$$Z = \sum_{j=1}^{n} p_j x_j \tag{1}$$

subject to,

$$\sum_{j=1}^{n} w_{ij} x_j \leq c_i, \qquad i = 1, ...m \tag{2}$$

$$x_j \in \{0, 1\}, \qquad j = 1, ...n \tag{3}$$

Equation (1) calculates the total profit of selecting item j and equation (2) ensures each knapsack constraint is satisfied. Equation (3) is the binary selection requirement.

2.2 The Multiple Knapsack Problems (MKP)

Given n items, we seek to pack m knapsacks with capacities c_i, $i = \{1, ..., m\}$. Each item j has a profit p_j and weight w_j. The problem is to select $i \in m$ disjoint subsets of items, such that subset i fits into the capacity c_i of knapsack i. The objective is to maximize the total profit of the selected items placed in the knapsacks. The MKP is formulated as:

Maximize,

$$Z = \sum_{i=1}^{m} \sum_{j=1}^{n} p_j x_{ij} \tag{4}$$

subject to,

$$\sum_{j=1}^{n} w_j x_{ij} \leq c_i, \qquad i = 1, ..., m \tag{5}$$

$$\sum_{i=1}^{m} x_{ij} \leq 1, \qquad j = 1, ..., n \tag{6}$$

$$x_{ij} \in \{0, 1\}, \qquad \text{forall i, j} \tag{7}$$

Equation (4) provides the total profit of assigning item j to knapsack i ($x_{ij} = 1$). Equation (5) ensures each knapsack constraint is satisfied while equation (6) ensures each item is assigned to at most one knapsack.

2.3 The Multiple Choice Knapsack Problems (MCKP)

The MCKP adds disjoint multiple-choice constraints [19]. Given m mutually disjoint classes $(N_1, ..., N_m)$ of items, pack representative items from these disjoint classes into a single knapsack of capacity c. Each item $j \in N_i$ has a profit p_{ij} and a weight w_{ij}. The objective of the problem is to maximize the profit of a feasible solution. The MCKP problem is formulated as:

Maximize,

$$Z = \sum_{i=1}^{m} \sum_{j \in Ni} p_{ij} x_{ij} \tag{8}$$

subject to,

$$\sum_{i=1}^{m} \sum_{j \in Ni} w_{ij} x_{ij} \leq c, \tag{9}$$

$$\sum_{j \in Ni} x_{ij} = 1, \qquad i \in \{1, ...m\} \tag{10}$$

$$x_{ij} \in \{0, 1\}, \qquad i \in \{1, ...m\}, \qquad j \in N_i \tag{11}$$

Equation (8) calculates the profit of an assignment of items, a value to be maximized. Equation (9) ensures the single knapsack capacity is not exceeded while equation (10) ensures selection of a single item from each of the m disjoint subsets of items.

2.4 Multiple-choice Multi-dimensional Knapsack Problems (MMKP)

The MMKP considers m classes of items, where each class has n_i items. Each item j of class i has profit value p_{ij}, and requires resources given by the weight vector $w_{ij} = (w_{ij1}, w_{ij2}, ..., w_{ijl})$. The amount of available resources are given by a vector $c = (c_1, c_2, ..., c_l)$. The MMKP selects an item from each class and is formulated as:
Maximize,

$$Z = \sum_{i=1}^{m} \sum_{j=1}^{n_i} p_{ij} x_{ij} \tag{12}$$

subject to,

$$\sum_{i=1}^{m} \sum_{j=1}^{n_i} w_{ijk} x_{ij} \leq c_k, \qquad k \in \{1, ...l\} \tag{13}$$

$$\sum_{j=1}^{n_i} x_{ij} = 1, \qquad i \in \{1, ...m\} \tag{14}$$

$$x_{ij} \in \{0, 1\}, i \in \{1, ...m\}, j \in \{1, ...n_i\} \tag{15}$$

Equation (12) provides the profit of selecting an item from every class, a value to be maximized. Equation (13) ensures the resource capacity of knapsack k is not exceeded while equation (14) ensures selecting a single item from each of the i classes.

2.5 Empirical Studies Involving Variants of Knapsack Problems

Researchers have developed heuristic algorithms to solve the KP and its variants. In [11], [14], and [17] branch and bound approaches are used to solve the MKP. The test problems generated in [11] and [14] were similar in structure with randomly generated profit and weight values. In [17] different types of test problem instances were obtained by varying the range of the distribution used for the coefficients.

Both [19] and [2] developed branch and bound algorithms to solve the MCKP while [16] solved the MCKP using dynamic programming. Both [19] and [6] randomly generate test problems by independently drawing the profit and weight values from a uniform distribution and explicitly avoiding the repetition of profit and weight values within the multiple-choice set. Similar to his work on MKP in [17], [16] generated the test problem instances by varying the distribution used for the coefficients.

Moser et al. [15] developed a heuristic algorithm using Lagrange multipliers to solve the MMKP. Akbar et al. [1] present two heuristic algorithms for solving the MMKP. Their initial heuristic (M-HEU), uses both improving and non-improving item selections. Extending M-HEU to an incremental method yields I-HEU when the number of classes in the MMKP is large. Khan et al. [12] develop a heuristic approach (HEU) that uses the concept of an aggregate resource for selecting items for the knapsack. They improve the solution using an item exchange approach. Moser et al. [15] generated their test sets by considering MDKP with profits and weights generated from a uniform distribution and divided among non-empty classes. They varied the number of classes and knapsacks. Both [1] and [12] randomly generated their test problems from a uniform distribution by varying the number of classes, items, and knapsacks.

Figures 1, 2, and 3 respectively summarize the design of various studies of the MKP, MCKP, and MMKP heuristics involving test problem generation. The Factors column is used to indicate the factors considered while generating the test problems and the Measures column indicates the performance measures used to compare the performance of the developed heuristics with other heuristics. These figures follow the format employed in Hill and Reilly [9] in their MDKP study.

Interestingly few of the studies cited in Figures 1-3 examine correlation structure within the test problems. This means any relationship among test problem coefficient sets is due to sampling error. As problems get larger, and seemingly more difficult, the correlation values converge to independence (which is generally an easier type of problem). When correlation is considered, the correlation ranges induced are not representative of the entire range of problem structure instances.

Author	Factors					Measures			
	m	n	D	S	Σ	Tm	Acc/Err	OpS	Iter
Hung and Fisk (1978)	x	x	x	x		x		x	
Martello and Toth (1985)	x	x	x			x			
Pisinger (1999)	x	x		x	x	x			

m= number of knapsacks
n= number of items
D= Distribution of constraint coefficients
S= Slackness of constraints
Σ=Correlation induced between problem coefficients
Tm= CPU time taken to solve the problem
Acc/Err= Accuracy or error between heuristic solution value and optimal solution value
OpS= number of problems solved to optimality
Iter= number of iterations

Figure 1. Factors and Measures used in Empirical Analysis of the MKP Heuristics

Author	Factors					Measures			
	m	n_i	D	S	Σ	Tm	Acc/Err	OpS	Iter
Sinha and Zoltners (1979)	x	x	x		x				
Armstrong et al. (1983)	x	x	x		x				
Pisinger (1995)	x	x				x	x		
Kozanidis et al. (2005)	x	x				x			

m= number of classes
n= number of items in each class
D= Distribution of constraint coefficients
S= Slackness of constraints
Σ=Correlation induced between problem coefficients
Tm= CPU time taken to solve the problem
Acc/Err= Accuracy or error between heuristic solution value and optimal solution value
OpS= number of problems solved to optimality
Iter= number of iterations

Figure 2. Factors and Measures used in Empirical Analysis of the MCKP Heuristics

3 TEST PROBLEM CHARACTERISTICS

Problem generation approaches should systematically or randomly vary problem attributes. These attributes should include the number of variables, number of constraints, marginal distributions of objective and constraint coefficients, method of setting right-hand side values, correlation structure between the objective and every constraint, correlation structure between constraint coefficients,

Author	Factors						Measures			
	l	m	n	D	S	Σ	Tm	Acc/Err	OpS	Iter
Moser et al. (1997)	x	x	x				x	x	x	
Akbar et al. (2001)	x	x	x	x		x	x	x	x	
Khan et al. (2002)	x	x	x			x	x	x		

l= number of knapsacks

m= number of classes

n= number of items in each class

D= Distribution of constraint coefficients

S= Slackness of constraints

Σ=Correlation induced between problem coefficients

Tm= CPU time taken to solve the problem

Acc/Err= Accuracy or error between heuristic solution value and optimal solution value

OpS= number of problems solved to optimality

Iter= number of iterations

Figure 3. Factors and Measures used in Empirical Analysis of the MMKP Heuristics

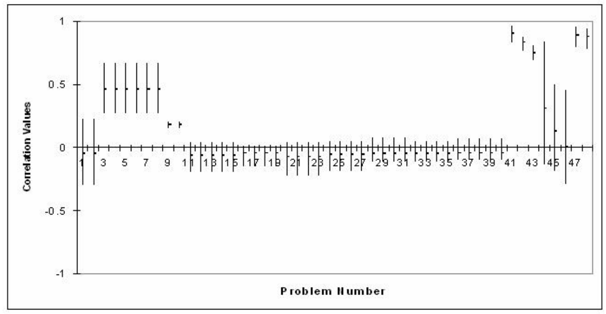

Figure 4. Range of Correlation Values Between Objective Function and Constraint Coefficients for MDKP Standard Test Problems

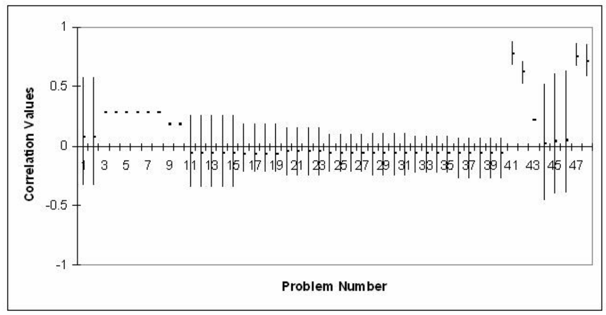

Figure 5. Range of Correlation Values Between Constraint Coefficients for MDKP Standard Test Problems

slackness ratios [8].

Test problems used for empirical examination of heuristic algorithms are either from libraries of standard problems or are synthetically generated. The standard test problem sets available at [3] lack sufficient diversity particularly in the correlation structure and the constraint slackness settings. Using test problems that provide an insufficient breadth of diversity leads to questionable heuristic performance generalization, particularly since the structure of industrial problems are ill-defined [4].

3.1 Problem Structure Analysis of Test Problems

3.1.1 Structure of MDKP Test Problems

On his website, Beasley [3] provides a set of 48 test problems for the MDKP extracted from the literature. Correlation values between the objective function and the constraint coefficients and the interconstraint correlation are calculated and plotted in Figures 4 and 5, respectively, for each of the 48 problems. The range of problem structure is very limited. Such limited range of problem instances lead to tenuous claims of general heuristic performance [8]. Further, the constraints have similar ratios of coefficients to right-hand side values.

3.1.2 Structure of MMKP Test Problems

Khan et al. [12] provide standard test problems available at [7]. The data available for each of the test problem instances include the number of resources, number of classes, number of items in each of the classes, number of constraints, resource capacities, solution values by exact, [15] and [12] heuristic approaches. Figures 6 and 7 plot the analysis of the test problems.

There are 13 test problems provided by [12] in 13 files from I01 to I13. Figure 6 displays the range of objective function to constraint coefficient values for the 13 test problems. The values are all positive, ignoring negative correlations and the range decreases as problem size increases; not good coverage of the range of correlation structures.

Figure 7 shows the range of interconstraint coefficient values for the same 13 test problems. Once again the range of values is not large in fact the values are mostly the same for each problem.

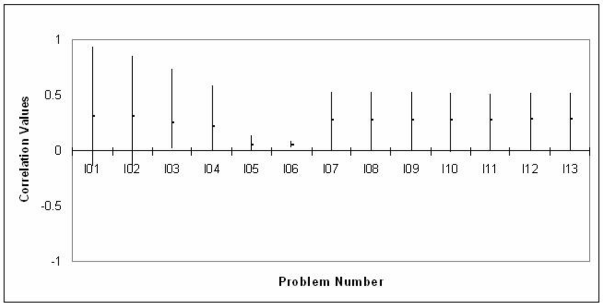

Figure 6. Range of Correlation Values Between Objective Function and Constraint Coefficients for MMKP Standard Test Problems

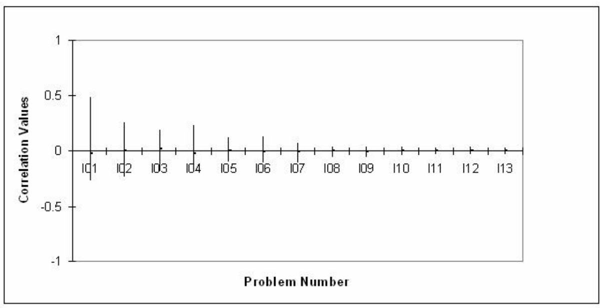

Figure 7. Range of Correlation Values Between Constraint Coefficients for MMKP Standard Test Problems

4 IMPLICATIONS OF POOR TEST PROBLEM SETS

Hooker [10] and Rardin and Uzsoy [18] are essential reads for anyone considering a heuristic empirical study. Test sets serve two purposes. One purpose is to promote heuristic performance insight. Test sets for insight should adhere to the principles provided in [18]. However, as our research shows, current problem sets lack the range of characteristics called for by an experimental design; this means insights gleaned may be lacking.

The second purpose of test sets are for competitive testing. Since we generally lack knowledge of the general structure of actual problems, we want competitive test problems to be random and sufficiently cover the full range of problem instances. Few such problem sets exist; the exception is [5]. The problem set in [5] generates 9 problem sets of 30 problems each for each combination of 50, 100, and 250 decision variables and 5, 10, and 25 problem constraints. Each constraint slackness ratio is randomly drawn from the continuous range $[0.20, 0.80]$. Each objective function coefficient set to constraint coefficient set correlation value is randomly drawn from the continuous range $[-0.90, 0.90]$, with inter-constraint coefficient set correlations set to ensure feasibility. Using randomly determined coefficient distribution functions, the random sample of problem coefficients is generated to attain the target (random) correlation structure. This final test set represents a truly random, sufficiently broad set of test problems appropriate for competitive testing.

Table 1. Comparing Heuristic Performance Against Standard and New Problems Using Percentage from Optimal

Standard Problem File	Toy	Koch	New Problem Set Size	Toy	Koch
mknapcb1	2.81	0.97	50-5	4.27	2.22
mknapcb2	2.09	0.44	100-5	5.51	1.77
mknapcb3	1.47	0.21	250-5	5.84	0.89
mknapcb4	3.89	1.81	50-10	6.55	3.40
mknapcb5	2.71	0.81	100-10	7.57	2.66
mknapcb6	1.91	0.32	250-10	10.46	2.45
mknapcb7	4.87	2.25	50-25	9.15	6.84
mknapcb8	3.74	1.39	100-25	10.52	5.75
mknapcb9	3.46	1.14	250-50	13.24	5.02

Two well established MDKP heuristics are given in [20] (Toy) and [13] (Koch). These are generally “better” performing heuristics. Table 1 depicts their performance on both the current standard MDKP problems and the new problems from [5]. Note the consistent performance against the standard problems but increasingly degraded performance as problem size increases against the new problems. This result has not been previously observed, begs the question of why does this occur, leads to research to overcome this behavior, and questions the generality of heuristic performance to real-world instances.

5 CONCLUSIONS

This paper reviews the various types of KP variants specifically examining the structure of the test problems employed during empirical testing. Our findings indicate these problems are not particularly broad in range. The problem with an insufficient set of problems was demonstrated using two well-established MDKP heuristics against a standard problem set and a newly generated robust set of problems. Our results demonstrate our concern regarding current empirical testing results. Future work will extend our problem generation methods and heuristic performance insight experimentation to the range of KP variants.

REFERENCES

[1] M. M. Akbar, O. Ergun, and A. P Punnen. Heuristic solutions for the multiple-choice multi-dimensional knapsack problem. In *International Conference on Computer Science*, San Francisco, USA, 2001.

[2] R. D. Armstrong, P. Sinha, and A. A Zoltners. The multiple-choice nested knapsack model. *Management Science*, 28(1):34–43, 1982.

[3] J. E Beasley. Problems available at the website. http://people.brunel.ac.uk/ mastjjb/jeb/info.html, 2006.

[4] N. S. Cherbaka, R. D. Mueller, and K. P. Ellis. Multidimensional knapsack problems and their application to solving manufacturing insourcing problems. In *Proceedings of the Annual Industrial Engineering Research Conference*, 2004.

[5] Y. K Cho. *Developing New Multidimensional Knapsack Heuristics Based on Empirical Analysis of Legacy Heuristics*. AFIT/DS/ENS/05-01, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, 2005.

[6] M. E Dyer, N Kayal, and J Walker. A branch and bound algorithm for solving the multiple-choice knapsck problem. *Journal of Computational and Applied Mathematics*, 11:231–249, 1984.

[7] M Hifi. Mmkp problems available at the website. http://www.laria.u-picardie.fr/hifi/OR-Benchmark/MMKP/, 2006.

[8] R. R Hill. An analytical comparison of optimization problem generation methodologies. In D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan eds, editors, *Proceedings of the 1998 Winter Simulation Conference*, pages 609–615, 1998.

[9] R. R Hill and C. H Reilly. The effects of coefficient correlation structure in twodimensional knapsack problems on solution procedure performance. *Management Science*, 46(2):302–317, 2000.

[10] J. N Hooker. An empirical science of algorithms. *Operations Research*, 42(2):201–212, 1994.

[11] M. S Hung and J. C Fisk. An algorithm for zero-one multiple knapsack problems. *Naval Research Logistics Quarterly*, 24:571–579, 1978.

[12] S. Khan, K. F. Li, E. G. Manning, and M Akbar. Solving the knapsack problem for adaptive multimedia systems. *Studia Informatica, Special Issue on Combinatorial Problems*, 2(2):157–178, 2002.

[13] G. Kochenberger, B. McCarl, and F. Wyman. A heuristic for general integer programming. *Decision Sciences*, 5(1):36–44, 1974.

[14] S. Martello and P Toth. Algorithm 632 a program for the 0-1 multiple knapsack problem. *ACM Transactions on Mathematical Systems*, 11(2):135–140, 1985.

[15] M. Moser, D. P. Jokanovic, and N Shiratori. An algorithm for the multidimensional multiple-choice knapsack problem. *IEICE Transactions in Fundamentals*, E80-A(3):582–589, 1997.

[16] D Pisinger. A minimal algorithm for the multiple-choice knapsack problem. *European Journal of Operational Research*, 83:394–410, 1995.

[17] D Pisinger. An exact algorithm for large multiple knapsack problems. *European Journal of Operational Research*, 114:528–541, 1999.

[18] R. L Rardin and R Uzsoy. Experimental evaluation of heuristic optimization algorithms: A tutorial. *Journal of Heuristics*, 7(3):261–304, 2001.

[19] P. Sinha and A. A Zoltners. The multiple-choice knapsack problem. *Operations Research*, 27(3):503–515, 1979.

[20] Y Toyoda. A simplified algorithm for obtaining approximate solutions to zero-one programming problems. *Management Science*, 21(12):1417–1427, 1975.